

The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey

Brittany Bond, J. David Brown,
Adela Luque, and Amy O'Hara

November 5, 2013

Disclaimer

Any opinions and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information on individual firms is disclosed.

Motivation

- Record linkage can enrich data and improve its quality
- Surveys and administrative data can be linked via Protected Identification Key (PIK)
- Census Bureau is unable to reliably assign a PIK to every record
- What is the nature of the bias introduced by incomplete PIK assignment?
- Can record linkage research lead to higher validation rates and less bias?

Research Strategy

- The American Community Survey (ACS) contains rich data on persons and households
- We study how person and household characteristics are associated with validation
- Results are compared from two different validation procedures

Person Processing

- Person Identification Validation System (PVS)
- Data processed through modules depending on completeness of personally identifiable information on input file
 - 2009 Modules: Geosearch, Namesearch
- Reference file should contain people at addresses from the same reference period as input file to PVS
- Parameters are set to minimize false matches

2010 PVS Enhancements

- Adds ITIN data to reference files
- Geosearch includes more neighboring zip codes in search area
- Adds Household Composition Module

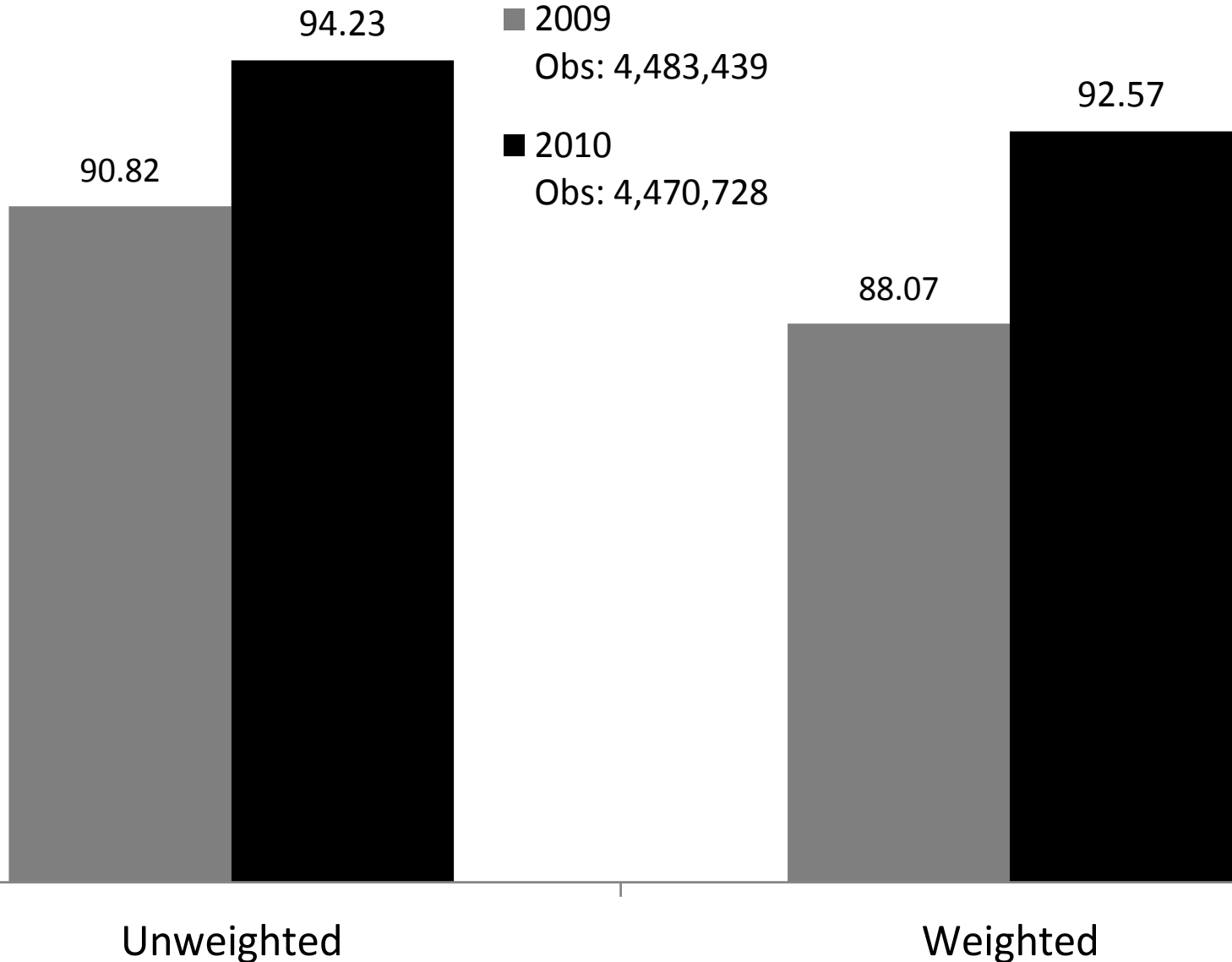
Factors Potentially Affecting Validation Rates

- **Mobility**
 - Age, migration, housing structure type, rent vs. own, living with or without family
- **Societal integration**
 - Income, education, employment, health insurance, race, ethnicity, U.S. citizenship, English language ability
- **Government program participation**
 - Military, govt. employment, social security, public assistance, public health insurance

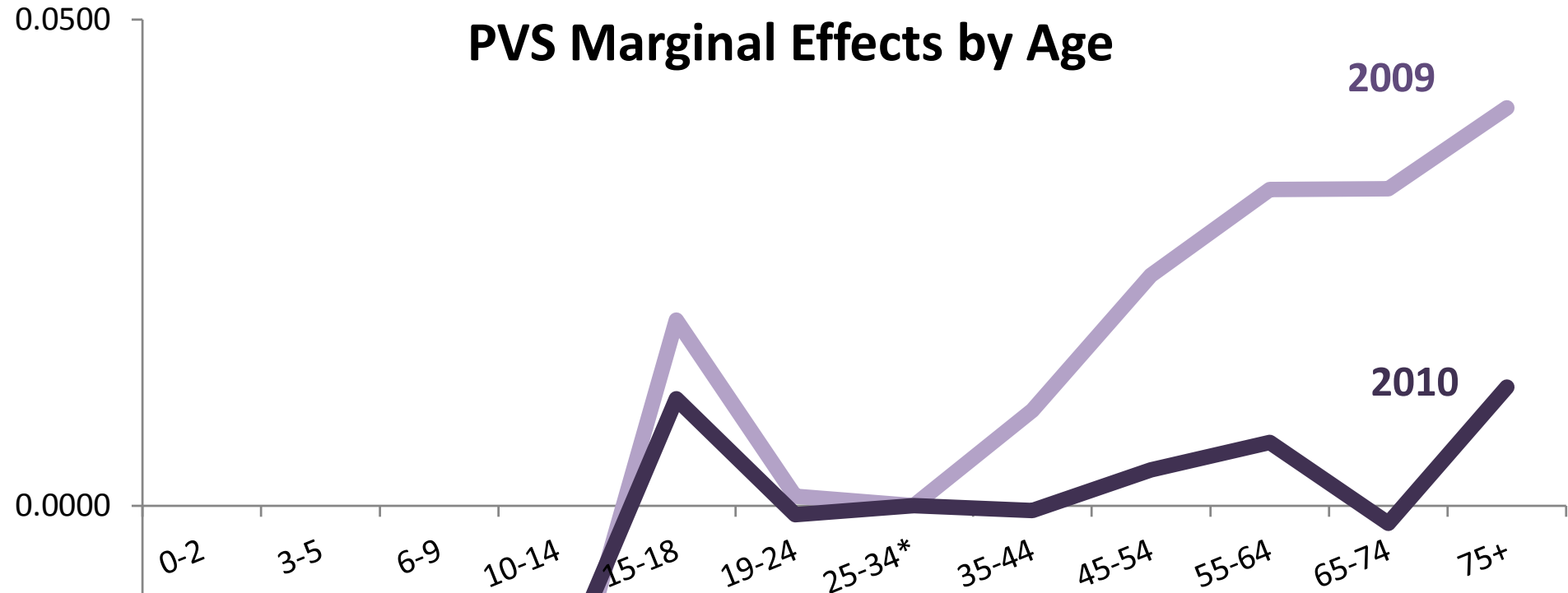
Factors Potentially Affecting Validation Rates

