Unbiased Variance Estimator of the Randomised Response Techniques for Population Mean

Ondřej Vozár[1] | *Prague University of Economics and Business, Prague, Czech Republic*

Received 6.9.2022 (revision received 21.11.2022), **Accepted** (reviewed) 7.12.2022, **Published** 17.3.2023

Abstract

Antoch, Mola and Vozár (2022) proposed recently new randomized response technique for population mean or total of a quantitative variable. The aim of the paper is to solve the open problem to derive unbiased variance estimator of these procedures. In their proposal, unlike other randomized response techniques for population mean or total the randomized response is not a linear function of a sensitive variable. However, standard techniques to derive variance estimators in this setting are based on this assumption. That is why an interviewer needs also to know values pseudorandom numbers (i.e., results of individual randomization experiments). Respondents might perceive this relaxation of privacy protection negative. The performance of the approximate two-sided confidence intervals of distributions with different shape including their coverage is assessed by a simulation study for simple random sampling without replacement.

Keywords	***DOI***	***JEL code***
Unbiased variance estimator, randomized response techniques, survey sampling, Horvitz-Thompson estimator, simple random sampling without replacement, population mean	*https://doi.org/10.54694/stat.2022.38*	*C83*

INTRODUCTION

The concern of respondents of statistical surveys on their privacy on the one hand and growing interested of survey sponsors on sensitive issues (like drug consumption, behavior deviation from society or legal norms like corruption etc.) on the other hand make methodological challenges for survey statisticians. It is natural that participants (respondents) of such a survey tend to refuse to participate (non-response) or to provide strongly biased answers (severe measurement error). As mentioned by Särndal et al. (1992, p. 547), if we ask sensitive or pertinent questions in a survey, conscious reporting of false values would often occur. These issues can not be resolved by standard techniques like model-based regression imputation or reweighting (see also Särndal et al., 1992, p. 547). For detailed discussion of reweighting and imputation see monograph Särndal and Lundström (2005).

[1] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: vozo01@vse.cz. Also the Czech Statistical Office, Unit for Mathematical and Statistical Methods and Statistical Quality, Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: ondrej.vozar@czso.cz.

Randomized response techniques (RRTs) are designed to resolve both the high nonresponse and bias due to fabricated answers. Instead of direct questioning, each respondent makes independently a random experiment (its result is unknown to an interviewer) and then answers depending on its result. Eriksson (1971) proposed the first RRT for mean of a quantitative variable. A respondent provides value of a sensitive variable multiplied by a number on the card drawn from a deck of cards. This problem was also studied by T. Dalenius and his colleagues (Dalenius and Vitale, 1979). All the methods are based on providing linear transformation of a sensitive variable instead of its true value (for example Chaudhuri, 1987). The state-of-art of RRT is presented in recent monographs by Chaudhuri (2017) and Chaudhuri et al. (2016).

Notions of survey sampling theory and randomized response techniques for mean are presented in Section 1. Unbiased variance estimator of two mean estimators using RRT by Antoch et al. (2022) are derived in Section 2. Performance of the variance estimators and coverage of approximate normal confidence intervals is assessed by simulation study in Section 3. The last section consists of conclusions of this paper.

1 NOTIONS OF RRTS FOR MEAN OF QUANTITATIVE VARIABLES

Let $U = \{1, \dots , N\}$ be a finite population of N identifiable units unambiguously identified their labels. The purpose of the survey is to estimate the population total or mean of a sensitive variable Y denoted as $\bar{t}_Y = \frac{\sum_{i\in U} Y_i}{N}$. Therefore, we select a random sample s with probability $p(s)$ describing a sampling plan with a fixed sample size n. Let us define π_i as the probability of inclusion of the i^{th} element of the population in the sample s, i.e., $\pi_i = \sum_{s\ni i} p(s)$. The $\pi_{i,j}$ is defined as the probability of inclusion of the i^{th} and j^{th} element in the sample s, i.e., $\pi_{i,j} = \sum_{s\ni i\cap s\ni j} p(s)$ (note, that $\pi_{i,i} = \pi_i$). In depth overview of survey sampling is given by monograph of Särndal et al. (1992), mathematically rigorous treatment by Tillé (2020).

If the variable of interest is difficult to survey by the direct questioning due to its sensitivity, the interviewers collect randomized response Z correlated to Y. Randomization of the responses must be done for each unit in the sample independently. The randomization procedure should not also depend on the sampling plan $p(s)$.

