Estimating Gross Domestic Product for Tribal Nations: A Blended Administrative Record and Survey Approach

Carl Sanders
U.S. Census Bureau

Randall Akee*
Harvard University and NBER

Carla Medalia
U.S. Census Bureau

Cass Dorius
U.S. Census Bureau

Sallie Keller
U.S. Census Bureau

March 30, 2026

Abstract

Income measurement on American Indian Reservations is notoriously difficult, with small survey sample sizes and inconsistent administrative records for complete and timely coverage. Difficulties in measurement are one reason National Accounts do not measure or report income generated on American Indian Tribal Lands. Recent research has established that linking household survey and administrative tax record data to produce individual income statistics can mitigate measurement error from both sources. This paper may be the first to apply this innovative approach to examining income for individuals and businesses located on federally-recognized tribal reservations, a unique geographic population that spans the United States and has different demographic and socioeconomic characteristics than the nation as a whole. Using the American Community Survey linked to Internal Revenue Service W-2 and 1040 filings from 2021 to 2023, we examine discrepancies between survey and tax record reports of wages and income. First, building on research developed by the Census Bureau's National Experimental Wellbeing Statistics (NEWS), we compute and compare income estimates for residents of tribal reservations to the nation by socioeconomic and demographic characteristics. Second, we will calculate the flow of business revenues and estimated profits using the Census Longitudinal Business Database for non-governmental business establishments located on Tribal Nation lands for the same time period. This will establish a lower bound for two major components of GDP for American Indian Tribal Nation economies: annual income flows to individual households and income to firms located on tribal lands. We discuss why the Tribal Nation government sector will necessarily be excluded from this analysis. The proposed methods will provide the only estimates of Tribal Nation Gross Domestic Product on an annual and disaggregated basis.

Keywords: American Indians, Earnings, Measurement Error, Administrative Records, GDP, Tribal Governments
JEL classification: C80, D01, O15

Acknowledgments

Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. The U.S. Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-7503840 (P-7535195); Disclosure Review Board (DRB) approval number: CBDRB-FY25-SEHSD003-116).
*Corresponding author.

Email: rakee@hks.harvard.edu (Randall Akee)

1. Introduction

Income measurement on American Indian Reservations is notoriously difficult, with small survey sample sizes and inconsistent administrative record population coverage. For this and related reasons, National Accounts calculations in the United States do not release GDP estimates for Tribal Nations at any level of aggregation.[1] Identifying methods that improve the reporting and accuracy of income and other characteristics of historically underrepresented populations is a target of current U.S. Census Bureau research. In this paper, we examine how combining administrative record and survey information can help document income patterns on American Indian Reservations. In particular, we focus on American Indian Tribal Reservation lands in the lower 48 states for federally-recognized tribal nations, which make up approximately 56 million acres of land[2], over 2 percent of U.S. territory, with a total population exceeding 1 million individuals (Akee and Medford, 2025). Our analysis examines all prime working-age adults residing on these tribal lands, regardless of their tribal enrollment status, race, or ethnicity.

In this paper, we take the first step in the process of estimating the GDP of Tribal Nations using the Income approach to measurement (i.e. Gross Domestic Income[3]). This less-commonly used method of measuring economic activity is constructed by summing the total income associated with production: wages paid to workers, profits (net depreciation) paid to capital owners, and net taxes on production paid to governments. The standard method used to produce National GDP relies on data not available at the level of Tribal Nations often simply due to sample sizes: for example, the Monthly Wholesale Trade Survey[4] samples approximately 4,200 firms, providing only minimal coverage of businesses located on Tribal Nations. In contrast, the GDI approach used here to measure personal income exploits large household surveys (here, the American Community Survey) and universe-level administrative data set (IRS W-2 and 1040 records) that have significant coverage of individual wage and business income for workers and sole proprietors on Tribal Nations. For measurement of national accounts at smaller levels of aggregation, our approach can avoid pitfalls of expensive data requirements, and contributes to the literature on disaggregated economic accounts (Andersen et al., 2026).

In this work, we begin the process of estimating GDP/GDI by documenting the characteristics of personal income on Federally-recognized American Indian reservations in the lower 48 states. Our results compare the income estimates from the American Community Survey and IRS tax records for 36,000 on-reservation workers and 4.7 million off-reservation workers, allowing us to document measurement error in estimating the compensation received by employees and other individual income items for workers and households located on American Indian Reservations.

We proceed in several stages in our comparison of survey and administrative record income. First, we examine the differences in reporting of earnings (and later income) between administrative record and survey reports on- versus off-reservation. To do this, we link survey data from the American Community Survey (ACS) to administrative tax records from the Internal Revenue Service (IRS) for the period 2021 to 2023 to measure earnings, and income in future analyses. Once workers are matched, we estimate differences in administrative tax records and survey-reported ACS for the same individual in a given year, extending the U.S. Census Bureau's National Experimental Wellbeing Statistics (NEWS) estimates of measurement-error corrected income (Bee et al., 2023) into smaller geographies.