- Estimate probit models with dependent variable = 1 if record successfully validated
- Run separately for 2009 and 2010 ACS person records
- Weight by ACS person weights
- Report marginal effects

Overall PVS Rates, Weighted and Unweighted



PVS Marginal Effects by Age



PVS RATES

<u>Age</u>	<u>2009</u>	<u>2010</u>	<u>Age</u>	<u>2009</u>	<u>2010</u>
0-2	76.03	91.98	35-44	87.52	92.11
3-5	86.31	92.23	45-54	91.11	93.46
6-9	86.48	92.03	55-64	92.82	94.23
10-14	87.00	92.32	65-74	93.77	94.48
15-18	87.63	92.30	75+	93.56	94.83
19-24	83.98	89.99	TOTAL	88.07	92.57
25-34*	84.10	90.93			

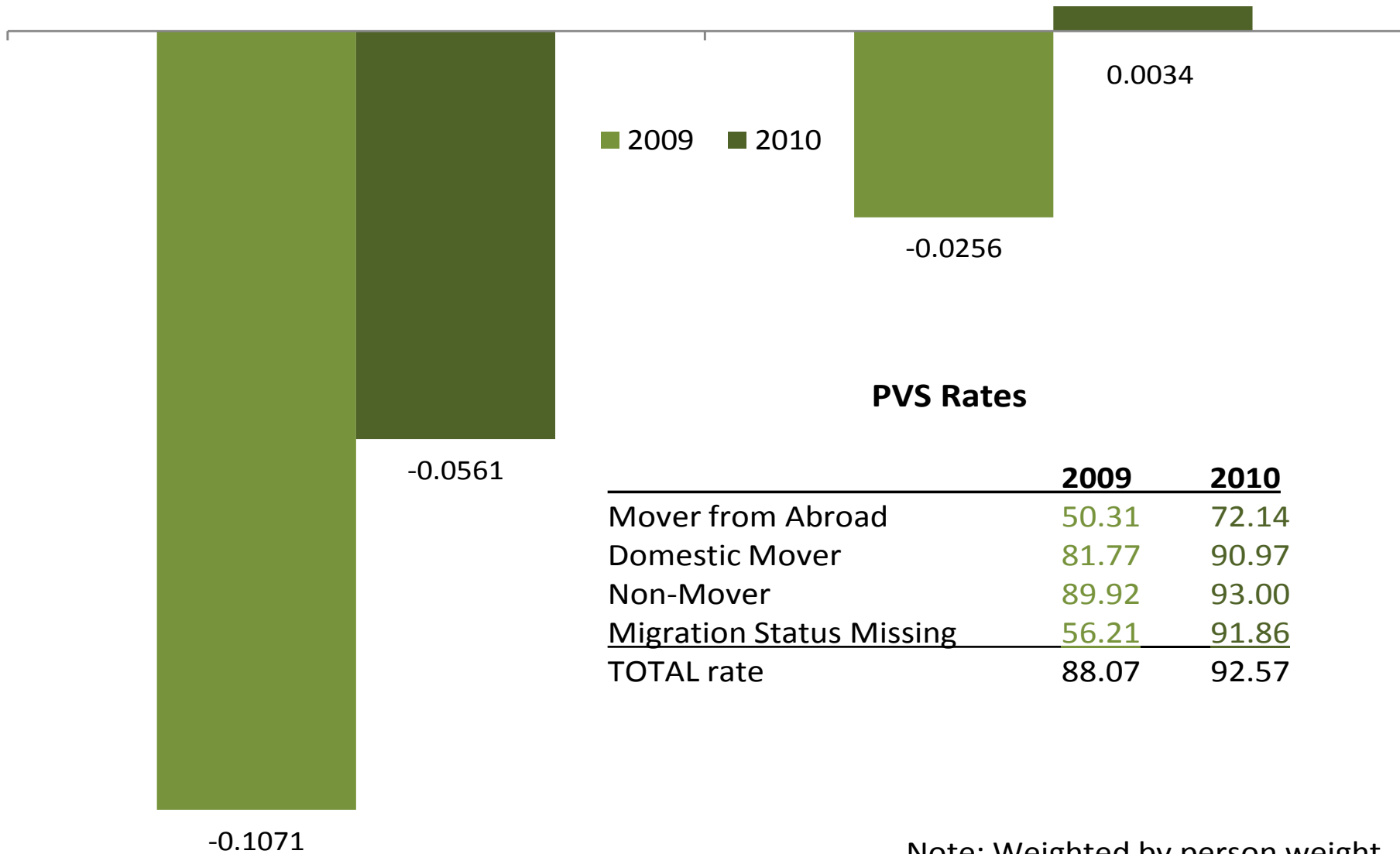
Note: Weighted by person weight.

* Base Category: Age 25-34.

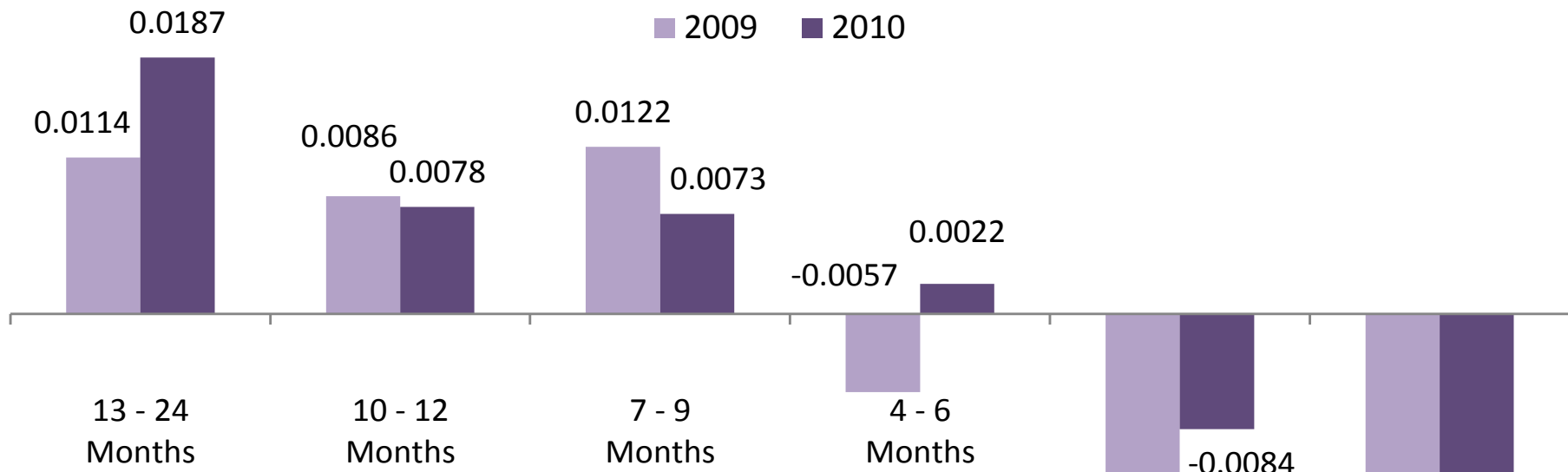
PVS Marginal Effects by Migration Status