Then, the survey has two phases. In the first phase a sample s is selected from population U. Then in the sencond phase, given sample s, responses Z_i are obtained using the chosen RRT. The corresponding probability distributions are denoted as $p(s)$ and $q(r|s)$. The notions of expected value, unbiasedness, and variance are derived by a two-fold averaging process. The first averaging is done over all possible samples s that can be selected using the sampling plan $p(s)$. The second averaging is done over all possible randomized response sets r that can be realized given sample s under the randomized response distribution $q(r|s)$. We use the notation of Arnab (1994), Chaudhuri (1992) and denote the expectation operators with respect to these distributions by E_p and E_q.

In survey using direct questioning, if $\pi_i > 0$, $\forall\ i \in U$ and known population size N, the population mean $\bar{t}_Y$ is often estimated by linear unbiased Horvitz-Thompson estimator:

$$\bar{t}_s^{HT} = \frac{1}{N} \sum_{i\in s} \frac{Y_i}{\pi_i} \tag{1}$$

with variance:

$$Var(\bar{t}_s^{HT}) = \frac{1}{N^2} \sum_{i\in U} \sum_{j\in U} (\pi_{i,j} - \pi_i\pi_j) \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j}. \tag{2}$$

Moreover, if $\pi_{i,j} > 0$, $\forall\, i, j \in U$, then the variance can be unbiasedly estimated as:

$$\hat{Var}(\bar{t}_s^{HT}) = \frac{1}{N^2} \sum_{i\in s} \sum_{j\in s} \frac{(\pi_{i,j} - \pi_i\pi_j)}{\pi_{i,j}} \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j} . \tag{3}$$

In a survey using RRT instead of direct questioning, the values of Y_i from the sample s are unknown to the interviewers. Instead, they collect randomized responses Z_i correlated with Y_i. Randomized responses Z_i are further transformed into another variables R_i. The transformed variables are more fit to construct the unbiased population mean Horvitz-Thompson's type estimator. It is constructed as:

$$\bar{t}_s^{HT,R} = \frac{1}{N} \sum_{i\in\in s} \frac{R_i}{\pi_i} . \tag{4}$$

We assume that transformed randomized responses R_i follow model of Chaudhuri (1992):

$$E_q(R_i) = Y_i,\ Var_q(R_i) = \phi_i,\ \forall i \in U,\ Cov_q(R_i, R_j) = 0_i,\ \forall i, j \in U : i \neq j . \tag{5}$$

The variance function ϕ_i of a transformed randomized response R_i is a function of Y_i.

The estimator $\bar{t}_s^{HT,R}$ of the population mean is defined as conditionally unbiased if it holds $E_q(\bar{t}_s^{HT,R}|s) = \bar{t}_s^{HT}$. It means, that conditional mean of $\bar{t}_s^{HT,R}$ given the sample s equals to the Horvitz-Thompson estimator $\bar{t}_s^{HT}$, if only direct questioning would take place. If the estimator $\bar{t}_s^{HT,R}$ is conditionally unbiased and estimator $\bar{t}_s^{HT}$ is unbiased, then estimator $\bar{t}_s^{HT,R}$ is also unbiased, becuase $E(\bar{t}_s^{HT,R}) = E_p(E_q(\bar{t}_s^{HT,R}|s) = E_p(\bar{t}_s^{HT}) = \bar{t}_Y$. Then the variance can be expressed using the standard probability formula as:

$$\begin{aligned} Var(\bar{t}_s^{HT,R}) &= E_p(Var_q(\bar{t}_s^{HT,R}|s)) + Var_p(E_q(\bar{t}_s^{HT,R}|s)) = E_p(Var_q(\bar{t}_s^{HT,R}|s)) + Var_p(\bar{t}_s^{HT,R}) \\ &= \sum_{i\in U} \frac{\phi_i}{\pi_i} + Var_p(\bar{t}_s^{HT,R}) . \end{aligned} \tag{6}$$

The variance given by (6) is decomposed into two terms. The second term is the variance of the estimator if RRT was not used. The first term represents the increase of the variance caused by the randomization.

2 MAIN RESULTS

Two estimators of mean using RRT by Antoch et al. (2022) are briefly presented. Unbiased plug-in variance estimators using ideas of Chaudhuri (1992) and Arnab (1994) are then derived.

