Appendix B: Native Areas Geography and Data Sources[1]

The IDDA Native Areas Supplement provides statistics derived from IRS Forms W-2 and 1040 within an aggregate geography that includes all American Indian and Alaska Native/Native Hawaiian (AIANNH) areas as delineated in 2017. The Census Master Address File identifies whether a given address (MAFID) falls within a Census block that is designated with a 4-character AIANNH area census code (aiannhce code). The aiannhce code specifies individual non-overlapping reservations, trust lands, and other legal and statistical entities as defined by the Census Bureau (Table B1).[2]

Table B1: Definitions for AIANNH Areas Included in the Aggregate Native Areas Geography

Area	Census Bureau Definition
American Indian reservations (Federal)	Areas that have been set aside by the United States for the use of tribes and whose boundaries are defined by tribal treaties, agreements, executive orders, federal statutes, secretarial orders, or judicial determinations. The Census Bureau recognizes federal reservations (and associated off-reservation trust lands) as territory over which American Indian tribes have primary governmental authority.
Off-reservation trust lands	Areas for which the United States holds title in trust for the benefit of a tribe (tribal trust land) or for an individual American Indian (individual trust land).
Hawaiian Home Lands	Areas held in trust for Native Hawaiians by the State of Hawaii.
Oklahoma Tribal Statistical Areas	Statistical areas identified and delineated by the Census Bureau in consultation with federally recognized American Indian tribes that had a former reservation in Oklahoma.
Alaska Native Village Statistical Areas	Statistical geographic entities representing permanent and/or seasonal residences of Alaska Natives who are members of, or receive governmental services from, the defining Alaska Native village. ANVSAs are intended to include only an area where Alaska Natives, especially members of the defining Alaska Native Village, represent a substantial proportion of the population during at least one season of the year.
Tribal Designated Statistical Areas	Statistical entities identified and delineated for the Census Bureau by federally recognized American Indian tribes that do not currently have a federally recognized land base (reservation or off-reservation trust land). Generally encompasses a compact and contiguous area that contains a concentration of individuals who identify with a federally recognized American Indian tribe and in which there is structured or organized tribal activity.
American Indian reservations (State)	Reservations established by some state governments for tribes recognized by the state.
State Designated Tribal Statistical Areas	Statistical entities for state-recognized American Indian tribes that do not have a state-recognized land base (reservation). Generally encompasses a compact and contiguous area that contains a concentration of individuals who identify with a state recognized American Indian tribe and in which there is structured or organized tribal activity

[1] The opinions and conclusions expressed here are those of the authors should not be interpreted as reflecting the views of the U.S Census Bureau, the Federal Reserve Board of Governors, the Federal Reserve Bank of Minneapolis, or any other person associated with the Federal Reserve System. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7511151; Disclosure Authorization Numbers CBDRB-FY23-0277, CBDRB-FY23-0373, CBDRB-FY23-CES014-019, and CBDRB-FY23-CES014-016.

[2] Definitions are taken from the glossary of Census geographic programs and products, publicly available on the Census website: https://www.census.gov/programs-surveys/geography/about/glossary.html

The aiannhce code takes a value of 9999 if the Census block is not designated as a Native area. We convert this information to a 0/1 variable which flags whether individuals in the tax sample resided in a Native area in a given year based on their MAFID in that year.

Demographic variables are drawn from the same sources as in the core IDDA dataset with the exception of race/ethnicity. The core dataset reports income statistics for individuals who identify as non-Hispanic American Indian or Alaska Native only and for individuals who identify as Native Hawaiian or Pacific Islander only. The Native Areas Supplement uses a more expansive notion of race and Native identity that includes all individuals who identify one of their races as American Indian or Alaska Native or Native Hawaiian or Pacific Islander, regardless of Hispanic ethnicity. We use self-reported primary and secondary races from the most recent Decennial Census (2010 and 2000) or American Community Survey a record appears in. If a record does not appear in either the American Community Survey or Decennial Census, we use the Census Best Race file which draws information from a variety of administrative sources to determine a single race and Hispanic ethnicity for individuals in the Census Bureau data system.