Mover From Abroad in 12 Months before Interview
Month

Domestic Mover in 12 Months before Interview
Month



PVS Marginal Effects by Timing of Recent Move, by number of previous months



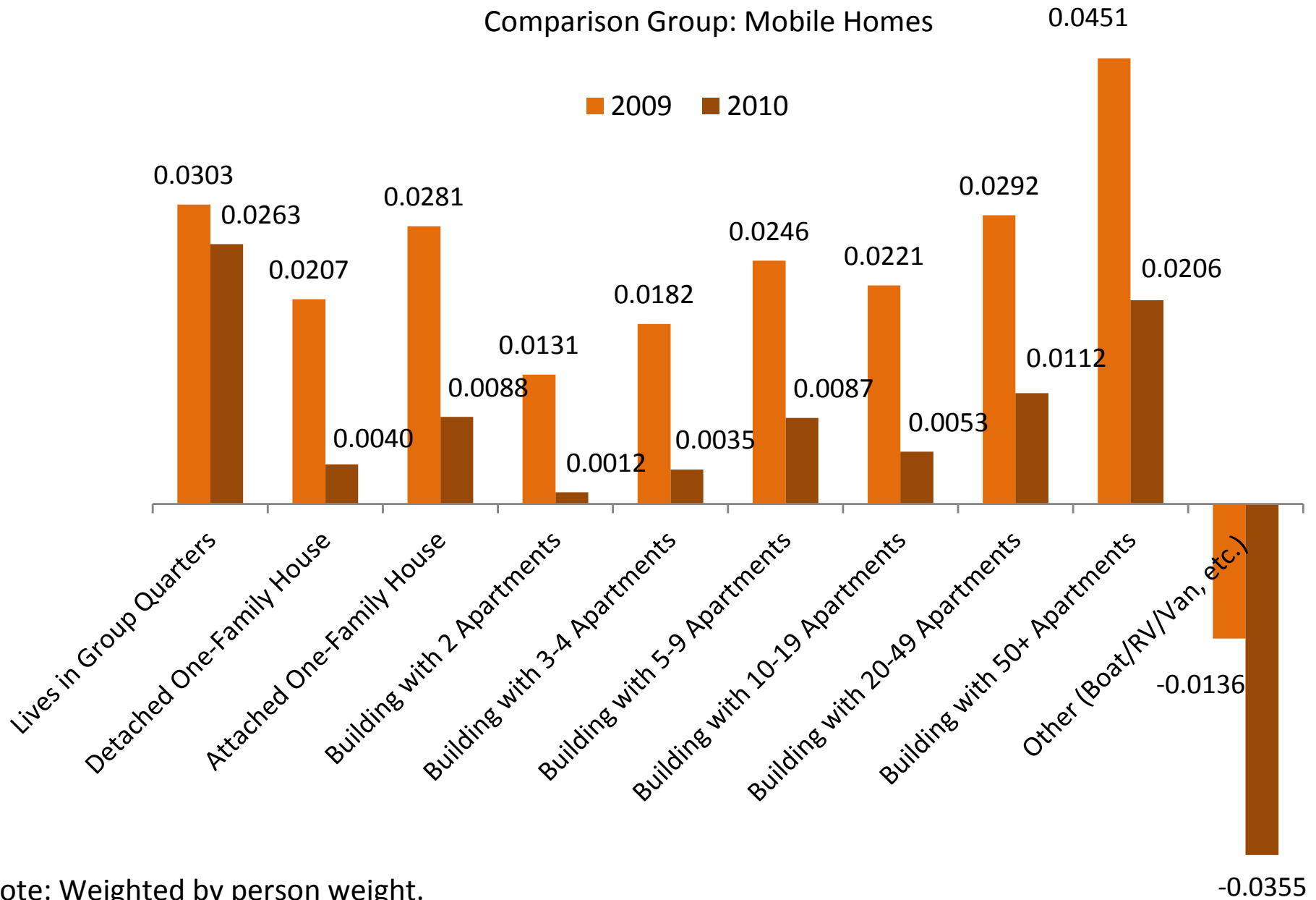
PVS Rates

	2009	2010
More than 24 months	89.58	92.97
13-24 Months	87.21	93.01
10-12 Months	84.27	91.31
7-9 Months	83.16	91.28
4-6 Months	79.80	90.33
1-3 Months	75.25	88.86
<u>Month of Interview</u>	<u>75.25</u>	<u>87.27</u>
TOTAL rate	88.07	92.57

Note: Weighted by person weight.

PVS Marginal Effects by Type of Housing

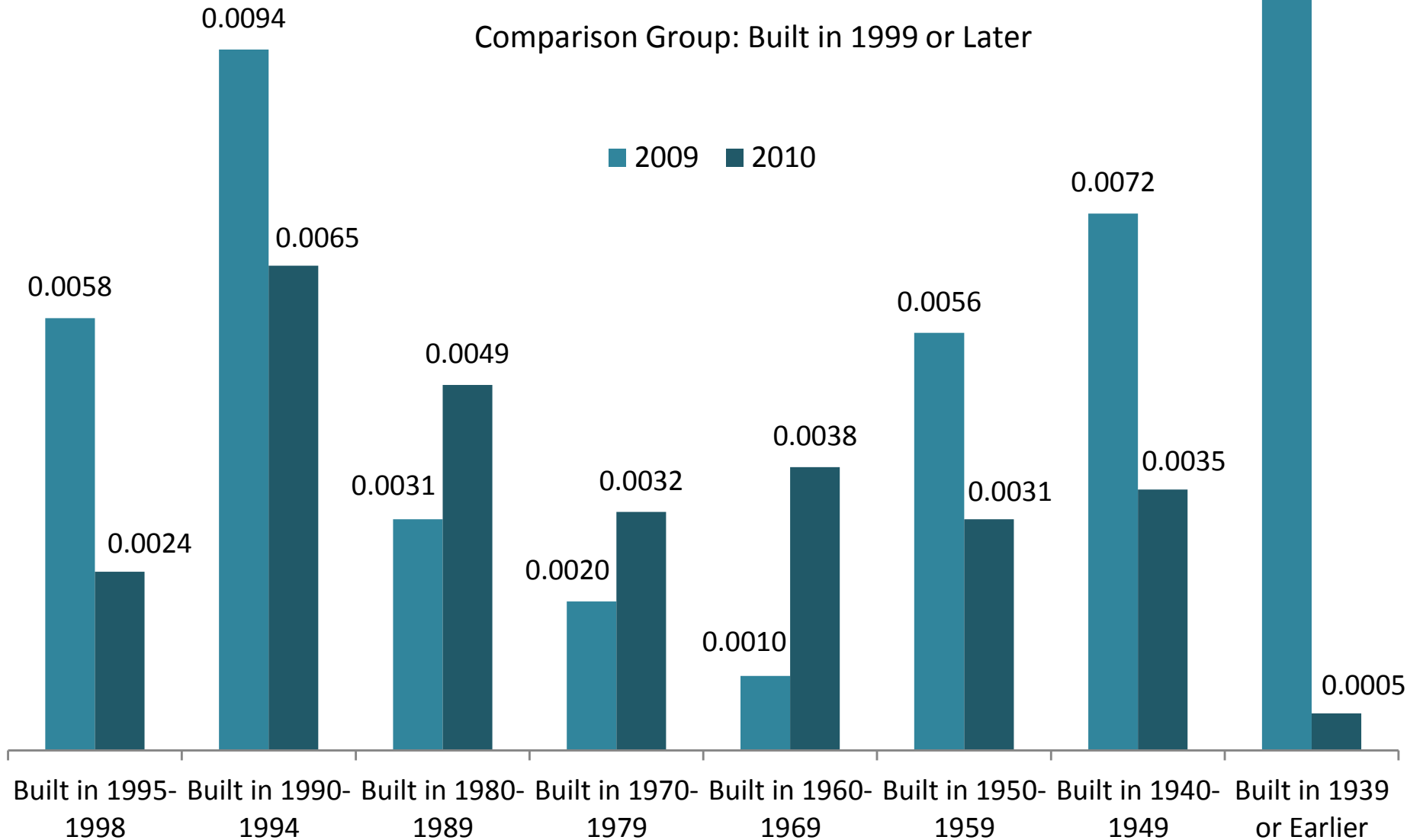
Comparison Group: Mobile Homes



Note: Weighted by person weight.

