Journal of Official Statistics
Vol. 5. No. 1, 1989, pp. 23–39

The Optimal Design of Quality Control Samples to Detect Interviewer Cheating

Paul P. Biemer and S. Lynne Stokes[1]

"He is not cheated who knows he is being cheated."
Sir Edward Coke, Institutes (1628)

Abstract: Without interviewer quality control, interviewer cheating can seriously affect the accuracy of survey results. This paper proposes a method for designing quality control samples which maximizes the probability of detecting cheating for a fixed cost. First, data on interviewer cheating from a recent U.S. Bureau of the Census study are presented. Then a statistical model for describing dishonest interviewer behavior is proposed which assumes cheating is a random event governed by a probability distribution whose parameters depend on the interviewer. These parameters control the frequency and intensity of cheating as well as the geographic clustering of the falsified units.

A general quality control sample design and several associated cost models are proposed. A procedure for optimally choosing the sample design parameters according to specific types of interviewer behavior is described. Finally, the procedure is applied to optimize the interviewer quality control system used by the U.S. Bureau of the Census for the Current Population Survey and other current surveys.

Key words: Current Population Survey; National Crime Survey; reinterview; "curbstoning;" nonsampling error; survey costs.

1. Introduction

Every survey data collection organization, especially those that conduct personal interviews, must deal with the problem of interviewer cheating. The most blatant example of cheating occurs when an interviewer fabricates the responses for an entire questionnaire. Sometimes, however, cheating takes a more subtle form. For example, an interviewer may ask some questions in an interview and fabricate the responses to others. An interviewer may deliberately deviate from prescribed procedures, such as conducting a telephone interview where a face to face interview was indicated or conducting the interview with a willing but inappropriate respondent.

One of the most common methods used for detecting interviewer cheating in personal interview surveys is the verification method. For

[1]Paul Biemer is Head, Department of Experimental Statistics, and Director, University Statistics Center, New Mexico State University, Las Cruces, NM 88003-3130, U.S.A. S. Lynne Stokes is Assistant Professor, Department of Management Science and Information Systems and Center for Statistical Sciences, University of Texas, Austin, TX 78712, U.S.A.

this method, a sample of an interviewer's assignment is recontacted in order to verify that an interview was conducted as required and that (at least) the critical components of the questionnaire were obtained accurately.

The question we address in this paper is how to design the verification sample in order to maximize the probability of detection of a cheating interviewer at least once during a specified time period. The methodology developed is appropriate for organizations whose interviewer staff is stable and whose interviewers participate regularly in surveys in which they have similar workloads. Although the emphasis here is on "in-the-field" interviewing (i.e., face to face and decentralized telephone interviewing), the methodology is adaptable to centralized telephone interviewing. In that case verification takes the form of a system of unobtrusive telephone monitoring.

Since the resources allocated to this aspect of a survey's quality control program is generally quite limited, only a small portion of the interviewer's workloads can be verified. The competing choices we allow to be made concerning the verification design are (a) how often the interviewer is chosen for verification (b) how much of his/her assignment is inspected when he/she is chosen and (c) what size the sampling units (persons, households, or groups of households) should be. The optimal choice depends on an individual interviewer's cheating behavior and the cost of the design choices.

In Section 2 we review information which has appeared in the literature concerning interviewer cheating behavior. In addition, the data resulting from a program implemented by the U.S. Bureau of the Census in 1982 to collect such data are reported. In Section 3, a model for interviewer cheating behavior is suggested and the probability of detection for a given verification scheme is derived. Section 4 gives some empirical rules for optimal design for our cheating model for two typical cost function forms. Finally, in Section 5 the model is used to develop a verification sample design for the Current Population Survey (CPS), the largest demographic survey run by the U.S. Bureau of the Census. This application motivated the development of our model.

2. Interviewer Cheating Behavior

Interviewer cheating has long been recognized as a problem among survey organizations. Crespi (1945) conjectured that "almost every interviewer will eventually succumb [to cheating] ... if fabrication is made to appear the only practicable solution to the problems facing the interviewer." He suggested several factors, related to either the questionnaire or the administration of the survey, which may operate to demoralize the interviewer. Related to the questionnaire were: (1) questionnaire length and respondent burden, (2) poor questionnaire design, e.g., apparent repetition of the same questions, and (3) difficult or antagonistic questions. Among the administrative demoralizers are: (1) overly difficult assignments or inadequate remuneration, (2) improper or inadequate training, (3) use of part-time interviewers for whom the unpredictable demands of interviewing may compete with the necessities of another job and (4) external factors such as the weather, bad neighborhoods, roads, etc. which may operate to encourage cheating.

Crespi's proposed solution to the cheater problem is the dual strategy of (a) eliminating the demoralizers by careful and intelligent survey design and administration with ample opportunity for interviewer advisement (for example, the present-day "quality circles") and (b) using the verification method to deter cheating. Bennett (1948 a and b), Sheatsley (1951), Boyd and Westfall (1955), and Evans (1961) appeared subsequently which provided suggestions to help the survey practitioner implement part (a) of this strategy.

There is little guidance in the literature, however, on implementing (b). Making the problem worse is the paucity of published data on characteristics of interviewers who cheat or how they do it. This is understandable because cheaters are difficult to detect, and few studies have made the collection of such data their major goal. The first reported data was made for the National Opinion Research Center (Sheatsley (1951)) and included characteristics of the small number of interviewers who were dismissed for cheating between 1941 and 1949.

In 1982, the U.S. Bureau of the Census began a program to collect information on all confirmed or suspected cases of cheating by interviewers in their current surveys. The purpose of the data collection was to aid in the modeling of cheating behavior and ultimately in the selection of an optimal verification design. Cheating problems targeted were complete or partial fabrication of the survey responses as well as other improper interviewer conduct, such as use of proxy respondents in situations where self-response was required.

The first results available from this study covered the period September 1982 through August 1985 (Bureau of the Census (1986)). During that time, it was established that 140 interviewers (about 3–5 % of all interviewers) committed some form of cheating, and an additional 31 interviewers were suspected of cheating. Of the 140 confirmed cases, 100 were identified through the reinterview verification program. The remaining 40 cases were detected by other means, such as inspection of the returns, information provided by other interviewers, etc. Overall, most of the cheating (72 %) involved complete fabrication of interviews. The next most frequent violation was the misclassification of units as vacant when they were, in fact, occupied (17 %). Indeed, this form of cheating is just as damaging as falsifying an entire interview since the unit is then erroneously regarded as out-of-scope for the survey. In the National Crime Survey (NCS), which requires that each respondent answer for him/herself, 20 out of the 26 confirmed cases of cheating involved the violation of this self-response rule, often accompanied by other infractions as well.

Some further results of this study are now summarized.