Sample Selection

The Native areas statistics are produced from a subset of the final, valid W-2 and 1040 samples used to construct the core IDDA dataset. In the cross-sectional data, they are produced from the subset of W-2 or 1040 records whose address identifier links to an aiannhce code for a native area in the given year. Records that do not link to a valid mafid or whose mafid is missing an aiannhce code are excluded from analysis.

Like in the core dataset, the native areas supplement includes income dynamics over 1- and 5-year time horizons. Records appear in the longitudinal files if the individual is in the valid W-2 (or 1040) sample in one of the two years and reside in a native area in one of the two years. Thus, individuals can migrate in/out of the native areas sample similarly to the way that individuals can age in/out of the prime-aged working sample in the core dataset. Individuals are considered “out-of-sample” if they are in the relevant tax sample but do not reside in a native area, and their income values in that year are set to missing. Individuals that are not in the relevant tax sample are considered “missing” in that year regardless of whether they live in a native area. Statistics in the native areas supplement do not include a prime-aged working subsample.

Income variables

Three income concepts are available in the Native areas supplement: household adjusted gross income and household nonwage income from Form 1040 and individual total compensation from Form W-2 (see section 3).

As in the overall W-2 and 1040 samples, some individuals in the native areas sample are affiliated with a tribe and some are not. IRS Publication 5424 “Income Tax Guide for Native American Individuals and Sole Proprietors” provides detailed income reporting instructions for individuals with tribal affiliations. The following sources of income may be particularly relevant in the Native Areas sample:

- Schedule C self-employment income is included in household nonwage income.

- The taxable portion of Alaska Permanent Fund dividends are included in household nonwage income.
- Per-capita distributions (for example, dividends from trust lands or gaming activity) made by tribal governments are included in household nonwage income.
- General welfare payments made by tribal governments to individuals are not reported on Form 1040.

Statistics modules

The Native Areas supplement includes the same five statistics modules as the core dataset: Percentiles of Income, Top Income Shares, Top Income Population Shares, Income Change Distributions, and an augmented version of the Income Transition Matrix module that includes migration. Like in core IDDA, statistics in the native areas supplement exist at the *ktlg* level, where

• k is a sample and corresponding income concept (in this case, the native areas-household-1040 or native areas-individual-W-2 sample),
• t is the time horizon (1998 to 2019 + 1- and 5-year horizons),
• l is the geography (native areas aggregate geography), and
• g is a demographic group (overall, age, sex, Native identity, or an intersection).

All modules adhere to the rule that statistics are always calculated within k, t, and l. That means all income bins and cutpoints—for example, when population shares are reported for the top 10 percent of the income distribution—are taken within the native areas geography and sample, not across the full U.S. population. The statistics included in the five modules are defined in section 4.

Augmented Transition Matrix

In the main IDDA dataset, the Income Transition Matrix Module gives the probability than an individual in a given demographic group and initial income bin moves to a different income bin, or out of the W-2 or 1040 sample, over 1 or 5 years (from y0 to y1). Geography is defined in the base year, so an individual who moves from state A to B is included in the transition matrix for state A and their subsequent year income is placed within a quartile based on the distribution in state A.

The native areas transition matrices track migration as well as income mobility. Individuals are included in the transition matrix if they reside in a native area in either the base or subsequent year, and their income quartile is considered “out-of-sample” in the year they do not reside in a native area. These augmented transition matrices provide three main types of probabilities:

- The probability that an individual in a particular demographic group and initial income quartile is in another quartile of the native areas income distribution, or not in the relevant tax sample, in y1.
- The probability that an individual starting in a given quartile of the native areas income distribution does not live in the native areas geography in y1.
- The distribution across y1 income quartiles of recent movers to the native areas geography.

Table B2 summarizes the interpretation for each of these types of transitions, as well as transitions in and out of the W-2 and 1040 data.