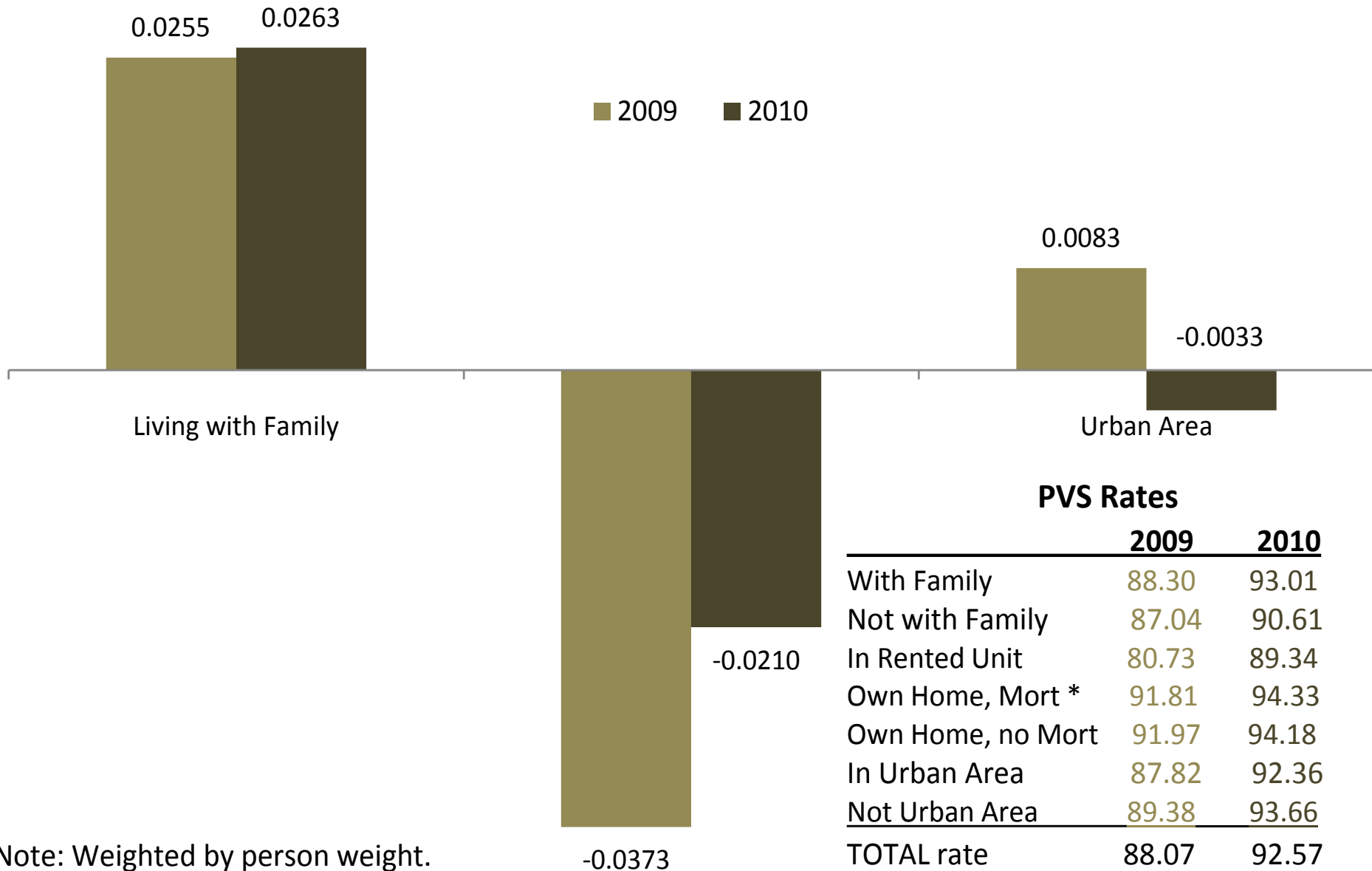
PVS Marginal Effects by When Housing Unit was Built

Comparison Group: Built in 1999 or Later



Note: Weighted by person weight.

PVS Marginal Effects by Where Living

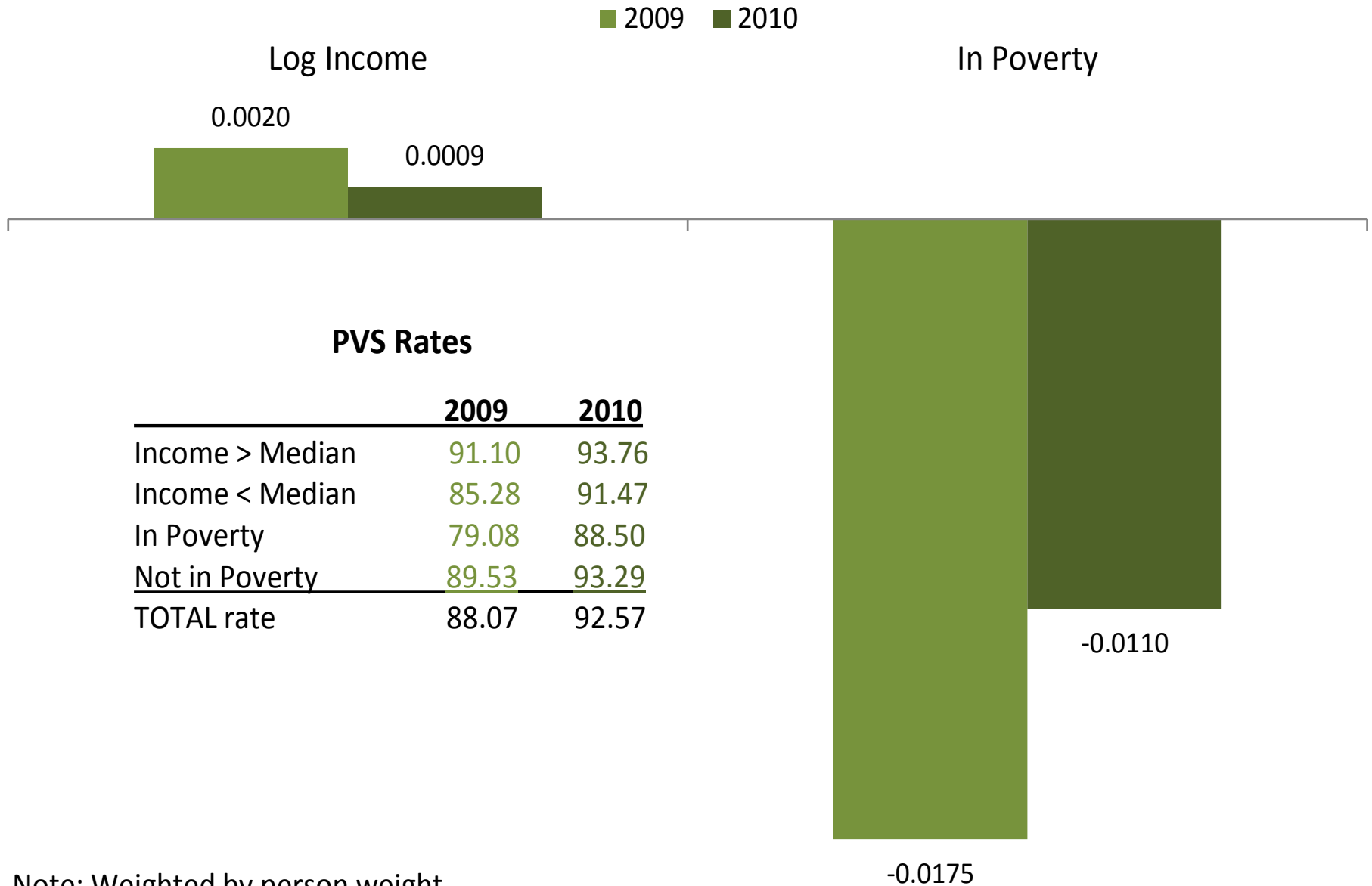


PVS Rates

	2009	2010
With Family	88.30	93.01
Not with Family	87.04	90.61
In Rented Unit	80.73	89.34
Own Home, Mort *	91.81	94.33
Own Home, no Mort	91.97	94.18
In Urban Area	87.82	92.36
Not Urban Area	89.38	93.66
TOTAL rate	88.07	92.57

Note: Weighted by person weight.
 *with a mortgage.

