Replication of Nagler (1994) and Alvarez and Brehm (1995) with Special Attention to Numeric Accuracy

October 31, 2001

Abstract

In the course of an ongoing research project testing numerical accuracy of statistical techniques and software frequently used by political scientists we replicate two studies, Nagler (1994) and Alvarez and Brehm (1995). We find numeric inaccuracies, errors in formulas, and data problems that reduce the strength of the substantive conclusions of the studies. Our replication analyses reveal there are a number of factors that we may take for granted in the normal course of our analysis that affect the ability to faithfully replicate. More attention to these details may improve replication of research in the discipline.

1 Introduction

In an on-going research project into the numerical accuracy of statistical programs frequently used by political scientists (Altman and McDonald 2001), we replicate research articles employing complex maximum likelihood functions. In two replication studies, presented here, we are unable to accurately replicate published results. We find that the inability to replicate is caused by the presence of numerical inaccuracy in the statistical programs used to estimate the original results, errors in the published results and formulas, undocumented choices of starting values for maximum likelihood estimations, failure to completely archive and document the version of these data, and errors in data coding. These discrepancies demonstrate that replication depends on a number of factors that we may normally take for granted, and that more attention to detail may aid subsequent replication.

These replication studies will proceed as follows. We present replication results of studies by Nagler (1994) and Alvarez and Brehm (1995). (We would like to thank these authors for their attention to providing replication datasets.) We first attempt to duplicate the published results with archived data and archived code. We then run independent analysis on other statistical platforms using the archived data. Finally, we run each analysis using reference data from ICPSR.

We tnote any discrepancies, attempt to identify their source, and resolve them. We then conclude with comments on our findings and recommendations for political scientists who wish to produce replicable results.

2 Replication Study One: Nagler (1994)

Our first replication study is Jonathan Nagler's (1994) article "Scobit: An Alternative Estimator to Logit and Probit" published in *American Journal of Political Science* 38(1): 230-55. Replication data are drawn from the 1984 Current Population Survey, Voter Supplement File published by the Census Bureau (ICPSR #8457), and methods for replication are given in Nagler (1994). We note below some exceptions to these data and model specification that may prevent accurate replication.

The scobit estimator relaxes the assumption of probit and logit models that an individual with a probability of .5 of choosing between two alternatives is the most sensitive to changes in the independent variables. Nagler applies the scobit model to voting turnout by replicating studies by Wolfinger and Rosenstone (1980) and Nagler's (1991) own respecification of Wolfinger and Rosenstone's model. Readers interested in the substantive nature of the research are encouraged to read these sources.

To develop the scobit model, Nagler starts with the familiar binary response model, whose log-likelihood is given by:

$$\log L = \sum (y \log(F(\cdot)) + (1-y)\log(1-F(\cdot)))$$

To allow the shape of the response curve to shift the steepest response from .5, Nagler proposes $F(\cdot)$ to be equal to the cdf of the Burr-10 distribution (Burr 1942):

$$F(\cdot) = (1 + \exp(X\beta))^{-\alpha}$$

Note that when $\alpha = 1$, the scobit and logit models are equivalent.

We attempt at the outset to recreate data analyzed by Nagler from a reference data source. Nagler documents that his data are drawn from the 1984 Current Population Survey, Voter Supplement File, which is published by the Census Bureau. Nagler does not specify the rules to construct his data, so we assume that he follows rules laid out in the study he re-analyzes, by Wolfinger and Rosenstone (1980). Wolfinger and Rosenstone select responses from persons identifying themselves as citizens and those persons who provide a "yes" or "no" response to the vote question; those that could not recall if they voted are excluded from the analysis. When we follow these rules, we arrive at 99,673 observations, not the 98,857 reported in Nagler's analysis. We attempted other selections to trim 816 observations to no avail. When contacted, the author provided us with the original data extract, but reported an inability to reproduce the extract from a reference source.

The inability to reproduce the number of observations may be a result of changes to the CPS data itself. CPS datafiles are often corrected following release, no explicit versioning system is used when these updates occur, and no attempt is made to propagate updates to other data archives.[1] The International Consortium for Political Science Research (ICPSR) does record dates that data in their holdings are received and updated, but does not provide access to previous versions. Since Nagler does not document the version or modification date of the data he uses and there are no other case identifiers, such as respondent ids, we unable to investigate the source of the discrepancies more closely. In the analysis that follows we will present two sets of results, one replicating Nagler's work using data supplied to us and an additional run using the most recent version of ICPSR data.

We then attempt to reproduce Nagler's analysis. The program and results file using SHAZAM (v 6.2) were generously supplied to us by Jonathan Nagler. We ran the original program on the most recently available, at the time of this writing, of SHAZAM (v 9.0) running on a Linux platform.[2] In the interest of numerical accuracy, we replicate the analysis on two independent statistical packages, Stata (v 6.0) running on Windows and Gauss (v 3.2.43) running on a HP-Unix platform. We chose Stata because scobit has been implemented as a built-in function *scobit* within that program and Gauss because of its popularity among political methodologists.

Different optimization routines may find different maxima. One important option of many optimization routines is the use of analytic gradients, often requiring pain-staking coding by the user, or internally generated numerical approximations. The original implementation code of scobit in SHAZAM, provided in Nagler (1994), uses a numerical approximation for the gradient and Hessian. Stata's implementation employs an analytic gradient and Hessian. For Gauss's *maxlik* package, we code an analytic gradient and use a numeric approximation for the Hessian.

The initial attempt at replication failed dramatically in Gauss: The MLE failed to converge, or converged to completely different parameters. This failure to replicate, however, was a result of a mis-specification of the published analytic gradient of the scobit model, necessary for some maximum likelihood search algorithms. The correct gradient is given by:

$$\frac{\partial \log L}{\partial \beta} = \left(\frac{(1-y)}{F(\cdot)} - \frac{y}{1-F(\cdot)} \right) \times (-\alpha) \times \exp(X\beta) \times (1+\exp(X\beta))^{-(\alpha+1)} \times X$$

$$\frac{\partial \log L}{\partial \alpha} = \left(\frac{(1-y)}{F(\cdot)} - \frac{y}{1-F(\cdot)} \right) \times \log(1+\exp(X\beta)) \times (1+\exp(X\beta))^{-\alpha}$$

[1] Personal communication with CPS staff, April 10, 2000.

[2] We have noted in other analyses that replication can be platform dependent, due to different implementations of calculations, such as precision of the numbers or choice of psuedo-random number generators, on different platforms.

Starting values for the MLE may also affected the performance of optimization routines. Stata's implementation of scobit uses logit estimates as starting values for the scobit model. Using this methodology, Stata finds a credible maximum to the likelihood function. By default, Gauss assigns a vector of zeros as starting values if no other starting values are provided by the user, and scobit would not converge using these defaults. When we followed the lead of Stata to provide a vector of logit results, Gauss converged, albeit slowly.[3]

In Table 1 we report the results of our best attempt at replication of scobit. The first column of Table 1 presented the results published in Table 5 of Nagler (1994: 246) or obtained from Nagler's output file. In column two, we present the replication using SHAZAM code and data supplied to us by Nagler, changing only the version of SHAZAM to the latest version. In columns three and four, we present replication results using data supplied to us by Nagler estimated with Stata and Gauss, respectively. Discrepancies between the published results and our replication studies at the second digit, that cannot be attributed to rounding of variation of the third digit, are bolded.[4]

Comparing the coefficients among all the programs, using the same data as the published results, reveals that the programs do not agree on the value of the fourth digit and in some cases they do not agree on the third digit, however, there is general agreement for the first two digits. Confirming the agreement, the calculation of the log-likelihood is consistent across all three software packages. Exceptions occur with Gauss, which calculated coefficients for *Closing Date* equal to roughly half of the coefficients from the other packages and for *Gubernatorial Election* which was roughly one-and-a-half times greater than the coefficients calculated by the other packages. The value of the log-likelihood is slightly lower for the Gauss solution. Given the weak explanatory power of these variables in the equation, it may be that Gauss simply choose coefficients among the many likely values occurring in the flat surface of the likelihood function with respect to these variables.