2.1 Two mean estimators using RRT by Antoch et al. (2022)

In the majority of RRT, a respondent masks true value by some algebraic operations (multiplication, respectively adding number for a randomly drawn card). To avoid guessing true value of sensitive variable, Antoch et al. (2002) proposed much safer technique. They assume, that the surveyed variable Y is non-negative and bounded from above: $0\leq Y\leq M$; the upper bound M of the variable Y is also known. Each respondent independently on the others generates pseudorandom number U_i from the uniform distribution on interval $(0,M)$. To protect privacy, the respondent then answers a question: "Is the value of Y greater than U"?

The response $Z_{i,(0,M)}$ of the *ith* respondent follows the alternative distribution with the parameter $\frac{Y_i}{M}$. It holds $E(Z_{i,(0,M)}) = \frac{Y_i}{M}$ and $Var(Z_{i,(0,M)}) = \frac{Y_i}{M}(1 - \frac{Y_i}{M})$. Then we transform $Z_{i,(0,M)}$ to $R_{i,(0,M)} = MZ_{i,(0,M)}$, for which it holds:

$$E(R_{i,(0,M)} = Y_i,\ \phi_{i,(0,M)} = Var(R_{i,(0,M)} = Y_i(M - Y_i)\ . \tag{7}$$

The unbiased mean estimator takes form:

$$\bar{t}^{\,HT,R}_{(0,M)} = \frac{1}{N}\sum_{i\in s}\frac{R_{i,(0,M)}}{\pi_i}\ . \tag{8}$$

Antoch et al. (2022) also proposed an estimator on the assumption, that interviewer also knows a value of pseudorandom number U_i (i.e., the question to which a respondent answer). An interviewer transforms the answer of i^{th} respondent as:

$$Z_{i,\alpha,(0,M)} = \begin{cases} 1 - \alpha + 2\alpha\,\frac{U_i}{M}, \text{ if } U_i < Y_i \\ -\alpha + 2\alpha\,\frac{U_i}{M}, \text{ otherwise} \end{cases} \quad 0 \le \alpha < 1,\ i \in s\ , \tag{9}$$

where α is a tuning parameter. It is set a priori by an interviewer and its value is unknown to respondents. It holds $E(Z_{i,\alpha,(0,M)}) = \frac{Y_i}{M}$ and $Var(Z_{i,\alpha,(0,M)}) = \frac{1-2\alpha}{M^2}\,Y_i(M - Y_i) + \frac{\alpha^2}{3}$. We further transform $Z_{i,\alpha,(0,M)}$ to $R_{i,\alpha,(0,M)} = MZ_{i,\alpha,(0,M)}$ with:

$$E(R_{i,\alpha,(0,M)} = Y_i,\ \phi_{i,\alpha,(0,M)} = Var(R_{i,\alpha,(0,M)} = (1 - 2\alpha)\ Y_i(M - Y_i) + \frac{\alpha^2M^2}{3}\ . \tag{10}$$

The unbiased mean estimator takes form:

$$\bar{t}^{\,HT,R}_{\alpha,(0,M)} = \frac{1}{N}\sum_{i\in s}\frac{R_{i,\alpha,(0,M)}}{\pi_i}\ . \tag{11}$$

2.2 Unbiased variance estimator of two mean estimators by Antoch et al. (2022)

Let us assume, it holds $\pi_{i,j} > 0, \forall\ i, j \in U$, then an unbiased variance estimator of Horvitz-Thompson's type estimator for the population mean takes form:

$$V\hat{a}r(\bar{t}^{\,HT,R}_s) = \frac{1}{N^2}\sum_{i\in s}\frac{\hat{\phi}_i}{\pi_i^2} + \frac{1}{N^2}\sum_{i\in s}\sum_{j\in s}\frac{(\pi_{i,j} - \pi_i\pi_j)}{\pi_{i,j}}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}, \text{ where } E_q(\hat{\phi}_i) = \phi_i, \forall\ i, j \in U\ . \tag{12}$$