PVS Marginal Effects by Income

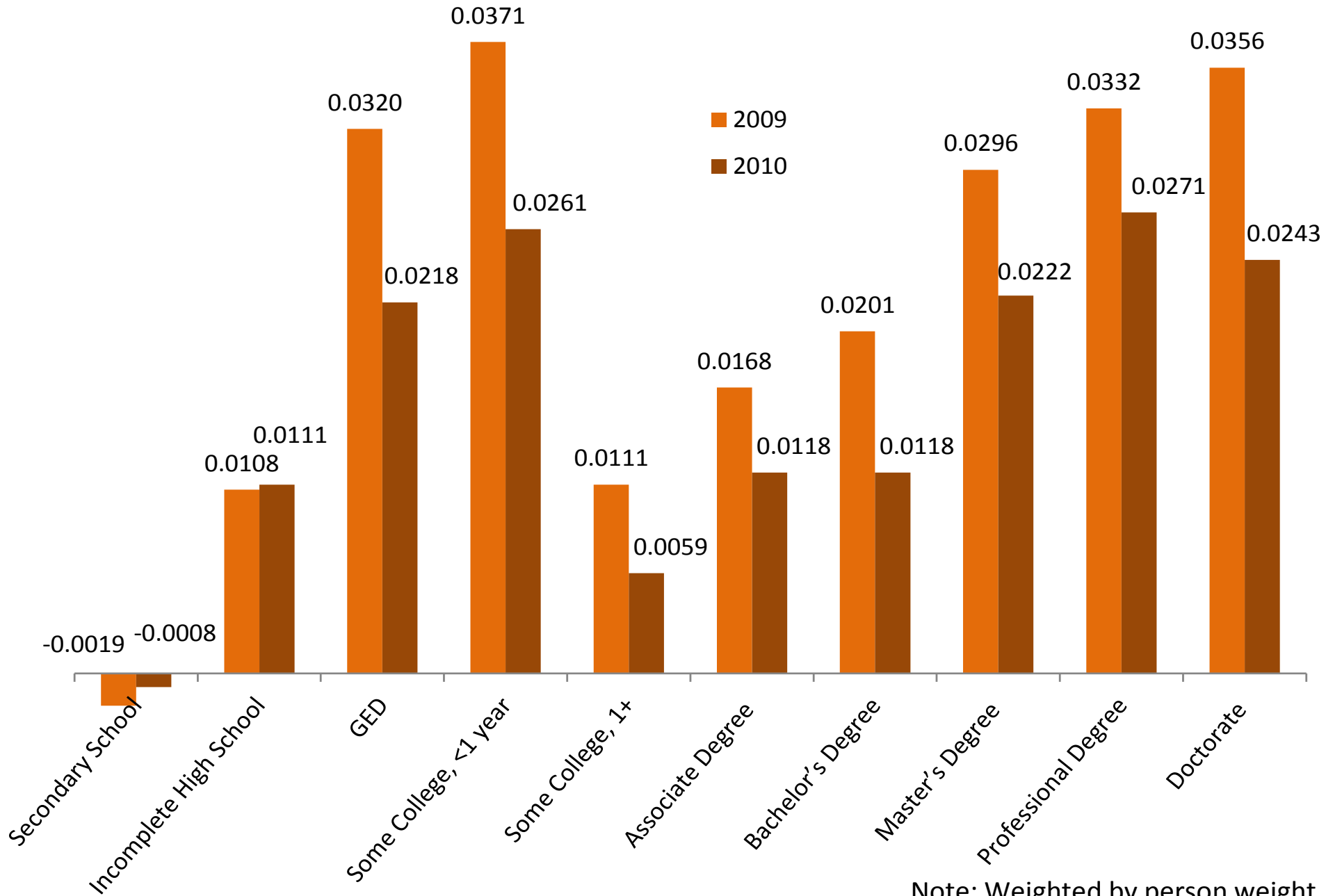


PVS Rates

	2009	2010
Income > Median	91.10	93.76
Income < Median	85.28	91.47
In Poverty	79.08	88.50
Not in Poverty	89.53	93.29
TOTAL rate	88.07	92.57

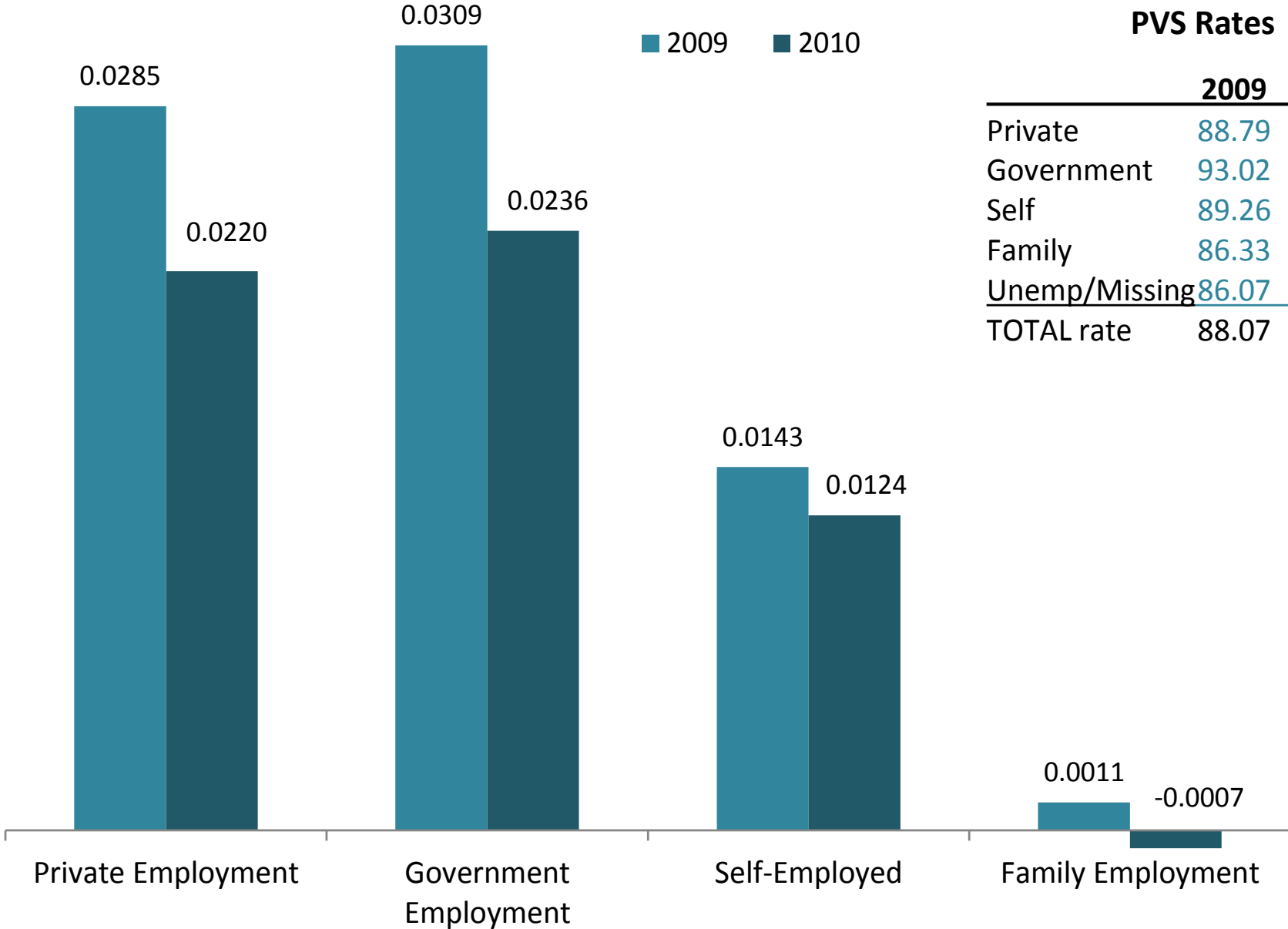
Note: Weighted by person weight.

PVS Marginal Effects by Highest Education Achieved



Note: Weighted by person weight.

