Native Americans in the Historical Census: New Data and Applications

Christian Dippel* Dustin Frye†

March 27th 2020

Abstract

The digitized historical Full Count Census waves from 1900–1940 are a rich source of information for individual- or household-level quantitative research on the Native American population, with the average census wave containing more than 300,000 Native American individuals. Without the missing information on reservation, however, there is no treatment variation in any of the major historical policies that Native Americans were exposed to, such as Indian boarding schools and land allotment. We describe the construction of a stable reservation-to-individual crosswalk that assigns a reservation to over ninety percent of individuals in the historical Native American population, and apply this crosswalk to answering some long-standing questions on within-reservation inequality.

Keywords:
JEL Codes:

*University of California, Los Angeles, and NBER.
†Vassar College.

1 Introduction

American Indians are the poorest Census-defined racial group in the United States; they have the highest rates of violent victimization (Bureau of Justice Statistics, 2019), and they have the poorest health outcomes, including the worst substance abuse problems (Center for Disease Control, 2019).[1] A large body of literature in institutional, cultural, and development economics and related fields suggests that these socio-economic outcomes (income, health, and well-being) are shaped by historical factors. In short, history matters.

In the Native American context, there is great potential for quantitative historical research using the historical waves of the 1900–1940 Full Count Census waves.[2] The 1900–1940 Census waves are a rich data source. The average Census wave contains more than 300,000 individuals whose race is enumerated as being Native American, and the data include rich demographic, and in later years also economic, information. Furthermore, 1900–1940 covers almost the entirety of the "Assimilation Era," during which Native American tribes were exposed to a range of assimilationist policies such as boarding schools and land allotment, the consequences of which reverberate to the present day (Treuer, 2012, 2019).[3]

Unfortunately, the micro data contain no information on the reservation an individual Native American belonged to. This is a critical limitation because reservations varied dramatically in the experience of their formation (Dippel, 2014; Feir, Gillezeau, and Jones, 2019; Anderson, 2020), in their exposure to different policies during the Assimilation Era (Carlson, 1978, 1981; Golenko, 2010; Feir, 2016; Gregg, 2018; Leonard, Parker, and Anderson, 2020) , and in their political recognition and post-Assimilation era experience (Anderson and Parker, 2008; Akee, Jorgensen, and Sunde, 2015; Frye and Parker, 2019). It merits clarifying that today's federally or state-recognized 'administrative tribes' are almost always reservations, with the exception of a few reservations that jointly form a federally recognized tribe, and a few recognized tribes without a reservations.

[1] The percentage of Native Americans who experienced violence in 2013 was 2.8%, relative to 1.3% for both blacks and Hispanics, and 1.1% for whites.

[2] The 1900 starting year is determined by the fact that "Indians not taxed" (i.e. the vast majority who were living on reservations at this time) were not enumerated in the Population Census before 1890, and the 1890 Census was lost to a fire. By historical estimates, these constituted roughly ninety percent of the Native American population in 1860, 1870 and 1880. See National Archives (2019). According to the "rule of 72", the 1950 wave will gradually be made available to researchers starting in 2022.

[3] The beginning of the era is usually marked by the General Allotment (or 'Dawes') Act of 1887, and its end by the Indian Reorganization (or 'Howard-Wheeler') Act of 1934.

Reservations can almost always be clearly traced to a single cultural/ancestral tribe (e.g., White Earth is an Ojibwe reservation, Rosebud is a Sioux reservation), for which the Census equally had no information. The absence of reservation/tribe information is a combination of two factors: In most Census waves, reservation/tribe was never enumerated in the first place; and in the 1930 wave where it was enumerated, it is not among the variables that have up to now been digitized in the Full Count micro data.[4]

We address this problem through the construction of a stable reservation-to-individual crosswalk. This crosswalk is based on a combination of spatial matching techniques using Census-enumerated variables. First, we do text-analysis of the official descriptions of 'enumeration districts' to identity those described on reservations. Second, we geo-locate all reported towns in the Population Census and overlay them with historical reservation maps. This allows us to identify many additional enumeration districts which contain towns that are located in or immediately adjacent to reservations. Combining these two exercises, we can create a stable reservation-to-individual crosswalk that assigns over 90% of individual Native Americans in the 1930 and 1940 Census to their reservation.