To derive variance estimator by standard methods, the second moment of the transformed randomized response R_i must be a function of Y_i^2. Unfortunately, it does not hold for original proposal (8) by Antoch et al. (2022) (without knowledge of pseudorandom numbers). The reason is, that it holds $(R_{i,(0,M)})^n = M^n Z_{i,(0,M)}$, which implies $E(R_{i,(0,M)})^n = M^{(n-1)} Y_i$. An interviewer needs to know values of U_i to estimate variance of $\bar{t}^{HT,R}_{(0,M)}$, because it holds:

$$E(2R_{i,(0,M)}U_i) = 2\int_0^{Y_i} U_i dU_i = Y_i^2 .$$

Then, using values of U_i, Y_i^2 can be unbiasedly estimated as $2R_{i,(0,M)}U_i$. Variance of $R_{i,(0,M)}$ can be unbiasedly estimated as:

$$\hat{\phi}_{i,(0,M)} = R_{i,(0,M)}(M - 2U_i) . \tag{13}$$

A drawback of that proposal is that if $Y_i > \frac{M}{2}$, the value of $\hat{\phi}_{i,(0,M)}$ can be negative, even if it estimates non-negative quantity. In the same manner, we can construct unbiased variance estimator as:

$$\begin{aligned} V\hat{a}r(\bar{t}^{HT,R}_{(0,M)}) &= \frac{1}{N^2}\sum_{i\in s}\frac{R_{i,(0,M)}(M-2U_i)}{\pi_i^2} + \frac{1}{N^2}\sum_{i\in s}\sum_{j\in s}\frac{(\pi_{i,j}-\pi_i\pi_j)}{\pi_{i,j}}\frac{R_{i,(0,M)}}{\pi_i}\frac{R_{j,(0,M)}}{\pi_j} \\ &+ \frac{1}{N^2}\sum_{i\in s}(1-\pi_i)\frac{R_{i,(0,M)}(2U_i - R_{i,(0,M)})}{\pi_i^2} . \end{aligned} \tag{14}$$

For the second proposal of Antoch et al. (2022) using knowledge of pseudorandom numbers (11) the variance estimation technique of Chaudhuri (1992), because for $\alpha \neq 0$ variable Y_i^2 is one of term of the second moment of $R_{i,\alpha,(0,M)}$, namely:

$$E(R^2_{i,\alpha,(0,M)}) = 2\alpha Y_i^2 + (1-2\alpha)\, MY_i + \frac{\alpha^2 M^2}{3} . \tag{15}$$

Using $R^2_{i,\alpha,(0,M)}$and $R_{i,\alpha,(0,M)}$ as unbiased estimator of Y_i, the unbiased variance estimator is:

$$\hat{\phi}_{i,\alpha,(0,M)} = \frac{-(1-2\alpha)\, R^2_{i,\alpha,(0,M)} + (1-2\alpha)\, MR_{i,\alpha,(0,M)} + \frac{\alpha^2M^2}{3}}{2\alpha} . \tag{16}$$

The unbiased variance estimator is then:

$$\begin{aligned} V\hat{a}r(\bar{t}^{HT,R}_{(0,M)}) &= \frac{1}{N^2}\sum_{i\in s}\frac{\hat{\phi}_{i,\alpha,(0,M)}}{\pi_i^2} + \frac{1}{N^2}\sum_{i\in s}\sum_{j\in s}\frac{(\pi_{i,j}-\pi_i\pi_j)}{\pi_{i,j}}\frac{R_{i,\alpha,(0,M)}}{\pi_i}\frac{R_{j,\alpha,(0,M)}}{\pi_j} \\ &+ \frac{1}{N^2}\sum_{i\in s}\frac{(1-\pi_i)}{\pi_i^2}\left(\frac{(1-2\alpha)(R^2_{i,\alpha,(0,M)} - MR_{i,\alpha,(0,M)}) - \frac{\alpha^2M^2}{3}}{2\alpha}\right). \end{aligned} \tag{17}$$

3 SIMULATION STUDY

In the simulation study, we study coverage of approximate confidence intervals:

$$\bar{t}^{HT,R}_{(0,M)} = \pm\, t_{1-\alpha/2}\sqrt{V\hat{a}r(\bar{t}^{HT,R}_{(0,M)})}\,, \qquad (18)$$

where $t_{1-\alpha/2}$ is the quantile of student distribution with $n - 1$ degrees of freedom.

All simulations and computations were done by the statistical freeware *R*, version 4.2.1 (R Core Team, 2022). We run simulation for three different distributions of variable *Y* with different shape (see Table 1).

