

Douglas Yeo, Harold Mantel, and Tzen-Ping Liu

Douglas Yeo, Statistics Canada, Tunney's Pasture, Ottawa, ON K1A 0T6, Canada; douglas.yeo@statcan.ca

Key Words: Longitudinal surveys, Public-use microdata files, Confidentiality

1. Introduction

The National Population Health Survey (NPHS) is a longitudinal household survey conducted by Statistics Canada, instituted in 1994 to survey the health of Canadians along with its determinants. Wave 1 of the survey provided a panel of approximately 17,000 respondents to be contacted every two years for up to twenty years; data are collected at four points in time during the survey year. Panel respondents were chosen by randomly selecting one person per surveyed household. The sample design of the NPHS is mostly based upon the Labour Force Survey (LFS) sampling methodology. The LFS design generally selects a stratified two-stage sample of dwellings selected within clusters (except in some rural, remote, and apartment strata) with six clusters per stratum. NPHS strata were created by grouping LFS strata, keeping some or all of the LFS selected clusters, but selecting fresh sample dwellings within those clusters. In Quebec the NPHS sample is taken from households that participated in the *Enquête Sociale et de Santé* (ESS), a health survey conducted by Santé Québec in 1992-93. Its design was similar to that of the LFS. (For more details on the NPHS design, see Tambay and Catlin, 1995.)

This paper discusses the new bootstrap variance estimation program put in place for wave 2, carried out in 1996-97 and released in 1998. Section 2 details the reasons behind the change, alternative methods examined, and the final choice. The bootstrap methodology is described in section 3, the results of empirical studies in section 4. Section 5 describes the implementation of the bootstrap weight program. Conclusions and areas for future work are provided in section 6.

2. Background

Given the variety of NPHS data products, the challenge is to produce useable data for quite disparate groups of users and to provide access to a variance estimation system. Two wave 2 public-use microdata files (PUMFs) were released, the first file containing general demographic and health information for all members of the sampled households, the second with detailed health information for the selected respondents only. (See National Population Health Survey: Public use microdata file documentation, 1996-97.)

Statistics Canada must provide sampling variances for all surveys, but for multistage surveys such as the NPHS, resampling or Taylor methods are required, and analysts require access to design information. However, stratum and cluster information cannot appear on PUMFs for confidentiality reasons, due to the detailed level of geography they represent. If cluster membership were available on the PUMF, users could potentially determine household membership and thus significantly increase the probability of identifying individuals. On the other hand, if users do not have access to this design information, they cannot calculate design-based variance estimates.

In wave 1 of the NPHS, approximate coefficient of variation (CV) tables were disseminated with the PUMFs, allowing users to look up a CV based on the size of the estimate calculated from the survey data. The CVs were derived using the variance formula for simple random sampling and incorporating a representative design effect, and are meant to apply to all characteristics. This design effect was determined by taking the 75th percentile of the design effects for a wide range of characteristics. Tables were produced only by province and by age group, and could only be used for simple estimates such as totals and proportions. Users doing more complex analyses, where CV tables are not appropriate, had to contract with Statistics Canada for jackknife variance estimates.

The challenge, then, is to preserve the confidentiality of respondents while giving users enough design information to calculate reliable variance estimators. For wave 2, several options were examined to better deal with this problem. Random renumbering of the stratum and cluster identifiers did not solve the problem. Since detailed cluster information would still be available on the file, a user could identify persons within the same cluster or even household, and could use this variable as a match key in an attempt to link PUMFs over time.

Another option considered was the dissemination of collapsed superstrata and superreplicates (Mayda *et al.*, 1996) so that users would be provided with enough information to calculate a reasonable variance estimate while preserving the confidentiality of the data. Advantages of this method were that the superstrata and superreplicates contained a larger group of individuals than the original strata and clusters (replicates), and that unbiased variance estimates could be produced. The user would not be able to identify the original clusters with certainty, and the superstrata and superreplicates would mimic the original design.

