Data Appendix
CCC Data

Contents

1. Overview of data sources 2
   a. CCC Archives 2
   b. CCC Camps 2
   c. Colorado Name Index 3
   d. FamilySearch (BYU Record Linking Lab) 3
2. CCC Regulations 4
3. Assignment of individual ids for multiple records 5
4. Imputing camp numbers for CO data 6
5. Construction of camp location and characteristics from historical records (incomplete) 6
6. Matching Individuals to 1940 Census and WWII Enlistment Records 8
   a. Introduction to Matching Approach 8
   b. Overview of Matching Procedure 9
   c. Implementation 11
   d. Census Matching Results 12
   e. WW2 Matching Results 15
7. Imputing Hispanic Origin 19
8. Imputing Probability of Survival 19
9. Calculation of Marginal Value of Public Funds 20
10. Instrumental Variables Approach 21
11. Special Acknowledgements 25
12. Additional Tables and Figures 26

Data Appendix Table 1: Comparison of Counties of Enrollees vs Whole State 26
Data Appendix Table 2: Instrumental Variables Results 27
Data Appendix Figure 1: History of CCC program 28
Data Appendix Figure 2: Example Colorado enrollment record 29
Data Appendix Figure 3: Example New Mexico discharge record 31
Data Appendix Figure 4: CO and NM data completeness 32
Data Appendix Figure 5: Cohort eligibility and participation in CCC 34
Data Appendix Figure 6: CCC enrollees in CO and MN are more disadvantaged than enrollees nationwide 35
Data Appendix Figure 7: Histogram of Instrument and First Stage 38

1. Overview of data sources

Data used for CCC is assembled from various sources. The major sources of data are:

1) **Archival documents** that include application and discharge forms newly digitized by us and various information about CCC camps of primarily New Mexico and Colorado.
2) **FamilySearch / Ancestry.com** data that links the individuals found in the archival files to various historical sources available online from familysearch.com and ancestry.com, assembled by the BYU Record Linking Lab
3) **Social Security Administration Death Master File** data where we use the SSN, death date, and birth dates found in (2) to link people to correct identifiers

These sources are combined to create the final record-level data. Because some records in the archive belong to the same individual, the record-level data contain more observations than the number of individuals. We tag records so that records belonging to the same individual are assigned the same *PersonID*. We detail the procedure in Section 6.

We use the person-level data and add in additional sources of data to complete the Analysis Sample. The records we link are:

1) **1940 Census** that we machine-match for demographic and family characteristic variables
2) **WWII enlistment records** that we machine-match for demographic variables

The individuals in the Analysis Sample are uniquely identified by variables *state* (of enrollment) and *PersonID*. This is the final dataset used for analysis.

More details on each section:

a. CCC Archives

Colorado (CO)
The Colorado data is from transcriptions of following records: (i) *Certificate of Selection for the Civilian Conservation Corps*, (ii) *Application for the Enrollment*, (iii) *Discharge Form* (Unofficial name). The records are found in the Colorado State Archive under the title "Civilian Conservation Corps Enrollments (Statewide) 1936-1942."

New Mexico (NM)
The New Mexico data is from transcriptions of *Civilian Conservation Corps, New Mexico District records.* (Citation number: collection 1959-030)

b. CCC Camps

The opening and closing dates of CO camps come from Robert W. Audretsch, who supplied us with a list of camps, their associated companies, and the beginning and start dates of the company numbers within the camps.

The CO camp location comes from various historical records that we hand-coded.

The camp type code information comes from http://www.ccclegacy.org/CCC_Camp_Lists.html.

c. Colorado Name Index

Colorado Name Index contains information on a subset of enrollees and their camp assignment that was retrieved from searching through mentions of enrollees' names in contemporary local newspaper articles. Local newspapers often announced young men in their area who enrolled in the CCC and contained basic information about their enrollment. We have used this information to impute camp numbers in cases we were missing them. The procedure is detailed in Section 4.

The *Colorado Name Index* is from the following book:

A Colorado Civilian Conservation Corps Enrollee Name Index
by Robert W. Audretsch
Publisher: CreateSpace Independent Publishing Platform; 1 edition (April 5, 2017)
ISBN-10: 1545102910
ISBN-13: 978-1545102916
Amazon link: https://www.amazon.com/Colorado-Civilian-Conservation-Corps-Enrollee/dp/1545102910

d. FamilySearch (BYU Record Linking Lab)

After the records from the state archives were transcribed and cleaned, individuals in the data were sent to the BYU Record Linking Lab to be found in various historical genealogy websites including Ancestry.com and FamilySearch.org. Their date of death and social security numbers were collected. The individuals' names, date of birth, place of birth, allottee (usually a family member) names were used to find these individuals. The match is performed by trained historians, using records from multiple data sources and information from CCC.

The BYU Record Linking Lab found two major variables:

i. SSN

Social security numbers were mostly found on Ancestry.com. The sources of the SSNs on Ancestry are:

1) Ancestry.com. *U.S., Social Security Death Index, 1935-2014* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2011. Original data: Social Security Administration. *Social Security Death Index, Master File*. Social Security Administration.

2) Ancestry.com. *U.S., Social Security Applications and Claims Index, 1936-2007* [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2015. Original data: Social Security Applications and Claims, 1936-2007.

Note: SSN is only available for those who have been dead for 10 years. Therefore, we cannot find SSN for those who died before 2005/2006.

For reference, see:
SSDI: http://search.ancestry.com/search/db.aspx?dbid=3693
SSACI: http://search.ancestry.com/search/db.aspx?dbid=60901

ii. Death Dates

Death dates were found using various sources including the aforementioned social security administration data, Find A Grave Index, and other sources.

1) Ancestry.com. *U.S., Social Security Death Index, 1935-2014* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2011. Original data: Social Security Administration. *Social Security Death Index, Master File*. Social Security Administration.

2) Ancestry.com. *U.S., Find A Grave Index, 1600s-Current* [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2012. Original data: *Find A Grave*. Find A Grave. http://www.findagrave.com/cgi-bin/fg.cgi.

2. CCC Regulations

The rules and regulations regarding the operation of CCC camps as well as allotment of funds to CCC employees changed from the program's inception in 1933 to its closure in 1942. Below is a compilation of CCC regulations that are pertinent to our research.

1. Each employee of the CCC was given a serial number which was composed in the following:
    a. Serial numbers started with the letters "CC" to denote the Civilian Conservation Corps as opposed to other emergency relief programs. The letters "CC" were followed by the number of the area corps number. In the case of Colorado and New Mexico, the area number was 8. See the map below (source: National Parks Service. https://www.nps.gov/parkhistory/online_books/ccc/ccc/chap2.htm).

   b. Serial numbers then contain information on the company number. "In order that the numerical designation of the company may indicate its origin by corps area, blocks of numbers are assigned in accordance with the following system: 100-199 to First Corps Area, 201-299 to Second Corps Area, 901-999 to Ninth Corps Area.. When this series becomes exhausted, 1,000 will be added to each block of numbers; e.g. 1101-1199 to First Corps Area, 1201-1299 to Second Corps Area, and so on" (quote found here: https://babel.hathitrust.org/cgi/pt?id=mdp.39015020215433;view=1up;seq=17).
2. Allocation of funds received by CCC employees:
   a. Our data show that there was variation in the amounts received by CCC employees. This is consistent with the regulations found here: https://babel.hathitrust.org/cgi/pt?id=mdp.39015020215433;view=1up;seq=25 In particular, enrollees without special status (such as leaders or assistant leaders) were paid $30 per month. Of the $30 received, enrollees were required to pay at least $22 to their families.
3. Enrollment over time:
   a. CCC enrollment: Our data mostly contain information for those who enrolled after 1937. The most likely reason for this is that the CCC changed from being a program that was part of the Emergency Conservation Work program to its own entity known as the Civilian Conservation Corps in 1937. See quote here: "There are hereby transferred to the Corps all enrolled personnel, records, papers, property, funds, and obligations of the Emergency Conservation Work established under the Act of March 31, 1933 (48 Stat. 22), as amended; and the Corps shall take over the institution of the camp exchange heretofore established and maintained, under supervision of the War Department, in connection with and aiding in administration of Civilian Conservation Corps work camps conducted under the authority of said Act as amended: Provided, That such camp exchange shall not sell to persons not connected with the operation of the Civilian Conservation Corps" (source here: https://www.nps.gov/parkhistory/online_books/ccc/cccaa.htm)

3. Assignment of individual ids for multiple records

Individuals can generate multiple records in the CCC record-keeping system. For example, a person who enrolled twice could generate two records: one enrollment form for each time he enrolled. Because our raw data consists of records of enrollment and discharge, our raw data is in the record-level, not in the individual-level. We convert the record-level raw data into an individual-level data by using the information in the records to assign records to unique individuals.

We use the following information in each record to determine whether records belong to the same individual: enrollee's first and last names, birth dates, CCC serial number, social security number (if available in the original records for CO), allottee's first and last names, and allottee's relation to the participant. All of these fields in each record are subject to transcription and record-keeping errors. In addition, SSN data is only sparsely available for CO enrollees.

Therefore, we first use a "fuzzy" matching algorithm for each record to group records with similar field values. Then, we verify the matches manually. Additional information from the BYU Record Linking Lab allowed them to tag more records as coming from the same individuals.

Records vs Individuals Statistics

	CO	NM
Number of Records	21,538	10,713
Number of Individuals	18,644	9,699
Number of Individuals with…		
- 1 record	16,082	8,746
- 2 records	2,263	894
- 3 records	269	57
- 4 records	27	2
- 5 records	3	0

4. Imputing camp numbers for CO data

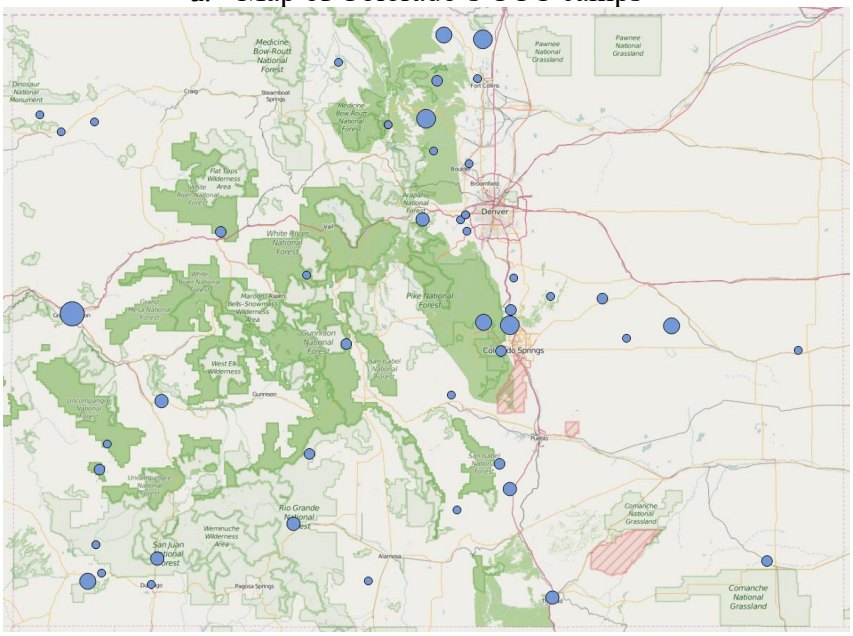
We have used various sources to impute camp numbers for individuals that do not have camp information in the CO data.

1) **Company Numbers**: For some enrollees, we have company numbers but not camp numbers. The correspondence between company and camp numbers were obtained from Robert W. Audretsch, who documented the company number assigned to specific camps over time.
2) **CCC Serial Numbers**: Each enrollee was assigned a serial number when they first enrolled. The serial number contains the area of enrollment (as described in Section 2) and the company number they were assigned to. The company numbers were then used to impute the camp of assignment.
3) **Colorado Name Index:** For enrollees with enrollment date information but no camp information (either directly from the records or that could be imputed from the serial numbers), we supplemented the camp information through the *Colorado Name Index*. As described in Section 1, the Index contains information from local newspapers on enrollees and their camps at a point in time (when the article was published). We used enrollees' first and last names, place of birth or place of enrollment application, and their enrollment and discharge date to manually match the enrollee to a newspaper record in the Index. Then, we assigned the camp information from the Index as the enrollee's first camp of assignment.