While our replication analysis confirms the published coefficients, we find discrepancies between the published standard errors and those reported by our three replications. In the case of SHAZAM we see that the standard errors are uniformly *smaller*, by approximately ten percent, in the newer version than the published results. For both Stata and Gauss, the standard errors are generally consistent between the two packages and are roughly ten percent *greater* than the published standard errors. These discrepancies are not simply a choice of analytic or numeric to calculate the Hessian, which is in turn to calculate standard errors, as our replication of SHAZAM and Gauss both use a numeri-

[3]Gauss provides a number of different maximum-likelihood algorithms: the steepest descent method, BFGS (Broyden, Fletcher, Goldfarb, Shanno), DFP (Davidson, Fletcher, Powell), Newton-Raphson, BHHH (Berndt, Hall, Hall, and Hausman), and the Polak-Ribiere Conjugate Gradient method. (Aptech 1999). In our use of Gauss, we used the default algorithm, BFGS. SHAZAM, used in the Nagler's original analysis, provides BFGS and DFP, but Nagler did not specify which was used for scobit.

[4]In some cases, we bold discrepancies that occur at precision levels less than what we report in our table of results.

	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	SHAZAM	(Stata)	(Gauss)
Constant	-5.3290	-5.3290	-5.3289	-5.3486
	(.3068)	**(.2753)**	**(.3326)**	**(.3340)**
Education	.3516	.3516	.3516	.3591
	(.1071)	**(.0788)**	**(.1175)**	**(.1175)**
Educ. Squared	.0654	.0654	.0654	.0647
	(.0135)	**(.0097)**	**(.0146)**	**(.0145)**
Age	.1822	.1822	.1822	.1824
	(.0098)	**(.0102)**	**(.0107)**	**(.0108)**
Age Squared	-.0013	-.0013	-.0013	-.0013
	(.0001)	**(.0001)**	**(.0001)**	**(.0001)**
South	-.2817	-.2817	-.2817	-.2817
	(.0239)	**(.0209)**	**(.0313)**	**(.0314)**
Gubernatorial	.0003	.0003	.0003	**.0004**
Election	(.0219)	**(.0178)**	**(.0313)**	**(.0314)**
Closing Date	-.0012	-.0012	-.0012	**-.0006**
	(.0088)	**(.0065)**	**(.0095)**	**(.0095)**
Closing Date ×	-.0055	-.0055	-.0055	**-.0058**
Education	(.0038)	**(.0028)**	**(.0042)**	**(.0042)**
Closing Date ×	.0002	.0002	.0002	**.0003**
Educ. Squared	(.0004)	**(.0003)**	**(.0005)**	**(.0005)**
Alpha	.4150	.4150	.4151	.4147
	(.0324)	**(.0330)**	**(.0341)**	**(.0342)**
Log-Likelihood	-55283.16	-55283.16	-55283.16	-55283.17
No. of Observations	98,857	98,857	98,857	98,857

Table 1: Scobit Replication Results

	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	SHAZAM	Stata	(Gauss)
Constant	-5.3290	-5.3471	-5.3471	-5.3471
	(.3068)	**(0.2923)**	**(.3356)**	**(.3357)**
Education	.3516	.3494	.3494	.3494
	(.1071)	**(.0956)**	**(.1187)**	**(.1187)**
Educ. Squared	.0654	**.0664**	**.0664**	**.0664**
	(.0135)	**(.0124)**	**(.0148)**	**(.0148)**
Age	.1822	.1838	.1838	.1838
	(.0098)	**(.0099)**	**(.0109)**	**(.0109)**
Age Squared	-.0013	-.0013	-.0013	-.0013
	(.0001)	**(.0001)**	**(.0001)**	**(.0001)**
South	-.2817	**-.2976**	**-.2976**	-.2976
	(.0239)	**(.0228)**	**(.0334)**	**(.0334)**
Gubernatorial	.0003	**-.0016**	**-.0016**	**-.0016**
Election	(.0219)	**(.0184)**	**(.0325)**	**(.0325)**
Closing Date	-.0012	**-.0024**	**-.0024**	**-.0024**
	(.0088)	**(.0071)**	**(.0096)**	**(.0096)**
Closing Date ×	-.0055	**-.0053**	**-.0053**	**-.0053**
Education	(.0038)	**(.0033)**	**(.0042)**	**(.0042)**
Closing Date ×	.0002	.0002	**.0002**	**.0002**
Educ. Squared	(.0004)	(.0004)	**(.0005)**	**(.0005)**
Alpha	.4150	.4105	.4105	.4105
	(.0324)	**(.0311)**	**(.0338)**	**(.0338)**
Log-Likelihood	-55283.16	-55723.35	-55723.35	-55723.39
No. of Observations	98,857	99,673	99,673	99,673

Table 2: Scobit Replication Results Using Reference Data Source

cally calculated Hessian. We further note the numeric calculation of Gauss is close to the analytic calculation of Stata, leading us to speculate that numeric inaccuracies are present in SHAZAM's numeric approximation of the Hessian. However, without knowing the true answer in this case, we cannot state this finding with certainty.

If we take the estimated results from the newer version of SHAZAM at face value, we reach a different substantive conclusion that is the main point of Nagler's (1994) article. Nagler wishes to determine the validity of Wolfinger and Rosenstone's (1980) finding that persons of lower education are disproportionately affected by early closing dates for voter registration. In his published scobit model analysis, Nagler estimates a t-stat of -1.45 for this variable and concludes, "attempts to find variable-specific interactive effects between education and closing remain futile" (1994: 251). Using the newer version of SHAZAM, we estimate that the interaction term between education and registration closing date is nearly significant (t-stat -1.94), raising doubts about Nagler's finding. However, Stata and Gauss estimate a smaller t-stat than the published results, confirming with greater confidence Nagler's original conclusion.

We present our replication results of Nagler's model using a reference data source, the Current Population Survey, currently available from ICPSR, in Table 2. Not surprisingly, different data yields different results, but only slightly so. For those coefficients estimated with tight standard errors in the proceeding analysis, the new data generally confirms the coefficients and standard errors estimated by our Stata and Gauss replication using the original data. We still find persistent differences in standard errors between SHAZAM and Stata and Gauss, similar to the analysis of the archived data.

Interestingly, analysis of the reference data source tends to support Nagler's substantive conclusion. Stata and Gauss produce similar estimates using the archived and reference data sources for the interaction term between education and registration closing date. SHAZAM, however, estimates a larger standard error using the reference data than the archived data.