We find that on-reservations, mean survey and mean administrative records reports of earnings are approximately equal. In contrast, off-reservation IRS W-2 earnings tend to be about 5 percent higher than survey-reported ACS values. These averages mask a higher degree of survey-administrative records disagreement on tribal reservations: we find about 8 percent higher disagreement between survey and

[1]The lowest aggregation where the Bureau of Economic Analysis currently releases data is U.S. states, https://www.bea.gov/data/gdp/gdp-state.

[2]https://www.bia.gov/faqs/what-federal-indian-reservation

[3]See McCulla and Mead (2007) for a discussion of the methodological differences between GDP and GDI and https://www.bea.gov/data/income-saving/gross-domestic-income for recent National-level estimates.

[4]https://www.census.gov/wholesale/data/index.html

administrative records earnings reports for individuals living on-reservations than off-reservations when we look at differences either above or below the administrative records report. Further, on-reservation workers are 4 percentage points more likely to have a difference greater than 10% between their W-2 and the ACS log earnings reports. These initial findings will allow us to adjust future estimates of Tribal GDP for individuals and households.

Socioeconomic and demographic characteristics play an important role in differences in earnings reports between on and off-reservations. We find that there are larger absolute wage gaps for individuals with lower educational attainment or who are younger, female, and American Indians or Alaskan Natives (AIAN), which attenuate the relationship between on- and off- reservation status and wage gaps. In future work, we plan to examine the potential reasons for the various associations between individual demographics and the administrative record and survey earnings gap, including labor force attachment rates.

Our findings that there is more on-reservation disagreement than off-reservation disagreement between survey and administrative records of earnings have important implications for how earnings are measured on reservations. If administrative records are viewed as the "true" earnings measures as in Bollinger et al. (2019), this would suggest that survey earnings measurement error is greater on-reservation; whereas if both survey and administrative records are potentially measured with error as in Abowd and Stinson (2013), this would mean we are more uncertain about actual income levels on reservations than we previously believed. The next step of this paper consists in using the Bee et al. (2023) approach to corrected for measurement error and aggregating income to the reservation level, as well as including self-employment and other individual income components.

Future work, forthcoming in a seperate paper, will focus on the income flows to business establishments located on American Indian Tribal reservation lands. Building on previous work (Akee et al., 2018, 2019), we geocode the Integrated Longitudinal Business Database and the Longitudinal Business Database for Tribal Reservation locations and will estimate total revenue and estimated profits for non-governmental business establishments in a given year. Finally, we'll combine these flows with the data from individuals and household income in a given year to estimate a lower bound of Tribal Nation GDP for individual reservation locations. It is important to note that we may not necessarily release individual Tribal Nation GDP estimates, but instead, we may aim to disclose Tribal GDP at a region or state-level of aggregation. We will determine these final levels of aggregation in the process of our analysis in consulation with Tribal government officials and leaders.

2. Background and Literature Review

Extensive research in labor economics has studied measurement error in survey-reported wage and salary earnings and the possibility of reducing bias and improving precision by using administrative earnings records (Alvey and Cobleigh, 1975; Duncan and Hill, 1985; Bound and Krueger, 1991; Bound et al., 1994; Pischke, 1995; Bollinger, 1998; Bound et al., 2001; Roemer, 2002). Recent work has established that even income estimates derived from administrative records alone may be biased, and that using both administrative records and survey reports together can improve data quality (Bee et al., 2023, 2025; Medalia et al., 2019; Meyer and Mittag, 2015; Kapteyn and Ypma, 2007; Gottschalk and Huynh, 2010; Meijer et al., 2012; Abowd and Stinson, 2013; Bee et al., 2019; Imboden et al., 2019; Jenkins and Rios-Avila, 2023).

There are few studies that examine the differences in self-reported and administrative records reports for earnings or income data for Native Americans or other Indigenous Peoples. This is due, in large part, to the lack of large sample sizes for Tribal Areas in most surveys. Therefore, most studies cited above tend to examine measurement error at the national-level. Differences in sub-populations may either be completely omitted or averaged out in these larger sets of analyses. In one paper, Akee (2011) examines differences in earnings data for a relatively small population of workers in the Federated States of Micronesia using administrative data and self-reported earnings at the individual level. The finding indicates that, on average, in 1993 and 1999 there was over and under-reporting of earnings

relative to the administrative record reports. Additionally, controlling for prior earnings volatility (such as job loss or other shocks) helps to explain some of the reported errors in earnings; this argues that recall bias may play an important role in accuracy. Also, employment in specific sectors (government employment) or a longer tenure length with an employer resulted in lower reported differences in earnings reports.

This project follows the Statistical Products First innovation cycle used by the Census Bureau to develop new products that are designed from the beginning to meet data users' needs (Jarmin, 2024; Keller et al., 2024). The cycle begins with stakeholder engagement, which serves to identify the purposes and uses that statistical products must support. Data discovery follows, involving the identification and evaluation of relevant data sources across surveys, administrative records, and other domains. Next data and methodological gaps are identified that may require developing new approaches to data integration and modeling. A dissemination stages follows, focusing on translating these insights into usable statistical products. Finally, the value of these products is assessed through feedback and evaluation, informing subsequent iterations. Whereas this has been described as a sequence of activities, the steps in the SPF process are highly interactive.

