A response surface approach to assessing the relative efficiency of adaptive cluster sampling
by Philip James Turk

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics
Montana State University

Abstract:
In this paper, we provide motivation and background behind adaptive cluster sampling (ACS). Several initial sampling designs are discussed with respect to theory and comparison to conventional sampling. A review of the literature for ACS is given for the years 1990 through 2002. Factors and issues determining the relative efficiency of ACS are discussed. We discuss further developments including initial unequal probability sampling, other types of estimators, detectability, bootstrapping, and limiting sampling effort and cost.

Using a designed experiment, multiple factors known to influence the efficiency of ACS were studied via response surface methodology where data were generated through computer simulation. This approach allowed us to characterize significant interaction and quadratic effects as opposed to ignoring them as has been done in the literature. Using simple random sampling without replacement as a comparison, two responses were considered: the relative efficiency for the Horvitz-Thompson estimator of T (REHT) and the relative efficiency for the Hansen-Hurwitz estimator of T (REHH). Response surface plots at given conditions and grid searches were used to find factorial settings yielding variance-optimal relative efficiencies. Conditions that optimize REHT were harder to characterize than for REHH. Different simulation approaches from the literature were compared with ours. A concern regarding any investigation and the validity of conclusions regarding important factors is the dependence on the simulation approach used to generate the study area population. The two estimation procedures were compared both through the response surfaces and through distributions of data gathered under varying population settings. Horvitz-Thompson estimation was relatively more efficient than Hansen-Hurwitz estimation. However, the variability of REHT was greater than the variability of REHH, most notably due to extremely low Var [THT} in some populations that were neither patchy nor rare. This low variance came at the price of a large final sample size. Finally, other factors and issues are presented with respect to ideas towards future research.

A RESPONSE SURFACE APPROACH TO ASSESSING THE RELATIVE EFFICIENCY OF ADAPTIVE CLUSTER SAMPLING

by

Philip James Turk

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2003

APPROVAL

of a dissertation submitted by

Philip James Turk

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. John Borkowski John J Borkowski (Signature) 4/18/03 Date

Approved for the Department of Statistics

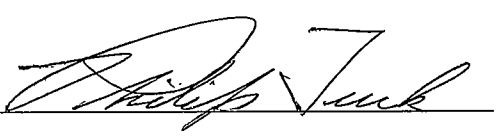
Dr. Ken Bowers Kenneth J. Bowers (Signature) 4/18/03 Date

Approved for the College of Graduate Studies

Dr. Bruce McLeod Bruce R. McLeod (Signature) 4-21-03 Date

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature 

Date 4/18/03

ACKNOWLEDGEMENTS

The preparation of a thesis is a long, grueling endeavor which is facilitated by the help of others. I would like to thank my advisor, Dr. John Borkowski, for his patient wisdom in being able to know when to offer guidance and when to let me explore various frontiers on my own. Dr. Robert Boik and Dr. Bill Quimby were readers, a tedious task for which I am appreciative. Thanks go to Dr. Steve Cherry, Dr. Warren Esty, and Dr. Jodi Carson for serving on my committee. Finally, I would be remiss if I did not give thanks for my "support system"; my friends who never let me give up.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In this paper, we provide motivation and background behind adaptive cluster sampling (ACS). Several initial sampling designs are discussed with respect to theory and comparison to conventional sampling. A review of the literature for ACS is given for the years 1990 through 2002. Factors and issues determining the relative efficiency of ACS are discussed. We discuss further developments including initial unequal probability sampling, other types of estimators, detectability, bootstrapping, and limiting sampling effort and cost.

Using a designed experiment, multiple factors known to influence the efficiency of ACS were studied via response surface methodology where data were generated through computer simulation. This approach allowed us to characterize significant interaction and quadratic effects as opposed to ignoring them as has been done in the literature. Using simple random sampling without replacement as a comparison, two responses were considered: the relative efficiency for the Horvitz-Thompson estimator of τ (RE_{HT}) and the relative efficiency for the Hansen-Hurwitz estimator of τ (RE_{HH}). Response surface plots at given conditions and grid searches were used to find factorial settings yielding variance-optimal relative efficiencies. Conditions that optimize RE_{HT} were harder to characterize than for RE_{HH}. Different simulation approaches from the literature were compared with ours. A concern regarding any investigation and the validity of conclusions regarding important factors is the dependence on the simulation approach used to generate the study area population. The two estimation procedures were compared both through the response surfaces and through distributions of data gathered under varying population settings. Horvitz-Thompson estimation was relatively more efficient than Hansen-Hurwitz estimation. However, the variability of RE_{HT} was greater than the variability of RE_{HH}, most notably due to extremely low $Var[\hat{\tau}_{HT}]$ in some populations that were neither patchy nor rare. This low variance came at the price of a large final sample size. Finally, other factors and issues are presented with respect to ideas towards future research.

CHAPTER 1

INTRODUCTION

Motivation and Background

Adaptive sampling is a type of sampling scheme in which the procedure for selecting units to include in the sample may depend on values of the variable of interest observed during the survey, i.e. the sampling is "adapted" to the data (Thompson and Seber 1996). A specific type of adaptive scheme is adaptive cluster sampling (ACS). ACS operates under the rule that when the observed value of a randomly selected unit satisfies some condition of interest, C (for example $y_i \geq c_a$, where c_a is a fixed *critical value*), other additional units in some pre-defined accompanying neighborhood are added to the sample. In turn, if these additional units satisfy C, then their unit neighborhoods are added to the sample as well, and so on. This process stops when no further units satisfying C are encountered. "In general terms, it means that if you find what you are looking for at a particular location you sample in the vicinity of that location with the hope of obtaining even more information." (Salehi and Seber 1997a, p. 959).

An ACS design is distinguished from the classical probability based sampling designs in that the sample selection depends on the observed unit values of the variable of interest. Historically, the ACS designs arose from informative designs (Cassel,

Särndal, and Wretman 1977), sequential statistical methods, network sampling and snowball sampling and this history, along with references, is well summarized in Thompson (1992). The most important application for ACS occurs for those populations where the number of individuals is small and the individuals show a tendency to cluster. In these cases, were one to use classical designs, the majority of measurements will be 0 and many clusters will be missed. Consequently, variances associated with estimation of the population mean or total will be large and resulting confidence intervals for corresponding parameters will be wide (Seber and Thompson 1994). Fields of study where ACS has been effectively and successfully used and where such populations naturally occur include, but are not confined to, ecology, biology, epidemiology, environmental sciences, demography and geology (Brown 1994; Dryver and Thompson 1998; Seber and Thompson 1994; Thompson 1990; Thompson and Seber 1996). For example, many populations of animals and plants have aggregation tendencies due to such factors as schooling, flocking, dispersal patterns, and environmental patchiness (Thompson 1991a).

There are numerous differences between ACS and the classical designs. With the classical designs, the entire selection of units in the sample is made prior to making any observations while with ACS the selection of units in the sample depends on observed values of the variable of interest. Let the sample s be a set or sequence of labels identifying the units for observation and $P(s)$ be the sample's selection probability. For many classical designs, $P(s)$ is nonzero and constant, i.e. it does not depend on the

data. On the other hand, for ACS, $P(s)$ is sometimes zero or nonconstant, i.e. it does depend on the data. Moreover, ACS assigns higher sample selection probabilities to samples that include units where $y_i \geq c_a$. A similar situation exists regarding the unit marginal inclusion probabilities, or π_is. Many classical plans are self-weighting i.e. $\pi_i = \pi \ \forall \ i = 1, 2, \ldots, N$ units. However, for ACS, this is no longer the case because the $\pi_i s$ depend on the data (through C) and are therefore no longer constant (Seber and Thompson 1994). Another important difference is that classical estimators used in conjunction with ACS are biased. However, modified estimators exist for ACS that are design-unbiased for the population mean along with design-unbiased estimators of their variances. That is, the unbiasedness depends on the way the sample is selected and does not depend on any assumptions about the population. In other words, if t is the estimator and t_s is the estimate computed when sample s is selected, then we have:

$$E[t] = \sum_s t_s P(s) = \mu$$

where the summation is taken over all possible samples. Also, the Rao-Blackwell method has been used to obtain smaller variance design-unbiased estimators for ACS. It should be noted that although design-unbiased estimators are the ones generally used when conducting ACS, estimators based on models are used on occasion. Frequentist model-based and likelihood model-based estimation along with stochastic population theory is thoroughly discussed in Thompson and Seber (1996). Only the

design-based approach is discussed in this dissertation and the reader should bear this in mind when terms like "unbiased", for example, are used.

There are numerous advantages that ACS can and/or does have over the classical designs. In many instances ACS can be a more efficient design, i.e. the variances of estimators are smaller for an equivalent amount of sampling effort. Estimators derived via ACS make use of unit labels which can result in lower variance than estimators that do not make use of labels (Seber and Thompson 1994; Thompson and Seber 1996). Indeed, virtually every paper cited in this dissertation deals to some extent with the conditions under which ACS is a more efficient design for an equivalent amount of sampling effort. Because the location and shape of clusters of individuals is oftentimes not known prior to the survey, ACS can be used in situations where stratification may not be possible (Thompson 1991b). ACS also can find local maxima or areas of high abundance, thus increasing the yield of the sample. This attribute can be of particular importance, for example, when sampling rare or endangered species where gathering as much information as possible about individuals is desired (Seber and Thompson 1994). Another major advantage of the ACS design is its flexibility in terms of its construction. One can modify the initial sample size, the unit size, the configuration of the neighborhood, and the sampling condition C relative to the situation to achieve the most efficient design. Furthermore a variety of methods for choosing the initial sample exist, such as simple random sampling

with replacement (SRSWR) and simple random sampling without replacement (SRSWOR) (Thompson 1990), strip sampling, systematic sampling (Thompson 1991a), stratified sampling (Thompson 1991b), sampling via probability proportional to size (Pontius 1997; Roesch 1993; Smith, Conroy, and Brakhage 1995), and simple latin square sampling (Borkowski 1999). Finally, two other potential advantages from a cost standpoint are that the average distances between sampled units are lessened and the quadrat locations are easier to find (Brown and Manly 1998; Salehi and Seber 1997a).

There are, however, some disadvantages to ACS relative to the classical designs. The final number of distinct sampled units and both initial and final number of distinct sampled networks are random. Consequently, it can be difficult to determine and control the total sampling effort and cost of the survey in advance, which is necessary for some surveys. Several papers in recent years have dealt with this issue and they are discussed in the next chapter (Brown and Manly 1998; Christman and Lan 1998; Felix-Medina and Thompson 1999; Salehi and Seber 1997a,b). Work has been done in developing the theory for using a pilot study to design an ACS survey with a given efficiency or expected cost (Salehi and Seber 1997b). Ironically, it is the very flexibility of ACS that makes finding variance optimal designs complicated. For example, the selection of C can be critical to the efficiency of the ACS design. As with the classical designs, there exists a positive probability that a sample will yield a zero abundance estimate with this issue having been addressed by Christman and

Lan (1998). Another important distinction between the two types of designs is that not all the information from sampled units is used in ACS. Specifically, *edge units*, or those sampled units that do not satisfy C but are on the boundary of a unit that does, are used only if they are part of the initial sample. This has been addressed by Dryver and Thompson (1998). Finally, the modified ACS estimators are not optimal (from a variance standpoint) among the set of possible estimators that use unit labels.

Fixed-Population Sampling Theory and Notation

We will now consider the setting of sampling from a fixed, finite population. Most of the following discussion, in terms of theory and notation, is attributed to Thompson and Seber (1996).

Consider a region divided into N units $(u_1, u_2, \ldots, u_N)$ of equal area indexed by their respective labels $(1, 2, \ldots, N)$. Assume that the units form a partition of the region. Here, a label identifies the location of the respective unit. Unit i will be associated with a response of interest, $y_i, i = 1, 2, \ldots, N$. Thus, we have a vector of population y-values , $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$. Moreover, in the fixed population view, the population (or study population) is composed of $\mathbf{y}$ where $\mathbf{y}$ is considered to be a fixed set of unknown constants. We may write $\boldsymbol{\theta}= \mathbf{y}$ to represent the values as a vector of unknown parameters.

A selection-ordered sample of size n is a sequence $s_0 = (i_1, i_2, \ldots, i_n)$ of n of the unit labels where the set of all possible samples is denoted by $\mathcal{S}$. The objective under

this setting is to select a sample, observe the y-values for the units in the sample, and estimate some function $z(\boldsymbol{\theta})$. Typical examples of population quantities to be estimated include the population total $\phi(\boldsymbol{\theta}) = \sum_{i=1}^{N} y_i = \tau$, the population mean $\phi(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} y_i = \mu$, and the finite-population variance $\phi(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i-\mu)^2}{N-1} = \sigma^2$.

The *data* d_0 consists of the y-values for the units in the sample, together with the associated unit labels, namely the selection-ordered pairs. Thus, $d_0 = (s_0, \mathbf{y}_0)$, where $\mathbf{y}_0$ is the set of selection-ordered sample y-values; that is, $\mathbf{y}_0 = (y_i : i \in s_0)$. Furthermore, we will also be interested in s, consisting of the reduced set of ν *distinct* labels in s_0 sequentially-ordered from the smallest to largest label for uniqueness. Additionally, $\mathbf{y}_s$ is the corresponding set of y-values in the same order as s. Also by definition, $d_s = (s, \mathbf{y}_s)$. It is clear that given s_0 we need only $\mathbf{y}_s$ to provide all the information about d_0. Thus, in some cases it will be convenient to redefine d_0 to be equal to $(s_0, \mathbf{y}_s)$ since both representations contain exactly the same information.

The sampling design is the procedure by which the sample is selected and is the sample selection probability function $P_\theta(s_0)$ where the notation emphasizes the dependence of the procedure on $\boldsymbol{\theta}$. The sampling design also may be specified as the conditional probability $P(s_0|\mathbf{y})$ of selecting the sample s_0 such that $P(s_0|\mathbf{y}) \geq 0$ and $\sum_{s_0 \in \mathcal{S}} P(s_0|\mathbf{y}) = 1$ for all for all $s_0 \in \mathcal{S}$. In our case, we are only interested in adaptive designs, i.e. those designs such that $P_\theta(s_0) = P(s_0|\mathbf{y}) = P(s_0|\mathbf{y}_s)$. In other words, the procedure for selecting units depends on values of the variable of interest, but only through units included in the sample.

Also it is helpful to consider the *unordered* (with respect to both sequence and selection), reduced set $s_R = i_1, i_2, \ldots, i_\nu$ of ν distinct labels in the sample. Additionally, $\mathbf{y}_R$ is the corresponding set of y-values and, by definition, $d_R = (s_R, \mathbf{y}_R)$. Note that the *reduction* from d_0 to either d_s or d_R is achieved by ignoring information about order and the multiplicities of the units in the sample.

Let $\boldsymbol{\theta} \in \Theta$, where the parameter space Θ is a subset of $\Re^N$. A parameter vector $\boldsymbol{\theta}$ is said to be *consistent* or *compatible* with datum d_s (or, equivalently, with datum d_R) if the ith component of $\mathbf{y}_s$ (or, equivalently, of $\mathbf{y}_R$), equals the s_ith component of $\boldsymbol{\theta}$ for all units i in the sample. That is, $\boldsymbol{\theta}$ and d_s (or d_R) are consistent if the y and θ values coincide for all sample units, i.e. $\boldsymbol{\theta}$ gives rise through the design to the given d_s. Consistency depends only on the values of the distinct units in the sample, not on the order or multiplicity of selection. Therefore, if $\boldsymbol{\theta}$ is consistent with the reduced data d_s (or with d_R) it also will be consistent with the original data d_0. We associate with every d_s a subset Θ_{d_s} of Θ whose vectors are consistent with d_s i.e. Θ_{d_s} is the set of $\boldsymbol{\theta}$ that give rise through the design to the given s. Since consistency depends only on the values of the distinct units in the sample, consistency with d_0 is equivalent to consistency with the reduced data d_s (and d_R), and $\Theta_{d_0} = \Theta_{d_s} = \Theta_{d_R}$.

An example demonstrating the use of all the aforementioned notation can be found in Thompson and Seber (1996, p. 33-34).

We note for ACS that although the final sample size is random, the probability of getting the final sample is the same as that for the initial sample since the augmented

or adapted part of the final sample is not randomly selected and is chosen through C.

Previously, it was stated that $P_\theta(s_0)$ depends on $\boldsymbol{\theta}$. To clarify this statement, we note that on the set Θ_{d_R}, $P_\theta(s_0)$ is constant and positive for a given d_0, but 0 otherwise. Thus, $P_\theta(s_0)$ does not truely directly depend on $\boldsymbol{\theta}$. Rather, it depends on $\boldsymbol{\theta}$ only through Θ_{d_R}. As an example for an adaptive design, suppose we use simple random sampling and continue sampling until a y-value exceeds some prechosen level c. The probability of obtaining a particular sample will depend on $\boldsymbol{\theta}$, specifically the y-values in the sample. Suppose that just the first unit in the population has a y-value exceeding c. Therefore, $P(s_0|\mathbf{y}_s) = 1/N$ for $s_0 = (1)$. However, if $\boldsymbol{\theta}$ was such that only the second unit in the population had a y-value exceeding c, then $P(s_0|\mathbf{y}_s) = 0$ for $s_0 = (1)$ as $\boldsymbol{\theta}$ would no longer be compatible with the sample consisting of just the first unit. Hence for each d_0 there exists a Θ_{d_R} such that $P(s_0|\mathbf{y}_s)$ does not depend on $\boldsymbol{\theta}$ for $\boldsymbol{\theta} \in \Theta_{d_R}$ and $P(s_0|\mathbf{y}_s) = 0$ for $\boldsymbol{\theta} \notin \Theta_{d_R}$. Since $P(s_0|\mathbf{y}_s)$ does not depend on unobserved components of $\boldsymbol{\theta}$, it is constant as a function of $\boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \Theta_{d_R}$.

With adaptive designs, the probability of obtaining a specific datum $d_0 = (s_0, \mathbf{y}_s)$ is $P_\theta(D_0 = d_0)$ where:

$$P_\theta(D_0 = d_0) = P(s_0|\mathbf{y}_s) I_{\Theta_{d_R}}(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta \tag{1.1}$$

where $I_{\Theta_{d_R}}(\boldsymbol{\theta})$ is an indicator function taking the value 1 whenever $\boldsymbol{\theta} \in \Theta_{d_R}$ and 0 otherwise. We can also view this probability distribution as the likelihood function of the parameter $\boldsymbol{\theta}$ i.e. $L(\boldsymbol{\theta}; d_0) = P_\theta(D_0 = d_0)$. Note that given the value of d_0,

$P(s_0|\mathbf{y}_s)$ is fixed and the likelihood takes only two values. Thus, the likelihood is maximized when $\boldsymbol{\theta} \in \Theta_{d_R}$ so that any value of $\boldsymbol{\theta}$ consistent with the data will be a maximum likelihood estimator.

We now state the following theorems.

THEOREM 1.1. *For an adaptive design satisfying equation 1.1, D_R is a minimally sufficient statistic for $\boldsymbol{\theta}$.*

THEOREM 1.2. *Let $T(D_0)$ be any estimator of $\phi(\boldsymbol{\theta})$ that is not a function of D_R. Define $T_{RB} = E[T|D_R]$. Then:*

1. *$T_{RB}(D_R)$ is an estimator of $\phi(\boldsymbol{\theta})$.*

2. *$E[T_{RB}] = E[T]$*

3. *$MSE[T_{RB}] \leq MSE[T]$ with strict inequality for all $\boldsymbol{\theta} \in \Theta$ such that $P_\theta(T \neq T_{RB}) > 0$.*

The reader is refered to Thompson and Seber (1996) for the proofs of these two theorems.

A corollary to the preceding theorem is:

COROLLARY 1.3. *If T is unbiased for $\phi(\boldsymbol{\theta})$, then $Var[T_{RB}] \leq Var[T]$ with strict inequality for all $\boldsymbol{\theta} \in \Theta$ such that $P_\theta(T \neq T_{RB}) > 0$.*

The reader will recognize the preceding corollary as being the Rao-Blackwell theorem (Blackwell 1947; Rao 1945).

Several Rao-Blackwell estimators have been proposed over recent years (Dryver and Thompson 1998; Salehi 1999; Salehi and Seber 1997a; Thompson and Seber 1996). Unfortunately, none of these estimators are uniformly better than the others with respect to their variances. The following theorem helps explain why.

THEOREM 1.4. *The statistic D_R is not complete.*

The reader is referred to Cassel et al. (1977) for a proof.

If we consider any sampling designs in the fixed-population setting (adaptive and conventional), an UMVUE does not exist because it would have to have zero variance with this being due to the fact that D_R is not complete (Lehmann 1983). The incompleteness of D_R in the finite population sampling situation can be attributed to the presence of the unit labels (Thompson and Seber 1996). Lehmann (1983) showed that if the data do not include labels, then the order statistics are complete and so an UMVUE exists for those sampling schemes where the probabilities $P_\theta(s_R)$ are known and positive for all s_R. However, deliberately ignoring the labels is not advised and has been a topic of controversy (Cassel et al. 1977; Chaudhuri and Vos 1988). For the purposes of this dissertation, this controversy is a moot point for two reasons. First, the only estimators we will be focusing on are label dependent. In other words, these estimators depend on specific knowledge of which labels are sampled because these

unit labels make it possible to determine which survey units to adaptively sample, i.e. where the survey units are, and how to calculate the estimates. Second, the estimators from the ACS designs that are label dependent generally perform much better than those estimators that do not use labels, the presence of a UMVUE within the latter class of estimators not withstanding (Thompson and Seber 1996).

ACS With an Initial SRSWOR - Theory

Consider a population consisting of N rectangularly-shaped units that can be arranged in a grid format. Assume an initial simple random sample of size n_1 is taken without replacement from the population. In the Motivation and Background section of this chapter, the underlying process of ACS is described. Recall that the process stops when no further units satisfying C are encountered. At this point, a *cluster* of units is obtained that contains a boundary of units called *edge* units, where an edge unit is any unit not satisfying C but in the neighborhood of one that does (Thompson 1990). The final sample then consists of n_1 clusters. These clusters are not necessarily distinct since two non-edge units in the same cluster could have been selected in the initial sample. If a unit selected in the initial sample does not satisfy C, then it is trivially considered to be a cluster of size 1.

Neighborhoods can be defined in a variety of ways although the first-order neighborhood consisting of the unit itself and the four adjacent units sharing a common boundary (to the north, south, east and west) is by far the most prevalent. This

form of neighborhood is appropriate for studies in which the y-values that satisfy C tend to cluster and are not oriented in any particular direction (Christman 2000). Technically, the units in the neighborhood don't have to be physically contiguous. For example, the neighborhood may be defined through a social relationship between units (Dryver and Thompson 1998). However, the neighborhood relationship is assumed to be symmetric. Also, the neighborhoods do not depend on $\boldsymbol{\theta}$.

The *network* A_i for unit i is defined to be the cluster generated by unit i but with its edge units removed. A selection of any unit in A_i leads to the selection of all of A_i, i.e. the network is symmetric. By definition, any unit that does not satisfy C is trivially considered to be a network of size 1, i.e. is a *trivial network*. Thus, any cluster of unit size greater than 1 can be decomposed into a network whose units satisfy C and further networks (edge units) of size 1 that do not satisfy C. It is important to note the distinction between clusters and networks. Clusters are not necessarily disjoint since they may have overlapping edge units. On the other hand, the entire population of N units can be partitioned into a set of disjoint and exhaustive networks.

Once again, note that if the y-value of a sampled unit satisfies a certain condition C, for example $y_i \geq c_a$, then the rest of the unit's neighborhood is added to the sample. Therefore, unit i will be included in the final sample either if any unit of A_i is selected as part of the initial sample or if any unit of a network for which unit i is

an edge unit is selected. Thus, we have the marginal inclusion probability for unit i:

$$\pi_i = 1 - \left[\binom{N - m_i - a_i}{n_1} \Big/ \binom{N}{n_1} \right] \tag{1.2}$$

where m_i denotes the number of units in A_i, and a_i denotes the number of units in networks of which unit i is an edge unit. If unit i satisfies C, then $a_i = 0$, while if unit i does not satisfy C, then $m_i = 1$.

The problem immediately becomes apparent that a_i is unknown for a unit i not satisfying C. Thus, π_i, the probability that unit i is included in the sample, cannot be computed from the sample data. Instead, we consider the empirically derived "partial" marginal inclusion probability:

$$\pi_i' = 1 - \left[\binom{N - m_i}{n_1} \Big/ \binom{N}{n_1} \right] \tag{1.3}$$

Consequently, we are now dealing with a sample of n_1 networks, not necessarily distinct. We can interpret π_i' as being the probability that at least one unit in the initial sample intersects network A_i. Later we will see that the estimators used in ACS make use of observations not satisfying C only when they are included in the initial sample. Consequently, π_i' can also be interpreted as being the probability that unit i is used in the estimator.

In 1990, Thompson (1990) developed an estimator based on a modification of the estimator of Horvitz and Thompson (1952) using the partial inclusion probabilities:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i I_i}{\pi_i'} \tag{1.4}$$

where I_i takes on the value 1 with probability π_i' if the initial sample intersects A_i, and 0 otherwise. It can be used when sampling is either WR or WOR.

Importantly, note that observations that do not satisfy C are ignored if they are not included in the initial sample.

Another way of expressing equation 1.4 is:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* I_k}{\alpha_k} \tag{1.5}$$

where y_k^* is the sum of the y-values for the kth network, K is the total number of distinct networks in the population and I_k takes a value 1 with probability α_k if the initial sample intersects the kth network , and 0 otherwise. For each unit i in the kth network, π_i' will be the same, i.e. equal to α_k. If there are x_k units in the kth network, then:

$$\alpha_k = 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1} \right] \tag{1.6}$$

It should be emphasized, for the same reason as previously mentioned, that α_k is not the actual network inclusion probability. Rather, α_k is referred to as the *marginal initial intersection probability*, or the probability that at least one unit in the initial sample intersects A_k.

If there are x_k units in the kth network and x_j units in the jth network, then the probability (or *joint initial intersection probability*) that networks j and k are both intersected by at least one unit each in the initial sample is:

$$\alpha_{jk} = 1 - \left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k}{n_1} \right] \Big/ \binom{N}{n_1} \tag{1.7}$$

The following results were derived by Thompson (1990).

$$E[\hat{\mu}_{HT}] = \mu \tag{1.8}$$

$$Var[\hat{\mu}_{HT}] = \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k}\right)\right] \tag{1.9}$$

with unbiased estimator of the variance:

$$\widehat{Var}[\hat{\mu}_{HT}] = \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_{jk}\alpha_j \alpha_k}\right) I_j I_k\right] \tag{1.10}$$

where, by definition, $\alpha_{jj} = \alpha_j$.

It should be noted that $\hat{\mu}_{HT}$ will have low variance if $\alpha_k \propto y_k^*$. In this case, $\hat{\mu}_{HT}$ will be relatively constant and will therefore have little variability (Thompson and Seber 1996). Also note that α_{jk} must be greater than 0 for all pairs of networks $j \neq k$.

Thompson (1990) also developed an estimator based on a modification of the estimator of Hansen and Hurwitz (1943), which can be used when sampling is either WR or WOR, namely:

$$\hat{\mu}_{HH} = \frac{1}{n_1}\sum_{i=1}^{N} \frac{y_i f_i}{m_i} \tag{1.11}$$

where f_i is defined to be the number of times that the ith unit in the sample appears in the estimator. The modification is necessary because the draw-by-draw selection probabilities, $p_i = (m_i + a_i)/N$, will generally not be known for every unit in the sample. As was the case with $\hat{\mu}_{HT}$, observations that do not satisfy C are ignored if they are not included in the initial sample. Another expression of $\hat{\mu}_{HH}$ written in

terms of the n_1 networks (not necessarily unique) intersected by the initial sample is:

$$\hat{\mu}_{HH} = \frac{1}{n_1}\sum_{i=1}^{n_1} w_i = \bar{w} \tag{1.12}$$

where $w_i = \frac{1}{m_i}\sum_{j\in A_i}(y_j)$, is the mean of the m_i observations in A_i.

The following results were derived by Thompson (1990).

$$E[\hat{\mu}_{HH}] = \mu \tag{1.13}$$

$$Var[\hat{\mu}_{HH}] = \frac{N-n_1}{Nn_1(N-1)}\sum_{i=1}^{N}(w_i-\mu)^2 \tag{1.14}$$

An alternative expression is:

$$Var[\hat{\mu}_{HH}] = \frac{N-n_1}{Nn_1}\sigma_B^2 \tag{1.15}$$

where σ_B^2 is the between-network variance.

An unbiased estimator of equation 1.14 is:

$$\widehat{Var}[\hat{\mu}_{HH}] = \frac{N-n_1}{Nn_1(n_1-1)}\sum_{i=1}^{n_1}(w_i-\hat{\mu}_{HH})^2 \tag{1.16}$$

ACS With an Initial SRSWR - Theory

All the preceding estimators from the previous section can be used with only minor modifications when the initial random sample is conducted with replacement. Specifically, both the marginal and joint initial intersection probabilities will be modified for Horvitz-Thompson estimation and $\widehat{Var}[\hat{\mu}_{HH}]$ will be modified for Hansen-Hurwitz estimation. The reader is referred to Thompson and Seber (1996, p. 100) for further details.

ACS Versus Conventional Sampling - Simple Random Sampling

The sample mean of the *final* ACS sample, $\bar{y}$, as well as the average of the *final* ACS sample cluster means, are both biased estimators of μ since each unit has a different selection probability (Thompson 1990). Further, the sample variance of the *final* ACS sample, is biased for σ^2. Of course, the *initial* sample mean selected by simple random sampling, $\bar{y}_1$, selected with or without replacement, is an unbiased estimator of μ. Note that $\bar{y}_1$ also does not make use of the unit labels.

Thompson (1990) demonstrated, via a specific example with an initial sample of fixed size n_1 selected via SRSWOR from a small population, that an ACS strategy can be more efficient than a simple random sampling strategy. In other words, the variance of the estimators $\hat{\mu}_{HT}$ in equation 1.9 and $\hat{\mu}_{HH}$ in equation 1.14 can both be less than the variance of $\bar{y}$ with a SRSWOR design. For comparative purposes, the expected effective final sample size $E[\nu]$, under ACS, was calculated and used in lieu of n_1 in the formula for the variance of $\bar{y}$:

$$Var[\bar{y}_{SRS}] = \left(\frac{N-n_1}{N}\right)\frac{\sigma^2}{n_1} \tag{1.17}$$

In another example, Thompson used a rare, clustered population produced via simulation, setting C as $y_i \geq 1$. He compared the two design strategies as in the previous paragraph for a variety of initial sample sizes n_1. Relative efficiency of a given estimator $\hat{\mu}_.$ was calculated as:

$$\text{efficiency}(\hat{\mu}_.) = Var[\bar{y}]/Var[\hat{\mu}_.] \tag{1.18}$$

This formula will also hold for $\hat{\tau}_.$ since $\hat{\tau}_. = N\hat{\mu}_.$. Obviously, $\hat{\mu}_.$ is considered a more efficient estimator when efficiency$(\hat{\mu}_.) > 1$. The results showed that both $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$ became increasingly more efficient as n_1 increased, with the efficiency increase being far greater for $\hat{\mu}_{HT}$ than for $\hat{\mu}_{HH}$. Next, the y-values were set to indicate the presence or absence of point objects in the units. Interestingly, the efficiency of $\hat{\mu}_{HH}$ was worse (less than 1) regardless of the initial sample size. However, the efficiency of $\hat{\mu}_{HT}$ was worse only for smaller n_1, but for larger n_1 the efficiency of $\hat{\mu}_{HT}$ was better (greater than 1) once again.

Although the design-unbiasedness of ACS does not depend on the type of population being sampled, the relative efficiency of ACS does. Thompson (1990) showed that ACS using a simple random sample without replacement of size n_1 and $\hat{\mu}_{HH}$ will have lower variance than the sample mean of a simple random sample of size n if and only if:

$$\left(\frac{1}{n_1} - \frac{1}{n}\right)\sigma^2 < \frac{N-n_1}{Nn_1(N-1)}\sum_{k=1}^{K}\sum_{i\in B_k}(y_i - w_{k(i)})^2 \tag{1.19}$$

$$= \frac{N-n_1}{Nn_1}\sigma_W^2 \tag{1.20}$$

where $k(i)$ is the kth network that includes unit i, B_k denotes the set of units that comprise the kth network, and $w_{k(i)}$ is the average of the y-values of the units in the kth network that includes unit i. Note that in this context n is also the size of the final adaptive sample. We see that the term on the right side of equation 1.19 contains the within-network variance component of the total population variance, or

σ_W^2. Thus, equation 1.19 says that ACS using $\hat{\mu}_{HH}$ will be more efficient than simple random sampling if the within-network variance of the population is "high". This is similar to the case with the decision to use conventional cluster sampling. One would want to form clusters such that the y-values within each cluster are as variable as possible but the cluster total values are as similar as possible. Smith et al. (1995) expressed Equation 1.19 as follows:

$$\sigma^2 < \frac{f_D(1-f_1)}{(f_D-f_1)}\sigma_W^2 \tag{1.21}$$

where the initial and final sampling fractions are $f_1 = \frac{n_1}{N}$ and $f_D = \frac{n}{N}$.

In light of equation 1.19, we now have an explanation for the result that $\hat{\mu}_{HH}$ is less efficient than $\bar{y}$ for all initial sample sizes when the population consisted of y-values that were either 0 or 1. For this population, the within-network variance would be 0 since every network would consist of a single unit with $y_i = 0$ or a group of one or more units each with $y_i = 1$. Thus, for this type of population, the ACS design using $\hat{\mu}_{HH}$ cannot do better than the simple random sample using $\bar{y}$. Similarly, Christman (1996b) demonstrated that, using an initial SRSWOR, $\hat{\mu}_{HT}$ under ACS cannot be more efficient than $\bar{y}$ when all the networks are of size 1 or C is so restrictive that no unit meets the condition. This result could be attributed to the same aforementioned reason.

ACS With an Initial Sample of Primary Units - Theory

Two ACS designs utilizing both primary and secondary units were described by Thompson (1991a), each using modifications of the estimators of Horvitz and Thompson (1952) and Hansen and Hurwitz (1943). The first ACS design occurs when an initial SRSWOR is taken of n_1 primary units where the primary units are strips or clusters (from a conventional standpoint) of an equal number of secondary units. All secondary units are sampled and further adaptive sampling is done at the secondary level.

The second ACS design occurs when an initial simple random sample without replacement is taken of n_1 primary units where the primary units are formed from k systematic selections. These primary units contain an equal number of non-contiguous secondary units, all of which are sampled. Further adaptive sampling is done at the secondary level.

Let N be the number of of primary units, each consisting of M secondary units. Unit (i, j) represents the jth secondary unit in the ith primary unit with associated y-value, y_{ij}. Some classical results are as follows (Scheaffer, Mendenhall, and Ott 1990).

A conventional unbiased estimator of μ is:

$$\bar{y}_1 = \frac{1}{Mn_1} \sum_{i=1}^{n_1} y_{i.} \tag{1.22}$$

where $y_{i.} = \sum_{j=1}^{M} y_{ij}$. Its variance is:

$$Var[\bar{y}_1] = \frac{1}{M^2 n_1}\sigma_Y^2 \left(1 - \frac{n_1}{N}\right) \tag{1.23}$$

where $\sigma_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_{i.} - M\mu)^2$. An unbiased estimator of the variance in equation 1.23 is:

$$\widehat{Var}[\bar{y}_1] = \frac{1}{M^2 n_1}s_Y^2 \left(1 - \frac{n_1}{N}\right) \tag{1.24}$$

where $s_Y^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(y_{i.} - M\bar{y}_1)^2$. Note that for an initial systematic sample with only one starting point ($n_1 = 1$), an unbiased estimator of the variance is not available. It should be noted that both the conventional sample mean and the estimator in equation 1.22 used in conjunction with the *final* ACS sample are biased for μ.

The modified Horwitz-Thompson estimator used in this setting is:

$$\hat{\mu}_{HT} = \frac{1}{MN}\sum_{k=1}^{K}\frac{y_k^* I_k}{\alpha_k} \tag{1.25}$$

There are, however, some differences from the simple random sampling case presented earlier in this chapter. Referring to equation 1.6, now, x_k represents the number of *primary* units in the population that intersect the kth network. I_k is an indicator function taking value 1 with probability α_k if at least one primary unit of the initial sample intersects the kth network, and 0 otherwise, where α_k is defined as before. α_{jk} is the probability that at least one primary unit of the initial sample intersects both networks j and k where x_{jk} is defined to be the number of primary units that

intersect both networks:

$$\alpha_{jk} = 1 - \frac{\left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k + x_{jk}}{n_1}\right]}{\binom{N}{n_1}} \tag{1.26}$$

The following results were derived by Thompson (1991a):

$$E[\hat{\mu}_{HT}] = \mu \tag{1.27}$$

$$Var[\hat{\mu}_{HT}] = \frac{1}{M^2 N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k}\right)\right] \tag{1.28}$$

with unbiased estimator:

$$\widehat{Var}[\hat{\mu}_{HT}] = \frac{1}{M^2 N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_{jk}\alpha_j \alpha_k}\right) I_j I_k\right] \tag{1.29}$$

For an initial systematic sample where $n_1 = 1$, $\alpha_{jk} = 0$ for some j and k, underscoring the fact that an unbiased estimator of the variance is not available for such a design (Thompson 1991a).

In this setting, the modified Hansen-Hurwitz estimator used is:

$$\hat{\mu}_{HH} = \frac{1}{M n_1}\sum_{k=1}^{K} \frac{b_k y_k^*}{x_k} \tag{1.30}$$

where b_k is the number of times network k is intersected by the initial sample of primary units.

The following results were also derived by Thompson (1991a):

$$E[\hat{\mu}_{HH}] = \mu \tag{1.31}$$

$$Var[\hat{\mu}_{HH}] = \frac{\sigma_W^2}{n_1}\left(1 - \frac{n_1}{N}\right) \tag{1.32}$$

where $\sigma_W^2 = \frac{1}{N-1}\sum_{i=1}^{N}(w_i - \mu)^2$, where $w_i = \frac{1}{M}\sum_{k=1}^{K}\frac{I_{ik}y_k^*}{x_k}$, where $I_{ik} = 1$ if the ith primary unit intersects the kth network, and 0 otherwise.

An unbiased estimator of $Var[\hat{\mu}_{HH}]$ is:

$$\widehat{Var}[\hat{\mu}_{HH}] = \frac{s_W^2}{n_1}\left(1 - \frac{n_1}{N}\right) \tag{1.33}$$

where $s_W^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(w_i - \hat{\mu}_{HH})^2$.

ACS Versus Conventional Sampling - Cluster and Systematic Sampling

Thompson (1991a) used a rare, clustered population produced via simulation to examine two primary unit structures. Whenever one or more point objects were encountered in the primary unit, further adaptive sampling was conducted at the secondary level. The first example used initial long, thin strip plots while the second example used a spatial systematic initial sample with two randomized starting points. Once again for comparative purposes, the variance was computed, via equation 1.23, for the sample mean $\bar{y}$ of a simple random sample of primary units with sample size equal to the expected effective primary unit sample size under the adaptive design in place of n_1. Comparisons were done for various initial sample sizes of primary units ranging from 1 up to a sampling fraction of .5.

The first example showed results similar to those from the initial simple random sample results. For all initial sample sizes, both $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$ were relatively more efficient than $\bar{y}$ and this disparity became more pronounced as n_1 increased. $\hat{\mu}_{HT}$

was more efficient than $\hat{\mu}_{HH}$ for all initial sample sizes greater than size 1 and this disparity became more pronounced as n_1 increased.

The second example yielded dramatic results. Variances using initial systematic samples were much lower than those variances obtained using initial long, thin strip primary plots. This was attributed by Thompson to the spatial process used to simulate the population data (Thompson 1991a). Adaptive strategies using $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$ were far more efficient relative to simple random sampling without replacement for the population used and this disparity became more pronounced with increasing n_1. In fact, the relative efficiency of $\hat{\mu}_{HT}$ was 1.52 for $n_1 = 1$ and went to infinity as n_1 increased to 8. The reason this occurred was because the adaptive strategy had zero variance for $n_1 > 6$ because the α_ks became 1 for all the networks in the population used in this example. Once again, the results showed $\hat{\mu}_{HT}$ was more efficient relative to $\hat{\mu}_{HH}$, with the disparity becoming more pronounced with increasing n_1.

Stratified ACS - Theory

Stratification is used when there is a known source of variability such that homogeneous units can be formed into "blocks" or strata. Used appropriately, stratification can reduce the variance of an estimator. We assume simple random samples without replacement are taken of the units in *each* stratum. One can then adaptively sample at the unit level. A problem occurs when a cluster lies in more than one stratum. One possible solution might be that if such clusters are truncated at the stratum

boundaries, then the stratum estimators are independent and can be combined to provide an overall weighted estimator of μ. However, it will be shown that truncating clusters leads to a loss of efficiency and consequently using an estimator that allows for clusters to overlap stratum boundaries is desired (Thompson 1991b).

Let the population of N units be partitioned into H strata, with N_h units in the hth stratum, $h = 1, 2, \ldots, H$. Unit (h, i) is defined to be the ith unit in the hth stratum with y-value y_{hi}. If an initial simple random sample without replacement of n_h units are taken from each stratum h, then we define the initial total sample size to be $n_0 = \sum_{h=1}^{H} n_h$. Some classical results are as follows (Scheaffer et al. 1990).

An unbiased estimator of μ is:

$$\bar{y}_0 = \sum_{h=1}^{H} \frac{N_h}{N} \bar{y}_h \tag{1.34}$$

where $\bar{y}_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$. Its variance is:

$$Var[\bar{y}_0] = \frac{1}{N^2} \sum_{h=1}^{H} N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} \tag{1.35}$$

where $\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \mu_h)^2$, where μ_h is the population mean for stratum h. Were $\bar{y}_0$ to be used in conjunction with the *final* ACS sample, then it would be biased for μ.

Thompson (1991b) proposed several estimators and derived a variety of results to be used in conjunction with stratified ACS. The first ACS estimator we will consider is:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* I_k}{\alpha_k} \tag{1.36}$$

where the K distinct networks are indexed *without* regard to stratum boundaries and I_k equals 1 with probability α_k if the initial sample of size n_0 intersects network k, and 0 otherwise. The derivation of α_k, the marginal initial intersection probability, is now more complex because we now must consider probabilities of intersecting network k with initial samples in *each* of the strata. Because an initial unit in a given stratum can only intersect that part of each network in the stratum, we define x_{hk} to be the number of units in stratum h that lie in network k. Thus we have:

$$\alpha_k = 1 - \left[\prod_{h=1}^{H} \binom{N_h - x_{hk}}{n_h} \Big/ \binom{N_h}{n_h}\right] \tag{1.37}$$

and the probability that the initial sample intersects both networks j and k, or the joint initial intersection probability for both networks j and k:

$$\alpha_{jk} = 1 - (1 - \alpha_j) - (1 - \alpha_k) + \left[\prod_{h=1}^{H} \binom{N_h - x_{hj} - x_{hk}}{n_h} \Big/ \binom{N_h}{n_h}\right] \tag{1.38}$$

The following result was derived:

$$E[\hat{\mu}_{HT}] = \mu \tag{1.39}$$

We also have the result that $Var[\hat{\mu}_{HT}]$ and its unbiased estimator are the same as equations 1.9 and 1.10, respectively.

In order to discuss the estimators based on a modification of Hansen and Hurwitz (1943), we will need additional notation. Let A_{hi} be the network containing unit (h, i), A_{ghi} be the part of A_{hi} in stratum g, and m_{ghi}be the number of units in A_{ghi}.

We then have the first of three similar estimators:

$$\hat{\mu}_{HH} = \sum_{h=1}^{H} \frac{N_h}{N} \bar{w}_h \tag{1.40}$$

where $\bar{w}_h = \frac{\sum_{i=1}^{n_h} w_{hi}}{n_h}$, where $w_{hi} = \frac{n_h}{N_h} \frac{Y_{hi}}{\sum_{g=1}^{H} \frac{n_g}{N_g} m_{ghi}}$, where Y_{hi} is the sum of the y-observations in A_{hi}.

Thompson (1991b) derived the following results:

$$E[\hat{\mu}_{HH}] = \mu \tag{1.41}$$

$$Var[\hat{\mu}_{HH}] = \frac{1}{N^2} \sum_{h=1}^{H} N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} \tag{1.42}$$

where $\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (w_{hi} - \bar{W}_h)^2$, where $\bar{W}_h = \frac{\sum_{i=1}^{N_h} w_{hi}}{N_h}$.

An unbiased estimator of $Var[\hat{\mu}_{HH}]$ is obtained by simply replacing σ_h^2 in equation 1.42 with $s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (w_{hi} - \bar{w}_h)^2$.

A second estimator, based on the stratified "multiplicity" estimator of network sampling, was proposed by Thompson, who also gave a variety of literature citations on the history of this type of estimator (Thompson 1992):

$$\hat{\mu}'_{HH} = \sum_{h=1}^{H} \frac{N_h}{N} \bar{w}'_h \tag{1.43}$$

where $\bar{w}'_h = \frac{\sum_{i=1}^{n_h} w'_{hi}}{n_h}$, where $w'_{hi} = \frac{Y_{hi}}{\sum_{g=1}^{H} m_{ghi}}$.

This estimator also is unbiased for μ. $Var[\hat{\mu}'_{HH}]$ and its unbiased estimator, $\widehat{Var}[\hat{\mu}'_{HH}]$, are calculated via equation 1.42, using w'_{hi}, instead of w_{hi}.

Finally, there is a third estimator which ignores any units added through crossing stratum boundaries (truncates clusters):

$$\hat{\mu}''_{HH} = \sum_{h=1}^{H} \frac{N_h}{N} \hat{\mu}_h \tag{1.44}$$

where $\hat{\mu}_h = \frac{\sum_{i=1}^{n_h} w''_{hi}}{n_h}$ and w''_{hi} is the total of the y-values in the intersection of stratum h with A_{hi} divided by the number of units in the intersection.

