REEVALUATING GENERALIZED VARIANCE MODEL PARAMETERS FOR THE NATIONAL CRIME VICTIMIZATION SURVEY

Thomas R. Krenzke[1]
U.S. Bureau of the Census, Washington, D.C. 20233

KEY WORDS: Jackknife, Design Effect, Regression

1. Introduction

The National Crime Victimization Survey (NCVS) is a large household survey conducted by the Census Bureau to measure victimization totals and rates for many types of crimes and for many subpopulations. A useful measure of reliability of the estimates is the relative variance (relvariance), which is the variance of an estimate divided by the estimate squared. However, to provide relvariances for each of the published estimates would be impractical due to the large sample and complex design. This leads to calculating relvariances for a practical number of estimates, then modeling the relationship between the estimated relvariances and the estimated crime totals. Estimated variance model parameters are provided to the user in which they can enter an estimate of interest to get the resulting variance estimates. The conductors of the survey benefit from savings in publication costs and costs relating to variance calculations, and it has the appealing characteristic of smoothing the directly calculated variance estimates.

The NCVS implemented new survey methods and the 1993 NCVS crime estimates are the first to reflect the enhancements. The new methods include redefinition of crime categories, more use of computer-assisted telephone interviewing (CATI), and the addition of new questions. As a result, respondents reported more crime incidents than under the old methods. The last set of variance model parameters were developed from data collected in 1988 under old survey methods. Since the redesign may have caused changes in the clustering of person and property crimes, new generalized variance function (GVF) parameters were needed from 1994 data, the first year when a full sample was available using new methods.

2. NCVS Sample Design and Estimation

The NCVS uses a stratified, multi-stage, cluster sample where the target population includes persons aged 12 and older, excluding crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates (McGinn and Monahan 1990). The first stage of sample selection is the selection of primary sampling units (PSUs). The 89 metropolitan areas with over 250,000 population are selected with certainty and are called self-representing (SR) PSUs. The remaining 153 sample PSUs are called non-self-representing (NSR) PSUs. Homogenous groups of NSR PSUs are assigned to the same strata. Within a stratum, one NSR PSU is taken with probability proportionate to its population size. With a sampling interval (SI) of one housing unit in about 1,800 across the entire nation, the second-stage sampling units are selected. This stage is conducted in two steps. First, Enumeration Districts (EDs), which contain about 300 housing units on average, are selected systematically from a list of EDs where its probability of selection is proportionate to its population size. Within a selected ED, segments of about four housing units each were randomly selected. Basically, segments are constructed within the selected EDs through address lists from the decennial census, lists of new permits for new construction, area sampling approaches, and lists of special places like dormitories. Each housing unit within the selected segment is included in the sample and each person aged 12 and over are interviewed. About 10,000 housing units are interviewed monthly as part of an ongoing group of housing units called a panel rotation group. A panel rotation group is interviewed seven times (every six months) within three years and then rotated out of the sample.

3. Choice of Variables for Variance Purposes

The model's parameters for the NCVS are estimated using estimated relvariances and estimates of crime totals from a subset of the many possible survey variables. These variables are specifically chosen for the model building process and the choice of variables is important since they have an effect on the model parameters. It seems best to have a set of data points that are equally distributed across subpopulations and across equal length intervals in the range of the estimated totals. Keeping this in mind for NCVS variance purposes, variances were estimated directly for 840 total variance estimates, which consisted of 520 person estimates and 320 property estimates. Person estimates were associated with crimes committed against people (i.e., assault, robbery, rape, etc.). Property estimates were associated with crimes committed to properties (i.e., theft, motor vehicle theft, burglary, etc.). Of the 520 person estimates, 26 were overall person crimes, and the remaining 494 were called person domain estimates, which were overall person crime estimates broken out by 19 person domains (domains were males, hispanics, reported to police, urban, etc.). Of the 320 property estimates, 16 were overall property crimes estimates, and the remaining 304 were called property domain estimates, which were the overall property crimes estimates broken out by 19 property domains.

