Tax Alternative Model: Reporting Results by Race and Ethnicity

May 2022

Sara del Moral, Tax Policy Specialist
Research & Fiscal Analysis

Department of
Revenue
Washington State

"Working together to
fund Washington's future"

Overview

Background ⟩ Method ⟩ Discussion
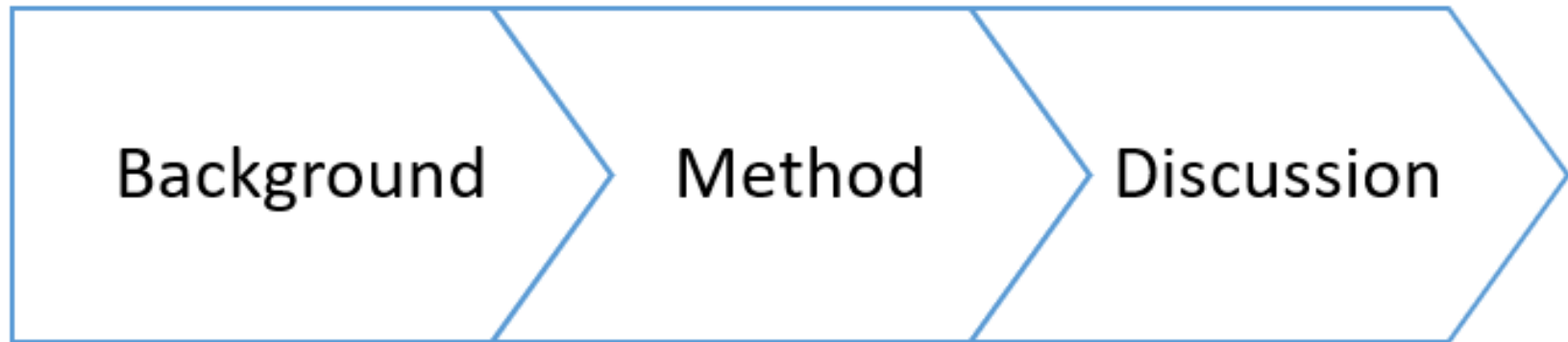
Background

Legislative mandate

A 2021 budget proviso directed us to….

"… conduct analysis of the current tax structure and proposed alternatives to estimate the impact on taxpayers, **including tax paid as a share of household income for various racial and ethnic groups** as reported in the most current census data available, American community survey, or other similar data sources."

Source: ESSB 5092

Methods observed in the tax literature

For imputing race or ethnicity for federal tax returns:

- **Direct imputation for each return,** using surname and geographic location [1].

- Use **Census Current Population Survey** (CPS), construct simulated tax units [2][3].

> **For imputing other variables, researchers have linked tax returns data with survey data (e.g., CPS, Medical Expenditure Panel Survey) [4]**

1. Fisher et al., 2021; 2. Goldin & Michelmore, 2022; 3. Hardy et al., 2022; 4. Joint Committee on Taxation (2015).

Tax Literature Review Highlights

<u>Typical question</u>: Does a given tax preference for the federal individual income tax favor certain racial groups?

<u>Method</u>:

1. **Create dataset** composed of tax returns with a race / ethnicity variable.
2. Estimate **participation rate and magnitude** of preference, in aggregate, comparing racial groups.
3. Compare racial groups **within the same income bracket**.

Tax Literature Review Highlights

Example results:

- "… we document **striking disparities in eligibility and benefits by income and race**. … Approximately three-quarters of white, non-Hispanic and Asian children are eligible for the full Child Tax Credit, compared to only about half of Black and Hispanic children. [1]"

- "The Earned Income Tax Credit **lowers overall inequality** by 5-10 percent in a typical year, improving the incomes of Black households relative to White households in the bottom half of the distribution. [2]"

1. Goldin & Michelmore, 2022; 2. Hardy et al., 2022

Interview highlights

U.S. Joint Committee on Taxation:
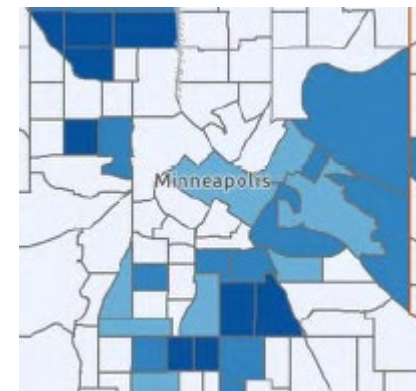
- Impute race using zip code, surname, and Census data [1].
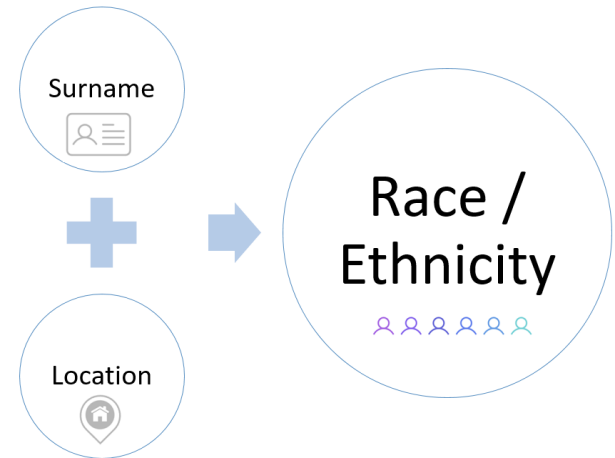
Minnesota:

- Overlay census tract race frequencies with tax statistics on a map.
  - ✓ Also considering surname / location method.