This estimator also is unbiased for μ. $Var[\hat{\mu}''_{HH}]$ and its unbiased estimator, $\widehat{Var}[\hat{\mu}''_{HH}]$, are calculated via equation 1.42, using w''_{hi}, instead of w_{hi}.

ACS Versus Conventional Sampling - Stratified Random Sampling

Thompson (1991b) once again used a rare, clustered population produced via simulation to compare ACS and conventional sampling under stratification. The design divided the population into two strata where the intial SRSWORs were of equal size. A unit satisfied the condition C if $y_i \geq 1$. One network straddled the boundary of the two strata. A variety of sample sizes were examined. The variance of the stratified sample mean of the conventional stratified random sample with total sample size equal to $E[\nu]$, the expected effective sample size under ACS, was calculated for comparative purposes.

The results showed the following relationship for all n_0:

$$Var[\bar{y}_0] > Var[\hat{\mu}''_{HH}] > Var[\hat{\mu}'_{HH}] = Var[\hat{\mu}_{HH}] > Var[\hat{\mu}_{HT}] \tag{1.45}$$

This relationship of inequalities became more pronounced with increasing n_0. The equality $Var[\hat{\mu}'_{HH}] = Var[\hat{\mu}_{HH}]$ is by construction because the two estimators are identical if stratum and sample sizes are equal. Thompson (1991b) showed that the variances of $\hat{\mu}_{HH}$, $\hat{\mu}_{HH}'$, and $\hat{\mu}_{HH}''$ were fairly similar. For this population, however, this result was not surprising because a couple of the aggregations were far away from the stratum boundary. He also noted that substantial variance reduction can be achieved relative to SRS when ACS is used, with the reduction oftentimes being dramatic.

Next, using the same population, the estimates were compared when the initial sample sizes in the two strata were in the ratio 2:1. The conclusions were identical to when the initial sample sizes were equal except that now, $Var[\hat{\mu}'_{HH}] > Var[\hat{\mu}_{HH}]$.

Finally, the same population was partitioned into 16 strata where the initial strata samples were of equal size for a range of n_0. The conclusions were identical to equation 1.45. However, the gain in relative efficiency was much more so than in the 2 strata case. Thompson (1994) speculated that this was because the initial conventional design itself, stratified sampling with many small strata, is efficient for spatially aggregated populations and that ACS improved the efficiency even further.

CHAPTER 2

A REVIEW OF ADAPTIVE CLUSTER SAMPLING: 1990-2002

Practical Issues Determining Relative Efficiency of ACS

In general, neither ACS nor conventional sampling designs are uniformly better than the other over every possible population configuration (Thompson and Seber 1996). Smith et al. (1995) warn that blind application of ACS may lead to a sampling design no better or worse than SRS in terms of precision. Christman (2000) also noted this in her research. A number of factors involved in determining the efficiency of ACS relative to SRSWOR already have been mentioned. We now take an in-depth look at a variety of these factors. As was discussed in the Introduction, ACS with initial designs other than simple random sampling have been compared through examples with corresponding conventional designs, but analytic expressions of component factors influencing relative efficiency have not been obtained.

Choice of Sampling Design

Choosing an appropriate sampling design can be important and complicated particularly if the population is rare and/or clustered. The reader is referred to Christman (2000) for a review of quadrat-based sampling of rare and clustered populations

as it relates to basic survey sampling issues. Additionally, in this review, a simulation study was done comparing the different classes of sampling designs across three rare and clustered populations with varying degrees of within-network variance. Five thousand samples of fixed sample size were drawn. Furthermore, different schemes for allocating the initial sample size were considered for both conventional and ACS stratified sampling designs. The designs that had the lowest variances of the estimators of τ, the population total, were those in which the population was stratified such that one stratum contained all of the rare elements and the other contained none of the rare elements and optimal or disproportional allocation of stratum SRS sizes was employed. If stratification was done arbitrarily, as is oftentimes necessary, then use of stratified ACS resulted in a dramatic reduction of variance relative to stratified SRSWOR, regardless of the method of stratum sample size allocation. However, a conventional systematic sampling design gave even lower variance than one using stratified ACS and had a smaller expected sample size. When systematic ACS was conducted, the variance dropped lower still although the expected sample size reached its highest level. Consequently, the major result was that if stratification based on location of the rare elements is not possible a priori and sampling cost is not an issue, then systematic ACS sampling may provide the best option from a variance efficiency standpoint.

Within-Network Variability and Related Issues

By partitioning the total sum of squares for the variable of interest into between-network and within-network components, it can be shown that relative efficiency, not corrected for unequal sampling effort, can be expressed as:

$$\frac{Var[\hat{\mu}_{HH}]}{Var[\bar{y}]} = \left[\frac{m}{n_1}\left(\frac{N-n_1}{N-m}\right)\right]\left[1 - \frac{\sum_{k=1}^{K}\sum_{i\in B_k}(y_i - w_{k(i)})^2}{\sum_{i=1}^{N}(y_i-\mu)^2}\right] \tag{2.1}$$

where m is the size of the SRS, n_1 is the initial ACS sample size, B_k is the set of units that comprise the kth network, and $w_{k(i)}$ is the average of the y-values of the units in the network that includes unit i (Thompson and Seber 1996). It should be noted that the relative efficiency $Var[\hat{\mu}_{HT}]/Var[\bar{y}]$, although straightforward to derive, involves more computation than equation 2.1 and does not lend itself to easy interpretation. On the other hand, as we saw in the Introduction, the relative efficiency can generally be improved through the use of $\hat{\mu}_{HT}$ anyway.

The last, or second, term in the square brackets of equation 2.1 contains a fraction that is the proportion of total population variance that is contained within networks. For given equivalent sample sizes m and n_1, it is (as stated in the previous chapter) clear that ACS will be a more efficient design when the within-network variation is a large proportion of the total variation. Such a population could be described as *clustered* or as having aggregation tendencies. Thus, the relative efficiency depends on the partitioning of the population into K networks, which depends on the spatial distribution of the population, condition C, the neighborhood structure, the quadrat

size, and the size and the selection of the initial sample. Unfortunately, many of these factors are interrelated which complicates finding the most efficient design. (Brown 1996; Christman 1996a,b; Smith et al. 1995; Thompson and Seber 1996). The results of Smith et al. (1995) suggest that for the Horvitz and Thompson estimator, the magnitude of the within-network variance relative to the total population variance may be the most important factor determining efficiency.

Christman (1996b, 1997) computationally verified the result that $\hat{\mu}_{HH}$ peforms well relative to $\bar{y}$ from a SRS when the within-network variance is a large proportion of the total variance and there exists a first-order neighborhood. Similar results were noted for $\hat{\mu}_{HT}$.

Critical Value

Christman (1996b, 1997) demonstrated that $\hat{\mu}_{HT}$ is an efficient estimator for an array of populations relative to SRS if the critical value $c_a = 1$. For first-order neighborhoods, the efficiency performance of $\hat{\mu}_{HT}$ became worse when the critical value c_a was set to 5. Christman's result (Christman 1996b) that $\hat{\mu}_{HH}$ peforms well relative to $\bar{y}$ from a SRS when the within-network variance is a large proportion of the total variance and there exists a first-order neighborhood was less pronounced with $c_a = 5$ than with $c_a = 1$. It was speculated that this was in part due to a decrease in within-network variability imposed by the more restrictive condition.

In looking at $\hat{\mu}_{HT}$, when $c_a = 2$, Brown (1996) reported an improvement in relative efficiencies for ACS for high density populations. The improvement occurs because increasing c_a has the effect of reducing the size of networks and thus reducing the number of superfluous edge units. However, for low density populations, relative efficiencies for ACS actually became worse because few networks were formed and few units were selected adaptively.

Christman (2000) also noted that the efficiency of ACS relative to simple random sampling of equivalent sample size is inversely related to the choice of C. Moreover, as c_a increases, within-network variance decreases. Hence, choosing a relatively large value for c_a in order to control the overall sample size has the undesired effect of reducing the efficiency of ACS. Furthermore, edge units likely will contain nonzero counts but will not be in the estimation procedure, which leads to a loss of information.

A small value of c_a results in a selected large cluster being sampled intensively while a large value of c_a results in many small clusters being sampled, but each at a lower intensity. The choice of C will therefore depend on whether sampling effort should be concentrated on sampling individual clusters or on sampling many clusters which ultimately depends on where the largest source of variation is, within or between clusters (Brown 1994).

Neighborhood Definition

As was the case with the selection of C, altering the definition of the neighborhood can change the focus of sampling to within or between clusters (Brown 1994). Christman (1996b) examined ACS for several populations using three types of neighborhoods. The result was that the most efficient ACS designs were those that utilized neighborhoods based on units that were physically contiguous. In fact, Christman (1996b, 1997) proposed a non-contiguous, "four-away" neighborhood that was constructed such that if unit(i, j) was selected and $y_{ij} \geq c_a$, then units lying two over to the north, south, east and west were sampled. For this type of neighborhood, both $\hat{\mu}_{HH}$ and $\hat{\mu}_{HT}$ had a higher variance than $\bar{y}$ from an SRSWOR of size $E[\nu]$ for almost every initial sample size for all populations simulated in the research!

There is also some evidence suggesting that a truncated first-order neighborhood, where only two adjacent units are sampled, may be as or more efficient than the usual first-order neighborhood. This would depend on the population structure, particularly if the study region was two-dimensional. This result is not surprising because, in general, a neighborhood consisting of only 2 neighbors rather than 4 neighbors lowers the effective sample size, which we will learn is a condition that has been shown to be favorable for ACS. Also, for this reason, the truncated first-order neighborhood might be attractive if cost is a concern (Christman 1996b, 1997). If clusters tend to form

in a specific direction, then the neighborhood definition should reflect the expected orientation (Christman 2000).

Brown (1996) achieved better relative efficiencies for ACS when a smaller first-order neighborhood was used instead of a second-order neighborhood. It was proposed that this was because relatively more units were used in the estimator $\hat{\mu}_{HT}$ when there was a smaller neighborhood definition because the number of superfluous edge units was minimized.

Estimator Type

As was shown in the first chapter, $\hat{\mu}_{HT}$ is a more efficient estimator than $\hat{\mu}_{HH}$ under a variety of ACS designs. Thus, $\hat{\mu}_{HT}$ should be more efficient than $\hat{\mu}_{HH}$ relative to $\bar{y}$ from an SRSWOR of size $E[\nu]$. Shortly, we will see that greater efficiency in ACS may be achieved through use of the Rao-Blackwell improvement of either $\hat{\mu}_{HH}$ or $\hat{\mu}_{HT}$.

Christman (1996b, 1997) compared the variances and efficiencies of $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$ in a computer simulation study. The results showed that while $\hat{\mu}_{HT}$ had the lowest variance under most of the simulation situations explored, it was much more sensitive to changes in the population structure and in the sampling conditions discussed in this chapter. Thus, the sampling design should be carefully planned when using $\hat{\mu}_{HT}$ and potential modification of the sampling effort anticipated. This will be discussed in greater detail in the later section "Limiting Sampling Effort and Cost in ACS".

Unfortunately, the efficiency of $\hat{\mu}_{HT}$ often comes at the cost of a larger effective sample size.

Unit Size

The choice of unit size determines to some extent the size of networks. As with the choice of C, if the surveyor had prior knowledge of the population structure, then he or she could increase the efficiency of ACS using $\hat{\mu}_{HH}$ by selecting the unit size accordingly (Christman 1996b, 1997). These results were corroborated by the simulation study done by Smith et al. (1995) who compared the efficiency of several sampling designs using Horvitz-Thompson estimation of the abundance of several species of wintering waterfowl. They noted that the results varied according to species, C, and unit size, with some combinations leading to increased ACS precision compared to SRSWOR and some not. Size of the unit affected the number and size of networks and the within-network variability.

Sample Size and Cost - Theory

It is also clear from equation 2.1 that ACS will be a more efficient design than SRSWOR when the first of the two terms in square brackets is small. Define:

$$b(n_1, m, N) = \frac{m}{n_1}\left(\frac{N - n_1}{N - m}\right) \tag{2.2}$$

The following results can be shown via basic calculus. For fixed n_1, b is an increasing function of m. When $m = n_1$, $b = 1$. For most comparisons of interest, $m \geq n_1$. ACS

will be more efficient if m is not much bigger than n_1. If $m \leq n_1$, it is obvious upon inspection of the previous equation and equation 2.1 that ACS will always be more efficient than simple random sampling. If m and n_1 both increase at a constant ratio, that is, $m = a_1 k$ and $n_1 = a_2 k$, b is an increasing function of k. Finally, with n_1 and m fixed, b is a decreasing function of N if $m > n_1$ and increasing if $m < n_1$.

Thompson (1994) proposed the following cost function for ACS:

$$c_T = c_0 + (c_1 - c_2)n_1 + c_2 E[\nu] \tag{2.3}$$

where c_T is the total cost, c_0 is an initial fixed cost, c_1 is the marginal cost for each unit obtained by simple random sampling, c_2 is the marginal cost for each adaptively added unit and $E[\nu]$ is the expected final sample size when the initial sample size is n_1. Note this is only one of many possible cost functions. In some studies sampling an edge unit may take less time, and therefore be less costly than sampling a network unit. The cost function would then include separate components for both unit types (Brown 1996; Thompson 1994).

For the same total cost, the conventional simple random sampling design could use a sample size m_c satisfying:

$$c_T = c_0 + c_1 m_c \tag{2.4}$$

Thus:

$$m_c = \left(1 - \frac{c_2}{c_1}\right) n_1 + \frac{c_2}{c_1} E[\nu] \tag{2.5}$$

In general, $c_2 < c_1$ since there is less cost associated with sampling units within clusters due to distance and time considerations. Intuitively, this is not an unreasonable assumption. Analytically, Brown (1996) showed in a simulation study that the average distance travelled for ACS was less than the average distance for SRSWOR across an array of population models. Hence, m_c is a weighted average of n_1 and $E[\nu]$ where:

$$n_1 \leq m_c \leq E[\nu] \tag{2.6}$$

Since the variance for the SRS decreases as m increases, we have for a fixed expected cost:

$$\frac{Var[\hat{\mu}_{HH}; n_1]}{Var[\bar{y}; n_1]} \leq \frac{Var[\hat{\mu}_{HH}; n_1]}{Var[\bar{y}; m_c]} \leq \frac{Var[\hat{\mu}_{HH}; n_1]}{Var[\bar{y}; E[\nu]]} \tag{2.7}$$

Thompson and Seber (1996, pp. 110-111) showed that the left side of inequality 2.7 is always less than or equal to 1. The right side of inequality 2.7 can serve as a conservative assessment of efficiency giving the benefit of the doubt to the conventional strategy and was the basis of the comparisons discussed in the Introduction.

Define $g(n_1)$ to be b from equation 2.2 evaluated at $E[\nu]$:

$$g(n_1) = \frac{E[\nu]}{n_1}\left(\frac{N - n_1}{N - E[\nu]}\right) \tag{2.8}$$

For a fixed N, $E[\nu]$ is a nondecreasing function of n_1 and that $E[\nu]/n_1$ is decreasing in n_1 (Thompson and Seber 1996, p. 153-154). The relationship between $E[\nu]$ and n_1 is discussed in greater detail in the next section on geographic rarity. Additionally, however, $g(n_1)$ may be either increasing or decreasing in n_1 (Thompson and Seber

1996, p. 154-155). As a result, $g(n_1)$ and hence the upper bound on the relative efficiency, are similarly influenced by the choice of n_1, in addition to the population structure.

Brown (1996) showed, using the Horvitz-Thompson estimator, a general improvement in relative efficiencies for ACS when n_1 was increased, particularly in high density populations. It was speculated that increasing n_1 also increases the expected number of selected networks. This in turn would have the effect of increasing the proportion of within-network variation of the estimated variance, $\widehat{Var}[\hat{\mu}_{HT}]$.

Christman (1996b, 1997) found that in order for $\hat{\mu}_{HT}$ to be an efficient estimator for populations with "large" numbers of clusters, a larger n_1 is needed. Unfortunately, as a result, there can be a dramatic increase in $E[\nu]$. Across an array of populations and design strategies, the efficiency of $\hat{\mu}_{HH}$ tended to improve with increasing n_1 but generally was less efficient than $\bar{y}$ under SRSWOR of size $E[\nu]$. It was speculated that the reason this occurred was that while $Var[\hat{\mu}_{HH}]$ decreased with increasing n_1, the increase in n_1 was such that $Var[\bar{y}]$ decreased at a relatively quicker rate.

Geographic Rarity

A population with a given clustering or aggregation pattern has more *geographic rarity* if the given clusters or aggregations are scattered over a large study region than one of the same population size and number of clusters with the clusters crowded into a smaller study region. In the former case, only a few units will contain elements,

most of the other units will contain none, and these units that do contain elements are spatially clustered in a contiguous fashion. It can be shown for a population of a fixed number of elements and fixed spatial aggregation pattern that:

$$lim_{N\to\infty} E[\nu] = n_1 \tag{2.9}$$

In other words, the expected effective sample size approaches the initial sample size with increasing geographic rarity (Thompson and Seber 1996, p. 156). This was also shown analytically in two simulation studies done by Brown (1994) and Brown and Manly (1998). Furthermore, due to this result, the right side of equation 2.7 has been shown with increasing geographic rarity, or as N goes to infinity, to approach an uncorrected version of the last, or second, term in square brackets on the right side of equation 2.1 (Thompson and Seber 1996, p. 156-157). Thus, this says that if there is any within-network variation whatsoever, aggregated populations, if geographically rare, will be more efficiently sampled with ACS than with conventional designs. Smith et al. (1995) claimed that the relative efficiency of ACS for Horvitz-Thompson estimation was highest when $\frac{E[\nu]}{N}$ was "close" to $\frac{n_1}{N}$. They determined that $E[\nu]$ depended on the same factors that were previously mentioned to have determined within-network variance and that although the uncertainity of the final sample size is a disadvantage of ACS, a small difference between the final and initial sampling fractions is characteristic of an efficient design. However, while agreeing that the claim was valid for Hansen-Hurwitz estimation, Christman (1996b,

1997) seemed to dispute this claim for Horvitz-Thompson estimation. This topic will be explored further in the Discussion.

Christman (1996a), in an experimental simulation study using a fixed C and n_1, found that only those populations where the number of clusters was small typically yielded relative efficiencies favorable for ACS. As the number of clusters (and consequently, $E[\nu]$) was increased, the distributions of relative efficiencies became very skewed with the majority of simulated populations yielding relative efficiencies favorable for SRSWOR. Occasional "outliers", or relative efficiencies favorable for ACS, were attributed to high within-network variability.

Brown (1996) also reported similar results, stating that ACS "is not a suitable technique for sampling populations that do not have a very patchy spatial pattern". To discern differences between patchiness, indices of aggregation were used to discriminate between the populations grouped according to the appropriateness of using ACS and to construct classification functions for applied use.

Further Developments

A careful and thoughtful preliminary investigation of the population structure should be conducted before utilizing ACS. More work needs to be done to help develop guidelines in setting the levels of the factors that influence ACS surveys. It has been suggested that this goal may be accomplished via a pilot study (Brown 1996; Smith et al. 1995). An approach suggested by Brown (1996) would be to use a small unit

size, a small critical value, and a large neighborhood definition in a pilot survey. The data could then be reanalyzed using various levels of the aforementioned factors to help define an appropriate design for the population.

Another approach suggested by Brown (1996) would be to sample the population in phases. At the initial phase, educated guesses are made for the levels of the sample design factors and an initial area is surveyed. The results from the initial phase are used to see if levels of the design factors need to be changed. Sampling would then continue in the next phase with the revised sample design. Again, the sample design would be re-evaluated and, if necessary, revised further for the subsequent phase. Separate estimates from each phase would be combined to give one estimate for the total area.

Christman (1997) also suggested that a sequential component be added to the sample design in which statistics that describe the aggregation and rarity of the clusters would be calculated from the sample data and the results used to modify further sampling effort.

Initial Unequal Probability Sampling

There are some circumstances in designing a survey where each unit will not have the same probability of inclusion in the initial sample. Perhaps the quadrats may simply be of unequal size. For example, in some ecological studies, the usual

rectangular grid of secondary units is not adequate to cover the study site (Pontius 1996, 1997).

Roesch (1993) used PPS, or probability proportional to size selection, with replacement to randomly drop points in the population study area in a simulation of ACS in a forest inventory. Here, the auxilliary variable was the circular area of trees proportional to the trunk radius at basal height. He found this strategy to be reasonable when sampling for a rare characteristic that tended to be clustered in groups of trees. Both the theory and methodology of the Roesch paper are well described in Thompson and Seber (1996, p. 101-108).

Smith et al. (1995) used PPS in their oft-cited simulation study for estimating the abundance of several species of wintering waterfowl. They examined ACS with the initial sample selected by SRSWOR and PPS with replacement and compared it to conventional sampling: SRSWOR and PPS with replacement. For PPS sampling under ACS, the Horvitz-Thompson estimator was used with modified initial intersection probabilities that were based on the area of available habitat in a sampling unit. The results indicated that PPS under ACS had better efficiency relative to SRSWOR under ACS for certain conditions (unit size and C) when the auxilliary variable habitat was appropriately defined.

ACS with an initial sample of primary units has been extended to situations where the primary units have unequal numbers of equally sized secondary units and

consequently falls under PPS (Pontius 1996, 1997). This research centered on selection of primary units with replacement in conjunction with the Horvitz-Thompson estimator. Two cases were considered: the strip ACS design considered in the Introduction and strip ACS under stratification of primary units. In the latter case, selection of primary units was done independently between strata. Modifications of both the marginal and joint initial intersection probabilities were derived for both cases in great detail and functions to calculate them in S+ were presented (Pontius 1996).

Rao-Blackwell Estimators

Perhaps one of the more interesting developments in ACS has been the application of the Rao-Blackwell method. Note that the unbiased estimators $\bar{y}_1$, $\hat{\mu}_{HT}$, and $\hat{\mu}_{HH}$ applied to an adaptive cluster sample all depend on the order of selection because they depend on which n_1 units are in the initial sample. Additionally, $\hat{\mu}_{HH}$ depends on repeat selections and if the initial sample was selected WR, then all three estimators would depend on repeat selections. Thus, since all three estimators are not functions of D_R, we can apply the Rao-Blackwell Theorem (Theorem 1.2 and Corollary 1.3 from the Introduction). From Thompson and Seber (1996), we have the Rao-Blackwell estimator of $\phi(\boldsymbol{\theta})$:

$$T_{RB} = \frac{1}{\xi} \sum_{g=1}^{G} t_g I_g. \tag{2.10}$$

For an initial SRSWOR, $G = \binom{\nu}{n_1}$, the number of possible combinations of n_1 distinct units from the ν in the final sample, t_g denotes the value of any one of the three estimators when the initial sample consists of combination g, for $g = 1, 2, \ldots, G$, I_g is 1 when the gth combination gives rise to (is compatible with) d_R, and 0 otherwise, and $\xi = \sum_{g=1}^{G} I_g$, the number of compatible combinations. By construction, Rao-Blackwell estimators do utilize edge units even if they are not selected in the initial sample, whereas $\hat{\mu}_{HH}$ and $\hat{\mu}_{HT}$ utilize observations not satisfying C only when they are selected as part of the initial sample.

If T is any one of the three estimators, we know:

$$Var[T_{RB}] = Var[T] - E[Var[T|d_R]]. \tag{2.11}$$

Notice $Var[T]$ is a function of $\boldsymbol{\theta}$ and $\widehat{Var}[T]$, an unbiased estimator of $Var[T]$, is not a function of D_R. Thus, we apply the Rao-Blackwell theorem to obtain an unbiased estimator with smaller variance:

$$\begin{aligned} \widehat{Var}_{RB}[T] &= E[\widehat{Var}[T]|d_R] && (2.12) \\ &= \frac{1}{\xi}\sum_{g=1}^{G} \widehat{Var}_g[T] I_g. && (2.13) \end{aligned}$$

Furthermore:

$$\begin{aligned} Var[T|d_R] &= E[(T - T_{RB})^2|d_R] && (2.14) \\ &= \frac{1}{\xi}\sum_{g=1}^{G} (t_g - T_{RB})^2 I_g. && (2.15) \end{aligned}$$

Consequently:

$$\widehat{Var}[T_{RB}] = \frac{1}{\xi}\sum_{g=1}^{G}[\widehat{Var}_g[T] - (t_g - T_{RB})^2]I_g. \tag{2.16}$$

Unfortunately, it is possible for $\widehat{Var}[T_{RB}]$ to take on negative values for some sets of data. If the initial sample is selected WR, then the previous formulas hold with $G = \nu^{n_1}$, the number of permutations of n_1 units, distinguishing order and allowing repeats, from the ν distinct units.

Next, consider the statistic:

$$D_f = \{(i, y_i, f_i) : i \in s_R\}. \tag{2.17}$$

In other words, D_f is D_R along with f_i where f_i is the frequency of intersection of the initial sample with the network A_i, and A_i denotes the network containing unit i for each unit i in the final sample.

Looking for a moment at just the modified Hansen-Hurwitz estimator, we now state the following theorem from Thompson and Seber (1996) without proof.

THEOREM 2.1. *Suppose that sampling is WR or WOR and $\bar{y}_1$ is the initial sample mean. Then:*

1. *D_f is sufficient for $\boldsymbol{\theta}$.*

2. *$E[\bar{y}_1|D_f = d_f] = \hat{\mu}_{HH}$.*

3. *$E[\bar{y}_1|D_R = d_R] = E[\hat{\mu}_{HH}|D_R = d_R] = \hat{\mu}_{RBHH}$.*

Two results become apparent. First, the resulting estimator $\hat{\mu}_{RBHH}$ is the same when Rao-Blackwell theory is applied to both $\bar{y}_1$ and $\hat{\mu}_{HH}$. Second, by parts 1 and 2 of the theorem, $Var[\hat{\mu}_{HH}] \leq Var[\bar{y}_1]$ when both have sample size n_1.

Recall the examples from Thompson (1990) that were discussed in the section "ACS versus Conventional Sampling - Simple Random Sampling" in the Introduction. In the first example an initial sample of fixed size n_1 was selected via SRSWOR from a small population. Thompson (1990) calculated $\hat{\mu}_{RBHH}$ and $\hat{\mu}_{RBHT}$ for all possible samples and showed that $\hat{\mu}_{RBHT}$ had the lowest variance among the five unbiased estimators ($\bar{y}_1$, $\hat{\mu}_{HT}$, $\hat{\mu}_{RBHT}$, $\hat{\mu}_{HH}$, and $\hat{\mu}_{RBHH}$).

The Rao-Blackwell method (equations 2.10 and 2.16) can also be used in stratified ACS. Rao-Blackwell estimation for this situation is complicated by two facts which were discussed in the Introduction. First, when there is stratification, there is substantially more computation involved to determine overall initial compatible stratified samples. Second, there are four distinct Rao-Blackwell estimators, including $\hat{\mu}_{RBHT}$, and the three different modified Hansen-Hurwitz estimators ($\hat{\mu}_{HH}$, $\hat{\mu}'_{HH}$, and $\hat{\mu}''_{HH}$) that exist for stratified ACS. Extending the use of theorem 2.1, it can be shown that the Rao-Blackwell version of the usual initial weighted stratified sample mean of equation 1.34, $\bar{y}_o$, is the same as $\hat{\mu}''_{RBHH}$ (Thompson and Seber 1996). Thompson (1991b) showed an example where initial samples of fixed and equal size for each stratum ($n_1 = n_2$) were selected from a small population with unequal stratum sizes. All possible samples were generated. When the four Rao-Blackwell estimates

were obtained for all of the samples, $\hat{\mu}_{RBHT}$ had the lowest variance among any of the estimators ($\hat{\mu}_{HT}$, $\hat{\mu}_{HH}$, $\hat{\mu}'_{HH}$, and $\hat{\mu}''_{HH}$) and each of these estimators was improved by the Rao-Blackwell method.

In the case where ACS is conducted using an initial sample of primary units (for example, the strip sampling discussed in the Introduction), the Rao-Blackwell method can still be used but is unlikely to result in any improvement. The reason is that it is unlikely that more than one initial selection of primary units will lead to the same d_R (Seber and Thompson 1994). Consequently, $\xi = 1$, $T_{RB} = t_g$, and $Var[T|d_R] = 0$.

One potential problem that arises from the use of Rao-Blackwell method is the complexity of computation. Computational aspects of the Rao-Blackwell estimators deserve further study, because the number of terms in equations 2.10 and 2.16 are based on the samples compatible with d_R and therefore the number of ξ estimates to evaluate from the compatible samples is potentially huge.

In an effort to get around the aforementioned problem, Dryver and Thompson (1998) introduced Rao-Blackwell estimators utilizing *sample* edge units. A sample edge unit is an edge unit whose network that classifies it as an edge unit is intersected by the initial sample, even if the sample edge unit itself is intersected by the initial sample. To obtain the Rao-Blackwell estimators, they conditioned $\hat{\mu}_{HH}$ and $\hat{\mu}_{HT}$ on D^+, a statistic that is sufficient but not minimal, rather than D_R. They claimed that the effort entailed in computing these Rao-Blackwell estimators is reduced. Define

the statistic:

$$D^+ = \{(i, y_i, f_i) : i \in s_c, (j, y_j) : j \in s_{\bar{c}}\}. \tag{2.18}$$

Here, s_c is the set of all the distinct units in the final sample satisfying C, and $s_{\bar{c}}$ is the set of all the distinct units in the final sample not satisfying C. The new estimators, $\hat{\mu}_{RBHH+}$ and $\hat{\mu}_{RBHT+}$, are developed by considering the average of the sample edge units in the final sample.

One estimator is:

$$\hat{\mu}_{RBHH+} = \frac{1}{n}\sum_{i=1}^{n} w_i' \tag{2.19}$$

Here, $w_i' = w_i(1 - e_i) + \bar{y}_e e_i$, where w_i is the usual average value of a unit in the network which contains unit i, where e_i is an indicator variable that takes on the value 1 if unit i is a sample edge unit and 0 otherwise, and where $\bar{y}_e$ is the average y-value for the sample edge units in the final sample. Derivation of $Var[\hat{\mu}_{RBHH+}]$, and both an unbiased estimator $\widehat{Var}[\hat{\mu}_{RBHH+}]$ and unbiased Rao-Blackwell estimator $\widehat{Var}_{RB}[\hat{\mu}_{RBHH+}]$ are given in Dryver and Thompson (1998, p. 730).

The second estimator is:

$$\hat{\mu}_{RBHT+} = \frac{1}{N}\sum_{k=1}^{K} \frac{y_k' I_k}{\alpha_k} \tag{2.20}$$

Here, y_k' is the usual sum of the network y-values if the network satisfies C while for a sample edge unit (a network of size one) it is defined to be $\bar{y}_e$. Derivation of $Var[\hat{\mu}_{RBHT+}]$, both an unbiased estimator $\widehat{Var}[\hat{\mu}_{RBHT+}]$ and unbiased Rao-Blackwell

estimator $\widehat{Var}_{RB}[\hat{\mu}_{RBHT+}]$, and an example are given in Dryver and Thompson (1998, p. 730-731). It should be noted that since we are conditioning on a statistic that is sufficient but not minimal, then for either the Horvitz-Thompson or Hansen-Hurwitz estimator we have:

$$Var[\hat{\mu}_{RB.}] \leq Var[\hat{\mu}_{RB.+}] \leq Var[\hat{\mu}_{.}]. \tag{2.21}$$

Salehi (1999) introduced closed forms of both $\hat{\mu}_{RBHH}$ and $\hat{\mu}_{RBHT}$ along with unbiased estimators of their variances which can be easily computed and revealed some other interesting results. For instance, it is frequently the case for many studies that C is $y_i \geq 1$. The author showed that if the y-values for the sample edge units are zero, then $\hat{\mu}_{HT}$ cannot be improved using Rao-Blackwell theory and $\hat{\mu}_{RBHT} = \hat{\mu}_{HT}$. On the other hand, $\hat{\mu}_{HH}$ can be improved using $\hat{\mu}_{RBHH}$ when the y-values of sample edge units are zero. It was further shown that applying Rao-Blackwell theory to $\hat{\mu}_{HH}$ will result in a greater relative improvement (variance reduction) than for the case of applying the theory to $\hat{\mu}_{HT}$.

Salehi illustrated his computations, notation, and conclusions via a small population example and an analysis of a set of data from Smith et al. (1995). For the latter, he deliberately selected an initial sample involving the most computation possible. In both of these examples, $\hat{\mu}_{RBHH}$ was more efficient than $\hat{\mu}_{RBHT}$. Even though these results suggest otherwise, Thompson and Seber (1996) caution us that neither $\hat{\mu}_{RBHH}$ nor $\hat{\mu}_{RBHT}$ is uniformly better than the other. The reason, as we recall from corollary 1.3 and theorem 1.4 of the Introduction, is due to the incompleteness of D_R.

Other Estimators

Christman and Lan (1998) considered ACS based on sequentially selecting initial sampling units until some predefined number of non-zero Y-values are obtained in order to estimate τ. The idea is that by stopping when a certain fixed number of k units are sampled satisfying C one will have controlled the sample size and also avoid estimates of zero abundance. Note the number of networks intersected by the initial sample will then be less than or equal to k. Because of the sequential sampling, the usual ACS Hansen-Hurwitz estimator from the Introduction will be positively biased. Thus, Christman and Lan introduced a new sequential unbiased ACS estimator of τ, called $\hat{\tau}_{IA}$, based on a modification, via the use of network means, of a previously described conventional sequential estimator. For further comparative purposes, the ACS Hansen-Hurwitz estimator was modified such that n_1 was set equal to the expected value of the number of sequentially sampled, but not adaptively sampled, units and the sample was taken randomly. Note that not only is this form of the estimator not sequential, it is now unbiased as well. Simulations of samples from a small population were taken according to the sequential adaptive strategy under both with- and without-replacement strategies. Results showed that adding an adaptive component reduced the variance associated with conventional sequential sampling. The comparative Hansen-Hurwitz ACS estimator mentioned above had smaller variance than $\hat{\tau}_{IA}$. However, with the former, there is a positive probability

of returning an estimated value of 0 for τ. Unfortunately, the distribution of $\hat{\tau}_{IA}$ was highly skewed right. Thus, Christman decided to compare efficiencies using the more resistant median absolute deviations of the estimators instead of variances. Using this criterion, the results showed that $\hat{\tau}_{IA}$ was more efficient than the ACS Hansen-Hurwitz estimator. Therefore, the decision over which type of ACS design to use, sequential versus non-sequential, amounts to considering the tradeoff between the risk of obtaining a very high outlying estimate versus the risk of obtaining a zero estimate. This decision could be facilitated by considering the rarity of the population because, by construction, the more rare the population the lower the probability of obtaining a very high $\hat{\tau}_{IA}$.

Salehi and Seber (1997a) introduced a modification of ACS in which networks are sampled only once, i.e. WOR. They presented an unbiased version of the conventional PPS WOR estimator $\hat{t}_{Raj}$ of τ due to Raj (1956) along with its variance and an unbiased estimator of its variance. Subsequently, $\hat{t}_{Raj}$ was modified according to Murthy (1957) to yield another unbiased estimator $\hat{t}_M$ of τ along with its variance and an unbiased estimator of its variance. These two estimators were compared to the usual modified ACS Hansen-Hurwitz and Horvitz-Thompson estimators from the Introduction, their Rao-Blackwell estimators and the SRS estimator via a small population example for a fixed n_1. Results showed some interesting phenomena. Using $\hat{\mu}_{RBHH}$ resulted in a greater relative improvement (variance reduction) than for $\hat{\mu}_{RBHT}$, not a surprising result given the discussion in the previous section on the

work of Salehi (1999). Ignoring the Rao-Blackwell estimators, for the design with the networks selected WOR, $\hat{t}_M$ was more efficient than the other estimators. However, $\hat{\mu}_{RBHH}$ was the most efficient estimator overall for the example. Next, the authors analyzed one of the data sets from Smith et al. (1995) utilizing computer simulation. Comparisons were made among the estimators except for the Rao-Blackwell versions. Once again, $\hat{t}_M$ was more efficient than the other estimators and its efficiency gain over the next best estimator $\hat{\mu}_{HT}$ increased with n (number of sampled networks). Although the authors felt these results would hold for many populations after a review of the literature, they suggested further research. Unfortunately, one disadvantage of $\hat{t}_M$ is that it requires substantial computation. Finally, $\hat{t}_{Raj}$, and consequently $\hat{t}_M$, can be improved using the Rao-Blackwell method in conjunction with D_R. The reader is referred to Salehi and Seber (1997a, p. 212-214) for derivation of this estimator ($\hat{t}_{RBM}$) along with its variance and corresponding unbiased estimator.

Felix-Medina and Thompson (1999) combined ACS and conventional double sampling to come up with a new design called Adaptive Cluster Double Sampling (ACDS) utilizing a regression estimator. The reader is referred to the subsequent section in this chapter "Double Sampling and Multivariate ACS" for more discussion on this topic.

Order Statistics

The selection of c_a for criterion C may be difficult or impossible to ascertain in some survey situations. Perhaps the investigator also wants to search for high values of y in addition to estimating μ or τ. In such cases, the criterion for additional sampling may be made relative to the observed sample values by basing it on the sample order statistics. To illustrate, if there is an initial sample of size n_1, additional units would then be added to the sample in the neighborhoods of a predetermined number of sites, $n_1 - r$, with the largest order statistics of the variable of interest, $y_{(r+1)}, y_{(r+2)}, \ldots, y_{(n_1)}$, where $r \in \{0, 1, \ldots, n-1\}$. If the y_is of any of the additional units meets or exceeds $y_{(r+1)}$ in the initial sample, then we add its neighborhood too, i.e. c_a becomes $y_{(r+1)}$. Thus, this scheme differs radically from the previous schemes in which C is fixed and determined before the survey is begun. Using order statistics, the clusters are determined by the initial data rather than in advance (Thompson and Seber 1994).

The final sample average is a biased estimator of μ. However $\hat{\mu}_{ord}$, an unbiased estimator of μ, used in conjunction with the order statistics of an adaptive cluster sample, is the usual Hansen-Hurwitz estimator from the Introduction (see equation 1.12) (Thompson 1996; Thompson and Seber 1996). Because α_k (see equation 1.6) depends on the sample, it can be shown that $\hat{\mu}_{HT}$ is not unbiased when ACS is based on order statistics (Seber and Thompson 1994).

Thompson and Seber (1996) showed that $\hat{\mu}_{ord}$ is also the expectation of $\bar{y}_1$, the initial sample mean, conditional on a sufficient statistic D_f where:

$$D_f = \{(i, y_i, f_{is}) : i \in s_R\}, \tag{2.22}$$

f_{is} is the number of units from A_{is} included in the initial sample, and A_{is} denotes the network containing unit i for each unit i in the final sample. The subscript s is a reminder that the network configuration is now determined, in this case, by the sample. Thus, by the Rao-Blackwell theory of the Introduction, $\hat{\mu}_{ord}$ will be more efficient than $\bar{y}_1$ when both have sample size n_1.

Both the variance of $\hat{\mu}_{ord}$ and its unbiased estimator have been derived using conditional arguments. Unfortunately, this estimator is computationally complex and can be negative for some samples. Two biased estimators of the variance of $\hat{\mu}_{ord}$ exist as well that guarantee a nonnegative result (Thompson and Seber 1996, p. 167).

The Rao-Blackwell method can also be used to improve $\hat{\mu}_{ord}$, since $\hat{\mu}_{ord}$ obviously depends on the order in which the sample is selected, and therefore is not a function of D_R. In fact, it can be shown that the resulting Rao-Blackwell estimator, $\hat{\mu}_{RBord}$, is the expectation of either $\hat{\mu}_{ord}$ or $\bar{y}_1$ conditional on D_R, which we recall is a minimally sufficient statistic (Thompson and Seber 1996, p. 168). Thus, using the Rao-Blackwell theory from the previous section (equation 2.10), we have:

$$\hat{\mu}_{RBord} = \frac{1}{\xi}\sum_{g=1}^{G}\hat{\mu}_{ord,g}I_g = \frac{1}{\xi}\sum_{g=1}^{G}\bar{y}_{1g}I_g. \tag{2.23}$$

Explicit formulas for the variance of $\hat{\mu}_{RBord}$ and its unbiased estimator derived from equations 2.10 and 2.16 are given in Thompson and Seber (1996, p. 167-169).

Thompson and Seber (1996) worked through an example which consisted of drawing all possible SRSWORs from a small population and letting $c_a = y_{(n)}$. They then calculated the initial sample total, $\hat{\tau}_1$, $\hat{\tau}_{ord}$, and $\hat{\tau}_{RBord}$ for each sample and the variance of these estimators were calculated as well. Also, the variance of $\hat{\tau}_{ord}$ was calculated through the use of the three estimators mentioned previously (again, one unbiased and two biased). Interestingly, upon comparison, $Var[\hat{\tau}_{RBord}] < Var[\hat{\tau}_1] < Var[\hat{\tau}_{ord}]$, where $Var[\hat{\tau}_1]$ was calculated using $E[\nu]$ and $MSE[\widehat{Var}[\hat{\tau}_{ord}]]$ was lower for the two biased estimators of $Var[\hat{\tau}_{ord}]$ than for the unbiased estimator.

Double Sampling and Multivariate ACS

In some ACS surveys, there is more than one variable of interest. Previously, C depended on some function of a single variable. In the multivariate setting, C may depend on any one of the variables, all of the variables, the sum of all the variables, etc. Thompson and Seber (1996, chap. 8) extensively discussed multivariate aspects of ACS and the reader is referred to that monograph for theory and details. Per the monograph, unbiased estimates, $\hat{\mu}_{HH}$ and $\hat{\mu}_{HT}$ still exist for μ of each variable irrespective of the choice of C. Also, an unbiased estimate of the variance-covariance matrix of these estimates exists. Univariate results from the Introduction can be applied to each of the p-variables, even though C may depend on other variables.

The ACS strategy will be efficient relative to the conventional strategy if the within-network variation is large relative to the total variation. Analagous to the univariate case, the relative efficiency will ultimately depend on the partitioning of the population into networks, which, in turn, depends on the myriad of issues already discussed.

As a special case, we now examine surveys that make use of a rapid assessment variable. In this case, it may be desirable to base ACS on one variable, such as estimated abundance based on rapid assessment, in order to estimate τ for another variable, such as actual abundance. For example, Felix-Medina and Thompson (1999) combined ACS and conventional double sampling to come up with a new design called Adaptive Cluster Double Sampling (ACDS) appropriate when C (or C_x) is based on an easy-to-measure auxilliary variable x which is correlated with the difficult-to-measure (or more expensive-to-measure) survey variable y. They speculated that the final sample size of y-values could be reduced while obtaining a good estimate of μ_y. First, they defined the auxilliary variable to be a binary variable which takes the value 1 if it is speculated that $y \geq c_y$ and 0 otherwise, i.e. x is comprised of values that are quick estimates of the actual y-values. An ordinary adaptive cluster sample based on C_x is taken where, in this case, $c_x = 1$. Since x is easy-to-measure, n_1 can be large, thus increasing the probability of detecting clusters with elements satisfying C_x. Next, a conventional subsample of the networks is taken from the initial sample. From each of these subsampled networks, conventional and independent subsamples of units are taken and their y-values are measured. This sampling design allows the

number of y-values to be fixed beforehand. Also, when $x_i = 0$ for any adaptively added unit, the more expensive variable of interest y_i does not have to be measured for that (edge) unit. Thus, the total cost of sampling is reduced.

The goal is to use regression to model the relationship between x and y with the adaptive cluster sample based on x. Felix-Medina and Thompson (1999) constructed regression estimators $\hat{\mu}_R$ based on $\hat{\mu}_{HT}$ and $\hat{\mu}_R^*$, the solution of a system of estimating equations, with these estimators not necessarily being unbiased. They also presented asymptotic variances for these estimators along with estimators of their variances through use of the Delta Method. To compare the performance of ACDS with that of ACS, a simulation study was carried out. Two thousand SRSWORs were selected from a simulated population and three sampling designs were used. The first was the usual ACS design. The second design was ACDS I for which the maximum number of y-measurements to be recorded was set in advance. Subsampling was conducted after the adaptive sample based on x had been completed, thus implying that sampled networks would be visited twice. For ACDS II, the third design considered, the number of y-measurements was not set in advance. Subsampling was conducted as soon as the x-values in the networks were measured so that sampled networks would be visited once. Cost functions were constructed and defined for each design and initial sample sizes were determined so that the expected costs were equivalent. Results showed that the two ACDS designs yielded lower MSEs than the ACS design.

Two issues, however, were raised from the study that might affect the implementation of ACDS. First, the results showed that the performance of ACDS could be improved when n_1 was increased. This improvement occurs since the between-network variance was high for the particular simulated population, in which case the best strategy is one that emphasizes sampling more networks. Second, the performance of the ACDS design was very poor when the study was repeated on a data set from Smith et al. (1995). The problem is that the ACDS estimator is not resistant to outliers, i.e. it is seriously and negatively affected by high within-network variation. In both cases, the researcher would therefore have to have some prior knowledge of the nature of the decomposition of the total variance of the population to set appropriate sample sizes for each phase.

Before leaving this section, it should be noted that for a mean vector $\boldsymbol{\mu}$ (or total vector $\boldsymbol{\tau}$) there exist multivariate estimators based on the Rao-Blackwell method. Not surprisingly, these are simply extensions of the univariate case discussed earlier in this chapter. The reader is referred to Thompson and Seber (1996, p. 206-207) for a straightforward derivation of the variance-covariance matrix for the Rao-Blackwell estimators of $\boldsymbol{\mu}$ along with the unbiased estimator of this matrix.