4. Direct Variance Estimates for the NCVS

Due to the complexities in the sample design and for the reasons cited in NCVS variance research documentation by Weidman and Williams (1992), direct variance estimates using 1994 new methods data were calculated by the stratified jackknife technique as they were when using 1988 old methods data. This is a replication procedure where reduced sample estimates are attained by dropping one standard error computational unit (SECU) or cluster within a pseudostratum or stratum, and reweighting the remaining SECUs within the pseudostratum. The variance among these reduced sample estimates is measured. SECU codes and pseudostratum numbers were assigned differently depending on the component of variance being calculated. Code assignment procedures follow closely to what is written in Weidman 1993. Three components of variance were initially estimated,

1) within (or total) SR PSU variance, 8223 replicates,

2) total NSR PSU variance, 153 replicates,

3) within NSR PSU variance, 5865 replicates.

With these variance components, the following were derived (see partial results in Table 1),

1) total variance = within SR PSU variance + total NSR PSU variance,

2) between PSU variance = total NSR PSU variance - within NSR PSU variance.

An overall measure of how much the sampling scheme affects the total variance is the design effect (DEFF). The design effect is the variance estimate from the complex design divided by the variance estimate from a simple random sample (SRS). When DEFF is computed, one knows that in order to have the same precision with the complex approach as with the SRS of size n, it would take DEFF * n individuals. Basically, in the NCVS, the DEFF is greater for types of crime that occur more frequently as shown in Table 1. A further look into the largest 20 DEFFs shows that the design affects variances associated with geographic domains the greatest since 10 out of the largest 11 DEFFs are for rural area estimates and 15 out of the largest 20 DEFFs are associated with geographic estimates. The traditional two-parameter GVF for the NCVS assumes that there is a constant design effect. The new three parameter model attempts to address the varying DEFF and improves the fit, however there is still some room for improvement. One problem is that both models use a binomial variance instead of a multinomial variance, that is, both models assume that a person reports zero or one incident when in reality there can be multiple incidents reported per person. Components of the DEFF were examined for a better understanding of why the DEFF varies, and to give basis for sample design improvements and for sample design development for supplemental surveys. In addition, it may result in an improved generalized variance model. Alexander and Hubble, 1990, list and explain several reasons for the varying DEFF across crime categories. This work is an attempt to quantify three of the effects: between PSU variance, multiplicity of crimes, and clustering of crimes within segments. Other effects include systematic sampling, rotation pattern, and adjustments made in survey weighting.

1) Between PSU variance -- The DEFF due to between PSU variance was calculated to show how the design affects the variances with regards to between PSU variance. This can be calculated as the ratio of total variance to the total variance without between PSU variance. The results in Table 1 show that the DEFFs due to between PSU variance seem to vary as the overall DEFFs do, therefore explaining some of the variation in DEFF among crime categories.

2) Effect of multiplicity of crimes -- A person or household may be victim of a crime more that once. Due to this multiplicity, the crime rate is not a true proportion. The multiplicity of crimes has an increasing effect on variances, and we can measure the effect using the ratio of the variance from the

multinomial distribution to the variance from the binomial distribution. The results in Table 1 show that the effect of multiplicity seems to vary the most across crime categories, therefore explaining some of the variation in DEFF among crime categories.

3) Effect of clustering of crimes within segments -- The intracluster correlation coefficient in this case refers to the correlation among persons within segments for person crimes, and the correlation among housing units within segments for property crimes. Correlation among elements within the cluster has an increasing effect on the variance. The underlying theory of intracluster correlation can be found in Hansen, Hurwitz, and Madow (1953), where the DEFF due to clustering is, $DEFF_{(clus)} = 1 + (\bar{n} - 1)$ roh, where $\bar{n}$ is the average cluster size (7.65 for person crimes, 3.59 for property crimes), and roh is the estimated intracluster correlation coefficient. The results in Table 1 show that although the effect of clustering seems to explain only a small portion of the variability of DEFFs among crime categories, it explains a good portion of the overall DEFF for each crime category.

Since most of the crime categories have a residual DEFF slightly greater than one, there are still some effects that are not measured that may explain the remaining difference from the variance under simple random sampling. Details of these calculations are presented in the longer version of this paper.

5. Generalized Variance Function Parameters

The directly calculated variance estimates are generalized into variance models. The objectives of the variance modeling process were the following:

1) Create the model such that the estimates that are the most reliable receive the most weight. The most reliable estimates tend to be crime estimates that are large.

2) Keep the model simple. Two or three parameters should be sufficient to keep the predicted relative variances within about 10% of the observed relative variances, on average.