1. For the two largest demographic surveys, the CPS and the NCS, 87 % of the falsified interviews occurred in urban areas, only 13 % in rural areas. Since roughly 70 % of the sample is located in urban areas for these surveys, there is evidence (statistically significant at the 5 % level of significance) of a higher degree of cheating in urban areas.
2. Table 1 shows the distribution of cheaters (CPS and NCS only) by years of service with the Bureau of the Census. Almost half of the confirmed violators have less than one year of service, while only 23 % of all interviewers have less than one year of service. These data indicate a substantial and highly significant tendency for relatively inexperienced interviewers to cheat more frequently than interviewers having one or more years of experience. Alternatively, the data may indicate an adeptness of more experienced interviewers for escaping detection of cheating.

Table 1. Distribution of CPS and NCS interviewers found cheating by years of experience

Length of service	Cheaters (%)	All interviewers (%)
Less than 1 year	46	23
1–2 years	13	28
3 years or more	43	48

3. Experienced interviewers (i.e., those with a year or more of experience) who were detected cheated at an average rate of 19 % of the households in their assignments for CPS and NCS. Furthermore, less than 13 % of these cheaters were involved in fabrication of responses. By contrast, interviewers with

less than one year of service displayed a tendency to cheat at a much higher rate, viz., an average of 30 % of the households in their assignments, with roughly half of the cheaters being involved with the complete fabrication of interviews.

3. The Model

In Section 3.1, a simple model is proposed that describes interviewer cheating behavior. This model views cheating as a random event governed by a specified probability distribution which depends on several parameters that control the frequency and pattern of cheating. In conjunction with this model, a general design for the verification sample is described. It is a variation of the one previously used by the U.S. Bureau of the Census for the CPS and is a generalization of the one in current use. This design is described in Section 3.2. In Section 3.3, the probability of detection of cheating by an interviewer for any specific verification design is derived. This probability will depend, of course, on the interviewer's cheating parameters. Finally, in Section 4.4, two models for the cost of verification sampling are proposed. If all parameters of the interviewer cheating and cost models were known, an optimal design could be selected. Such a design is defined to be one which maximizes the probability of detection.

One problem with this approach is that interviewers do not all behave alike, and therefore a design which is optimal for one may not be optimal for another. A solution to this problem is to divide the interviewers into strata defined by their frequency and pattern of cheating. For example, the data described in Section 2 suggest that interviewer experience would be a good stratifying variable for CPS interviewers. The optimal verification design parameters could be determined for each stratum.

The second problem with the approach, however, is that the parameters of the cheating model for individual interviewers or groups of interviewers are not known in advance. Furthermore, it is difficult to obtain enough information to estimate the parameters of the model proposed in Section 3.1. During the study described in Section 2, the interviewer was either fired or resigned in 97 % of the cases in which he or she was detected fabricating an interview. Information about patterns of cheating is not available when the behavior can be observed only once.

Nevertheless, the model developed can be useful for determining a verification design. Two possibilities for its use follow. First, one might optimize the design against the most damaging violators. This may be, for example, interviewers who falsify more than some specified fraction of their assignments. Second, one might use the model to choose a design, if one exists, which is nearly optimal against a wide range of likely interviewer behavior. The latter of these is the use made of the model for the CPS application described in Section 5.

3.1. *Model for interviewer cheating*

Consider a complex survey with any probability sampling design for which ultimate stage sampling units (USUs) are clusters of m interview units. For example, USUs may be geographical segments of m housing units or they may be households of m individuals. The time period for the survey during which interviewers are to be evaluated will be referred to as the observation cycle. Every interviewer is to be inspected at least once during an observation cycle. Let f denote the number of times the survey is repeated during one observation cycle and refer to these repetitions as interviewing periods.

Consider a particular interviewer for some interviewing period within an observation cycle. Let n denote the number of USUs in the interviewer's assignment and let the random variables b_h $(h = 0, \ldots, m)$ denote the number of USUs having exactly h misrepresented (or fabricated) interview units. Hence $\sum_{h=0}^{m} b_h = n$.

In the next paragraph, we propose a probability distribution for $\mathbf{b}' = (b_0, \ldots, b_m)$ which is a mixture of two distribution functions: a Bernoulli distribution with parameter π and a multinomial distribution with parameter vector $\mathbf{p}' = (p_0, \ldots, p_m)$.

There are numerous factors which may influence an interviewer's decision to cheat during an interviewing period, as Crespi suggested. Some of these factors are always present, such as problems with a questionnaire. Others occur only occasionally, such as problems with the weather or interference from a part-time job. Consequently, some interviewers (those influenced by the ever-present problems) may be susceptible to cheating at all times. Others (those influenced only by irregularly occurring events) may cheat only when there is no other way to complete their assignments. We attempt to capture this behavior pattern in our model by defining a parameter π to be the probability that an interviewer considers or has some positive probability of cheating, and we refer to π as the frequency of cheating. Next we define $P(\mathbf{b})$ to be the probability distribution of $\mathbf{b}' = (b_0, \ldots, b_m)$ associated with the interviewer and assume that $P(\mathbf{b})$ is the multinomial distribution with parameters $\mathbf{p}' = (p_0, \ldots, p_m)$ and n. This multinomial assumption suggests that the probabilities of falsifying $h = 0, 1, \ldots, m$ units in a cluster is the same for every cluster. However, given that one or more units have been falsified in a cluster, the multinomial distribution allows us to change the probability that other units in that cluster will be falsified. That is, we may, through $\mathbf{p}$, arrange for a nonzero intracluster correlation for falsified units. (In the sequel, an interviewer whose cheating behavior is governed by the model with parameters $(\mathbf{p}, \pi)$ will be referred to as a $(\mathbf{p}, \pi)$ cheater.) Thus, p_h is the probability that h units in a given USU are misrepresented. We define $\bar{p}$, referred to as cheating intensity, to be the expected proportion of misrepresented interview units in a $(\mathbf{p}, \pi)$ cheater's assignment, given that he or she is susceptible to cheating; i.e.,

$$\bar{p} = \sum_{h=0}^{m} hp_h/m. \tag{3.1}$$

The decision to cheat for a particular unit may not be made independently of other units in an interviewer's assignment. For example, cheating may be concentrated within certain USU's which share characteristics that may influence the interviewer to cheat, such as nontelephone households, undesirable neighborhoods, or areas with difficult access. It is possible with this model to describe the strength of this clustering effect by appropriately defining the p_hs. If no clustering is present, for example, misrepresented units will be distributed among the interview units in each USU according to a binomial distribution, so that $p_h = \binom{m}{h}\bar{p}^h(1-\bar{p})^{m-h}$. Perfect clustering would mean that either all or no units in a USU are misrepresented; i.e., that $p_0 + p_m = 1$. One can show that these two extreme conditions yield extreme values (0 and 1 respectively) of the intracluster correlation

$$\delta = \mathrm{Cov}(y_{ij}, y_{ij'})/\mathrm{Var}(y_{ij}), \tag{3.2}$$

where $y_{ij} = 1$ if the jth unit in the ith USU is misrepresented and 0 otherwise. Correct specification of the magnitude of δ is important, since it has an impact on the optimal reinterview sample design.