We then apply this reservation-to-individual crosswalk to answering two long-standing questions on within-reservation inequality. The first pertains to the differential formation-experience of reservations, namely whether reservations were forcibly formed out of formerly politically disconnected tribal bands. The second pertains to the differential experience at the end of the assimilation era, namely the adoption of the Indian Reorganization Act (IRA).

This paper's contribution is to the field of quantitative research on Native American history. This research area has been slow to adopt the micro-data based research approaches that are becoming more common in the quantitative study of other historical populations; instead relying almost exclusively on reservation-aggregates that were reported in Annual Reports by the Bureau of Indian Affairs (BIA) from the 19th century until the Indian Reorganization Act (IRA) of 1934. Examples of such reservation-aggregate studies include Carlson (1978, 1981); Anderson and Parker (2008); Golenko (2010); Dippel (2014); Akee et al. (2015); Frye and Parker (2019); Leonard et al. (2020); and Feir et al. (2019). Exceptions we are aware of include Akee (2019), who uses

[4] Digitization efforts prioritize having a full count of people over having a wide data-set with all variables contained in the hard-copies.

county of residence to link Native Americans in 1900 and 1910 to one of two reservations in Minnesota (White Earth and Red Lake), as well as Dippel and Frye (2020).[5] By clarifying the main challenges that researchers will face when using historical micro-data on Native Americans, and proposing solutions to these challenges, we hope to enable more micro-based research on this important population.

We being by surveying the historical micro-data that exist for Native Americans in Section 2, before describing the construction of the reservation-to-individual crosswalk in Section 3. In Section 4, we then apply this crosswalk to answering some long-standing questions on within-reservation inequality.

[5] Miller (2015, 2016) also uses micro-data, but her focus is primarily on freed black slaves within the Cherokee nation.

2 Native Americans in the Historical Censuses

Native Americans were first enumerated as a separate racial group in 1860. However, until 1890 Census enumerators were explicitly instructed not to enumerate "Indians not taxed." Indians "not taxed" were all of those living on reservations, and these constituted roughly 90 percent of the overall Native American population. In 1890, the Census enumerated all Native Americans for the first time, but the 1890 micro-form was lost to a fire. From 1900 on, all Native Americans are in principle enumerated in the Census. However, the enumeration of individual Native Americans' tribal affiliation varied a lot across waves (as we discuss in Section 3.1), Furthermore, in most years tribe was not enumerated, as we discuss now

In 1900 and 1910, Native Americans were enumerated using a more detailed 'Indian Schedule' (Familysearch, 2019). which included questions about tribal affiliation (and blood quantum). However, the tribal affiliation that was enumerated was more akin to an individual's ethnic tribe than to their affiliation with a specific reservation. This is less useful in the sense that federal policy varied by reservation, not tribe, and because it was reservations that, after 1934, would become the polity that is today called a recognized tribe. From the researcher's point of view, this discussion is moot, because either way tribe is not an included variable in the digitized Full Count Census.

In 1920, there was no question on tribal affiliation.

In 1930, there were no special schedules, but tribal affiliation was included as a question on the regular Census schedule. In digitized form, this variable is contained in the 5% micro-sample available on IPUMS (See variable 'TRIBE' on IPUMS). However, it has not been digitized in the Full Count census, simply because the digitization is expensive and prioritizes having a full count of people over having more variables.

From 1940 on, the regular Census schedule again did not ask any tribal information.[6]

Not having tribe information severely limits the interesting questions that can be asked in the micro-data, simply because all consequential federal and state policies varied at the reservation level in whether and how they were applied. In the next section, we discuss how we successfully assigned over 95% of Native Americans in the historical census waves to their tribe.

[6] Beginning in 1970, the Census started again including tribal affiliation, but in 1970 and 1980, the tribal affiliation that was enumerated was again more akin to an individual's ethnic tribe than to their affiliation with a specific reservation. It is only since the 1990 census that tribal affiliation has been enumerated in the way that reflect's an individual's affiliation with a specific recognized tribe, i.e. a specific reservation.