Our implementation of the SPF cycle in this process is useful background. In the first step of the cycle, we engaged data users to understand what data they needed and how they intended to use it. Over the last two years, an interdisciplinary team from across the Census Bureau and academia attended tribal conferences and held numerous listening sessions and outreach events to better understand the data needs of tribal nations. In these conversations, we learned that tribal nations have less access to economic data about their communities than other groups, and this can put them at a competitive disadvantage when applying for grants, forecasting future needs, and evaluating the successes and gaps of their programs and services. Tribal leaders also indicated that a GDP-like measure could help tribes to better track their long-run changes in their economic conditions that don't already exist in a timely or accurately manner. To meet these needs, we identified the need for a GDP-like measure that is reproducible, scalable, comparable, and relatively simple to maintain over time.

In the second stage of the SPF process, we considered all of the information resources the Census Bureau can bring to bear to construct a Tribal GDP-like measure. We identified that existing GDP data is provided only at the national, state or county levels, and the GDP for tribal reservations are not calculated or publicly distributed. We determined that we would need to leverage the Census Bureau's Data Linkage Infrastructure (U.S. Census Bureau, 2025), which includes census and survey data, administrative records, third-party data, and public data sources, in order to develop a Tribal GDP-like measure.

In the third stage of the SPF process, we met with subject matter and data experts from throughout the Census Bureau, the Bureau of Economic Analysis, and academia. We discussed the potential methods for creating a new GDP-like measure and selected the income flow method, which measures all sources of income flows to a geographic unit. This includes income flows to 1) individuals, 2) business establishments, and 3) governments.

In the next step, we focused on exploring solutions and decided to examine individual income flows first, followed by income flows to business establishments. At this time, we we will not estimate tribal government incomes given tribal sovereignty and desired confidentiality from tribal nations (Carroll et al., 2019). As part of this process, we partnered with the Census Bureau's NEWS project to link the administrative records to survey microdata within the Census Bureau's Data Linkage Infrastructure.

In the final two steps of the SPF innovation cycle, we consider how to disseminate the statistical products we are developing and get feedback on the effectiveness and scalability of our products. The work in this paper represents a pilot project that is part of this larger effort to produce a Tribal GDP-like measure.

3. Data and Methods

3.1. Data

We use a "blended data" approach to measuring income, linking individual and household survey data with administrative records.

For the survey data, we use Title 13-protected microdata from the American Community Survey (ACS), the largest household survey conducted by the Census Bureau. The ACS is a nationally representative survey with coverage of tribal reservation lands. To ensure we have adequate sample sizes for the analyses, we use combined 1-year ACS data from 2021-2023. The ACS questionnaire, under the heading "Income in the past 12 months", asks income for each household member in a number of income categories. In particular, for each member of the household, the respondent fills in a field to "give your best estimate of the total amount [of income for this person] during the past 12 months". We use the response to the category "Wages, salary, commissions, bonuses, or tips from all jobs. (Report amount before deductions for taxes, bonds, dues, or other items)." Unless otherwise noted, all results use the ACS final person weights to make the numbers representative of the U.S. population.

For the administrative records, we use Form W-2 information returns records from the IRS[5]. Currently, we have access to the Box 1 (Wages, tips, other compensation) entries for the universe of W-2s. To be consistent with the survey data, we use these administrative records corresponding to tax years 2021-2023. For workers with multiple W-2s, we first drop any W-2s identified as duplicates and then sum across the forms to get total W-2 earnings.

There are expected differences between the ACS earnings report and the W-2 Box 1 numbers. For one, the reference periods are different: ACS earnings are asked for the last 12 months, with the survey administered throughout the year, whereas W-2s are specific to a tax (calendar) year. Second, the ACS question is explicitly pre-deductions, while Box 1 excludes employer-paid health insurance premiums; employer contributions to Flexible Spending Accounts, Health Savings Accounts, or dependent care benefits; some retirement plan contributions, and a variety of other (typically smaller) excludable items.

Individual-level records are linked using the Protected Identification Key (PIK). These unique and anonymous identifiers are assigned either probabilistically through the Person Identification Validation System (PVS) (U.S. Department of Commerce, 2022; Wagner and Layne, 2014) or deterministically through the Quick PIK process (U.S. Department of Commerce, 2022; Finlay, 2024). PIKs allow records from the ACS to be linked at the individual-level to records from the tax data and other administrative records.

We obtain geographic information for addresses in the survey or administrative records from the Census Bureau's Master Address File (National Academies of Sciences, Engineering, and Medicine, 2023), a frame of all addresses in the United States and includes information about living quarters. Finally, we use a crosswalk dataset produced by the Census Bureau and links American Indian Tribal Reservation lands to other Census geographies, including state, county, and census tract.