This method had several disadvantages. First, the exact variances corresponding to the original design could not be generated. Second, collapsing reduced the number of degrees of freedom, affecting the precision of variance estimates and reducing the number of covariates available for use in analyses—a major problem in small domains. Third, Valliant (1997) has shown that with collapsed strata, the balanced repeated replication (BRR) variance estimator may become inconsistent, while the asymptotic equivalence of the BRR and jackknife variance estimators shown in Rao and Wu (1988) indicates that the jackknife variance estimator may suffer from the same problem. Fourth, collapsing might be less effective in future waves of a longitudinal survey. Depending on how the collapsing is done and the number of dwellings in the resulting collapsed replicates, attrition, migration, and other events might affect the usability of future results. Finally, the provision of superstrata and superreplicates might facilitate malicious linking of PUMFs over time, if the collapsing were to be done deterministically.

The final option was to create bootstrap weights so users could calculate variance estimates closer to the exact estimates and for more complex analyses. The bootstrap has a number of advantages over the jackknife used in wave 1. First, with the bootstrap it is possible to give users a *reasonable* number of replicate weights, compared with 3,000 jackknife replicates. Disseminating replicate weights reduces the need for Statistics Canada to respond to requests from users for custom variance estimates. In fact, the original goal was to provide bootstrap weights with the wave 2 PUMFs. (Section 4 describes why this could not be achieved.)

Secondly, it was felt that the bootstrap program would be easier to run; the jackknife program was not feasible operationally given the very large cross-sectional supplemental samples that had been added in wave 2. As well, the jackknife had been programmed to be run province by province, which would not have taken into account interprovincial migration between survey years. Thirdly, we still wanted to produce the coefficient of variance (CV) tables for use in simple analyses, but we wanted to replace the process used to create the tables for wave 1, which was not easily extendible to longitudinal data, with bootstrap design effects and CVs. Finally, the bootstrap is recommended over the jackknife to estimate the variance of nonsmooth functions such as quantiles and the low income cut-off (LICO) point (Kovačević and Yung, 1997). Because of these advantages, it was decided to introduce the use of bootstrap weights in wave 2.

3. Bootstrap Methodology

This section describes the “standard” bootstrap used in the NPHS, as well as some of the problems encountered in its implementation.

3.1 The Standard Bootstrap

The bootstrap resampling method for the iid case has been extensively studied. (See Efron, 1982.) It was extended by Rao and Wu (1988) to stratified multistage designs and again by Rao, Wu and Yue (1992) to include nonsmooth statistics. This latter design was implemented in the NPHS. It assumes L design strata, where stratum h has N_h clusters and $n_h \geq 2$ sampled clusters. Subsampling within selected clusters (*e.g.*, selecting one person in a dwelling) is performed according to some probability sampling design with unbiased estimation of cluster totals, Y_{hi} , with $h=1, \dots, L$ and $i=1, \dots, n_h$.

An estimator of the total Y , for example, is obtained using the variable of interest, y_{hik} , and design weights, w_{hik} , associated with the k^{th} sample element in sample cluster i belonging to stratum h by

$$\hat{Y} = \sum_{(hik) \in s} w_{hik} y_{hik},$$

where s denotes the sampled elements. The NPHS design weights are then poststratified to ensure consistency with known demographic totals. Given that each element in the population belongs to a poststratum that can cut across the design strata, the total number of elements in the c -th poststratum is ${}_cM$, a known quantity. Letting ${}_c w_{hik}$ represent the poststratified or final weight defined by

$${}_c w_{hik} = \frac{{}_c M}{{}_c \hat{M}} w_{hik},$$

${}_c \hat{M} = \sum_{(hik) \in s} w_{hik} {}_c \delta_{hik}$ and ${}_c \delta_{hik}$ the poststratum indicator variable, the poststratified estimator is defined as

$$\hat{Y}_{ps} = \sum_c \sum_{(hik) \in s} {}_c w_{hik} y_{hik} {}_c \delta_{hik}.$$

The standard bootstrap variance estimator for $\hat{\theta} = g(\hat{Y})$, with \hat{Y} given above and g a known function, is calculated as follows:

