Religious disaffiliation and Internet use

Allen B. Downey*

August 2012

Abstract

This is a draft of a paper currently under consideration for publication. Please do not redistribute.

Using data from the General Social Survey, 2000-2010 we measure the relationship between Internet use and religious affiliation.

We find that Internet use is associated with decreased probability of religious affiliation; for moderate use (2 or more hours per week) the odds ratio is 0.82 (CI 0.69–0.98, p=0.01). For heavier use (7 or more hours per week) the odds ratio is 0.58 (CI 0.41–0.81, p¡0.001).

In the 2010 U.S. population, Internet use could account for 7.7 million people with no religious affiliation, or 30% of the increase in disaffiliation, relative to the 1980s.

From 1990 to 2010 the fraction of people in the United States with no religious preference increased from 8% to 18%, based on data from the General Social Survey (GSS). At the same time, the fraction of Protestants dropped from 62% to 51%; the fraction of Catholics and Jews did not change significantly; the fraction of other religions increased from 3.3% to 4.2%.

During the same period, the prevalence of Internet use increased from essentially zero to nearly 80%. Figure 1 shows Internet users per 100 people and the fraction of the population with no religious preference. In this paper we explore the effect of Internet use on religious affiliation.

Of course there are many possible explanations for increasing religious disaffiliation. Hout and Fischer investigate this trend (through 2000) and identify as causes political beliefs and generational effects, including a decrease in the fraction of people raised with no religion [1].

Smith and Kim study the decline in the Protestant majority, identifying as causes a decline in "the intergenerational retention rate for Protestants,

*Professor of Computer Science, Olin College of Engineering, Needham MA 02492. Email: allen.downey@olin.edu, Web: allendowney.com

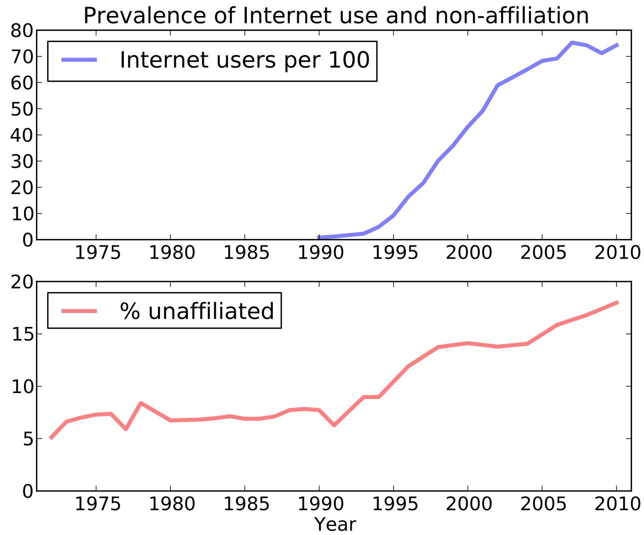

Figure 1: Internet users per 100 people and the fraction of the population with no religious preference, 1972–2010. Sources: World Bank, General Social Survey (GSS).

... shifts in immigration, [and] an increasing share of people identifying as generic Christians." [4].

Schwadel studies generational effects, especially people raised with no religion [2], and the effect of education [3].

Vargas investigates the effect of "political attitudes, religious skepticism, life stressors, and sociodemographic characteristics [5].

Wilcox et al investigate the differential decline among the high school graduates, relative to college graduates, and identify as causes "economic characteristics, current and past family characteristics, and attitudes toward premarital sex [6].

1 Methodology

We selected the following variables from the General Social Survey (GSS), available from `http://www3.norc.org/gss+website/`:

RELIG : "What is your religious preference?"

RELIG16 : "In what religion were you raised?"

AGE : Respondent's age when surveyed.

YEAR : Year of survey.

EDUC : "What is the highest grade in elementary school or high school that you finished and got credit for?"

INCOME : "In which of these groups did your total family income, from all sources, fall last year before taxes, that is?"

SEI : Respondent socioeconomic index (computed by GSS based on respondent's occupation).

SRCBELT : Classification of respondent's metropolitan statistical area as urban, suburban or rural (coded by GSS based on location of interview and U.S. Census data).

WWWHR : "Not counting e-mail, about how many minutes or hours per week do you use the Web?"

COMPWT : Respondent's computed sample weight. All statistics reported in this paper reflect these weights.

We use data from GSS survey years 2000, 2002, 2004, 2006 and 2010. In 2008 questions about Internet use were not asked.