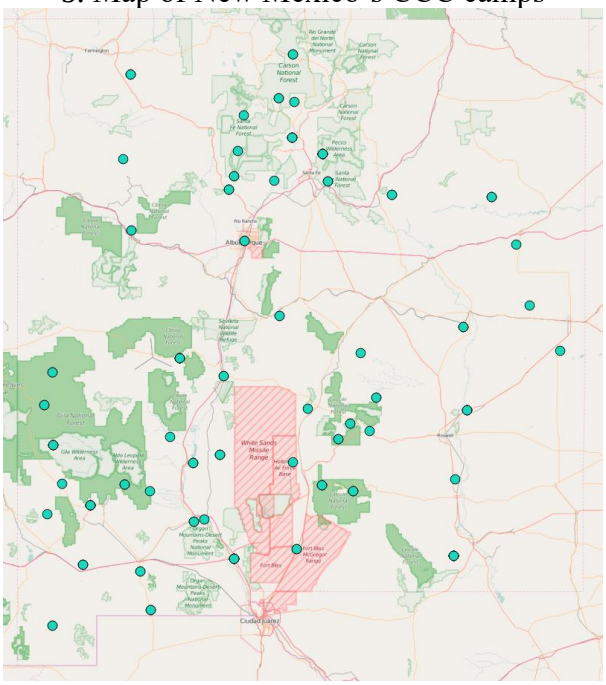
5. Construction of camp location and characteristics from historical records (incomplete)

Camp ID in administrative records merged with camp-information from multiple sources (see Appendix D). Dates of operation of camp were obtained from Robert W. Audretsch. Camp location was approximated by location descriptions in historical documents.

a. Map of Colorado's CCC camps



b. Map of New Mexico's CCC camps



Distance to closest town was computed taking the list of Colorado and New Mexico towns and their latitude and longitude from United States Geological Survey's Geographic Names

Information System (USGS GNIS). Pairwise distances from each camp to each city was calculated, then for each camp, the town with the smallest distance value was selected as the distance to closest town.

Camp weather information was obtained from historical weather data at the PRISM Climate Group at Oregon State University. The data contains minimum and maximum temperature and precipitation at the monthly level and covers the entire United States from 1985-1980 at the spatial scale of 4km x 4km. It is a climatologically aided interpolation and takes as first guess the long-term averages in the area. For more information, visit the PRISM website at http://prism.oregonstate.edu/historical/. We obtain the historical monthly weather data for each camp from the GIS raster files using camp location (longitude and latitude).

Camp peer characteristics are computed using information of individuals at each point in time in our dataset. The peer characteristics for enrollee i is the weighted average of demographic characteristics of other enrollees in our data who served in the same camp overlapped in service duration with i, where the days of overlap are used as the weights. Thus, enrollees that overlapped for a longer period of time get higher weights in the peer characteristics calculation.

In other words, the peer characteristics PX_i of enrollee i is calculated by,

$$PX_i = \sum_{j \in K_i} \frac{d_{ij}}{\sum_{j \in K_i} d_{ij}} x_j$$

where K_i is the set of enrollees that overlap with i, d_{ij} is the days of overlap between i and j, x_j is the demographic characteristic of j.

6. Matching Individuals to 1940 Census and WWII Enlistment Records

This appendix overviews the matching approach used to match CCC participants to Census and WW2 Army enlistment records. We rely on the Expectation Maximization approach to match records. Overall, the match rates are consistent with standard literature and the matches seem consistent. There seems to be some selection in terms of who is matched.

a. Introduction to Matching Approach

The matching approach follows "Linking Individuals Across Historical Sources: a Fully Automated Approach" by Ran Abramitzky, Roy Mill, and Santiago Perez (2018).[1] Any matching approach has to balance three competing goals:

1. Minimize false negatives (Type II errors)
2. Minimize false positives (Type I errors)
3. Create a representative sample

Ideally records would be identified by a unique administrative identifier that is stable across datasets (e.g., social security number). In most historical cases, we are forced to rely on a combination of less definitive information, such as year of birth, name, place of birth, and place

[1] Please see this article for a more detailed description of the approach

of residence to match records. Therefore, choosing how to match on these characteristics is a major decision. There are three major sources of variation in variables across records for a given individual. First, the respondent introduces variation. They could state the wrong age or change their name (e.g., ``Nick" instead of ``Nicholas"). This issue is especially prevalent in historical Censuses due to lower literacy and education levels. Secondly, the interviewer can make transcription errors (e.g., write the name as ``Brian" instead of ``Ryan"). Finally, additional errors are introduced during the digitization of physical Census rolls.

We choose to rely on the Expectation Maximization (EM) approach outlined in Abramitzky et al. (2018). Individuals are matched to 1940 Census and WWII enlistment records primarily using automated methods. One alternative approach would be to rely on exact matches but relying solely on exact matches would significantly lower match rates and increase Type II errors. There are significant transcription errors in these records and the EM approach allows some flexibility when dealing with errors.

The EM approach falls under the umbrella of automated methods. The advantages of automated methods include the fact that they are reproducible, rule-based, can compare all records, and are cheaper. The disadvantages are that they do not have the same contextual information that humans do (e.g., "Bill" is short for "William") and humans are better able to incorporate additional information in a flexible manner.

Bailey et al. (2018) raised substantive concerns about using automated methods as opposed to linking by hand. They find that automated linking algorithms produce high rates of incorrect matches ranging from 13 to 69 percent when assuming hand-linked sample is the ground truth. Match rates are especially poor when automated methods are combined with phonetic name cleaning. They tested three automated methods, Ferrie (1996), iterative method of Abramitzky et al. (2012 and 2014), and the regression prediction approach of Feigenbaum (2016), though not the EM approach. These results are an issue because poor matches can significantly attenuate estimates.

Abramitzky et al. (2018) find much better results for more modern automated methods, such as the EM approach, than the approaches tested in Bailey et al. (2018). Additionally, they find that automated methods perform similarly to hand-linking methods when the same information is used. Conservative EM methods tested by Abramitzky et al. had <10% false match rate, which was lower than hand-linking methods with the same information, though hand-linking methods also made significantly more matches. Moreover, when both methods made a match then there was greater than 90% agreement.

In order to address concerns of false matches we rely on conservative matching criteria and do not conduct phonetic cleaning or significant name standardization. Finally, we validate a subset of matches against hand matches provided by Family Search (FS).

b. Overview of Matching Procedure

There are several decisions to make before beginning any estimation. The first decision is which variables to match on. The standard approach is to match on pre-determined characteristics. Typically, this means birth year, place of birth, first name and last name.

The second decision is which variables to block on. The approach will only compute distance between individuals who are exact matches on certain characteristics. Fundamentally, blocking is used to reduce computational complexity by avoiding computing distances between every potential pair of individuals. For example, it is common to block on the first letter of the first name.

The third decision is how to measure string distance. Some approaches effectively use an indicator for whether names are an exact match or they combine this approach with a phonetic cleaning algorithm, such as the NYSIIS. Phonetic cleaning is especially useful if most errors are due to translating a heard name to a written one. Continuous string distance measures can also be used and are most useful when errors are due to transcription mistakes during digitization.

Now, we present the basic concept behind the Expectation Maximization algorithm (Dempster, Laird, and Rubin 1977; Winkler 1989). For any observation, there are many match candidate pairs, *i*. For each candidate pair we observe distances γ_i. Assume each of these candidates are drawn from one of two distributions. Each candidate pair has two associated probabilities: one for true matches, $P(\gamma_i|Match_i)$, and one for false matches, $P(\gamma_i|NotMatch_i)$. Using Bayes Rule, the probability that our candidate pair *i*, with distance γ_i, is a true match is given by:

$$P(Match|\gamma_i) = \frac{P(\gamma_i|Match_i)}{P(\gamma_i|Match_i)p_{m+}\ P(\gamma_i|NotMatch_i)(1-p_m)}$$

Using these expressions, we can take the following approach to estimate match probabilities:

1. Define distribution families for each of the distance variables to get $P(\gamma_i|Match_i)$ and $P(\gamma_i|NotMatch_i)$. Assume distances for each variable are independently distributed conditional on match status
2. Guess initial parameter values $\theta_m^{(t)}, \theta_{nm}^{(t)}$ for each distribution and the probability of a true match, $p_m^{(t)}$
3. Loop over the following two steps until convergence:
    a. Calculate for each pair the probability of a match, $w_i^{(t)} = P(Match|\gamma_i)$ for a given $(\theta_m^{(t)}, \theta_{nm}^{(t)}, p_m^{(t)})$
    b. Get updated parameter estimates $(\hat{\theta}_m^{(t+1)}, \hat{\theta}_{nm}^{(t+1)}, \hat{p}_m^{(t+1)})$ by maximizing:

$$\log L(\gamma, \theta, p_m) = \sum_{i=1}^{n} w_i^{(t)} \log p_m P(\gamma_i|\theta_m) + \left(1 - w_i^{(t)}\right) \log(1-p_m) P(\gamma_i|\theta_{nm})$$

Once we have the converged estimates then we can compute $P(\gamma_i|Match_i)$ for any candidate pair. The final major choice is choosing what qualifies as a match. There are two components to this decision:

1. The minimum threshold in order to qualify as a match
2. The maximum threshold for the second closest match

(1) means that if there are no ``good" matches then it is better not to declare any a match. (2) means that if there are at least two ``good" candidates then there is a high Type II error rate when

selecting one over the other. For the primary analysis, we take a conservative approach, setting a high threshold for (1) and (2).[2]

c. Implementation

One significant issue is that New Mexico CCC records do not contain data on the birthplace of participants. When matching to the 1940 Census and WW2 records we rely on a two step procedure to create matches:

- **First stage:** Colorado and New Mexico CCC participants are matched to 1940 Census and WW2 enlistment records
  - *Blocking variables:* State of residence, first letter of first name and first letter of last name
  - *Matching variables:* Year of birth and name distances
- **Second stage:** Next, we remove matched individuals and for unmatched individuals in the Colorado CCC we conduct a second round of matching
  - *Blocking variables:* Place of birth, first letter of first name and last name
  - *Matching variables:* Year of birth and name distances

In the first stage we look only within the current state of residence (e.g., only look at residents of Colorado in the 1940 Census for CO CCC participants). In the second stage, we use the additional information on place of birth for CO CCC participants to search across the United States.

The primary concern with using the state of residence is that we will miss migrants. There are two reasons that this should not be a major issue in our case. First, the 1940 Census, most CCC enlistment, and most WW2 enlistment take place in a relatively short time frame. Secondly, we can check the number of migrants in the Family Search hand-links. For both CO (91.4%) and NM (96.8%) most of the CCC participants are still in the same state during the 1940 Census. For New Mexico it seems very reasonable to only look within the state. The percentage is somewhat lower for Colorado, which is why we conduct the second stage and also match on place of birth.

Next, we decide to use the Jaro-Winkler string distance (Jaro 1989, Winkler 2006). The Jaro-Winkler string distance calculates the number of transpositions required to match two strings, weighting errors in the early part of the string more heavily. The distance is measured from 0 (no matching characters) to 1 (exact match). We invert this scale so that 0 is exact match and 1 is no matching characters so our measure is increasing in distance. In our case the largest concern is transcription errors during digitization so it makes sense to use a string distance measure.

The next choice is the creation of distributions for distance variables. We follow Abramitizky et al. (2018) and specify multinomial distributions for year of birth and name distances. Year of birth distances are segmented into groupings of 0, 1, or 2 years distance.[3] Name distances are segmented based on Jaro-Winkler scores into groupings: [0,0.067], (0.067,0.120],(0.120,0.250],(0.250,1]. These groups run from closest to farthest distance.