The sum of our replication analysis raises questions about Nagler's original substantive finding, that persons of lower income are not disproportionately affected by early closing dates of registration, although the analysis of the reference data source does restore some confidence of Nagler's original finding. However, perhaps we should rephrase Nagler's finding: "attempts to find *the correct estimates of* variable-specific interactive effects between education and closing remain futile."

3 Replication Study Two: Alvarez and Brehm (1995)

Our second replication study is R. Michael Alvarez and John Brehm's article "American Ambivalence Towards Abortion Policy: Development of a Hetreoskedastic Probit Model of Competing Values" in *American Journal of Political*

Science 39(4): 1055-89. Replication information and data for this article are available at the ICPSR replication archive as study #1113. We use data available in the replication archive and construct a separate dataset for analysis from a reference data source, a 1982 subset of the Cumulative General Social Survey. The cumulative data file used in the original analysis, 1972-1989, is no longer available at ICPSR but is contained in the 1972-1994 file available as ICPSR study #6492.

Alvarez and Brehm develop a model of response to policy questions when there is conflict among the underlying beliefs of the individual. Alvarez and Brehm argue that conflict begets ambivalence, or unequal variation in respondent's answers on policy questions between those who hold strong, consistent beliefs and those that are conflicted. The authors point out that such heteroskedasticity is well known in the linear regression framework, but has not been studied extensively in the binary choice framework of "yes/no" responses on survey instruments. The authors formulate a test and means to correct for heteroskedasticity in binary choice models and apply this methodology to survey questions on abortion policy issues. Readers interested in the details of the theory and data are referred to the published manuscript (Alvarez and Brehm 1995).

To develop the heteroskedastic probit model, Alvarez and Brehm start with the familiar binary response model, whose log-likelihood is given by:

$$\log L = \sum (y \log(F(\cdot)) + (1-y)\log(1-F(\cdot)))$$

To account for heteroskedasticity in the familiar probit model, Alvarez and Brehm propose:

$$F(\cdot) = \Phi(\frac{X\beta}{\exp(Z\gamma)})$$

Where Φ is the cumulative normal distribution, $X\beta$ is the linear form of the familiar probit, and $\exp(Z\gamma)$ is the variance component to account for heteroskedasticity.

The gradient of the heteroskedastic probit model, not provided by Alvarez and Brehm but used in our analysis, is given by:

$$\frac{\partial \log L}{\partial \beta} = \left(y(\frac{\phi(\cdot)}{\Phi(\cdot)})(\frac{X}{\exp(Z\gamma)}) + (1-y)(-\frac{\phi(\cdot)}{1-\Phi(\cdot)})(\frac{X}{\exp(Z\gamma)}) \right)$$

$$\frac{\partial \log L}{\partial \gamma} = \left(y(\frac{\phi(\cdot)}{\Phi(\cdot)})(-\frac{X\beta}{\exp(Z\gamma)}Z) + (1-y)(-\frac{\phi(\cdot)}{1-\Phi(\cdot)})(-\frac{X\beta}{\exp(Z\gamma)}Z) \right)$$

Where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the normal distribution.

We then attempt to replicate Alvarez and Brehm's analysis. The authors identify SHAZAM as the statistical package used in their estimation, and apparently use SHAZAM's algorithms to calculate a numeric gradient and Hessian. Using the Alvarez and Brehm replication code and data, we replicate the analysis on a newer version of SHAZAM (v 9.0) running on Linux, on Stata (v 6.0) running on Windows, and Gauss (v 3.2.43) running on HP-Unix. For our Stata replication, we use the heteroskedastic probit, *hetprob*, command programmed into Stata. This implementation uses an analytic gradient and Hessian. For our Gauss replication, we use the *maxlik* library programmed with an analytic gradient and a numeric Hessian. In addition to these replications, we generate an independent replication dataset following the rules in the published manuscript and the SAS code available in the replication archive and analyze the these constructed data with Stata.

The results of our replications, along with the original published coefficients and standard errors, are presented in Tables 4-10. Each table corresponds to one of seven dependent variables analyzed by Alvarez and Brehm (relating to a different hypothetical abortion situation). In the first column we present the published results, in the second column we present replication results using the archived code and data run on a newer version of SHAZAM than the published analysis, in the third and fourth columns we present the replications using the archived data for Stata and Gauss. (*SHAZAM returned final results with no indication of failure to converge or other serious problems. However, it did report over seventy thousand runtime warnings about the log function when *Mothers' Health* was used as a dependent variable.)

Among our three replications using the replication data set, we note general agreement in estimated coefficients and standard errors between the newer version of SHAZAM, Stata, and Gauss, even though the two use different methods to calculate the Hessian (used to compute standard errors, though not used to find the optimum of the likelihood function). Where the statistical packages disagree on coefficients, the differences occur at the fourth digit. The log-likelihoods are also in general agreement, confirming the estimated coefficients are within the neighborhood of the same optima. For standard errors, there is greater variation among our three replications, particularly for coefficients with weak statistical significance. Differences in standard errors occur at the third digit, though in no case is their disagreement between the three replications about the level of statistical significance of a coefficient.

The original published results are reported with only two significant digits to the right of the decimal, and we have no results files, so we are limited what we can say about the accuracy of the originally published results. However, we note discrepancies between the published results and our replications at the precision level reported in the article. First, we note that some published coefficients and standard errors are dramatically different by one or two orders of magnitude greater than those estimated by our independent replications. These errors appear in bold under the *Published* column of results. We suspect that these errors are the result of typos in publication of the original research – but we cannot confirm this, since the original output used by the authors to construct their table of results was not distributed in the replication archive. In the model with *Mothers' Health* as the dependent variable, coefficients on the variables *Know What ERA Means*, the interaction term *Pro Count* $\times$ *Con Count*, and *Importance* appear in print one to two orders of magnitude larger than estimated by Stata or Gauss. In the equation with *Too Poor* as the dependent variable, the coefficient and standard error on the constant appear to be off by an order of magnitude. In the equation with *No More Children* as the dependent variable, the standard error on the coefficient for *Religious Intensity* appear to be an order of magnitude too large. These errors are but small pieces of the overall puzzle that Alvarez and Brehm examine, so although the substantive interpretation of these coefficients are affected by these errors, we do not believe the errors affect the generalizations Alvarez and Brehm infer from their estimation.

Under the replication results, we bold discrepancies that differ from the published results that we cannot attribute to publication errors. Although the majority of differences that, after rounding, result difference from the second digit of the published results by .01, we identify ten sizeable discrepancies of standard errors that exceed this difference. The underestimation is not uniform, and tends to occur for standard errors in the same equation, such as *Mothers' Health* and *No More Children* as the dependent variables. This suggests to us that SHAZAM's algorithm used to calculate and invert the numeric Hessian encountered numeric inaccuracy for these equations, and that this inaccuracy is sensitive to task it is being put to.

When we replicate the analysis using a reference source, and rules laid out in the published article and the SAS code available in the replication archive purportedly used to generate the archived dataset, we are surprised to find the number of observations in our new analysis do not match the published number or the number in our replications using archived data. A close inspection of these data reveal the root of the discrepancy lies with the coding of a single variable, *Religious Intensity*. When we drop this variable from the analysis, these replication archive data and our own constructed data match in the number of observations and the heteroskadastic probit estimation results.