- (i) Independently for each stratum, select a simple random sample with replacement of m_h clusters from the n_h sampled clusters. With m_{hi}^* defined as the number of times the (hi) -th cluster is selected ($\sum_i m_{hi}^* = m_h$), define the bootstrap weights as

$$w_{hik}^* = \left[1 - \left(\frac{m_h}{n_h - 1} \right)^{1/2} + \left(\frac{m_h}{n_h - 1} \right)^{1/2} \frac{n_h}{m_h} m_{hi}^* \right] w_{hik}.$$

For the NPHS, m_h was set to $n_h - 1$, a commonly used value, ensuring that the bootstrap weights, w_{hik}^* , would be nonnegative and reducing the equation to

$$w_{hik}^* = \frac{n_h}{n_h - 1} m_{hi}^* w_{hik}.$$

- (ii) These weights are then poststratified in the same way as the design weights to obtain the final weights:

$${}_c w_{hik}^* = \frac{{}_c \hat{M}}{{}_c \hat{M}^*} w_{hik}^*, \quad \text{where } {}_c \hat{M}^* = \sum_{(hik) \in s} w_{hik}^* {}_c \delta_{hik}.$$

To calculate the bootstrap estimator of θ , $\hat{\theta}^*$, replace the survey weights, ${}_c w_{hik}$, with the final bootstrap weights, ${}_c w_{hik}^*$, in the formula for $\hat{\theta}$.

- (iii) Independently replicate these two steps a large number of times, B , and calculate the corresponding estimates, $\hat{\theta}_{(1)}^*, \dots, \hat{\theta}_{(B)}^*$.

The bootstrap variance estimator for $\hat{\theta}$ is then given by

$$v_B(\hat{\theta}) = \frac{1}{B} \sum_b \left(\hat{\theta}_{(b)}^* - \hat{\theta}_{(\cdot)}^* \right)^2, \quad \text{where } \hat{\theta}_{(\cdot)}^* = (1/B) \sum_b \hat{\theta}_{(b)}^*.$$

3.2 Missing Clusters

This implementation of the bootstrap takes into account only the final weighting step, poststratification. Ideally, all adjustments to the design weight should be included in the calculation of the bootstrap weights. For example, nonrespondents and respondents should be included in the initial file from which bootstrap subsampling is done. Only then should nonrespondents be dropped.

Since only survey respondents were included in the implementation of the bootstrap for the NPHS, the problem of missing clusters arose. If there are originally n_h clusters in a particular stratum, a simple random sample with replacement of n_h-1 clusters should be taken; the bootstrap weight expansion factor should be $n_h/(n_h-1)$. However, because of nonresponse, some of the smaller clusters had no respondents among their sampled households; *i.e.*, in stratum h , p_h clusters were missing. Given the decision to include only respondents in the bootstrap subsampling for the NPHS, there was no choice but to take a sample of n_h-p_h-1 of clusters from the remaining n_h-p_h clusters with responding households. Similarly, the bootstrap weight expansion factor in this case was $(n_h-p_h) / (n_h-p_h-1)$. This happened rarely enough that the impact on estimates was negligible.

Limited empirical study has suggested that ignoring nonresponse adjustments in the bootstrap algorithm does not have a large impact on the final variance estimates: they are negatively biased, but this bias is small. In any event, we plan to incorporate this weighting adjustment in the bootstrap algorithm in wave 3, properly accounting for entirely missing clusters and nonresponse within clusters, two potential sources of variation.

4. Empirical Study

Results of studies on the optimal values for the bootstrap sampling parameters, the type of bootstrap to use, and confidentiality issues are presented here.

4.1 Simulations

Many simulations were run to determine B , the number of bootstrap samples. There is a tradeoff here between precision and operational efficiency: B must be large enough to produce reliable variance estimates, but

not so large as to make the bootstrap weight files unwieldy. Using a subset of wave 1 data, variance estimates stabilised at around 200 bootstrap samples for estimates of totals and ratios and at around 400 for regressions. An empirical comparison of the bootstrap with 500 weights and the jackknife suggested that the bootstrap CVs of almost 95% of simple estimates and 75% of regression coefficient estimates were within $\pm 1\%$ of the jackknife CV. (See Appendix A for more comparisons.)