This simple model can describe a wide range of interviewer cheating behavior. The consistent, low-level cheater can be modeled by setting π large and $\bar{p}$ small, while the erratic cheater can be accomodated by setting π small. Clustering of the affected units can be modeled by the relative values of the p_hs.

Table 2. Notation

Symbol	Definition
n	number of USU's in an interviewer assignment
m	number of units in a USU
b_h	number of USUs having h falsified units; $\sum_h b_h = n$.
I	number of interviewers for the survey
f	number of interviewing periods in an observation cycle
s	number of interviewers selected for each supplementary sample
ℓ	number of USU's to be reinterviewed in each interviewer assignment
t	number of units to be reinterviewed in each sample USU
π	probability that an interviewer is susceptible to cheating
p_h	probability h units in a USU are falsified
δ	intracluster correlation coefficient for cheating; see (3.2).

3.2. *The verification sample design*

One major goal of the verification sample is to detect interviewer cheating; a second is to deter it. In order to meet the first goal, we search for a design which, for a given cost, will maximize the probability of selecting the units in a sample which are misrepresented. So that the design we choose will be a deterrent, we require the following:

1. Every interviewer must be selected at least once during an observation cycle.
2. The selection of interviewers and units must be unpredictable by the interviewers.

Many designs satisfy these criteria. One of them, which is now used by the U.S. Bureau of the Census, will illustrate the design optimization methodology. For simplicity, we assume a single population of interviewers having cheater parameters $(\mathbf{p}, \pi)$. For the case where interviewers are divided into strata, the procedure described below may be applied separately in each stratum.

Let the I interviewers for the survey be divided into f mutually exclusive and exhaustive groups. (Recall that f is the number of interview periods in an observation cycle.) To simplify the subsequent formulas, we assume that I is evenly divisible by f. Randomly order the groups. Then all interviewers in group i will be selected for verification in the ith interviewing period. This group will be referred to as the ith predesignated sample. In addition, let a specified number, say s, of interviewers be selected at random from the remaining $I\,(1 - 1/f)$ interviewers. We refer to this group as the supplementary sample. This group introduces more unpredictability into the selection of interviewers. From each selected interviewer's assignment a sample of ℓ USU's is randomly chosen. Within each USU, t units are reinterviewed. Thus, the total sample size in each interviewing period is $(I/f + s)\ell t$.

If cheating is observed for any of the ℓt interview units in an interviewer's assignment, we say that a cheater was detected. Our objective is to find the verification design parameters s, ℓ, and t which maximize the probability of detecting a cheater for a fixed total cost under a specified interviewer cheating model.

3.3. *Derivation of detection probability*

The probability of detecting a cheater with verification design parameters, (s, ℓ, t) will be denoted by $D(s, \ell, t)$. In Appendix 1, it is shown that, for a $(\mathbf{p}, \pi)$ cheater

$$D(s, \ell, t) = 1-[1-P_t(\mathbf{p}, \pi)]\left[1-\frac{fs}{I(f-1)}\right]P_t(\mathbf{p}, \pi)^{f-1}, \quad (3.3)$$

where

$$P_t(\mathbf{p}, \pi) = \pi(1-\eta_t^{\ell}) \text{ and } \eta_t = \binom{m}{t}^{-1} \sum_{h=0}^{m} \binom{m-h}{t} p_h, \quad (3.4)$$

where $\binom{m-h}{t} = 0$ if $t > m - h$. η_t can be described as the probability that a $(\mathbf{p}, \pi)$ interviewer is not observed cheating in a sampled USU during an interviewing period in which he is susceptible. $P_t(\mathbf{p}, \pi)$ is then the conditional probability that a $(\mathbf{p}, \pi)$ cheater is observed cheating in the ith interviewing period, given that he or she is selected in either the ith predesignated or corresponding supplementary samples. (Note that $P_t(\mathbf{p}, \pi)$ does not depend upon i.)

One special case of (3.4) that will be considered in detail is when all interview units are inspected within the sample USU's, or $t = m$. In that case $P_t(\mathbf{p}, \pi)$ becomes

$$P_m(p_0, \pi) = \pi(1 - p_0^{\ell}). \tag{3.5}$$

Then only π and p_0 need be specified to compute $D(s, \ell, t)$.

3.4. *A general cost model for design optimization*

The cost model described in this section is general enough to apply for many verification operations. Let $K(s, \ell, t)$ denote the variable costs for an observation cycle of a design having parameters s, ℓ, and t. Then, for constants C_1, C_2, C_3, and C_4,

$$K(s, \ell, t) = (I + fs)[C_1\ell + C_2\ell t + C_3(t)\sqrt{\ell} + C_4] \tag{3.6}$$

where C_1 is the cost associated with each sampled USU's (for example, a sampling cost); C_2 is the cost incurred for each unit inspected; $C_3(t)$ allows for a travel cost for situations in which interviewers must travel varying distances to the USU's. This cost will depend upon t if the cost of travel to inspect all units in a USU depends on the number of units verified in a USU. C_4 is the fixed cost associated with each sampled interviewer.

Two special cases of (3.6) will be considered in Section 4. The first is the case where cost is simply proportional to the total sample size, i.e., $C_1 = C_3 = C_4 = 0$, or

$$K_1(s, \ell, t) = (I + fs)C_2\ell t. \tag{3.7}$$

This cost function might be appropriate for a verification program which relies solely on telephone reinterviewing of a sample of survey respondents. There are no costs for travel or any other disaggregate costs associated with the number of sample USU's.

For designs in which all the interview units are revisited in person, but moving from one unit to another within a USU does not incur much additional travel costs, a reasonable cost function may be obtained setting $C_1 = 0$ in (3.6) and allowing $C_3(t) = C_3$. Then

$$K_2(s, \ell, t) = (I + fs)\ (C_2\ell t + C_3\sqrt{\ell} + C_4). \tag{3.8}$$

The cost of traveling to the ℓ USUs is proportional to $\sqrt{\ell}$ if the USUs are randomly distributed within the interviewer's assignment area (Pielou (1969, p. 111)).

4. Some Rules for Optimal Design

In this section, we state some general rules, some analytical and some empirical, concerning the optimal choice of a verification design when cost functions of the form given by (3.7) or (3.8) are appropriate. The optimal design is determined by maximizing the detection probability $D(s, \ell, t)$, subject to the constraint $K(s, \ell, t) = C_F$, where C_F is the total fixed cost for the verification sample.

Rules 1 through 3 deal with determining an optimal trade-off between the frequency that an interviewer is sampled (s) and the thoroughness of the inspection of his or her assignment (ℓ). The number to be selected from each USU is assumed fixed, which is equivalent to assuming a fixed value for η_t, defined in (3.4).