[2] The threshold for (1) is 0.8 and the minimum distance for the second best match (2) is 0.3
[3] Matches with larger distances are not considered

We also add in the hand-matches from Family Search. If the Family Search matches conflict with the automated methods then we use the Family Search match. Finally, we also conduct a tie-breaking procedure using additional information in cases where the best match clears the minimum threshold but the second best match is too close. If the first best match passes the tiebreak criteria and second best match fails then we count it as a match. Middle initial is used as a tiebreaker in both stages, while place of birth is used as a tiebreaker in the first stage for Colorado. For example, if the CCC record has middle initial "F", the first best match also has the middle initial "F" but the second best match has the middle initial "M" then it is counted as a match.

d. Census Matching Results

Matching Appendix Table 7-1: Match rates between CCC records and 1940 Census

Census Match Rates by Type	CO	NM	Overall
EM and FS	0.08	0.06	0.07
Only EM	0.34	0.22	0.30
Only FS	0.05	0.09	0.07
No match	0.53	0.63	0.56
Observations	18644	9699	28343

Note: Values represent match rates as percentages of column totals. Match rates are for CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by FamilySearch team

False Negatives (Type II error): Matching Appendix Table 7-1 shows that 44% of CCC participants have been matched to 1940 Census records. 30% of participants have been matched through EM only, 7% through FS only, and 7% through both methodologies. This match rate for the EM approach is in line with the literature. Additionally, there is an upper bound on potential matches. In order to find this upper bound for matches to the 1940 Census, Abramitzky et al. (2018) linked a copy of the 1940 Census digitized by Family Search and one digitized by Ancestry.com. Even in this case they can only link up to 67% of the Census due to individuals with similar attributes and ``brutally bad transcriptions" due to difficulties reading cursive.

Matching Appendix Table 7-2: Match consistency between EM and FS for CCC-1940 Census matches

Census Match Consistency	CO	NM	Overall
% of participants matched by both EM and FS	0.08	0.06	0.07
% of overlap matched to same individual	0.95	0.91	0.94
Observations	18644	9699	28343

Note: Values represent percentages of column totals. Match rates are for CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by Family-Search team. Consistent FS and EM match measures whether EM and FS approaches matched CCC participant to the same Census individual in cases when both approaches make a match

False Positives (Type I error): While we do not have an absolute ``ground truth" sample, one way of examining Type I errors is to see if the EM and FS approaches match the same individual when they overlap. As seen in Matching Appendix Table 7-2, there is a high degree of consistency when both methods made a match - 94% of the time they matched the same CCC participant to the same Census record. We can go a step further and examine the discrepancies to understand if there is a reason to prefer the EM approach or FS hand matches. We use additional information (e.g., county of residence) and classify the discrepancies. In about 1/3 of cases the EM match is preferred, in 1/3 of cases the FS match is preferred, and the remaining cases are indeterminate. Therefore, there does not seem to be a clear reason to prefer either method.

Representativeness: Finally, we check which individuals are matched by regressing an indicator of whether matched on CCC participant characteristics at the time of their first enrollment. If matches are at random then there should be no clear pattern.

Matching Appendix Table 7-3: Predictors of CCC-1940 Census matches by type of match for CO

	EM match	FS match	EM and FS	FS or EM match
Age at enrollment	0.00	-0.00**	-0.00	-0.00
	(1.00)	(0.01)	(0.19)	(0.29)
Age of death	0.00	-0.00*	-0.00**	0.00
	(0.15)	(0.04)	(0.01)	(0.13)
Enroll year	-0.00	-0.00	-0.00	-0.00
	(0.80)	(0.27)	(0.46)	(0.54)
Dist. to camp (mi)	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Born in CO	-0.01	0.00	-0.01	-0.00
	(0.14)	(1.00)	(0.09)	(0.60)
Height (in)	0.02*	0.01	0.01**	0.01*
	(0.01)	(0.08)	(0.01)	(0.03)
Weight (lb)	-0.00	-0.00	-0.00	-0.00
	(0.49)	(0.41)	(0.10)	(0.74)
BMI	0.01	0.01	0.01*	0.00
	(0.35)	(0.30)	(0.04)	(0.61)
Missing parent	-0.06***	-0.07***	-0.04***	-0.08***
	(0.00)	(0.00)	(0.00)	(0.00)
Farm	0.01	0.05***	0.03***	0.03
	(0.45)	(0.00)	(0.00)	(0.07)
Urban	-0.00	-0.02	-0.01	-0.01
	(0.76)	(0.18)	(0.15)	(0.65)
Years educ	0.01***	0.00*	0.00***	0.01***
	(0.00)	(0.02)	(0.00)	(0.00)
Unemployed	-0.02	-0.01	-0.02	-0.02
	(0.29)	(0.25)	(0.08)	(0.38)
Constant	1.17	5.37	2.47	4.07
	(0.87)	(0.29)	(0.54)	(0.58)
Observations	18644	18644	18644	18644
R^2	0.044	0.018	0.018	0.049

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for CO CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by FamilySearch team

Matching Appendix Table 7-3 shows the results for Colorado CCC participants broken out by match type. In general, CCC participants who were matched seem slightly better off. For example, matched individuals have higher education levels, less likely to be missing parents, and are taller on average. These differences do not seem to be large in absolute magnitude though, so it seems as though the matches are reasonably well representative.

Matching Appendix Table 7-4: Predictors of CCC-1940 Census matches by type of match for NM

	EM match	FS match	EM and FS	FS or EM match
Age at enrollment	-0.00	-0.00**	-0.00*	-0.00**
	(0.07)	(0.01)	(0.05)	(0.01)
Age of death	0.00	-0.00	-0.00	0.00
	(0.74)	(0.15)	(0.05)	(0.86)
Enroll year	-0.00	0.00	-0.00	-0.00
	(0.42)	(0.92)	(0.66)	(0.65)
Dist. to camp (mi)	-0.00	0.00	0.00	-0.00
	(0.20)	(0.35)	(0.48)	(0.40)
Constant	4.98	-0.24	1.47	3.27
	(0.39)	(0.96)	(0.63)	(0.60)
Observations	9699	9699	9699	9699
R^2	0.019	0.003	0.003	0.016

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for NM CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by FamilySearch team

Matching Appendix Table 7-4 shows the results for New Mexico matches. For New Mexico, we have significantly fewer indicators of participant characteristics; however, there again seems to not be large differences in terms of the type of individual matched.

e. WW2 Matching Results

Matching Appendix Table 7-5: Match rates between CCC records and WWII Enlistment records

WW2 Match Rates	CO	NM	Overall
EM match	0.31	0.24	0.29
EM match (Adj)	0.78	0.59	0.72
Observations	18644	9699	28343

Note: Values represent match rates percentages of column totals. Adjusted values are scaled by state-age cohort enlistment percentages

False Negatives (Type II error): Matching Appendix Table 7-5 shows that 29% of CCC participants are matched to WW2 army enlistment records. There are two primary reasons that this match rate is lower than the 1940 Census. First, there is no supplementary source of matches to augment the EM approach with (FS matches). Secondly, the Census has universal coverage while only a subset of men will be in the WW2 army enlistment records. We can compute an adjusted match rate by estimating the percentage of men in each state-year of birth cell that are in the records.[4] This procedure assumes CCC participants are no more likely to enlist than other of the same age in the same state. Based on these calculations we would expect 40% of the Colorado CCC participants and 41% of the New Mexico CCC participants to have be in the WW2 army enlistment records. The adjusted match rates (match percentage of those we expect to find) and is 78% for Colorado and 59% for New Mexico. Note that these adjusted match rates seem high but cannot account for whether CCC individuals were more likely to serve in the

[4] Using state-year of birth-years of education cells does not substantively alter the results

Army. For example, CCC camps typically involved significant Army administration which could increase the likelihood to serve due to familiarity with the military.

False Positives (Type I error): Without another source of matches for the WWII data it is difficult to conduct any sort of consistency analysis. Therefore, we rely on the findings of high consistency in the CCC to 1940 Census matches in order to support the EM approach in this case.

Representativeness: We repeat the regression of match status on characteristics, but the interpretation is slightly complicated in this case. There are two forms of selection: first, selection into who is drafted (and meets minimum standards) or enrolled in the Army, and secondly there is selection through who is matched.

Matching Appendix Table 7-6: Predictors of CCC-WWII enlistment matches by type of match for CO

	EM match
Age at enrollment	-0.01***
	(0.00)
Age of death	0.00
	(0.94)
Enroll year	0.01***
	(0.00)
Dist. to camp (mi)	-0.00*
	(0.01)
Born in CO	0.02**
	(0.00)
Height (in)	-0.01
	(0.15)
Weight (lb)	0.00
	(0.14)
BMI	-0.01
	(0.12)
Missing parent	-0.01
	(0.63)
Farm	0.02
	(0.22)
Urban	-0.02
	(0.19)
Years educ	0.01***
	(0.00)
Unemployed	0.01
	(0.68)
Constant	-26.86***
	(0.00)
Observations	18644
R^2	0.041

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for CO CCC participants to WW2 enlistment records. EM stands for Expectation Maximization approach

Matching Appendix Table 7-6 shows the results for Colorado CCC participants. Matched individuals are again better educated, but most indicators are not statistically significant. Matching Appendix Table 7-7 shows the results for New Mexico CCC participants.

Matching Appendix Table 7-7: Predictors of CCC-WWII enlistment matches by type of match for CO

	EM match
Age at enrollment	0.00**
	(0.00)
Age of death	-0.00
	(0.79)
Enroll year	-0.02***
	(0.00)
Dist. to camp (mi)	-0.00
	(0.06)
Constant	40.74***
	(0.00)
Observations	9699
R^2	0.035

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for NM CCC participants to WW2 enlistment records. EM stands for Expectation Maximization approach

7. Imputing Hispanic Origin

We follow the approach of Fryer and Levitt (2004) to construct a Hispanic name index for any first or last name using the 1940 Census. The name index is constructed using the Hispanic indicator variable in the 1940 Census. Each first and last name is given a value (0-1) based on:

$$HispIndex_i = \frac{\#\ of\ indiviuals\ with\ name\ who\ are\ Hispanic}{\#\ of\ individuals\ with\ name}$$

Individuals were not directly asked whether they are Hispanic during the Census until 1980 so an algorithm was used to classify individuals in prior Censuses retroactively. Eight rules were used, but at their most basic they are:

1. Individual or their parents/grandparents were born in a Hispanic area
2. Individual has a Spanish surname and was born in the US
3. Individual is a relative or spouse of someone who qualifies by (1) or (2)

Once the indexes are created, they are matched to CCC participants. There is an index for first name, last name, and a combined index created by combining them. Individuals above certain thresholds are classified as likely Hispanic.

8. Imputing Probability of Survival

Probability of survival of individuals are imputed in two ways. First, we can impute the probability as 0 for those with missing age of death (presuming that they are dead). Second, we can take a more sophisticated approach using the fact that the person was at least alive at the time of discharge and using the conditional probably of survival after having survived to age a_d at the time of discharge. This probability of survival uses information of survival probabilities from age a_d to a desired age threshold, e.g. $\bar{a}$ = 70. These rates can be obtained from the corresponding cohort life tables put out by the SSA (Bell and Miller 2005) for each enrollee's birth cohort, b.

We estimate survival models where we make various assumptions about the missing data. We concentrate on survival to age 70, which is slightly below the median age at death (73). Because the number 70 is a round multiple of ten, it avoids issues of age heaping. Appendix Table 4 shows the results. We start by estimating survival models using only the sample without missing data for reference (Panel A). Panel A shows the same basic patterns we found in Table 2: those who trained longer were more likely to survive and the estimates are very stable. In the last specification, the results imply that one more year of training increased the probability of survival to age 70 by about 4.6% relative to the mean. Panel B shows the results when we impute the probability of survival using life tables and information on the age at the time of training. Here, we find that the effect of training duration (once we add all controls) is somewhat lower (2.3. instead of 3 percentage points) but still statistically significant.