We can reconstruct what happened by examining the descriptive statistics for *Religious Intensity*. According to the Appendix of Alvarez and Brehm (1995: 1078-1079), the Religious Intensity variable "...was coded from the follow-up question to religious preference, 'Would you call yourself a strong (PREFERENCE NAMED) or not a strong (PREFERNCE NAMED)?' This variable was set to 1 for *strong preference*, .33 for not strong preference, .67 for not very strong preference, and 0 for no religious preference." When we produce frequencies for the archived and our own constructed datasets, in Table 2, we see that although the archived SAS code that purportedly was used to construct the archived dataset does implement the rule as stated, the rule was apparently not followed in generating the archived dataset and the data used in the published results (since the published number of observations matches our replication analysis of the archived data). In these archived data we observe what appears to be a permutation of the rules in the data appendix and the SAS

Religious Intensity			
Archived Data		Constructed Data	
0	155	0	132
.5	758	.33	155
1	760	.67	758
		1	760
Missing	187	Missing	55

Table 3: Frequencies for Religious Intensity

code; not a strong preference is coded as 0, not a very strong preference is coded as .5, strong preference is coded as 1, and no religious preference is coded as missing. In other words, those persons who responded with no religious preference to this question are listwise deleted from the analysis, and the three remaining responses are re-scaled onto the [0,1] interval.

Not surprisingly, when we introduce persons with no religious preference into the model we produce results different from those published, presented in bold in the Tables 11-17. Considering that listwise deletion occurs on a subset of persons who might have substantially different views of abortion than the rest of the sample, all of the models (except Mother's Health, described below) generally do not dramatically change and still support the published table of results. The exception is the variable *Religious Intensity*, the subject of the coding error, for which we see an across the board increase in the size of the coefficient. The standard error also tends to increase, but not as greatly, boosting the variable into statistical significance for five of seven equations where before it only reached that threshold for one equation in the published results. Curiously, the text of the published article agrees with our findings and disagrees with the published table of results: "*Religious Intensity* is consistently negative: those who have strong religious preferences are more likely to reject abortion under all seven scenarios, and to a statistically significant degree in all but two of those cases" (Alvarez and Brehm 1995: 1070).

When we replicated the analysis from reference data, we saw dramatic differences in SHAZAM's estimate of the model with *Mother's Health* as the dependent variable. When we estimated the model with the latest version of SHAZAM, we once again observed that thousands of warning were generated by SHAZAM's logarithmic function, although SHAZAM again reported no error in the final results. Unlike the analysis on the archived data, the coefficients estimated by SHAZAM on reference data were often wildly different different than those estimated by Stata and Gauss (which were similar to each other). SHAZAM's coefficients differed by orders of magnitude, significance, and sign. We also note that log-likelihood for SHAZAM's estimation was -324.6417, which is lower than -317.4736 that estimated by Stata and -317.4743 estimated by Gauss. Despite this, we speculate that SHAZAM has not found the global optimum, but has succumbed to numerical inaccuracies in intermediate calculations. Ironically, even though Alvarez and Brehm's ran their analysis on an older version, and generally less accurate, version of SHAZAM using incorrectly coded data, they produced what are probably better results than they would have using the most recent SHAZAM software and correctly coded data.

Other than the *Mother's Health* model, Shazam, Stata and Gauss produced roughly the same results when run on the reference data. Comparing the different replications, we note general agreement on coefficients. But, as with the replication using archived data, we find that SHAZAM continues to estimate different standard errors than Stata and Gauss, which are in general agreement. The discrepancies are not uniform. Sometimes the statistical packages are in agreement, and other times SHAZAM is higher or lower than Stata and Gauss. Again, this leads us to believe that numerical inaccuracies affect SHAZAM calculation of the standard errors, but without knowledge of the true answer, we cannot state this finding with certainty.

We also see changes in coefficients and standard errors among variables with a low ratio of the coefficient to the standard error. Although we observe change signs in two cases, we do not dramatically increase the level of confidence for the estimated coefficients on any of these variables. Drawing upon the similar exercise of introducing a new version of data in the previous replication of Nagler's research, the lesson is that coefficients with low confidence are those that seem to be most sensitive to changes in data.

Overall, we believe that our constructed data and the estimates we derive from it more faithfully replicate the analysis as presented in the text of Alvarez and Brehm. The coding error that we identify fortunately does not dramatically change the results and, from the text of the article, we believe that Alvarez and Brehm were looking at the results we produce using our constructed data when writing the description of their results. Still, like our replication of Nagler, we discover indication of numeric instability among the published standard errors, but only for certain cases suggesting the inaccuracies are sensitive to the task SHAZAM is put to. And again, these discrepancies are troubling, but while softening the published conclusions of Alvarez and Brehm, do not change their overall results (we have no way of telling how these inaccuracies may affect the results Alvarez and Brehm discuss in their article since these results are not published).

4 Conclusion

Our two replication studies reveal that replication has a number of hidden pitfalls. Some are simply what appear to be errors in publication, and could be fixed with more attention paid by authors to the final page proofs. Others go deeper into the enterprise of replication, such as numerical accuracies across different statistical software packages, careful coding of variables, and issues related to data archiving.

The case of Nagler (1994) shows attention should be paid to numerical inaccuracies. If Nagler had estimated his scobit model using his archived data on the most recent version of SHAZAM he would have reached a different substantive conclusion than the originally published results, estimated with an earlier version of SHAZAM. If he had run his analysis on Stata or Gauss, he would have more confidence in his results. Since we do not have the luxury of knowing the true answer, we are left in the uncomfortable position of being unable to confirm or deny Nagler's original research. However, our analysis of a reference data source restores confidence in his conclusions.

Our replication of Alvarez and Brehm (1995) using the same software shows that the discrepancy between statistical software is not systematic. In five of the seven Alvarez and Brehm models, all three statistical software programs generally produce the same estimates, while in the remaining two there are numerous discrepancies between SHAZAM, and Stata and Gauss. In one model, using a reference data source, we find SHAZAM settles on a local optima, while both Stata and Gauss find a solution with a higher log-likelihood.

Finally, we see that additional care may be needed with replication data. Authors who wish their work to be replicable over an extended period should carefully document the source, version, and modification date of the data that they use. Authors who provide a separate replication dataset can ensure that later changes to the original data source does not effect their results, but should including respondent (or other) identifier information in order to match the replication dataset to the original data source. We have seen that this may not be sufficient, as errors can even creep into replication archives. Providing the code to manipulate a primary source of data is useful on this account, as this allows researchers to follow in the footsteps of those that created the data, even when versions of data change.

One should not take this article as a particular criticism of the authors we replicate. Quite the opposite, Alvarez, Brehm and Nagler have taken extraordinary care to make their research reproducible and are to be applauded for it. The lesson to take from this enquiry is that ensuring reproducible results is more difficult than is commonly understood, and requires more methodological attention.

References

[1] Alvarez, R. Michael and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Hetreoskedastic Probit Model of Competing Values." *American Journal of Political Science* 39(4): 1055-89.

[2] Micah Altman and Michael P. McDonald. 2001. "Choosing Reliable Statistical Software." Manuscript.

[3] Burr, Irving W. 1942. "Cumulative Frequency Functions." *Annals of Mathematical Statistics* 13: 215-32.

[4] McCullough, Bruce D. 1999. "Econometric Software Reliability: Eviews, LIMDEP, SHAZAM, and TSP." *Journal of Applied Econometrics* 14:191-202.