Detectability in ACS

In the discussion so far, it has been assumed that y_i is recorded without error for each unit in the sample. Clearly, this might not be a realistic assumption. The

probability that an object in a selected area is observed is called its *detectability* (Thompson 1992). Imperfect detectability is a prevalent source of nonsampling error in many surveys which will lead to an underestimate of τ. For example, Smith et al. (1995) admitted that detectability was an issue in their research. To estimate τ in a survey with imperfect detectability, both the sampling design and the detection probabilities must be taken into account. The complexity of this situation is more difficult with ACS designs than with conventional designs because the probability of inclusion of an object depends not only on its detection probability but also on the detection probability of objects in nearby units. Fortunately, results that adjust for detectability and extensions of this theory to ACS for a variety of cases have been worked out in a straightforward fashion (Thompson and Seber 1994, 1996). Two main ideas are utilized. First, an estimator of τ, unbiased under the assumption of perfect detectability, can be adjusted by the detection probability or its estimate. Second, for derivation purposes, expectations and variances with respect to the initial sample conditional on the detectability are taken followed by expectations with respect to detectability. Cases include one where the probability of detection, g, is assumed to be known, constant, and independent for each object in the population and one where g is constant, unknown, and has to be estimated. Additional cases include modification of the previous two cases such that $g = g_{ij}$, i.e. the detection probability is not constant and can vary among j objects in i units in the population. Obviously, the level of complexity increases accordingly in these latter situations. The reader is

referred to Thompson and Seber (1996, chap. 9) where the appropriate estimators, their variances and estimated variances have been derived and the theory is discussed.

It should be added that the effect of imperfect detectability on the relative efficiency of simple random sampling and ACS would be to add an identical amount to the variance under each strategy, so that the relative efficiency, whether greater than or less than 1, would move slightly closer to 1 (Thompson and Seber 1996).

Applications of Bootstrapping

Confidence intervals for τ are typically based on asymptotic normality of the sampling distribution. However, for any ACS design, the sampling distribution of $\hat{\tau}$ can be very asymmetric (Christman 1997; Christman and Pontius 2000) and discontinuous (Christman 2000). Thus, a confidence interval based on the assumption of normality is inappropriate. As an option, bootstrapping can be used to generate an empirical sampling distribution for $\hat{\tau}$ and interval estimates can be computed.

Pontius and Christman (1997) noted that any bootstrapping procedure would have to take into account the fact that the population size is finite and attempted to address this by modifying the bootstrapping algorithm of Sitter (1992a,b). They conducted a small ACS simulation study and compared three types of confidence intervals for τ using their modified finite population bootstrapping algorithm in conjunction with $\hat{\tau}_{HT}$. Sitter's bootstrap algorithm worked reasonably well for the specific population and the network sizes used which yielded bootstrapped estimates of τ that were,

in general, close to the sample estimates. However, the authors cautioned that the network sizes were small relative to N and that the variability in interval estimates might increase with larger networks. All three types of confidence intervals yielded minimal biases and similar interval widths for all given sample sizes. Additionally, all three types of confidence intervals yielded estimated coverages slightly less than the actual coverage. The confidence interval based on the normal approximation had, not surprisingly, the worst coverage for all sample sizes, whereas the bias-corrected and accelerated confidence interval always had the best coverage.

Christman and Pontius (2000) addressed these aforementioned topics using $\hat{\tau}_{HH}$. They did not, however, consider bootstrapping interval estimators based on $\hat{\tau}_{HT}$ due to theoretical complications arising from an unequal probability sampling design that consists of a combination of with- and without-replacement selections for networks. Four nonparametric, finite population bootstrap methods and three types of confidence intervals were investigated where the input sample consisted of n_1 network totals. The results depended on the type of population being simulated and n_1. Results also were obtained using the data from Smith et al. (1995). Bootstrap percentile intervals generally provided better results when compared to the aymptotic normal and bias-corrected and accelerated intervals. Moreover, the authors recommended using the Bootstrap With Replacement method of McCarthy and Snowden (1985) with percentile intervals.

Di Battista and Di Spalatro (1999) presented a resampling procedure, which they called the Bootstrap for Adaptive Cluster Sampling (BACS), for estimating the bias of the sample mean and sample variance computed on the final sample produced by ACS when the initial sample is a SRSWR. They showed through proof and a simulation study that the BACS method is consistent, i.e. as $n \to \infty$, bias estimated via the BACS method converges to the actual bias generated via ACS.

Optimality

By definition, a sampling strategy for estimating or predicting a population quantity $z(\mathbf{y})$ consists of a sampling design $P(s|\mathbf{y})$ and an estimator $\hat{z}(d_s)$. It was stated in the Introduction that no UMVUE exists for the class of all adaptive and conventional designs in the fixed-population setting. Consequently, and as we have seen throughout this paper, there is no optimal strategy in the fixed-population strategy such that:

$$Var[\hat{z}_{opt}(d_s)|\mathbf{y}] = \min_{(P,\hat{z})\in B} Var[\hat{z}(d_s)|\mathbf{y}] \tag{2.24}$$

for all $\mathbf{y} \in \mathcal{Y}$, where B is the class of all design-unbiased strategies (Thompson and Seber 1996).

Limiting Sampling Effort and Cost in ACS

In summary, we have seen that the situation in which ACS would be least efficient relative to simple random sampling is a population in which the units satisfying C are

all separate from each other, so that the within-network variance is low and adaptive sampling adds only edge units, which increase the sample size but do not contribute to the precision of estimators (Thompson 1994). Much attention has been given to devising ways to limit the (random) size of the final sample in ACS. This can be particularly important from a cost standpoint. Although the final sample size can be somewhat controlled by the appropriate selection of the factors discussed so far, prior knowledge of the population structure is required. Sometimes this knowledge is not available and prior planning of resources is necessary.

On occasion, when c_a is set too low, it becomes clear that continued sampling would lead to the addition of more units than time or cost allow. If sampling were curtailed prematurely, biases may arise. We have already discussed one remedy, i.e. the use of order statistics. Nonetheless, Thompson (1994) offered the following additional guidelines to deal with this type of situation where the need may arise to limit sampling effort.

The study region can be stratified ahead of time, with the decision whether to use ACS and what C should be for any subsequent stratum based on the results for the strata already surveyed. In each stratum, design-unbiased estimators of τ_h and its variance are calculated depending on the design used. The stratified design-unbiased estimate of τ will be the sum of the individual τ_hs with an analogous result holding for the variance assuming independence between strata. As previously mentioned in this chapter, an efficient design would be one where the study region is stratified into two

strata such that one stratum contains most or all of the rare elements and the other contains few or none of the rare elements. Christman (2000) gives an example of such a design where SRSWOR is used in the "common" stratum and ACS in the "rare" stratum. Unfortunately, this type of stratification is not always possible beforehand. In this same context, poststratification could be used as a design-unbiased procedure although Christman (2000) discouraged this practice. For example, one stratum could represent the part of the population where only the conventional initial sample was completed, while another stratum could represent the part of the population where enough resources existed such that ACS was able to be conducted. Perhaps C might have been modified several times during the course of the study as more results became available. Then, the region would be poststratified into strata based on the modification of C.

The author would be remiss in neglecting to note that the choice of strata sample sizes is in itself a substantive and important topic. The reader is encouraged to consult Christman (2000) for an overview and Thompson and Seber (1996, chap. 7) for a detailed discussion of adaptive allocation strategies in stratified sampling.

Finally, another guideline would be to modify the neighborhood definition. Neighborhoods could be truncated at some predefined region or block boundary to limit the size of networks. Also, defining the neighborhood to be a non-contiguous systematic grid will constrain networks to be spatially sparse.

Brown (1994) and Brown and Manly (1998) presented a restricted ACS design that allows the final sample size to be defined prior to sampling. Units for the "initial" sample are sequentially selected until a desired final sample size is reached. Moreover, the neighborhood of a selected unit is sampled if C is met. If the cumulative total sample size is below a defined number of units, another initial unit is selected. If this unit, and associated units in the cluster, result in the cumulative total sample size exceeding the defined limit, the cluster is included but sampling ceases. This methodology restricts the selected initial sample to one that produces a final sample size at or just over the defined limit and has the effect of reducing variation in the final sample size. However, using $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$, from the previous chapter, for this design results in a positive bias with the severity of the bias depending on the population structure. In some certain types of populations this bias can be accurately estimated via bootstrapping and adjustments can be made to the estimators. Using the usual equations for estimating the variance of both $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$ under the restricted ACS setting yielded biased, but in all cases, "reasonable" results (Brown 1994; Brown and Manly 1998).

In a simulation study across an array of population models, Brown (1994) showed results that generally favored restricted over unrestricted ACS for patchy populations in terms of MSE. However, in a similar and more extensive study, Brown and Manly (1998) discovered that restricted ACS was worse than unrestricted ACS in terms of relative efficiency for all population models under consideration!

One potential drawback of restricted ACS is that if c_a is set too low, then all the units selected will be from the same cluster since the final sample size will be immediately obtained. Poor coverage of the study area will be the result. Again, prior knowledge of the population structure is very helpful. Another issue of concern is that prior development of a travel plan between units in the initial sample cannot be done. If travel between units is difficult, then this might be a significant problem. For some studies, a possible solution might be to stratify the study area and set a desired final sample size within each stratum (Brown and Manly 1998).

Salehi and Seber (1997b) described a two-stage ACS design that used an SRSWOR of primary units, and then a sub-SRSWOR of secondary units within each of the selected primary units. Two variations on this design were proposed. The first was a scheme where the clusters were allowed to cross primary unit boundaries while in the second scheme, the clusters were truncated at the boundaries. The estimators and their variances from the Introduction were modified for both schemes. Their unpublished results suggested that the nonoverlapping scheme appeared to be more efficient in terms of estimation, although the comparison was dependent on the population structure. The authors also stated that they preferred the use of their Horvitz-Thompson estimator. However, Felix-Medina and Thompson (1999) cautioned that if the clusters are large relative to the size of the primary units, then only their portions that are intersected by the primary units in the initial sample are sampled and a reduction in efficiency of this sampling design would be expected.

When the sampling fraction of primary units was small, Salehi and Seber (1997b) constructed an upper bound on $Var[\hat{\mu}_{HH}]$. They went on to derive a formula for the number of primary units, m, to be sampled to achieve the desired upper bound on $Var[\hat{\mu}_{HH}]$. Statistical input for the formula came from an initial pilot study. They also developed a cost function for the nonoverlapping scheme. Using m in the cost function would allow the investigator to plan for the total cost of the survey for a given desired level of precision.

In an effort to control and plan sample size, Salehi and Seber (1997a) introduced a modification of ACS in which networks are sampled only once, i.e. WOR. They claimed that if the main expense is the cost of travelling to the sites of the initial sample units, then cost could be controlled better if the number of networks, ν, was fixed a priori. Christman and Lan (1998) considered ACS based on sequentially selecting initial sampling units until some predefined number of non-zero Y-values are obtained in order to estimate τ. The reader is referred to the previous section in this chapter "Other Estimators" for more discussion of these two papers. Felix-Medina and Thompson (1999) combined ACS and conventional double sampling to generate a new design called Adaptive Cluster Double Sampling (ACDS) utilizing a regression estimator. They speculated that the final sample size and total cost could be reduced while obtaining a good estimate of μ. The reader is referred to the previous section in this chapter "Double Sampling and Multivariate ACS" for more discussion on this topic.

Some Problems With Current Thinking

There are problems associated with the manner in which several of the studies cited throughout the last two chapters were conducted. Consequently, results and subsequent conclusions generated from the studies may be suspect. This is particularly true for Horvitz-Thompson estimation which has not been so nearly and neatly theoretically characterized as has been Hansen-Hurwitz estimation.

The first, and what the author feels is the most important problem is the widespread use of "one-at-a-time" experimentation. Although several studies have examined multiple factors at once, the researchers often assume these factors do not interact and the possible existence of significant interaction effects are simply ignored. Even when the possibility of two factors interacting is tacitly acknowledged, no attempt is made to quantify the interaction effect in any meaningful way. Perhaps this is done to facilitate the ease of the data analysis. For whatever reason, we will see that any attempt to characterize the effect a variable might have on relative efficiency will be dependent on other variables.

The simulation approach employed in this dissertation differs from the approaches used in the previously cited research with respect to protocol and the generation of populations. Because of this lack of uniformity of simulation approaches and because it is known that the sampling distributions of the ACS estimators can be quite skewed and variable, there can be a significant impact on the conclusions drawn, especially if the amount of replication is small. We also note in our research the use of the actual

true variances, $Var[\hat{\tau}_{HT}]$ and $Var[\hat{\tau}_{HH}]$, as opposed to estimating them as has often been done.

It is not surprising that upon reading the introductory primer presented in Chapters 1 and 2, the reader is left with the impression that the conclusions from a variety of studies appear to contradict one another. Consequently, an approach was sought that would allow for more quantifiable and reproducible results regarding relative efficiency of ACS. An ideal tool for this approach is the analysis of a designed experiment via response surface methodology. In such a fashion we can accomplish the following goals:

(1.) We would be able to characterize and describe those conditions invoking change in relative efficiency, and we would also be able to describe the change itself in terms of direction and magnitude. Using grid searches, we then attempt to describe those conditions which optimize relative efficiency and to provide fitted models useful to the applied practitioner or researcher.

(2.) Using a designed experiment approach allows us to simultaneously look at many more variables than has been done in prior research projects. Significant interaction and quadratic effects can also be characterized and the corresponding response surfaces can be catalogued.

(3.) The relative efficiency of both the Horvitz-Thompson and Hansen-Hurwitz estimators will be compared and contrasted with respect to an array of variables known or suspected to affect relative efficiency.

(4.) Several simulation approaches used in the literature will be compared and contrasted to the one used in our research.

CHAPTER 3

METHODS

Multiple factors known to influence the efficiency of ACS were considered as candidates for input into the experimental design. *Controllable factors* are defined to be those factors that can be set by the researcher before conducting the survey. *Fixed covariates* are defined to be those factors that cannot be set by the researcher before conducting the survey, but can be fixed for the purposes of this analysis. Finally, *random covariates* are defined to be those factors that cannot be set by the researcher, neither before conducting the survey nor for the purposes of this analysis. Table 1 gives a short summary of the actual factors selected and used in the study. Levels for the controllable factors and the fixed covariates were chosen after careful consideration of all previously conducted simulations found in the literature cited throughout the Introduction.

We digress briefly to discuss the random covariate, Rarity, or the rarity index, $\frac{E[\nu]}{n_1}$. It is clearly seen from Equation 2.9 that:

$$\lim_{N\to\infty} \frac{E[\nu]}{n_1} = 1 \tag{3.1}$$

It was discussed in Chapter 2 that the literature had suggested that relative efficiency of ACS is highest when $\frac{E[\nu]}{N}$ is relatively close to $\frac{n_1}{N}$. Consequently, a diagnostic for an efficient ACS design would be one whose rarity index was close to 1. Note the rarity

Table 1. Factors used in the study.

Factor Name	Factor Description	Factor Type	Factor Levels	Units
Area	square study area dimension, $r(c)$	controllable	10, 20, 30	units
Critical	critical value, c_a	controllable	1, 2	elements
Sample	initial sample size, n_1	controllable	10, 30, 50	units
Shape	shape of cluster, $\frac{\sigma_y^2}{\sigma_x^2}$	fixed covariate	1, 4	
Direct	within-cluster correlation, ρ	fixed covariate	0, .7	
Spread	absolute spread of cluster, σ_x	fixed covariate	.1, .5	units
Parents	number of parents, λ_p	fixed covariate	1, 3, 5	clusters
Kids	number of offspring, λ_o	fixed covariate	10, 30, 50	elements
Pervarw	proportion within-network variability, $\frac{\sigma_W^2}{\sigma^2}$	random covariate		
Predp	predicted number of parents	random covariate		clusters
Numnets	number of study area networks	random covariate		networks
Rarity	rarity index, $\frac{E[\nu]}{n_1}$	random covariate		

index is a relative measure of "closeness" between these two fractions. Unfortunately, taking the difference between these two fractions is an absolute measure. Using the rarity index avoids those situations where two different populations might have an equal difference between the initial and final sample sizes, yet relatively speaking, the expected effective sample size might be much "closer" to the initial sample size for one population than the other. For example, consider $n_1 = 5$ and $E[\nu] = 10$ for one population and $n_1 = 85$ and $E[\nu] = 90$ for the other population where both populations consist of $N = 100$ units.

All programming to implement the simulations was done in R, version 1.6.2. Jeffrey Pontius of Kansas State University, Mohammad Salehi Marzijarani of Isfahan University of Technology, Roger Bivand of the Norwegian School of Economics and Business Administration, and Giovanni Petris of the University of Arkansas provided useful background functions and/or recommendations. Additionally, Jim Robison-Cox and Jeffrey Banfield of Montana State University provided useful assistance in writing code. All functions used in the programming to implement the simulations are included in Appendix A. Some of the functions have been "commented" in order to expedite the flexibility of the simulations. A description of the purpose of the function and its arguments is included in each function header. Also, the functions are listed in sequence of usage in the program. Verification of the accuracy of the program was done by running the program using known data sets, most notably the ringneck data set from Smith et al. (1995) and the small example from Thompson (1990).

Rowlingson and Diggle (1993, 1996) discussed Splancs, an enhancement package to S-Plus consisting of a set of functions used for the display and analysis of spatial point pattern data. Splancs was adapted and packaged for R by Bivand and Gebhardt (2000) and is documented in the R help library. This package was used as a tool in writing the programming for the simulations conducted here.

Departure of the objects from complete spatial randomness towards a clustered spatial point pattern from a Poisson cluster process was discussed in Buzas and Hayek

(1997). We verified this departure through visual inspection of a graph of the study region and by insuring the ratio of the population variance to the population mean was well beyond 1. For each of 72 trials in the experimental design, 250 populations were generated as realizations of a variation of the Neyman-Scott process, a special case of a Poisson cluster process. A discussion of the Neyman-Scott process is given in Cressie (1993, p. 661-666). The choice of 250 will be discussed shortly. First, a bounded region B was defined. A 0-truncated distribution generated from a spatially homogeneous Poisson process with mean λ_p was used in order to randomly generate numbers of parent or cluster events. The two-dimensional location of each parent was randomly generated using two independent uniformly distributed realizations within B. The variation of the Neyman-Scott process lies in the fact that Cressie (1993) used a spatially nonhomogeneous Poisson process in order to generate the parent locations. For each parent, independent random numbers of offspring were drawn from a 0-truncated distribution generated from a spatially homogeneous Poisson process with mean λ_o. The location of each offspring was independently and randomly generated around its parent according to a bivariate normal distribution with the mean centered on the parent's location and a covariance matrix corresponding to the trial settings.

In order to avoid what is known as an edge effect, the parent generating process was conducted on a region B containing a smaller region A. Only those offspring that fell into region A were used to generate the population. In other words, A became the study area. Offspring falling into region A from parent events outside of A were not

ignored (the edge effect) but included in the population as part of a smaller cluster. A vector of population y-values, $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}'$, was then obtained by dividing region A into N equal sized units and counting the number of offspring in each unit. The vector is formed into a matrix with dimensions depending on the number of rows and columns in region A. Several different approaches exist in the literature regarding the handling of the edge effect and they will be discussed in further detail in the Discussion section.

At this point, there are a couple of important details to mention. Unlike Smith et al. (1995), we have fixed the size of the units to have dimension 1 x 1 even though the size of the study area is allowed to vary in our study, Furthermore, the spread of the cluster is an absolute measure, not a relative measure. In other words, the spread of the cluster is not relatively scaled to the size of the study area.

To ascertain the impact of different simulation approaches on relative efficiency measures, two items of interest will be examined. First, a portion of an experimental study by Christman (1996a) is replicated using the simulation approach described here and 500 populations are generated with the settings in Table 2, in order to compare the results between studies. Second, hypothetical populations are constructed and results under three different simulation approaches are compared.

The R code for the *genpop* function is located in Appendix A. In order to better understand the study simulation process, the following simple example (Figure 1)

Table 2. Settings used to replicate study of Christman (1996a).

Shape	Direct	Spread	Parents	Area	Critical	Kids	Sample
1	0	1	1	10	1	50	10

demonstrates how the *genpop* function is used. The discussion proceeds in order of operational implementation within the function.

We first call the *genpop* function with inputs $\lambda_p = 1$, $\lambda_o = 3$, $\sigma_x^2 = .1$, $\sigma_y^2 = .1$, and $\sigma_{xy} = 0$ where x and y are coordinates in the Cartesian plane. Remaining inputs define region B to have dimension 4 x 4 and region A to have dimension 2 x 2 where each unit is a 1 x 1 square. The coordinates that define the boundary of region B are $(-1,-1),(3,-1),(3,3)$ and $(-1,3)$, while the coordinates that define the boundary of region A are $(0,0),(2,0),(2,2)$ and $(0,2)$.

We take a moment to note here three issues relating to the study. First, σ_y^2 is determined by Spread and Shape. Specifically, σ_y^2 is either 0.01 or 0.04 given σ_x^2 is equal to 0.01, or σ_y^2 is either 0.25 or 1.00 given σ_x^2 is equal to 0.25. Second, the covariance input, $\sigma_{xy} = \rho\sigma_x\sigma_y$, is determined by Direct. Specifically, $\sigma_{xy} = 0.000$ given Direct is equal to 0.00 or σ_{xy} will equal one of 0.007, 0.014, 0.175 or 0.350 given Direct is equal to 0.70. Third and finally, the number of population units are determined by Area, i.e. N will equal one of 100, 400 or 900, given Area is equal to 10, 20 or 30 respectively.

A number of outputs can be obtained. Because we are simulating a spatially homogeneous Poisson process:

$$\lambda_p = \mu_p a \tag{3.2}$$

where μ_p is a constant intensity parameter and a is a defined area. Thus, for our example, if $\lambda_p = 1$ and the area of region A is 4 units then μ_p equals 0.25 clusters per unit. Next, *genpop* takes μ_p and calculates λ_p^* for region B because the process is spatially homogeneous. Thus, for our example, $\lambda_p^* = (.25)(16) = 4$ clusters because region B has an area of 16 units.

A random value is then obtained from a 0-truncated distribution generated from a spatially homogeneous Poisson process with mean λ_p^*. Because a set of numbers is required by the R function as input to do this sampling, a discrete domain with an upper bound of 1,000 standard deviations was arbitrarily chosen where Chebychev's Inequality assures us that the choice of such an upper bound is more than adequate. The 0-truncated Poisson standard deviation for the count of clusters in region B is:

$$\sqrt{\frac{\lambda_p^{*2}+\lambda_p^*}{1-\exp(-\lambda_p^*)} - \left(\frac{\lambda_p^*}{1-\exp(-\lambda_p^*)}\right)^2}$$

Six parents were randomly selected in region B for our example, three of which fell into region A. Two independent realizations are obtained from a uniform distribution with parameters -1 and 3 for each parent and the subsequent locations of the parents within region B are printed. In our example, the R output generated is displayed in Table 3.

Table 3. Cluster center locations.

	Location	
Parent	x	y
1	0.4604359	1.3898835
2	1.7439836	1.6698143
3	1.1893717	0.1771077
4	1.9276435	-0.6523074
5	-0.8657158	0.1679549
6	0.7506571	-0.4235404

These locations are denoted with the symbol '+' in Figure 1. Given this actual number of parents, we use Equation 3.2 to update our intensity parameter and solve $6 = (\mu_p^*)(16)$ to yield $\mu_p^* = .375$. Thus, we would have predicted that our study area, region A, contained $(.375)(4) = 1.5$ parents. This is the predicted number of parents factor referred to in Table 1.

Next, for each parent, a random value is taken from a 0-truncated distribution generated from a spatially homogeneous Poisson distribution with mean λ_o and the upper bound on the domain set to 1,000 standard deviations. For our example, in region B, the numbers of offspring are 3, 2, 5, 1, 1, and 1 (for a total of 13) for parents 1 through 6, respectively.

Using the *genpop* inputs, the appropriate covariance matrix for offspring location is input and the locations of the offspring are generated. In our example, the R output generated is displayed in Table 4.

Of the 13 offspring in region B, the *genpop* function reports that 7 offspring lie in our study area, i.e. $\tau = 7$. Of course this is also easily verified upon inspection

Table 4. Original offspring locations.

	Location	
Offspring	x	y
1	0.3651039	0.85697650
2	0.1976724	1.48334720
3	0.2061313	1.36428806
4	1.9593322	2.11580510
5	2.4503648	1.66633790
6	1.5113465	0.41102936
7	1.4328085	-0.20994952
8	1.1178397	0.20209536
9	1.1233880	0.02632157
10	1.5698909	0.44219412
11	2.1255528	-0.47803488
12	-0.6778012	-0.12203441
13	0.6875375	-0.13082528

of the offspring locations, denoted with the symbol 'O' in Figure 1. The code then forms the appropriate grid of units, generates the matrix of population y-values, and displays the matrix in Table 5.

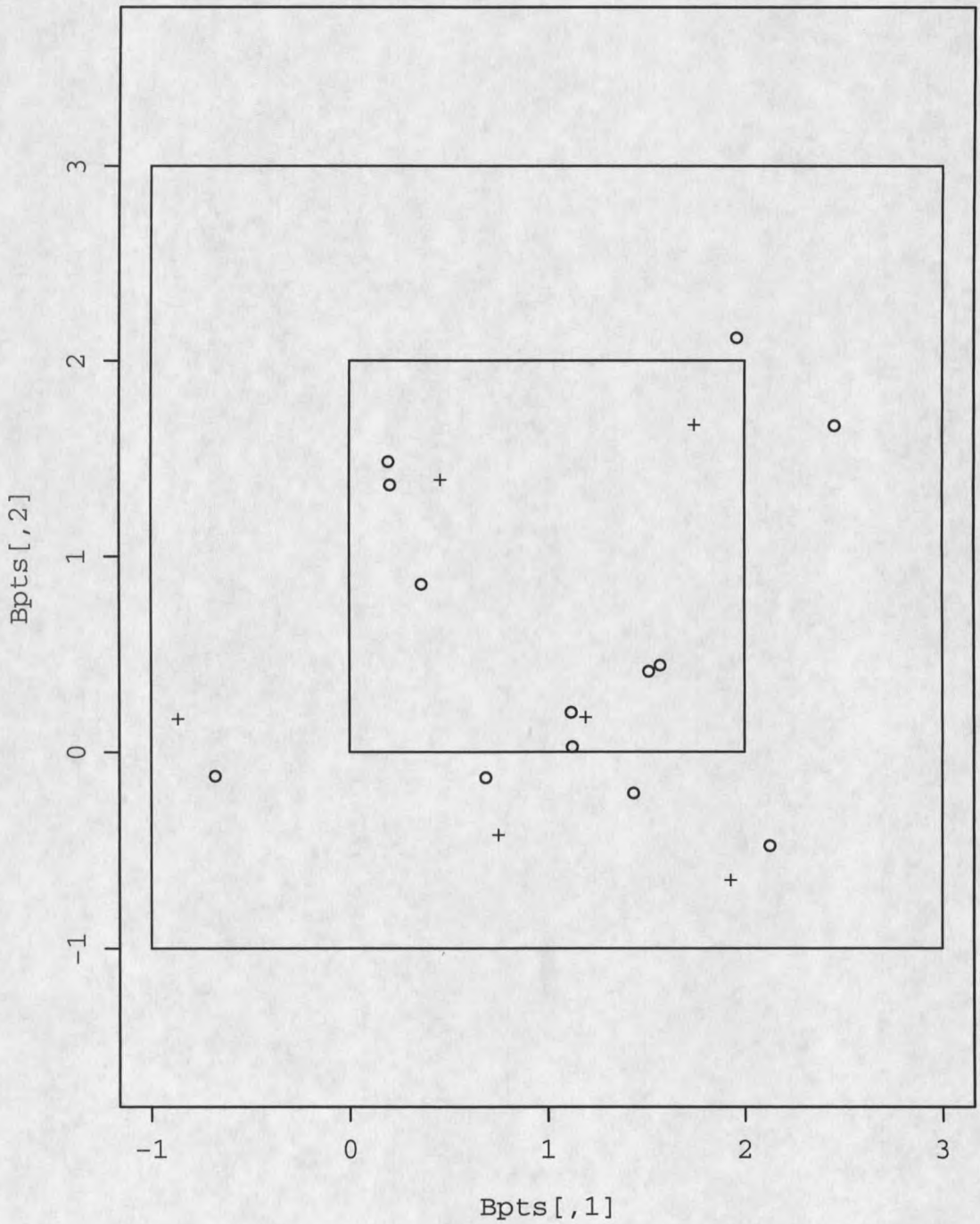
Table 5. Matrix generated for example population (Figure 1).

2	0
1	4

The last portion of code in the *genpop* function was inserted for the rather rare case where only one or just a few parent locations are generated in the study area but all the offspring are located outside the study area. In this event, if there are

no offspring detected within the study area, then the simulation is terminated and *genpop* is reinitialized.

Figure 1. An Example Population.



The Initial Study and Experimental Setup

An initial study was done whereby multiple populations were generated for each of three different types of settings. The settings or classes of population are listed in Table 6.

Table 6. Settings used in the initial study.

Setting	Shape	Direct	Spread	Parents	Area	Critical	Kids	Sample
Nominal	1	0.0	0.1	5	10	2	50	10
Extreme 1	4	0.0	0.5	5	10	1	50	30
Extreme 2	1	0.0	0.1	1	30	1	10	30

Using the factor levels for this study (Table 1), the Extreme 1 setting emulates a population with many offspring dispersed throughout the smallest (10 x 10) study area while the Extreme 2 setting emulates a population with few offspring clustered in the largest (30 x 30) study area. The Nominal setting was chosen to emulate what was felt to be an intermediate population. There were two purposes motivating this initial study. First and foremost was to empirically assign a reasonable number of populations to be generated per experimental trial for the designed experiment. This was done by generating 500 populations at a setting. A size of 500 was chosen as an arbitrary upper bound after a thorough review of the simulation history in the literature. Also, computer resources were freely available. A plot of the accumulating

median, henceforth known as an M_a plot, was made for each of a variety of responses against the sequence of $N = 500$ populations. The choice and use of the median will be discussed further in the Discussion. Here, the accumulating median, M_a, is defined to be:

$$M_a = \begin{cases} X_{(\frac{a+1}{2})} & a \text{ is odd} \\ (X_{(\frac{a}{2})} + X_{(\frac{a+2}{2})})/2 & a \text{ is even} \end{cases}$$

where X is one of the response measurements, and $a = 1, 2, \ldots, 500$ populations. Thus, we have one M_a corresponding to each of the 500 populations as they are sequentially generated. Each plot was examined to see at which population the graph appeared to flatten out or "stabilize". When visually appropriate, the y-axis was assigned approximate, albeit crude, limits of $\pm$ 3 sample standard deviations of the 500 response values.

The second purpose of the initial study was to examine and compare the distributions of a variety of responses at the three different population settings through the use of numerical summaries and density histograms. To allow for better comparison, when visually appropriate, the density histograms for each setting were standardized with respect to the scale of both the x and y-axes.

To avoid problems with multicollinearity, the experiment and analysis were split into two parts (henceforth called the FCA, or fixed covariate analysis, and the RCA, or random covariate analysis). In the FCA, the experimental design was run using both the controllable factors and fixed covariates. In the RCA, realizations of the

random covariates were obtained. For both the FCA and RCA, two responses were obtained:

(i.) the relative efficiency for the Horvitz-Thompson estimator of τ (RE_{HT})

(ii.) and the relative efficiency for the Hansen-Hurwitz estimator of τ (RE_{HH})

RE_{HT} and RE_{HH} were calculated using Equation 1.18 where the appropriate expected effective sample size $E[\nu]$, under ACS, was used in lieu of n in the formula for $Var[\hat{\tau}_{SRS}]$. One of the major objectives for this experiment was to generate fitted response surface models and then use these models to determine those areas, or factor settings, where RE_{HT} and RE_{HH} are maximized.

The experimental design was a 72 trial design generated via Proc Optex in SAS using the D-optimality criterion option. Seventy-two trials were chosen because this ensures a good G-efficiency and was similar to the number of trials used in the more conventional central composite design. The experimental design matrices for both analyses were centered and scaled between -1 and $+1$. For each of the 72 trials, 250 populations were generated. For each population, the four random covariate values and the two relative efficiencies were obtained. Next, the median value for each of these six variables was calculated from the data for the 250 populations. These values correspond to our use of the terms Pervarw, Predp, Numnets, Rarity, RE_{HH} and RE_{HT}.

The design matrix of uncoded settings for the FCA is shown in Table 7. SAS reported a G-efficiency of 82.3772%. The resulting uncoded design matrix for the

RCA and the data for both responses, RE_{HH} and RE_{HT}, is shown in the results. The analysis was performed using SAS, version 6.12, and some representative examples of accompanying SAS code are contained in Appendix B.

Table 7: Design matrix of uncoded settings used in the FCA

Trial	Shape	Direct	Spread	Parents	Area	Critical	Kids	Sample
1	1	0.0	0.1	1	10	1	50	10
2	1	0.0	0.1	1	20	2	10	10
3	1	0.0	0.1	1	30	1	10	30
4	1	0.0	0.1	1	30	2	50	50
5	1	0.0	0.1	5	10	1	10	50
6	1	0.0	0.1	5	10	2	10	30
7	1	0.0	0.1	5	20	2	30	50
8	1	0.0	0.1	5	30	1	50	10
9	1	0.0	0.5	1	10	1	10	10
10	1	0.0	0.5	1	10	2	30	50
11	1	0.0	0.5	1	30	1	50	50
12	1	0.0	0.5	3	20	2	50	30
13	1	0.0	0.5	3	30	2	10	10
14	1	0.0	0.5	5	10	1	50	10
15	1	0.0	0.5	5	10	2	50	50
16	1	0.0	0.5	5	30	1	10	50
17	1	0.0	0.5	5	30	2	30	10
18	1	0.7	0.1	1	10	1	50	50
19	1	0.7	0.1	1	10	2	10	50
20	1	0.7	0.1	1	30	1	50	10
21	1	0.7	0.1	1	30	2	10	50
22	1	0.7	0.1	3	10	2	30	10
23	1	0.7	0.1	3	30	1	30	50
24	1	0.7	0.1	5	10	2	50	50
25	1	0.7	0.1	5	20	1	10	10
26	1	0.7	0.1	5	30	2	50	10
27	1	0.7	0.5	1	10	1	10	50
28	1	0.7	0.5	1	10	2	50	10
29	1	0.7	0.5	1	30	1	10	10
30	1	0.7	0.5	1	30	2	30	30
31	1	0.7	0.5	5	10	1	30	30
32	1	0.7	0.5	5	10	2	10	10

continued on next page

Table 7: continued from previous page

Trial	Shape	Direct	Spread	Parents	Area	Critical	Kids	Sample
33	1	0.7	0.5	5	30	1	50	50
34	1	0.7	0.5	5	30	2	10	50
35	4	0.0	0.1	1	10	1	50	50
36	4	0.0	0.1	1	10	2	50	30
37	4	0.0	0.1	1	20	1	10	50
38	4	0.0	0.1	1	30	1	30	10
39	4	0.0	0.1	3	10	2	10	50
40	4	0.0	0.1	3	30	1	50	10
41	4	0.0	0.1	5	10	1	10	10
42	4	0.0	0.1	5	10	2	50	10
43	4	0.0	0.1	5	30	1	50	50
44	4	0.0	0.1	5	30	2	10	10
45	4	0.0	0.1	5	30	2	10	50
46	4	0.0	0.5	1	10	2	10	10
47	4	0.0	0.5	1	20	1	50	10
48	4	0.0	0.5	1	30	2	10	50
49	4	0.0	0.5	1	30	2	50	10
50	4	0.0	0.5	3	10	1	50	50
51	4	0.0	0.5	5	10	2	10	30
52	4	0.0	0.5	5	20	1	30	50
53	4	0.0	0.5	5	30	1	10	10
54	4	0.0	0.5	5	30	2	50	50
55	4	0.7	0.1	1	10	1	10	10
56	4	0.7	0.1	1	20	2	50	10
57	4	0.7	0.1	1	30	2	10	10
58	4	0.7	0.1	1	30	2	50	50
59	4	0.7	0.1	5	10	1	50	50
60	4	0.7	0.1	5	10	2	10	50
61	4	0.7	0.1	5	20	1	50	10
62	4	0.7	0.1	5	30	1	10	50
63	4	0.7	0.5	1	10	1	50	10
64	4	0.7	0.5	1	10	2	10	30
65	4	0.7	0.5	1	10	2	50	50
66	4	0.7	0.5	1	30	1	10	50
67	4	0.7	0.5	1	30	1	50	50
68	4	0.7	0.5	3	20	2	10	10
69	4	0.7	0.5	5	10	1	10	50
70	4	0.7	0.5	5	10	2	50	10

continued on next page

Table 7: continued from previous page

Trial	Shape	Direct	Spread	Parents	Area	Critical	Kids	Sample
71	4	0.7	0.5	5	30	1	30	10
72	4	0.7	0.5	5	30	2	50	30

The Fixed Covariate Analysis (FCA)

The initial general linear model used for the FCA was a regression model containing all eight predictor variables and their pairwise interactions. Additionally, quadratic terms were included for those factors containing three levels, namely: number of parents, number of offspring, study area, and initial sample size. Thus, including the intercept, there are 41 terms in the initial model

$$RE_{\cdot} = \beta_0 + \sum_{i=1}^{8} \beta_i F_i + \sum\sum_{i<j} \beta_{ij} F_i F_j + \sum_{k} \beta_{kk} F_k^2 + \epsilon \tag{3.3}$$

where $i = 1, \ldots, 8$, $j = 1, \ldots, 8$, $k = 1$, 3, 7, and 8, and

$$F_1 = \text{Area}, \qquad F_2 = \text{Critical},$$

$$F_3 = \text{Sample}, \qquad F_4 = \text{Shape},$$

$$F_5 = \text{Direct}, \qquad F_6 = \text{Spread},$$

$$F_7 = \text{Parents}, \qquad F_8 = \text{Kids}.$$

The design matrix was checked for multicollinearity problems through the use of variance inflation factors. Next, variable selection and model-building was accomplished via stepwise and backwards regression procedures. Also, best subsets of all-possible-regression models were examined using the criteria adjusted R^2, AIC, MSE, and

Mallow's C_p. The strong principle of hierarchy was obeyed. In other words, the presence of interaction and second-order terms required the inclusion of all lower-order terms contained within those of higher order.

Once a final reduced model (henceforth referred to as the model) was selected, it too was checked for multicollinearity problems. The efficacy of the model was checked by examining influence diagnostics, namely: residuals, studentized residuals, leverages, dffits, and dfbetas. The assumption of constant variance for each model was checked by plotting the residuals versus predicted responses, whereas the normality assumption was checked via a normal probability plot in conjunction with Shapiro-Wilk's test.

Global minimum and maximum predicted responses in the region of operability were obtained by examining predicted responses over a grid of level settings. The FCA utilized all possible permutations (1,296 for RE_{HT} and 648 for RE_{HH}) of those factor level settings used in the analysis in addition to the 72 trials used in the experimental design. The corresponding level settings for all the factors and any differences between these level settings were identified. It was then determined if either of the two settings that yielded the optima were actually observed. The predicted responses and prediction variances were verified and confidence intervals for each respective $\mu_{y|\mathbf{x}_0}$ were generated.

Three-dimensional response surface plots were generated for each significant interaction term in the model using a grid of previously determined, reasonable level

settings (1,681 total for both responses), with all other variables set to 0 (coded centerpoint level). If necessary a two-dimensional plot was generated for any factor involved in a quadratic term (41 level settings). The respective ranges of response values for the three-dimensional plots were standardized. Depending on the response, these two- and three-dimensional plots were searched for areas where relative efficiencies were above or below 1 and for the minimum or maximum. The corresponding optimal level settings were identified for the two factors of the interaction term, and, if necessary, for the one factor in any quadratic term. Any differences in the results between the responses were noted. Collectively, this analysis will henceforth be referred to as the *centerpoint* FCA and any related two- or three-dimensional plot will be referred to as a *centerpoint plot* from the FCA.

The next part of the FCA entailed setting reasonable levels of the fixed covariates, generating a set of predicted values over a grid of previously determined levels of the three controllable factors, and then selecting the minimum and maximum predicted response from this set. Specifically, the fixed covariates levels were set as Shape = 4, Direct = 0, Spread = .1, Parents = 3, and Kids = 30. The levels for the controllable factors were set as 10 to 30 by 1 for Area, 1 and 2 for Critical, and 10 to 50 by 2 for Sample. Consequently, there were 882 possible permutations used to construct the grid for both responses. The corresponding level settings for the controllable factors yielding both the minimum and maximum predicted response were identified. Once again, both the predicted responses and their variances were verified. Confidence

intervals for each respective $\mu_{y|\mathbf{x}_0}$ were generated. Two- and three-dimensional plots from grids of previously determined, reasonable level settings (42 or 441 permutations) were generated for the three interaction terms involving the controllable factors even if the corresponding interaction effect was not significant. The respective ranges of response values for the three-dimensional plots were standardized. Depending on the response, these plots were searched for areas above or below relative efficiencies of 1 and for both the minimum and maximum. The remaining (third) controllable factor was set to a level based on the levels that yielded both the minimum and maximum predicted response. The particular settings yielding the minimum and the maximum values from the grid search were recorded. Collectively, this analysis is referred to as the *conditional* FCA and any related three-dimensional plot is referred to as a *conditional plot* from the FCA.

The Random Covariate Analysis (RCA)

The initial general linear model used for the RCA was a full second-order regression model for the four random covariates

$$RE_{\cdot} = \beta_0 + \sum_{i=1}^{4} \beta_i R_i + \sum\sum_{i<j} \beta_{ij} R_i R_j + \sum_{i=1}^{4} \beta_{ii} R_i^2 + \epsilon \qquad (3.4)$$

where $i = 1, \ldots, 4$, $j = 1, \ldots, 4$, and

$$R_1 = \text{Pervarw}$$

$$R_2 = \text{Predp}$$

$$R_3 = \text{Numnets}$$

$$R_4 = \text{Rarity.}$$

Because multicollinearity was found to be a problem with the random covariates, information from correlation matrices, variance inflation factors, and ridge regression analyses were used to reduce the number of candidate terms considered for model selection. The RCA was conducted much like the FCA. The RCA used predictions from the model to perform a global grid search analysis which utilized all possible permutations (1,331 for RE_{HT} and 1,296 for RE_{HH}) of arbitrarily chosen sequences of those factor level settings used in the analysis in addition to the 72 trials used in the experimental design. A *centerpoint* RCA was performed in the same manner as the centerpoint FCA and any related two- or three-dimensional plot will be referred to as a *centerpoint plot* from the RCA.

Predictions from the model were used to perform a grid search analysis for each pair of random covariates where the reasonable level(s) of the other remaining random covariate(s) was (were) determined based on the results from identifying the settings yielding the global minimum and maximum. This was done regardless of whether or not the pair of random covariates was involved in a significant interaction effect. Next,

two- and three-dimensional plots based on a grid of previously determined, reasonable level settings (441 permutations) were generated, and the plots were searched for areas where relative efficiencies were above or below 1 and both the minimum and maximum predicted response. Respective ranges of response values for the three-dimensional plots were standardized. The level settings of the two factors corresponding to the minimum and maximum predicted responses for the interaction term were identified. The particular settings yielding the minimum and maximum values from the global grid search were recorded. Any differences between the results for the two responses were noted and recorded. If necessary a two-dimensional plot was generated for any factor involved in a quadratic term (21 level settings) and the preceding methodology was again followed. Collectively, this analysis will henceforth be referred to as the *conditional* RCA and any related two- or three-dimensional plot will be referred to as a *conditional plot* from the RCA.

Confirmation Trials

To conclude our experiment, confirmation trials were run at several checkpoints to validate all four models. The checkpoint settings were the centerpoints and those that gave the minimum and maximum RE_{HT} and RE_{HH} values from the global grid search for the FCA. A single median observation from 250 populations for each response was generated from the simulation program and compared to the 95% prediction interval from the corresponding FCA models. Subsequently, values were obtained for the four

random covariates and were used to generate 95% prediction intervals utilizing the models from the RCA. The same single median observations were then compared to these respective intervals.

CHAPTER 4

RESULTS

The resulting design matrix of uncoded levels for the RCA along with the median data for both responses, RE_{HT} and RE_{HH}, is shown in Table 8.