3 Constructing a Stable Reservation-to-Individual Crosswalk

3.1 A Note on the 1900–1920 Data

From 1900 – 1920, the Office of Indian Affairs consistently reported larger Native American populations than the Full Count Census estimates from the Census Bureau and consequently, the available IPUMS full count sample. Table 1 reports population estimates for Native Americans from 1860 – 1940 reported separately by the Census Bureau and the Office of Indian Affairs.

The final column includes the counts available through IPUMS. The estimated number of Native Americans fluctuates considerably by source. The large systemic discrepancy between Census/IPUMS and the Office of Indian Affairs from 1860–1880 is explained by the fact that the Population Census did not enumerate "Indians not taxed" (i.e. the ninety percent who were living on reservations at this time) before 1890 (National Archives, 2019).

The Office of Indian Affairs (OIA) noted the discrepancies in population counts between their office and the Census Bureau (Office of Indian Affairs, 1935, p. 62–73). They believed that the enumerations in 1900 and 1910, which featured additional Indian supplements, included personnel from the OIA that were familiar with reservations and Native Americans, which resulted in a more accurate count. They attributed the general under counting to resource constraints in reaching remote Native American communities and felt that the large difference in 1920 population counts to the exclusion of OIA personnel from the enumeration process, resulting in Native Americans being enumerated as non-Native. Because of these large fluctuations in enumeration and contemporary recording of Native Americans we focus our paper on 1930 and 1940, when Census/IPUMS enumeration closely aligns with the Office of Interior estimates.

3.2 Constructing a Stable Reservation-to-Individual Crosswalk for 1930 and 1940

The most readily available spatial information in the historical Census is county. Unfortunately, reservations tend to be relatively tightly clustered in parts of the Pacific Northwest, the Southwest, the Plains, and around the border area of Minnesota and Michigan. These within-county clusters prohibit accurate matching of individuals to reservations.[7]

[7] An additional concern is that county boundaries were also in flux over this period. Between 1900 and 1940 there were nearly 750 county boundary adjustments, with many of these occurring in western states with American Indian reservations (Siczewicz, 2011).

Table 1: Population Counts of American Indians by Decade

Decade	Census Bureau	Office of Indian Affairs	IPUMS
1860	44,021	339,421	40,150
1870	25,731	313,712	21,052
1880	65,407	306,543	7,039
1890	248,253	249,278	
1900	237,196	270,544	174,145
1910	265,683	304,950	283,975
1920	244,437	336,337	236,799
1930	332,397	340,541	330,223
1940	333,969	327,958	308,303

Notes: Census Bureau figures from the National Archives (2019) and US Census Bureau (1951). Office of Indian Affairs figures from Office of Indian Affairs (1935) and Office of Indian Affairs (1941).

Instead of county, we therefore relied on the city (variable `STDMCD`) and enumeration district variables (variable `ENUMDIST`) . However, there is considerable variation in the usefulness of these variables across census waves. Specifically, the spatial information in the 1930 Census is better than in the 1940 census on both variables. Because of this, our approach is to create a complete `ENUMDIST`-to-reservations crosswalk for 1930 (which incorporates the `STDMCD` information), and then to apply this to 1940 through a `ENUMDIST-1930`-to-`ENUMDIST-1940` crosswalk, which we web-scraped from Morse and Weintraub (2019).

The city variable `STDMCD`: In 1940, city is often available only for large metropolitan areas, rendering them unhelpful for determining reservation assignment. The 1930 Census is unique in that it has city or township information that can be used to assign reservations to enumeration districts.[8] Because `STDMCD` in 1930 includes the many small towns one finds on or near reservations, we identified cities associated with almost the entire universe of reservations, which we could then link to the enumeration districts that contained these towns.

We focus our search on the 5,154 townships that included a Native American resident in 1930. We geolocated this set of townships using both the BING Maps and Google Maps APIs. Both APIs deliver specific coordinates for each township with the level of accuracy or precision of each match. We spatially join each geolocated township to modern reservation boundaries to assign each township to a reservation. To identify adjacent townships, we calculate the distance from

[8] `STDMCD` is not available on the IPUMS/NBER version of the Full Count, but it is available on the *Center for Population Research* (CCPR) server at UCLA version.

each township to the closest reservation boundary. A small set of townships are identified by their location in the Public Land Surveillance System (PLSS). For this subset, we assign reservation location using the spatial overlap of reservations to the PLSS.