[5] Nagler, Jonathan. 1991. “The Effect of Registration Laws and Education on U.S. Voter Turnout.” *The American Political Science Review* 85 (4): 1393-1405.

[6] Nagler, Jonathan. 1994. “Scobit: An Alternative Estimator to Logit and Probit.” *American Journal of Political Science* 38(1): 230-55.

[7] U.S. Dept. of Commerce, Bureau of the Census. 1986. CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE, 1984 [Computer file]. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census [producer], 1985. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

[8] Wolfinger, Raymond E. and Steven J. Rosenstone. 1980. *Who Votes?* New Haven: Yale University Press.

	Mothers' Health			
	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	2.55	2.5478	2.5478	2.5477
	(.46)	**(.4891)**	**(.4861)**	**(.4861)**
Black	-.51	-.5118	-.5118	-.5118
	(.14)	**(.1639)**	**(.1641)**	**(.1641)**
Male	-.08	-.0792	-.0792	-.0792
	(.11)	**(.1263)**	**(.1251)**	**(.1251)**
Catholic	-.52	-.5205	-.5205	-.5205
	(.13)	**(.1524)**	**(.1513)**	**(.1513)**
Religious Intensity	-.39	-.3910	-.3910	-.3908
	(.20)	**(.2217)**	**(.2194)**	**(.2194)**
Attend Church	-1.04	-1.0382	-1.0382	-1.0381
	(.25)	**(.2747)**	**(.2748)**	**(.2748)**
Know What	**-.18**	-.0178	-.0178	-.0182
ERA Means	(.17)	**(.2007)**	**(.1990)**	**(.1990)**
Support ERA	.33	.3261	.3261	.3263
	(.17)	**(.2107)**	**(.2083)**	**(.2083)**
Variance Model				
Pro Count	-.14	-.1399	-.1399	-.1400
	(.07)	**(.0751)**	**(.0763)**	**(.0763)**
Con Count	.17	.1689	.1689	.1688
	(.09)	**(.0959)**	**(.0984)**	**(.0984)**
Pro Count ×	**-.44**	-.0440	-.0440	-.0440
Con Count	(.04)	(.0437)	(.0444)	(.0444)
Importance	**.51**	.0051	.0051	.0053
	(.15)	**(.1390)**	**(.1442)**	**(.1442)**
Information	.37	.3712	.3712	.3711
	(.13)	(.1265)	(.1335)	(.1335)
Firmness of Opinion	-.37	-.3705	-.3705	-.3706
	(.16)	**(.1652)**	**(.1713)**	**(.1713)**
Number of Obs.	1312	1312	1312	1312
Log Likelihood	Not Reported	-312.4701	-312.4701	312.4695

Table 4: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Rape			
	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	1.92	1.9227	1.9227	1.9226
	(.40)	**(.4065)**	(.3964)	(.3965)
Black	-.47	-.4669	-.4669	-.4669
	(.13)	(.1307)	(.1299)	(.1300)
Male	-.20	**-.2059**	**-.2059**	**-.2059**
	(.09)	(.0902)	(.0883)	(.0883)
Catholic	-.15	-.1503	-.1503	-.1503
	(.10)	(.0997)	(.0959)	(.0959)
Religious Intensity	-.17	-.1685	-.1685	-.1685
	(.14)	(.1437)	**(.1331)**	**(.1331)**
Attend Church	-.99	-.9859	-.9859	-.9858
	(.23)	(.2341)	(.2307)	(.2307)
Know What	-.14	-.1385	-.1385	-.1385
ERA Means	(.15)	(.1483)	(.1463)	(.1463)
Support ERA	.12	.1166	.1166	.1166
	(.14)	(.1381)	(.1342)	(.1342)
Variance Model				
Pro Count	-.19	-.1917	-.1917	-.1917
	(.09)	(.0890)	(.0902)	(.0902)
Con Count	.20	.1960	.1960	.1960
	(.12)	(.1226)	(.1226)	(.1226)
Pro Count ×	-.03	-.0269	-.0269	-.0269
Con Count	(.05)	(.0520)	(.0523)	(.0523)
Importance	.17	.1666	.1666	.1666
	(.15)	(.1507)	(.1534)	(.1534)
Information	-.13	-.1277	-.1277	-.1276
	(.14)	(.1370)	(.1398)	(.1398)
Firmness of Opinion	-.58	-.5824	-.5824	-.5825
	(.17)	(.1700)	(.1739)	(.1739)
Number of Obs.	1302	1302	1302	1302
Log Likelihood	Not Reported	-461.2838	-461.2838	-461.2843

Table 5: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Birth Defect			
	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	2.02	2.0173	2.0173	2.0173
	(.40)	(.4001)	(.4019)	(.4020)
Black	-.54	-.5385	-.5385	-.5385
	(.15)	(.1485)	(.1492)	(.1492)
Male	-.21	-.2071	-.2071	-.2071
	(.11)	(.1079)	(.1054)	(.1054)
Catholic	-.33	-.3279	-.3279	-.3279
	(.12)	(.1191)	(.1186)	(.1186)
Religious Intensity	-.51	-.5064	-.5064	-.5064
	(.19)	(.1893)	(.1875)	(.1875)
Attend Church	-.91	-.9081	-.9081	-.9081
	(.24)	(.2381)	(.2377)	(.2377)
Know What	.01	.0091	.0090	.0091
ERA Means	(.16)	(.1615)	(.1588)	(.1588)
Support ERA	.40	.4008	.4008	.4007
	(.18)	(.1766)	**(.1710)**	**(.1709)**
Variance Model				
Pro Count	-.06	-.0612	-.0612	-.0612
	(.08)	(.0795)	(.0804)	(.0804)
Con Count	.37	.3706	.3706	.3706
	(.12)	(.1233)	(.1248)	(.1247)
Pro Count $\times$	-.09	-.0931	-.0931	-.0931
Con Count	(.05)	(.0506)	(.0511)	(.0511)
Importance	-.14	-.1387	-.1387	-.1387
	(.16)	(.1615)	**(.1657)**	**(.1657)**
Information	.05	.0453	.0453	.0455
	(.14)	(.1440)	**(.1459)**	**(.1459)**
Firmness of Opinion	-.61	-.6102	-.6102	-.6101
	(.16)	(.1587)	(.1607)	(.1607)
N	1294	1294	1294	1294
Log Likelihood	Not Reported	-488.0805	-488.0805	-488.0800