Table 8: Design matrix of uncoded levels used in the RCA. All data shown are median values, $N = 250$

Trial	$\frac{\sigma_W^2}{\sigma^2}$	number of networks	$\widehat{\lambda}_p$	$\frac{E[\nu]}{n_1}$	RE_{HT}	RE_{HH}
1	0.189	99	1.4	1.122	1.069	1.022
2	0.000	400	1.7	1.020	0.990	0.990
3	0.069	899	1.35	1.015	1.061	1.052
4	0.123	899	1.8	1.014	1.132	1.111
5	0.121	97	4.9	1.159	0.970	0.820
6	0.049	98	4.9	1.220	0.862	0.800
7	0.102	397	5.0	1.102	1.007	0.983
8	0.217	895	5.3	1.075	1.190	1.186
9	0.214	96	1.4	1.402	0.980	0.891
10	0.228	96	1.4	1.113	3.373	0.986
11	0.471	892	1.8	1.135	1.904	1.569
12	0.352	387	2.5	1.384	1.268	1.054
13	0.074	897	2.6	1.061	1.011	1.004
14	0.547	73	4.9	4.521	0.598	0.290
15	0.414	79	4.9	1.449	5.834	0.639
16	0.26	885	4.4	1.218	1.190	1.069
17	0.275	881	5.3	1.292	1.071	1.050
18	0.178	99	1.4	1.056	1.396	1.033
19	0.000	100	1.4	1.040	0.970	0.951
20	0.191	899	0.9	1.015	1.209	1.204
21	0.000	900	0.9	1.009	0.996	0.996
22	0.132	98	3.5	1.245	0.913	0.892
23	0.191	898	2.6	1.034	1.194	1.169
24	0.184	97	4.9	1.179	1.083	0.829
25	0.100	397	5.0	1.107	0.990	0.988

continued on next page

Table 8: continued from previous page

Trial	$\frac{\sigma_W^2}{\sigma^2}$	number of networks	$\widehat{\lambda}_p$	$\frac{E[\nu]}{n_1}$	RE_{HT}	RE_{HH}
26	0.151	897	4.85	1.061	1.111	1.108
27	0.231	97	1.4	1.111	2.602	0.980
28	0.326	95	1.4	1.506	1.020	0.918
29	0.240	896	0.9	1.059	1.240	1.218
30	0.231	896	0.9	1.060	1.283	1.218
31	0.485	78.5	4.9	2.076	1.920	0.499
32	0.108	94	4.9	1.701	0.681	0.626
33	0.484	871	5.3	1.503	1.449	1.225
34	0.081	895	5.3	1.096	1.016	0.983
35	0.329	98	1.4	1.085	2.052	1.179
36	0.221	99	1.4	1.112	1.274	1.019
37	0.143	399	1.7	1.034	1.191	1.103
38	0.295	898	0.9	1.029	1.388	1.377
39	0.040	99	2.8	1.096	0.980	0.886
40	0.357	896	2.6	1.067	1.454	1.441
41	0.206	95	4.9	1.572	0.785	0.743
42	0.275	94.5	4.9	1.666	0.821	0.770
43	0.363	893	5.3	1.100	1.461	1.400
44	0.058	897	5.3	1.053	1.006	1.004
45	0.055	898	4.4	1.044	1.016	1.000
46	0.015	99	1.4	1.182	0.892	0.867
47	0.413	388	0.8	1.673	1.088	0.982
48	0.018	899	0.9	1.028	1.010	0.990
49	0.251	892	0.9	1.167	1.164	1.130
50	0.486	73	2.8	1.510	39.196	0.635
51	0.074	93	4.9	1.478	0.828	0.584
52	0.377	362	5.0	2.176	1.321	0.616
53	0.195	881	4.4	1.387	0.918	0.900
54	0.284	868	4.4	1.541	1.079	0.879
55	0.112	99	1.4	1.122	0.971	0.971
56	0.252	399	0.8	1.047	1.250	1.228
57	0.000	899	1.8	1.015	0.996	0.996
58	0.244	899	1.8	1.024	1.324	1.273
59	0.365	93	4.9	1.277	1.748	0.872
60	0.073	97	4.9	1.174	0.945	0.776
61	0.357	393	5.0	1.221	1.241	1.220
62	0.158	896	4.4	1.060	1.131	1.104

continued on next page

Table 8: continued from previous page

Trial	$\frac{\sigma_W^2}{\sigma^2}$	number of networks	$\widehat{\lambda}_p$	$\frac{E[\nu]}{n_1}$	RE_{HT}	RE_{HH}
63	0.431	90	1.4	2.306	1.052	0.661
64	0.017	99	1.4	1.129	0.968	0.871
65	0.280	93	1.4	1.186	6.838	0.947
66	0.167	895	0.9	1.075	1.209	1.091
67	0.422	890	0.9	1.215	1.740	1.366
68	0.036	397	3.3	1.138	0.923	0.912
69	0.284	82	4.9	1.477	2.907	0.498
70	0.392	74	4.9	4.258	0.496	0.254
71	0.384	863	5.3	1.918	0.877	0.844
72	0.302	871	5.3	1.536	1.015	0.921

The results obtained for RE_{HT} include some unusually large values. For example, see trial 50 in Table 8. Consequently, via a SAS macro, a Box-Cox transformation procedure was used. For both analyses, an inverse transformation was suggested for RE_{HT} (henceforth called RE^*_{HT}) whereas no transformation was required for RE_{HH}. Consequently, from a relative efficiency standpoint, we will be interested in minimizing RE^*_{HT}.

Figures 2, 3, and 4 show examples of nominal, extreme 1, and extreme 2 populations (see Table 6) from the initial study. The actual data for these populations are in Appendix C. Table 9 contains the resulting parameters for these specific three populations.

From the initial study, the density histograms for RE_{HH}, RE_{HT}, and RE^*_{HT} under the three different population settings for $N = 500$ are in Figures 5 through 7. For the Extreme 1 setting (Figure 6), the distribution for RE_{HT} was so severely

Figure 2. A Nominal Population.

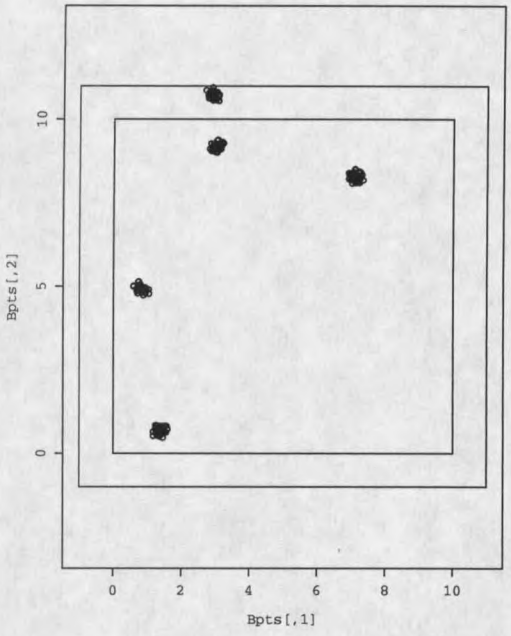


Figure 3. An Extreme 1 Population.

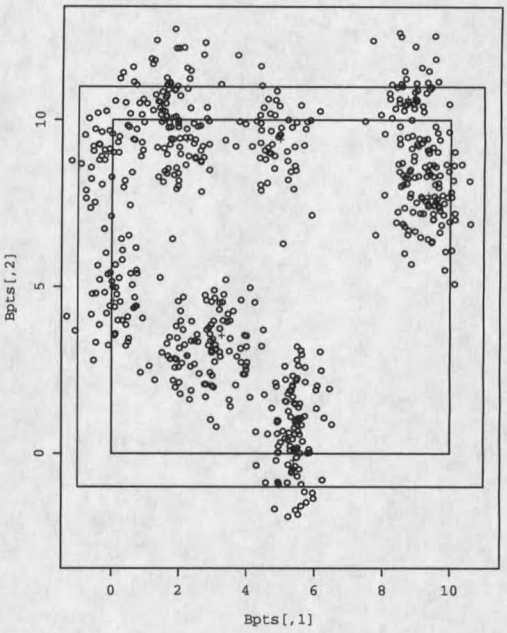


Figure 4. An Extreme 2 Population.

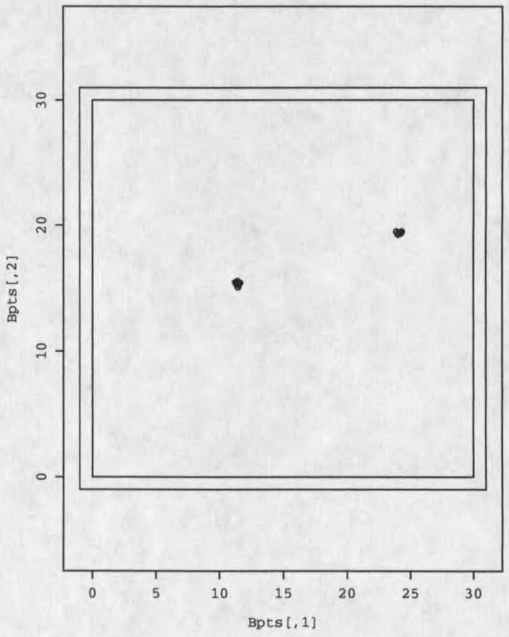


Table 9. Initial study example population characteristics.

Study Population	Nominal	Extreme 1	Extreme 2
τ	204	419	20
σ^2	74.604	31.832	0.229
σ_B^2	59.689	10.839	0.215
σ_W^2	14.916	20.994	0.014
$\frac{\sigma_W^2}{\sigma^2}$	0.200	0.660	0.061
Number of Networks	96	39	899
$\widehat{\lambda}_p$	3.5	8.3	1.8
$E[\nu]$	14.133	92.145	30.574
$\frac{E[\nu]}{n_1}$	1.413	3.072	1.019
$Var[\hat{\tau}_{HH}]$	53719.758	2529.044	5604.839
$Var[\hat{\tau}_{HT}]$	52213.595	102.318	5593.275
$Var[\hat{\tau}_{SRS}]$	45325.795	271.354	5851.761
RE_{HH}	0.844	0.107	1.044
RE_{HT}	0.868	2.652	1.046

skewed right that including a histogram was impractical. Thus, a histogram for the distribution of RE_{HT} (labeled edited RE_{HT}) only for those values less than the third quartile (26.732) was made. The density histograms for the four random covariates under the three different population settings for $N = 500$ are in Figures 8 through 10. Table 10 contains five-number summaries for the distributions of RE_{HH}, RE_{HT}, RE^*_{HT}, and the four random covariates under the three different population settings for $N = 500$. Note Table 10 also contains a column ($\% < 1$) showing the proportion of the respective 500 relative efficiencies having a value strictly less than 1.000.

A summary of the simulation results comparing the approach of Christman (1996a) to the one described in this paper is given in Table 11.

Figures 11 and 12 contain the distributions for RE_{HT}, RE_{HH}, and the four random covariates for all 72 trials while Table 12 contains the resulting five-number summaries. For comparison's sake, Figure 11 also contains the distribution for RE^*_{HT} and a modified distribution for RE_{HT}. The modified distribution was created by removing the four values for RE_{HT} greater than 3 (3.373, 5.834, 6.838, and 39.196). A five-number summary for RE^*_{HT} is in Table 12 which also contains a column ($\% < 1$) showing the proportion of the 72 respective relative efficiencies having a value strictly less than 1.000. Figure 13 is a plot of RE^*_{HT} versus RE_{HH} for all 72 trials.

Figure 5. Relative Efficiency Distributions, Nominal Setting, $N = 500$.

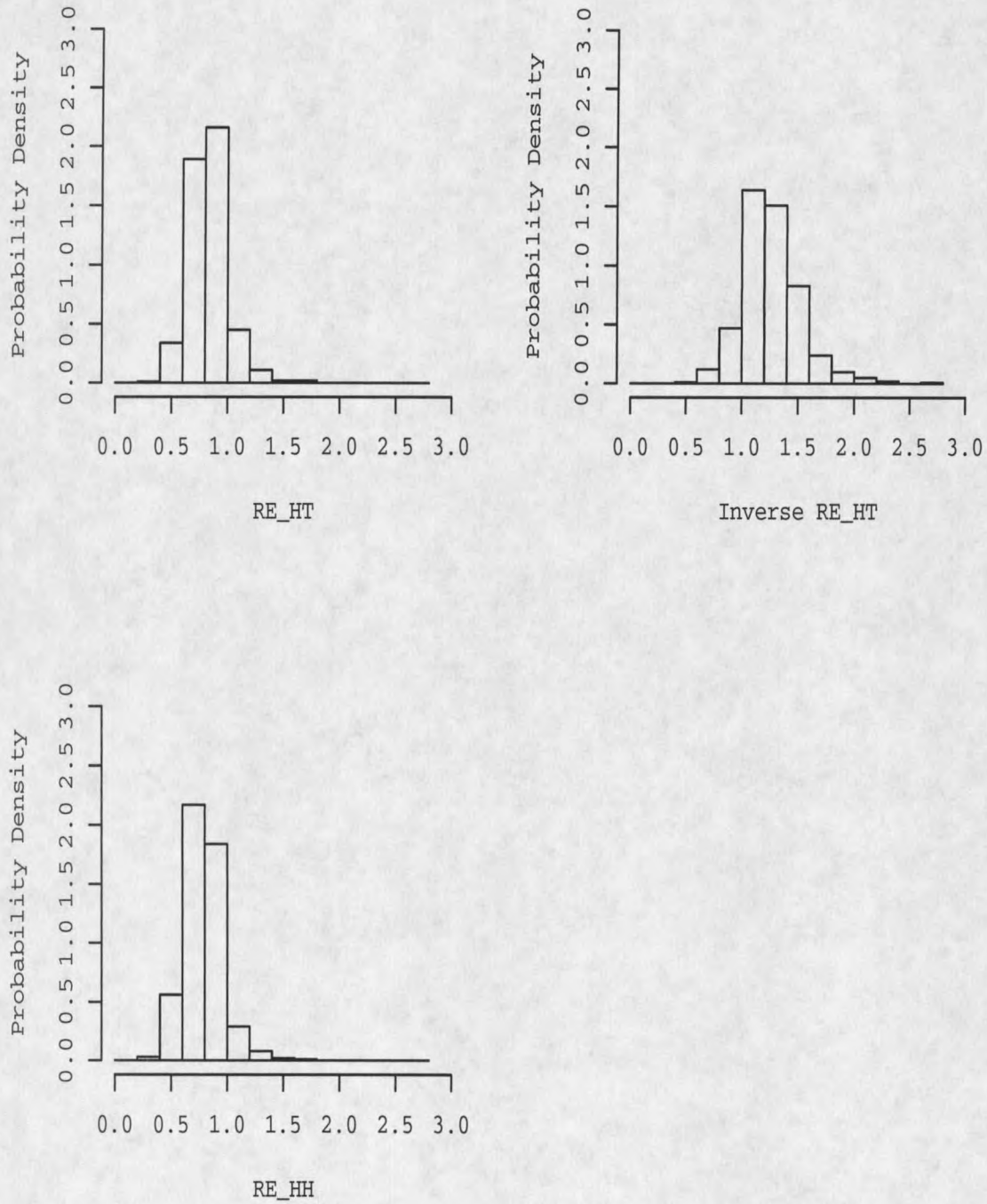


Figure 6. Relative Efficiency Distributions, Extreme 1 Setting, $N = 500$.

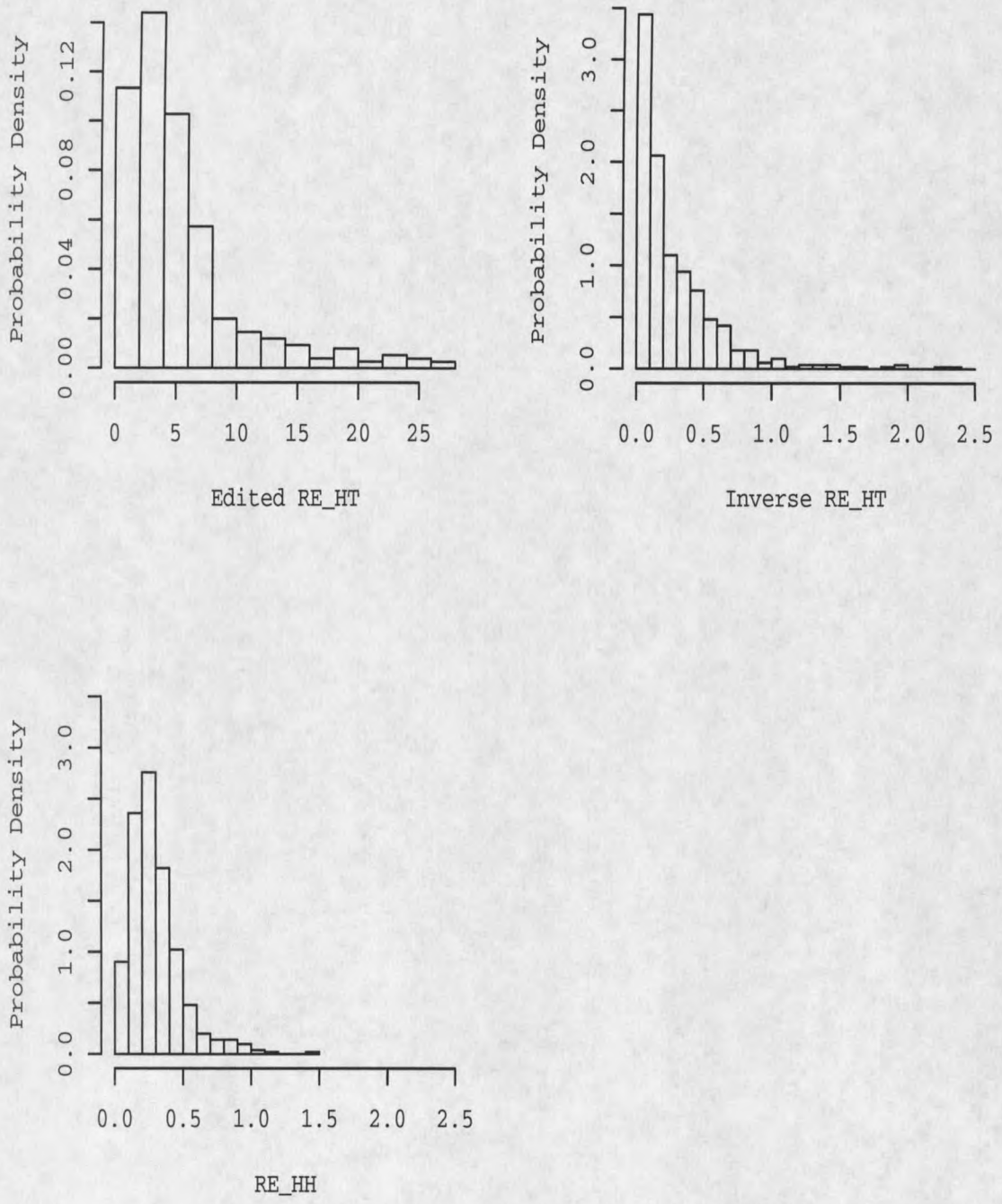


Figure 7. Relative Efficiency Distributions, Extreme 2 Setting, $N = 500$.

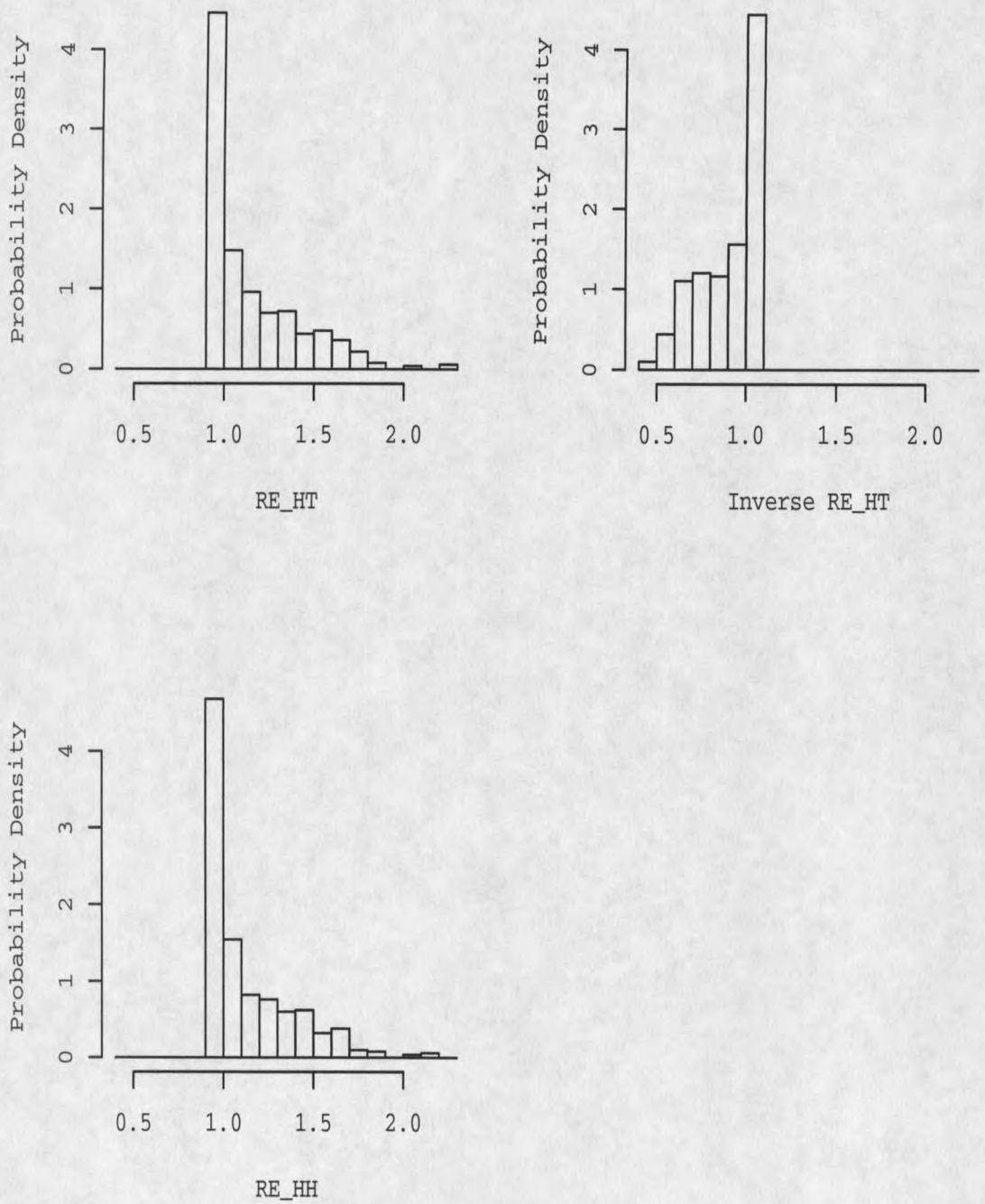


Figure 8. Random Covariate Distributions, Nominal Setting, $N = 500$.

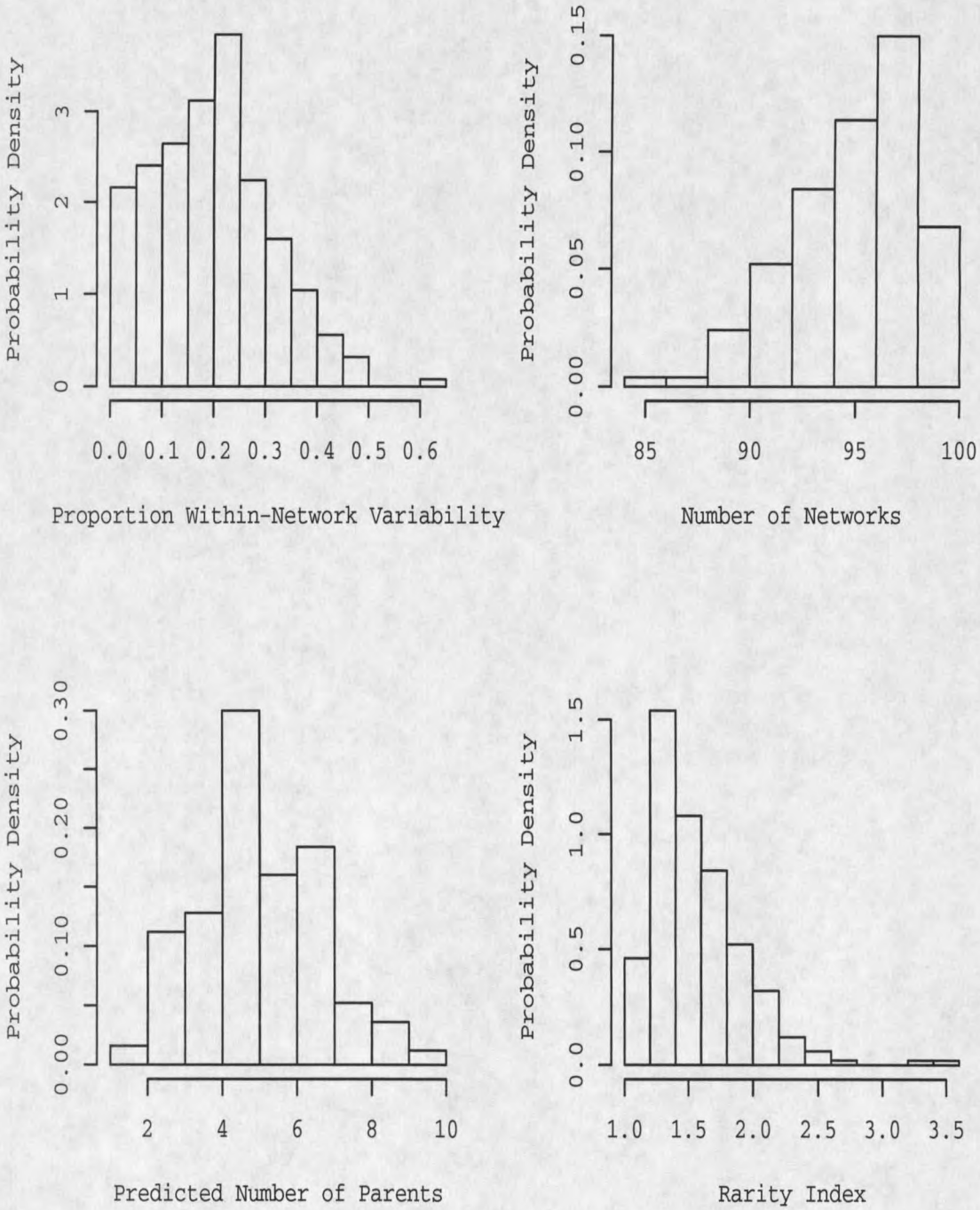


Figure 9. Random Covariate Distributions, Extreme 1 Setting, $N = 500$.

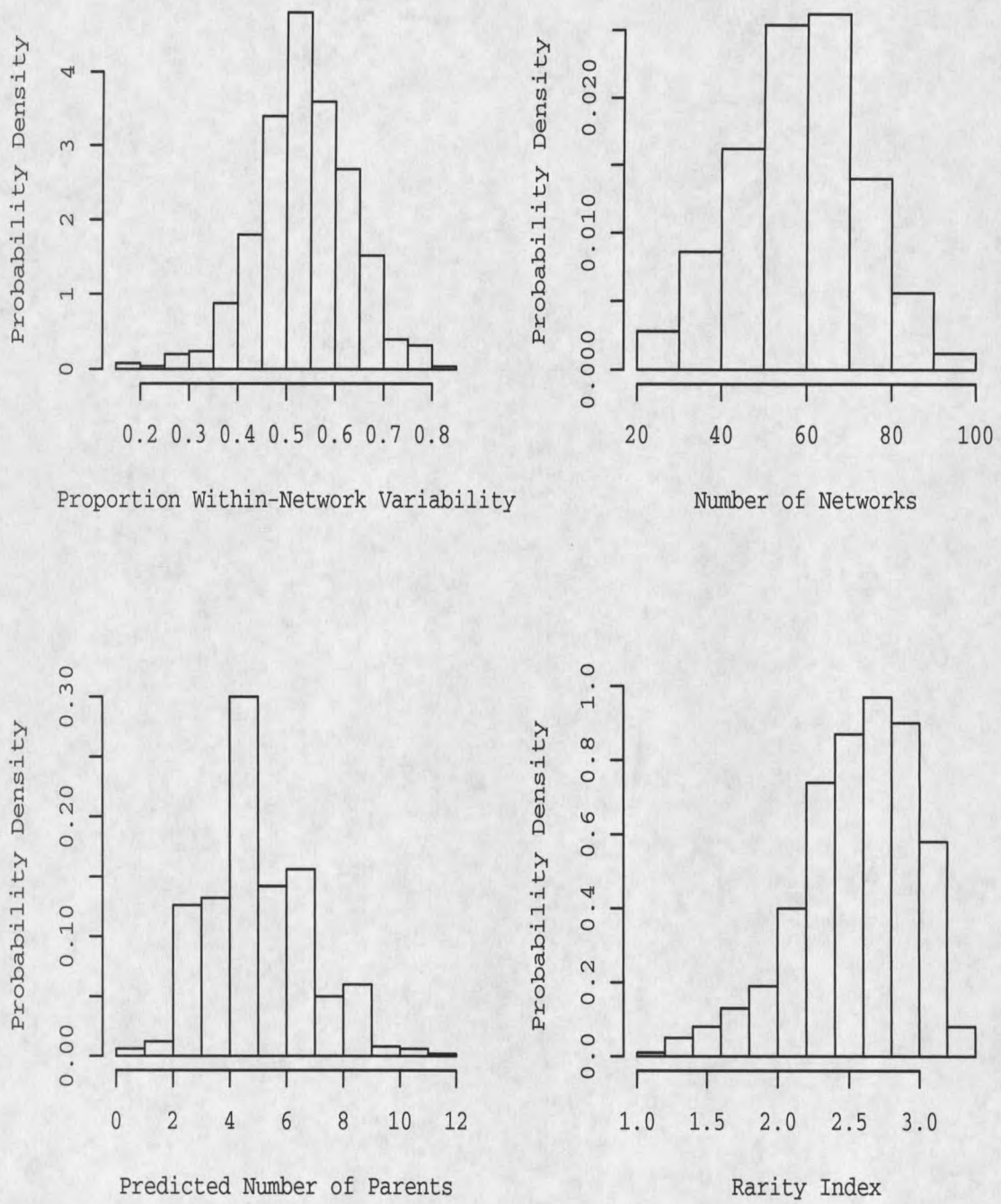


Figure 10. Random Covariate Distributions, Extreme 2 Setting, $N = 500$.

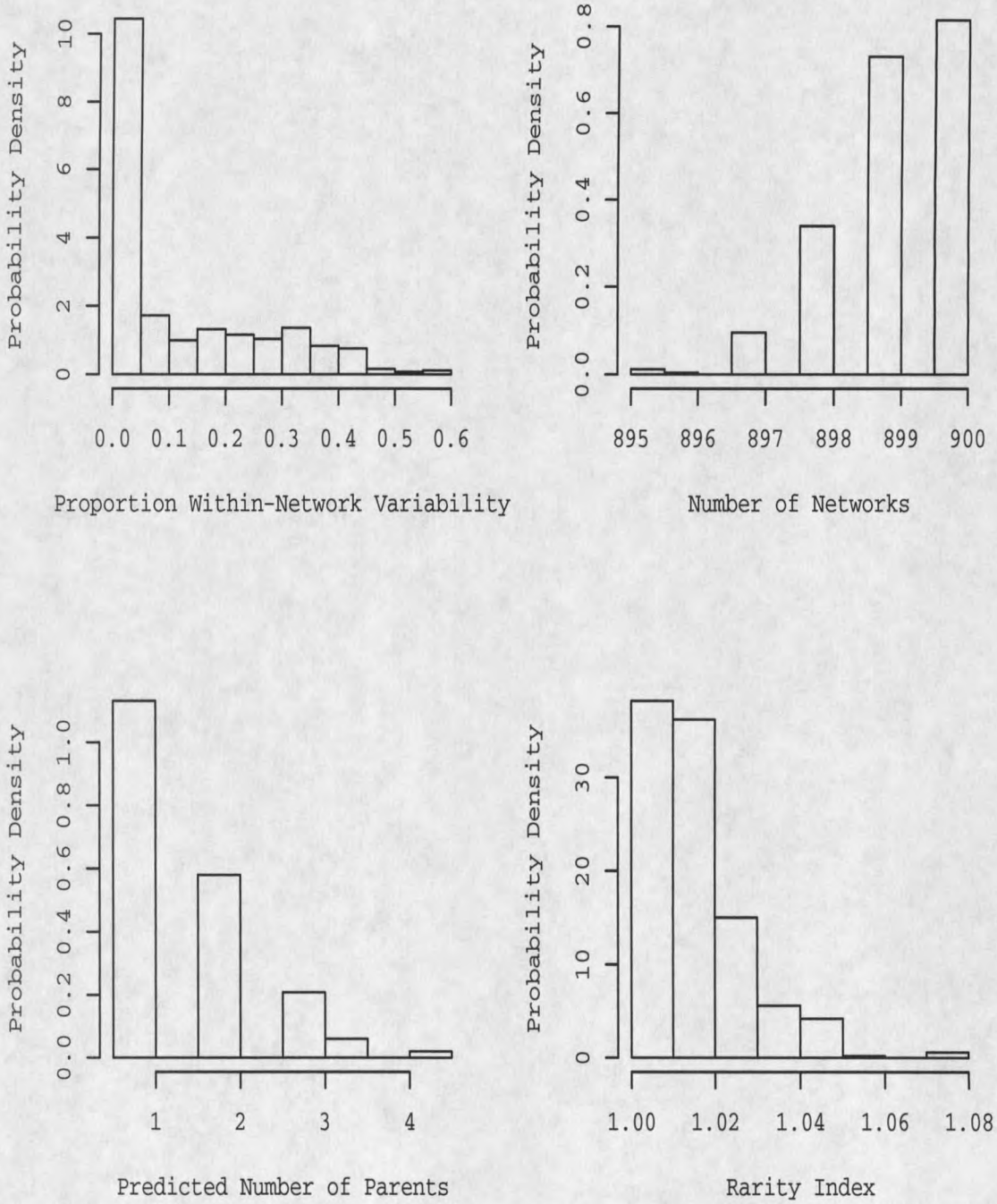


Table 10. Numerical summary of relative efficiencies and random covariates from simulation for initial study where $N = 500$. Here, na = not applicable. For the RE_{HT} distribution under the Extreme 1 setting, μ and Max where so large in magnitude due to the presence of outliers, they were not included in the table.

Setting	Parameter	Min	Q_1	M	μ	Q_3	Max	% < 1
Nominal	RE_{HT}	0.380	0.715	0.818	0.828	0.912	1.771	88.0
	RE^*_{HT}	0.565	1.096	1.223	1.259	1.398	2.632	12.0
	RE_{HH}	0.279	0.682	0.786	0.788	0.875	1.685	91.8
	$\frac{\sigma^2_W}{\sigma^2}$	0.000	0.110	0.198	0.197	0.272	0.615	na
	Number of Networks	85.0	94.0	96.0	95.5	98.0	100.0	na
	$\widehat{\lambda}_p$	1.4	3.5	4.9	5.0	6.2	9.7	na
	$\frac{E[\nu]}{n_1}$	1.036	1.320	1.512	1.569	1.748	3.536	na
Extreme 1	RE_{HT}	0.420	2.533	5.448		26.732		3.8
	RE^*_{HT}	0.000	0.037	0.184	0.278	0.395	2.381	96.2
	RE_{HH}	0.025	0.171	0.258	0.299	0.379	1.415	99.2
	$\frac{\sigma^2_W}{\sigma^2}$	0.152	0.482	0.537	0.538	0.599	0.804	na
	Number of Networks	21.0	49.8	60.0	59.2	68.0	97.0	na
	$\widehat{\lambda}_p$	0.7	3.5	4.9	5.0	6.2	11.1	na
	$\frac{E[\nu]}{n_1}$	1.152	2.314	2.604	2.556	2.857	3.287	na
Extreme 2	RE_{HT}	0.975	0.996	1.022	1.165	1.284	2.256	44.4
	RE^*_{HT}	0.443	0.779	0.979	0.888	1.004	1.026	55.4
	RE_{HH}	0.974	0.996	1.011	1.153	1.266	2.182	46.4
	$\frac{\sigma^2_W}{\sigma^2}$	0.000	0.000	0.038	0.118	0.228	0.553	na
	Number of Networks	895.0	899.0	899.0	899.1	900.0	900.0	na
	$\widehat{\lambda}_p$	0.9	0.9	0.9	1.5	1.8	4.4	na
	$\frac{E[\nu]}{n_1}$	1.002	1.004	1.015	1.016	1.022	1.078	na

Table 11. Numerical summary of relative efficiencies from replication of Christman (1996a).

Study	Parameter	Min	Q_1	M	μ	Q_3	Max	% < 1
Christman	RE_{HT}	0.33	1.18	1.53	1.63	1.93	4.21	11.5
Turk	RE_{HT}	0.21	0.72	0.99	1.18	1.22	45.36	51.4
Turk	RE_{HH}	0.06	0.24	0.37	0.46	0.61	1.44	93.6

Figure 11. Relative Efficiency Distributions for 72 Trials.

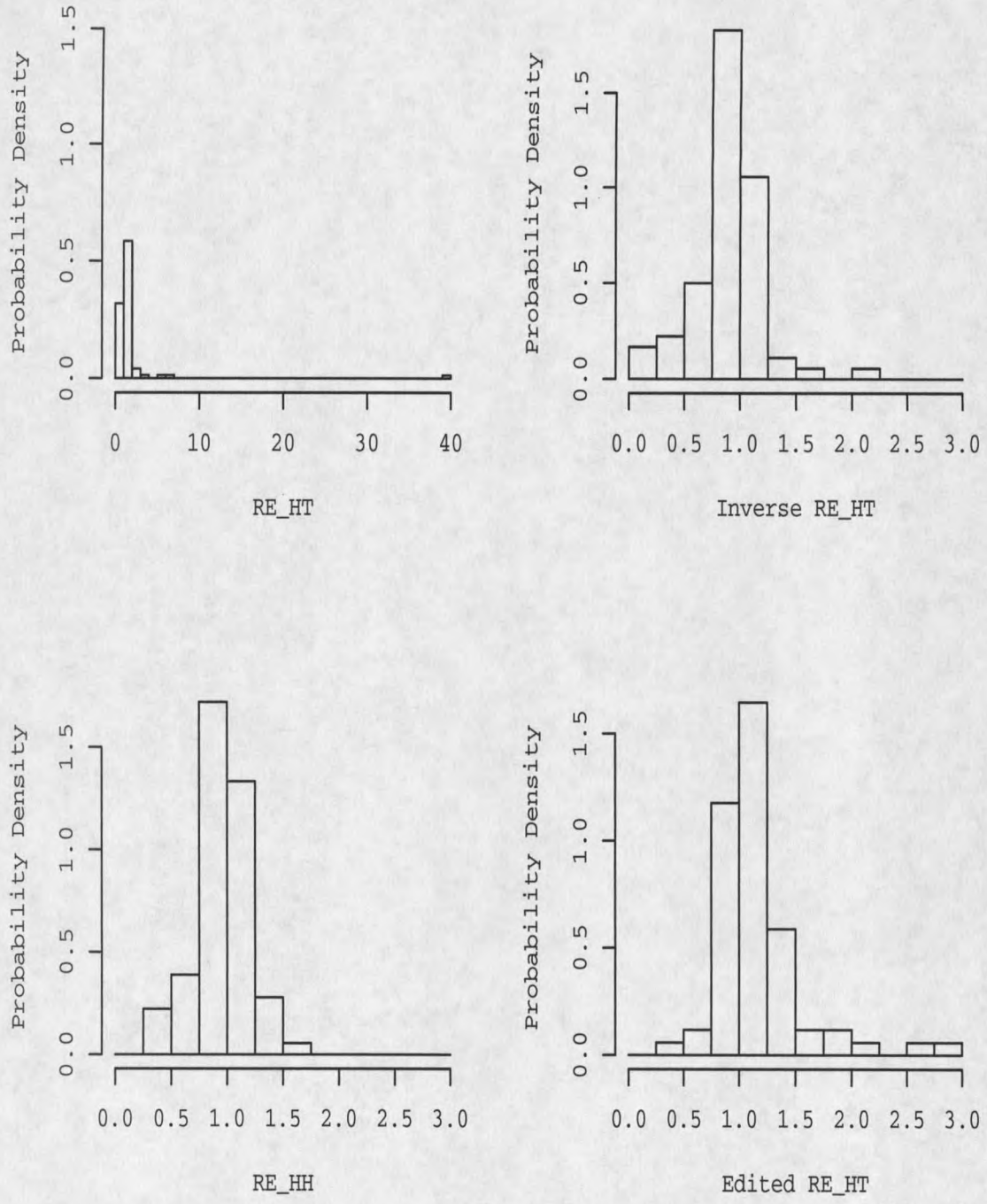


Figure 12. Random Covariate Distributions for 72 Trials.

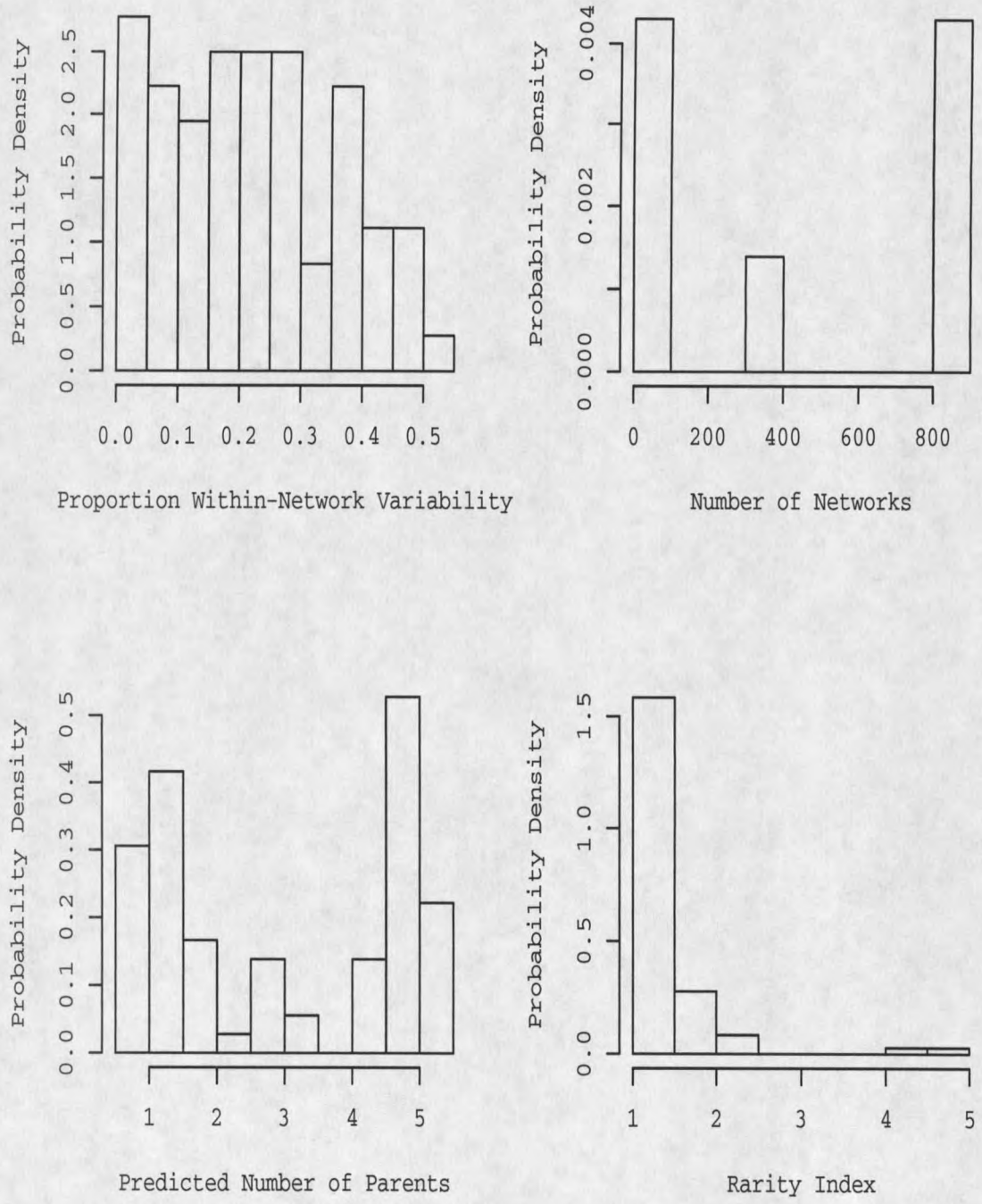


Figure 13. Plot of RE^*_{HT} Versus RE_{HH} for 72 Trials With Reference Line $RE^*_{HT} = 1/RE_{HH}$.

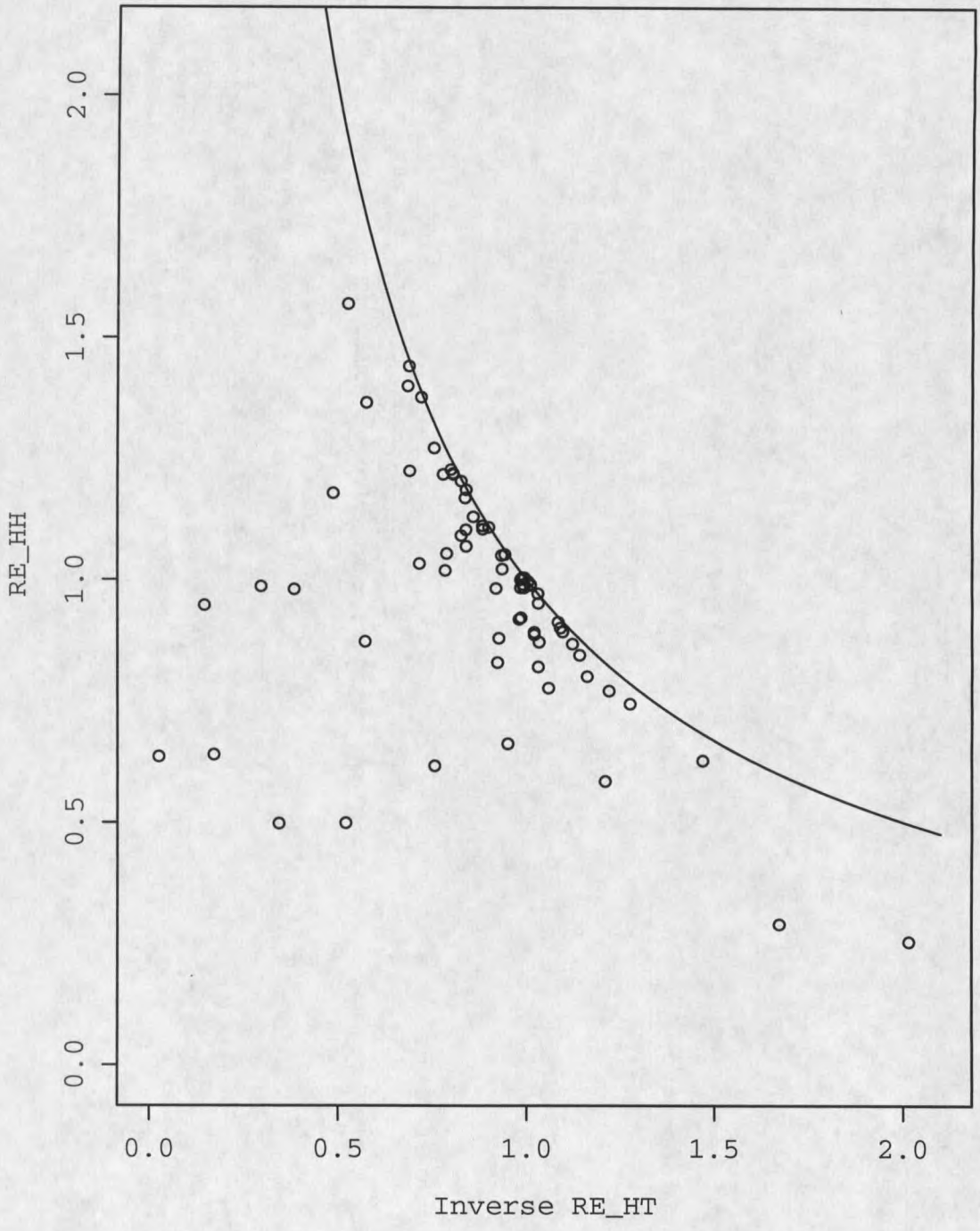


Table 12. Numerical summary of simulation results for 72 trials. na = not applicable.