The analytical sample is limited to the prime working age population, aged 25-54, who reside at residential addresses and not Group Quarters (GQs), and all following results implement this restriction. Table 1 shows that across the three years in total, the resulting sample size for Off-Reservation areas is just under 4.7 million individuals and is 36,000 for the On-Reservation areas. The weighted percentage of individuals who are successfully assigned PIKS is quite high for both geographic groups: 93.1 percent of the observations are assigned a PIK for the Off-Reservation areas and the PIK rate for the On-Reservation population is 89.3 percent. We also show that 74.6 percent of people living in Off-Reservation areas receive a W-2 and approximately 64.7 percent of people living On-Reservation receive a W-2. The total population living on Tribal reservations was approximately 1,070,561 according to the 2022 American Community Survey data for all races (Akee and Medford, 2025).

[5]Future versions of this paper will include results from additional tax forms such as the 1040, 1099-Gs, and 1099-Rs.

Table 1: Unweighted Count and weighted sample match statistics

Variable	Off-Reservation	On-Reservation
Unweighted ACS Count	4,693,000	36,000
Weighted PIK Rate	0.931	0.893
Weighted W-2 Receipt	0.746	0.647

Source: 1-Year American Community Survey linked to Internal Revenue Service W-2 file, 2021-2023.
Note: Approved for release under DMS: P-7503840 (P-7535195) and DRB: CBDRB-FY25-SEHSD003-116.

3.2. Independent variables

We use both individual- and household-level characteristics from the ACS to identify socioeconomic differences between on- and off-reservation areas that may contribute to differences in reported wages. Educational attainment is measured as 5 discrete categories: no high school, high school degree, some college, bachelor's (BA) degree, and master's (MA) degree or higher. Age is measured as an continuous variable, ranging from 25 to 54. Sex is defined as being either male or female. We include one race variable indicating whether or not the individual is an American Indian or Alaskan Native (AIAN). The total number of children in the household is also included as a continuous variable.

From IRS form W-2, we create include an independent variable indicating whether or not the individual received a W-2. Future analyses will extend this work and include independent variables from IRS forms 1040 and 1099-R indicating whether or not the individual and/or household filed that form.

Table 2 provides the average characteristics of the off and on-reservation population. Concerns about non-random assignment of PIKs (Bond et al., 2014) lead us to include sample information with measures separated into workers who received PIKs and those who did not. The first five rows provide mutually-exclusive educational attainment categories. On average, the off-reservation population that has been assigned a PIK are more likely to be represented in the upper educational categories of Some College and higher. Alternatively, this means that the individuals who reside off of reservation lands and have not been assigned a PIK are more likely to be represented in the lower educational attainment categories such as No High School and High School. This same condition holds for the on-reservation population as well.

The next categories in this table provide some demographic characteristics of these populations. We find that there is little discernible difference in the average ages of individuals residing either on or off reservations and whether they have been assigned a PIK or not. However, it does appear that males are more likely to be represented in the No PIK group for both the on and off-reservation populations. American Indians residing on reservations are more likely to be represented in the No PIK group; but they are equally likely to be in either the PIK or No PIK categories for the off-reservation population. Finally, the differences in the number of kids does not appear to differ too much among these four different populations.

3.3. Dependent variables

Our outcomes of interest include three different measures of wage gaps for individuals defined by comparing ACS wage income to wages reported on the W-2. First, the "wage gap" is the log total W-2 earnings for tax year t minus their log reported ACS earnings over the last 12 months measured sometime during year $t + 1$. Second, "absolute wage gap" is defined by taking the absolute difference of the log wage gap. Third, we create a dummy variable indicating whether the absolute difference of the log wage gap is greater than or equal to 0.10 log points.

Future analyses will explore differences in income gaps, in addition to wages. To do this, we will compare total money income, as defined in the ACS, to the corresponding boxes on IRS Form 1040. We will also explore differences in retirement income between the ACS and IRS Form 1099-R.

Table 2: Proportions for the On or Off-Reservation Populations and PIK Match Categories

	Off-Reservation		On-Reservation	
Variable	**Has PIK**	**No PIK**	**Has PIK**	**No PIK**
No High School	0.079	0.246	0.115	0.181
High School	0.228	0.301	0.327	0.426
Some College	0.284	0.213	0.353	0.289
BA degree	0.254	0.165	0.138	0.073
MA or More	0.155	0.075	0.067	0.030
Age	39.220	38.620	39.360	38.660
Male	0.496	0.523	0.497	0.510
AIAN Race	0.025	0.029	0.477	0.701
Number of Kids	0.963	0.922	1.260	1.257

Source: 1-Year American Community Survey linked to Internal Revenue Service W-2 file, 2021-2023.
Note: All variables are weighted with ACS person weights. Approved for release under DMS: P-7503840 (P-7535195) and DRB: CBDRB-FY25-SEHSD003-116.