Table 1 Distribution of variable *Y*

Indicator	**Mean** E(*Y*)	**Standard deviation** sd(*Y*)	*P* (0 ≤ *Y* ≤ 3)
Exponential (*λ* = 1)	1.0	1.00	0.950
Normal (*μ* = 1.5, *σ* = 0.5)	1.5	0.50	0.997
Uniform (*U*(0,3)	1.5	0.87	1.000

Source: Own construction

For each distribution, we generate 1 000 populations of size *N*. From each realization of the population, we then draw 1 000 samples of size *n* by simple random sampling without replacement. We generate populations of size *N* = {200, 400, 1 000} from which we select samples of size ranging from 20 to 100. Populations and samples of these sizes are common in applications of RRT (size of community, village) or any strata from business or social statistics survey. Pseudorandom numbers come from uniform distribution on interval (0,3). This interval covers most of the range of values of the variable *Y* (see the last column of Table 1). Since there is no prior information on distribution of *Y*, we use the default value of the tuning parameter $\alpha = 0{,}75$.

The performance of the variance estimators is assessed by covering the approximate confidence interval with the chosen confidence level. Proportion of simulated samples with approximate confidence interval covering the population mean is compared with prescribed confidence level. All simulation results of coverage of two-sided 90%, 95% and 99% confidence interval for the mean are summarized in Tables 2–4 (see the Annex). Variance estimators works well for symmetric distributions (normal, uniform) for both methods. Approximate confidence intervals then work well even for small sample and population sizes. For distribution skewed from the left (exponential) the coverage of approximate is a bit underestimated. This effect is observed for estimator $\bar{t}^{HT,R}_{(0,M)}$ using the variance estimator by Formula (14). However, the same effect is also present in confidence intervals for estimator (17) with using pseudorandom values $\bar{t}^{HT,R}_{\alpha,(0,M)}$. Preliminary analysis indicates that for skewed sensitive variable its variance estimates are skewed in the same direction. It implies for exponentially distributed sensitive variable, that too many confidence intervals might be too narrow, even if the population mean and its estimate are very closed. Also, presence of negative values of transformed randomized responses *R* or values of *R* exceeding its upper bound *M* of the sensitive surveyed variable *Y* might be the reason.

DISCUSSION AND CONCLUSION

We derived unbiased variance estimators for mean estimators using new RRT proposed by Antoch et al. (2022). Unfortunately, for this, it is necessary to breach respondents' privacy slightly for the first estimator, because the value of the pseudorandom numbers *U* must be available to an interviewer (they also know the questions asked). The reason is that, unlike other RRTs, the randomized response does not contain a transformed sensitive value *Y* (they only provide Yes/No responses instead of a numeric value). For the second estimator, the respondents' privacy can breach less, if the respondents take some effort and calculate the transformed randomized response R by themselves and provide them to an interviewer.

The simulation study was designed to study the coverage of two-sided intervals of the mean for different populations (symmetric, skewed from the let, uniform), using reliability, population and sample size resembling the setting in real-life applications. For symmetric distributions, approximate confidence intervals using proposed variance estimators work well even for small sample and population sizes.

However, performance of variance estimators and confidence intervals seems to get worse if distribution is skewed. For the distribution skewed to the left, the coverage of confidence intervals seems to be smaller than the chosen reliability. The relationship between shape of the distribution and the coverage of approximate confidence intervals is the topic for further research.

ACKNOWLEDGMENT

This work of was supported by Institutional Support to Long-Term Conceptual Development of Research Organization, the Faculty of Informatics and Statistics of the University of Economics, Prague. The author is grateful to RNDr. Radka Lechnerová, Ph.D. from the Czech Statistical Office and the unknown reviewers for their valuable comments that considerably improved the contents of this paper.

References

ANTOCH, J., MOLA, F., VOZÁR, O. (2022). New Randomized Response Technique for Estimating the Population Total of a Quantitative Variable [online]. *Statistika: Statistics and Economy Journal*, 102(2): 205–227. <https://doi.org/10.54694/stat.2022.11>

ARNAB, R. (1994). Nonnegative variance estimation in randomized response surveys [online]. *Communication in Statistics – Theory and Methods*, 23(6): 1743–1752. <https://doi.org/abs/10.1080/03610929408831351>.

CHAUDHURI, A. (1987). Randomized response surveys of a finite population: a unified approach with quantitative data [online]. *J. Statistical Planning and Inference*, 15: 157–165. <https://doi.org/10.1016/0378-3758(86)90094-7>.