Parameter	Min	Q_1	M	μ	Q_3	Max	% < 1
RE_{HT}	0.496	0.980	1.081	1.877	1.292	39.196	31.9
RE_{HH}	0.254	0.870	0.987	0.968	1.105	1.569	56.9
RE^*_{HT}	0.026	0.774	0.925	0.883	1.020	2.016	68.1
$\frac{\sigma^2_W}{\sigma^2}$	0.000	0.102	0.216	0.219	0.327	0.547	na
Number of Networks	73.0	97.0	397.0	478.2	895.0	900.0	na
$\widehat{\lambda}_p$	0.8	1.4	2.7	3.1	4.9	5.3	na
$\frac{E[\nu]}{n_1}$	1.009	1.060	1.132	1.338	1.414	4.521	na

Results for RE^*_{HT} and RE_{HH} Data - The Fixed Covariate Analysis

The FCA results of the variable selection process described in the Methods yielded models with ANOVA results summarized in Table 13 for RE^*_{HT} and Table 14 for RE_{HH}. No unusual variance inflation factors were noted at any time in the model building process for either response. An examination of the influence diagnostics revealed several "unusual" points, which was not surprising given the amount of data and the number of influence diagnostics. None of these points were determined to be sources of concern. There was nothing to suggest any potential violations of the assumptions of normality and constant variance.

Table 15 contains the FCA global grid search results for both the maximum and minimum predicted RE^*_{HT} and RE_{HH}. For the minimum RE^*_{HT} response, a negative predicted value was obtained. Consequently, the minimum RE^*_{HT} response was set

Table 13. SAS computer output for the reduced FCA model for RE^*_{HT}. Effects unique to the model for this response are proceeded by #.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	5.77814	0.30411	15.32	<.0001
Error	52	1.03209	0.01985		
Corrected Total	71	6.81023			
Root MSE	0.14088	R-Square	0.8484		
Dependent Mean	0.88324	Adj R-Sq	0.7931		
Coeff Var	15.95061				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > $\|t\|$
Intercept	1	0.88441	0.01667	53.06	<.0001
Shape	1	0.00808	0.01673	0.48	0.6310
Direct #	1	-0.00531	0.01680	-0.32	0.7532
Spread	1	-0.03757	0.01686	-2.23	0.0302
Parents	1	0.08970	0.01780	5.04	<.0001
Area	1	-0.01409	0.01813	-0.78	0.4407
Critical	1	0.05272	0.01684	3.13	0.0029
Kids	1	-0.07409	0.01809	-4.10	0.0001
Sample	1	-0.17211	0.01796	-9.58	<.0001
Shape*Spread	1	0.04331	0.01681	2.58	0.0129
Shape*Critical #	1	0.03825	0.01682	2.27	0.0271
Direct*Critical #	1	0.02652	0.01704	1.56	0.1256
Spread*Parents	1	0.03844	0.01788	2.15	0.0362
Spread*Area	1	0.05640	0.01820	3.10	0.0031
Spread*Sample	1	-0.11522	0.01812	-6.36	<.0001
Parents*Area	1	-0.03816	0.01893	-2.02	0.0490
Parents*Kids	1	0.03607	0.01907	1.89	0.0641
Parents*Sample #	1	-0.04257	0.01900	-2.24	0.0294
Area*Sample #	1	0.14727	0.01945	7.57	<.0001
Kids*Sample	1	-0.05809	0.01936	-3.00	0.0041

Table 14. SAS computer output for the reduced FCA model for RE_{HH}. Effects unique to the model for this response are proceeded by #.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	4.05844	0.21360	55.27	0.0001
Error	52	0.20098	0.00386		
Corrected Total	71	4.25942			
Root MSE	0.06217	R-Square	0.9528		
Dependent Mean	0.96754	Adj R-Sq	0.9356		
Coeff Var	6.42544				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > $\|t\|$
Intercept	1	1.02431	0.02009	50.99	0.0001
Shape	1	-0.02036	0.00739	-2.75	0.0081
Spread	1	-0.07106	0.00742	-9.57	0.0001
Parents	1	-0.11183	0.00783	-14.28	0.0001
Area	1	0.16433	0.00797	20.61	0.0001
Critical	1	-0.01743	0.00742	-2.35	0.0227
Kids	1	0.03989	0.00799	5.00	0.0001
Sample	1	0.02972	0.00804	3.70	0.0005
Shape*Spread	1	-0.05638	0.00746	-7.56	0.0001
Spread*Parents	1	-0.05410	0.00788	-6.86	0.0001
Spread*Area	1	0.04131	0.00802	5.15	0.0001
Spread*Critical #	1	0.02290	0.00746	3.07	0.0034
Spread*Kids #	1	-0.03092	0.00803	-3.85	0.0003
Spread*Sample	1	0.01932	0.00801	2.41	0.0194
Parents*Area	1	0.04394	0.00840	5.23	0.0001
Parents*Kids	1	-0.02343	0.00841	-2.79	0.0074
Area*Critical #	1	-0.04299	0.00801	-5.37	0.0001
Area*Kids #	1	0.05023	0.00861	5.84	0.0001
Kids*Sample	1	0.02407	0.00851	2.83	0.0066
Area*Area #	1	-0.06809	0.02172	-3.13	0.0028

to 0 for the confidence interval and range calculations. The 95% confidence interval for $\mu_{y|\mathbf{x}_0}$ was calculated using

$$\hat{y}(\mathbf{x}_0) \pm t^* SE[\hat{y}(\mathbf{x}_0)]$$

where

$$SE[\hat{y}(\mathbf{x}_0)] = \sqrt{\hat{\sigma}^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

and t^* is the 97.5th percentile of the t distribution with $n - p - 1$ where p is the number of predictor variables.

Table 15. FCA global grid search results for maximum and minimum predicted RE^*_{HT} and RE_{HH}. A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. Range is the range of predicted responses. no = not observed; na = not applicable; CI = confidence interval.

Factor	Minimum RE^*_{HT}	Maximum RE^*_{HT}	Factor	Minimum RE_{HH}	Maximum RE_{HH}
Shape*	1	4	Shape*	4	1
Direct	.7	.7	Direct	na	na
Spread	.5	.5	Spread	.5	.5
Parents*	1	5	Parents*	5	1
Area	10	10	Area*	10	30
Critical*	1	2	Critical	1	1
Kids	50	50	Kids	50	50
Sample*	50	10	Sample*	10	50
$\hat{y}(\mathbf{x}_0)$	-0.020	1.632	$\hat{y}(\mathbf{x}_0)$	0.207	1.483
$SE[\hat{y}(\mathbf{x}_0)]$	0.082	0.077	$SE[\hat{y}(\mathbf{x}_0)]$	0.036	0.034
$y(\mathbf{x}_0)$	no	2.016	$y(\mathbf{x}_0)$	no	1.569
95% CI	[0, 0.144]	[1.476, 1.787]	95% CI	[0.135, 0.279]	[1.414, 1.551]
range$[\hat{y}(\mathbf{x}_0)]$	1.632		range$[\hat{y}(\mathbf{x}_0)]$	1.276	

Figures 14 through 35 are the three-dimensional interaction response surface plots for the condition where all other variables are fixed to the centerpoint (fixed to 0).

Figure 36 is a two-dimensional quadratic response surface plot where all other variables are fixed to the centerpoint. Tables 16 and 17 summarize the important results gathered from all the aforementioned plots.

Table 16. Centerpoint FCA of RE^*_{HT}. Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted RE^*_{HT}. The levels are uncoded and are in the factor order given in the Effect column.

Effect	Minimum RE^*_{HT}	Minimum Levels	Maximum RE^*_{HT}	Maximum Levels	Range RE^*_{HT}	Maximum RE^*_{HT} less than 1?
Shape*Spread	0.795	(1,.5)	0.957	(1,.1)	0.162	yes
Shape*Critical	0.802	(4,1)	0.983	(4,2)	0.182	yes
Direct*Critical	0.800	(.7,1)	0.958	(.7,2)	0.158	yes
Spread*Parents	0.719	(.5,1)	0.975	(.5,5)	0.256	yes
Spread*Area	0.805	(.5,10)	0.992	(.1,10)	0.188	yes
Spread*Sample	0.560	(.5,50)	1.134	(.5,10)	0.575	no
Parents*Area	0.771	(1,10)	1.026	(5,10)	0.256	no
Parents*Kids	0.685	(1,50)	1.012	(5,10)	0.328	no
Parents*Sample	0.665	(1,50)	1.189	(5,10)	0.524	no
Area*Sample	0.579	(10,50)	1.218	(10,10)	0.639	no
Kids*Sample	0.580	(50,50)	1.073	(10,10)	0.492	no

Table 18 contains the grid search results conditioned at fixed covariate levels for both the maximum and minimum predicted RE^*_{HT} and RE_{HH}. Figures 37 through 54 are the two- and three-dimensional interaction FCA response surface plots with the fixed covariates set to reasonable levels.

For the two-dimensional plots, the *region of favorability* is defined to be the area such that conditions would favor using ACS over SRS. For RE^*_{HT}, this would be the area that yields predicted values less than or equal to 1, while for RE_{HH}, it would be the area that yields predicted values greater than or equal to 1. The region of

Figure 14. FCA Centerpoint Plot of Shape*Spread for RE^*_{HT}.

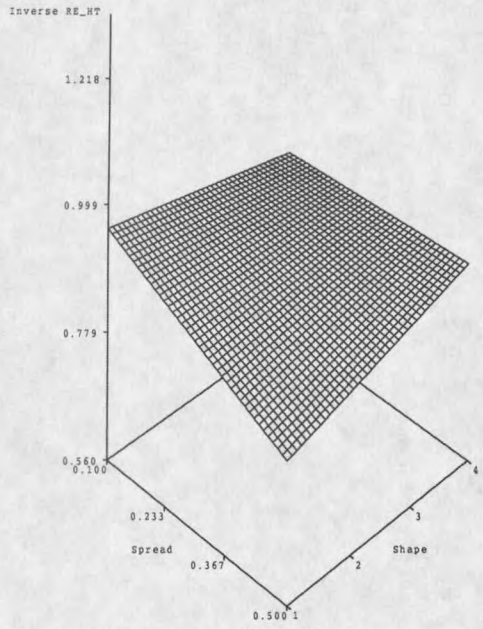


Figure 15. FCA Centerpoint Plot of Shape*Critical for RE^*_{HT}.

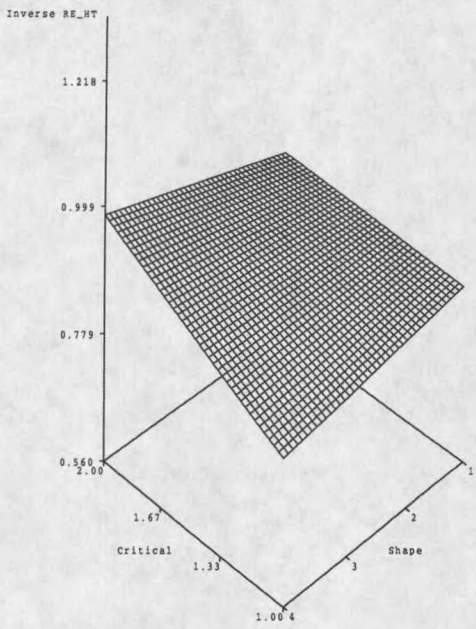


Figure 16. FCA Centerpoint Plot of Direct*Critical for RE^*_{HT}.

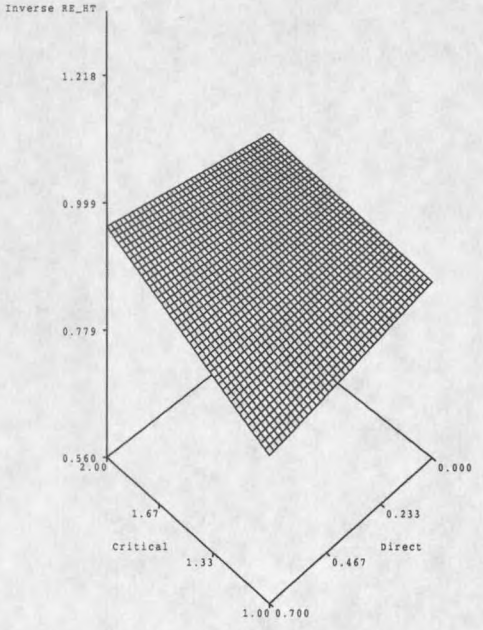


Figure 17. FCA Centerpoint Plot of Spread*Parents for RE^*_{HT}.

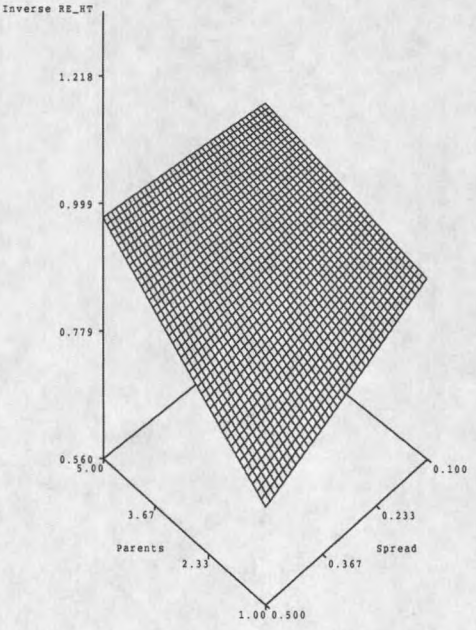


Figure 18. FCA Centerpoint Plot of Spread*Area for RE^*_{HT}.

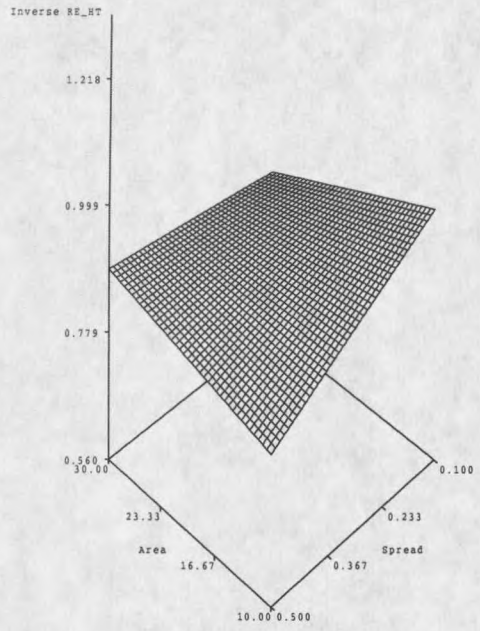


Figure 19. FCA Centerpoint Plot of Spread*Sample for RE^*_{HT}.

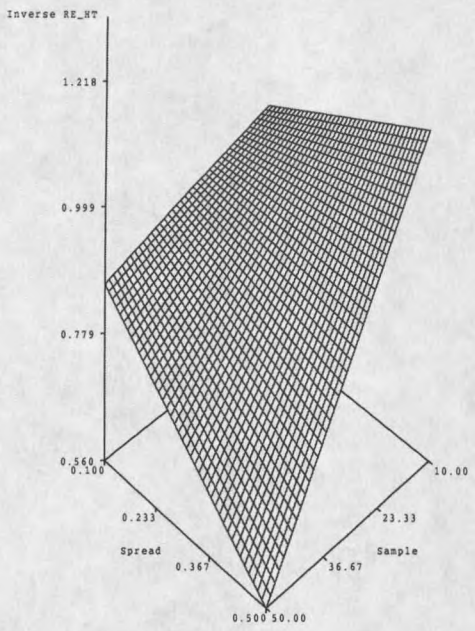


Figure 20. FCA Centerpoint Plot of Parents*Area for RE^*_{HT}.

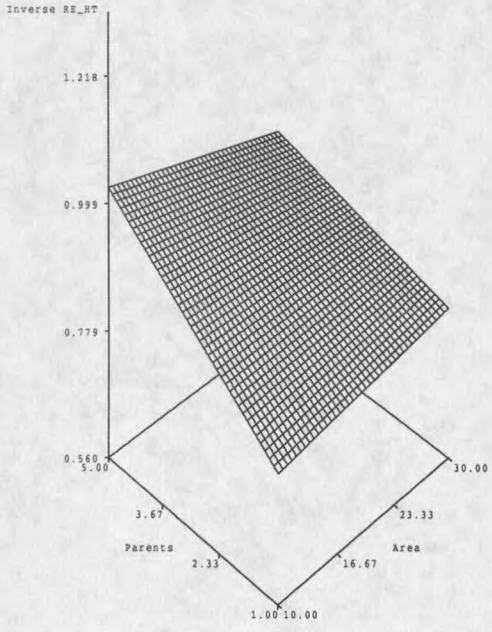


Figure 21. FCA Centerpoint Plot of Parents*Kids for RE^*_{HT}.

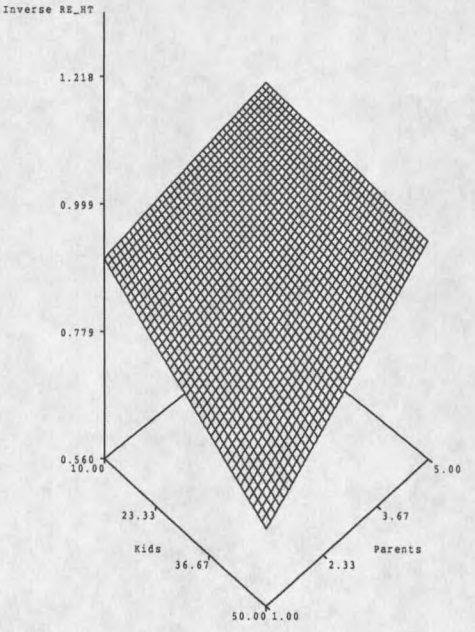


Figure 22. FCA Centerpoint Plot of Parents*Sample for RE^*_{HT}.

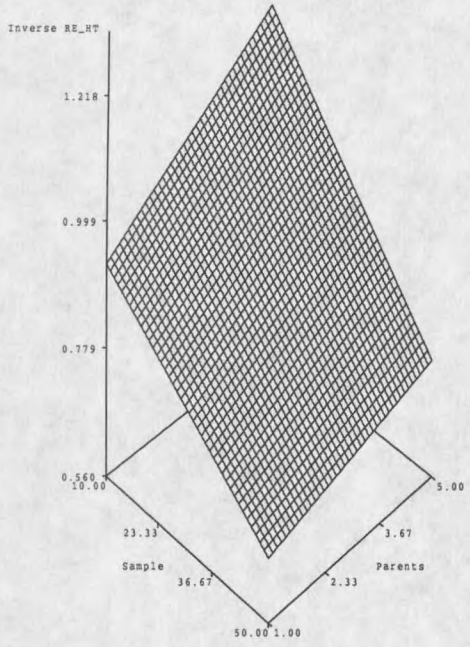


Figure 23. FCA Centerpoint Plot of Area*Sample for RE^*_{HT}.

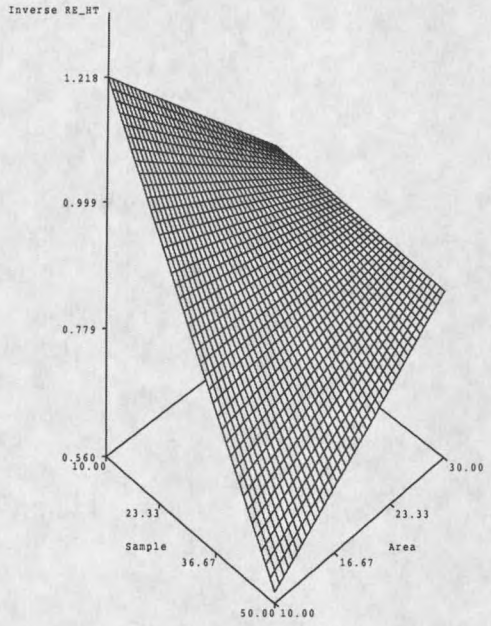


Figure 24. FCA Centerpoint Plot of Kids*Sample for RE^*_{HT}.

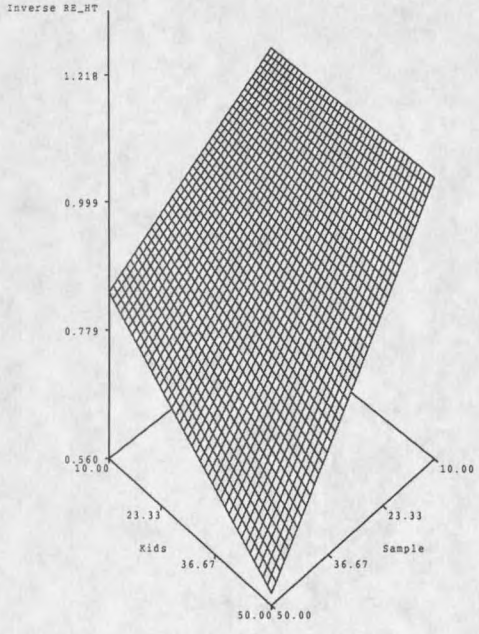


Figure 25. FCA Centerpoint Plot of Shape*Spread for RE_{HH}.

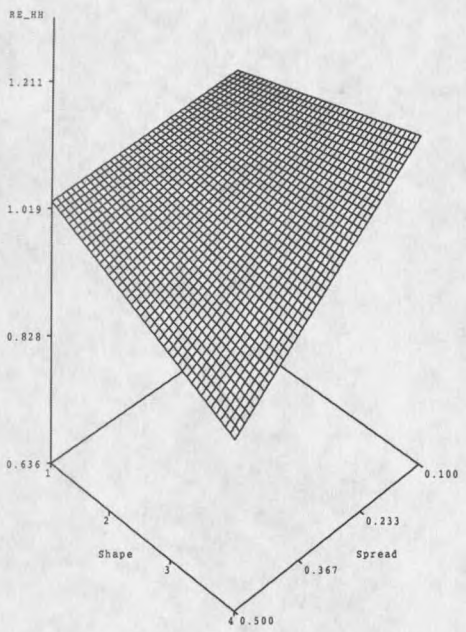


Figure 26. FCA Centerpoint Plot of Spread*Parents for RE_{HH}.

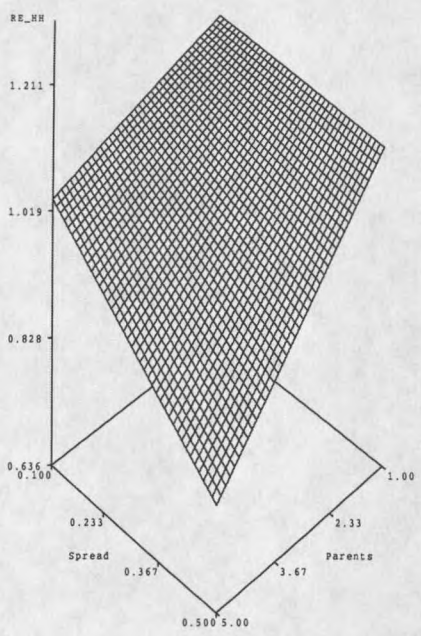


Figure 27. FCA Centerpoint Plot of Spread*Area for RE_{HH}.

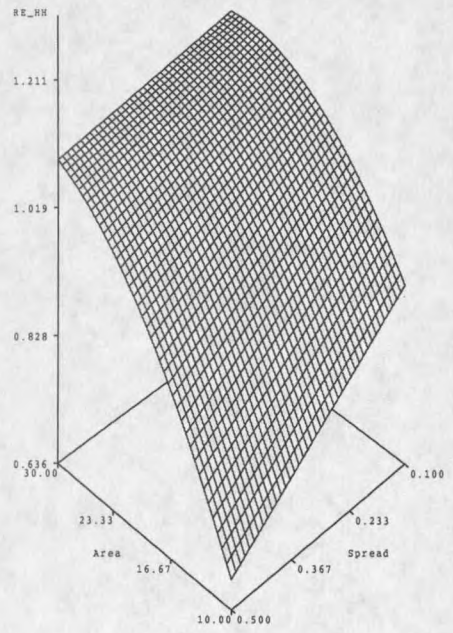


Figure 28. FCA Centerpoint Plot of Spread*Critical for RE_{HH}.

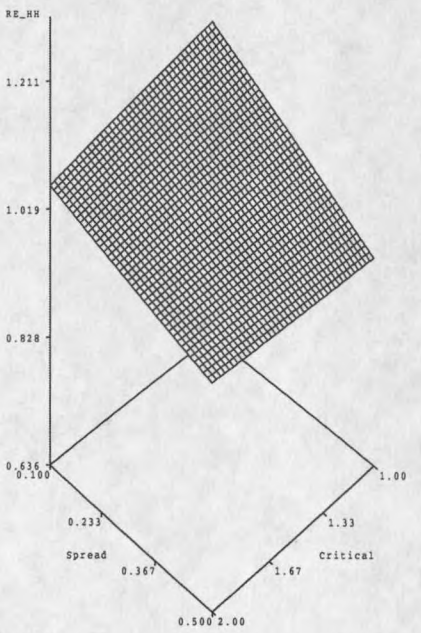


Figure 29. FCA Centerpoint Plot of Spread*Kids for RE_{HH}.

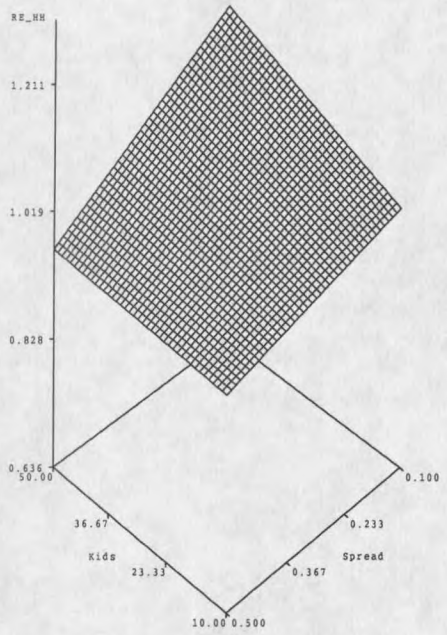


Figure 30. FCA Centerpoint Plot of Spread*Sample for RE_{HH}.

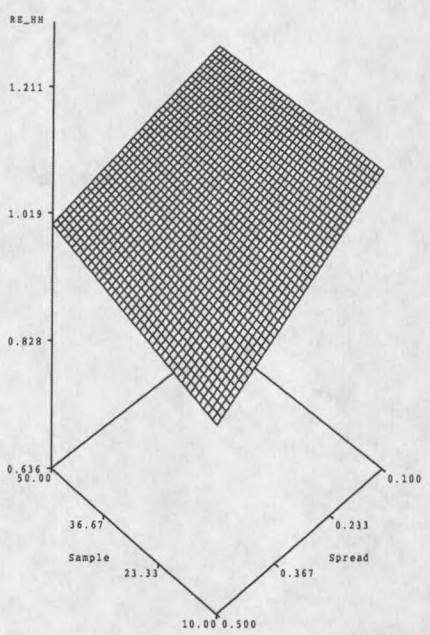


Figure 31. FCA Centerpoint Plot of Parents*Area for RE_{HH}.

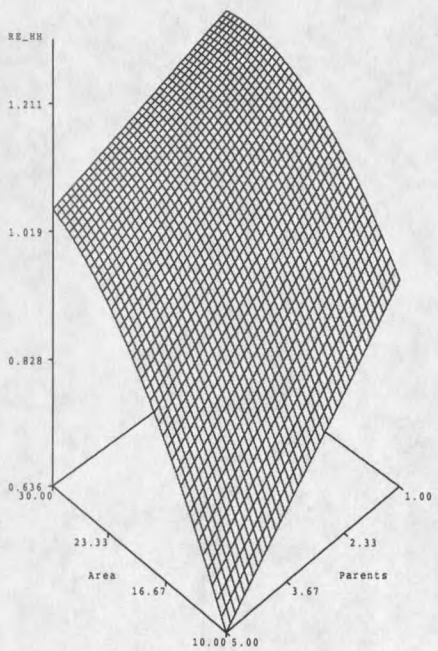


Figure 32. FCA Centerpoint Plot of Parents*Kids for RE_{HH}.

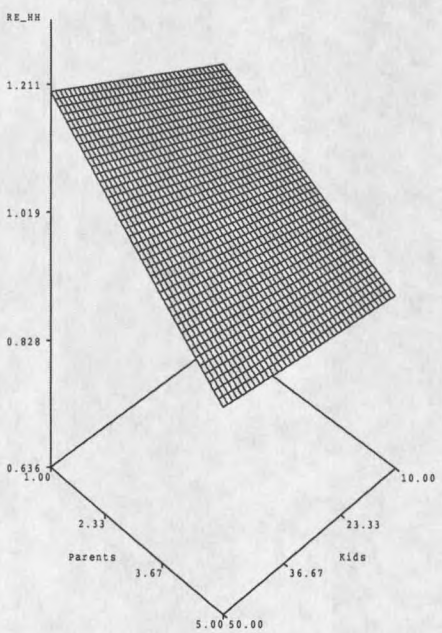


Figure 33. FCA Centerpoint Plot of Area*Critical for RE_{HH}.

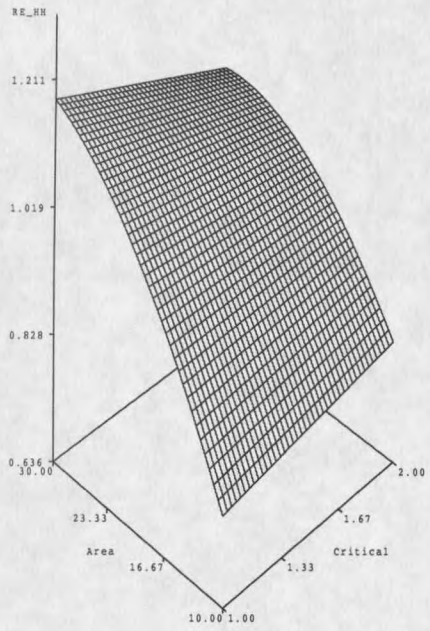


Figure 34. FCA Centerpoint Plot of Area*Kids for RE_{HH}.

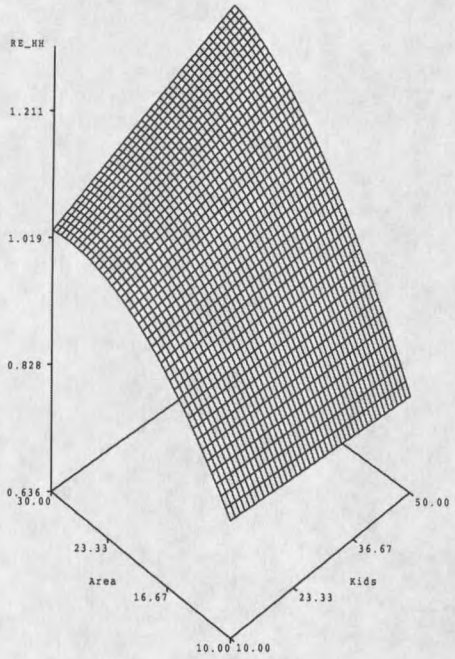


Figure 35. FCA Centerpoint Plot of Kids*Sample for RE_{HH}.

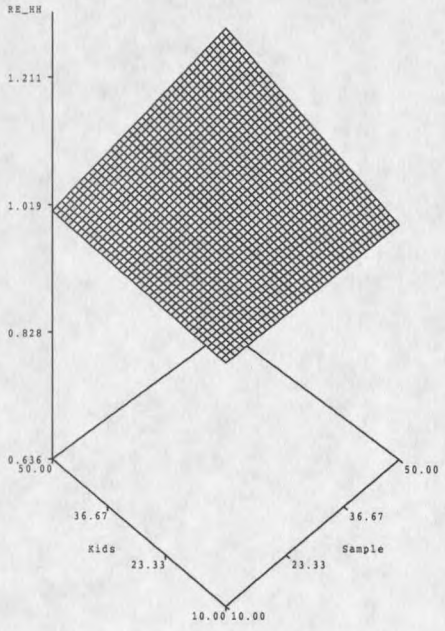


Figure 36. FCA Centerpoint Plot of Area for RE_{HH}.

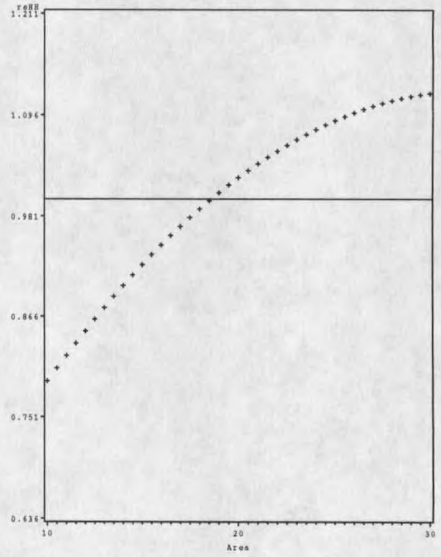


Table 17. Centerpoint FCA of RE_{HH}. Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted RE_{HH}. The levels are uncoded and are in the factor order given in the Effect column.

Effect	Minimum RE_{HH}	Minimum Levels	Maximum RE_{HH}	Maximum Levels	Range RE_{HH}	Maximum RE_{HH} less than 1?
Shape*Spread	0.877	(4,.5)	1.131	(4,,1)	0.255	no
Spread*Parents	0.787	(.5,5)	1.153	(.1,1)	0.366	no
Spread*Area	0.680	(.5,10)	1.151	(.1,29)	0.471	no
Spread*Critical	0.948	(.5,1)	1.136	(.1,1)	0.188	no
Spread*Kids	0.944	(.5,10)	1.166	(.1,50)	0.222	no
Spread*Sample	0.904	(.5,10)	1.106	(.1,50)	0.202	no
Parents*Area	0.636	(5,10)	1.189	(1,29)	0.553	no
Parents*Kids	0.896	(5,10)	1.199	(1,50)	0.303	no
Area*Critical	0.766	(10,1)	1.181	(30,1)	0.415	no
Area*Kids	0.782	(10,50)	1.211	(30,50)	0.429	no
Kids*Sample	0.979	(10,10)	1.118	(50,50)	0.139	no
Area*Area	0.792	(10,10)	1.121	(30,30)	0.329	no

favorability is highlighted in the two-dimensional contour plots by the enlargened, solid contour lines. No two-dimensional plots were made for those cases where either all or no conditions were favorable for ACS. Table 19 and 20 summarize the important results gathered from all the aforementioned plots.

Table 18. Conditional grid search results of the FCA with fixed covariate levels for both the maximum and minimum predicted RE^*_{HT} and RE_{HH}. A controllable factor marked with an asterisk indicates a change occurred between minimum and maximum level settings for the predicted response. Range is the range of predicted responses. no = not observed; na = not applicable; CI = confidence interval.

Factor	Minimum RE^*_{HT}	Maximum RE^*_{HT}	Factor	Minimum RE_{HH}	Maximum RE_{HH}
Shape	4	4	Shape	4	4
Direct	0	0	Direct	na	na
Spread	.1	.1	Spread	.1	.1
Parents	3	3	Parents	3	3
Area*	30	10	Area*	10	30
Critical*	1	2	Critical	1	1
Kids	30	30	Kids	30	30
Sample	10	10	Sample*	10	50
$\hat{y}(\mathbf{x}_0)$	0.667	1.231	$\hat{y}(\mathbf{x}_0)$	0.927	1.280
$SE[\hat{y}(\mathbf{x}_0)]$	0.059	0.063	$SE[\hat{y}(\mathbf{x}_0)]$	0.026	0.026
$y(\mathbf{x}_0)$	no	no	$y(\mathbf{x}_0)$	no	no
95% CI	[0.549, 0.784]	[1.104, 1.359]	95% CI	[0.875, 0.979]	[1.229, 1.331]
range	0.564		range	0.353	

Table 19. Conditional FCA of RE^*_{HT} with fixed covariates set to reasonable levels. Factor levels are for the respective minimum and maximum predicted RE^*_{HT}. The levels are uncoded and are in the factor order given in the Effect column. The setting of the remaining third controllable factor is noted in the Effect column. "@" indicates the effect was included in the model. "*" indicates a correspondence with either the maximum or minimum obtained from the conditional grid search (see Table 18).

Effect	Minimum RE^*_{HT}	Minimum Levels	Maximum RE^*_{HT}	Maximum Levels	Range RE^*_{HT}	Maximum RE^*_{HT} less than 1?
Area*Critical, Sample = 10	0.667*	(30,1)	1.231*	(10,2)	0.564	no
@ Area*Sample, Critical = 1	0.667*	(30,10)	1.102	(10,10)	0.435	no
@ Area*Sample, Critical = 2	0.796	(30,10)	1.231*	(10,10)	0.435	no
Critical*Sample, Area = 10	0.694	(1,50)	1.231*	(2,10)	0.537	no
Critical*Sample, Area = 30	0.667*	(1,10)	0.976	(2,50)	0.309	yes

Figure 37. FCA Conditional Plot of Area*Critical for RE^*_{HT}, Sample = 10.

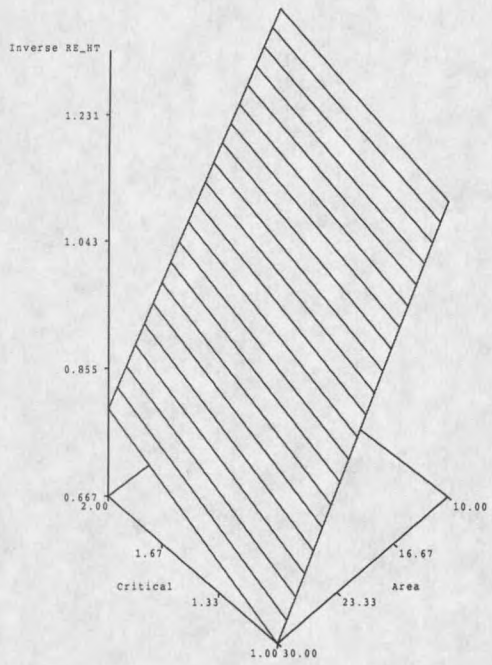


Figure 38. FCA Conditional Plot of Area*Critical for RE^*_{HT}, Sample = 10.

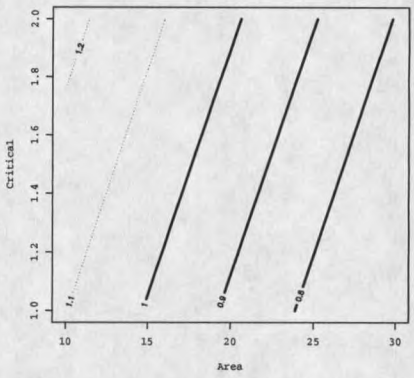


Figure 39. FCA Conditional Plot of Area*Sample for RE^*_{HT}, Critical = 1.

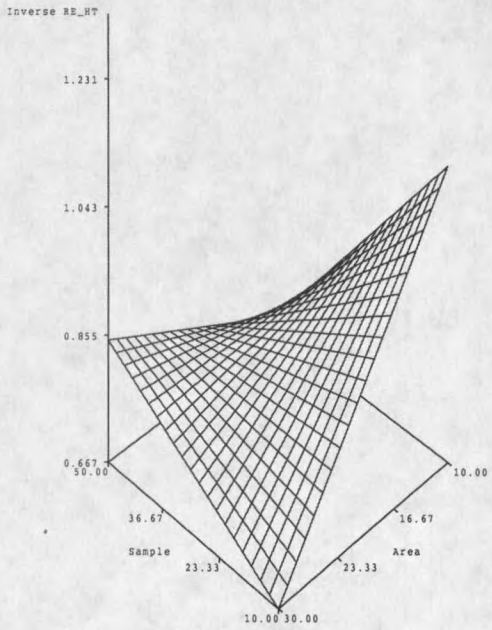


Figure 40. FCA Conditional Plot of Area*Sample for RE^*_{HT}, Critical = 1.

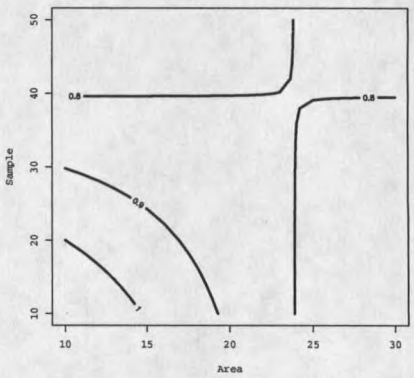


Figure 41. FCA Conditional Plot of Area*Sample for RE^*_{HT}, Critical = 2.

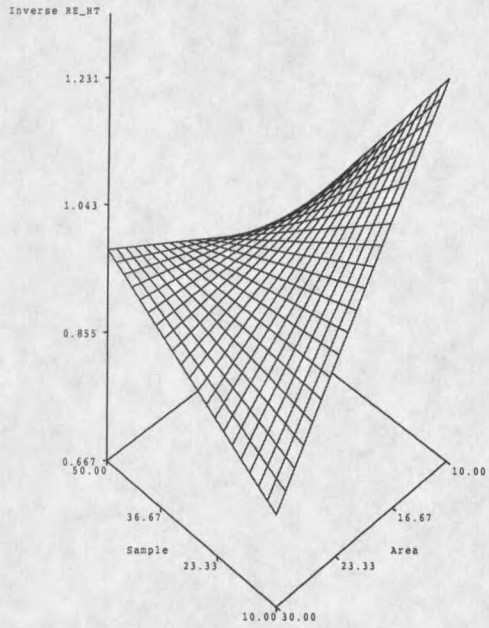


Figure 42. FCA Conditional Plot of Area*Sample for RE^*_{HT}, Critical = 2.

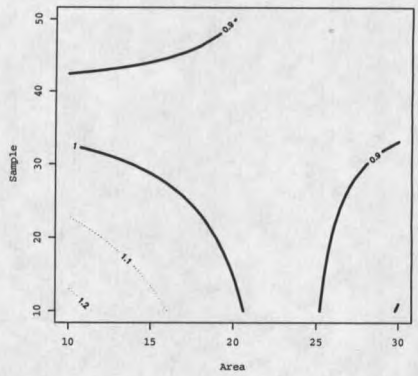


Figure 43. FCA Conditional Plot of Critical*Sample for RE^*_{HT}, Area = 10.

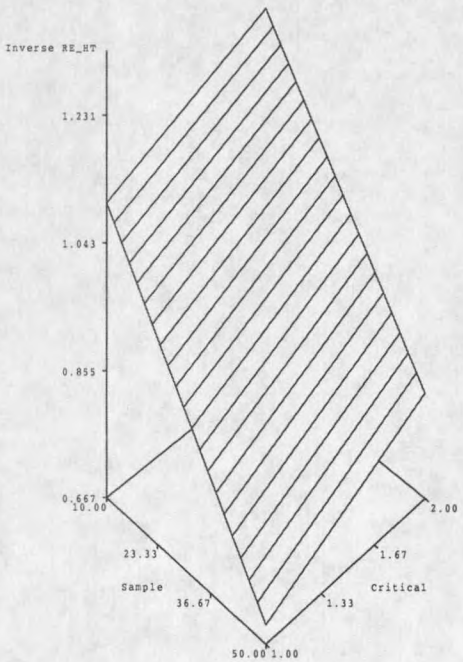


Figure 44. FCA Conditional Plot of Critical*Sample for RE^*_{HT}, Area = 10.

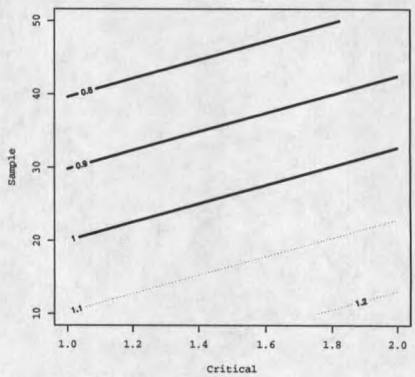


Figure 45. FCA Conditional Plot of Critical*Sample for RE^*_{HT}, Area = 30.

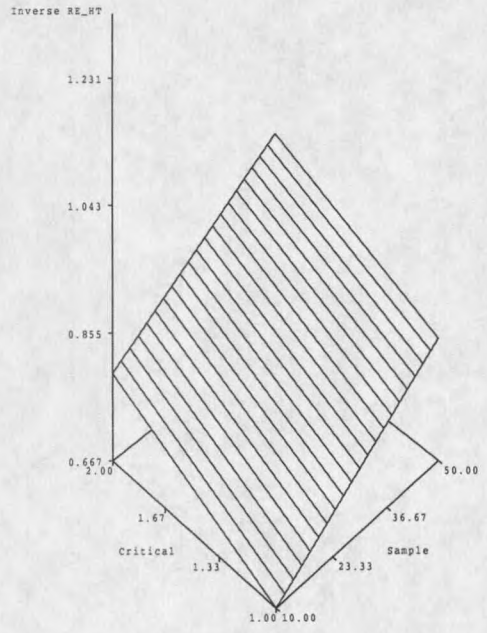


Figure 46. FCA Conditional Plot of Critical*Sample for RE^*_{HT}, Area = 30.