PVS Marginal Effects by Employment

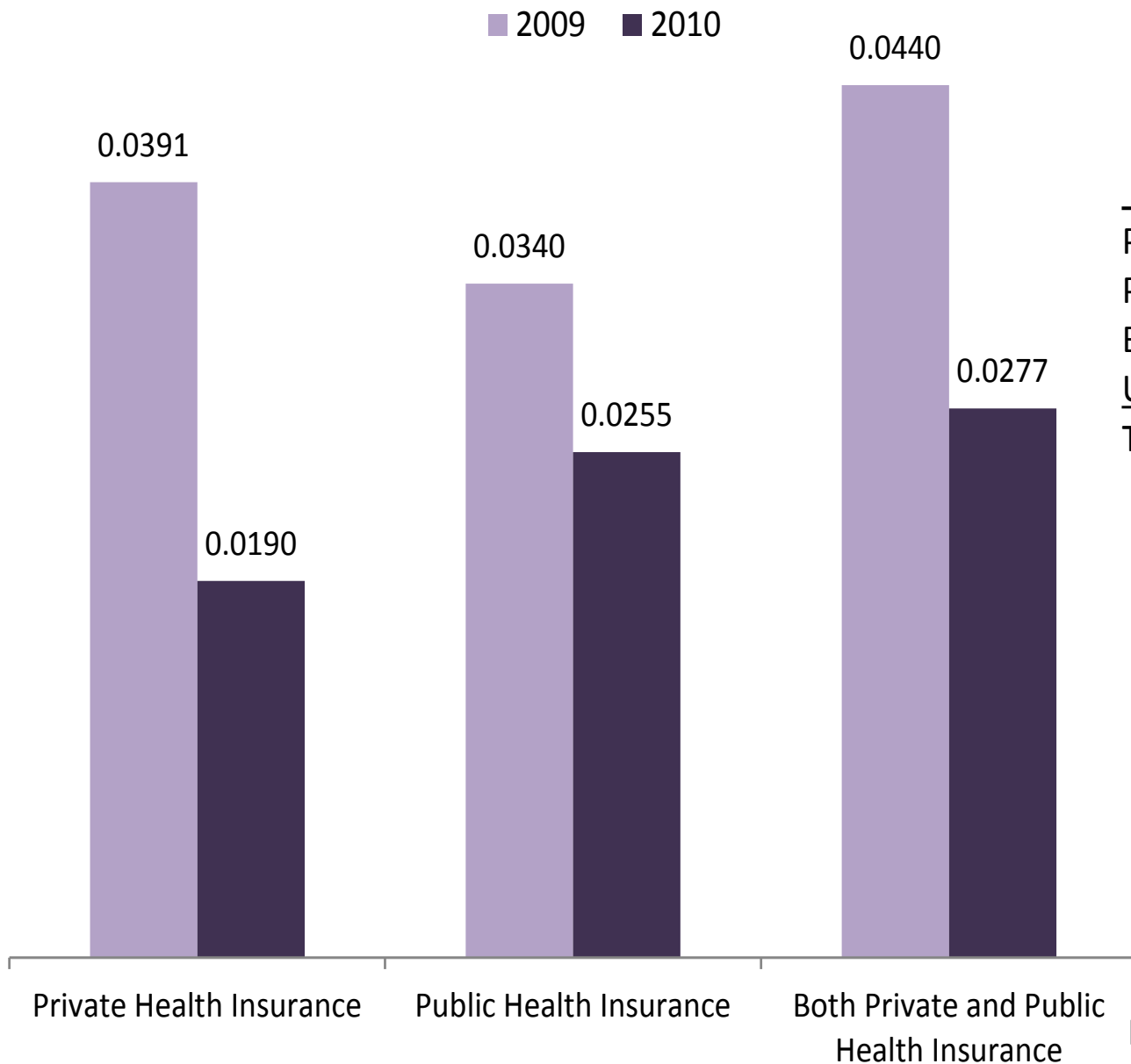


PVS Rates

	2009	2010
Private	88.79	92.81
Government	93.02	94.90
Self	89.26	92.61
Family	86.33	90.81
<u>Unemp/Missing</u>	<u>86.07</u>	<u>91.83</u>
TOTAL rate	88.07	92.57

Note: Weighted by person weight.

PVS Marginal Effects by Health Insurance Coverage



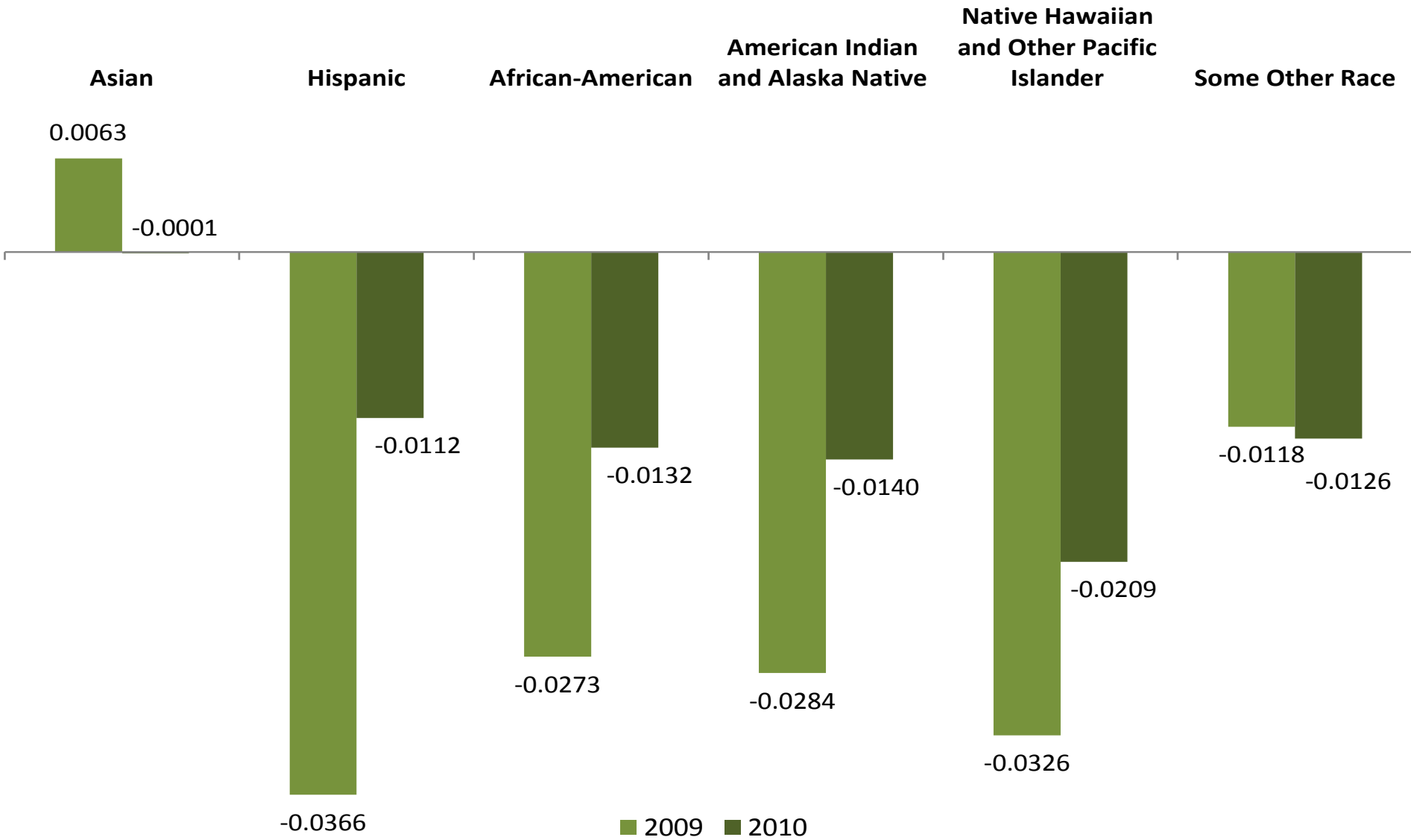
PVS Rates

	2009	2010
Private	90.94	93.87
Public	85.79	92.27
Both	94.25	95.37
Uninsured	75.83	86.53
TOTAL rate	88.07	92.57

Note: Weighted by person weight.