These survey years include 14 948 respondents. For analysis, we selected 12 360 respondents who reported being raised in a religion (RELIG16). By the design of the GSS, not all respondents are asked all questions. We excluded respondents who were not asked or did not answer one or more of the relevant questions, yielding 8265 respondents.

Using Python programs available from `https://github.com/AllenDowney/internet-religion`, we compute the following recodes:

has_relig : 1 if the respondent reported any religious affiliation when interviewed as an adult, or 0 if the respondent reported "None".

had_relig : 1 if the respondent reported being raised in a religion, 0 otherwise.

top80_income : 1 if the respondent reports annual household income of $25,000 or more, which is the highest bracket in the survey. About 81% of respondents exceed this threshold.

Variable	Odds ratio	Probability	p-value
`(Intercept)`	8.5 (7.3, 10)	90 (88, 91)	*
`top80_income`	1.3 (1.1, 1.6)	92 (91, 93)	*
`born_from_1960`	0.82 (0.79, 0.86)	90 (89, 92)	*
`educ_from_12`	0.79 (0.7, 0.88)	88 (87, 90)	*
`www2`	0.7 (0.6, 0.81)	84 (82, 85)	*

Table 1: Model 1: odds ratios and cumulative probabilities, with 95% confidence intervals; * indicates statistical significance at $p < 0.001$

born_from_1960 : year the respondent was born minus 1960 (subtracting 1960 makes it easier to interpret the results of the regression).

educ_from_12 : number of years of school completed, minus 12; so 0 indicates a high school graduate; 4 or more indicates a college graduate.

www2 : 1 if the respondent reports using the Internet 2 of more hours per week, 0 otherwise. 53% of respondents exceed this threshold.

www7 : 1 if the respondent uses the Internet more than 7 hours per week, 0 otherwise. 25% of respondents exceed this threshold.

We compute logistic regressions with `has_relig` as the dependent variable. To control for income, age, and education, we use `top80_income`, `born_from_1960` and `educ_from_12` as control variables. We also tested `SEI` and and `SRCBELT`, but did not find statistically significant associations with these variables, so we omit them from the models.

2 Results

Table 1 shows results from Model 1, which estimates the effect of Internet use after controlling for income, year born, and education. For `born_from_1960` the odds ratio is for someone born in 1970, relative to someone born in 1960. For `educ_from_12` the odds ratio is for a college graduate relative to a high school graduate.

Odds ratios can be difficult to interpret; the cumulative probabilities are intended to help. If we start with a hypothetical person raised in some religion, with income in the lowest quintile, born in 1960, with high school education but no college, and low Internet use (less than 2 hours per week), the probability that this person has a religious affiliation as an adult is 90%. Now we change one variable at a time:

Variable	Odds ratio	Probability	p-value
`(Intercept)`	8.5 (7.3, 10)	90 (88, 91)	*
`top80_income`	1.3 (1.1, 1.6)	92 (91, 93)	*
`born_from_1960`	0.83 (0.79, 0.86)	90 (89, 92)	*
`educ_from_12`	0.8 (0.71, 0.89)	88 (87, 90)	*
`www2`	0.82 (0.69, 0.98)	86 (85, 88)	p=0.01
`www7`	0.71 (0.6, 0.83)	81 (80, 83)	*

Table 2: Model 2: odds ratios and cumulative probabilities, with 95% confidence intervals; * indicates statistical significance at $p < 0.001$

- If this person's income were in the top 80%, that would increase the chance of religious affiliation to 92%
- If (in addition) that person were born 10 years later (in 1970) the chance would drop to 90%.
- If (in addition) that person went to college, the chance would drop to 88%.
- If (in addition) that person used the Internet 2 or more hours per week, the chance would drop to 84%.

We also tested Model 2, which includes two levels of Internet use, `web2` and `web7`; see Table 2. Both `web2` and `web7` are statistically significant. Combining `web2` and `web7`, using the Internet 7 or more hours per week has and odds ratio of 0.58 and decreases the chance of religious affiliation by 5 percentage points (for the hypothetical case above).

To measure the information content of the models, we compute the self-information of partitioning (see Methodological Notes, below) and compare it with a null model that contains the same control variables and replaces `web2` with a randomly-generated variable.

We find that the null model captures 105 bits of information about respondents' religious affiliation; Model 1 captures 120 bits. This difference is significant with $p = 0.009$.

Model 2 captures 132 bits of information, which is significantly better than Model 1 augmented with a random variable (p=0.012). Comparing Model 2 to the null model indicates that information about Internet improves predictive value by 26% relative to the control variables.

More concretely, we can use this model to estimate the overall impact of Internet use on religious affiliation. As a baseline, we use Model 2 to fit, for each respondent, the probability of having a religious affiliation, and add up the total probability. In this simulation, 12.4% of respondents would have no religious affiliation.