Table B2: Native Areas Transition Matrix

pctl_y0	pctl_y1	In y0 native areas sample?	In y1 native areas sample?	Interpretation
miss	lt25-gt75	No	Yes	The individual is not in the relevant tax sample in y0, and may or may not live in a native area. In y1, the individual is in the tax sample, lives in a native area, and has income in quartile pctl_y1 of the income distribution in native areas in y1.
out	lt25-gt75	No	Yes	The individual is in the relevant tax sample in y0, but does not live in a native area. In y1, the individual is in the tax sample, lives in a native area, and has income in quartile pctl_y1.
out	miss	No	No	The individual is in the relevant tax sample in y0, but does not live in a native area. In y1, the individual lives in a native area and is in the IRS dataset but not the relevant tax sample.
out	out	No	No	Not reported
miss	miss	No	No	Not reported
miss	out	No	No	The individual is not in the relevant tax sample in y0, but is in the IRS dataset and lives in a native area. In y1, the individual is in the relevant tax sample but does not live in a native area.
lt25-gt75	out	Yes	No	In y0, the individual is in the tax sample, lives in a native area, and has income in quartile pctl_y0 of the income distribution in native areas in y0. In y1, the individual is in the tax sample but not in a native area.
lt25-gt75	miss	Yes	No	In y0, the individual is in the tax sample, lives in a native area, and has income in quartile pctl_y0 of the income distribution in native areas in y0. In y1, the individual is not in the tax sample and may or may not live in a native area.
lt25-gt75	lt25-gt75	Yes	Yes	In y0, the individual is in the tax sample, lives in a native area, and has income quartile pctl_y0 of the income distribution in native areas in y0. In y1, the individual is in the tax sample, lives in a native area, and has income quartile pctl_y1 of the income distribution in native areas in y1.

Coverage

Table B3 provides availability rates for statistics in the IDDA native areas supplement by module and sample. Statistics are suppressed using the same method and minimum cell size as the core dataset. The "income levels" availability rate is the percent of total defined statistics available in the percentiles of income, top income shares, and top income population shares modules.

Table B3: Coverage in IDDA Native Areas Supplement

	1998-2004		2005-2019	
Native areas-1040	Defined	Available	Defined	Available
Income Levels	5,180	100%	11,100	100%
Transition Matrix	8,568	100%	14,688	100%
Income Changes	6,048	100%	10,368	100%
Native areas-W2				
Income Levels	-		7,170	98%
Transition Matrix	-		8,976	100%
Income Changes	-		6,336	100%

Coverage by race/ethnicity (Native vs. non-Native identity) and its intersections is 100 percent across all modules. The IDDA native areas supplement includes statistics by age, sex, Native identity, the intersection of age and Native identity, and the intersection of sex and Native identity. Table B4 shows the demographic disaggregations available by module and sample. The income concepts, demographic groups, and income percentiles generally match those included in the state-level IDDA data.

Table B4: Demographic Disaggregations Available in IDDA Native Areas Supplement

Static Measures:

Module	Sample and income concept (k)		Demographic disaggregations (g)	y0 quartiles+
Percentiles of Income	Native areas-Individual W-2	TC	xall, xaged, xaiannh, xsex, xagedXaiannh, xaiannhXsex	p = 10, 25, 50, 75, 90, 95, 98
	Native areas-Household-1040	GI, NW	xall, xaged, xaiannh, xagedXaiannh	p = 10, 25, 50, 75, 90, 95, 98
Income and Population Shares	Native areas-Individual W-2	TC	xall, xaged, xaiannh, xsex, xagedXaiannh, xaiannhXsex	p = 0 (across only), 90, 95, 98
	Native areas-Household-1040	GI, NW	xall, xaged, xaiannh, xagedXaiannh	p = 0 (across only), 90, 95, 98

Dynamic measures:

Module	Sample and income concept (k)		Demographic disaggregations (g)	y0 quartiles+	y1 quartiles+
Income Change Distributions	Native areas-Individual W-2	TC	xall, xaged, xaiannh, xsex	lt25, 25t50, 50t75, gt75	10, 25, 50, 75, 90, mean
	Native areas-Household-1040	GI, NW	xall, xaged, xaiannh	lt25, 25t50, 50t75, gt75	10, 25, 50, 75, 90, mean
Transition Matrix	Native areas-Individual W-2	TC	xall, xaged, xaiannh, xsex	miss, out, lt25, 25t50, 50t75, gt75	miss, out, lt25, 25t50, 50t75, gt75
	Native areas-Household-1040	GI, NW	xall, xaged, xaiannh	miss, out, lt25, 25t50, 50t75, gt75	miss, out, lt25, 25t50, 50t75, gt75