Table 6: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Too Poor			
	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	**.02**	.0922	.0922	.0922
	(.01)	(.0823)	(.0798)	(.0798)
Black	-.09	-.0940	-.0940	-.0940
	(.06)	(.0573)	(.0576)	(.0576)
Male	-.04	-.0417	-.0417	-.0417
	(.04)	(.0400)	(.0392)	(.0392)
Catholic	.01	.0100	.0100	.0100
	(.04)	(.0401)	(.0395)	(.0395)
Religious Intensity	-.17	-.1703	-.1703	-.1703
	(.10)	(.0955)	(.0943)	(.0943)
Attend Church	-.35	-.3483	-.3483	-.3483
	(.17)	(.1714)	(.1717)	(.1717)
Know What	.10	.0958	.0958	.0958
ERA Means	(.08)	(.0743)	(.0734)	(.0734)
Support ERA	.22	.2188	.2188	.2188
	(.12)	(.1164)	(.1155)	(.1155)
Variance Model				
Pro Count	-.25	-.2472	-.2472	-.2472
	(.22)	(.2252)	(.2185)	(.2185)
Con Count	-.50	-.4982	-.4982	-.4983
	(.19)	(.1874)	(.1921)	(.1921)
Pro Count ×	.19	.1880	.1880	.1880
Con Count	(.11)	(.1094)	(.1087)	(.1087)
Importance	-.16	-.1629	-.1629	-.1629
	(.31)	**(.3047)**	(.3095)	(.3095)
Information	-.32	-.3168	-.3168	-.3168
	(.29)	(.2915)	**(.2983)**	**(.2983)**
Firmness of Opinion	.60	.6010	.6010	.6011
	(.58)	(.5839)	(.5751)	(.5751)
N	1291	1291	1291	1291
Log Likelihood	Not Reported	-822.8984	-822.8984	-822.8989

Table 7: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	No More Children			
	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	.03	.0304	.0304	.0304
	(.08)	(.0760)	(.0747)	(.0747)
Black	-.11	-.1118	-.1118	-.1118
	(.06)	(.0627)	(.0620)	(.0620)
Male	-.02	-.0227	-.0227	-.0227
	(.03)	(.0344)	(.0337)	(.0337)
Catholic	.02	.0242	.0242	.0242
	(.04)	(.0408)	(.0399)	(.0399)
Religious Intensity	-.13	-.1299	-.1299	-.1299
	(.69)	(.0741)	(.0733)	(.0733)
Attend Church	-.43	-.4303	-.4303	-.4304
	(.17)	**(.1896)**	**(.1889)**	**(.1889)**
Know What	.09	.0851	.0851	.0851
ERA Means	(.07)	(.0667)	(.0666)	(.0666)
Support ERA	.31	.3068	.3068	.3068
	(.13)	(.1341)	(.1332)	(.1332)
Variance Model				
Pro Count	-.26	-.2621	-.2621	-.2621
	(.18)	(.1820)	(.1813)	(.1813)
Con Count	-.58	-.5762	-.5762	-.5762
	(.17)	**(.1751)**	**(.1753)**	**(.1754)**
Pro Count ×	.25	.2468	.2468	.2467
Con Count	(.09)	**(.0957)**	**(.0957)**	**(.0957)**
Importance	-.18	-.1782	-.1782	-.1782
	(.26)	(.2646)	**(.2699)**	**(.2699)**
Information	-.28	-.2794	-.2794	-.2793
	(.25)	(.2527)	.(2521)	(.2521)
Firmness of Opinion	.47	.4734	.4734	.4731
	(.43)	**(.4551)**	**(.4365)**	**(.4364)**
Number of Obs.	1289	1289	1289	1289
Log Likelihood	Not Reported	-798.6787	-798.6787	-798.6786

Table 8: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Single			
	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	.11	.1114	.1114	.1114
	(.09)	(.0918)	(.0890)	(.0890)
Black	-.23	-.2260	-.2260	-.2260
	(.10)	(.0991)	(.0988)	(.0988)
Male	-.06	-.0623	-.0623	-.0623
	(.05)	(.0471)	(.0464)	(.0464)
Catholic	-.03	-.0263	-.0263	-.0263
	(.04)	**(.0471)**	(.0423)	(.0423)
Religious Intensity	-.18	-.1764	-.1764	-.1764
	(.09)	(.0893)	(.0892)	(.0892)
Attend Church	-.47	-.4711	-.4711	-.4711
	(.20)	(.2018)	(.2008)	(.2008)
Know What	.09	.0929	.0929	.0929
ERA Means	(.08)	(.0780)	(.0748)	(.0748)
Support ERA	.31	.3121	.3121	.3121
	(.13)	(.1337)	(.1328)	(.1328)
Variance Model				
Pro Count	-.34	-.3424	-.3424	-.3424
	(.17)	(.1659)	(.1670)	(.1670)
Con Count	-.41	-.4060	-.4060	-.4060
	(.16)	(.1611)	(.1619)	(.1619)
Pro Count ×	.21	.2116	.2116	.2116
Con Count	(.08)	(.0823)	(.0826)	(.0826)
Importance	-.24	-.2364	-.2364	-.2364
	(.25)	**(.2353)**	**(.2615)**	**(.2615)**
Information	-.28	-.2817	-.2817	-.2817
	(.24)	(.2372)	**(.2455)**	**(.2455)**
Firmness of Opinion	1.81	1.8081	1.8081	1.8081
	(.67)	**(.6565)**	**(.6554)**	**(.6554)**
N	1293	1293	1293	1293
Log Likelihood	Not Reported	-793.6072	-793.6072	-793.6072

Table 9: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Any Reason			
	Published	Replication with Archived Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	-.07	-0.0704	-.0704	-.0704
	(.13)	(.1333)	(.1303)	(.1303)
Black	-.15	-.1527	-.1527	-.1527
	(.09)	(.0895)	(.0879)	(.0879)
Male	-.13	-1268	-.1268	-.1267
	(.07)	(.0675)	(.0650)	(.0650)
Catholic	.05	.0502	.0502	.0502
	(.07)	(.0719)	(.0707)	(.0707)
Religious Intensity	-.22	-.2152	-.2152	-.2152
	(.12)	(.1160)	**(.1138)**	**(.1138)**
Attend Church	-.79	-.7902	-.7902	-.7901
	(.26)	**(.2541)**	**(.2505)**	**(.2504)**
Know What ERA Means	.12	.1155	.1155	.1155
	(.10)	(.1017)	(.0990)	(.0990)
Support ERA	.51	.5120	.5120	.5120
	(.17)	(.1714)	(.1684)	.1684)
Variance Model				
Pro Count	-.22	-.2237	-.2237	-.2237
	(.15)	(.1516)	(.1521)	(.1521)
Con Count	-.48	-.4832	-.4832	-.4831
	(.14)	(.1432)	(.1440)	(.1440)
Pro Count × Con Count	.22	.2185	.2185	.2185
	(.08)	(.0755)	(.0770)	(.0770)
Importance	-.30	-.3008	-.3008	-.3009
	(.25)	(.2524)	(.2547)	(.2547)
Information	.68	.6760	.6760	.6759
	(.23)	(.2310)	(.2345)	(.2345)
Firmness of Opinion	.63	.6315	.6315	.6315
	(.38)	(.3800)	(.3805)	(.3805)
N	1295	1295	1295	1295
Log Likelihood	Not Reported	-768.9695	-768.9695	-768.9697