Rule 1. Assume the cost function $K_1(s, \ell, t)$ in (3.7) and $\pi = 1$. Then for t held fixed, any values of s and ℓ satisfying $K_1(s, \ell, t) = C_F$ will provide a sample design that is either optimal or near optimal.

Rule 1 is supported by Theorem 1, which is stated and proved in Appendix 2. Theorem 1 says that, under the conditions of Rule 1, the maximum detection probability for fixed cost can be achieved by choosing s at either of its extremes (i.e., $s = 0$ or $s = \frac{f-1}{f}I$). Numerical investigations suggest further that the detection function $D(s, \ell, t)$ is very flat over the entire range of s for likely choices of C_F and $\mathbf{p}$.

Therefore, when the cost model described in (3.7) is appropriate, the probability of detection of the consistent cheater varies little with the choice of s and ℓ. The verification design choice can safely be made, then, on the basis of other considerations. For example, a design which is optimal for erratic cheaters could be implemented, and the survey managers could be assured that it would be near optimal for the consistent ones as well.

Rule 2. Assume the cost function $K_2(s, \ell, t)$ in (3.8) and $\pi = 1$. Then for t held fixed, choosing s small but non-zero such that $K_2(s, \ell, t) = C_F$ will provide an optimal or near optimal design.

Rule 2 is supported by Theorem 2, which is stated and proved in Appendix 2. Theorem 2 says that, under the conditions of Rule 2, the probability of detection obtained by choosing s at its minimum ($s = 0$) is always larger than that obtained by choosing s at its maximum ($s = \frac{f-1}{f}I$). Numerical investigations further suggest that a choice of $s = 0$ actually maximizes the detection probability for likely choices of $\mathbf{p}$, C_F and for a wide range of cost function parameters C_2, C_3, and C_4.

In practice, however, the choice of no supplementary sample ($s = 0$) eliminates the unpredictability of the verification design. Since unpredictability was one of its specifications, this choice is not acceptable. The best strategy, as Rule 2 suggests, is therefore to choose s small, but non-zero.

When $\pi < 1$, numerical investigations have shown that no verification design exists which will come close to maximizing the probability of detection over all reasonable specifications of the cheating and cost models for either form of the cost function. Examples can be found for which detection probability is maximized at either extreme or at intermediate values of s. Furthermore, the loss in detection probability from a poor choice of a survey design is sometimes large. This observation suggests the following rule.

Rule 3. No optimal verification design exists for detecting an interviewer having $\pi = \pi_0 < 1$. Instead, the best strategy is sensitive to $\mathbf{p}$ and parameters of the cost models. Therefore, the best possible information about these parameters should be collected and sensitivity analyses performed to aid in the choice of design.

An application of Rule 3 is illustrated in Section 5.

Our discussion so far has dealt solely with the choice of s and ℓ. Now we turn to the problem of determining the optimal choice for t. It is affected by the magnitude of δ, defined in (3.2). Rules 4 and 5 address the choice of t when δ assumes one of its extreme values.

Rule 4. If there is no clustering of misrepresented units (i.e., $\delta = 0$) and the cost function is $K_1(s, \ell, t)$ defined in (3.7), then any choice of t is equally good. However, if the cost of sampling a new USU is greater than the cost of sampling a comparable number of units within existing sample USUs, as for $K_2(s, \ell, t)$, then $t = m$ is optimal.

Rule 5. If misrepresented units are perfectly clustered (i.e., $\delta = 1$), then the best strategy is to choose $t = 1$ regardless of the cost function.

Rules 4 and 5 follow from the observations about δ made in Section 3. If there is no clustering within a USU, an identical amount of information can be obtained from sampling two units within the same USU as from different USU's. Therefore the optimal choice depends on which is cheaper. Since for $K_1(s, \ell, t)$, either choice has identical cost, the choice of t is unimportant. For $K_2(s, \ell, t)$, there is a saving of costs associated with remaining within a USU, so the best strategy is to choose t as large as possible. When there is perfect clustering of misrepresented units, selecting more than one unit per USU buys you no information and thus is wasteful, if it costs anything at all.

For those frequent cases in which δ is between the two extremes, the optimal choice of t is not so easy to make. However, the two rules together suggest that when cost is directly proportional to the number of units verified (as for K_1), $t = 1$ should be chosen. When K_2 is the appropriate cost function, an analysis such as that undertaken in Section 5 is required to determine the optimal t.

5. An Application to the Current Population Survey

5.1. *Description of the Current Population Survey*

The CPS provides the official labor force statistics for the United States. The survey is conducted monthly by the U.S. Bureau of the Census and has consisted of between 50 000 and 60 000 household interviews per month. In addition to data on employment and unemployment, the survey also provides information on annual household and individual income.

The sample is a stratified multistage cluster sample having USU's which are typically clusters of four neighboring households. The segments are selected at random within primary units which are essentially counties or groups of counties. Interviewer assignments generally average about 12 segments or about 48 housing units. Due to the rotational design of the sample, about an eighth of the housing units in an assignment are new to the program, an eighth are being interviewed for the second time, and so on up to an eighth being interviewed for the eighth and last time. Between 30 % and 40 % of the interviews are conducted face to face while the remainder are conducted by telephone.

The CPS interview quality control program consists of three components: (a) a reinterview survey to detect interviewer cheating, (b) an inspection of all the interview forms by clerks who are specially trained to detect interviewer errors in completing the forms, and (c) an annual on-site observation of the interviewer by a supervisory representative as the interviewer completes an assignment. The purpose of the annual observation is to provide an expert evaluation of the interviewer's interviewing technique. The remainder of the section will be concerned with the sample design of the quality control reinterview survey.

5.2. *The quality control reinterview*

For a sample of households, the reinterviewer, who is typically a senior interviewer and supervisory representative, re-asks some or all of the questions on the original questionnaire. Any discrepancies between the original interview and the reinterview are reconciled with the respondent. The reinterviewer also determines whether the discrepancy was the fault of the interviewer or the respondent. In addition to detecting interviewer cheating, the reinterview also serves as a device for detecting both deliberate and unintentional errors that occurred in the original interview. The number of errors found in the interviewer's assignment that are the fault of the interviewer are tallied and these results are immediately reported to the interviewer for corrective action. Whenever possible, the reinterviews are conducted by telephone. Otherwise, a face to face reinterview is conducted and travel costs are incurred.

Prior to 1982, the CPS quality control reinterview program required that each interviewer be randomly selected for reinterview twice per year, once during the first six months of the year and again in the last six months. Each time an interviewer was selected, all the households in a randomly chosen third of the approximately 12 USU's in his/her assignments were reinterviewed. Thus, about one eighteenth of all CPS households (i.e., one sixth of the interviewers and one-third of each interviewer's assignment) were reinterviewed each month. This design was flawed since the time of reinterview was somewhat predictable. For example, an interviewer selected for reinterview in January could not be selected again until July. Many interviewers were aware of this pattern so that the reinterview was less effective as a deterrent.