3) Construct the process so that the relvariances associated with the overall crime estimates are not overestimated by a large amount nor consistently underestimated. These estimates are used the most in press releases since they are of national interest.

4) Create a model for each differing domain, keeping the number of models to a minimum for the main publication. Since there are special reports on domains, generate variance model parameter estimates for these domains.

In the construction of GVFs, Wolter (1985) states that the resulting variance model should be the one that provides the best empirical fit to the data. In this work, diagnostic tests generally consisted of investigating outliers and plotting the standardized residuals with the predicted values. Generally speaking, cases with an absolute standardized residual greater than 3 were outlying observations. If an outlying case was influential on the model (i.e., absolute value of DFFITS > .25) then its exclusion from the model building process was considered. Most of the influential cases that were excluded were estimates associated with rural areas. As mentioned before, rural cases have the largest DEFFs and the largest proportion of total variance explained by between PSU variance, and this explains why data points relating to rural areas are different from the main stream of the data.

Models were evaluated according to tools such as the adjusted R^2 and the mean absolute relative deviation (ARD). Another application of the ARD can be found in Bieler and Williams (1990) where it was applied to the 1988 National Household Survey on Drug Abuse. The ARD is the absolute value of the prediction error, $(V^2_{predicted} - V^2_{observed}) / V^2_{observed}$, where, $V^2_{observed}$ is the observed relative variance, and $V^2_{predicted}$ is the predicted relative variance.

To achieve objective 1) of giving the most weight to the estimated totals that we have the most confidence in, iterative weighted least squares (WLS) regression was used. The initial weights were set at $1/(V^2_{observed})^2$ using the observed relative variances. In the subsequent iterations, the weights were $1/(V^2_{predicted})^2$. The maximum number of iterations allowed was five since there are small marginal improvements or no improvement thereafter. For addressing the 2nd objective of keeping the model simple, two of the models evaluated were:

A) Traditional NCVS model, $V^{2\prime}\{X'\} = a + b / X'$,

B) Resulting NCVS model from 1988 (under old survey methods), $V^{2\prime}\{X'\} = a + b / X' + c / X^{1/2\prime}$.

Assuming the crime rate is a proportion, where $p = X' / Y$ (this is not necessarily accurate due to multiplicity of crimes), and assuming Y is a constant, the variance of p can be written as a function of a constant DEFF, Var(p) = DEFF {p (1 - p) / n}, then substituting (X' / Y) for p, and after some algebra, the relvariance of X' is, V^{2}'{X'} = -(1/n) + (Y/n)(DEFF/X'), which is of the form of Model A. So Model A assumes a constant DEFF. The third term in Model B attempts to adjust Model A in order to take into account the nonconstant DEFF. For groups of estimates that have the same base, Y, the only element that changes in the calculation of DEFF is the crime estimate, X'. Therefore, the third term of Model B should help improve the fit for these groups.

Models were fit to all 840 estimates. The results of the WLS regression runs showed that Model B had the lowest mean ARD (17.97) (19.82 for Model A) and the highest adjusted R^2 (.9178) (.8926 for Model A), therefore Model B had the most potential and the decision was made to use Model B. This was the form of the variance model that resulted in an independent study using 1988 data under the old survey methods.

There was a logical split in the list of 840 variance estimates between the person and property estimates. However, the decision to separate person and property estimates was based on the data. A plot illustrated a curve fit to all 840 estimates (excluding outliers) with data points associated with person estimates shown. The curve fits the person estimates well, however, another plot exposed that the curve missed the main stream of the property estimates. This was supported by the fact that 79.6% of the observed relvariances associated with property estimates were overestimated by the regression curve. Thus the decision was made to treat person and property estimates separately. Prediction curves for the person estimates model and the property estimates model were plotted and it was clear that the property estimates curve predicts relvariances smaller than that of the person curve for estimates of the same magnitude. This is due mostly to the estimated totals relating to property crimes being larger in magnitude as a group than estimated totals relating to person crimes. The fit of Model B for the property estimates had improved since the mean ARD for the property estimates in the all-estimates model was 25.20 and the mean ARD for the property estimates in the property-estimates specific model was 17.25. Since the all-crimes model fit the person estimates well, there was no real improvement to the fit by developing a model specifically to the person crime estimates.