Table 10: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Mothers' Health			
	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	2.55	**2.2697**	**2.7275**	**2.7275**
	(.46)	**(.2433)**	**(.5184)**	**(.5184)**
Black	-.51	**-.2614**	**-.5035**	**-.5035**
	(.14)	**(.0325)**	**(.1579)**	**(.1579)**
Male	-.08	**.9516**	**-.0732**	**-.0732**
	(.11)	**(.1266)**	**(.1204)**	**(.1204)**
Catholic	-.52	**-.2326**	**-.5174**	**-.5174**
	(.13)	**(.0275)**	**(.1467)**	**(.1467)**
Religious Intensity	-.39	**-.3105**	**-.5837**	**-.5837**
	(.20)	**(.2561)**	**(.2943)**	**(.2943)**
Attend Church	-1.04	**-1.8679**	**-1.0726**	**-1.0726**
	(.25)	**(.1876)**	**(.2720)**	**(.2720)**
Know What	**-.18**	**.0867**	**-.0156**	**-.0156**
ERA Means	(.17)	**(.1457)**	**(.1923)**	**(.1923)**
Support ERA	.33	**.1498**	**.3094**	**.3094**
	(.17)	**(.0764)**	**(.2004)**	**0.2004**
Variance Model				
Pro Count	-.14	**-1.3138**	**-.1531**	**-.1531**
	(.07)	**(.0000)**	**(.0752)**	**(.0752)**
Con Count	.17	**1.5610**	**.1535**	**.1535**
	(.09)	**(.0367)**	**(.0954)**	**(.0954)**
Pro Count ×	**-.44**	**-.6336**	**-.0373**	**-.0373**
Con Count	(.04)	**(.0000)**	**(.0434)**	**(.0435)**
Importance	**.51**	**1.2511**	**-.0227**	**-.0227**
	(.15)	**(.1762)**	**(.1412)**	**(.1412)**
Information	.37	**1.0113**	**.3930**	**.3931**
	(.13)	**(.1762)**	**(.1301)**	**(.1301)**
Firmness of Opinion	-.37	**-1.5552**	**-.3710**	**-.3710**
	(.16)	**.2426**	**(.1704)**	**(.1704)**
Number of Obs.	1312	1422	1422	1412
Log Likelihood	Not Reported	-324.6417	-317.4736	-317.4743

Table 11: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Rape			
	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	1.92	**1.9537**	**1.9537**	**1.9538**
	(.40)	(.3933)	(.3938)	(.3938)
Black	-.47	**-.4644**	**-.4644**	**-.4644**
	(.13)	**(.1239)**	**(.1237)**	**(.1237)**
Male	-.20	-.2009	-.2009	-.2009
	(.09)	**(.0827)**	**(.0826)**	**(.0826)**
Catholic	-.15	-.1505	-.1505	-.1506
	(.10)	**(.0936)**	**(.0918)**	**(.0918)**
Religious Intensity	-.17	**-.2471**	**-.2471**	**-.2470**
	(.14)	**.1670**	**(.1669)**	**(.1669)**
Attend Church	-.99	**-.9487**	**-.9487**	**-.9488**
	(.23)	**(.2185)**	**(.2174)**	**(.2174)**
Know What	-.14	**-.1513**	**-.1513**	**-.1513**
ERA Means	(.15)	**(.1421)**	**(.1397)**	**(.1397)**
Support ERA	.12	.1161	.1161	.1161
	(.14)	**(.1332)**	**(.1270)**	**(.1270)**
Variance Model				
Pro Count	-.19	**-.2180**	**-.2180**	**-.2180**
	(.09)	(.0875)	(.0874)	(.0874)
Con Count	.20	**.1861**	**.1861**	**.1862**
	(.12)	**(.1139)**	**(.1163)**	**(.1163)**
Pro Count ×	-.03	**-.0212**	**-.0212**	**-.0213**
Con Count	(.05)	(.0501)	(.0504)	(.0504)
Importance	.17	**.1470**	**.1470**	**.1466**
	(.15)	**(.1403)**	(.1458)	(.1458)
Information	-.13	**-.0935**	**-.0935**	**-.0933**
	(.14)	**.1304**	**(.1320)**	**(.1320)**
Firmness of Opinion	-.58	**-.6004**	**-.6004**	**-.6005**
	(.17)	(.1655)	(.1712)	(.1712)
Number of Obs.	1302	1412	1412	1412
Log Likelihood	Not Reported	-478.4871	-478.4871	-478.4873

Table 12: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

		Birth Defect		
	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	2.02	2.0248	2.0248	2.0251
	(.40)	**(.3874)**	(.3991)	(.3992)
Black	-.54	-.5413	-.5412	-.5414
	(.15)	**(.1400)**	**(.1434)**	**(.1434)**
Male	-.21	**-.1859**	**-.1859**	**-.1860**
	(.11)	**(.0972)**	**(.0974)**	**(.0974)**
Catholic	-.33	**-.3105**	**-.3105**	**-.3106**
	(.12)	**(.1115)**	**(.1117)**	**(.1117)**
Religious Intensity	-.51	**-.5979**	**-.5979**	**-.5979**
	(.19)	**(.2128)**	**(.2266)**	**(.2266)**
Attend Church	-.91	**-.8448**	**-.8448**	**-.8448**
	(.24)	**(.2178)**	**(.2195)**	**(.2195)**
Know What	.01	-.0081	-.0081	-.0081
ERA Means	(.16)	**(.1431)**	**(.1485)**	**(.1485)**
Support ERA	.40	**.4329**	**.4329**	**.4329**
	(.18)	**(.1609)**	**(.1640)**	**(.1640)**
Variance Model				
Pro Count	-.06	**-.0900**	**-.0900**	**-.0899**
	(.08)	(.0794)	(.0788)	(.0788)
Con Count	.37	**.3608**	**.3608**	**.3608**
	(.12)	(.1238)	(.1203)	(.1203)
Pro Count ×	-.09	-.0894	-.0894	-.0894
Con Count	(.05)	(.0519)	(.0499)	(.0499)
Importance	-.14	-.1362	-.1362	-.1361
	(.16)	**(.1495)**	**(.1535)**	**(.1535)**
Information	.05	**.0622**	**.0622**	**.0623**
	(.14)	**(.1303)**	(.1362)	(.1362)
Firmness of Opinion	-.61	**-.6306**	**-.6306**	**-.6306**
	(.16)	(.1594)	(.1575)	(.1575)
N	1294	1405	1405	1405
Log Likelihood	Not Reported	-509.3861	-509.3861	-509.3856

Table 13: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Too Poor			
	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	**.02**	**0.2911**	**.2911**	**.2911**
	(.01)	**(.1396)**	**(.1388)**	**(.1389)**
Black	-.09	**-.1371**	**-.1371**	**-.1371**
	(.06)	**(.0713)**	**(.0711)**	**(.0711)**
Male	-.04	**-.0539**	**-.0539**	**-.0539**
	(.04)	**(.0475)**	**(.0467)**	**(.0467)**
Catholic	.01	.0123	.0123	.0123
	(.04)	**(.0505)**	**(.0489)**	**(.0489)**
Religious Intensity	-.17	**-.4003**	**-.4003**	**-.4003**
	(.10)	**(.1693)**	**(.1679)**	**(.1679)**
Attend Church	-.35	**-.4495**	**-.4494**	**-.4495**
	(.17)	(.1728)	(.1729)	(.1729)
Know What	.10	.1009	.1009	.1009
ERA Means	(.08)	(.0828)	(.0818)	(.0818)
Support ERA	.22	.3086	**.3086**	.3086
	(.12)	**(.1317)**	**(.1313)**	**(.1313)**
Variance Model				
Pro Count	-.25	**-.0645**	**-.0645**	**-.0645**
	(.22)	**(.1735)**	**(.1755)**	**(.1755)**
Con Count	-.50	**-.3891**	**-.3891**	**-.3891**
	(.19)	**(.1561)**	**(.1596)**	**(.1596)**
Pro Count ×	.19	**.0848**	**.0848**	**.0848**
Con Count	(.11)	**(.0864)**	**(.0875)**	**(.0875)**
Importance	-.16	**.0685**	**.0685**	**.0685**
	(.31)	**(.2556)**	**(.2646)**	**(.2646)**
Information	-.32	**-.2309**	**-.2309**	**-.2310**
	(.29)	**(.2173)**	**(.2325)**	**(.2325)**
Firmness of Opinion	.60	**.1366**	**.1366**	**.1366**
	(.58)	**(.4458)**	**(.4043)**	**(.4044)**
N	1291	1401	1401	1401
Log Likelihood	Not Reported	-875.9104	-875.9104	-875.9108