In 1982, a redesigned CPS quality control reinterview program was implemented to correct this deficiency. Prior to implementation a study was conducted to aid in the selection of an improved sample design (Biemer, Judkins, Schreiner, and Stokes (1982)). The sampling scheme described in Section 3 was adopted for the CPS and subsequently for all the Bureau's continuing demographic surveys. The observation cycle was chosen to be twelve months, so that the predesignated sample consisted of $I/12$ interviewers. The design options to be determined, then, were the number of USU's to be sampled from each interviewer's assignment (ℓ) and the number of households to sample from each USU (t). The size of the supplementary sample (s) was determined by the cost constraint. We begin with a description of the cost function for the reinterview program.

5.3. *Cost function*

The cost of reinterviewing an interviewer's assignment can be decomposed into several components. There will be costs associated with the time involved in conducting face to face reinterviews and different costs for telephone reinterviews. Other costs arise from reconciling differences between interview and reinterview responses, completing the reinterview forms, and the time involved in discussing the results with the interviewers. For face to face reinterviews, mileage costs and cost for reinterviewer time while traveling and while conducting the reinterviews will be incurred. For telephone reinterviews, there will be no travel costs; however, telephone toll charges may be incurred.

A detailed analysis of the CPS reinterview costs was conducted and it was determined to be well-described by (3.6), but with $C_1 = 0$. Then the cost coefficients C_2, $C_3(t)$, and C_4 were estimated. The details of this analysis are documented in a Bureau of the Census report (Biemer et al. (1982)). In addition to a national cost function, a number of subnational cost functions were developed corresponding to urban, suburban, and rural areas where travel costs differ substantially. For the present illustration, only the national model will be considered. For a reinterview survey with design parameters (s, ℓ, t), the annual variable cost is given by the following model:

$$K(s, \ell, t) = (I + 12s)(2.41\, \ell t + C_3(t)\sqrt{\ell} + 79.58), \quad (5.1)$$

where

$$C_3(t) = 7.19\sqrt{1 - .85^t}\,. \quad (5.2)$$

The choice of $C_3(t)$ is explained by noting that the per USU cost depends on the average number of visits required for each USU in the reinterview sample, information which was not directly available. However, it may be assumed that a visit to a USU was made only if one of the t sampled households within it required a personal visit. Since roughly 85% of the CPS reinterview households have face to face reinterviews, we estimated by assuming a binomial distribution that approximately $(1 - .85^t)\ell$ USUs out of the ℓ sampled USUs would contain at least one such household. Thus, travel costs increase in proportion to the square root of this expression (see comment under (3.8)). Data from the main CPS survey were used to estimate the travel costs associated with each segment visited. The coefficient 7.19 in (5.2) includes the cost for travel time as well as mileage.

The coefficient $C_2 = 2.41$ is an average cost for reinterview time associated with each sample unit. C_2 includes the cost of the time spent for a face to face reinterview and the cost of the time for a telephone reinterview averaged over all reinterviewer assignments. Finally, $C_4 = 79.58$ is the average cost for reinterviewer travel from his/her home base to the area of the interviewer's assignment.

The optimization results which follow were found, through sensitivity analyses, to be quite robust to absolute errors in these coefficients. More critical are the relative errors among the coefficients, i.e., the proportion of total costs accounted for by each cost component. Yet even here quite substantial changes in the relative sizes of C_2, $C_3(t)$, and C_4 had only moderate impact on the optimal design.

5.4. *Determining the "near" optimal design of the CPS*

Since virtually nothing was known at the outset about interviewer cheating behavior, it was not possible to select an optimal reinterview design with the model described in Section 3. Instead, the model was first used to determine to what extent the detection probability was sensitive to the design choice over likely ranges for the parameters $\mathbf{p}$ and π. Rule 2 of Section 4 suggests that the trade-off between s and ℓ is not critical for consistent interviewers ($\pi = 1$). However, it was believed that the conditions causing cheating by CPS interviewers were temporary, and thus erratic cheaters ($\pi < 1$) should be considered the target. Rule 3 says that in this case, the optimal choice of ℓ and s is highly variable. That knowledge led to the data collection program whose results are described in Section 2.

We did know from the start that misrepresented units were not perfectly clustered within USU's. What we did not know was if there was some or no clustering. Rules 4 and 5 of Section 4 tell us that if $\delta = 0$, the design already in use ($t = m = 4$) was best, but if even a slight amount of clustering is present, that design might be inefficient. Therefore we made sure that the data about cheaters was collected in such a way that δ could be estimated. Since such data would take years to amass, however, a program was also begun to collect data to allow estimates of correlation to be made for characteristics believed to be associated with misrepresented households, such as telephone ownership, income, and employment status.

After some information from these data collection efforts became available, the model was again used to aid in the selection of a reinterview design. The goal was to select a design whose loss in detection probability from that of the optimal choice would be small over the range of π and $\mathbf{p}$ we believed to be likely. A further goal was to compare this near optimal design with that of the design in use ($\ell = 4$, $t = 4$). This procedure is now described.

The data collected since 1982 gives a small amount of information about $\bar{p}$ and δ. From Section 2, recall that $\bar{p}$, the proportion of misrepresented units in a cheating interviewer's assignment, was observed to be 30% or less. Therefore, we varied $\bar{p}$ in the interval $0 < \bar{p} \leq .30$. We also found that estimates of δ for characteristics believed to be associated with cheating were generally fairly small (less than .5), but non-zero. This led us to restrict our investigation to the values $0 \leq \delta \leq .5$. Further practical considerations limited the design choice by restricting ℓ to the range $2 \leq \ell \leq 6$, since there was concern that sampling more than six (out of a possible 12) USU's in an interviewer's assignment might adversely affect the interviewer's cooperation rate in the assignment for the subsequent months of interviewing. Sampling less than two would not provide adequate work and compensation to employ a reinterviewer.

Within these constraints, a search was begun for a near-optimal design. The first choice made was that of t, for which the behavior of

$$M(\delta; t, \pi) = \max_{2 \leq \ell \leq 6} D(s, \ell, t)$$

was studied. Figure 1 illustrates with $\bar{p} = .10$ and $\pi = .1$, .5, and .9 the type of results obtained. In order to completely specify $\mathbf{p}$, p_2 and p_3 were taken to have values corresponding to completely random cheating and p_0, p_1, and p_4 were then chosen to satisfy the constraints stated above for $\bar{p}$ and δ, along with $\sum_j p_j = 1$. Except for $\pi = .9$, $t = 3$ appears to yield uniformly higher detection probabilities than $t = 4$, while $t = 4$ is preferred when π is .9. Note, however, that $M(t, \delta, \pi)$ increased less than 4 % for $t = 4$ relative to $t = 3$ in that case. When $\bar{p}$ was varied over the range (.05, .30), only the level of detection probability changed, not the choice of design parameters. Therefore, a choice of $t = 4$ was made since it was near optimal and maintained the status quo.