Surname

＋ Location → Race / Ethnicity
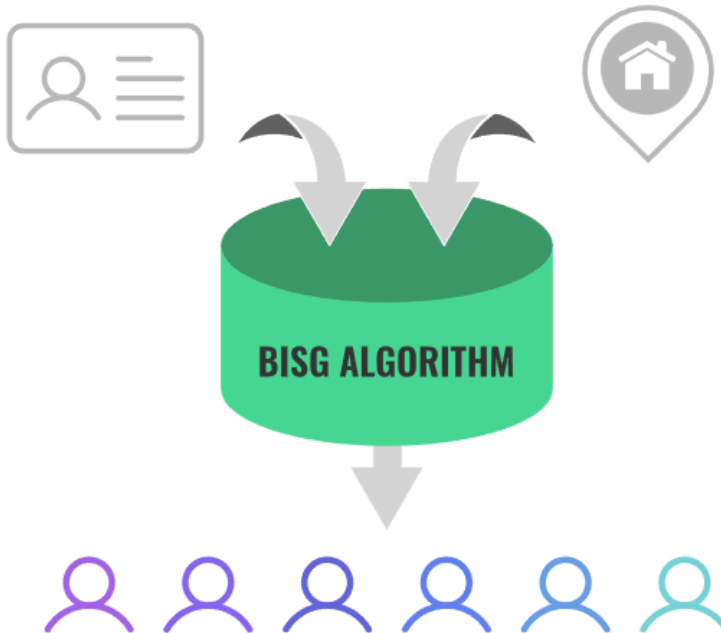
Unclaimed Working Family Credit

1. Zanazanian, A., personal communication, April 20, 2022; 2. Willette, 2021

Method

BISG: Impute using location and name

Administrative surname data   **Residential address data**

BISG ALGORITHM

Racial and ethnic probability for each data point
American Indian/Alaska Native, Asian and Pacific Islander, Black, Hispanic, Multiracial, White

Data Sources:

- Federal tax returns with **surname**, **zip code**, **census tract**

- U.S. Census:
  - **Surname** lookup to race probabilities
  - Racial composition for each **zip code** or **census tract**

Graphic Source: RAND Corporation (n.d.).
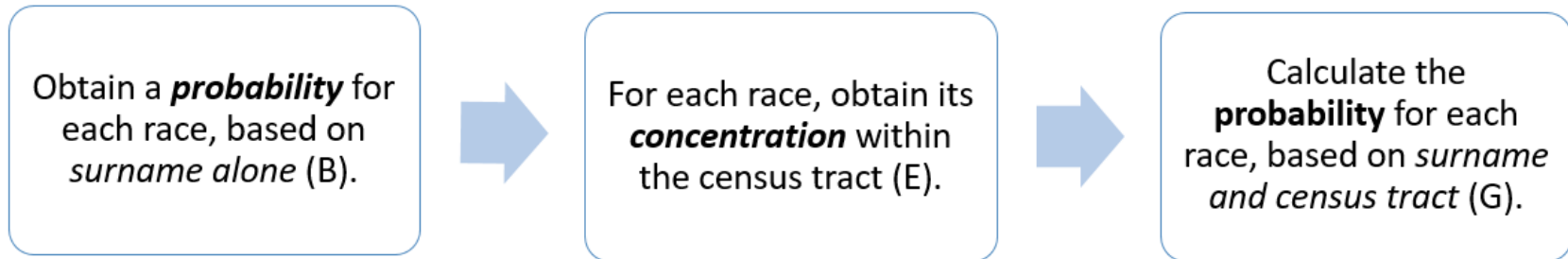
Method is supported by empirical research.

Estimates are **strongly predictive of self-reported race and ethnicity for the four largest racial and ethnic groups** in the U.S. [1]

The BISG method is **intended to estimate differences at the group or population level**; greater caution should be used in classifying specific individuals' race/ethnicity. [2]

[BISG results] **correlated highly with self-reported race-ethnicity** (0.76)... [However,] **imputing Native American and multiracial identities** from surname and residence **remains challenging**. [3]

1. RAND Corporation (n.d.); 2. Fremont et al., 2016 ; 3. Elliott et al., 2009

Step 1: Obtain P(race) for each category

Example for surname *Hernandez* in a low-Hispanic census tract.

Obtain a **probability** for each race, based on *surname alone* (B).

➡️

For each race, obtain its **concentration** within the census tract (E).

➡️

Calculate the **probability** for each race, based on *surname and census tract* (G).