4.2 Mean Bootstrap vs. Standard Bootstrap

The dissemination of these standard bootstrap weights with the PUMF puts respondents at risk of disclosure, since records in a given cluster will all have a bootstrap weight of zero if that cluster is not selected in a particular bootstrap sample. (For these records, $m_{hi}^* = 0$, resulting in a bootstrap weight of zero; multiplicative weight adjustments such as poststratification will not change the situation.) This problem arises because entire clusters are being resampled rather than single units as in the classical bootstrap. In a single bootstrap replicate, these zero weights do not pose a disclosure risk. However, cluster membership can be identified by examining the pattern of zero weights over *all* B bootstrap replicates. Even if cluster location is not on the PUMF, the use of other PUMF variables may allow users to deduce this, breaching confidentiality.

We did a number of simulation studies using wave 1 data and the *mean bootstrap* described in Yung (1997), to attempt to overcome this problem. In the mean bootstrap, R (perhaps 20) bootstrap samples are produced and the m_{hi}^* 's are averaged over the R samples. As long as each cluster appears in at least one of the R bootstrap samples, the averages will all be non-zero. Mean bootstrap weights and estimates are created as follows:

- (i) Independently for each stratum, select a simple random sample of n_h-1 clusters with replacement from the n_h sample clusters.
- (ii) Repeat step (i) R times. Let $m_{hi}^*(r)$ be the number of times the (hi) -th cluster is selected in the r -th bootstrap sample and $m_{hi(\cdot)}^* = (1/R) \sum_r m_{hi}^*(r)$ be the average number of times the (hi) -th cluster is selected over the R bootstrap samples. Define the mean bootstrap weight as

$$w_{hik(\cdot)}^* = \frac{n_h}{n_h - 1} m_{hi(\cdot)}^* w_{hik}.$$

- (iii) Poststratify the mean bootstrap weights $w_{hik(\cdot)}^*$ to obtain ${}_c w_{hik(\cdot)}^*$, the bootstrap final weights. Calculate $\hat{\theta}^*$ using the bootstrap final weights.
- (iv) Independently replicate steps (i) to (iii) a large number of times, B , and calculate the corresponding estimates, $\hat{\theta}_{(1)}^*, \dots, \hat{\theta}_{(B)}^*$.

The mean bootstrap variance estimator is then given as

$$v_{MB}(\hat{\theta}) = \frac{R}{B} \sum_b \left(\tilde{\theta}_{(b)}^* - \tilde{\theta}_{(.)}^* \right)^2, \text{ where } \tilde{\theta}_{(.)}^* = (1/B) \sum_b \tilde{\theta}_{(b)}^*.$$

Note that this procedure does not eliminate the possibility of a zero weight. R should be chosen to be large enough that the chance of $m_{hi(r)}^* = 0$ for all $r = 1, \dots, R$ is very small, but not so large that drawing $R \times B$ bootstrap samples becomes computationally difficult. In the NPHS implementation, we used a two-step procedure. We chose $R = 20$ to produce the initial set of bootstrap replicates; if any replicates contained zero weights, we rejected them and produced a new set of replicates. The alternative would have been to leave R variable and continue steps (i) and (ii) until no zero weights remained, but the procedure used was straightforward, quickly obtained the desired sets of bootstrap replicates, and simplified the variance formula.

4.3 Confidentiality Concerns

The mean bootstrap reduces confidentiality concerns somewhat. However, records in the same cluster for a particular bootstrap sample will still share the same weight, albeit not zero. However, when the method was originally developed, it was hoped that sufficient noise might be added by final poststratification of the mean bootstrap weights to alleviate this problem.