To simulate a world with no Internet use, we set `wwwhr` to 0 for all respondents and recompute the total probability. In this world, 9.9% of respondents have no religious affiliation. The difference between these simulations is an estimate of the impact of Internet use.

Based on the 2010 population of the United States, these percentages indicate that in the U.S. population during the survey years 2000–2010, Internet use might account for an excess of 7.7 million people with no religious affiliation.

Finally, we can quantify how much of the increased disaffiliation in the U.S. might be due to the Internet. Based on GSS respondents partitioned by decade, the fraction of people in the U.S. with no religious affiliation was 7.1% in the 1980s, 10.2% in the 1990s, and 15.3% in the 2000s. The difference between the 1980s and the 2000s is 8.2 percentage points, or 25 million people in the 2010 population. If 7.7 million people have no religious preference due to Internet use, that accounts for 30% of the observed increase.

3 Discussion

As always, statistical association does not prove causation, but in this case there are reasons to believe that Internet use causes disaffiliation from religion.

First, it is easy to imagine at least two ways Internet use could contribute to disaffiliation. For people living in homogeneous communities, the Internet provides opportunities to find information about people of other religions (and none), and to interact with them personally. Also, for people with religious doubt, the Internet provides access to people in similar circumstances all over the world.

Conversely, it is harder (but not impossible) to imagine plausible reasons why disaffiliation might cause increased Internet use.

Finally, although a third factor could cause both disaffiliation and Internet use, that factor would also have to be new and rising in prevalence, like the Internet, during the 1990s and 2000s (see Figure 1).

With appropriate caution, there is evidence here for causation, and not just statistical association.

4 Limitations

The GSS data imposes some limitations on our analysis.

- Annual income data includes 11 bins between \$1000 and \$25000, and one bin for higher incomes. Since 81% of respondents fall into the highest bin, we cannot measure the effect of higher incomes, if any.
- In 2010 the GSS used a new protocol to screen questions about Internet use. As a result, it is not possible to analyze changes in Internet use after 2006. The same protocol is being used in 2012, so the next data point for time series analysis will be available in 2015 at the earliest.

5 Methodological notes

Confidence intervals and p-values in this paper are estimated by resampling, which is parsimonious and robust. Specifically, we draw random samples from the observed sample, using respondent weights so that each generated sample is representative of the general population. We compute regressions for 1001 generated samples and compute the median, 95% confidence interval, and p-value for each parameter.

To compare models, we use the self-information of partitioning (SIP) to compute the total surprisal of each model. For example, if the model predicts that the probability of the dependent variable for the ith respondent is p_i and we learn that the actual value is 1, we have learned $-log_2(p_i)$ bits of information. If we learn that the actual value is 0, that's $-log_2(1-p_i)$ bits.

So if we compute values of p_i and then learn the actual data, the total number of bits we learn is:

$$-\sum x_i \log_2(p_i) + (1-x_i)\log_2(1-p_i)$$

With a perfect model $p_i = x_i$ and the total information of the data is 0. As the quality of the model decreases, the information remaining in the data increases. So the better model is the one that leaves less information in the data.

To adjust for the number of parameters, we compare each model to a baseline that contains the same number of parameters, but where the explanatory variables have been replaced with random values. Each SIP value we report is the difference between the self-information of an estimated model and a null model that has no information about respondents.

Details of these methods are in the comments and code available from `https://github.com/AllenDowney/internet-religion`.

References

[1] M. Hout and C.S. Fischer. Why more americans have no religious preference: Politics and generations. *American Sociological Review*, pages 165–190, 2002.

[2] P. Schwadel. Period and cohort effects on religious nonaffiliation and religious disaffiliation: A research note. *Journal for the Scientific Study of Religion*, 49(2):311–319, 2010.

[3] P. Schwadel. The effects of education on americans' religious practices, beliefs, and affiliations. *Review of Religious Research: The Official Journal of the Religious Research Association*, pages 1–22, 2011.

[4] T.W. Smith and S. Kim. The vanishing protestant majority. *Journal for the Scientific Study of Religion*, 44(2):211–223, 2005.

[5] N. Vargas. Retrospective accounts of religious disaffiliation in the united states: Stressors, skepticism, and political factors. *Sociology of Religion*, 2011.

[6] W.B. Wilcox, A.J. Cherlin, J.E. Uecker, and M. Messel. No money, no honey, no church: The deinstitutionalization of religious life among the white working class. *Research in the Sociology of Work*, 23:227–250, 2012.