The 3rd objective is concerned with the overall crime estimates (i.e. the crime estimates that were not disaggregated by domains). Plots illustrate the improvement in the fit to the overall person crime estimates by developing a model specific to that group. Using the model fit to the 520 person estimates, the mean ARD was 15.86. Using the model fit specifically to the 26 overall person crimes, the mean ARD improved to 6.21. Using the model fit to the 320 property estimates, the mean ARD was 21.62. Using the model fit specifically to the 16 overall property crimes, the mean ARD improved to 5.36. The models fit specifically to the 26 overall person crimes and the 16 overall property crimes are the first two sets of variance parameters recommended. These overall crime estimates are found in Table 1 of the Bureau of Justice Statistics annual publication of NCVS estimates.

A curve was fit to the remaining 494 person estimates, called person domain estimates, and also to the remaining 304 property estimates, called property domain estimates. These correspond to the third and fourth sets of variance parameters recommended. The estimates to which these parameters apply appear in all tables of the annual publication except Table 1.

The 4th objective corresponds to generating variance model parameters for each domain of interest. It is not generally preferred in this survey to have models available for each domain of interest mostly because of the confusion of which model to use when domains are crossed. However, domain-specific models were constructed in order to have available these GVFs for the special domain-specific publications. The relationship between relvariances and totals is different for some domains. In order to identify differing domains, plots were used to see if any domains were consistently above or below the prediction curve. Differing domains were also identified by listing the percentage of observed relvariances that were overestimated by the model.

In Table 2, the highlighted parameters will be made available for the NCVS estimates for general public use. Table 2 also presents the GVF results for the various domains of interest. Large improvements in fit were seen in rural areas. However, as mentioned earlier, it had been decided that the number of domain specific models be limited, so these listed domain specific models will be retained for use in special reports only. The models for the 26 overall person crimes and the 16 overall property crimes were not verified because all possible estimates were already exhausted. A separate set of variances was estimated for the purpose of verifying models for person domain and property domain estimates and the mean ARD results for each curve were close to the mean ARD results for the modeled data.

The 1994 variance model parameters generally produce larger variance estimates than the parameters from 1988. Most of this increase can be explained by the new methods, especially the redefinition of crimes and the new questions that resulted in an increase in the number of reported crime incidents. Since the number of incidents reported increased, but the number of sampling units remained the same, the new methods seemed to have caused an increase in clustering, and consequently, an increase in variance.

6. Summary

Variables were chosen with care for the variance modeling process. Direct variances were calculated by the stratified jackknife method for selected crime totals. Three components of the design effects, namely, between PSU variance, effect of multiplicity, and effect of clustering, were produced and analyzed in an attempt to explain the nonconstant DEFF that exists in the NCVS. The resulting three-parameter model attempts to address the issue of the varying DEFF among crime categories. The three-parameter model is an improvement on the traditional two-parameter model, which assumes a constant DEFF. The direct variances were generalized using data collected in 1994, the first full year of new survey methods. Objectives for the process were listed. In meeting these objectives, four sets of variance parameters were generated using the three-parameter model for incidents relating to overall person crimes, person domain estimates, overall property crimes, and property domain estimates. GVFs were also developed for special reports on domains of interest. The models were verified by a subset of NCVS data.

7. Acknowledgements

This work could not have been completed without the helpful guidance from David Hubble, Branch Chief of the Victimization and Expenditure Branch (VEB) within the Demographic Statistical Methods Division (DSMD) of the Census Bureau, and Marjorie Dauphin, Supervisory Mathematical Statistician within VEB. Assistance with VPLX programming from George Train III of DSMD and the suggestions and comments from the reviewers is also greatly appreciated. The SAS system was used for all regression runs. VPLX, developed by Robert Fay of the Census Bureau, was used for calculating variance estimates.

[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

8. References

Alexander, C. and Hubble, D. (1990), "Sample Weights for the NCS and NCS Estimation," 1990 Workshop on the Design and Use of the National Crime Survey, Sponsored by the Bureau of Justice Statistics and the Committee on Law and Justice Statistics of the American Statistical Association.

Bieler, G. and Williams, R. (1990), "Generalized Standard Error Models for Proportions in Complex Design Surveys," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Hansen, Hurwitz, and Madow (1952), *Sample Survey Methods and Theory, Volume I*, John Wiley & Sons, Inc.