In Panel C, we impute all missing as zero (we assume that all the men for whom survival is missing died before age 70). The rationale for doing this is that the DMF and other sources of death tend to be complete starting in the 1970s (Hill and Rosenwaike, 2001). If most of the missing data is missing because of death certificates are not available to researchers (rather than due to errors in matching) then all the missing deaths occurred between the CCC training and 1970, much before our CCC men turned 70 (recall most of the men were born around 1920).

When we do this, we find that one more year of training is associated with about a 5% increase in survival relative to the mean.

9. Calculation of Marginal Value of Public Funds

We first calculate the cost of the program. The cost measure of MVPF incorporates both the direct cost to the government and various mitigation of cost. In particular, the CCC cost measure includes the following:

1. Upfront cost of the program
2. Increased social security payout from both the increase in pension amount and increase in longevity of enrollees
3. Cost mitigation from increased tax revenue from increased earnings of the enrollees
4. Goods produced during the program, namely conservation work

We get information on (1) from Levine 2010, who estimated the annual cost per enrollee to be $1,004. Assuming the figure is in 1939 dollars, using Consumer Price Index All Urban Consumers (CPI-U) January-to-January growth, that amounts to $14,384.81 for our average enrollee who served around 0.8 years (9.6 months).

For (2), we use the mortality profile from our regression results illustrated in Figure 6. We assume that enrollees survive to age 45 with probability 1. For each age $x > 45$, we take the average survival rate to age x of our regression sample to be the baseline survival rate, and the estimate of the coefficient on duration to be the increase in survival rate for an enrollee that served one year. Multiplying the estimate by the average duration gives us the increase in the rate of survival for our average enrollee to age x, for each age x from 46-90. We assume after age 90, the survival rate declines to 0 evenly until age 95.

The baseline person receives the average PIA amount of $437.70 per month, assumed to be in 1982 dollars, as 1982 is the year on which our average enrollee turns 62 when SSA starts calculating PIA using AIME. Converting that to an annual benefit amount in 2017 dollars gives us an annual benefit of $13,525.85. We assume that 65 is the claiming age for social security benefit. Multiplying i) the PIA with ii) the probability of survival to age x for each $x \geq 65$, iii) by the discount factor, and finally iv) summing the yearly amounts gives us the present value of the baseline social security benefit.

The average enrollee receives an extra $14.11 of PIA (Table 2, Column 6 and average duration), but also has increased survival rates to age x. Again, multiplying the baseline benefit by the total increased survival rate and by the discount rate gives the increase in the PV of benefits from increased rate of survival. Multiplying the total increased survival rate by the discount rate and the additional PIA amount gives the increase in the PV of benefits from increased PIA amount. Summing these two and subtracting it from the baseline PV of social security benefit gives us the final cost increase from increase in social security benefit over the lifetime. The final measure amounts to $2,514.17.

Calculating (3) is similar to the above, but instead of multiplying the PIA amount for ages above 65, we multiply the implied earnings from the PIA amount for ages below 65. We calculate the implied annual earnings to be the implied AIME increase multiplied by 12. The implied AIME increase is calculated by assuming that the PIA bend points used are the 1982 PIA bend points. This gives us the total PV of earnings and PV of earnings increase. We calculate the tax portion of this by assuming a tax rate of 33.6%, which is the CBO estimated average tax rate

for FPL 100-149% provided in Appendix G of Hendren and Spruce-Keyser (2019). The final measure comes out to be $4,846.28.

We abstract from (4), as we have no good estimate of the total value of conservation work provided by the program. Thus, our estimate could be thought of as an upper bound of the cost.

Now, on to the WTP (or value) of the program. CCC provided the following short- and long-term benefits to enrollees:

1. Willingness to pay (WTP) for increase in longevity
2. Increase in earnings
3. Monthly real wage of $66.25 while enrolled, which includes the benefits enrollees received during the program (BLS 1941)

Calculating (1) is again similar to the above cost calculation on increased social security payment. Instead of multiplying the PIA amount, we multiply the statistical value of life, assumed to be $150,000 in 2017 dollars (Lee et al 2009), for ages 45 to 95. We obtain an estimate of $25,456.40. For (2), since we already obtained the PV of earnings increase and the subsequent tax increase in calculating the cost, it is simply the after-tax portion of the PV of earnings calculated there. Therefore, we have $9,491.82 for the post-tax earnings benefit. (3) is straightforward, where we take the average amount enrollees received ($66.25 multiplied by average duration), which is $11,390.36 in 2017 dollars.

The final measures of cost and benefit are $21,745.25 and $46,338.58, respectively. Finally, MVPF is equal to the ratio of WTP to Cost, which is estimated to be 2.13. Without the WTP for increase in longevity, the MVPF comes out to be 0.96. At around a value of life of $5,100, the MVPF is around 1.

10. Instrumental Variables Approach

We pursue an instrumental variables strategy that exploits special features of the program, inspired by the literature exploiting examiner/judge leniency (Dahl et al. 2014, Dobbie et al. 2018). We posit that there are “good” and “bad” camps that differ in enrollee retention. We calculate the mean duration of training in camps as a measure of enrollee retention and use it as an instrument for their individual duration. More specifically, to take care of the mechanical correlation between enrollees and their camp mean duration, we calculate the leave-out mean duration by excluding the individual and their peers in the mean calculation.
We exploit the quasi-random nature of the assignment to camps to compare the duration of training and the outcomes of individuals assigned to “good” and “bad” camps. Although participation in the CCC was voluntary, once individuals signed up, they were allocated by the authorities to various camps—individuals were not allowed to choose which camp to train on:

> ***d. No promises to be made regarding camp location. -*** *The assignment of enrollees to camp is the responsibility of the Army following enrollment. The location of vacancies and the administrative convenience of the Corps are the primary bases upon which such assignments must be made. Under no circumstances, therefore, will a selecting agent*

make any definite and binding promises to prospective enrollees as to the camps to which they will be assigned.[5]

To which camp an individual was assigned was an important determinant of the duration of training. The historical evidence discussed in the previous section suggests that many camp characteristics—weather, distance to home, type of camp/work, the characteristics of one's peers and idiosyncratic characteristics of the Army personnel running the camps—had great effects on how long individuals trained. We show evidence that observable camp characteristics did indeed influence duration. While there is a myriad of factors that affect duration none of them is individually a large predictor of duration. But we demonstrate that the average duration in the camp an individual was assigned to is a powerful predictor of an individuals' duration. We rely on this evidence and make use of the average camp duration as an instrument for individual durations.

This IV strategy relies on multiple assumptions. In addition to being predictive of individual durations (satisfying the relevance condition), the leave-out mean camp duration must affect outcomes only through its effect on individual durations (exclusion restriction). If individuals of the same type trained together at the same time in the same camp, their durations and outcomes will be correlated. We rely on the historical and statistical evidence of the quasi-random nature of the camp assignment to argue this is unlikely to be the case.

However, we must also assume that these factors that affected camp duration had no impact on outcomes. One concerning possibility is that individuals formed network of friends in a given camp leading to longer durations and also affecting possibly labor market outcomes and well-being. To avoid this complication, we compute a special leave-out mean, that does not include the duration of the peers that the individual trained with. Essentially a person's duration is predicted using the average duration of individuals that did not overlap with the person in the camp. We also investigate how controlling for peer characteristics affects the results.

I. Instrumental Variable Results

We base our IV on a simple idea. As described in the empirical strategy section, individuals did not choose which camp to attend and the process by which they were assigned to camps was ad hoc, possibly quasi-random. If this is the case then individuals are sent to good or bad camps independently of their characteristics. So we can use camp characteristics as instruments for duration. Although many camp traits affect duration, the most predictive camp characteristic will be the average duration of training in a given camp: this summarizes whether altogether the camp was such that individuals were induced to train for long durations as a result of the camp conditions. We now more formally specify the model and assumptions under which this intuition holds.

a. Construction and Assumptions

Suppose we have the following model for person i assigned to camp j,

$$y_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 x_i + \beta_3 C_j + \gamma_i + \varepsilon_{ij}$$
$$D_{ij} = \pi_0 + \pi_1 Z_j + \pi_2 x_i + \pi_3 C_j + \gamma_i + \nu_{ij}$$

where y_{ij} is the age at death (or some other outcome), D_{ij} is the duration of CCC training, x_i is a vector of individual characteristics, γ_i is the enrollment cohort (county-quarter) dummy, c_j is vector of observable camp characteristics, and $\boldsymbol{Z_j}$ is a scalar summary measure of camp

[5] *Standards of eligibility and selection for junior enrollees*. Civilian Conservation Corps. June 15, 1939. pp 13.

characteristics that affect only duration, but not the outcome. Following Dobbie et al. (2018) we will construct the following "leave-out mean" using the residualized camp duration, after controlling for the enrollment cohort, which in our case is the unit of (quasi) randomization. First, we calculate the residual from regressing D_{ij} on γ_i which yields D_{ij}^*.

$$D_{ij}^* = \mathrm{D_{ij}} - \widehat{\gamma_1} = \mathrm{Z_j^*} + \mathrm{x_i^*} + \mathrm{c_j^*} + \mathrm{\nu_{ij}^*}$$

where x^* denotes variables uncorrelated with γ_i the cohort of enrollment. Then we calculate the leave-out mean of the residuals of duration, by summing over all individuals who trained in the camp, except for *i* as follows

$$\hat{\mathrm{Z}}_{\mathrm{ij}} = \sum_{\mathrm{k}\in\mathrm{K_j}} \frac{\mathrm{D_{kj}^*} - D_{ij}^*}{\left|\mathrm{K_j}\right| - 1} = \mathrm{Z_j^*} + \bar{u}_{-ij} = \mathrm{Z_j^*} + \overline{x^*}_{-ij} + \mathrm{c_j^*} + \overline{v^*}_{-ij}$$

where set K_j includes everyone who trained in camp j. If there is indeed conditional randomization, then this leave-out instrument is uncorrelated to ν_{ij} *so long as all other relevant camp characteristics are controlled for in both the first and second stages.* However, if we think that there are omitted camp characteristics in the first and second stage equations, this will cause problems for our instrument, as that term will belong in the residualized instrument.

A potentially important violation of this assumption arises if there are peer effects, that is if the characteristics of one's peers affect one's survival and also one's duration. Then the leave-out mean will be correlated with the mean duration but also with the error term. A simple solution consists in leaving out all the contemporary peers from the computation of the leave-out mean. This avoids the mechanical correlation between $\overline{v^*}_{-ij}$ (omitted peer characteristics) in the first stage equation and the leave-out mean. Then our instrument consists of the mean duration of individuals who trained in the same camp as *i* but did not overlap with *i*. In practice overlap is itself endogenous since it depends on one's duration, which in turn depends on one's peers. To avoid this, we exclude the duration of those who enrolled in the same quarter and year.[6]

The second concern is that there are other characteristics of the camp that affect duration and that are correlated with outcomes. If this is the case, we violate the "exclusion restriction" assumption. We can never be certain this assumption holds, but we can include camp characteristics directly as controls in the first and second stage to test the sensitivity of the results to these characteristics. For example, as discussed before, we find that weather has a (small) effect on duration. It might also have a long-term effect on health. We can control for weather and type of camp. However, we will have to assume that other unobserved camp characteristics are not correlated with the cohort leave-out duration mean at the camp.

In summary, we will proceed as follows. First, we compute a test statistic to examine for each randomization unit whether individuals that enrolled in the same time and county were as good as randomly assigned to camps. Second, we compute leave-out camp mean durations for

[6] We experimented with alternative definitions, for instance excluding peers in the same camp in the first month of training, or those who enrolled in the same county and quarter. The results from these alternatives are similar.

each individual *i*, leaving *i* and all of *i*'s peers out of the computation and use these as instruments. Lastly, we test the sensitivity of the results to controlling for observable camp characteristics.

b. Instrumental variable results

The historical evidence is consistent with the narrative that individuals could not chose their camp. The instruction booklet *Standards of eligibility and selection for junior enrollees* "The location of vacancies and the administrative convenience of the Corps are the primary bases upon which such assignments must be made". To test that individuals are "as good as randomly assigned" to camp, and thus that their characteristics are not determining camp durations, we conduct a step-down p-value test. The basic idea is that if individuals who showed up in a given time and place are randomly assigned to different camps, then the mean characteristics of those individuals should be the same across camps. We first verify that there is in fact variation in where individuals are sent: within randomization units there are in fact many camps individuals are assigned to. Out of 1,858 "randomization units, only 190 have a single camp where individuals were sent to. In the remainder the average number of camps is 4.4, and the median number of camps is 3.