The mean bootstrap weights were analysed for possible inclusion on the NPHS wave 2 PUMFs, to give users the opportunity to calculate their own variance estimates. Unfortunately, cluster analysis showed that a malicious user would be able to recreate, with 100% accuracy, stratum and cluster membership from patterns in the bootstrap weights. This would not identify the location of these strata and clusters, but other information on the PUMF might allow this user to deduce location for at least some strata and clusters. For this reason, priority was shifted to providing variance estimation capability in the remote access program. (See section 5.4.)

We hope to test the mean bootstrap weights further for possible inclusion with future PUMFs. The disclosure risk seems reasonably low; we shall try to determine if this level of risk is acceptable. Work is also being done on the mean bootstrap algorithm itself to try to reduce or eliminate this risk, by collapsing strata before resampling and through the addition of noise to the final bootstrap weights.

5. Implementation

5.1 Master Files

The main application of the bootstrap weights has been for use internally with the master files to produce variance estimates. For each survey weight on the master

file (four cross-sectional and three longitudinal weights), a set of 500 bootstrap weights was created and stored in a central location. SAS macros were distributed to analysts that calculate variance estimates for such statistics as totals, ratios, and linear and logistic regression parameters.

We also created sets of 500 bootstrap weights for each of the survey weights on the wave 1 master files: the general weight for all household members, the health weight for the selected person, and the weight for a special Health Canada supplement that had been done. These weights were used in the simulation studies noted above and have been used for some analyses. The long-term goal is to be able to handle variance estimation for analyses on all waves of the survey using the bootstrap weights.

For external users, approximate CV tables were again disseminated with the PUMFs, derived this time from bootstrap design effects. Using bootstrap weights greatly reduced the time and complexity of this process.

5.2 Share Files

A second set of survey data files, the "share files", is sent to Health Canada and the provincial health ministries after each wave, containing only those respondents who agree to share their data with these departments. As subsets of the master files, the share files must be reweighted and poststratified to demographic totals. One reason for switching to bootstrap weights was to provide external users such as these with a means of calculating their own variance estimates. With wave 2 share weights calculated, bootstrap share weights could also be created. These weights, along with SAS and SPSS variance estimation programs, were provided to these departments.

5.3 Supplemental Surveys

There is also the opportunity at each wave for survey supplements. In wave 2, Health Canada funded supplemental questions and sample in order to interview all 1996 NPHS respondents with asthma. The supplement surveyed severity, environmental triggers, and drug use associated with their asthma, and was conducted several months after each NPHS collection quarter. Master and share files were created, and bootstrap weights were calculated for the two weights (general and health) on each file. This implementation of the bootstrap program was more complete than that of the core survey, in that the asthma bootstrap weights were adjusted for second-phase nonresponse. As well, because respondents with asthma are a relatively small domain, we were able to increase the number of bootstrap replicates to 1,000. The share file bootstrap weights and SAS programs to calculate variance estimates were made available to the client.

5.4 Remote Access to the NPHS Master Data Files

For reasons of confidentiality, the NPHS PUMFs contain significantly less information than the master files. Many of the original survey variables, available on the master files, are not placed on the PUMF, while others are modified (*e.g.*, regrouped, top-coded) to minimise the risk of disclosure. As well, a longitudinal PUMF has not been released for wave 2. To overcome these data gaps, a remote access program was instituted to allow users to run analyses using the master microdata file, but in such a way that the data remain confidential. Analysts write their own analysis programs and send them to Statistics Canada where they are run with the master microdata file as input. Output is then checked for confidentiality before being returned to the analyst. It is the analysts' responsibility to ensure that their analysis programs run properly. To this end we provide them with dummy files that they can use for development and testing of their programs. The dummy files have the same format as the master files, but contain fake data and only a subsample of records.

The data in the dummy files had to be consistent with the skip patterns induced by the questionnaire and approximately preserve the marginal distributions of variables and the relationships between closely related variables from the master microdata files. This was particularly important in the case of the longitudinal dummy file, since no longitudinal PUMF is yet available. To meet these objectives, the dummy files were created by randomly swapping blocks of variables among respondents within classes. Following the skips in the questionnaire, these classes were generally based on age/sex categories by province with a minimum class size to ensure confidentiality after data swapping. The blocks were formed by dividing variables into fifty groups of highly dependent variables, which were then further collapsed based on questionnaire-induced dependencies between them. As a final step to help ensure realistic estimates from the dummy files, the weights were poststratified within the classes to master file totals.