Enumeration Districts `ENUMDIST`: `ENUMDIST` in 1930 is a numeric value, than when combined with state and county, uniquely identifies an enumeration district in 1930. In 1940, enumeration districts are identified using state, supervisor district, and the enumeration district.[9] These enumeration districts have not been geo-referenced, and we therefore rely on the text of their official descriptions to make use of them. We web-scraped these descriptions from (Morse and Weintraub, 2019). Unfortunately, in 1940 these descriptions rarely distinguish reservations. In contrast, in 1930 these descriptions were much richer. In 1930, the official descriptions contained explicit references to being on a reservation for 513 enumeration districts covering 100 reservations.

Assigning Reservation in 1930: Our final location assignment is based reservation information from five sources. First, we manually searched the enumeration district descriptions that we web-scraped from Morse and Weintraub (2019) and assigned a specific reservation if it was listed in the location description. These descriptions identified 513 enumeration districts that connected to 100 unique reservations. Second, we used the spatial overlap from the PLSS. Third, we applied the spatial overlap and adjacency from the APIs. We restricted ourselves to assignments with a high degree of precision. Fourth, for cases where there was only one reservation present in the county, we assigned every resident of the county to the reservation. After sequentially applying the first four steps, we were left with 309 unassigned townships, which we manually matched to reservations or specified as outside of a reservation boundary.

The process resulted in a reservation assignment for nearly 330,000 of the 402,000 individuals living in households with a Native American. Table 2 shows the distribution of reservation assignment by source. The reservation population comes from nearly 330 different reservations. Over 2/3 of the off-reservation population is comprised of non-Natives. Among the Native American population, over 90 percent were still living on or adjacent to reservations in 1930.

Assigning Reservation in 1940: The 1940 Census does not contain the township information nor the reservation specific descriptions we used to assign reservations in 1930. However,

[9] There were again differences in data availability between different server versions. `ENUMDIST` is available on the IPUMS/NBER version in 1940, but not in 1930. Fortunately, `ENUMDIST` was available for 1930 on the *Center for Population Research* (CCPR) server at UCLA version.

Table 2: Reservation Assignment by Source and Decade

Source	1930	1940
Enumeration District Descrip	85,563	97,046
PLSS Overlap	20,071	22,212
Township API Overlap	116,822	78,843
Adjacent Township API	72,083	53,295
Single County	16,092	12,141
Hand Link	16,356	13,112
Total Reservation Population	326,987	276,649
Off-Reservation Population	75,225	83,178
Total Population	402,212	359,827

we constructed a crosswalk between enumeration districts in 1930 and 1940 with content web-scraped from Morse and Weintraub (2019). This `ENUMDIST-1930`-to-`ENUMDIST-1940` crosswalk allows us to map the 1930 reservation assignment at the enumeration district level to enumeration districts in 1940. The last column of Table 2 summarizes the results of the enumeration district crosswalk assignment.

4 Two Applications

In this section, we apply the stable reservation-to-individual crosswalk to answering some long-standing questions on within-reservation inequality.

We investigate two questions, the first pertains to the differential formation-experience of reservations, namely whether reservations were forcibly formed out of formerly politically disconnected tribal bands, the second pertains to the differential experience at the end of the assimilation era, namely the adoption of the Indian Reorganization Act (IRA).

The Indian Reorganization Act of 1934 is often viewed as the end of the assimilation era because it ended allotment, but it also included provisions for tribes to adopt tribal government constitutions that were modeled on corporate and municipal charters. These constitutions typically did not conform to the traditional political structures of tribes, included too few checks and balances, and accentuated sub-tribal factions on reservations (Cornell and Kalt 2000, 2005; Treuer 2012, p144-145; Dippel 2014; Akee et al. 2015; **?**).

References

Akee, R. (2019). Land titles and dispossession: Allotment on american indian reservations. *Journal of Economics, Race, and Policy*, 1–21.

Akee, R., M. Jorgensen, and U. Sunde (2015). Critical Junctures and Economic Development–Evidence from the Adoption of Constitutions Among American Indian Nations. *Journal of Comparative Economics 43*(4), 844–861.