3.4. Empirical Methodology

We aim to estimate the relationship between the administrative records/survey earnings gap on on- versus off-reservation status while controlling for individual- or household-level characteristics. The regression specifications for the wage gap and absolute wage gap are shown in Equation 1 and Equation 2, respectively, and the specification for the probability of the absolute wage gap is greater than or equal to 0.10 log points is shown in Equation 3, where Λ is the logistic function.

$$E[WageGap_i|X, s, t] = \alpha_0 + X'\beta + \theta_s + \theta_t \tag{1}$$

$$E[AbsWageGap_i|X, s, t] = \alpha_0 + X'\beta + \theta_s + \theta_t \tag{2}$$

$$Pr(|AbsLogDiff|\geq 0.1|X, s, t) = \Lambda(\alpha_0 + X'\beta + \theta_s + \theta_t) \tag{3}$$

The three outcomes for the three different wage gap measures are at the individual level in a given year for the dependent variables. We include state and year fixed effects, denoted by θ_s and θ_t, respectively. Equations 1 and 2 are estimated by OLS, and equation 3 is estimated using logistic regression. All regressions use the ACS internal-use person weights and all standard errors are clustered at the county-level following Abadie et al. (2023). The coefficients of interest, as defined in Subsection 3.2, are in the vector β.

4. Results

First, we consider the extent to which our demographic and socioeconomic variables in our X's differ systemically between on- and off-reservation workers. Again, our sample is restricted to individuals 25-54 who live in the lower 48 states.

In Table 3, we show the estimated marginal effect of being on-reservation for each component of our X control variables. These marginal effects and associated p-values are estimated with the ACS person weights, standard errors clustered at the county level, and state and year fixed effects included. In Panel A, we show the estimated coefficients for five mutually-exclusive categories of educational attainment. In general, we find that the on-reservation population is more represented in the lower educational categories including less than a high school degree, a high school degree and some college. The on-reservation population has a lower probability of being represented in the BA degree and the MA or more categories of educational attainment. These differences are all statistically significant at the 0.1 percent level.

In Panel B, we show the estimated differences between on- and off-reservation level for our demographic characteristics. The average age is not statistically significantly different between the on or off reservation populations; the average age represented in our data is approximately 39 years. The gender distribution is also approximately similar for the on and off reservation and is on average about 50 percent. American Indians are 45 percentage points more likely to be more highly represented for the on-reservation population. The on-reservation population is also more likely to have more children on average than the off-reservation population. Finally, the on-reservation population are less likely to have W-2 forms.

Table 3: Differences in Control Variables On- versus Off-Reservation: Regression Results

Panel A: Education Variables

	No HS	**HS**	**Some College**	**BA**	**MA+**
On Reservation	0.036	0.112	0.038	-0.111	-0.075
	(0.008)	(0.013)	(0.009)	(0.012)	(0.008)
P-Value	≤ 0.001	≤ 0.001	≤ 0.001	≤ 0.001	≤ 0.001
Off Reservation Mean	0.088	0.233	0.281	0.249	0.150

Panel B: Demographic and Other Variables

	Age	**Male**	**AIAN**	**No. Kids**	**W-2**
On Reservation	0.310	-0.007	0.452	0.251	-0.113
	(0.135)	(0.005)	(0.071)	(0.039)	(0.023)
P-Value	0.022	0.104	≤ 0.001	≤ 0.001	≤ 0.001
Off Reservation Mean	39.180	0.498	0.026	0.967	0.754
Off-Reservation Count			4,693,000		
On-Reservation Count			36,000		

Source: 1-Year American Community Survey linked to Internal Revenue Service W-2 file, 2021-2023. All specifications are OLS regressions with the column name as the dependent variable and an on-reservation dummy as the independent variable. We include state and year fixed effects, person weights, and standard errors are clustered at the county level. Approved for release under DMS: P-7503840 (P-7535195) and DRB: CBDRB-FY25-SEHSD003-116.

In the Table 4, we evaluate the differences between the linked administrative record reports and self-reported earnings for the prior year. In the first column of the table, we use the difference between administrative records and the ACS survey-reported earnings. We control for state and year fixed effects; the data is weighted at the person level and we cluster the standard errors at the county level for all regressions.

We find that the coefficient on on-reservation status is -0.042 log points; this indicates that for individuals residing on American Indian reservation lands, on average, the difference in W-2 reported earnings and ACS survey reports are about 4 percent lower than the corresponding difference for off-reservation populations. This estimated difference is statistically significant at the 1 percent level. Given that the off-reservation mean difference in the W-2 and ACS earnings reportings is 0.049, we can conclude that the mean difference for the on-reservation population is 0.007 or effectively zero. This means that, on average, on-reservation populations are reporting approximately similar earnings as compared to what is shown in the administrative data.

In the next column, we provide the same analysis for the absolute difference between the ACS self-report of earnings and the W-2 administrative record report. The estimated coefficient for being located on an American Indian reservation location is 0.081 log points, so the disagreement between W-2 and ACS reports is significantly higher on reservation. This indicates that the simple raw difference measure obscures positive and negative disagreements, which cancel each other out in aggregate for reservation residents. The estimated coefficient is statistically significant at the 1 percent level and we note that the off-reservation average for the absolute difference in the two reports of wages is 0.475 log points.

In the third column, the outcome variable is the probability that the absolute log difference in the two earnings reports will differ by more than 10 percent. The estimated coefficient in this regression shows that individuals residing on reservation lands are more likely to have an absolute difference in the self-reported and administratively reported earnings that is greater than 10 percent. These results are also statistically significant at the 1 percent level.