Since cost is fixed, only one other parameter, either s or ℓ, need be determined in order to completely specify the reinterview sample design. Furthermore, as can be noted from (3.5), the amount of clustering does not affect this choice when $t = 4$. In fact, only p_0 and π need be specified in order to completely determine the cheater behavior model when $t = 4$.

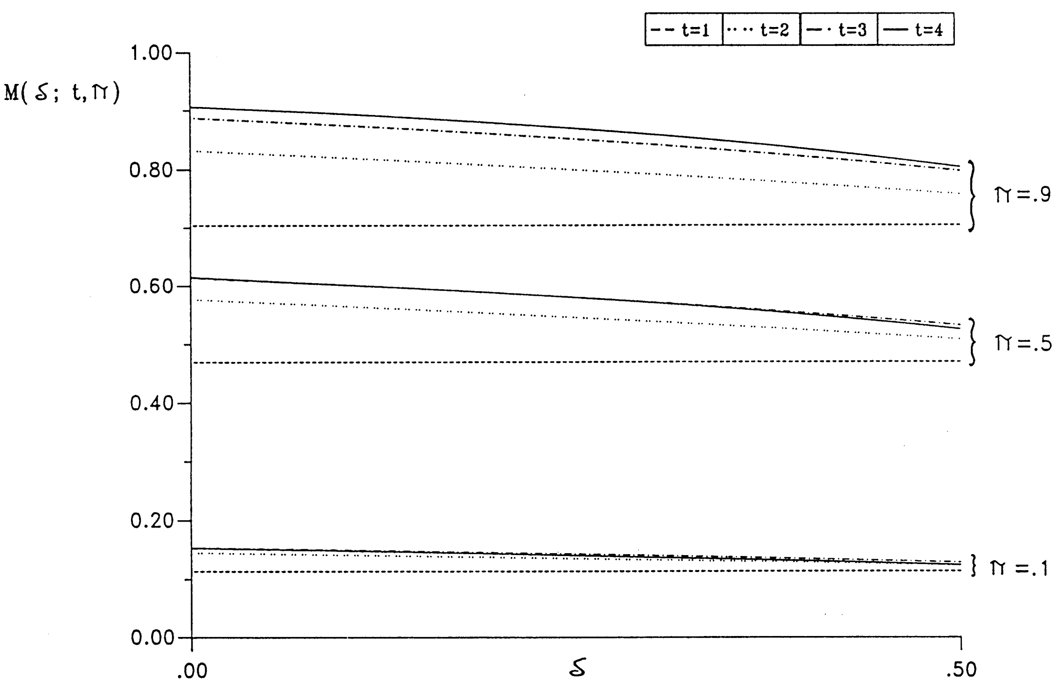

Fig.1. Maximum detection probabilities for given t as a function of δ for π = .1, .5 and .9

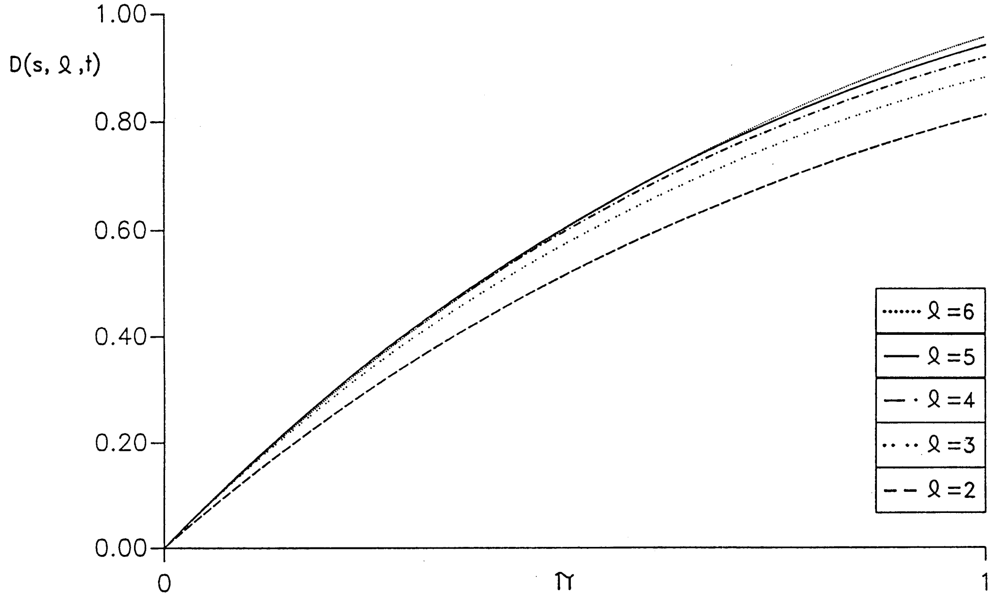

Fig. 2. Detection probabilities for given ℓ and $t = 4$ as a function of π

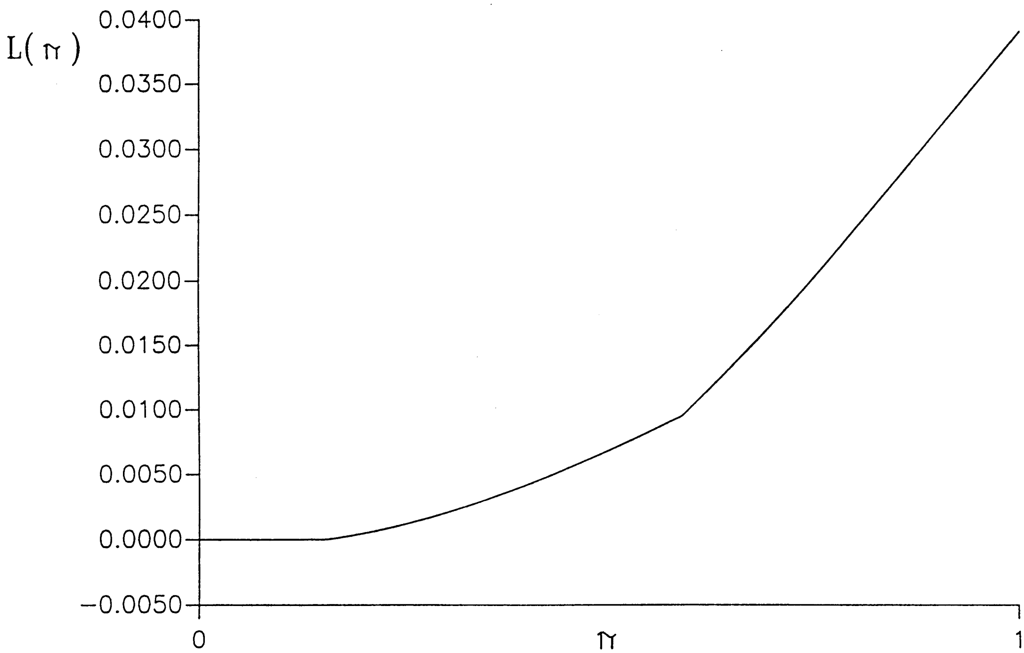

Fig. 3. Loss in detection probability for $\ell = 4$, $t = 4$ as a function of π

Figure 2 illustrates with $p_0 = .657$ the behavior of the detection probability $D(s, \ell, 4)$ as a function of π for $2 \le \ell \le 6$. ($p_0 = .657$ when $\delta = 0$ and $\bar{p} = .1$.) It suggests that for low cheating frequency ($\pi < .15$), there is little advantage in moving from the status quo value $\ell = 4$. For more consistent cheaters ($\pi > .15$), the advantage is more pronounced. When p_0 was varied over the range $.4 \le p_0 \le .8$, a similar pattern emerged.