There is another challenge in implementing this test: we have relatively small samples within a randomization unit, so that simple means tests will suffer from small sample biases. We solve the first problem by conduct exact inference. Out of 1437 randomization units, there are 197 for which the test can be conducted (more than 30 observations), and 152 that pass the test. In terms of observations, there are 10,597 observations for which the test can be conducted and 7,301 that pass the test. For our estimation, we will consider how restricting the sample to this smaller sub-sample affects our results. We consider two sub-samples: 1) "Randomized" sample that only takes county-of-enrollment*quarter-of-enrollment (CQE) cells that pass the randomize test, and 2) "Randomized Large" sample that assumes the CQE cell to pass if the test statistic cannot be computed due to its small sample (< 30).

Data Appendix Figure 7 shows the distribution of our instrument residualized for other controls in our specification (histogram) and the nonparametric first stage (line) calculated using an Epanechnikov kernel. There is a first stage positive relationship and it is monotonically increasing. Table 7 shows that the leave-out mean is a strong predictor of duration (after controlling for all covariates) of individual durations: regardless of the set of controls we use, or the estimation sample we consider, the camp leave-out mean is a statistically significant and economically significant predictor of individual duration. The magnitude changes from around 1 when not controlling for camp and peer controls to about 0.3 when controlling for camp and peer controls. The F-statistic is large, greater than 30 in most cases, well above the threshold of 10 required for weak instruments.

Data Appendix Table 2 shows the IV results of estimating survival to age 70 as a function of duration, and instrumenting individual training duration using the cohort-leave out mean duration. We show results with two sets of controls (individual controls and peer/camp characteristics added) and three samples ("Randomized Large" that pass the test or the test statistic cannot be computed, "Randomized" that pass the test, and the full sample). The results are very similar across specifications. The IV coefficients are imprecisely estimated, and while they are not different than our OLS estimates they are also not different from zero.

11. Special Acknowledgements

A. Dirk Van Hart provided us with dates for which New Mexico camps were open for enrollment.
B. Robert W. Audretsch provided us with dates for which Colorado camps were open for enrollment.

12. Additional Tables and Figures

Data Appendix Table 1: Comparison of Counties of Enrollees vs Whole State

Year	1930				1940			
State	CO		NM		CO		NM	
Geography	State	CCC	State	CCC	State	CCC	State	CCC
Variables								
Share Urban	0.50	0.40	0.25	0.22	0.53	0.42	0.33	0.28
Share in Farm	0.27	0.33	0.37	0.38	0.22	0.28	0.32	0.35
Share Owns Home	0.50	0.50	0.59	0.64	0.47	0.48	0.61	0.65
Mean Rent	38.88	37.60	26.39	23.09	102.99	95.43	219.27	271.40
Mean Age	29.57	28.35	25.26	25.24	31.40	30.12	26.14	25.84
Share Male	0.51	0.52	0.52	0.52	0.51	0.51	0.51	0.51
Share White	0.98	0.99	0.92	0.95	0.99	0.99	0.93	0.96
Share Mexican	0.06	0.07	0.14	0.11	0.07	0.13	0.34	0.44
Share Ever Married	0.51	0.49	0.45	0.44	0.54	0.52	0.47	0.45
Share Students	0.24	0.25	0.25	0.25	0.21	0.23	0.25	0.26
Share Foreign-born	0.10	0.09	0.06	0.05	0.07	0.06	0.03	0.02
Mean Occscore	21.78	20.59	19.05	18.34	22.54	21.38	20.10	19.19
Share Employed	0.90	0.90	0.93	0.92	0.90	0.89	0.88	0.85
Mean Income					392.11	332.25	326.73	277.49
Mean Educ Years					7.75	7.25	5.86	5.45
Share Hisp Origin					0.08	0.13	0.34	0.44

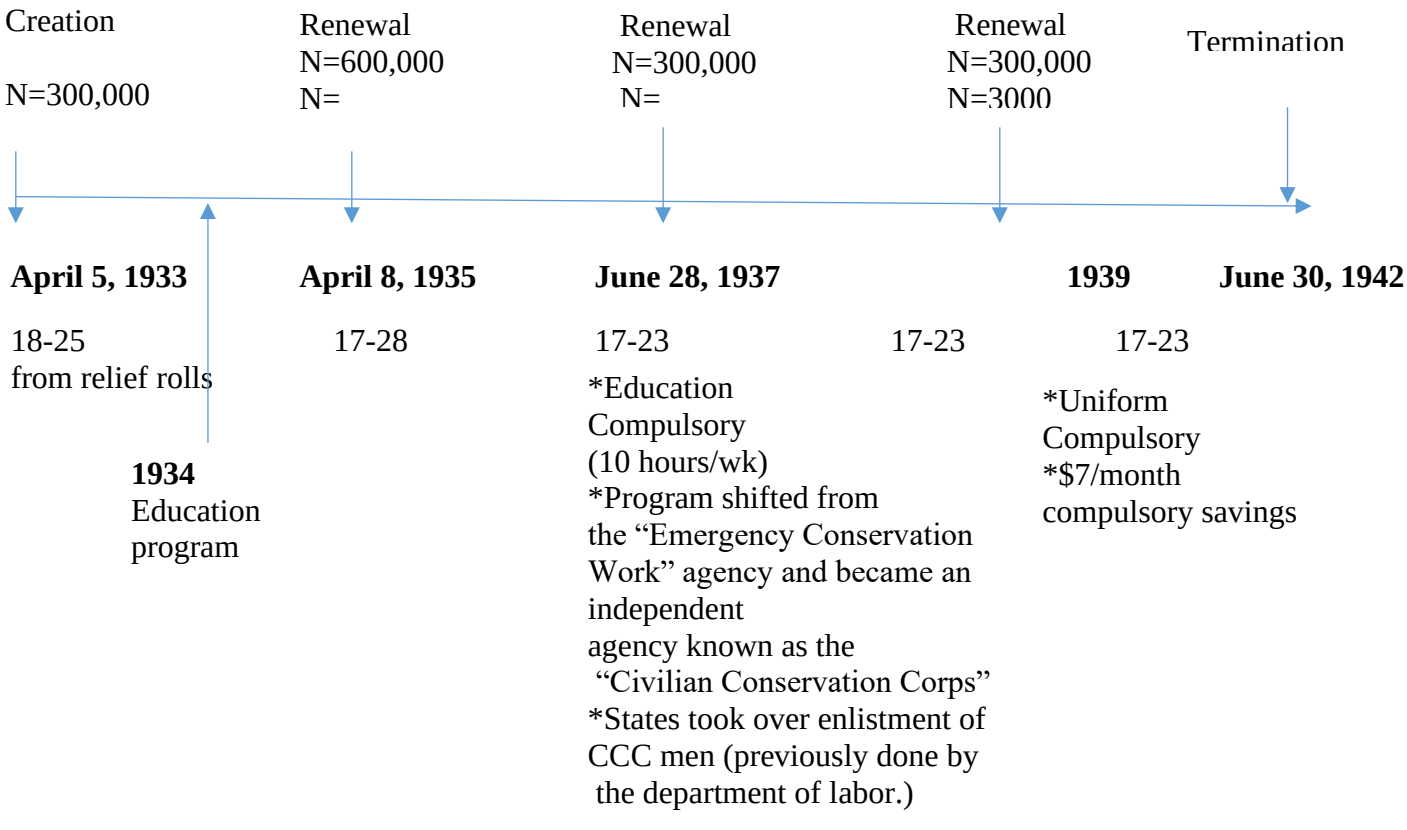
Note: Columns "State" are the state average of variables in each row. Columns CCC is the weighted average of county characteristics, where the weights are the share of CCC enrollees in our data enrolling from each county.

Data Appendix Table 2: Instrumental Variables Results

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	pass randomization test or randomization test cannot be computed		pass randomization test		All	
Dependent variable: survived to age 70, imputed (mean:)	Indiv Controls only	Add Camp + Peer Chars	Indiv Controls only	Add Camp + Peer Chars	Indiv Controls only	Add Camp + Peer Chars
Instrumental Variables Estimate						
Duration of service : Enrollment 1	0.0129	0.0484	0.00748	0.00734	0.0245	0.0761
	(0.0348)	(0.0699)	(0.0556)	(0.117)	(0.0347)	(0.0705)
Observations	19,097	19,097	6,513	6,513	21,195	21,195
Number of groupayq	1,794	1,794	150	150	1,839	1,839
First Stage						
Enrollment Quarter leave-out mean duration of camp	1.073***	0.582***	1.060***	0.573***	0.986***	0.522***
	(0.0951)	(0.0926)	(0.131)	(0.114)	(0.0944)	(0.0954)
F-statistic	127.12	39.47	65.14	25.15	108.95	29.89
F-statistic corrected using Hull's correction	.7	.22	.39	.15	.6	.17
OLS						
Duration of service : Enrollment 1	0.0222***	0.0268***	0.0335***	0.0416***	0.0226***	0.0275***
	(0.00545)	(0.00608)	(0.00798)	(0.00868)	(0.00526)	(0.00584)
Reduced Form						
Enrollment Quarter leave-out mean duration of camp	0.0138	0.0282	0.00793	0.00421	0.0241	0.0397
	(0.0375)	(0.0410)	(0.0590)	(0.0668)	(0.0342)	(0.0365)

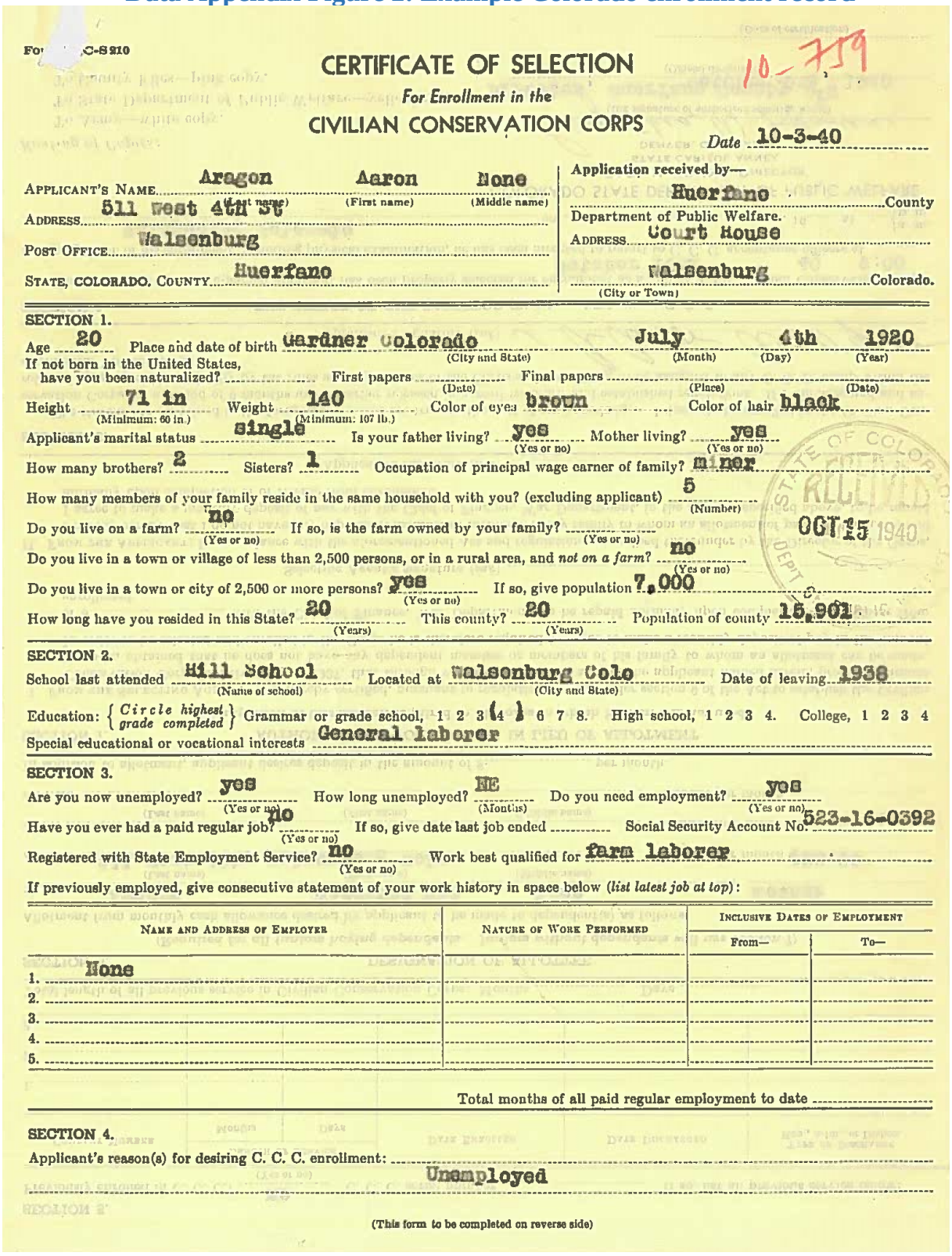
Note: the leave out mean excludes men that enrolled in the same quarter the enrollee started. This table uses the life table imputations for P70.