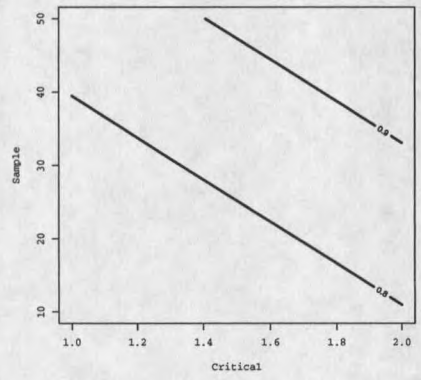


Figure 47. FCA Conditional Plot of Area*Critical for RE_{HH}, Sample = 10.

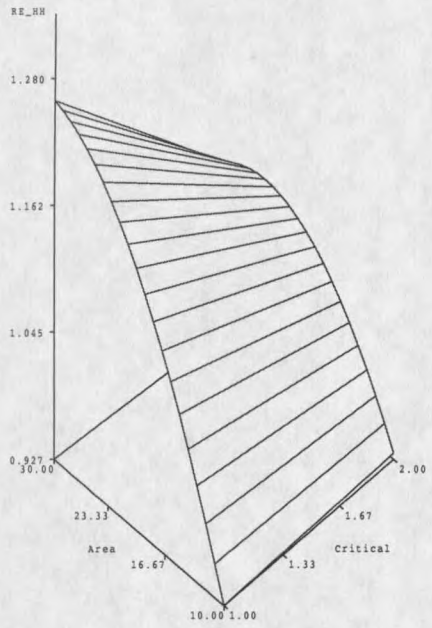


Figure 48. FCA Conditional Plot of Area*Critical for RE_{HH}, Sample = 10.

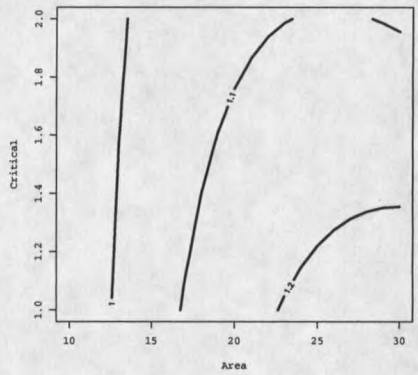


Figure 49. FCA Conditional Plot of Area*Critical for RE_{HH}, Sample = 50.

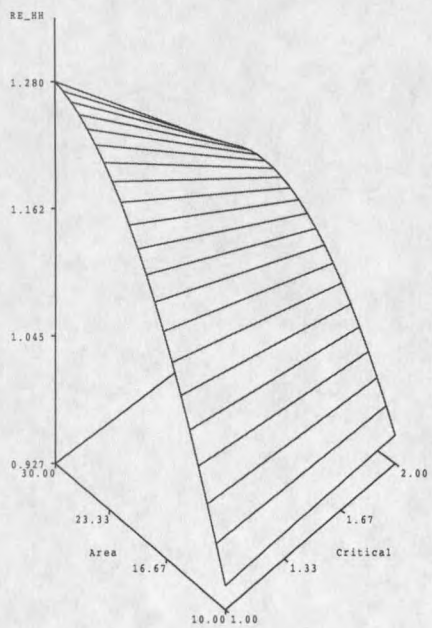


Figure 50. FCA Conditional Plot of Area*Critical for RE_{HH}, Sample = 50.

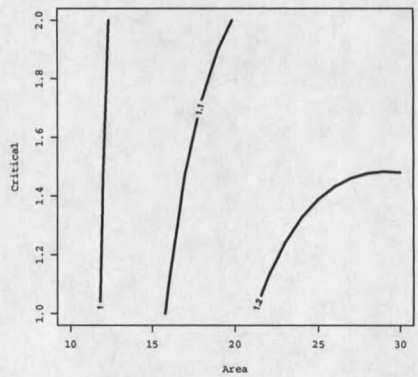


Figure 51. FCA Conditional Plot of Area*Sample for RE_{HH}, Critical = 1.

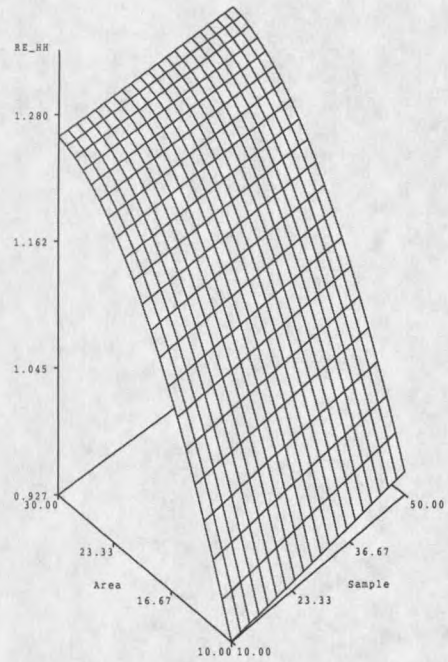


Figure 52. FCA Conditional Plot of Area*Sample for RE_{HH}, Critical = 1.

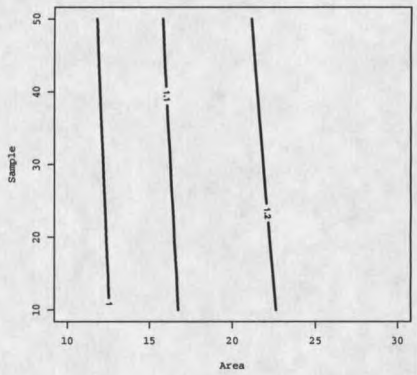


Figure 53. FCA Conditional Plot of Critical*Sample for RE_{HH}, Area = 10.

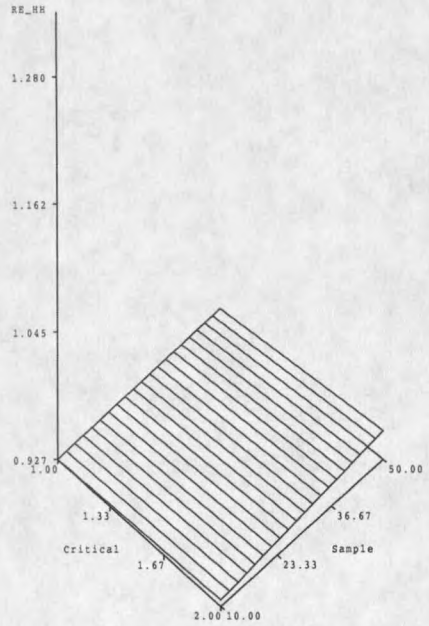


Figure 54. FCA Conditional Plot of Critical*Sample for RE_{HH}, Area = 30.

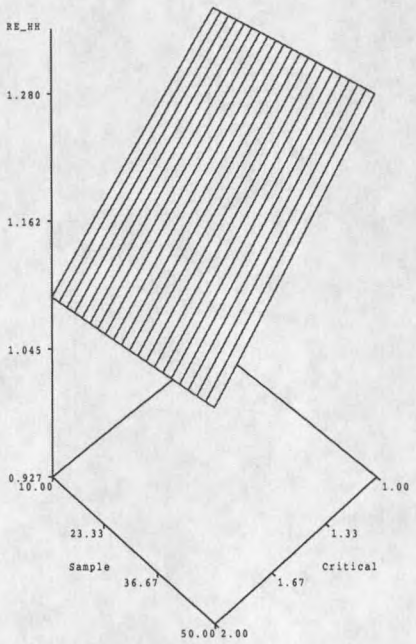


Table 20. Conditional FCA of RE_{HH} with fixed covariates set to reasonable levels. Factor levels are for the respective minimum and maximum predicted RE_{HH}. The levels are uncoded and are in the factor order given in the Effect column. The setting of the remaining third controllable factor is noted in the Effect column. "@" indicates the effect was included in the model. "*" indicates a correspondence with either the maximum or minimum obtained from the conditional grid search (see Table 18).

Effect	Minimum RE_{HH}	Minimum Levels	Maximum RE_{HH}	Maximum Levels	Range RE_{HH}	Maximum RE_{HH} less than 1?
@ Area*Critical, Sample = 10	0.927*	(10,1)	1.259	(30,1)	0.332	no
@ Area*Critical, Sample = 50	0.948	(10,1)	1.280*	(30,1)	0.332	no
Area*Sample, Critical = 1	0.927*	(10,10)	1.280*	(30,50)	0.353	no
Critical*Sample, Area = 10	0.927*	(1,10)	0.953	(2,50)	0.026	yes
Critical*Sample, Area = 30	1.093	(2,10)	1.280*	(1,50)	0.187	no

Results for RE^*_{HT} and RE_{HH} Data - The Random Covariate Analysis

As mentioned in the Methods, multicollinearity issues required a reduction in the number of candidate terms considered for the variable selection process. Results from the use of correlation matrices, variance inflation factors and ridge regression suggested the following starting set of terms for use in the model selection process for RE^*_{HT} : Pervarw, Numnets, Rarity, Pervarw*Numnets, Pervarw*Pervarw, Numnets*Numnets, and Rarity*Rarity. For the RE_{HH} response, the starting set was: Pervarw, Numnets, Predp, Rarity, Pervarw*Numnets, Pervarw*Predp, Numnets*Predp, and the four associated quadratic terms. Results of the variable selection process

described in the Methods yielded the models summarized in Table 21 for RE^*_{HT} and Table 22 for RE_{HH}.

Table 21. SAS computer output for the reduced RCA model for RE^*_{HT}. Effects unique to the model for this response are proceeded by #.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5.38018	1.07604	49.66	0.0001
Error	66	1.43005	0.02167		
Corrected Total	71	6.81023			
Root MSE	0.14720	R-Square	0.7900		
Dependent Mean	0.88324	Adj R-Sq	0.7741		
Coeff Var	16.66569				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > $\|t\|$
Intercept	1	1.666106	0.06525	25.54	0.0001
Pervarw	1	-0.53805	0.04693	-11.46	0.0001
Numnets	1	0.11943	0.02284	5.23	0.0001
Rarity	1	1.03294	0.07200	14.35	0.0001
Pervarw*Numnets	1	0.18649	0.03980	4.69	0.0001
Pervarw*Pervarw #	1	-0.14388	0.06669	-2.16	0.0346

For both relative efficiency responses, there were no unusual variance inflation factors noted after the model building process. However, an examination of the influence diagnostics revealed several "unusual" points, most notably two large Studentized residual values. Upon running the analysis with and without these two observations, negligible changes occurred with respect to the parameter estimates and the standard

Table 22. SAS computer output for the reduced RCA model for RE_{HH}. Effects unique to the model for this response are proceeded by #.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	3.98695	0.44299	100.80	0.0001
Error	62	0.27247	0.00439		
Corrected Total	71	4.25942			
Root MSE	0.06629	R-Square	0.9360		
Dependent Mean	0.96754	Adj R-Sq	0.9267		
Coeff Var	6.85158				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > $\|t\|$
Intercept	1	0.50079	0.04631	10.81	0.0001
Pervarw	1	0.22096	0.01980	11.16	0.0001
Numnets	1	0.13171	0.01085	12.14	0.0001
Predp #	1	-0.06092	0.01214	-5.02	0.0001
Rarity	1	-0.35039	0.03380	-10.37	0.0001
Pervarw*Numnets	1	0.12401	0.01820	6.81	0.0001
Pervarw*Predp #	1	-0.03806	0.02106	-1.81	0.0755
Numnets*Predp #	1	0.02456	0.01151	2.13	0.0368
Numnets*Numnets #	1	-0.08150	0.02609	-3.12	0.0027
Rarity*Rarity #	1	0.39028	0.04414	8.84	0.0001

errors. Consequently, it was decided to include the two observations. There was nothing to suggest any potential violation of the assumption of constant variance. In light of the aforementioned unusual points, evidence of a departure from the normality assumption for the errors was noted, more so for RE^*_{HT}, but it was felt that it was not severe enough to warrant corrective action.

Table 23 contains the RCA global grid search results for both the maximum and minimum predicted RE^*_{HT} and RE_{HH}. As was the case for the FCA, for the minimum

RE^*_{HT} response, a negative predicted value was obtained and was then set to 0 for the range calculations. In fact, the confidence interval yielded a range of strictly negative values and was also set to 0. Also, for the minimum RE_{HH} response, a negative lower confidence limit was obtained for $\mu_{y|\mathbf{x}_0}$ and so was set to 0.

Table 23. The RCA global grid search results for maximum and minimum predicted RE^*_{HT} and RE_{HH}. A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. Range is the range of predicted responses. no = not observed; na = not applicable; CI = confidence interval.

Factor	Minimum RE^*_{HT}	Maximum RE^*_{HT}	Factor	Minimum RE_{HH}	Maximum RE_{HH}
Pervarw*	0.547	0.000	Pervarw*	0.000	0.547
Numnets	73.0	73.0	Numnets*	73.0	900.0
Predp	na	na	Predp*	5.3	0.8
Rarity*	1.009	4.521	Rarity*	3.819	1.009
$\hat{y}(\mathbf{x}_0)$	-0.355	3.160	$\hat{y}(\mathbf{x}_0)$	0.073	1.711
$SE[\hat{y}(\mathbf{x}_0)]$	0.1156	0.1509	$SE[\hat{y}(\mathbf{x}_0)]$	0.066	0.047
$y(\mathbf{x}_0)$	no	no	$y(\mathbf{x}_0)$	no	no
95% CI	0	[2.859, 3.462]	95% CI	[0.000, 0.206]	[1.618, 1.804]
range$[\hat{y}(\mathbf{x}_0)]$	3.160		range$[\hat{y}(\mathbf{x}_0)]$	1.638	

Figures 55 through 62 are the three-dimensional RCA interaction response surface plots for the condition where all other variables are fixed to the centerpoint (fixed to 0). These include several two-dimensional quadratic response surface plots. Tables 24 and 25 summarize the important results gathered from all the centerpoint plots.

Figures 63 through 96 are the two- and three-dimensional interaction response surface plots where the levels of the random covariates are determined by the results of the global analysis. Once again, the region of favorability is highlighted in the two-dimensional contour plots by the enlargened, solid contour lines. As with the FCA, for

Table 24. The centerpoint RCA of RE^*_{HT}. Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted RE^*_{HT}. The levels are uncoded and are in the factor order given in the Effect column.

Effect	Minimum RE^*_{HT}	Minimum Levels	Maximum RE^*_{HT}	Maximum Levels	Range RE^*_{HT}	Maximum RE^*_{HT} less than 1?
Pervarw*Numnets	0.678	(0.547,73.0)	2.127	(0.000,73.0)	1.449	no
Pervarw*Pervarw	0.984	(0.547)	2.060	(0.000)	1.076	no
Rarity	0.633	(1.009)	2.699	(4.521)	2.066	no

Table 25. The centerpoint RCA of RE_{HH}. Variables not included in the effect are set to centerpoint levels. Factor levels are for the respective minimum and maximum predicted RE_{HH}. The levels are uncoded and are in the factor order given in the Effect column.

Effect	Minimum RE_{HH}	Minimum Levels	Maximum RE_{HH}	Maximum Levels	Range RE_{HH}	Maximum RE_{HH} less than 1?
Pervarw*Numnets	0.191	(0.000,73.0)	0.896	(0.547,900.0)	0.705	yes
Pervarw*Predp	0.257	(0.000,5.3)	0.821	(0.547,0.8)	0.564	yes
Numnets*Predp	0.202	(73.0,5.3)	0.597	(755.3,0.8)	0.395	yes
Numnets*Numnets	0.288	(73.0)	0.554	(817.3)	0.266	yes
Rarity*Rarity	0.422	(3.555)	1.241	(1.009)	0.819	no

several cases, two-dimensional plots were not generated when it was determined that there were either no conditions favorable for ACS, all the conditions were favorable for ACS, or there existed only a single set of conditions favorable for ACS. Tables 26 and 27 summarize the important results gathered from all of the aforementioned plots.

Table 26. Conditional RCA of RE^*_{HT}. Any remaining random covariate not included in the effect is set to a reasonable level. Factor levels are for the respective minimum and maximum predicted RE^*_{HT}. The levels are uncoded and are in the factor order given in the Effect column. The setting of any remaining random covariate is noted in the Effect column. "@" indicates the effect was included in the model. "*" indicates a correspondence with either the maximum or minimum obtained from the global grid search (see Table 23).

Effect	Minimum RE^*_{HT}	Minimum Levels	Maximum RE^*_{HT}	Maximum Levels	Range RE^*_{HT}	Maximum RE^*_{HT} less than 1?
Pervarw*Rarity, Numnets = 73.0	-0.355*	(0.547,1.009)	3.160*	(0.000,4.521)	3.160	no
Numnets*Rarity, Pervarw = 0.000	0.960	(900.0,1.009)	3.160*	(73.0,4.521)	2.200	no
Numnets*Rarity, Pervarw = 0.547	-0.355*	(73.0,1.009)	2.323	(900.0,4.521)	2.678	no
@ Pervarw*Numnets, Rarity = 1.009	-0.355*	(0.547,73.0)	1.094	(0.000,73.0)	1.449	no
@ Pervarw*Numnets, Rarity = 4.521	1.711	(0.547,73.0)	3.160*	(0.000,73.0)	1.449	no
@ Pervarw, Numnets = 73.0 Rarity = 1.009	-0.355*	(0.547)	1.094	(0.000)	1.449	no
@ Pervarw, Numnets = 73.0 Rarity = 4.521	1.711	(0.547)	3.160*	(0.000)	1.449	no
@ Rarity, Numnets = 73.0 Pervarw = 0.000	-0.355*	(1.009)	1.711	(4.521)	2.066	no
@ Rarity, Numnets = 73.0 Pervarw = 0.547	1.094	(1.009)	3.160*	(4.521)	2.066	no

Figure 55. RCA Centerpoint Plot of Pervarw*Numnets for RE^*_{HT}.

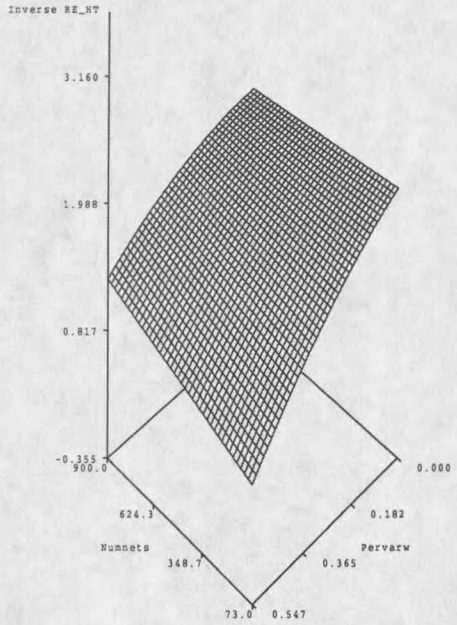


Figure 56. RCA Centerpoint Plot of Pervarw for RE^*_{HT}.

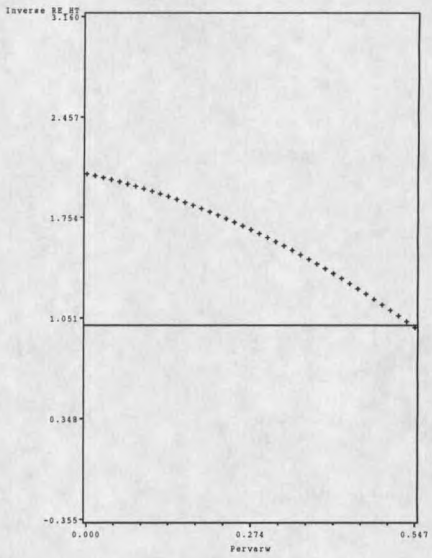


Figure 57. RCA Centerpoint Plot of Rarity for RE^*_{HT}.

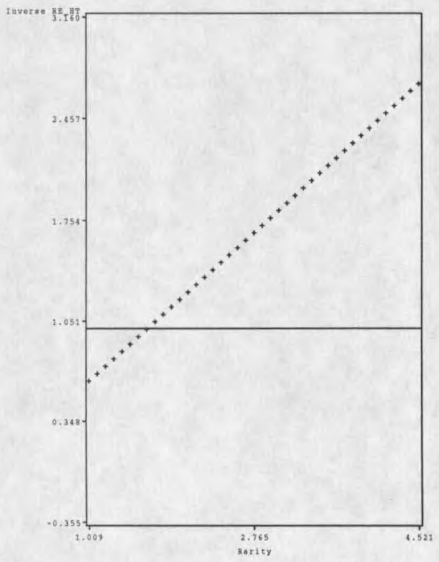


Figure 58. RCA Centerpoint Plot of Pervarw*Numnets for RE_{HH}.

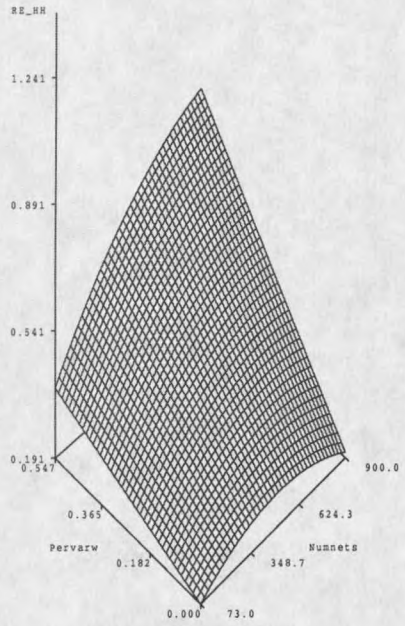


Figure 59. RCA Centerpoint Plot of Pervarw*Predp for RE_{HH}.

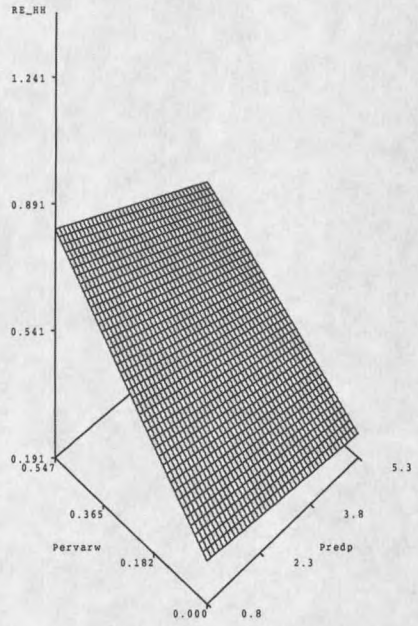


Figure 60. RCA Centerpoint Plot of Numnets*Predp for RE_{HH}.

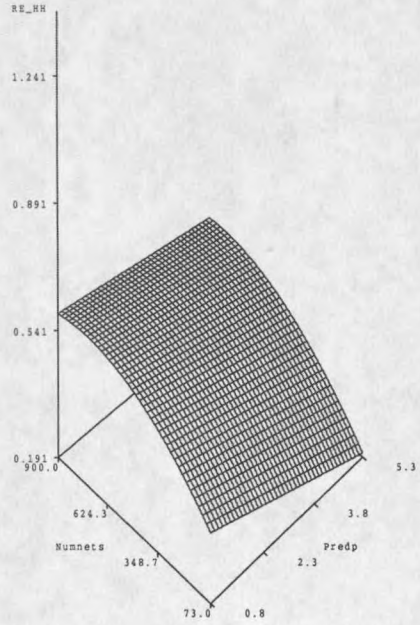


Figure 61. RCA Centerpoint Plot of Numnets for RE_{HH}.

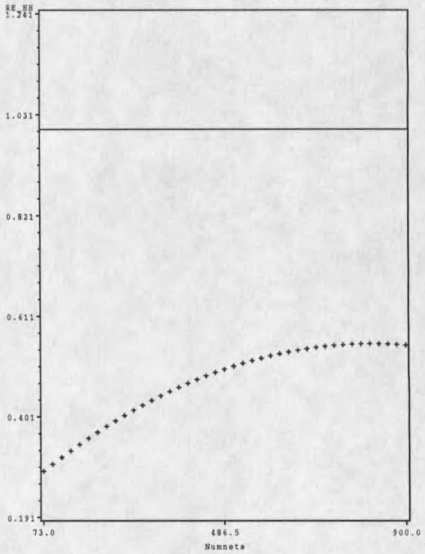


Figure 62. RCA Centerpoint Plot of Rarity for RE_{HH}.

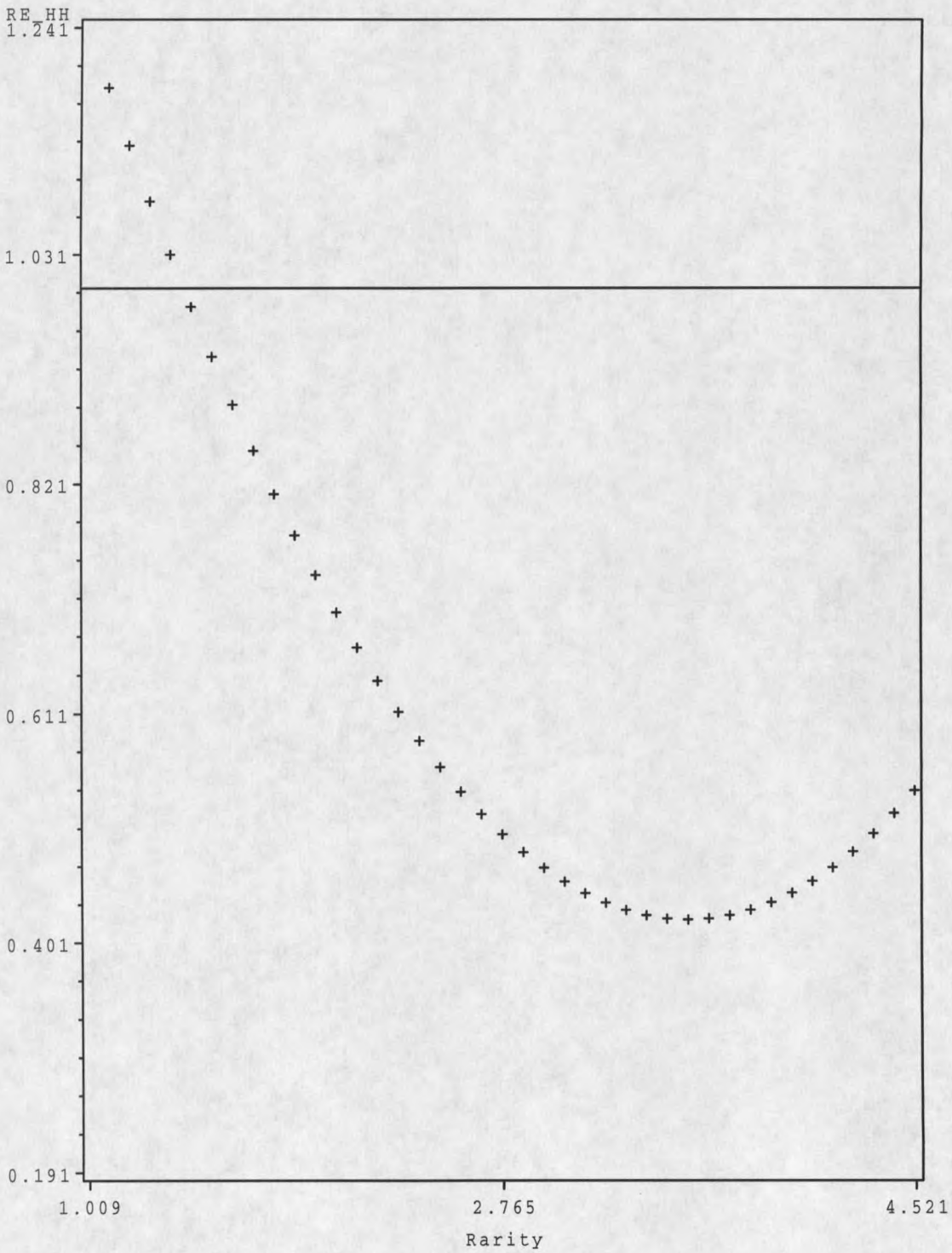


Figure 63. RCA Conditional Plot of Pervarw*Rarity for RE^*_{HT}, Numnets = 73.0.

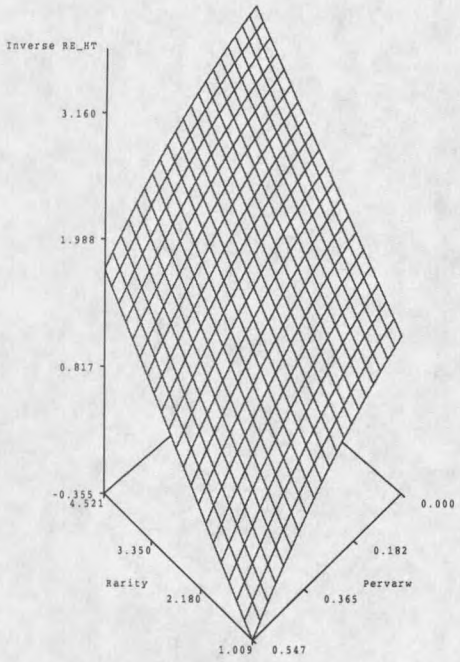


Figure 64. RCA Conditional Plot of Pervarw*Rarity for RE^*_{HT}, Numnets = 73.0.

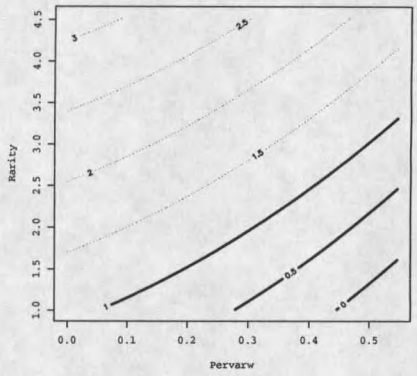


Figure 65. RCA Conditional Plot of Numnets*Rarity for RE^*_{HT}, Pervarw = 0.000.

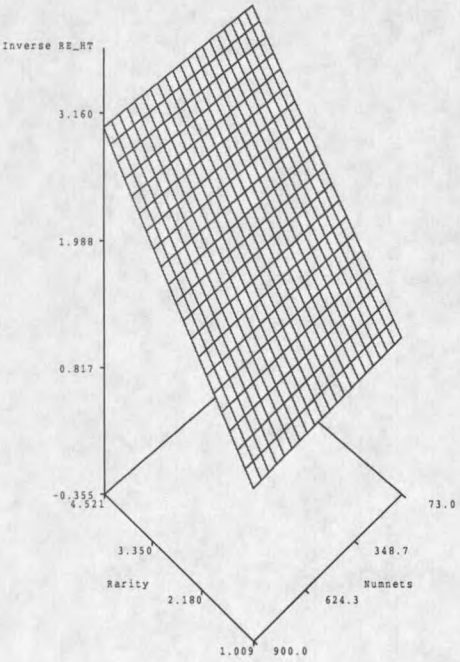


Figure 66. RCA Conditional Plot of Numnets*Rarity for RE^*_{HT}, Pervarw = 0.000.

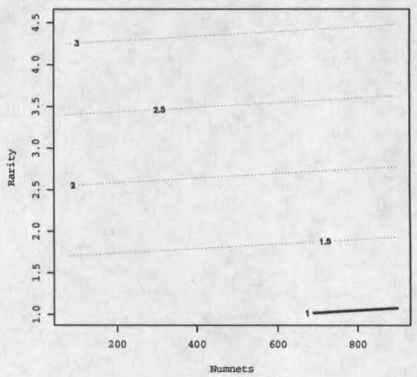


Figure 67. RCA Conditional Plot of Numnets*Rarity for RE^*_{HT}, Pervarw = 0.547.

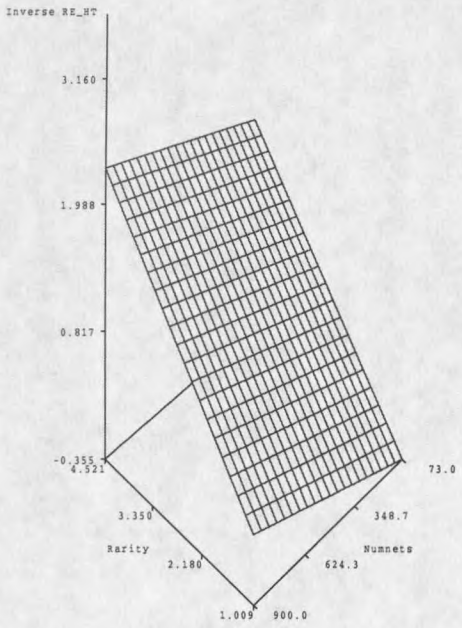


Figure 68. RCA Conditional Plot of Numnets*Rarity for RE^*_{HT}, Pervarw = 0.547.

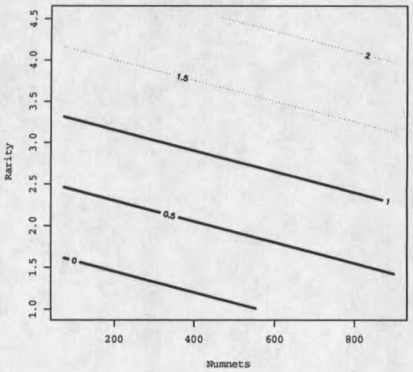


Figure 69. RCA Conditional Plot of Pervarw*Numnets for RE^*_{HT}, Rarity = 1.009.

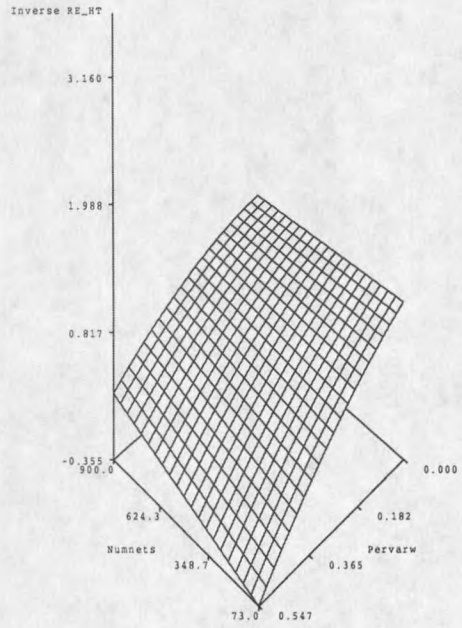


Figure 70. RCA Conditional Plot of Pervarw*Numnets for RE^*_{HT}, Rarity = 1.009.

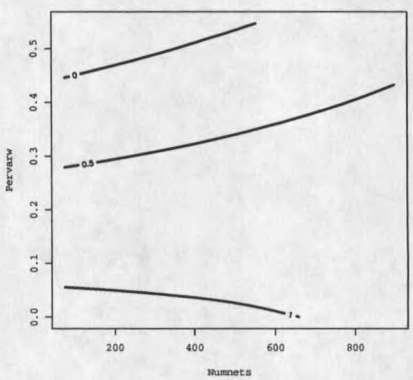


Figure 71. RCA Conditional Plot of Pervarw*Numnets for RE^*_{HT}, Rarity = 4.521.

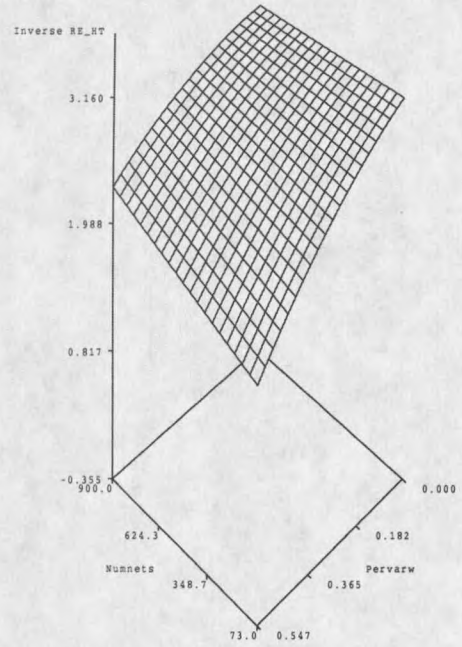


Figure 72. RCA Conditional Plot of Pervarw for RE^*_{HT}, Numnets = 73.0, Rarity = 1.009.

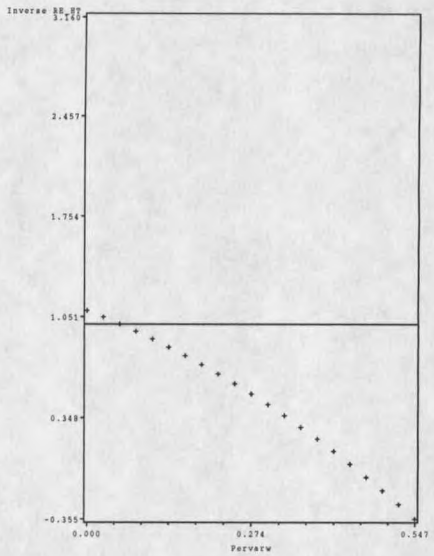


Figure 73. RCA Conditional Plot of Pervarw for RE^*_{HT}, Numnets = 73.0, Rarity = 4.521.

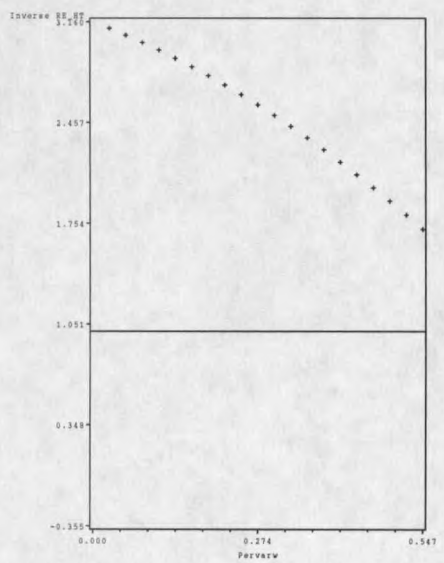


Figure 74. RCA Conditional Plot of Rarity for RE^*_{HT}, Numnets = 73.0, Pervarw = 0.000.

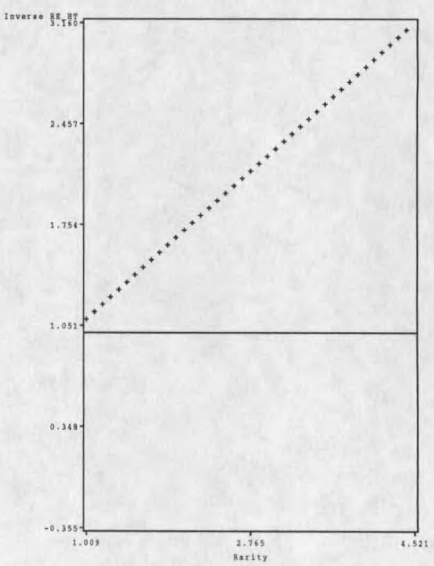


Figure 75. RCA Conditional Plot of Rarity for RE^*_{HT}, Numnets = 73.0, Pervarw = 0.547.

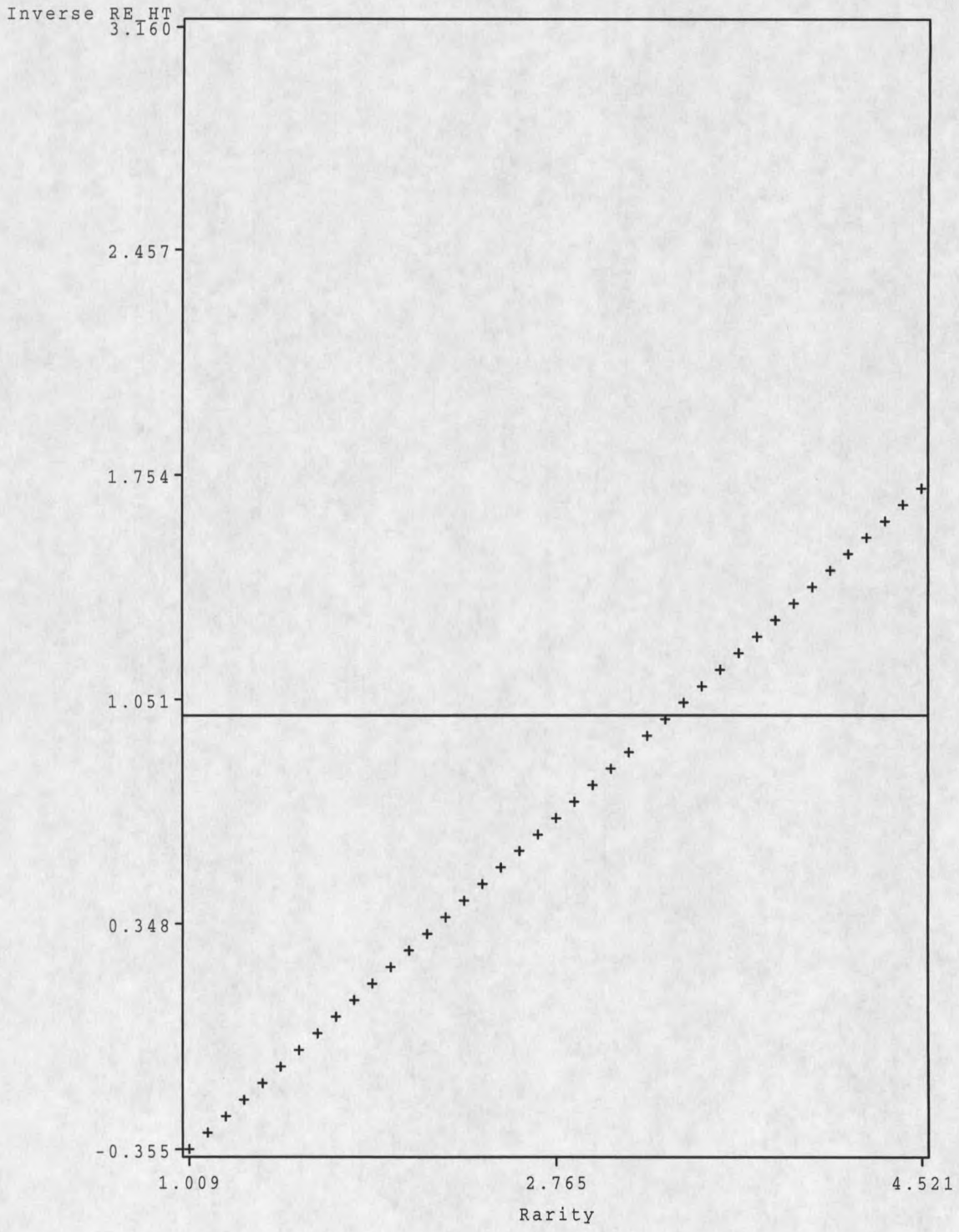


Figure 76. RCA Conditional Plot of Pervarw*Numnets for RE_{HH}, Predp = 0.8, Rarity = 1.009.

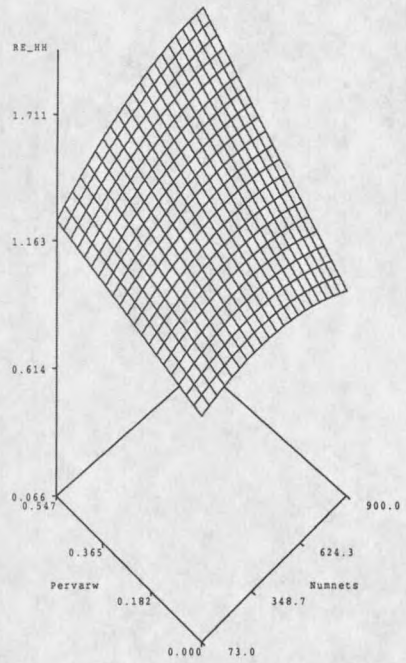


Figure 77. RCA Conditional Plot of Pervarw*Numnets for RE_{HH}, Predp = 0.8, Rarity = 1.009.

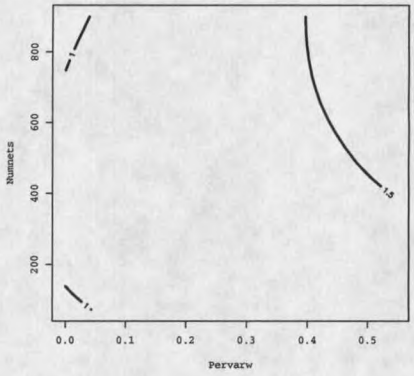


Figure 78. RCA Conditional Plot of Pervarw*Numnets for RE_{HH}, Predp = 5.3, Rarity = 3.819.

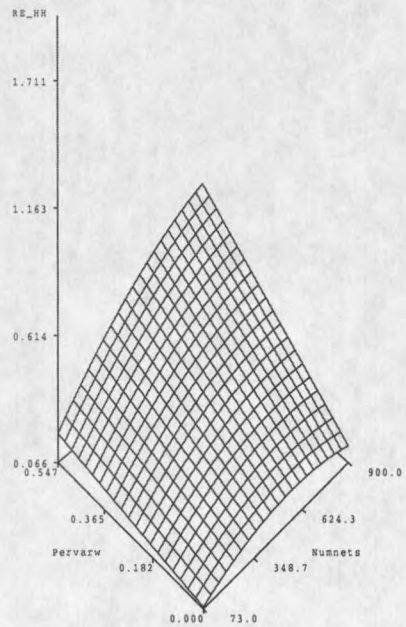


Figure 79. RCA Conditional Plot of Pervarw*Predp for RE_{HH}, Numnets = 900.0, Rarity = 1.009.