	B	C	D	E = D / C	F = B * E	G = F / SUM(F)
Race / Ethnicity	**Probability of Surname**	**County Population**	**Census Tract (CT) Population**	**Concentration (CT / County)**	**Surname Probability * Concentration**	**P (Race)**
White	0.04	585,961	13,964	0.024	0.001	0.07
Hispanic	0.95	185,551	2,491	0.013	0.013	0.93
American Indian	0.00	1,327	11	0.008	0.000	0.00
Two or More	0.00	46,741	270	0.006	0.000	0.00
Asian / Hawaiian	0.01	216,643	1,083	0.005	0.000	0.00
Black	0.00	107,306	509	0.005	0.000	0.00
TOTAL	1.00	1,143,529	18,328	TOTAL	0.014	1.00

Percentage of county population residing in census tract.   0.016

Step 2: Impute race for each return

Example for surname *Hernandez* in a low-Hispanic census tract.

Assign a random number between zero and one to each return.

- Less than 0.93 → impute as *Hispanic*.
- Otherwise, impute as *white*.

Single imputation (rather than multiple)

ID	Surname	Random	Race / Ethnicity
777	Hernandez	0.45	Hispanic
787	Hernandez	0.95	White
797	Hernandez	0.11	Hispanic
807	Hernandez	0.25	Hispanic
817	Hernandez	0.16	Hispanic
827	Hernandez	0.34	Hispanic
837	Hernandez	0.15	Hispanic
847	Hernandez	0.55	Hispanic
857	Hernandez	0.41	Hispanic
867	Hernandez	0.32	Hispanic
877	Hernandez	0.48	Hispanic
887	Hernandez	0.43	Hispanic
897	Hernandez	0.66	Hispanic
907	Hernandez	0.08	Hispanic
917	Hernandez	0.80	Hispanic
927	Hernandez	0.91	Hispanic
937	Hernandez	0.46	Hispanic
947	Hernandez	0.36	Hispanic
957	Hernandez	0.25	Hispanic
967	Hernandez	0.66	Hispanic

"… both single and multiple imputations result in unbiased estimates of study associations. But single imputation results in too small estimated standard errors, whereas multiple imputation results in correctly estimated standard errors and confidence intervals." (Donders et al., 2006)

Data Match Challenge

Some tax returns cannot be imputed based on surname and census tract.

8 percent of tax return addresses did not geocode but do have zip codes.

- Impute using zip code and surname.

Some surnames will likely not match with Census dataset.

- Impute using census tract or zip code only.

Report Results

Which method?

Method	Number of Returns
A - Surname & Zip Code	x,xxx,xxx
B - Surname Only	x,xxx,xxx
C - Zip Code Only	xxx,xxx
D - Other	xx,xxx
Total	3,840,698

Race / Ethnicity	Number of Returns
White	x,xxx,xxx
Hispanic	x,xxx,xxx
American Indian	x,xxx,xxx
Two or More	x,xxx,xxx
Asian / Hawaiian	x,xxx,xxx
Black	x,xxx,xxx
Total	3,840,698

Describe Results

*We will also report results on **income** and **county**.*

Compare

with summary tables from **U.S. Census**

References

Anson-Dwamena, R., Pattah, P. & Crow, J. (2020). Imputing missing race and ethnicity data in COVID-19 cases. Virginia Department of Health. https://www.vdh.virginia.gov/content/uploads/sites/182/2020/08/Race-Ethnicity-Imputation-COVID-19.pdf

Donders, A.R.T., Van der Heijden, G.J., Stijnen, T., & Moons, K.G. (2006). A gentle introduction to imputation of missing values. Journal of clinical epidemiology, 59(10), 1087-1091. https://www.uvm.edu/~rsingle/JournalClub/papers/Donders+2006-JClinEpi_imputation.pdf

Fisher, R., Lin, E., & Costello, R. (2021, November 17-19). Measuring marriage penalties and bonuses by race and ethnicity: An application of race and ethnicity re-weighting to tax data [Presentation]. 114th Annual National Tax Association Conference, Virtual proceedings. https://nta.confex.com/nta/2021/meetingapp.cgi/Day/2021-11-19

Goldin, J., & Michelmore, K. (2022). Who benefits from the Child Tax Credit?. National Tax Journal, 75(1), 000-000. https://www.nber.org/system/files/working_papers/w27940/w27940.pdf

Hardy, B., Hokayem, C., & Ziliak, J.P. (2022). Income inequality, race, and the EITC. National Tax Journal, 75(1), 000-000. http://www.bradleyhardy.com/wp-content/uploads/2022/03/ntj-ineq-race-eitc_HHZ_102921-full.pdf

Joint Committee on Taxation (2015). Estimating changes in the federal individual income tax: Description of the individual tax model. JCX-75-15. Congress of the United States. https://www.jct.gov/publications/2015/jcx-75-15/

RAND Corporation (n.d.) RAND Bayesian Improved Surname Geocoding: Advancing equity through data science. https://www.rand.org/health-care/tools-methods/bisg.html

Willette, E. (2021, October 16-20). Alternative approaches to analyzing tax data by race [Presentation]. Federation of Tax Administrators: Revenue Estimating and Tax Research Conference, Kansas City, MO. https://www.taxadmin.org/assets/docs/Meetings/21rev_est/willette.pdf

Discussion