CHAUDHURI, A. (1992). Randomized response: estimating mean square errors of linear estimators and finding optimal unbiased strategies. *Metrika*. 39: 341–357.

CHAUDHURI, A. (2017). *Randomized Response and Indirect Questioning Techniques in Surveys.* New York: Chapman and Hall/CRC. ISBN 978-11-3811542-2

CHAUDHURI, A., CHRISTOFIDES, T., RAO, C. (2016). *Handbook of Statistics 34. Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques.* Amsterdam: Elsevier. ISBN 978-0444-63570-9

DALENIUS, T., VITALE, R. (1979). A new randomized response design for estimating the mean of a distribution. In: JUREČKOVÁ, J. (eds.) *Contributions to Statistics*, Prague: Academia, 54–59.

ERIKSSON, S. (1973). A new model for randomized response [online]. *International Statistical Review*, 41: 101–113. <https://doi.org/10.2307/1402791>.

R CORE TEAM (2022). R: *A language and environment for statistical computing* [online]. Austria, Vienna: R Foundation for Statistical Computing. <https://www.R-project.org>.

SÄRNDAL, C., LUNDSTRÖM, S. (2005). *Estimation in Surveys with Nonresponse.* Chichester: J. Wiley and Sons. ISBN 978-0470-01133-1

SÄRNDAL, C., SWENSSON, B., WRETMAN, J. (1992). *Model Assisted Survey Sampling.* Heidelberg: Springer. ISBN 978-0387-40620-6

TILLÉ, Y. (2020). *Sampling and Estimation from Finite Populations.* New York: J. Wiley and Sons. ISBN 978-0470-68205-0

ANNEX

Table A1 The coverage of two-sided 90% confidence interval for mean

N	*n*	$\bar{t}_{(0,M)}^{HT,R}$			$\bar{t}_{\alpha,(0,M)}^{HT,R}$		
		Y ~ Exp	***Y ~ Norm***	***Y ~ Unif***	***Y ~ Exp***	***Y ~ Norm***	***Y ~ Unif***
200	20	0.885	0.903	0.901	0.894	0.899	0.901
	50	0.877	0.900	0.901	0.886	0.900	0.901
400	20	0.889	0.903	0.894	0.895	0.890	0.900
	50	0.878	0.900	0.901	0.886	0.899	0.901
1 000	20	0.893	0.902	0.886	0.895	0.897	0.900
	50	0.880	0.901	0.900	0.887	0.899	0.901
	100	0.869	0.900	0.900	0.875	0.899	0.900

Source: Own construction

Table A2 The coverage of two-sided 95% confidence interval for mean

N	*n*	$\bar{t}_{(0,M)}^{HT,R}$			$\bar{t}_{\alpha,(0,M)}^{HT,R}$		
		Y ~ Exp	***Y ~ Norm***	***Y ~ Unif***	***Y ~ Exp***	***Y ~ Norm***	***Y ~ Unif***
200	20	0.929	0.933	0.950	0.948	0.948	0.952
	50	0.929	0.948	0.950	0.941	0.949	0.951
400	20	0.925	0.931	0.955	0.948	0.947	0.951
	50	0.930	0.948	0.950	0.941	0.949	0.951
1 000	20	0.918	0.928	0.958	0.942	0.947	0.951
	50	0.932	0.950	0.949	0.942	0.949	0.951
	100	0.926	0.949	0.950	0.932	0.949	0.950

Source: Own construction

Table A3 The coverage of two-sided 99% confidence interval for mean

N	*n*	$\bar{t}_{(0,M)}^{HT,R}$			$\bar{t}_{\alpha,(0,M)}^{HT,R}$		
		Y ~ Exp	***Y ~ Norm***	***Y ~ Unif***	***Y ~ Exp***	***Y ~ Norm***	***Y ~ Unif***
200	20	0.974	0.981	0.988	0.990	0.988	0.992
	50	0.978	0.987	0.990	0.987	0.989	0.991
400	20	0.974	0.981	0.988	0.990	0.988	0.991
	50	0.979	0.987	0.989	0.988	0.989	0.990
1 000	20	0.974	0.980	0.988	0.990	0.987	0.991
	50	0.979	0.989	0.989	0.987	0.989	0.990
	100	0.978	0.990	0.990	0.984	0.989	0.990

Source: Own construction