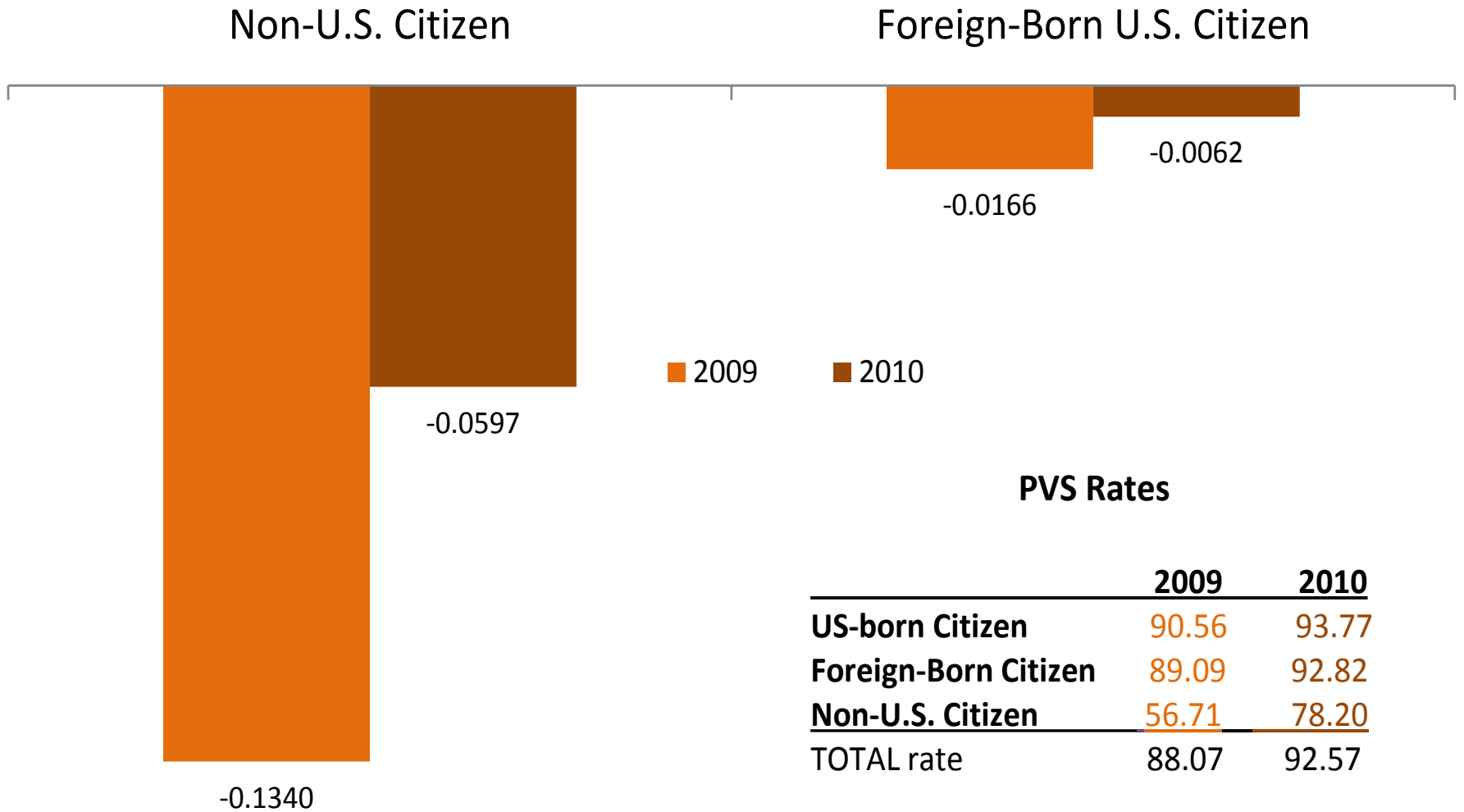
PVS Marginal Effects by Race/Ethnicity

Comparison Group: White for Race, Non-Hispanic for Ethnicity



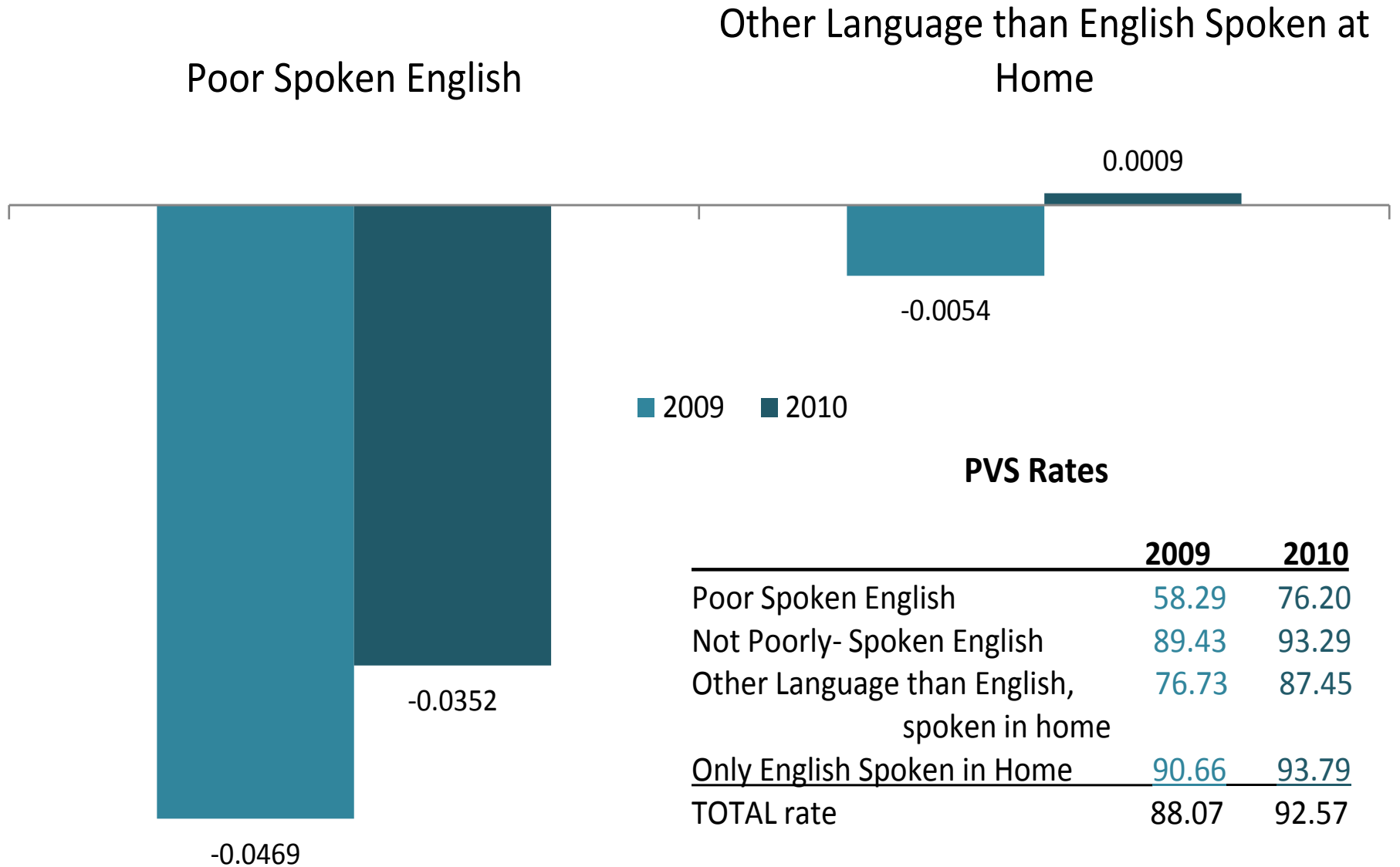
Note: Weighted by person weight.

PVS Marginal Effects by Citizenship Status



Note: Weighted by person weight.