On average, individuals residing on reservation lands report similar averages across the self-reported and administrative reports of earnings, in contrast with those residing off-reservation. However, the individual-level differences have larger absolute magnitude on-reservation. In the next set of analyses, we include additional control variables to identify which may be associated with larger or smaller gaps in earnings reports.

Table 4: Regressions Comparing Wage Reports between ACS and W-2 by On-/Off-American Indian Reservation

Dep Variables:	**Wage Gap**	**Abs Wage Gap**	**1(|Abs Log Diff|≥0.1)**
On Reservation	-0.042	0.081	0.041
	(0.013)	(0.022)	(0.012)
P-Value	0.001	≤ 0.001	≤ 0.001
Off Reservation Mean	0.049	0.475	0.703
On Reservation Mean (Est)	0.007	0.556	0.743
Off-Reservation Count		4,693,000	
On-Reservation Count		36,000	

Source: 1-Year American Community Survey linked to Internal Revenue Service W-2 file, 2021-2023. Note: All variables are weighted. Approved for release under DMS: P-7503840 (P-7535195) and DRB: CBDRB-FY25-SEHSD003-116, CDRB-FY26-SEHSD003-040. We include state and year fixed effects, person weights, standard errors are clustered at the county level. "Wage Gap" is W-2 earnings minus ACS earnings; "Abs Wage Gap" is |wage gap|, and "1(|Abs Log Diff|> 0.1)" using a dependent variable that is a dummy for if the two wage reports are greater than 0.10 apart with this empirical specification using a logistic regression.

Table 5 summarizes the regression results from running an OLS regression of the absolute value of the log differences in the reported earnings on a number of individual educational and demographic

characteristics; this is the same outcome variable from column 2 in Table 4. The estimated coefficient for residing on an American Indian reservation is 0.035 log points, significant at the 5% level, but is less than half the size of the previous coefficient on "On Reservation" in Table 4, indicating that these other characteristics are associated with a relatively large amount of the error in earnings reports. However, even given the other control variables we still find an economically and statistically significant relationship between reservation status and disagreements across earnings reports.

The next four variables measure the highest level of educational attainment from the ACS. The omitted category is a high school degree. In general, we find that higher levels of educational attainment are associated with smaller absolute differences in log wages. This may be due to the fact that individuals may have more steady employment and therefore this would result in less variation year-to-year which results in smaller earnings error reporting. In a similar fashion, we also find that older individuals are less likely to misreport their annual earnings. Older individuals may also tend to have steady employment and earnings amounts which would also result in a reduction in differences in earnings reports.

We also find that males are associated with a reduction in the absolute size of the absolute difference in log earnings; once again, this may be attributable to steady employment and long-term labor force participation. Additionally, we find that individual who are AIAN are more likely to have a larger absolute difference in wages than non-AIAN individuals. Finally, we see that having more children is positively associated with a larger absolute earnings difference.

Table 5: OLS Regression of Absolute Value of Log Wage Gap with Demographic Controls

Dep Variable: Abs Value of Log Wage Gap	**Estimate**	**P-Value**
On Reservation	0.035	0.041
	(0.017)	
No High School	0.133	≤ 0.001
	(0.005)	
Some College	-0.086	≤ 0.001
	(0.002)	
BA Degree	-0.154	≤ 0.001
	(0.003)	
MA Degree or Higher	-0.186	≤ 0.001
	(0.003)	
Age	-0.020	≤ 0.001
	(0.001)	
Age Squared	0.000	≤ 0.001
	(0.000)	
Male	-0.034	≤ 0.001
	(0.002)	
AIAN Race	0.041	≤ 0.001
	(0.004)	
Number of Kids	0.016	≤ 0.001
	(0.001)	
Off-Reservation Count	4,693,000	
On-Reservation Count	36,000	

Source: 1-Year American Community Survey linked to Internal Revenue Service W-2 file, 2021-2023. Note: We include state and year fixed effects, person weights, standard errors are clustered at the county level. Omitted category: High School. Approved for release under DMS: P-7503840 (P-7535195) and DRB: CBDRB-FY25-SEHSD003-116.

Table 6 provides a regression of the outcome variable is the probability that the absolute log

difference in the two earnings reports will differ by more than 10 percent on a number of individual educational and demographic characteristics; this is the same outcome variable from column 3 in Table 4. In the first row, we find that residing on a reservation is not related to the earnings reports differing by more than 10 percent; the estimated coefficient, which having the same sign as before, is not statistically significant at even a 10% level once these additional controls are included. We find in the next four rows that higher educational attainment is associated with a lower probability of the earnings reports differing by more than 10 percent. Age, just as we found in the previous table, also is associated with a reduction in the probability that difference in earnings would more than 10 percent different. Male respondents are associated with a lower probability of the difference in earnings being greater than 10 percent apart. Finally, the number of children a person has is positively related to a higher probability of the difference in earnings reports being larger than 10 percent.