For variance estimation, 500 sets of bootstrap weights were produced for each dummy file, allowing analysts to write their own variance estimation programs to be run at Statistics Canada using the master file and its associated bootstrap weights. Sample variance estimation programs using SAS and SPSS, adaptable to particular analyses, are also provided to users. This implementation exemplifies the power of the bootstrap program: its use can be extended easily to meet new survey demands. In this case, the dummy bootstrap weights were easily calculated using the dummy survey files and the original bootstrap SAS programs, resulting in a remote process much more efficient than the former case-by-case response to variance estimation.

6. Conclusions and Future Work

The dissemination of bootstrap weights allows users of NPHS data to now calculate their own correct design-based variance estimates for simple statistics such as proportions and totals as well as for more complex statistics such as regression coefficients. Although some technical knowledge is required, the calculation of these estimates is generally straightforward. The use of bootstrap weights should also prove fruitful for other surveys with complex sample designs, whether or not they release PUMFs.

We are working to improve the user interface of the variance programs and to include nonresponse and other weighting adjustments in the bootstrap algorithm for wave 3. These additions will ensure that the bootstrap captures more of the total variance and will eliminate the problems related to missing clusters noted in earlier. The mean bootstrap algorithm is also being revisited, to try to solve the confidentiality problem noted and thus to be able to disseminate bootstrap weights with the PUMFs.

REFERENCES

- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- Kovačević, M. and Yung, W. (1997). Variance estimation for measures of income inequality and polarization—An Empirical Study. *Survey Methodology*, 23, 41-52.
- Mayda, J.E., Mohl, C., and Tambay, J.-L. (1996). Variance estimation and confidentiality: They are related. *Proceedings of the Survey Methods Section, Statistical Society of Canada*, 135-141.
- National Population Health Survey 1996-97. Public use microdata file documentation, Statistics Canada Publication Number 82M0009XCB.
- Rao, J.N.K., and Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.
- Rao, J.N.K., Wu, C.F.J., and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 209-217.
- Tambay, J.-L. and Catlin, G. (1995). Sample design of the National Population Health Survey. *Health Reports (Statistics Canada Catalogue No. 82-003)*, Volume 7 (1), 33-42.
- Valliant, R. (1997). Limitations of balanced half-sampling. *Journal of Official Statistics*, 12, 225-240.
- Yung, W. (1997). Variance estimation for public use microdata files. *Proceedings of the Statistics Canada Symposium*, 91-95.

APPENDIX A: Bootstrap vs. Jackknife CVs

Using the survey weight from NPHS health PUMF for wave 1 (1994/95), 250 bootstrap files were created, 50 each with $B=100, 200, 300, 400,$ and 500 . These sets of bootstrap files were then used to calculate CVs for a large number of estimates, which were compared with the CVs from the full jackknife used in wave 1. Tables 1 and 2 show the results over 90 estimates of totals and ratios and over 25 regression estimates respectively. The average number of CVs over the B bootstrap samples within $x\%$ of the jackknife CV is given in the first line of each cell. The standard deviation of this average is also given. The second line (in parentheses) is the percentage over the 90 or 25 estimates.

For example, in Table 1, averaged over 100 bootstrap weights, 70.5 ± 3.8 of the 90 estimates had bootstrap CVs within 1% of the jackknife CV. In percentage terms, this is $78.4\% \pm 4.2\%$. If 500 bootstrap weights are used, this increases to 84.7 ± 2.3 or $94.3\% \pm 2.6\%$. For regressions, shown in Table 2, with only 100 bootstrap weights, $55.5\% \pm 7.6\%$ of the regression estimates had bootstrap CVs within 1% of the jackknife CV, which increased to $75.0\% \pm 7.0\%$ using 500 bootstrap weights. As expected, as the number of bootstrap weights is increased, not only does the average percentage of bootstrap CVs that are close to the full jackknife CV increase, but the standard deviation around that average percentage decreases. That is, the bootstrap CVs become more stable as B increases.