McGinn, L. and Monahan, M., (1990), "NCS Design," 1990 Workshop on the Design and Use of the National Crime Survey, Sponsored by the Bureau of Justice Statistics and the Committee on Law and Justice Statistics of the American Statistical Association.

Shapiro, G., (1982), "NCS Redesign: Overview of Intraclass Correlation Project and Specification for First Regression Run," Internal Memorandum to Robert T. O'Reagan, U.S. Bureau of the Census, August 2, 1982.

Weidman, L. (1993), "Variance Estimation Procedures for the National Crime Survey -- Collection Year 1988 -- Revised," Internal Memorandum to Charles Alexander, U.S. Bureau of the Census, March 29, 1993.

Weidman, L. and Williams, T. (1992), "Estimating Variance Components for a Complex Survey," *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Wolter, K. M. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.

Table 1. Variance Components for Crime Categories

Crime	Estimate	Standard Error	% of Total Variance Betw. PSU	With. NSR PSU	With. SR PSU	ROH	DEFF	DEFFs Due To Betw. PSU	Mult.	Clus.	Residual DEFF
All Pers.	11534068	282359	25.7	27.2	47.1	0.047	3.97	1.35	1.71	1.31	1.32
Violence	11031102	278034	27.3	27.3	45.3	0.045	4.02	1.38	1.71	1.30	1.32
Rape/SA	443509	37057	0.0	50.3	52.6	0.029	1.69	1.00	1.32	1.20	1.07
Robbery	1353267	67129	0.3	20.1	79.6	0.030	1.82	1.00	1.28	1.20	1.18
Assault	9234326	255804	28.5	27.8	43.6	0.040	4.03	1.40	1.66	1.27	1.36
All Prop.	31192210	450594	29.8	24.8	45.4	0.095	5.09	1.42	2.32	1.25	1.24
Burglary	5658726	143723	9.7	35.4	54.9	0.040	2.10	1.11	1.49	1.10	1.15
Mot.Veh.T	1887411	73048	1.5	17.0	81.5	0.020	1.57	1.02	1.18	1.05	1.24
Theft	23646072	371733	29.4	25.7	45.0	0.079	4.12	1.42	1.99	1.20	1.21

Table 2. GVF Results for Domains of Interest

Mean ARD* is for the subgroup applied to the next highest aggregate model.

Domains of Interest	a	b	c	Adj. R^2	Mean ARD	Mean ARD*	# Obs.
For Person Estimates							
Overall Person Crimes	.00021742	2362	0.777	.9840	6.21	15.86	26
All Person Domain Est.	-.00000335	2340	1.671	.9553	12.87	12.85	494
Low Income	-.00003558	2191	2.652	.9677	10.43	14.28	26
Medium Income	.00044261	2493	0.489	.9831	8.31	12.36	26
High Income	.00002932	2432	1.684	.9829	6.63	6.89	26
Hispanics	.00212027	2278	1.389	.9725	8.15	13.25	26
Urban	.00106910	2247	1.632	.9721	8.63	17.05	26
Suburban	.00057046	2391	1.136	.9806	7.88	11.19	26
Rural	.00784133	2194	1.449	.9535	9.42	33.83	26
Blacks	-.00006058	2921	1.776	.9825	6.76	16.21	26
Reported to Police	.00006827	2293	1.099	.9869	6.63	14.03	26
For Property Estimates							
Overall Property Crimes	.00009599	2526	0.202	.9894	5.36	21.62	16
All Property Domain Est.	.00005762	2536	0.599	.9359	14.39	15.97	304
Low Income	.00052980	2820	0.258	.9747	8.01	16.61	16
Medium Income	.00004521	2076	0.785	.9854	5.82	10.16	16
High Income	.00026055	2314	0.250	.9957	3.49	11.33	16
Hispanics	.00170683	2703	0.143	.9736	8.31	20.85	16
Urban	.00090965	2398	0.077	.9660	7.21	21.56	16
Suburban	.00046221	2122	0.649	.9859	5.43	16.95	16
Rural	.00675329	4955	-3.598	.9311	9.41	57.34	16
Blacks	.00105295	3356	-0.580	.9780	8.00	17.80	16
Reported to Police	.00006085	2692	-0.040	.9891	5.87	25.75	16