Table 6: The Probability that the Abs. Difference in Earnings Reports is Greater than 0.10: Logistic Regression Results

Dep Var: |Abs Log Diff|≥0.1	**Estimate**	**P-Value**
On Reservation	0.071	0.151
	(0.050)	
No High School	0.353	≤ 0.001
	(0.011)	
Some College	-0.275	≤ 0.001
	(0.005)	
BA Degree	-0.557	≤ 0.001
	(0.008)	
MA Degree or Higher	-0.662	≤ 0.001
	(0.010)	
Age	-0.080	≤ 0.001
	(0.003)	
Age Squared	0.001	≤ 0.001
	(0.000)	
Male	-0.082	≤ 0.001
	(0.005)	
AIAN Race	0.088	≤ 0.001
	(0.012)	
Number of Kids	0.049	≤ 0.001
	(0.002)	
Off-Reservation Count	4,693,000	
On-Reservation Count	36,000	

Source: 1-Year American Community Survey linked to Internal Revenue Service W-2 file, 2021-2023. Note: We include state and year fixed effects, person weights, standard errors are clustered at the county level. Omitted category: High School. Approved for release under DMS: P-7503840 (P-7535195) and DRB: CBDRB-FY25-SEHSD003-116.

5. Discussion

Our analysis is novel in linking administrative data and records to survey data for a relatively small, but important, geographic unit – American Indian reservation lands. We compare the differences in administrative and self-reported earnings to the off-reservation population as well. We find that

the on-reservation population (inclusive of all races and ethnic groups residing there) has relatively similar survey-reported and administrative record reports for earnings. On average, we find that the log wage gap is -0.042 relative to the off-reservation population which has an average wage gap of 0.049. Therefore, on average the difference in W-2 and ACS earnings at the individual level for the on-reservation population is approximately zero. This is a surprising finding, with possible sources for the difference including differences in the types of jobs worked and whether they issue W-2s; details of the ACS survey instrument on- versus off-reservation; or other currently unknown sources. For next steps, it we plan to expand the sample period and include more details about individuals workers and their jobs to see if these can explain these differences in the gaps.

On the other hand, we do note that there are larger amounts of disagreement in absolute earnings differences for the on-reservation population than for the off-reservation. These absolute differences in earnings reports are associated with a variety of individual-level characteristics such as educational attainment, gender, age, family size and individuals identifying as American Indian. While we do not have a causal explanation for all of these associations, it does appear that age and educational attainment may be related to the stability of employment and tenure in a particular position. Additionally, those individual characteristics would also be related to industry or occupation; there may be a differential composition of industry and occupations for an on-reservation population than for an off-reservation population. Specifically, if there is a higher proportion of part-time or "gig" employment for the on-reservation population than for the off-reservation population in general then this might explain why there are larger differences in earnings reports for the on-reservation population. Job stability and tenure are associated with a more constant level of earnings and therefore reporting is probably more consistent and accurate over time.

6. Conclusion

In this paper, we examined the extent to which self-reported earnings align with administrative reports of earnings for on-reservation populations. Our analysis finds that there are significant differences in the disagreement between survey and administrative record earnings for the on- versus off-reservation populations. These absolute differences in earnings are associated with lower educational attainment, younger ages, women and the American Indian population and individuals with more children. This analysis provides the first insight into the potential differences in earnings report for this geographic location. We have not had previous measures on how accurate the reporting of earnings were for this population previously. As a result, this analysis may inspire additional work that seeks to understand the degree to which the on-reservation population reports higher earnings than indicated from administrative records. It is possible that the on-reservation population, and American Indians in particular, are reporting other, potentially non-taxable sources of income as earnings in the American Community Survey. However, additional research will need to be conducted to examine this possibility.

This work is also the first step in a larger project that will attempt to combine several elements of Tribal Reservation Gross Domestic Product. Taking an "income" approach, we will assemble the individual (or household) incomes for the on-reservation population. We will include additional flows of income from business establishments located on reservation lands as well. While we will not be able to estimate Tribal Nation incomes, we will be able to provide a rough estimate of two key income flows on Tribal Reservation lands. We will examine, in future research, total money income by including tax data from IRS forms 1040 and 1099-R. These will serve as additional checks on the accuracy of income reporting for reservation geographies.

This project also demonstrates the Statistical Product First innovation cycle being used at the Census Bureau. Following this framework, we solicit input from stakeholders regarding the findings from this work in order to continue to develop statistical products to meet stakeholder needs.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics 138*(1), 1–35.

Abowd, J. M. and M. H. Stinson (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics 95*(5), 1451–1467.

Akee, Randall, E. M. and A. B. Medford (2025). Rsocial and economic changes in american indian reservations: A databook of the us census and the american community survey. Technical report, Harvard Kennedy School Project on Indigenous Governance and Development, Harvard University.

Akee, R. (2011). Errors in self-reported earnings: The role of previous earnings volatility and individual characteristics. *Journal of Development Economics 96*(2), 409–421.

Akee, R., E. Mykerezi, and R. M. Todd (2018). Reservation nonemployer and employer establishments: Data from us census longitudinal business databases. Technical report, Center for Economic Studies, U.S. Census Bureau.