Table 1: Number and (Percentage) of Bootstrap CVs Within $x\%$ of the Jackknife CV Estimates of Totals and Ratios (Number of Estimates=90)

$CV_B - CV_{Jackknife}$	Average \pm S.D. for Bootstrap				
	$B=100$	$B=200$	$B=300$	$B=400$	$B=500$
$\pm 1\%$	70.5 ± 3.8 ($78.4 \pm 4.2\%$)	77.6 ± 3.5 ($86.2 \pm 3.9\%$)	81.5 ± 2.7 ($90.6 \pm 3.0\%$)	83.9 ± 2.4 ($93.3 \pm 2.7\%$)	84.7 ± 2.3 ($94.3 \pm 2.6\%$)
$\pm 2\%$	84.1 ± 2.8 ($93.4 \pm 3.2\%$)	87.6 ± 1.5 ($97.4 \pm 1.7\%$)	88.9 ± 1.1 ($98.7 \pm 1.2\%$)	89.1 ± 1.0 ($99.0 \pm 1.1\%$)	89.2 ± 0.9 ($99.1 \pm 1.0\%$)
$\pm 4\%$	89.5 ± 0.7 ($99.5 \pm 0.7\%$)	89.9 ± 0.3 ($99.9 \pm 0.4\%$)	90.0 ± 0.1 ($100 \pm 0.2\%$)	90.0 ± 0.1 ($100 \pm 0.2\%$)	90.0 ± 0.0 ($100 \pm 0.0\%$)
$> 4\% $	0.5 ± 0.7 ($0.5 \pm 0.7\%$)	0.1 ± 0.3 ($0.1 \pm 0.4\%$)	0.0 ± 0.1 ($0.0 \pm 0.2\%$)	0.0 ± 0.1 ($0.0 \pm 0.2\%$)	0.0 ± 0.0 ($0.0 \pm 0.0\%$)

Table 2: Number and (Percentage) of Bootstrap CVs Within $x\%$ of the Jackknife CV Regressions (Number of Estimates=25)

$CV_B - CV_{Jackknife}$	Average \pm S.D. for Bootstrap				
	$B=100$	$B=200$	$B=300$	$B=400$	$B=500$
$\pm 1\%$	13.9 ± 1.9 ($55.5 \pm 7.6\%$)	16.4 ± 1.6 ($65.1 \pm 6.6\%$)	16.7 ± 1.6 ($66.8 \pm 6.5\%$)	18.5 ± 1.9 ($73.9 \pm 7.4\%$)	18.7 ± 1.8 ($75.0 \pm 7.0\%$)
$\pm 2\%$	18.1 ± 1.6 ($72.5 \pm 6.5\%$)	20.1 ± 1.4 ($80.6 \pm 5.7\%$)	20.6 ± 1.4 ($82.6 \pm 5.7\%$)	21.6 ± 1.3 ($86.5 \pm 5.4\%$)	22.1 ± 1.1 ($88.3 \pm 4.3\%$)
$\pm 4\%$	21.7 ± 1.2 ($86.7 \pm 4.9\%$)	23.0 ± 0.9 ($92.1 \pm 3.8\%$)	23.5 ± 0.7 ($93.8 \pm 2.9\%$)	23.5 ± 0.9 ($94.0 \pm 3.5\%$)	23.6 ± 0.9 ($94.5 \pm 3.4\%$)
$> 4\% $	3.3 ± 1.2 ($13.3 \pm 4.9\%$)	2.0 ± 0.9 ($7.9 \pm 3.8\%$)	1.5 ± 0.7 ($6.2 \pm 2.9\%$)	1.5 ± 0.9 ($6.0 \pm 3.5\%$)	1.4 ± 0.9 ($5.5 \pm 3.4\%$)