Of particular interest is the status quo value of $\ell = 4$. Figure 3 shows the loss in detection probability for $\ell = 4$ as a function of π with $p_0 = .657$, i.e., $L(\pi) = \max_{\ell} D(s, \ell, 4) - D(s, 4, 4)$, where as before s is chosen to satisfy the cost constraint. The greatest loss in detection probability is only about 4%. Because of the advantages of maintaining the current procedure, the reinterview design chosen was $t = 4$ and $\ell = 4$. s was then chosen to satisfy the cost constraint, which resulted in a supplementary sample of $s = I/12$.

6. A Concluding Note

The investigation to date has led us to conclude that no change in the basic design of the reinterview sample (except for the introduction of a supplementary sample) is warranted at this time. The increase in detection probability which might be gained from changing ℓ or t was shown by our model to be too small to justify the added expense and disruption of data collection which would follow a change in reinterview design parameters. Data on cheating cases are still being collected, however, and better estimates of parameters of the cheating model will eventually be available. When this data are available, the optimal design can be reassessed.

In addition to reassessing the optimal choice of s, ℓ, and t for the current sampling scheme, the next redesign of the Census Bureau reinterview program should investigate alternative sampling schemes, especially unequal probability sampling designs. For example, the data in Section 2 indicate that even higher detection probabilities may be realized if the less experienced interviewers or those working in urban areas, or both, were sampled at a higher rate than other interviewers. However, two issues to consider here are: (a) the effect on respondent cooperation for those interviewers sampled more frequently and (b) the effect of the design as a deterrent for those interviewers sampled less frequently. Our current model is insufficient for evaluating these effects.

Appendix 1

Let $P_t(\mathbf{p}, \pi)$ denote the conditional probability that a $(\mathbf{p}, \pi)$ cheater is detected given that he/she is selected for verification in an interviewing period using a verification design having parameters (s, ℓ, t). Define $\boldsymbol{\ell} = [\ell_0, \ldots, \ell_m]$ to be the sample analog to $\mathbf{b}$; i.e., $\ell_h (h = 1, \ldots, m)$ is the number of sample USU's in an interviewer's assignment containing h falsified units.

We assume that $\mathbf{b}$ is distributed as a multinomial random vector with parameters n and $\mathbf{p}$ and denote this distribution by $P(\mathbf{b})$. Thus, since the ℓ USU's are sampled using simple random sampling without replacement, $P(\boldsymbol{\ell}|\mathbf{b})$, the conditional distribution of $\boldsymbol{\ell}$ given $\mathbf{b}$, is the multivariate hypergeometric distribution.

Let A_t denote the event "no falsified interview units detected after inspecting t interview units in each of the ℓ sample USU's". Define the summation Σ' as the sum over all possible $\mathbf{b}$ such that $b_h \ge 0$ and $\Sigma b_h = n$. Likewise define Σ'' to be the sum over all possible $\boldsymbol{\ell}$ such that $0 \le \ell_h \le b_h$ and $\Sigma \ell_h = \boldsymbol{\ell}$. Let B be the event "the interviewer is susceptible to cheating" and let B^c denote the complement of B. Now, $P(A_t) = (1 - \pi) + \pi P(A_t|B)$ since $P(A_t|B^c) = 1$. Further,

$$P(A_t|B) = \sum{}' \sum{}'' P(A_t|B, \mathbf{b}, \boldsymbol{\ell})\, P(\boldsymbol{\ell}|\mathbf{b}) P(\mathbf{b}) \tag{A.1}$$

where by the above assumptions,

$$P(\mathbf{b}) = \binom{n}{b_0, \ldots, b_m} p_0^{b_0} \ldots p_m^{b_m}, \tag{A.2}$$

$$P(\boldsymbol{\ell} \mid \mathbf{b}) = \frac{\binom{b_0}{\ell_0} \cdots \binom{b_m}{\ell_m}}{\binom{n}{\ell}}, \tag{A.3}$$

and

$$P(A_t|B, \mathbf{b}, \boldsymbol{\ell}) = \prod_{h=0}^{m} \left[\frac{\binom{m-h}{t}}{\binom{m}{t}} \right]^{\ell_h}. \tag{A.4}$$

Substituting these into (A.1) yields

$$P(A_t|B) = \sum{}' \sum{}'' \prod_{h=0}^{m} \left[\frac{\binom{m-h}{t}}{\binom{m}{t}} \right]^{\ell_h} \frac{\ell!}{\ell_0! \ldots \ell_m!} \frac{(n-\ell)!}{(b_0-\ell_0)! \ldots (b_m-\ell_m)!} p_0^{b_0} \ldots p_m^{b_m}. \tag{A.5}$$

Letting $a_h = b_h - \ell_h$, $h = 1, \ldots, m$, we see that the term involving the b_h under Σ' is

$$\sum{}' \frac{(n-\ell)!}{a_0! \ldots a_m!} p_0^{(a_0+\ell_0)} \cdots p_m^{(a_m+\ell_m)} = p_0^{\ell_0} \cdots p_m^{\ell_m}. \tag{A.6}$$

Thus,

$$\begin{aligned} P(A_t|B) &= \sum{}'' \prod_{h=0}^{m} \left[\frac{\binom{m-h}{t}}{\binom{m}{t}} \right]^{\ell_h} \frac{\ell!}{\ell_0! \ldots \ell_m!} p_0^{\ell_0} \cdots p_m^{\ell_m} \\ &= \sum{}'' \frac{\ell!}{\ell_0! \ldots \ell_m!} \prod_{h=0}^{m} \left[\frac{\binom{m-h}{t}}{\binom{m}{t}} p_h \right]^{\ell_h} \\ &= \sum{}'' \frac{\ell!}{\ell_0! \ldots \ell_m!} \left(p_h^*\right)^{\ell_h} \end{aligned} \tag{A.7}$$

by the Multinomial Theorem, where $p_h^* = p_h \binom{m-h}{t} \Big/ \binom{m}{t}$. Therefore,

$$\begin{aligned} P_t(\mathbf{p}, \pi) &= 1 - \{(1-\pi) + \pi P(A_t|B)\} \\ &= 1 - \left\{ (1-\pi) + \pi \left[\binom{m}{t}^{-1} \sum_h \binom{m-h}{t} p_h \right]^{\ell} \right\}. \end{aligned} \tag{A.8}$$