Data Appendix Figure 1: History of CCC program[7]



[7] Information on the history of the CCC program used in Appendix Figure 1 come from the following sources: https://babel.hathitrust.org/cgi/pt?id=mdp.39015004052794;view=1up;seq=13 On June 28, 1937, the CCC was once again renewed with funding for three additional years according to Public Law No. 163 (effective on July 1, 1937) see here: https://www.nps.gov/parkhistory/online_books/ccc/cccaa.htm

Data Appendix Figure 2: Example Colorado enrollment record



Form C-S210

CERTIFICATE OF SELECTION

For Enrollment in the

CIVILIAN CONSERVATION CORPS

10-759

Date 10-3-40

APPLICANT'S NAME: Aragon (Last name) Aaron (First name) None (Middle name)

ADDRESS: 511 West 4th St

POST OFFICE: Walsenburg

STATE, COLORADO. COUNTY: Huerfano

Application received by—

Huerfano County

Department of Public Welfare.

ADDRESS: Court House

Walsenburg (City or Town) Colorado.

SECTION 1.

Age 20 Place and date of birth Gardner Colorado (City and State) July (Month) 4th (Day) 1920 (Year)

If not born in the United States, have you been naturalized? ____ First papers ____ (Date) Final papers ____ (Place) ____ (Date)

Height 71 in (Minimum: 60 in.) Weight 140 (Minimum: 107 lb.) Color of eyes brown Color of hair black

Applicant's marital status single Is your father living? yes (Yes or no) Mother living? yes (Yes or no)

How many brothers? 2 Sisters? 1 Occupation of principal wage earner of family? miner

How many members of your family reside in the same household with you? (excluding applicant) 5 (Number)

Do you live on a farm? no (Yes or no) If so, is the farm owned by your family? ____ (Yes or no)

Do you live in a town or village of less than 2,500 persons, or in a rural area, and *not on a farm*? no (Yes or no)

Do you live in a town or city of 2,500 or more persons? yes (Yes or no) If so, give population 7,000

How long have you resided in this State? 20 (Years) This county? 20 (Years) Population of county 15,901

STATE OF COLORADO RECEIVED OCT 15 1940 DEPT. OF PUBLIC WELFARE

SECTION 2.

School last attended Hill School (Name of school) Located at Walsenburg Colo (City and State) Date of leaving 1938

Education: { *Circle highest grade completed* } Grammar or grade school, 1 2 3 4 (5) 6 7 8. High school, 1 2 3 4. College, 1 2 3 4

Special educational or vocational interests General laborer

SECTION 3.

Are you now unemployed? yes (Yes or no) How long unemployed? NE (Months) Do you need employment? yes (Yes or no)

Have you ever had a paid regular job? no (Yes or no) If so, give date last job ended ____ Social Security Account No. 523-16-0392

Registered with State Employment Service? no (Yes or no) Work best qualified for farm laborer

If previously employed, give consecutive statement of your work history in space below (*list latest job at top*):

	Name and Address of Employer	Nature of Work Performed	Inclusive Dates of Employment From—	To—
1.	None			
2.				
3.				
4.				
5.				

Total months of all paid regular employment to date ____

SECTION 4.

Applicant's reason(s) for desiring C. C. C. enrollment: Unemployed

(This form to be completed on reverse side)

SECTION 5.

Previously enrolled in C. C. C.? no (Yes or no) C. C. C. serial number ..... If so, list all previous service below:

Company Number	Length of Service		Date Enrolled	Date Discharged	Type of Discharge Hon., Adm., or Dishon
	Months	Days			
1.					
2.					
3.					

Total length of all previous service in Civilian Conservation Corps: Months ..... Days .....

SECTION 6. **DESIGNATION OF ALLOTTEE**

(Required for all juniors having dependents. Juniors without dependents will use Section 7)

Allotment from monthly cash allowance desired by applicant to be made to dependent(s) as follows:

Name Aragon (Last name) Margaret Mrs (First name) None (Middle name) Relationship Mother

Address 511 West 4th, Walsenburg Colo Amount per month $22.00

Name ..... (Last name) ..... (First name) ..... (Middle name) Relationship .....

Address ..... Amount per month .....

In addition to allotment, applicant desires deposit in the amount of $..... per month.

SECTION 7. **AUTHORIZATION FOR DEPOSIT IN LIEU OF ALLOTMENT**

(Completion of this Section required in all cases in which Section 6 is not used)

I. From the Selecting Agency: It is hereby certified, pursuant to regulations issued under section 9 of the Act to establish the Civilian Conservation Corps effective July 1, 1937, that through verification of the status of the applicant named herein, proper assurance has been obtained that he does not have any dependent member or members of his family to whom an allotment can be made. In order to be selected and enrolled in the Corps he is therefore required to agree to make a monthly deposit of pay in the amount of $..... with the Chief of Finance, War Department, to be repaid normally upon completion of or release from enrollment.

Selecting Agent's signature (*ink*) .....

II. From the Applicant: In accordance with the aforementioned Act and regulations prescribed thereunder by the Director of the Corps, I hereby certify that I do not have any dependent member or members of my family to whom an allotment of pay can be made, and I agree to make a monthly deposit of pay with the Chief of Finance, War Department, in the amount specified above, to be repaid normally upon completion of or release from enrollment.

Applicant's signature (*ink*) .....

SECTION 8.

The statements contained in the foregoing Sections are true, to the best of my knowledge. I desire to be enrolled in the Civilian Conservation Corps for a period of 6 months unless earlier released in accord with law and established regulations. If I am accepted and enrolled, I agree to abide faithfully by the rules and regulations of the Corps and am willing to be assigned to any C. C. C. camp within the continental United States.

Applicant's signature (*ink*) J Aaron Aragon

SECTION 9. **THE OFFICE OF THE DIRECTOR (Division of Selection) C. C. C.**

Certifies that the above-named applicant has been properly selected for enrollment as a Junior in the Civilian Conservation Corps.

For completion of his enrollment, including physical examination, he has been directed to report to C. C. C. acceptance officers at October 10th 40 9:00 Walsenburg Colorado on ....., 19..... at ..... {a. m. / p. m.

COLORADO STATE DEPARTMENT OF PUBLIC WELFARE
EARL M. KOUNS, Director
STATE CAPITOL ANNEX
DENVER, COLORADO

Routing of Copies:

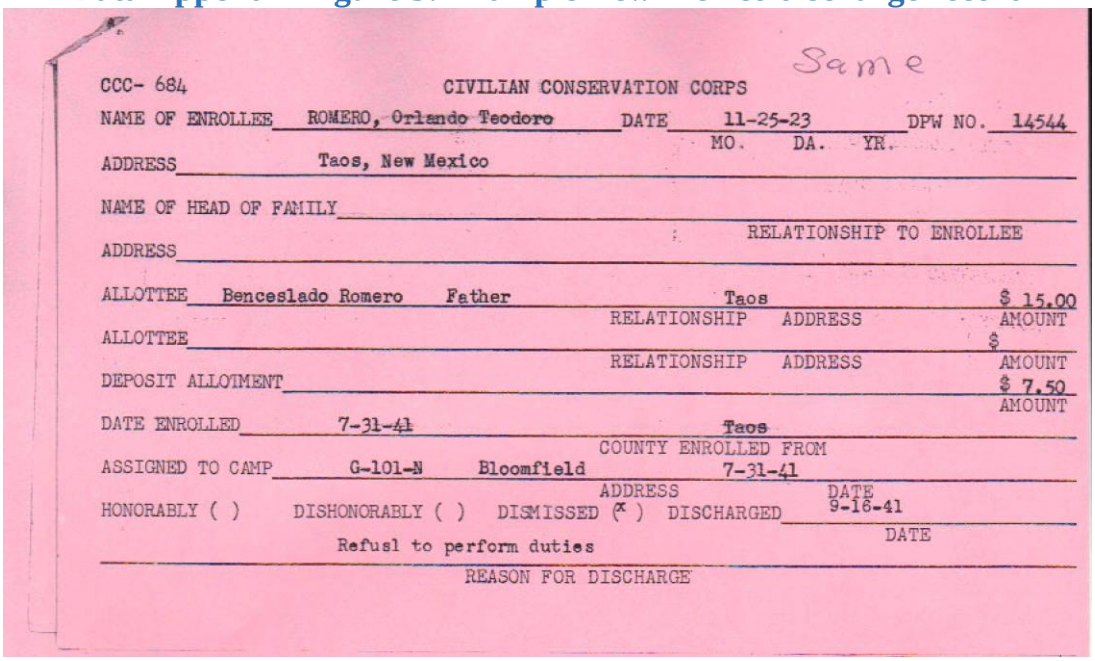
To Army—white copy.
To State Department of Public Welfare—yellow copy.
To County Files—pink copy.

By Cynthia W. Haines
(Ink signature of authorized selecting agent)

Director, Huerfano County, DPW
(Official designation)

October 9th, 1940

Data Appendix Figure 3: Example New Mexico discharge record



Same

CCC- 684 CIVILIAN CONSERVATION CORPS

NAME OF ENROLLEE ROMERO, ~~Orlando~~ Teodoro DATE 11-25-23 (MO. DA. YR.) DPW NO. 14544

ADDRESS Taos, New Mexico

NAME OF HEAD OF FAMILY _____ RELATIONSHIP TO ENROLLEE

ADDRESS _____

ALLOTTEE Benceslado Romero Father Taos $ 15.00
RELATIONSHIP ADDRESS AMOUNT

ALLOTTEE _____ $
RELATIONSHIP ADDRESS AMOUNT

DEPOSIT ALLOTMENT _____ $ 7.50
AMOUNT

DATE ENROLLED ~~7-31-41~~ ~~Taos~~
COUNTY ENROLLED FROM

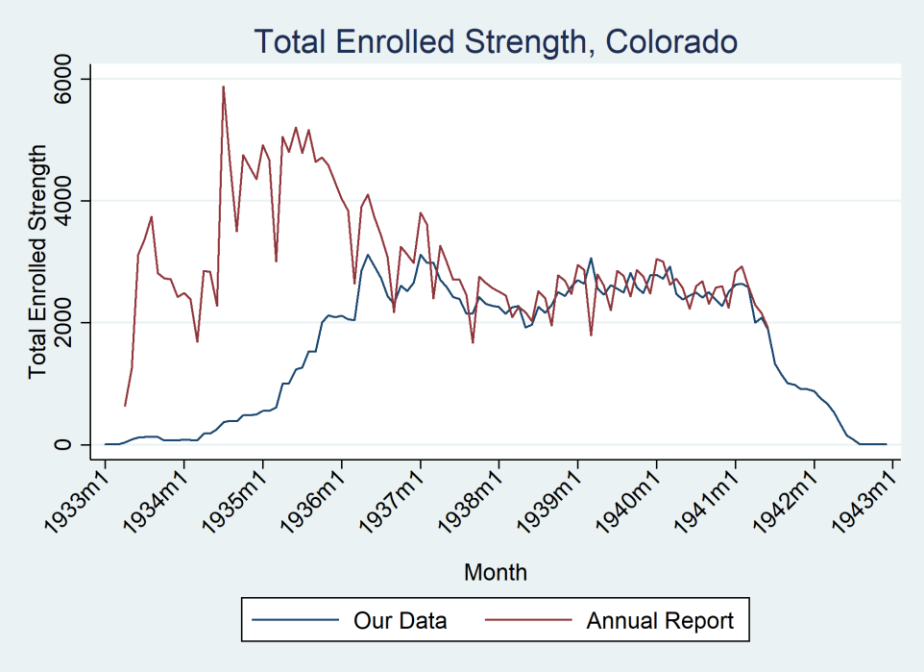
ASSIGNED TO CAMP G-101-N Bloomfield 7-31-41
ADDRESS DATE