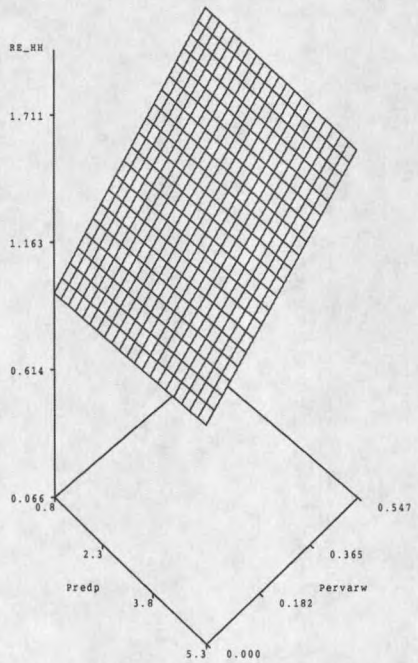


Figure 80. RCA Conditional Plot of Pervarw*Predp for RE_{HH}, Numnets = 900.0, Rarity = 1.009.

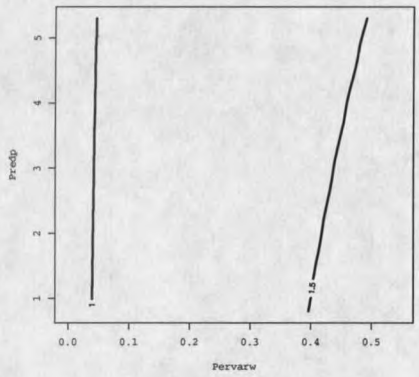


Figure 81. RCA Conditional Plot of Pervarw*Predp for RE_{HH}, Numnets = 73.0, Rarity = 3.819.

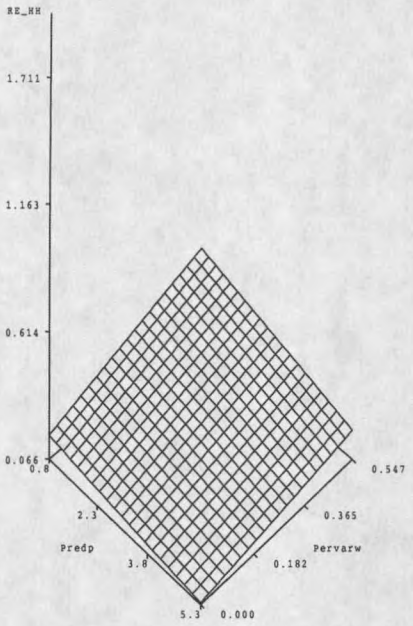


Figure 82. RCA Conditional Plot of Pervarw*Rarity for RE_{HH}, Numnets = 900.0, Predp = 0.8.

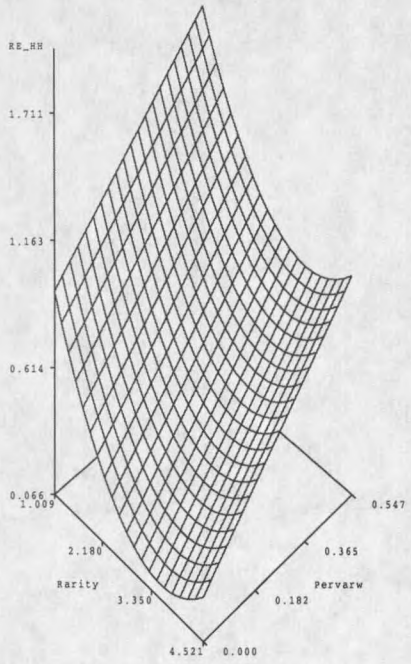


Figure 83. RCA Conditional Plot of Pervarw*Rarity for RE_{HH}, Numnets = 900.0, Predp = 0.8.

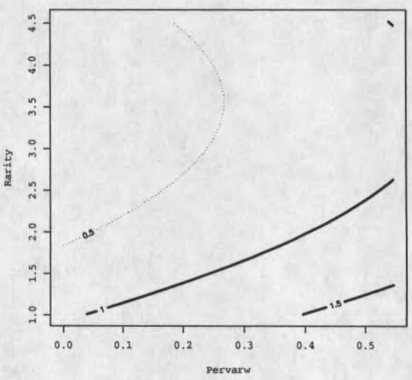


Figure 84. RCA Conditional Plot of Pervarw*Rarity for RE_{HH}, Numnets = 73.0, Predp = 5.3.

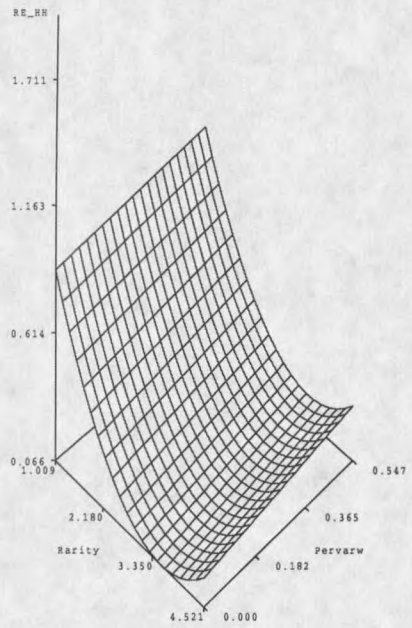


Figure 85. RCA Conditional Plot of Numnets*Predp for RE_{HH}, Pervarw = 0.547, Rarity = 1.009.

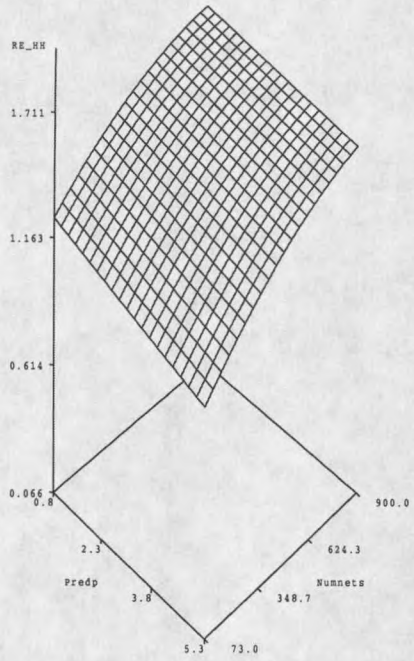


Figure 86. RCA Conditional Plot of Numnets*Predp for RE_{HH}, Pervarw = 0.000, Rarity = 3.819.

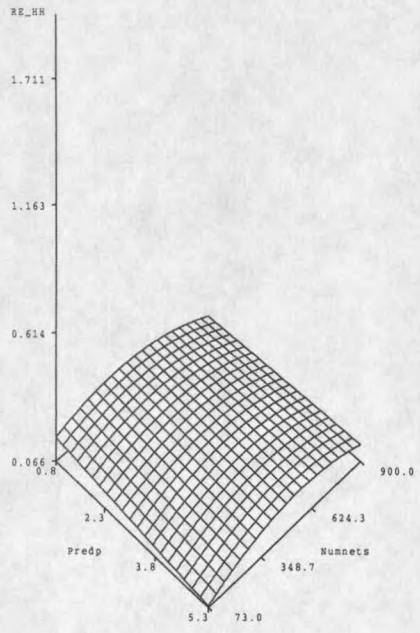


Figure 87. RCA Conditional Plot of Numnets*Rarity for RE_{HH}, Pervarw = 0.547, Predp = 0.8.

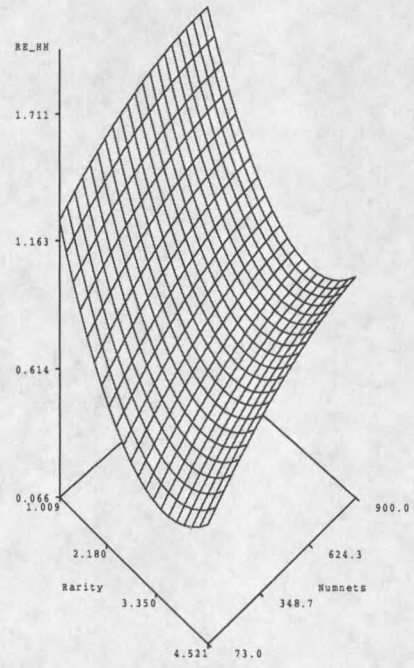


Figure 88. RCA Conditional Plot of Numnets*Rarity for RE_{HH}, Pervarw = 0.547, Predp = 0.8.

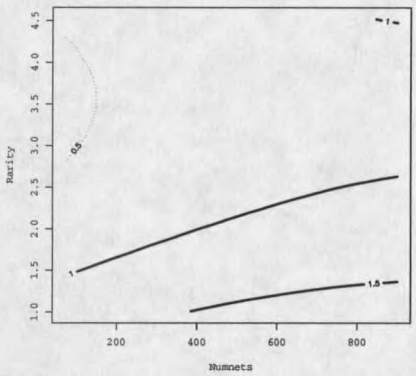


Figure 89. RCA Conditional Plot of Numnets*Rarity for RE_{HH}, Pervarw = 0.000, Predp = 5.3.

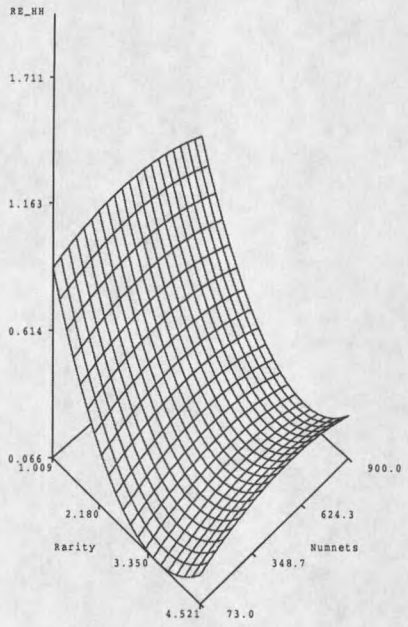


Figure 90. RCA Conditional Plot of Predp*Rarity for RE_{HH}, Pervarw = 0.547, Numnets = 900.0.

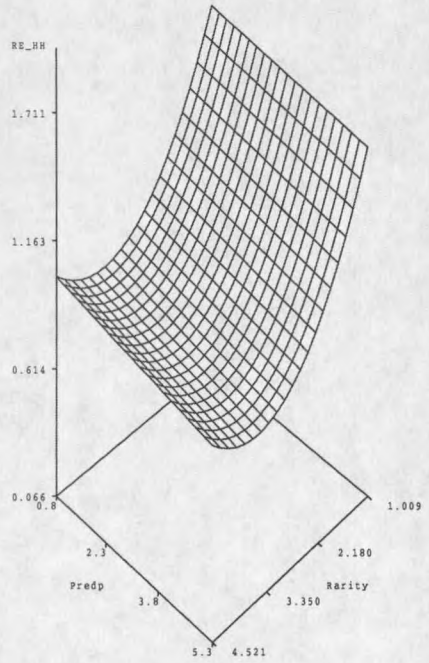


Figure 91. RCA Conditional Plot of Predp*Rarity for RE_{HH}, Pervarw = 0.547, Numnets = 900.0.

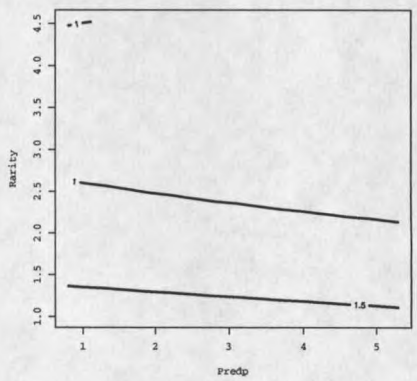


Figure 92. RCA Conditional Plot of Predp*Rarity for RE_{HH}, Pervarw = 0.000, Numnets = 73.0.

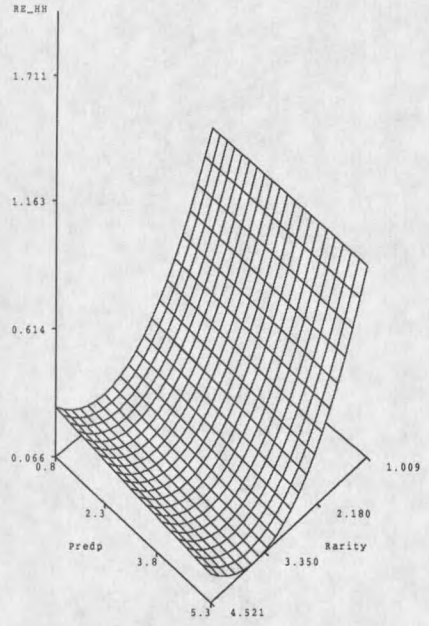


Figure 93. RCA Conditional Plot of Numnets for RE_{HH}, Pervarw = 0.547, Predp = 0.8, Rarity = 1.009.

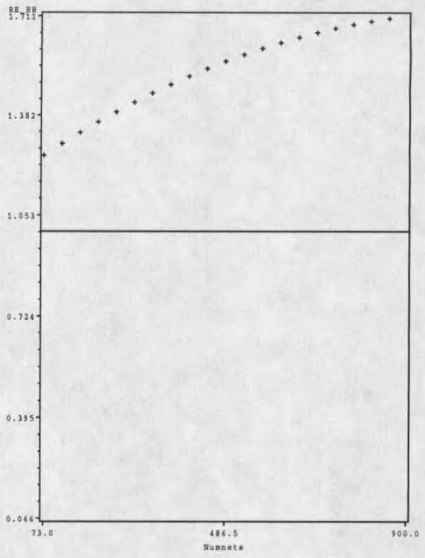


Figure 94. RCA Conditional Plot of Numnets for RE_{HH}, Pervarw = 0.000, Predp = 5.3, Rarity = 3.819.

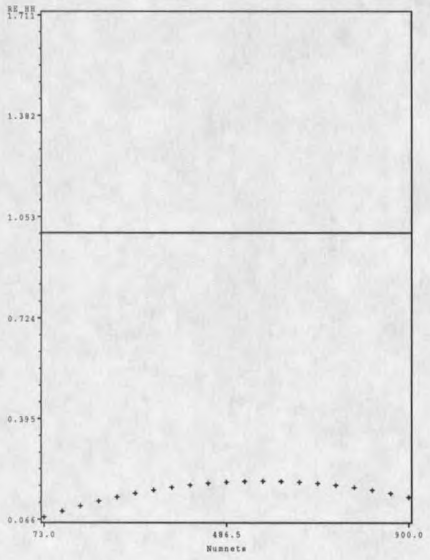


Figure 95. RCA Conditional Plot of Rarity for RE_{HH}, Pervarw = 0.547, Numnets = 900.0, Predp = 0.8.

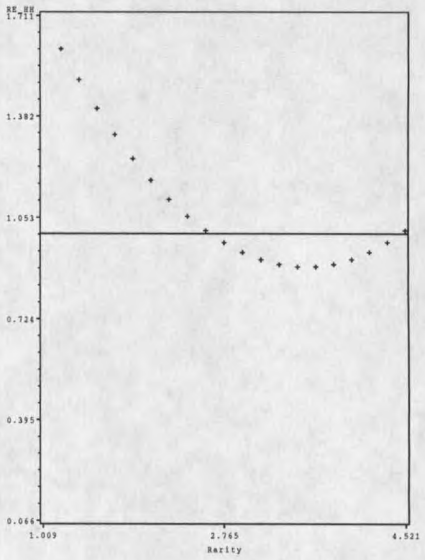


Figure 96. RCA Conditional Plot of Rarity for RE_{HH}, Pervarw = 0.000, Numnets = 73.0, Predp = 5.3.

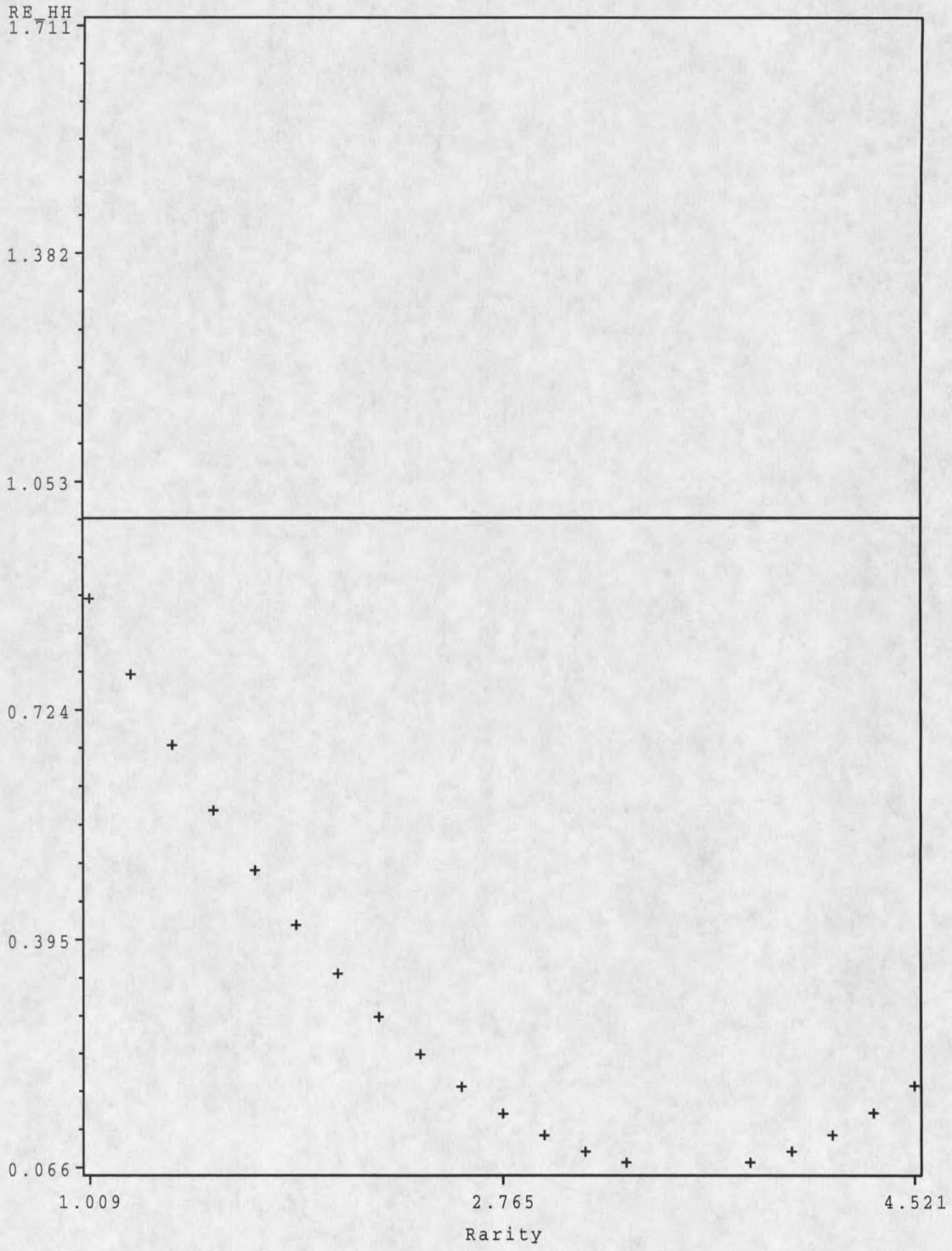


Table 27. Conditional RCA of RE_{HH}. Any remaining random covariate not included in the effect is set to a reasonable level. Factor levels are for the respective minimum and maximum predicted RE_{HH}. The levels are uncoded and are in the factor order given in the Effect column. The setting of any remaining random covariate is noted in the Effect column. "@" indicates the effect was included in the model. "*" indicates a correspondence with either the maximum or minimum obtained from the global grid search (see Table 23). "†" indicates a value that was actually lower than the minimum obtained from the global grid search.

Effect	Minimum RE_{HH}	Minimum Levels	Maximum RE_{HH}	Maximum Levels	Range RE_{HH}	Maximum RE_{HH} less than 1?
@ Pervarw*Numnets, Predp = 0.8 Rarity = 1.009	0.945	(0.000,900.0)	1.711*	(0.547,900.0)	0.766	no
@ Pervarw*Numnets, Predp = 5.3 Rarity = 3.819	0.073*	(0.000,73.0)	0.752	(0.547,900.0)	0.678	yes
@ Pervarw*Predp, Numnets = 900.0 Rarity = 1.009	0.945	(0.000,0.8)	1.711*	(0.547,0.8)	0.766	no
@ Pervarw*Predp, Numnets = 73.0 Rarity = 3.819	0.073*	(0.000,5.3)	0.438	(0.547,0.8)	0.365	yes
Pervarw*Rarity, Numnets = 900.0 Predp = 0.8	0.127	(0.000,3.467)	1.711*	(0.547,1.009)	1.584	no
Pervarw*Rarity, Numnets = 73.0 Predp = 5.3	0.066†	(0.000,3.467)	1.002	(0.547,1.009)	0.936	no
@ Numnets*Predp, Pervarw = 0.547 Rarity = 1.009	1.002	(73.0,5.3)	1.711*	(900.0,0.8)	0.709	no
@ Numnets*Predp, Pervarw = 0.000 Rarity = 3.819	0.073*	(73.0,5.3)	0.234	(445.2,0.8)	0.160	yes
Numnets*Rarity, Pervarw = 0.547 Predp = 0.8	0.430	(73.0,3.467)	1.711*	(900.0,1.009)	1.281	no
Numnets*Rarity, Pervarw = 0.000 Predp = 5.3	0.066†	(73.0,3.467)	1.001	(569.2,1.009)	0.935	no
Predp*Rarity, Pervarw = 0.547 Numnets = 900.0	0.744	(5.3,3.467)	1.711*	(0.8,1.009)	0.967	no
Predp*Rarity, Pervarw = 0.000 Numnets = 73.0	0.066†	(5.3,3.467)	0.979	(0.8,1.009)	0.913	yes
@ Numnets, Pervarw = 0.547 Predp = 0.8 Rarity = 1.009	1.249	(73.0)	1.711*	(900.0)	0.462	no
@ Numnets, Pervarw = 0.000 Predp = 5.3 Rarity = 3.819	0.073*	(73.0)	0.190	(569.2)	0.117	yes
@ Rarity, Pervarw = 0.547 Numnets = 900.0 Predp = 0.8	0.893	(3.467)	1.711*	(1.009)	0.818	no
@ Rarity, Pervarw = 0.000 Numnets = 73.0 Predp = 5.3	0.066†	(3.467)	0.884	(1.009)	0.818	yes

CHAPTER 5

DISCUSSION

The Choice of 250 Populations and the Use of the Median

Consider the settings of the factors for an individual trial to represent the parameters of a superpopulation and the simulation realizations of the modified Neymann-Scott process to represent the populations. After a review of the results from the initial study and a reading of the literature, some differences are noted between our approach and some of the ones previously used.

The distributions for the relative efficiencies of both the Horvitz-Thompson and Hansen-Hurwitz estimators were skewed, in some instances highly so. As a consequence, the use of only a single population in a simulation, a practice seen throughout the literature on ACS, could generate misleading results. It was determined that a far better approach would be to calculate the average of the distribution of relative efficiencies to assess the true relative efficiency associated with a superpopulation. Although multiple populations were used several times in the literature, the number of populations was small with no justification for the number. Thus, the question remained as to how many populations must be generated to obtain an estimate of the true relative efficiency.

Several simulations in the literature used an approach in which the variance of an ACS estimator was estimated using hundreds or perhaps thousands of Monte Carlo samples, and the estimated variance was used to derive the relative efficiency. Many times, no justification is given for the number of samples. Our simulation program has the option of being able to obtain simple random samples, return the statistics $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$, and to further calculate estimated variances for these two estimators. Most importantly, however, our program actually calculates the true variances $Var[\hat{\mu}_{HT}]$ and $Var[\hat{\mu}_{HH}]$. Because of the capability to calculate true variances and because computer resources were not a limitation, we were not constrained by the number of populations we could generate and examine. It was apparent that the use of the mean as a measure of central tendency for the population distribution would be most unwise due to the aforementioned skewness. Consequently, it was decided that the median would be a better, more robust choice as the measure of central tendency. As a case in point, compare the mean (unavailable) and the median (5.448) from the numerical summary for the Extreme 1 setting RE_{HT} population distribution (Table 10). The use of the median is not without precedent. For example, Christman and Lan (1998) used median absolute deviations in their research due to skewness and outliers in the sampling distribution of an adaptive estimator they were studying after arriving at the conclusion that "...variances are not the best method for comparing efficiencies of the estimators.".

Figures 97 through 102 are M_a plots for RE_{HT} and RE_{HH} under the three different population settings for $N = 500$. Figures 103 through 117 are M_a plots for RE^*_{HT} and the four random covariates under the three different population settings for $N = 500$.

In general, an examination of the M_a plots from the initial study (Figures 97 through 117) show an initial period of fluctuation. However, for all practical purposes, by approximately the 100th population generated, the plots have stabilized. It was decided that 250 populations would represent a conservative, yet efficient choice. Several M_a plots were generated using the inverse transformed RE^*_{HT}. These plots verified that 250 populations also would be sufficient and no issues of concern were noted.

Simulation Approaches and Edge Effect

A review of the literature concerning similar simulation studies on ACS revealed different methodologies in regards to the edge effect mentioned in the Methods section. Christman (1996a,b, 1997) randomly generated locations of parents from a Poisson process in a fixed-size study region. If the locations of the offspring fell beyond the study region, they were reflected back into the region. Felix-Medina and Thompson (1999) also randomly generated locations of parents from a Poisson process in a fixed-size study area and used the approach whereby if a coordinate of an offspring's location was outside the study region, that coordinate was set to a value on the border of the region. In an attempt to get around the edge effect, Brown (1994,

Figure 97. Plot of M_a for RE_{HT}, Nominal Setting, $N = 500$.

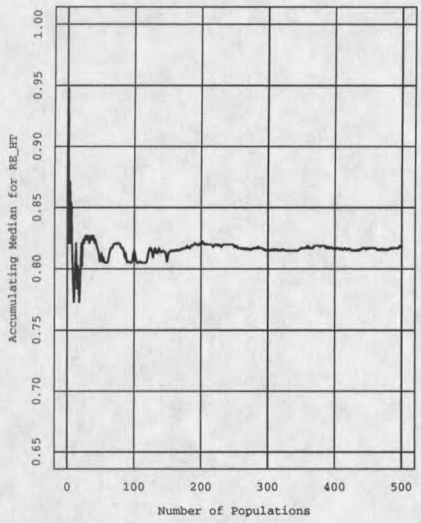


Figure 98. Plot of M_a for RE_{HH}, Nominal Setting, $N = 500$.

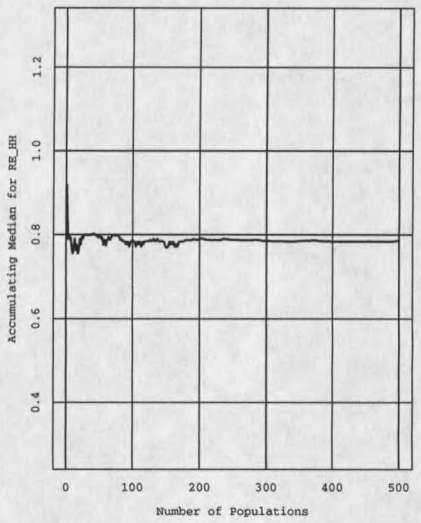


Figure 99. Plot of M_a for RE_{HT}, Extreme 1 Setting, $N = 500$.

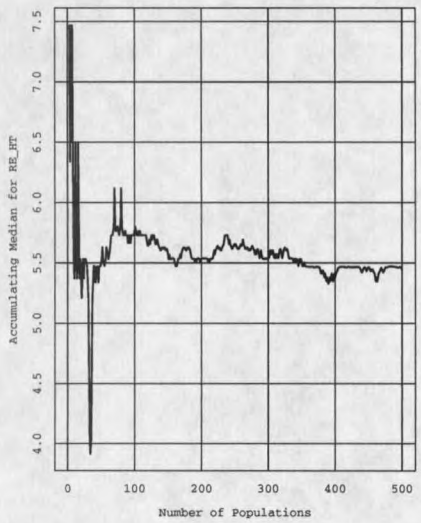


Figure 100. Plot of M_a for RE_{HH}, Extreme 1 Setting, $N = 500$.

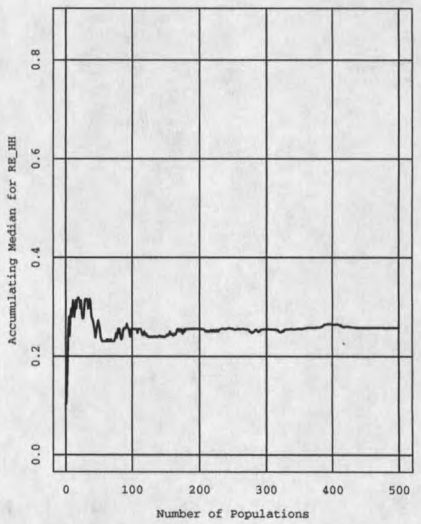


Figure 101. Plot of M_a for RE_{HT}, Extreme 2 Setting, $N = 500$.

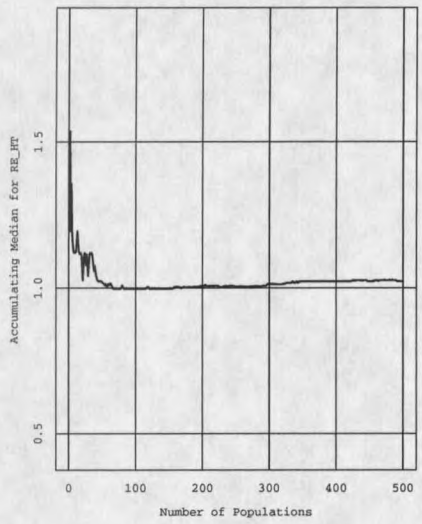


Figure 102. Plot of M_a for RE_{HH}, Extreme 2 Setting, $N = 500$.

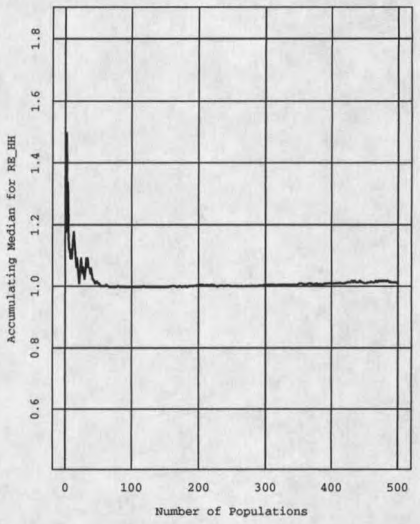


Figure 103. Plot of M_a for RE^*_{HT}, Nominal Setting, $N = 500$.

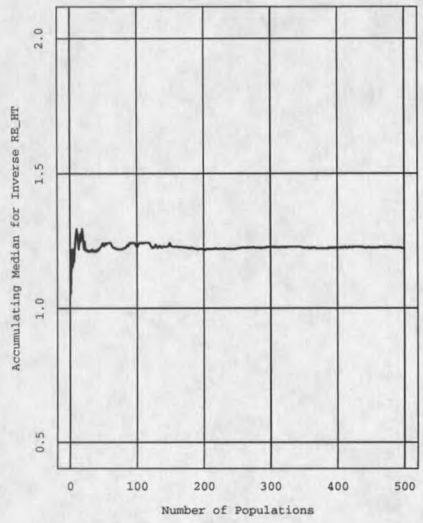


Figure 104. Plot of M_a for $\frac{\sigma^2_W}{\sigma^2}$, Nominal Setting, $N = 500$.

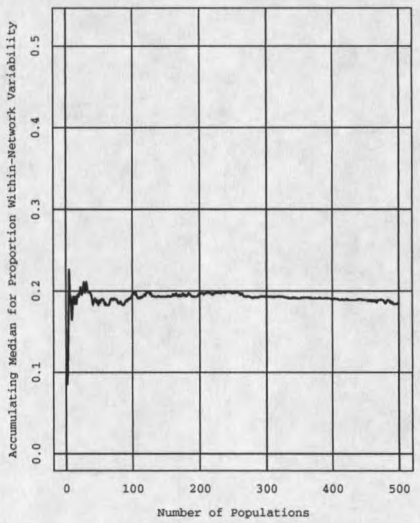


Figure 105. Plot of M_a for Numnets, Nominal Setting, $N = 500$.

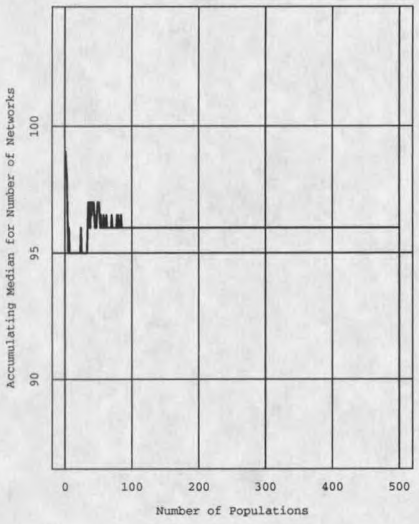


Figure 106. Plot of M_a for $\widehat{\lambda}_p$, Nominal Setting, $N = 500$.

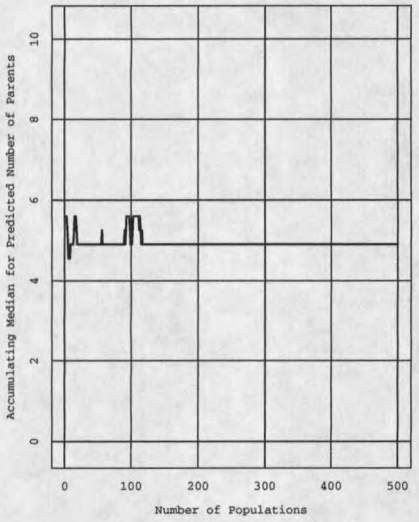


Figure 107. Plot of M_a for $\frac{E[\nu]}{n_1}$, Nominal Setting, $N = 500$.

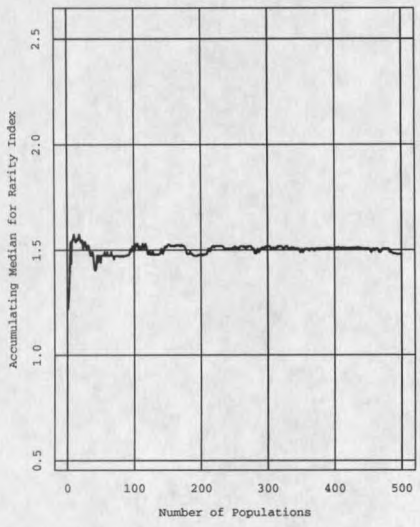


Figure 108. Plot of M_a for RE^*_{HT}, Extreme 1 Setting, $N = 500$.

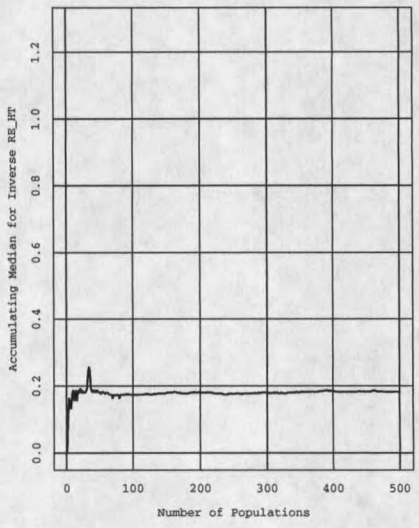


Figure 109. Plot of M_a for $\frac{\sigma_W^2}{\sigma^2}$, Extreme 1 Setting, $N = 500$.

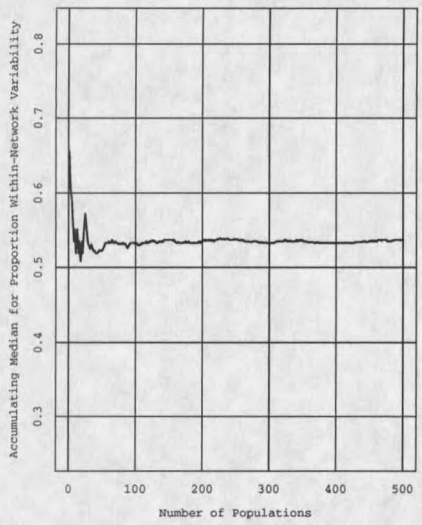


Figure 110. Plot of M_a for Numnets, Extreme 1 Setting, $N = 500$.

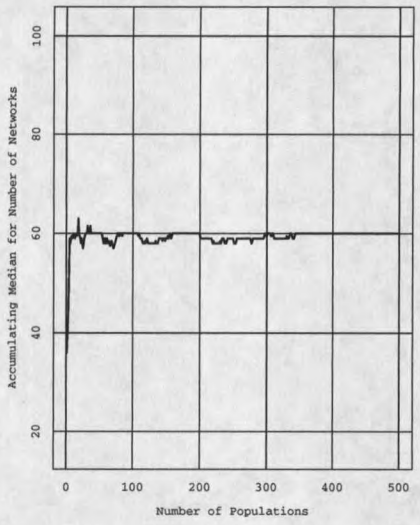


Figure 111. Plot of M_a for $\widehat{\lambda}_p$, Extreme 1 Setting, $N = 500$.

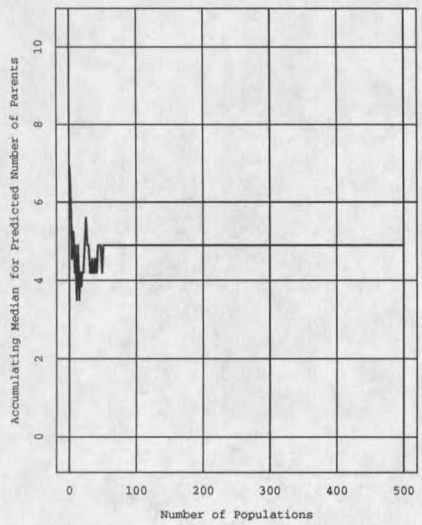


Figure 112. Plot of M_a for $\frac{E[\nu]}{n_1}$, Extreme 1 Setting, $N = 500$.

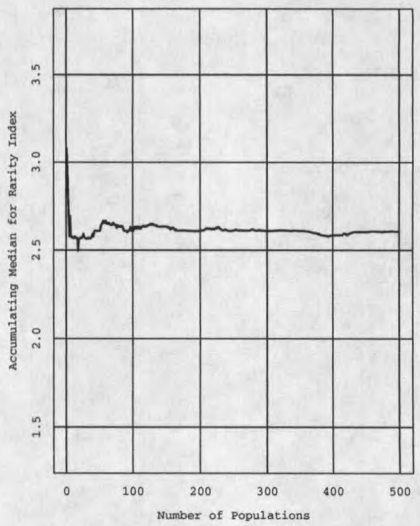


Figure 113. Plot of M_a for RE^*_{HT}, Extreme 2 Setting, $N = 500$.

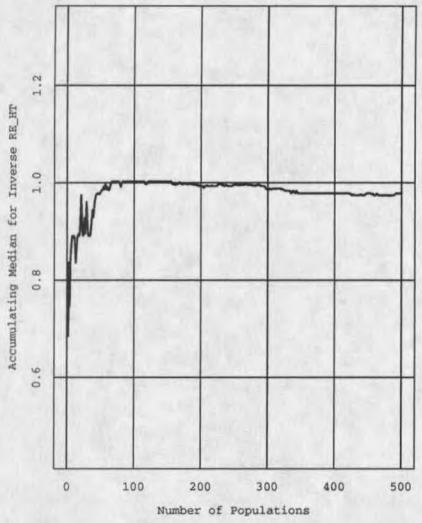


Figure 114. Plot of M_a for $\frac{\sigma^2_W}{\sigma^2}$, Extreme 2 Setting, $N = 500$.

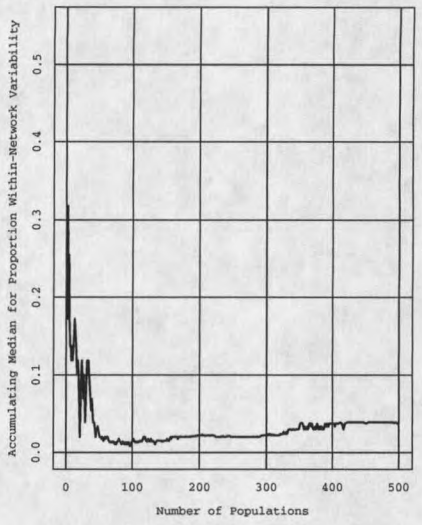


Figure 115. Plot of M_a for Numnets, Extreme 2 Setting, $N = 500$.

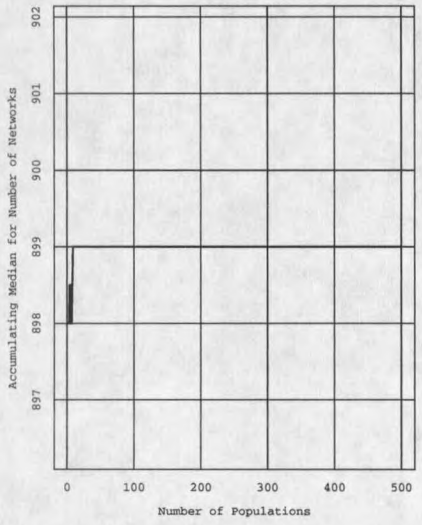


Figure 116. Plot of M_a for $\widehat{\lambda}_p$, Extreme 2 Setting, $N = 500$.

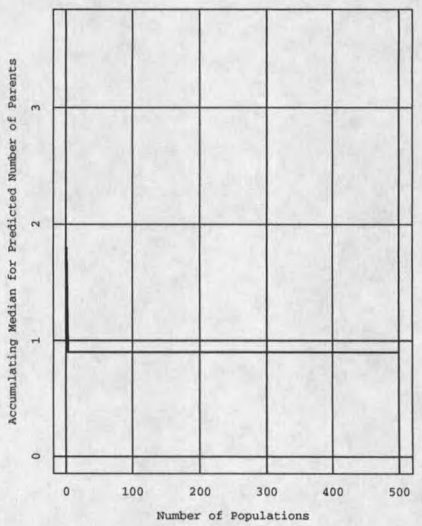


Figure 117. Plot of M_a for $\frac{E[\nu]}{n_1}$, Extreme 2 Setting, $N = 500$.

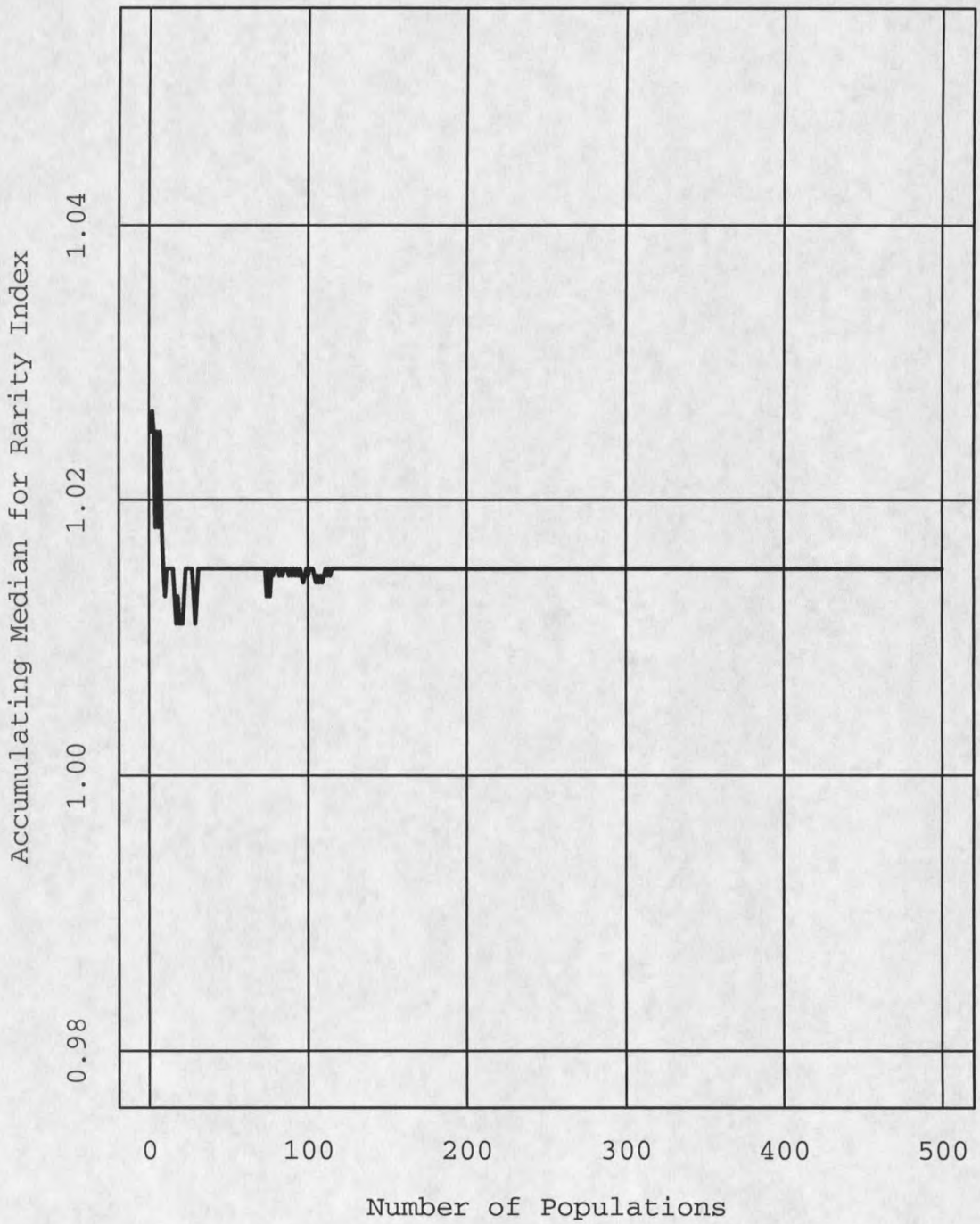
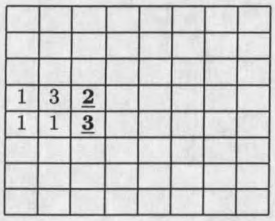


Figure 118. Study Population in Study Area B - Example 1.