Table 14: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	No More Children			
	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	.03	**0.0974**	**.0974**	**0.0974**
	(.08)	**(.0918)**	**(.0900)**	**(.0900)**
Black	-.11	**-.1549**	**-.1549**	**-.1549**
	(.06)	**(.0704)**	**(.0732)**	**(.0732)**
Male	-.02	-.0249	-.0249	-.0249
	(.03)	**(.0362)**	**(.0359)**	**(.0359)**
Catholic	.02	.0176	.0176	.0176
	(.04)	(.0425)	(.0413)	(.0413)
Religious Intensity	-.13	**-.2341**	**-.2341**	**-.2340**
	(.69)	**(.1122)**	**(.1119)**	**(.1119)**
Attend Church	-.43	**-.4580**	**-.4580**	**-.4579**
	(.17)	**(.1771)**	**(.1813)**	**(.1813)**
Know What ERA Means	.09	.0921	.0921	.0921
	(.07)	(.0682)	(.0703)	(.0703)
Support ERA	.31	**.3843**	**.3843**	**.3842**
	(.13)	**(.1457)**	**(.1490)**	**(.1490)**
Variance Model				
Pro Count	-.26	**-.1959**	**-.1959**	**-.1959**
	(.18)	**(.1550)**	**(.1610)**	**(.1610)**
Con Count	-.58	**-.4994**	**-.4994**	**-.4994**
	(.17)	**(.1570)**	**(.1619)**	**(.1619)**
Pro Count × Con Count	.25	**.1979**	**.1979**	**.1979**
	(.09)	**(.0821)**	**(.0848)**	**(.0849)**
Importance	-.18	**-.0187**	**-.0187**	**-.0187**
	(.26)	**(.2372)**	**(.2360)**	**(.2361)**
Information	-.28	**-.4311**	**-.4311**	**-.4311**
	(.25)	**(.2097)**	**(.2146)**	**(.2146)**
Firmness of Opinion	.47	**.4620**	**.4620**	**.4620**
	(.43)	**(.3751)**	**(.3850)**	**(.3850)**
Number of Obs.	1289	1399	1399	1399
Log Likelihood	Not Reported	-850.8463	-850.8463	-850.8466

Table 15: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Single			
	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	.11	**.1689**	**.1689**	**.1689**
	(.09)	**(.1097)**	**(.1114)**	**(.1114)**
Black	-.23	**-.2688**	**-.2688**	**-.2688**
	(.10)	(.1041)	**(.1098)**	**(.1098)**
Male	-.06	**-.0711**	**-.0711**	**-.0711**
	(.05)	(.0487)	(.0492)	(.0492)
Catholic	-.03	-.0284	-.0284	-.0284
	(.04)	**(.0459)**	**(.0462)**	**(.0462)**
Religious Intensity	-.18	**-.2508**	**-.2508**	**-.2508**
	(.09)	**(.1202)**	**(.1245)**	**(.1245)**
Attend Church	-.47	**-.5053**	**-.5053**	**-.5053**
	(.20)	**(.1895)**	(.1998)	(.1998)
Know What	.09	.0891	.0891	.0891
ERA Means	(.08)	(.0761)	(.0772)	(.0772)
Support ERA	.31	**.3852**	**.3852**	**.3852**
	(.13)	**(.1445)**	**(.1509)**	**(.1509)**
Variance Model				
Pro Count	-.34	**-.3223**	**-.3223**	**-.3223**
	(.17)	**(.1540)**	**(.1594)**	**(.1594)**
Con Count	-.41	**-.3759**	**-.3759**	**-.3759**
	(.16)	**(.1522)**	(.1573)	(.1573)
Pro Count $\times$	.21	**.1883**	**.1883**	**.1883**
Con Count	(.08)	(.0762)	(.0794)	(.0794)
Importance	-.24	**-.0936**	**-.0936**	**-.0936**
	(.25)	**(.2296)**	**(.2388)**	**(.2388)**
Information	-.28	**-.2686**	**-.2686**	**-.2686**
	(.24)	**(.2062)**	**(.2154)**	**(.2154)**
Firmness of Opinion	1.81	**1.7063**	**1.7063**	**1.7064**
	(.67)	**(.5554)**	**(.5790)**	**(.5790)**
N	1293	1403	1403	1403
Log Likelihood	Not Reported	-852.1331	-852.1331	-852.1331

Table 16: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy

	Published	Replication with Reference Data		
		Replication 1	Replication 2	Replication 3
Any Reason	(SHAZAM)	(SHAZAM)	(Stata)	(Gauss)
Constant	-.07	**.0354**	**.0354**	**.0353**
	(.13)	(.1296)	**(.1197)**	**(.1197)**
Black	-.15	**-.2049**	**-.2048**	**-.2048**
	(.09)	(.0926)	(.0900)	(.0900)
Male	-.13	-.1292	-.1292	-.1292
	(.07)	**(.0627)**	**(.0616)**	**(.0616)**
Catholic	.05	**(.0241)**	**.0241**	**.0241**
	(.07)	(.0675)	**.0634**	**(.0634)**
Religious Intensity	-.22	**-.3698**	**-.3698**	**-.3697**
	(.12)	**(.1529)**	**(.1497)**	**(.1497)**
Attend Church	-.79	**-.6880**	**-.6880**	**-.6880**
	(.26)	**(.2287)**	**(.2197)**	**(.2197)**
Know What	.12	**.1344**	**.1344**	**.1345**
ERA Means	(.10)	**(.1075)**	(.0950)	(.0950)
Support ERA	.51	**.5370**	**.5370**	**.5370**
	(.17)	(.1724)	(.1694)	(.1695)
Variance Model				
Pro Count	-.22	-.2249	-.2249	-.2249
	(.15)	**(.1447)**	**(.1449)**	**(.1449)**
Con Count	-.48	**-.4543**	**-.4543**	**-.4543**
	(.14)	**(.1405)**	**(.1399)**	**(.1399)**
Pro Count ×	.22	**(.1982)**	**.1982**	**.1982**
Con Count	(.08)	**(.0736)**	**(.0737)**	**(.0737)**
Importance	-.30	**-.2623**	**-.2623**	**-.2625**
	(.25)	**(.2397)**	**(.2400)**	**(.2400)**
Information	.68	**.4815**	**.4815**	**.4813**
	(.23)	**(.2054)**	**(.2102)**	**(.2102)**
Firmness of Opinion	.63	**.6824**	**.6823**	**.6823**
	(.38)	**(.3536)**	**(.3671)**	**(.3671)**
N	1295	1402	1402	1402
Log Likelihood	Not Reported	-832.4434	-832.4434	-832.4431

Table 17: Heteroskedastic Probit Estimates of Attitudes Toward Abortion Policy