HONORABLY ( ) DISHONORABLY ( ) DISMISSED (X) DISCHARGED 9-16-41
DATE

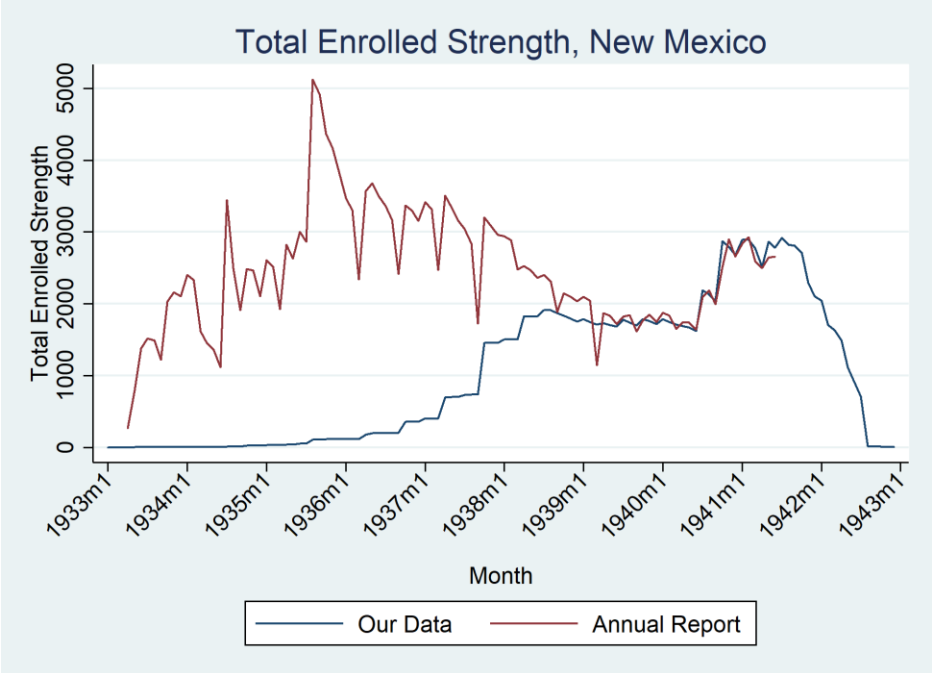
Refusl to perform duties
REASON FOR DISCHARGE

Data Appendix Figure 4: CO and NM data completeness

a. Archival data coverage in Colorado



a. Archival data coverage in New Mexico

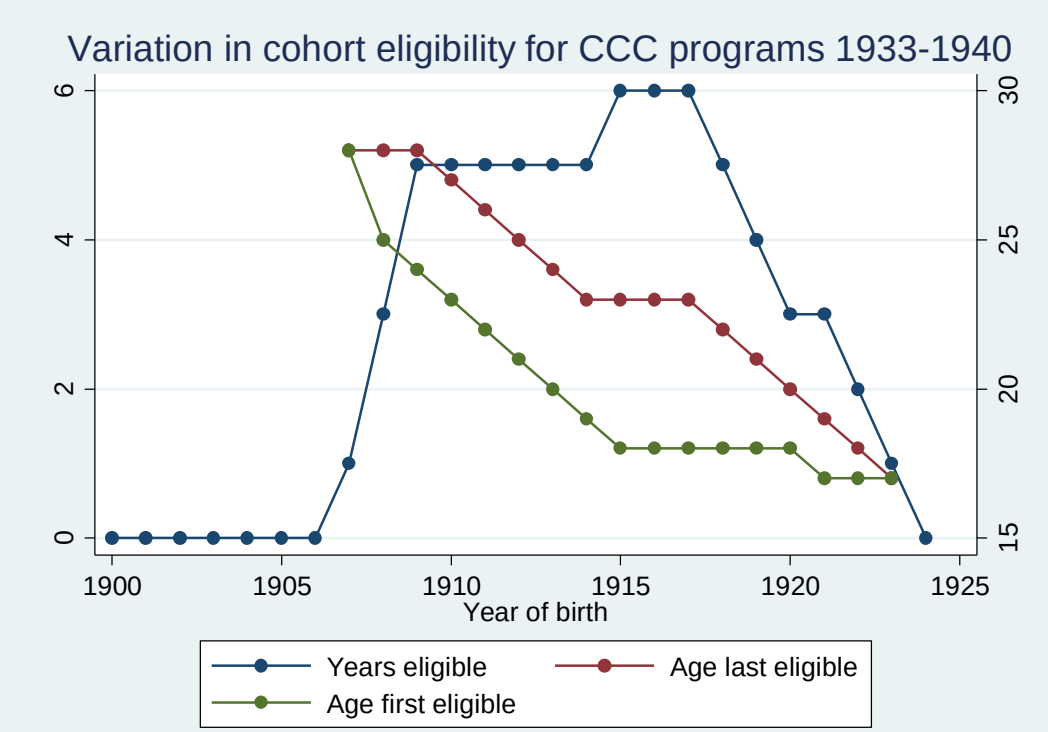


Note: Total enrolled strength is the number of enrollees at each month. Data from the Annual Report come from the following sources:

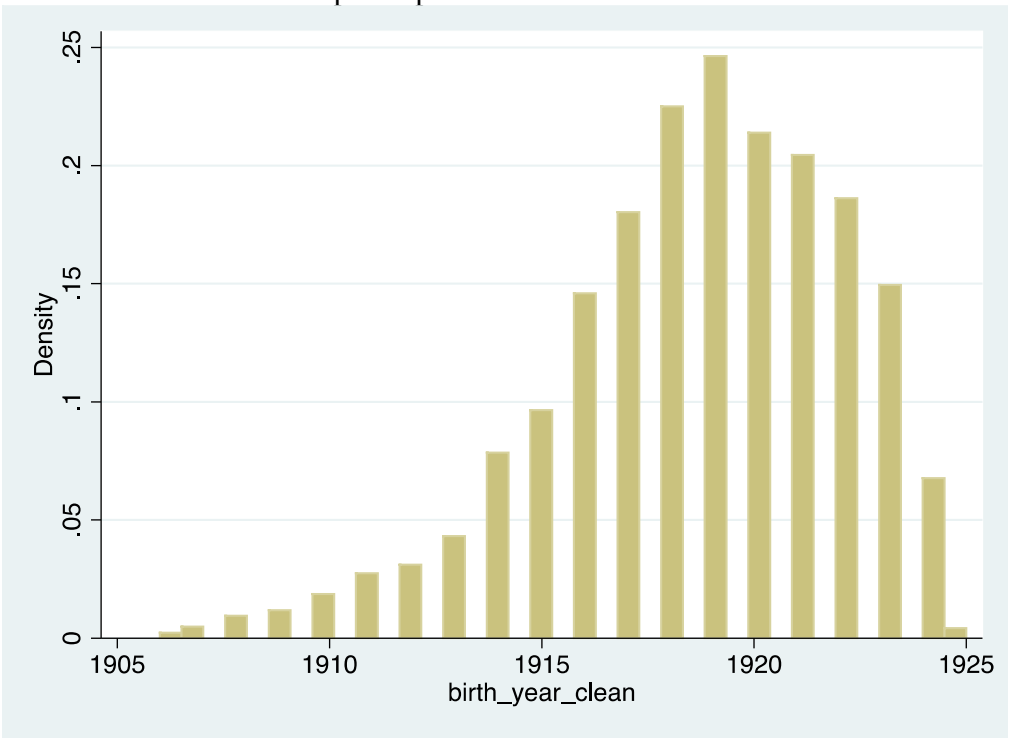
- Summary Report of the Director of Emergency Conservation Work on the Operations of Emergency Conservation Work: For the period extending from April 1933 to June 30, 1935, Appendix E
- Annual Report of the Director of Emergency Conservation Work: Fiscal Year Ending June 30, 1936, Appendix E
- Annual Report of the Director of Emergency Conservation Work: Fiscal Year Ending June 30 1937, Appendix D
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1938, Appendix E
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1939, Appendix I
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1940, Appendix E
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1941, Appendix E

Data Appendix Figure 5: Cohort eligibility and participation in CCC

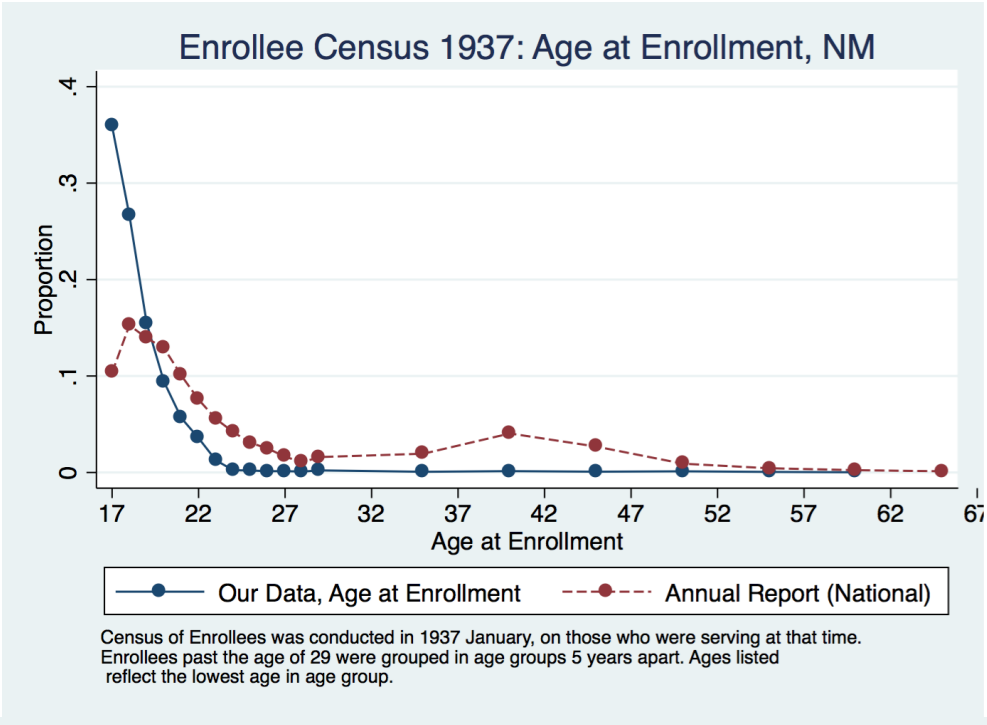
a. Variation in Cohort eligibility during the years CCC operated