Anderson, T. L. and D. P. Parker (2008). Sovereignty, Credible Commitments, and Economic Prosperity on American Indian Reservations. *The Journal of Law and Economics 51*(4), 641–666.

Anderson, W. (2020). Politics, Economics, and Native American Conflicts. *University of Michigan mimeo*.

Bureau of Justice Statistics (2019). American Indians and Crime. https://www.bjs.gov/content/pub/pdf/aic.pdf. Accessed: 2019-11-08.

Carlson, L. A. (1978). The Dawes Act and the Decline of Indian Farming. *The Journal of Economic History 38*(1), 274–276.

Carlson, L. A. (1981). *Indians, Bureaucrats, and Land: the Dawes Act and the Decline of Indian Farming*. Number 36. Praeger Pub Text.

Center for Disease Control (2019). Health of American Indian or Alaska Native Population. https://www.cdc.gov/nchs/fastats/american-indian-health.html. Accessed: 2019-11-08.

Cornell, S. and J. P. Kalt (2000). Wheres the glue? institutional and cultural foundations of american indian economic development. *The Journal of Socio-Economics 29*(5), 443–470.

Cornell, S. and J. P. Kalt (2005). Cultural evolution and constitutional public choice: institutional diversity and economic performance on american indian reservations. In *Uncertainty and Economic Evolution*, pp. 126–152. Routledge.

Dippel, C. (2014). Forced Coexistence and Economic Development: Evidence from Native American Reservations. *Econometrica 82*(6), 2131–2165.

Dippel, C. and D. Frye (2020). The Effect of Land Allotment on Native American Households During the Assimilation Era. *UCLA mimeo*.

Familysearch (2019). United States Census Indian Schedules. https://wiki.familysearch.org/en/United_States_Census_Indian_Schedules. Accessed: 2019-11-08.

Feir, D., R. Gillezeau, and M. Jones (2019). The Slaughter of the Bison and Reversal of Fortunes on the Great Plains. *R&R at the Review of Economic Studies*.

Feir, D. L. (2016). The long-term effects of forcible assimilation policy: The case of indian boarding schools. *Canadian Journal of Economics 49*(2), 433–480.

Frye, D. and D. Parker (2019). Local versus Central Governance: Long Run Effects of Federal Oversight over American Indian Reservations. *Vassar College mimeo*.

Golenko, A. (2010). Did allotment contribute to the diminishment of american indian incomes? a study of allotment policies and american indian incomes during 1912-1920.

Gregg, M. T. (2018). The long-term effects of american indian boarding schools. *Journal of Development Economics 130*, 17–32.

Leonard, B., D. P. Parker, and T. L. Anderson (2020). Land quality, land rights, and indigenous poverty. *Journal of Development Economics 143*.

Miller, M. (2015). Dawes cards and indian census data. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 48*(4), 214–229.

Miller, M. C. (2016). the righteous and reasonable ambition to become a landholder: Land and racial inequality in the postbellum south. *Review of Economics and Statistics*, 1–45.

Morse, S. P. and J. D. Weintraub (2019). The Unified Census Enumeration District Finder 1880–1940. https://stevemorse.org/census/unified.html. Accessed: 2019-11-08.

National Archives (2019). Native Americans in the Census, 1860-1890. https://www.archives.gov/publications/prologue/2006/summer/indian-census.html. Accessed: 2019-11-08.

Office of Indian Affairs (1935). Indian Land Tenure, Economic Status, and Population Trends. In *Part X of the Report on Land Planning*. Washington: Department of Interior.

Office of Indian Affairs (1941). *Statistical Supplement to the Annual Report of the Commissioner of Indian Affairs for the Fiscal Year Ended June 30, 1941*. Department of Interior.

Siczewicz, P. (2011). U.S. Historical Counties. Available online from https://publications.newberry.org/ahcbp/.

Treuer, D. (2012). *Rez Life: An Indian's Journey Through Reservation Life*. Grove/Atlantic, Inc.

Treuer, D. (2019). *The Heartbeat of Wounded Knee: Native America from 1890 to the Present*. Riverhead Books.

US Census Bureau (1951). Statistical abstract of the united states. *US Department of Commerce, Washington: Government Printing Office*.