PVS Marginal Effects by Language Spoken



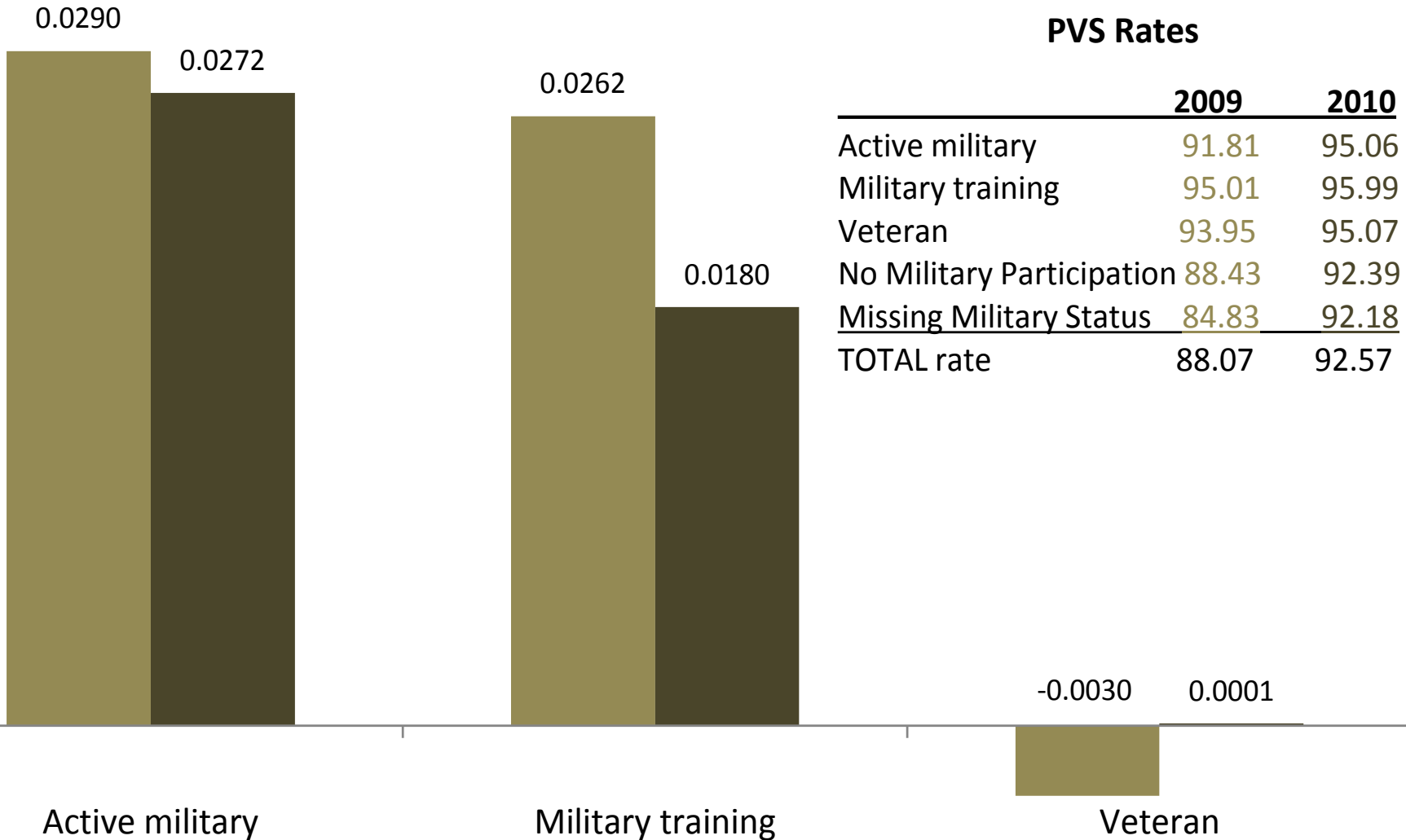
PVS Rates

	2009	2010
Poor Spoken English	58.29	76.20
Not Poorly- Spoken English	89.43	93.29
Other Language than English, spoken in home	76.73	87.45
Only English Spoken in Home	90.66	93.79
TOTAL rate	88.07	92.57

Note: Weighted by person weight.

PVS Marginal Effects by Military Participation

■ 2009 ■ 2010

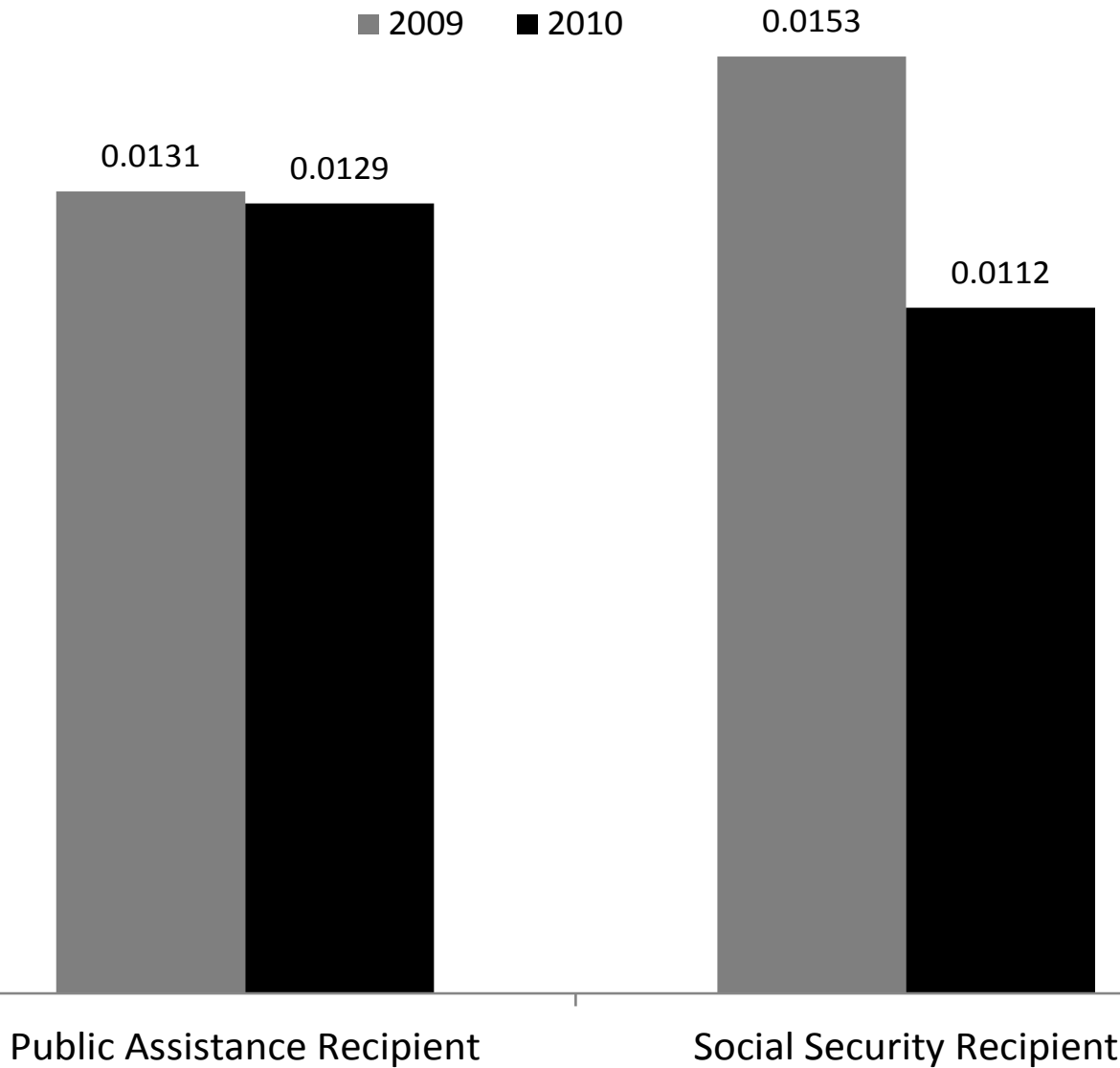


PVS Rates

	2009	2010
Active military	91.81	95.06
Military training	95.01	95.99
Veteran	93.95	95.07
No Military Participation	88.43	92.39
Missing Military Status	84.83	92.18
TOTAL rate	88.07	92.57

Note: Weighted by person weight.

PVS Marginal Effects by Public Assistance, Social Security



PVS Rates		
	2009	2010
Public Assistance	91.81	95.06
Social Security	95.01	95.99
No Public Assistance	93.95	95.07
No Social Security	88.43	92.39
TOTAL rate	88.07	92.57

Note: Weighted by person weight.

Conclusions

- Large differences in validation rates across groups in 2009 ACS
- Mobile persons, those on the fringes of society, non-participants in govt. programs less likely to be validated
 - Renters, recent movers, mobile homes
 - Less education, low income, non-employed, minorities, non-U.S. citizens, poor English
 - Military, social security, public assistance, public health insurance associated with validation

Conclusions

- Changes to validation system in 2010 increase validation rates
 - 4.5 percentage point increase
- Greatly reduces differences across groups
 - Drop in R-sq. from 0.138 to 0.061.