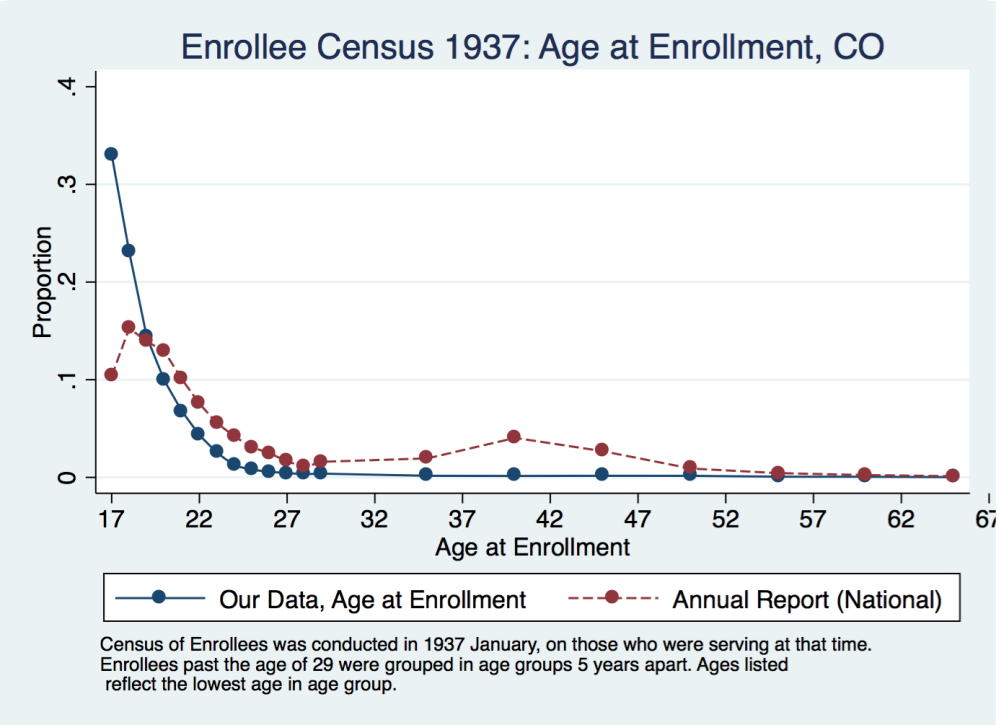


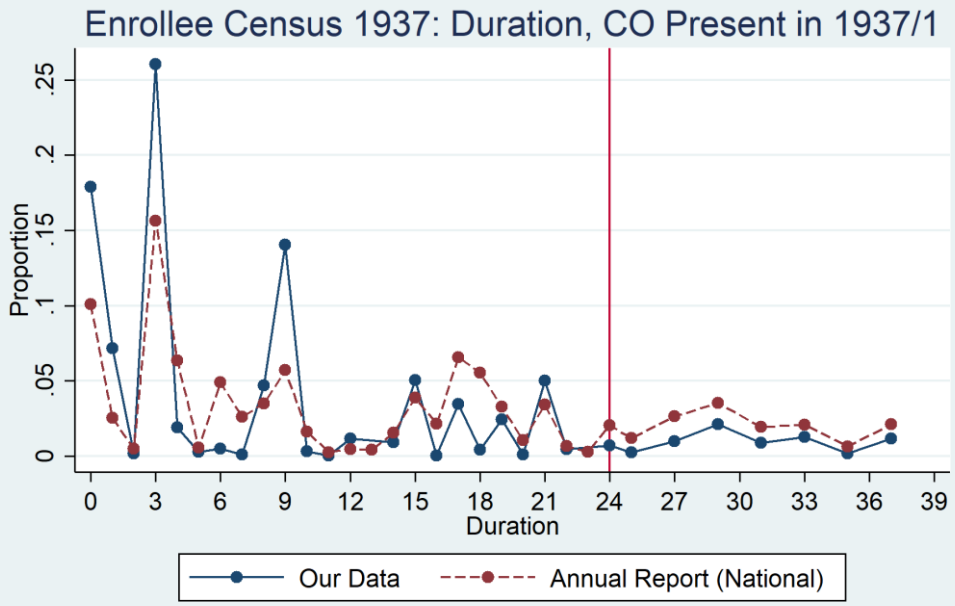
b. Cohort participation in CO and NM CCC records



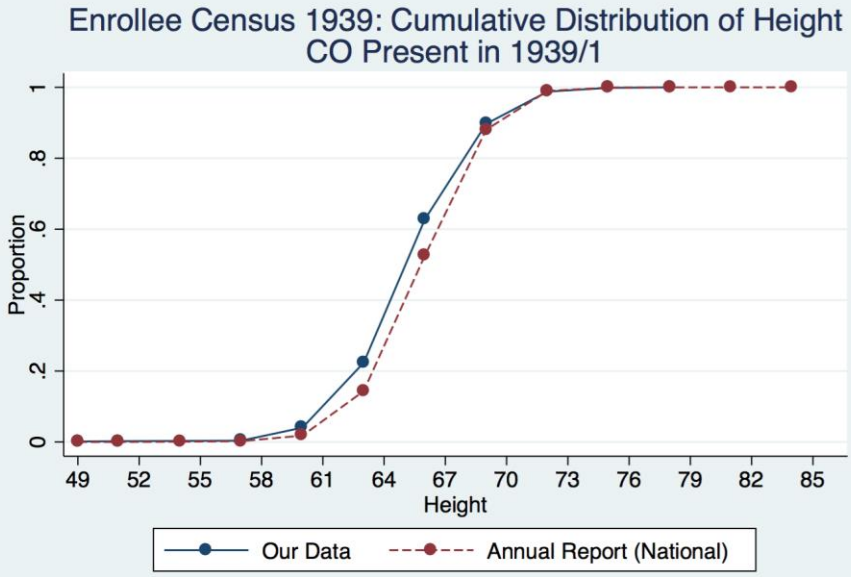
Data Appendix Figure 6: CCC enrollees in CO and MN are more disadvantaged than enrollees nationwide



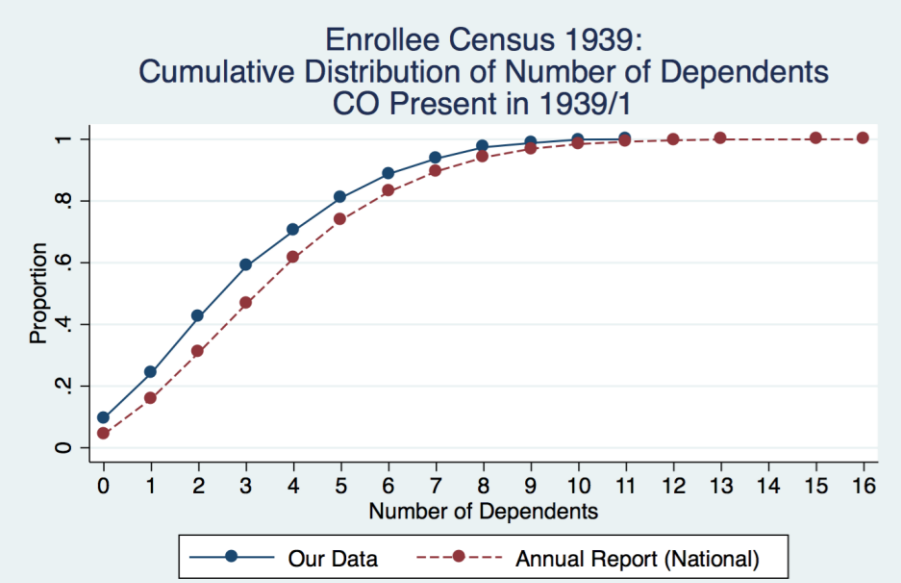


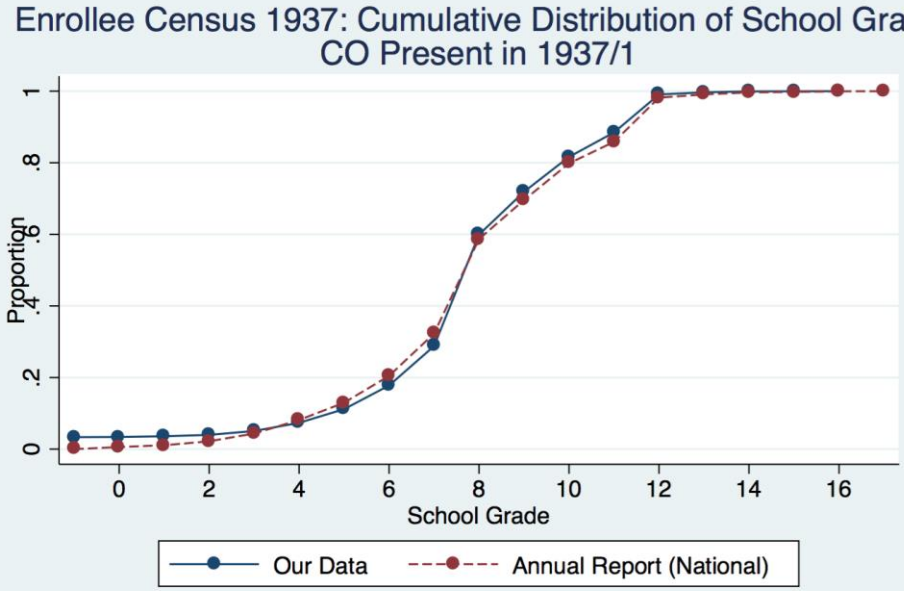


Census of Enrollees was conducted in 1937 January, so here we restrict the sample to be enrolled in 1937 Jan, and duration to be calculated analagously to the Report (duration until 1937 Jan

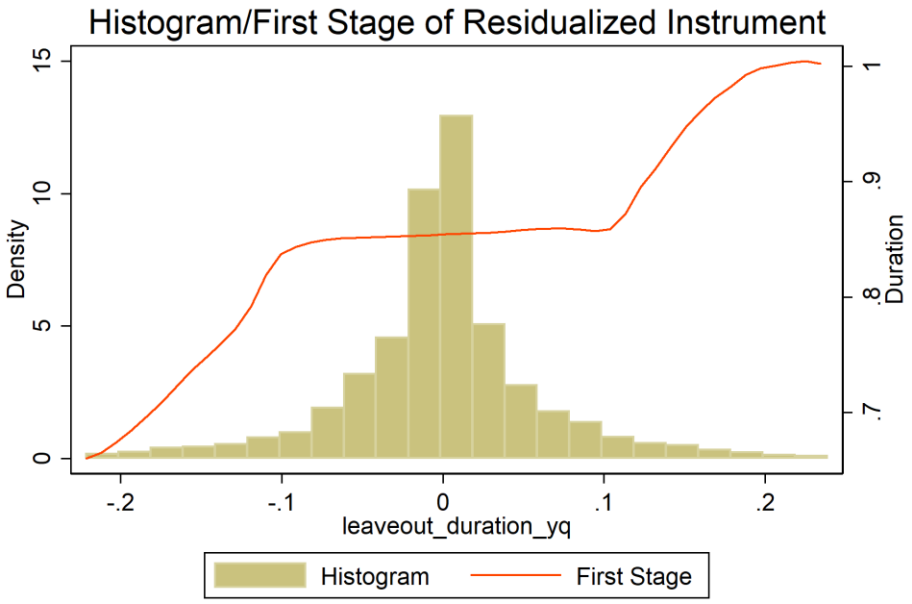


Height reported in ranges of 10. Lower bound of range is used to bucket height.


Enrollee Census 1939:
Cumulative Distribution of Number of Dependents
CO Present in 1939/1
Proportion
0
.2
.4
.6
.8
1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Number of Dependents
Our Data
Annual Report (National)
Number of Dependents presented for each dependent until over 13 dependents.
Lower bound of ranges of 2 from 13 was used to bucket 13 - 16 dependents.


Enrollee Census 1937: Cumulative Distribution of School Grad
CO Present in 1937/1
Proportion
0
.2
.4
.6
.8
1
0 2 4 6 8 10 12 14 16
School Grade
Our Data
Annual Report (National)
School Grade presented as ranges of 2 months. Lower bound of the range was used to bucket.
If number of unemploymed months missing, it is assumed enrollee never held a steady job.

Data Appendix Figure 7: Histogram of Instrument and First Stage



Values outside 1-99 percentiles are suppressed. Mean: 0.00, SD: 0.08.
First stage: Epanechnikov kernel with bandwidth 0.05