Akee, R., E. Mykerezi, and R. M. Todd (2019). Opportunities to diversify: Reservation workplaces and job numbers compared to nearby county areas. *Creating Private Sector Economies in Native America: Sustainable Development through Entrepreneurship*, 37–63.

Alvey, W. and C. Cobleigh (1975). Exploration of differences between linked Social Security and Current Population Survey earnings data for 1972. *Proceedings of the Social Statistics Section, American Statistical Association*.

Andersen, A., K. Huber, N. Johannesen, L. Straub, and E. T. Vestergaard (2026). Disaggregated economic accounts. *The Quarterly Journal of Economics*, qjag010.

Bee, A., J. Creamer, J. Mitchell, N. Mittag, J. Rothboaum, c. Sanders, L. Schmidt, and M. Unrath (2025). National experimental well-being statistics technical note on methodological changes for version 2. Technical report, U.S. Census Bureau.

Bee, A., J. Mitchell, N. Mittag, J. Rothbaum, C. Sanders, L. Schmidt, and M. Unrath (2023). *National experimental wellbeing statistics*. U.S. Census Bureau.

Bee, A., J. Mitchell, and J. Rothbaum (2019). Not so fast? How the use of administrative earnings data would change poverty estimates. *Unpublished U.S. Census Bureau Working Paper*.

Bollinger, C. R. (1998). Measurement error in the current population survey: A nonparametric look. *Journal of Labor Economics 16*(3), 576–594.

Bollinger, C. R., B. T. Hirsch, C. M. Hokayem, and J. P. Ziliak (2019). Trouble in the tails? what we know about earnings nonresponse 30 years after lillard, smith, and welch. *Journal of Political Economy 127*(5), 2143–2185.

Bond, B., J. D. Brown, A. Luque, A. O'Hara, et al. (2014). The nature of the bias when studying only linkable person records: Evidence from the american community survey. *Center for Administrative Records Research and Applications Working Paper 8*, 2–30.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics 12*(3), 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. In *Handbook of econometrics*, Volume 5, pp. 3705–3843. Elsevier.

Bound, J. and A. B. Krueger (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of labor economics 9*(1), 1–24.

Carroll, S. R., D. Rodriguez-Lonebear, and A. Martinez (2019). Indigenous data governance: strategies from united states native nations. *Data science journal 18*, 31.

Duncan, G. J. and D. H. Hill (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. *Journal of Labor Economics 3*(4), 508–532.

Finlay, K. (2024). Justice outcomes explorer (joe) technical documentation. Technical report, U.S. Census Bureau.

Gottschalk, P. and M. Huynh (2010). Are earnings inequality and mobility overstated? The impact of nonclassical measurement error. *Review of Economics and Statistics 92*(2), 302–315.

Imboden, C., J. Voorheis, and C. Weber (2019). Measuring systematic wage misreporting by demographic groups. *Unpublished U.S. Census Bureau Working Paper*.

Jarmin, R. (2024). Reflecting on years past and looking forward to 2024.

Jenkins, S. P. and F. Rios-Avila (2023). Reconciling reports: modelling employment earnings and measurement errors using linked survey and administrative data. *Journal of the Royal Statistical Society Series A: Statistics in Society 186*(1), 110–136.

Kapteyn, A. and J. Y. Ypma (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economics 25*(3), 513–551.

Keller, S., S. Shipp, V. Lancaster, and J. Salvo (2024). Advancing data science in official statistics–what is the curated data enterprise?

McCulla, S. and C. I. Mead (2007). An introduction to the national income and product accounts. *Methodology Papers: US National Income and Product Accounts, Bureau of Economic Analysis, US Departament of Commerce*.

Medalia, C., B. D. Meyer, A. B. O'Hara, and D. Wu (2019). Linking survey and administrative data to measure income, inequality, and mobility. *International journal of population data science 4*(1), 939.

Meijer, E., S. Rohwedder, and T. Wansbeek (2012). Measurement error in earnings data: using a mixture model approach to combine survey and register data. *Journal of Business & Economic Statistics 30*(2), 191–201.

Meyer, B. D. and N. Mittag (2015, October). Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness and holes in the safety net. Working Paper 21676, National Bureau of Economic Research.

National Academies of Sciences, Engineering, and Medicine (2023, November). *Assessing the 2020 Census: Final Report*, Chapter 5. Master Address File. Washington, D.C.: National Academies Press.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the psid validation study. *Journal of Business & Economic Statistics 13*(3), 305–314.

Roemer, M. (2002). Using administrative earnings records to assess wage data quality in the March Current Population Survey and the Survey of Income and Program Participation. *U.S. Census Bureau Center for Economic Studies Working Paper*.

U.S. Census Bureau (2025). Data linkage infrastructure.

U.S. Department of Commerce (2022). Best practices for monitoring and evaluating the arp, iija and other programs: Report of the department of commerce data governance working group.

Wagner, D. and M. Layne (2014). The person identification validation system (pvs): Applying the center for administrative records research and applications' (carra) record linkage software.