Finally, the probability of not being detected for an entire evaluation period is the probability of not being detected in the predesignated sample times the probability of not being detected in any supplementary sample. The former probability is simply $1 - P_t(\mathbf{p}, \pi)$. The latter probability is

$$\sum_{k=0}^{f-1} \Pr \text{ (interviewer is selected } k \text{ times and not detected each time)}$$

$$= \sum_{k=0}^{f-1} \binom{f-1}{k} \left(\frac{fs}{I(f\text{-}1)}\right)^k \left(1 - \frac{fs}{I(f\text{-}1)}\right)^{f-k-1} [1 - P_t(\mathbf{p}, \pi)]^k$$

$$= \left[1 - \frac{fs}{I(f-1)} P_t(\mathbf{p}, \pi)\right]^{f-1}.$$

Thus,

$$D(s, \ell, t) = 1 - \Pr \text{ (not being detected for an entire evaluation period)}$$

$$= 1 - [1 - P_t(\mathbf{p}, \pi)]\left[1 - \frac{fs}{I(f-1)} P_t(\mathbf{p}, \pi)\right]^{f-1}.$$

Appendix 2

Theorem 1. If t and K_1 (s, ℓ, t), defined in (3.7), are held fixed and $\pi = 1$, then $D(s, \ell, t)$ is maximized by either $s = 0$ or $s = \frac{f-1}{f} I$, where its values are equal.

Proof: Let us hold $K_1(s, \ell, t)$ fixed at C_F, i.e.,

$$\gamma(I + fs)\ell t = C_F \tag{B.1}$$

for some constant γ. Then it can be easily shown by substitution into (3.3) and (3.4) that $D(0, \ell, t) = D\left(\frac{f-1}{f} I, \ell, t\right) = 1 - \eta_t^{C_F/\gamma It}$.

Next, to establish that the maximum value of $D(s, \ell, t)$ over the range $0 \le s \le \frac{f-1}{f} I$ is $1 - \eta_t^{C_F/\gamma It}$, we will show that

$$\frac{1 - D(s, \ell, t)}{\eta_t^{C_F/\gamma It}} \ge 1 \tag{B.2}$$

for any fixed t, where $\ell = C_F/\gamma t(I + fs)$ from (A.1). To simplify notation, we write $r = s/[(f-1)/f]I$, so that $0 \le r \le 1$, and $g = C_F/\gamma It[1 + r(f-1)]$. Then the left hand side of (A.2) may be written

$$\left[\frac{1 - r(1 - \eta_t^g)}{\eta_t^{gr}}\right]^{f-1} = \left[(1 - r)\eta_t^{-gr} + r\eta_t^{g(1-r)}\right]^{f-1}. \tag{B.3}$$

Now consider the random variable X having probability function

$$h(x) = \begin{cases} r & \text{for } x = 1 - r \\ 1 - r & \text{for } x = -r \end{cases}$$

and the function $f(x) = \eta_t^{gx}$. Since $f(x)$ is convex, we know by Jensen's inequality that $E(f(X)) \ge f(E(X))$. Since $E(f(x))$ is given by the expression inside the brackets in (A.3) and since $E(X) = 0$, (A.2) is established.

Theorem 2. If t and $K_2(s, \ell, t)$, defined in (3.8), are held fixed and $\pi = 1$, then $D(0, \ell, t) \geq D(\frac{f-1}{f}I, \ell, t)$.

Proof: Define $h_0(\ell) = K_2(0, \ell, t) = I(C_2\ell t + C_3\sqrt{\ell} + C_4)$ (B.4)

$$\text{and } h_1(\ell) = K_2\left(\frac{f-1}{f}I, \ell, t\right) = f\,I\,(C_2\ell t + C_3\sqrt{\ell} + C_4). \quad \text{(B.5)}$$

Now let ℓ_0 be such that $h_0(\ell_0) = C_F$ and ℓ_1 be such that $h_1(\ell_1) = C_F$. Then

$$\begin{aligned} h_1(\ell_1) = C_F &= fI(C_2\ell_1 t + C_3\sqrt{\ell_1} + C_4) \\ &= I(fC_2\ell_1 t + fC_3\sqrt{\ell_1} + fC_4) \\ &> I(C_2 f\ell_1 t + C_3\sqrt{f\ell_1} + C_4) \text{ since } f > 1 \text{ and } C_3, C_4 > 0 \\ &= h_0(f\ell_1). \end{aligned}$$

Since $h_0(f\ell_1) < C_F$, $h_0(\ell_0) = C_F$, and $h_0'(\ell) > 0$, we know $f\ell_1 < \ell_0$. Thus by substitution into (3.3) and (3.4) we have

$$D(0, \ell_0, t) - D\left(\frac{f-1}{f}I, \ell_1, t\right) = \eta_t^{f\ell_1} - \eta_t^{\ell_0} > 0,$$

and the theorem is established.

7. References

Bennett, A. (1948a): Survey on Problems of Interviewer Cheating: Observations on the So-Called Cheater Problem Among Field Interviewers. International Journal of Opinion and Attitude Research, 2, pp. 89–96.

Bennett, A. (1948b): Toward a Solution of the "Cheater Problem" Among Part-Time Research Investigators. Journal of Marketing, 2, pp. 470–474.

Biemer, P., Judkins, D., Schreiner, I., and Stokes, S. L. (1982): Reinterview Redesign Report 1: Optimal Sampling Strategy for Interviewer Control. Bureau of the Census Technical Report, U. S. Bureau of the Census, Washington, D. C.

Boyd, H. and Westfall, R. (1955): Interviewers as a Source of Error in Surveys. Journal of Marketing, 19, pp. 311–324.

Crespi, L. P. (1945): The Cheater Problem in Polling. Public Opinion Quarterly, Winter, pp. 431–445.

Evans, F. B. (1961): On Interviewer Cheating. Public Opinion Quarterly, 25, pp. 126–127.

Pielou, E. C. (1969): An Introduction to Mathematical Ecology. John Wiley and Sons, Inc., New York.

Sheatsley, P. B. (1951): An Analysis of Interviewer Characteristics and Their Relationship to Performance, Part II. International Journal of Opinion and Attitude Research, 5, pp. 79–94.

U. S. Bureau of the Census (1986): Memorandum for General Distribution from I. Schreiner and K. Guerra, Subject "Falsification Study" (September 1982 through August, 1985).

Received March 1988
Revised March 1989