1996); Brown and Manly (1998) used an approach whereby within their study site (B), a central study area (A) was defined after the populations were simulated within B. However, no justification is given in these papers for the determination of the size of region A within the larger fixed-size region B. Furthermore, there is no adjustment of λ_p (which is considered in the methodology described in this paper). Consequently, the actual mean number of clusters generally would be less than the mean that was claimed.

To better understand the consequences of these approaches, several examples will be studied. Figure 118 shows a fictitious study population in an 8 x 8 study area B. Suppose study area A is defined to be the innermost 4 x 4 grid of units. For clarity, those y-values within A have been underlined and boldfaced.

We now generate study populations in A under several scenarios (see Figure 119). Scenario one would result via the methodology described in this paper. Scenario two resulted from applying the methodology of Felix-Medina and Thompson (1999), while scenario three resulted from applying the methodology of Christman (1996a,b, 1997).

Figure 119. Resulting Study Populations in Study Area A - Example 1.

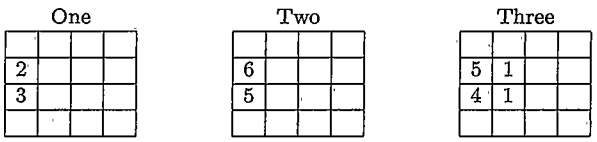


Table 28 shows the resulting parameters of the three study populations for example 1 where both c_a and $n_1 = 1$.

The results show a wide and surprising array of differences among the three populations with respect to their parameters, most importantly, relative efficiency. For example, population three had a much higher proportion of within-network variation, which from Chapter 2, is known to be a desired population characteristic for the implementation of ACS. We also know that in terms of implementing ACS, population three had the worst rarity index reading. Finally, note that population three yielded the worst relative efficiencies, the highest proportion of within-network variation not withstanding.

Figure 120 shows another fictitious study population in an 8 x 8 study area B. Again, let study area A be defined to be the innermost 4 x 4 grid of units. For clarity, those y-values within A have been underlined and boldfaced.

The resulting study populations in A for example 2 under the same several scenarios are shown in Figure 121. Table 29 shows the resulting parameters of the three study populations for example 2 where both c_a and $n_1 = 1$.

Table 28. Study population characteristics - example 1.

Study Population	One	Two	Three
τ	5	11	11
σ^2	0.763	3.563	2.363
σ_B^2	0.729	3.529	1.513
σ_W^2	0.033	0.033	0.850
$\frac{\sigma_W^2}{\sigma^2}$	0.044	0.009	0.360
Number of Networks	15	15	13
$E[\nu]$	1.625	1.625	3.250
$\frac{E[\nu]}{n_1}$	1.625	1.625	3.250
$Var[\hat{\tau}_{HH}]$	175.000	847.000	363.000
$Var[\hat{\tau}_{HT}]$	175.000	847.000	363.000
$Var[\hat{\tau}_{SRS}]$	107.923	504.231	148.292
RE_{HH}	0.617	0.595	0.409
RE_{HT}	0.617	0.595	0.409

Once again, wide differences exist among the populations with respect to the results. Differences can also be observed between the examples with population two now yielding the worst relative efficiencies. It is hard to draw broad conclusions based on two hypothetical examples. Nonetheless, the ramifications of using the simulation

Figure 120. Study Population in Study Area B - Example 2.

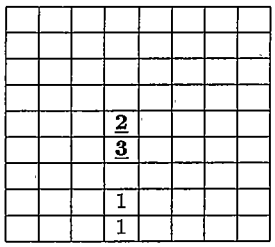
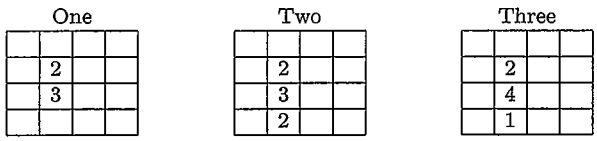


Figure 121. Resulting Study Populations in Study Area A - Example 2.



approaches that yielded populations two and three are not obvious and should be an area of future investigation. To further emphasize these points, consider the results of the study comparison of simulation results (Table 11). If we compare results for RE_{HT} between the two different studies, we would have reached two very different conclusions. For these settings (Table 2), the simulation approach from Christman (1996a) gives much more favorable results for RE_{HT} than does the approach described here. A serious concern regarding any future investigation and the validity of conclusions regarding the factors that are important in terms of relative efficiency is the dependence on the simulation approach used to generate the study area population.

An Initial and General Comparison of RE_{HT} Versus RE_{HH}

Table 29. Study population characteristics - example 2.

Study Population	One	Two	Three
τ	5	7	7
σ^2	0.763	0.929	1.196
σ_B^2	0.729	0.885	0.885
σ_W^2	0.033	0.044	0.311
$\frac{\sigma_W^2}{\sigma^2}$	0.044	0.048	0.260
Number of Networks	15	14	14
$E[\nu]$	1.875	2.688	2.688
$\frac{E[\nu]}{n_1}$	1.875	2.688	2.688
$Var[\hat{\tau}_{HH}]$	175.000	212.333	212.333
$Var[\hat{\tau}_{HT}]$	175.000	212.333	212.333
$Var[\hat{\tau}_{SRS}]$	91.907	73.642	94.777
RE_{HH}	0.525	0.347	0.446
RE_{HT}	0.525	0.347	0.446

We begin by examining the results from the initial study and comparing the distributions of RE_{HT} and RE_{HH} from the 500 populations generated for each of the three settings (see Table 10 and Figures 5 through 7). It is clear that any differences between the distributions depends on the parameters of the setting. The nominal

setting shows that both distributions are similar in terms of their relatively symmetric shape and the numerical summaries. The same conclusion is drawn for the extreme 2 setting except that both distributions are now highly skewed right. Given that this setting was constructed to emulate a rare and clustered population, it is no surprise that this setting gave the best overall results in terms of the percentage of relative efficiencies less than 1.00 (44.4 for RE_{HT} and 46.4 for RE_{HH}). For the extreme 1 setting, however, we see that a major difference has emerged. Both distributions are skewed right with the distribution for RE_{HT} exhibiting extreme skewness because it contains huge outliers. Christman (1996a) also noted severe skewness of the distribution for RE_{HT} among her least patchy populations. The distribution for RE_{HT} is shifted far to the right as compared to the distribution for RE_{HH}. Given this and the fact that an inverse transformation had to be done for RE_{HT}, suggests that the variability associated with RE_{HT} varies as a direct function of its size. From the numerical summaries we can see that in terms of relative efficiency performance, the Horvitz-Thompson estimator was only occasionally less than 1.00 (3.8 percent of the time) while the Hansen-Hurwitz estimator was almost always less than 1.00 (99.2 percent of the time). We are left with the overall impression that the Hansen-Hurwitz estimator will perform poorly relative to SRS unless the population is one that is rare and clustered. This result was also noted in the research of Christman (1996b).

The results for the 72 trials from this study (see Figure 11) indicate several issues which are also in agreement with the results from the initial study. First, the distribution for RE_{HT} is severely skewed right. Even after removal of the four outliers, the distribution for the edited RE_{HT} was still mildly skewed right. Upon inspection of the distribution for RE^*_{HT} , the transformation has pulled in the tails with the distribution now appearing roughly normal. The distribution for RE_{HH} was also roughly normal. Second, RE_{HT} is clearly much more variable than RE_{HH}. Approximate ranges were 2.5, 2.0 down to 1.3 for edited RE_{HT} , RE^*_{HT} and RE_{HH} respectively (see Table 12). Also, the results from the two global grid searchs (Tables 15 and 23) indicate the $SE[\hat{y}(\mathbf{x}_0)]$s and the range of predicted responses were much greater for RE^*_{HT} than RE_{HH}. The same conclusion is drawn for the grid search (Table 18) in the conditional FCA. This corroborates the assertion of Christman (1996a,b) that RE_{HT} is very sensitive to changes in the distribution of elements in the population with even minor changes leading to different relative efficiencies. Despite the relative robustness of Hansen-Hurwitz estimation, one is unfortunately confronted with a tradeoff due to our final observation. The Horvitz-Thompson estimator was more relatively efficient than the Hansen-Hurwitz estimator (see Figure 13), a result that has been seen many times in the published research on ACS ((Brown 1994; Christman 1996b)). Notice in almost every trial that $RE^*_{HT} < 1/RE_{HH}$, or equivalently we have the two inequalities, $RE_{HH} < RE_{HT}$ and $Var[\hat{\tau}_{HT}] < Var[\hat{\tau}_{HH}]$. In fact, there were only four trials where RE_{HH} was even equal to RE_{HT} (see Table 8). Note

that similar results also were seen in the results from the simulation to replicate the study of Christman (1996a) (Table 11). The median RE_{HT} value for the 72 trials was 1.081 indicating more often than not a trial result favorable for ACS. 68 % of the RE_{HT} values were greater than or equal to 1.000. On the other hand, the median RE_{HH} value of 0.987 indicated a majority of trial results were not favorable for ACS with 43.1 % of the RE_{HH} values being greater than or equal to 1.000. An examination of Figure 13 using the coordinate (1.0,1.0) as an origin shows that the 1st quadrant contains no plotted points indicating that no trial results were favorable for Hansen-Hurwitz estimation but not favorable for Horvitz-Thompson estimation! On the other hand, there are many points residing in the 3rd quadrant which means that there are conditions that favor Horvitz-Thompson estimation but are not favorable for Hansen-Hurwitz estimation.

Comparing the RE^*_{HT} and RE_{HH} responses with respect to the FCA and RCA models, one obvious result is that some of the significant effects were unique to the response. For example, in the FCA, the quadratic Area effect is significant for RE_{HH} (Table 14) but not for RE^*_{HT} (Table 13). Also, in the RCA, the quadratic Pervarw effect is significant for RE^*_{HT} (Table 21) but not for RE_{HH} (Table 22). Both of the models in both analyses had large coefficients of determination. Nonetheless, in terms of the strength of the linear association, we see that

(i.) for the FCA, RMSE = 0.1409 and $R^2 = 0.85$ for RE^*_{HT} and RMSE = 0.0622 and $R^2 = 0.95$ for RE_{HH}, and

(ii.) for the RCA, RMSE = 0.1472 and $R^2 = 0.79$ for RE^*_{HT} and RMSE = 0.0663 and $R^2 = 0.94$ for RE_{HH}.

Consequently, the model for RE_{HH} appears to be a better, or less noisier, one in both instances. This result is not surprising in light of the previous results and discussion concerning variability.

The Variability of RE_{HT}

Christman (1996b) stated that it was not easy to describe conditions under which Horvitz-Thompson estimation is more efficient than SRS. One reason is the excessive variability associated with Horvitz-Thompson estimation. We will now focus our attention on the reasons why RE_{HT} has greater variability than RE_{HH}. Christman (1996a) compared efficiencies of ACS in different populations for Horvitz-Thompson estimation. Both the set up of her research and the simulation approach used, however, were different than those used in our study. Furthermore, we have already discussed the potential differences between results that can occur by using different simulation approaches. Nonetheless, the author noted that for the least patchy populations, as the number of clusters increased, they tended to coalesce into a few large networks. Consequently, if a unit from any of these networks was sampled the result was that a very large proportion of the population appeared in the final sample. Smith

et al. (1995) also noticed, albeit tacitly, the same phenomenon occurring in their research, specifically for the ring-necked duck data set utilizing the largest-sized unit. A similar scenario was detected in our study and will be discussed via an example.

Recall that the highest RE_{HT} value of 39.196 occurred on trial 50 (Table 7). The factor levels are presented again in Table 30.

Table 30. Trial 50 settings used in the FCA.

Trial	Shape	Direct	Spread	Parents	Area	Critical	Kids	Sample
50	4	0.0	0.5	3	10	1	50	50

Upon inspection of the trial 50 data, it was noticed that there were several extraordinarily high results for RE_{HT}, i.e. several extraordinarily low results for $Var[\hat{\tau}_{HT}]$. A physical inspection of the 250 populations revealed the underlying reason for these results. For example, Figure 122 contains the 14th population and Table 31 contains the parameter values for trial 50 resulting from the R simulation program.

What occurred for population 14 was representative of what occurred for the other populations in this trial that yielded unusually high values for RE_{HT}. By virtue of the trial settings, most notably many offspring in a small study area, several networks coalesced into a single large network, or a *hypernetwork*.

Figure 122. The 14th Population for Trial 50 Displaying Only Units Containing One Or More Elements.

						1	6		
				1	1	4	9	5	
				1	5	4	10	5	
			1	1	11	5	5	3	
				1	5	3		1	
			1	3	1				
			5	7					
			4	5					
			5	4					
			1	2					

As mentioned earlier, if a unit from a hypernetwork is sampled the result will be that a very large proportion of the population will appear in the final sample. Consequently, one could anticipate a reduction in the variance of the Horvitz-Thompson estimator. However, a large final sample size does not, in and of itself, guarantee a low value for $Var[\hat{\tau}_{HT}]$. We regressed the 250 values of $Var[\hat{\tau}_{HT}]$ from the trial 50 populations on their respective values of $E[\nu]$ and found no evidence of either a significant linear association ($r^2 = 0.01$, $\mathrm{P}[F > 2.53] \approx .11$) or a significant quadratic association ($r^2 = 0.01$, $\mathrm{P}[F > 1.28] \approx .27$). Christman (1996a) had suggested that

Table 31. Trial 50 results for the 14th population.

τ	126
σ^2	5.649
σ_B^2	3.256
σ_W^2	2.393
$\frac{\sigma_W^2}{\sigma^2}$	0.424
Number of Networks	68
$E[\nu]$	78.000
$\frac{E[\nu]}{n_1}$	1.560
$Var[\hat{\tau}_{HH}]$	325.587
$Var[\hat{\tau}_{HT}]$	5.305396e-10
$Var[\hat{\tau}_{SRS}]$	159.328
RE_{HH}	0.489
RE_{HT}	300312420885

the high RE_{HT} values were due to high within-network variability. Thus, we also regressed the 250 values of $Var[\hat{\tau}_{HT}]$ from the trial 50 populations on their respective Pervarw (proportion of within-network variability) values and found significant evidence of only a weak linear association ($r^2 = 0.03$, $P[F > 7.92] \approx .01$) and a weaker quadratic association ($r^2 = 0.03$, $P[F > 3.97] \approx .02$). We now offer a simple

explanation for why these high RE_{HT} values occur by working through the derivation of $Var[\hat{\tau}_{HT}]$ for population 14. Although one might think that derivation of these results for such a large population would be a daunting task, the mathematics can be greatly simplified. First, note that the 68 networks can be divided into 67 trivial networks and the one hypernetwork. Also note that the trivial network sums (the y^*'s) will all, of course, be 0, while the hypernetwork sum will be $\tau = 126$. Consequently, if we expand equation 1.9 in order to derive $Var[\hat{\tau}_{HT}]$ (or $N^2 Var[\hat{\mu}_{HT}]$), extensive cancellation will occur yielding a single term relevant only for the hypernetwork:

$$Var[\hat{\tau}_{HT}] = y^{*2}\left(\frac{1-\alpha}{\alpha}\right) \tag{5.1}$$

The marginal initial intersection probability for the hypernetwork (α), will be extremely high. To be exact, application of equation 1.6 gives a value of $\alpha = 1$ - (3.3416e-14) ≈ 1 which implies a virtual certainity that the hypernetwork will be intersected by the initial sample! Thus, the last term of the right side of equation 5.1 will be extremely small and, consequently, $Var[\hat{\tau}_{HT}]$ itself will be extremely small. This issue should not be surprising in light of the sentence in the Introduction shortly after equation 1.10. $\hat{\mu}_{HT}$, and thus $\hat{\tau}_{HT}$, will have low variance if $\alpha_k \propto y_k^*$. Because the marginal initial intersection probabilities for any trivial network can easily be shown to be equal to .5 via equation 1.6, we clearly have a proportional relationship between the y^*'s and the α's. Again, by construction of the Horvitz-Thompson estimator (equation 1.5), $\hat{\tau}_{HT}$ will be relatively constant with little variability.

Lest one think the high RE_{HT} for population 14 of trial 50 does not come without a "heavy price", notice also the result for $E[\nu]$. On average, one could expect to obtain a final ACS effective sample size of 78 out of the 100 units in the population which would be unrealistic and prohibitive for many surveys.

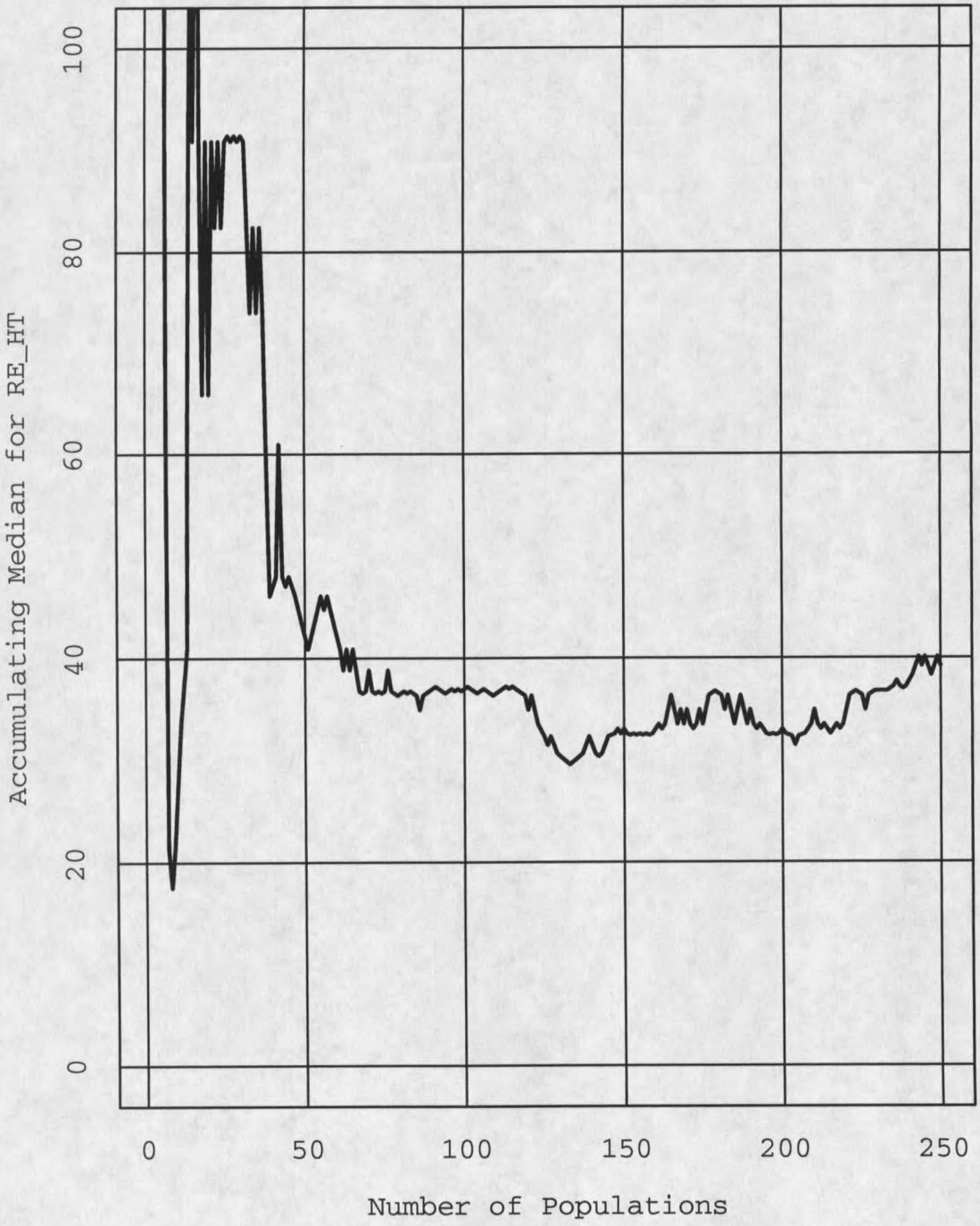
Figure 123 is an M_a plot for RE_{HT} for the 250 populations associated with trial 50. It is comforting to know that, despite the initial wild fluctuations, there is an eventual relative stabilization of the accumulating median shortly after the 50th population.

The phenomenon of the hypernetwork and its impact is also seen in the three populations from the initial study (Figures 2 through 4). The relevant portion of the results from the initial study (Table 9) have been reproduced in Table 32.

Table 32. Initial study example population characteristics.

Study Population	Nominal	Extreme 1
$E[\nu]$	14.133	92.145
$Var[\hat{\tau}_{HH}]$	53719.758	2529.044
$Var[\hat{\tau}_{HT}]$	52213.595	102.318
$Var[\hat{\tau}_{SRS}]$	45325.795	271.354
RE_{HH}	0.844	0.107
RE_{HT}	0.868	2.652

Figure 123. Trial 50 Results.



For the Extreme 1 example population, and a relatively large initial sample size of 30, the resulting expected effective final sample size of 92.145 out of 100 total units in the study area is impractically large and close to being a census. Although $Var[\hat{\tau}_{HH}]$ is reduced sharply as compared to the nominal example population, $Var[\hat{\tau}_{HT}]$ experiences a far greater relative reduction. Consequently, RE_{HH} is a terrible 0.107 while RE_{HT} is a misleadingly optimistic 2.652.

Brown (1996) examined factors that influenced RE_{HT} and presented results that seem to contradict ours. Most specifically, she claimed that RE_{HT} variation would be greatest for the most patchy populations. The results from our initial study (Table 10) suggest the opposite conclusion. We found between RE_{HT} variability to be much greater for the Extreme 1, or the least patchy, type of population due to the aforementioned hypernetworks. In addition to both studies being set up differently, there are several issues that bear mentioning that could explain this discrepancy. For example, Brown estimated the variance of $\hat{\tau}_{HT}$ with 200 samples; we calculate the variance directly. Only 20 populations were used in her study; 500 populations were used in ours. Finally, whereas we have demonstrated severe non-normality for the distribution of RE_{HT} under the Extreme 1 setting, Brown's use of the mean of the 20 RE_{HT} values and also the standard error, could be highly misleading. Even if one could assume normality, given that the variance is a function of the mean, the coefficient of variation might offer a better comparison, in which case we note

the coefficients of variation for her most patchy and least patchy populations were virtually identical.

In conclusion, many studies in the literature seem to embrace the use of the Horvitz-Thompson estimator based only on relative efficiency results. However, as the previous simple example has demonstrated, Horvitz-Thompson estimation may emphasizes variance reduction and disregard the resultant increase in the final ACS sample size. Consequently, even though one may obtain a large RE_{HT} value due to variance reduction, Horvitz-Thompson estimation may actually not be an "efficient" design if one also considers the cost of the final ACS sample size. Finally, for these examples, notice that using Hansen-Hurwitz estimation will produce the same large final ACS sample size but will yield only a relatively small reduction in variance of the ACS estimator.

A General Overview of the Grid Searches

We now examine and discuss the results from Table 15 for the global grid search from the FCA, a portion of which is reproduced in Table 33.

When comparing the optimum (minimum) RE^{*}_{HT} setting versus the worst (maximum) setting, the Direct, Spread, Area, and Kids variables remained the same while the Shape, Parents, Critical, and Sample variables switched settings from one extreme level to the other extreme level. When comparing the optimum (maximum) RE_{HH} setting versus the worst (minimum) setting, the Spread, Critical, and Kids variables

Table 33. FCA global grid search results for maximum and minimum predicted RE^*_{HT} and RE_{HH}. A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. na = not applicable.

Factor	Minimum RE^*_{HT}	Maximum RE^*_{HT}	Factor	Minimum RE_{HH}	Maximum RE_{HH}
Shape*	1	4	Shape*	4	1
Direct	.7	.7	Direct	na	na
Spread	.5	.5	Spread	.5	.5
Parents*	1	5	Parents*	5	1
Area	10	10	Area*	10	30
Critical*	1	2	Critical	1	1
Kids	50	50	Kids	50	50
Sample*	50	10	Sample*	10	50

remained the same while the Shape, Parents, Area, and Sample variables switched settings from one extreme level to the other extreme level. If we compare the optimum settings (excluding Direct which was not a term in the model for RE_{HH}), the only difference between the two response surface models was for Area, where the setting was 10 for RE^*_{HT} and 30 for RE_{HH}. When comparing the worst settings (again with the exclusion of Direct), the only difference between the two responses was for Critical, where the setting was 2 for RE^*_{HT} and 1 for RE_{HH}.

We now turn our attention towards the more stringent conditional grid search (Table 18), a portion of which is reproduced in Table 34.

Recall that in the conditional FCA, the fixed covariates are fixed to reasonable and more realistic (less extreme) levels. For example, a researcher may have some prior knowledge about the characteristics of the study population. This knowledge

Table 34. Conditional grid search results of the FCA with fixed covariate levels for both the maximum and minimum predicted RE^*_{HT} and RE_{HH}. A controllable factor marked with an asterisk indicates a change occurred between minimum and maximum level settings for the predicted response. na = not applicable.

Factor	Minimum RE^*_{HT}	Maximum RE^*_{HT}	Factor	Minimum RE_{HH}	Maximum RE_{HH}
Shape	4	4	Shape	4	4
Direct	0	0	Direct	na	na
Spread	.1	.1	Spread	.1	.1
Parents	3	3	Parents	3	3
Area*	30	10	Area*	10	30
Critical*	1	2	Critical	1	1
Kids	30	30	Kids	30	30
Sample	10	10	Sample*	10	50

could be used in a conditional grid search to help the researcher select the optimal settings for the controllable factors to be used for sampling.

For the three controllable factors (Area, Critical and Sample), and for RE^*_{HT}, the optimum and worst settings for both Area and Critical were opposite extreme levels. For RE_{HH}, the Critical setting remained unchanged while the optimum and worst Area and Sample settings were opposite extreme levels. If we compare the optimum settings between responses, Sample was equal to 10 for RE^*_{HT} but was equal to 50 for RE_{HH}. Examining the worst settings between responses, Sample now remained constant while Critical equaled 2 for RE^*_{HT} and 1 for RE_{HH}.

There is an important conclusion to be drawn here. Although the maximum RE_{HH}, minimum RE_{HH}, and maximum RE^*_{HT} factor settings are the same for the global and conditional FCA grid searches, this was not the case for minimum RE^*_{HT}. For RE^*_{HT}, the Area settings were 10 and 30 and the Sample settings were 50 and 10

for the global and conditional grid search results, respectively. Thus, it appears that while the results for RE_{HH} from the global grid search might be robust, the results for RE^*_{HT} are not, and optimization of RE^*_{HT} is dependent on the settings of the remaining fixed covariates. This conjecture is consistent with the work of Christman (1996b) in that the conditions that optimize RE_{HT} aren't as easily described as they are with RE_{HH}. The more detailed interaction analysis at specific population settings discussed in the next section will verify that RE_{HT} is indeed a slippery entity to characterize.

We now examine the results for the global grid search from the RCA (Table 23), a portion of which is reproduced in Table 35.

Table 35. The RCA global grid search results for maximum and minimum predicted RE^*_{HT} and RE_{HH}. A factor marked with an asterisk indicates a change took place between minimum and maximum level settings for the predicted response. na = not applicable.

Factor	Minimum RE^*_{HT}	Maximum RE^*_{HT}	Factor	Minimum RE_{HH}	Maximum RE_{HH}
Pervarw*	0.547	0.000	Pervarw*	0.000	0.547
Numnets	73.0	73.0	Numnets*	73.0	900.0
Predp	na	na	Predp*	5.3	0.8
Rarity*	1.009	4.521	Rarity*	3.819	1.009

First, a comparison of the optimum versus worst case settings for RE^*_{HT}, indicates the setting for Numnets stayed the same (73.0) while settings for the other two variables, Pervarw and Rarity, switched from one extreme level to the other. For

RE_{HH}, all four variables switched settings from one extreme level to the other extreme level while Rarity was also involved in a quadratic effect. Thus, although the optimum setting was 1.009, the worst case Rarity setting of 3.819 was slightly less than the largest value of 4.521. When comparing RE_{HH} and RE^*_{HT}, we note two differences. First, Numnets was a variable that switched settings for RE_{HH}, but not for RE^*_{HT}. Second, the variable Predp was a significant explanatory variable for the RE_{HH} model, but not the RE^*_{HT} model. As a result of these two points, differences occur when comparing the optimum and worst case settings between responses. The optimum number of networks setting for RE^*_{HT} is 73.0 versus 900.0 for RE_{HH}. Additionally, the optimum predicted number of parents, applicable of course only to RE_{HH}, is 0.8, while the worst predicted number of parents setting is 5.3. Finally, the worst Rarity setting for RE^*_{HT} is 4.521 versus the aforementioned 3.819 for RE_{HH}.

Interaction Analysis at Specific Population Settings

Examining the response surfaces for a pair of design variables while setting all other variables to fixed levels offers a good opportunity to examine the issue of optimality and to compare Horvitz-Thompson and Hansen-Hurwitz estimation. We have such opportunities here with both the FCA and RCA. Although the emphasis in this paper has and will be on those conditions that optimize relative efficiency, the intent is not to discount those conditions that yield the worst relative efficiency. It would be insightful to examine the response surfaces for those conditions that indicate the

worst performance to underscore the importance of the interaction and the complexity of their interpretation. This is left to the reader for the sake of brevity. We will also discover that the interpretation of the complex interaction structure can be very difficult. This is partially due to the number of variables in the models. It is also due to those instances where the interaction for a specific population displays antagonistic behavior, i.e. the direction of the conditional linear effects (first-order) are reversed for a given factor.

We will now proceed with an in-depth examination of these analyses. To facilitate discussion, the variables will be discussed either individually or in the context of some logical grouping. With the exception of Rarity, all the variables were involved in at least one significant interaction effect. Consequently, the linear effects for the variables should not be interpreted directly (i.e. ignoring relevant interactions) as it unfortunately is done all too often in the literature. For each variable, the significant interaction effects will be discussed in descending order of the absolute magnitude of their t values.

Controllable Factors - Sample Size, Critical Value and Area

Sample Size

We begin by considering initial sample size in the FCA. For both RE_{HH} and RE^{*}_{HT}, the optimum and worst settings from the centerpoint FCA (Table 16 and 17) matched the analagous settings from the global grid search (Table 15), i.e. $n_1 = 50$

and 10 respectively. The ranges from the centerpoint FCA were larger for RE^*_{HT} than RE_{HH} for the Spread*Sample and Kids*Sample interactions common to both models and for the situation from the conditional FCA where Area and Sample were allowed to vary and Critical was set to 1 (Tables 19 and 20). The optimum setting for RE^*_{HT} shifted to Sample = 10 for two population settings examined under the conditional FCA while both the optimum and the worst settings remained the same for RE_{HH} for one of these populations.

The significant Area*Sample interaction was unique to RE^*_{HT}. The centerpoint FCA showed that RE^*_{HT} improved with increasing n_1 with the linear effect being very strong for the smallest Area value (10) (Figure 23). Given the smallest Area value (10), minimum RE^*_{HT} was obtained for the largest Sample value (50), while all Sample settings were favorable for ACS for the largest Area value (30). However, the results of the conditional FCA were not consistent with the previous results. For our set of levels for the fixed covariates and across the levels of C, the direction of the linear effect of Sample actually reversed for the largest Area (Figures 39 through 42) although the absolute magnitude of the linear effect was still roughly the same. Obviously, this type of antagonistic behavior makes drawing conclusions more difficult. This result is also seen by an examination of the plots for Critical*Sample (not significant) across the range of settings for Area (Figures 43 through 46). From the optimization standpoint, no changes were noted. There was no significant Area*Sample interaction effect for RE_{HH} (Figures 51 through 52). Consequently, an examination of the response

surface for the population under the conditional FCA where $c_a = 1$ will not reveal any differences between the conditional linear effects with respect to direction and magnitude. Nonetheless, it is worthy to note the magnitude is very small relative to the effect of Area, that RE_{HH} performance increased with increasing initial sample size and that no conditions were favorable for ACS when Area = 10 while the opposite was true when Area = 30.

The models for both responses had a significant Spread*Sample interaction effect. For the centerpoint FCA, both RE^*_{HT} and RE_{HH} improved with increasing Sample with the linear effect being more pronounced at the largest value for Spread (Figures 19 and 30). The magnitude of the linear effects at both the highest (0.500) and lowest (0.100) setting for Spread was much greater for RE^*_{HT}. In terms of selecting optimal settings, both response surfaces showed all conditions as being favorable for ACS given the smallest Spread. When Spread was set to its highest setting, both responses were optimized at the largest Sample value. Christman (1996b) acknowledged the existence of a significant Spread*Sample interaction effect for RE_{HT} with its interpretation being in solid agreement with ours.

The models for both responses had a significant Kids*Sample interaction effect. Figures 24 and 35 from the centerpoint FCA show improvements in both RE^*_{HT} and RE_{HH} as initial sample size increases with the linear effect being more pronounced for a large value of Kids (50). The magnitude of the linear effects at both the highest

and lowest (10) setting for Kids was much greater for RE^*_{HT}. In terms of selecting optimal settings, the responses differed. When Kids = 10, no conditions were favorable for ACS for RE_{HH} while RE^*_{HT} had an optimum at the largest value for Sample. When Kids = 50, all conditions were now favorable for ACS for RE_{HH} while RE^*_{HT} once again was optimized at the largest Sample value. Christman (1996b) also acknowledged the existence of a significant Kids*Sample interaction effect for RE_{HT} with its interpretation being in rough agreement with ours.

Lastly, a significant Parents*Sample interaction effect was unique for RE^*_{HT}. From the centerpoint FCA, there was an improvement in RE^*_{HT} with increasing initial sample size and this linear effect was more pronounced for a large value of Parents (5) (Figure 22). All the conditions were favorable for ACS when Parents was set to its lowest value (1). When Parents was set to its highest value (5), an optimum was obtained for the largest value of Sample. Brown (1996) has detected a similarly interpreted significant Parents*Sample interaction effect in her research. Christman (1996b) acknowledged the existence of a significant Parents*Sample interaction effect with its interpretation being in rough agreement with ours.

Finally, from the conditional FCA, an examination of the plots for Area*Critical (Figures 47 through 50) across the range of settings for Sample, Area*Sample (not significant) where $c_a = 1$ (Figures 51 through 52), and Critical*Sample (not significant) across the range of settings for Area (Figures 53 and 54), all support the conclusion that RE_{HH} mildly improved with increasing initial sample size.

We can draw several conclusions from all of these analyses. The Sample variable appears to be quite important for the RE^*_{HT} response and not nearly so for RE_{HH}. In fact, judging by the magnitude of the t values and the number of significant interactions, a good case can be made that initial sample size is the most important variable in determining RE^*_{HT}. The magnitudes of the comparable linear effects were in all cases greater for RE^*_{HT}. In general, for both responses, large initial sample sizes would be preferred when using ACS, maybe more so for RE^*_{HT}. However, for RE^*_{HT}, this statement is not a generalization for all conditions since we have constructed a population where for large study areas a small initial sample size would be more desirable. Undoubtedly, the impact of hypernetworks is at play here, because the combination of large initial sample sizes in small study areas can so dramatically improve the performance of RE^*_{HT}. Consequently, the results of the global grid search and the centerpoint FCA could be merely a hypernetwork aberration while the results of the conditional FCA might be more realistic and practical. Christman (1996b) also found that the Horvitz-Thompson estimator was more efficient than the SRS sample mean for less clustered types of populations when the initial sample size was large. However, as we pointed out previously, this gain in efficiency often came at the cost of very large effective sample sizes. Christman (1996b) also found that RE_{HH} values tended to increase with increasing initial sample size.

Critical Value

Critical was involved in significant interaction effects for both responses. An examination of the t values and interaction effects (Tables 13 and 14) indicates that the critical value would not be considered a major variable in explaining relative efficiency, at least for the two levels considered here (1 and 2). From the FCA, there were no significant interaction effects held in common by the two models. From the centerpoint FCA, the setting $c_a = 1$ corresponded to both the optimum and worst RE_{HH} while the correspondence for RE^*_{HT} was $c_a = 1$ for the optimum and $c_a = 2$ for the worst case. For both responses, these settings matched those from the global grid search. Furthermore, these settings also appeared in the conditional FCA for two population settings when studying the Area*Critical term. From this analysis, when Sample was set to 10 (Table 19 and 20), we also note the range of predicted values for RE^*_{HT} was much greater than that for RE_{HH}.

The Area*Critical interaction effect was significant only in the case of RE_{HH}. For the centerpoint FCA, the slight curvature in the response surface is due to a significant quadratic effect for Area and is practically meaningless (Figure 33). There was a slight improvement in RE_{HH} for increasing values of c_a at the smallest value for Area (10), while a larger linear effect of an opposite nature was seen at the largest value for Area (30). In the Critical Value section of Chapter Two, Brown (1996) noted the same antagonistic behavior indicated by the directions of the conditional effects for RE_{HT} when she looked at critical values and population densities. Because

density is inversely proportional to area, this result is not surprising. As we shall soon see, the Area linear effect was quite strong for the RE_{HH} model. Thus, there were no conditions favorable for ACS when Area = 10 while all conditions were favorable given Area = 30. The response surfaces for two different scenarios from the conditional FCA yielded the same conclusions as with the centerpoint FCA, regardless of sample size (10 or 50) (Figures 47 through 50). The aforementioned conditional shift in the direction of the Critical linear effect was also detected in the plots from the conditional FCA for Critical*Sample (not significant) across the range of levels for Area (Figures 53 and 54). There was no significant Area*Critical interaction effect for RE^*_{HT}. Consequently, an examination of the response surface for the population under the conditional FCA where $n_1 = 10$ will not reveal any differences between the conditional linear effects with respect to direction and magnitude (Figure 37). Nonetheless, it is worthy to note the magnitude is small relative to the effect of Area, that RE^*_{HT} performance increased with decreasing values of c_a, and that no conditions were favorable for ACS when Area = 10 while the opposite was true when Area = 30.

The Spread*Critical interaction effect was significant for the RE_{HH} response only (Figure 28). An improvement in RE_{HH} was seen with decreasing values of c_a given the smallest level of Spread ($\sigma_x = .1$) while a very slight improvement was seen with increasing values of c_a given the largest level of Spread ($\sigma_x = .5$). Like the Area*Critical interaction, the linear effect of Critical is simply dominated, but this

time by the linear effect of Spread. Given Spread = .1, all the conditions were favorable for ACS while no conditions were favorable when Spread = .5. Christman (1996b) noted a similarly interpreted interaction between Pervarw and Critical which is noteworthy since the proportion of within-network variability is affected by the spread of offspring within a cluster.

For the RE^*_{HT} response, the Shape*Critical interaction and the Direct*Critical interaction were significant effects (Figures 15 and 16). For both interactions, there was an improvement in performance as the critical value decreased, particularly much more so when the cluster was cigar-shaped (Shape = 4) and tilted (Direct = 0.7). All the conditions were favorable for ACS for both interactions.

Choosing a value of c_a entails a tradeoff as alluded to in the section "Critical Value" of Chapter 2. If c_a is set low, although the Pervarw value would increase, so, unfortunately, would $E[\nu]$. On the other hand, if c_a is set high, the Pervarw value would decrease but so would $E[\nu]$. Because the literature suggests that a high Pervarw value is desired to optimize relative efficiency (and this study will corroborate this assertion), the tradeoff is, therefore, between relative efficiency and sample size and costs. This tradeoff has and will be a recurring theme in this paper.

We argue for a low critical value for several reasons. The two-level variable Critical appears to be a minor issue in the grand scheme of this study. We argue further that a review of the literature and the results of this paper will demonstrate the importance of a high value of Pervarw when optimizing relative efficiency, the

increase in expected effective sample size not withstanding. Examining our evidence, the performance of RE_{HH} was more dependent than RE^*_{HT} on the level settings for the other factors. An improvement in performance with decreasing values of c_a was established only for large-sized study areas and tight clusters which is a condition we will see is already conducive to ACS for RE_{HH}. The results of Christman (1996b) showed the best performance of Hansen-Hurwitz estimation occurred at the lowest setting of c_a, with this result being attributed to the higher proportion of within-network variability. The general conclusion drawn for RE^*_{HT} is that a mild case can be made that its performance is best at $c_a = 1$, as opposed to $c_a = 2$, which seems to support the current thinking in the literature ((Christman 1996b, 1997, 2000)). Additional supporting evidence for this assertion is seen in

(i.) the plots of Area*Critical (not significant) where Sample = 10 (Figures 37 and 38),

(ii.) the plots of Area*Sample across the range of settings for Critical (Figures 39 and 42),

(iii.) and the plots of Critical*Sample (not significant) across the range of settings for Area (Figures 43 and 46).

Although the optimum Critical setting for RE^*_{HT} might be considered an obvious choice, for RE_{HH}, the importance of Critical appears to be even less for RE^*_{HT} than for RE_{HH}.

Brown (1994) made the assertion that initial sample size and the size of the critical value were "interrelated". We note that no significant Sample*Critical interaction effect was detected for either response.

Finally, recall that Brown (1996) had showed the optimal choice of c_a, at least for the case of RE_{HT}, was directly related to the population density thus leaving the reader with the impression that a high, rather than a low, c_a value should be used in a population with high density. In this study, we did not find a significant Parents*Critical, Kids*Critical, or Area*Critical interaction effect for RE^*_{HT}. Furthermore, one might question the wisdom of using ACS in a population with high density (a common as opposed to rare population) in the first place.

Area

Area was involved in significant interaction effects for both responses. Furthermore, there was a significant quadratic Area effect for RE_{HH}. The Area linear effect, however, wasn't significant in the case of RE^*_{HT} which demonstrates the danger of ignoring interactions in surveying research. From the centerpoint FCA, the optimum and worst setting of Area for RE^*_{HT} was 10. For RE_{HH}, however, the Area settings for the optimum were either 30 or 29 and was 10 for the worst case. For both responses, these settings from the centerpoint FCA matched both the optimum and worst case settings from the global grid search with one obvious minor exception. The Parents*Area and Spread*Area effects had maximum RE_{HH} values when Area

was 29, slightly less than the global optimum of 30. This difference would be due to the significant Area quadratic effect, but it is practically meaningless.

Two additional population scenarios were studied under the conditional FCA: Area*Sample where c_a equaled either 1 or 2 and Area*Critical where n_1 equaled 10 or 50. The optimum and worst settings under these four scenarios were Area values of 30 and 10, respectively, for one or both responses. The Parents*Area and Spread*Area interaction effects were the two significant terms common to both models. The response surface ranges were much greater for RE_{HH} than for RE^*_{HT}. Also, for the Spread*Area effect, the range for RE^*_{HT} was sufficiently small so that all conditions were favorable for ACS while such was not the case for RE_{HH}. On the other hand, there were two situations from the conditional FCA where the ranges were greater for RE^*_{HT}. One was where Area and Sample varied with Critical set equal to 1 and the other was where Area and Critical varied and Sample equaled 10.

The significant Area*Sample interaction effect for RE^*_{HT} which has already been discussed in the context of initial sample size, will now be discussed in the context of the size of the study area. Figure 23, from the centerpoint FCA, shows that RE^*_{HT} improved with decreasing size of the study area given the largest setting for n_1 (50) but the opposite was true when the initial sample size was set to its lowest level (10). Both linear effects were roughly equivalent in terms of absolute magnitude. For the smallest sample size, the optimum RE^*_{HT} was obtained when Area = 30 while for the largest sample size, all conditions were favorable for ACS. Results from the conditional

FCA were generally consistent with these conclusions (Figures 39 through 42) for the two levels of Critical. There was no significant Area*Sample interaction effect for RE_{HH}. Consequently, an examination of the response surface for the population under the conditional FCA where $c_a = 1$ will not reveal any differences between the conditional linear effects with respect to direction and magnitude (Figures 51 through 52). Nonetheless, it is worthy to note the magnitude is very large relative to the effect of Sample, that RE_{HH} performance increased with increasing values of Area, and that no conditions were favorable for ACS when Area = 10 while the opposite was true when Area = 30. Virtually all conditions are favorable for ACS regardless of the initial sample size once the size of the area is larger than approximately 12 to 13.

Although the Area*Critical interaction effect was not significant for the RE^*_{HT} response, Figures 37 and 38 from the conditional FCA also support the conclusion that RE^*_{HT} performance is enhanced for larger study area sizes, say Area $\geq$ 20, across the levels of Critical, when the initial sample size is at its lowest level. Figures 43 through 46 show plots of the Critical*Sample interaction effect (not significant) for RE^*_{HT} across both extreme levels of Area (10 and 30). When Area = 10, RE^*_{HT} performance was best, in general, for large initial sample sizes, say $n_1 \geq 30$, regardless of the setting for critical value. However, when Area = 30, RE^*_{HT} performance improved with decreasing initial sample size regardless of the critical value level. In fact, all conditions were favorable for ACS.

There was a significant interaction effect for Area*Kids for RE_{HH} only. Figure 34 from the centerpoint FCA shows strong improvements in RE_{HH} as the size of the study area increases with the improvement being most dramatic when there were the most number of offspring (Kids = 50). Curvature in the response surface was due to a significant quadratic Area effect, but it is practically meaningless. An Area value of 30 gave the maximum RE_{HH} regardless of the value of Kids.

The significant Area*Critical interaction effect for RE_{HH} has already been discussed in the context of the critical value. In terms of the Area variable, a very large improvement in RE_{HH} occurs at both levels of Critical (more so for $c_a = 1$) for the centerpoint FCA (Figure 33) as the size of the study area increases. Maximum RE_{HH} values were obtained for the largest-sized study area regardless of Critical level. The same conclusions are reached from the conditional FCA regardless of the initial sample size (Figures 47 through 50). There was no significant Area*Critical interaction effect for RE^*_{HT} so consequently an examination of the response surface for the population under the conditional FCA where $n_1 = 10$ will not reveal any differences between the conditional linear effects with respect to direction and magnitude. Nonetheless, it is worthy to note that

(i.) the magnitude of the linear Area effect is very large relative to the effect of Critical,

(ii.) that RE^*_{HT} performance improved with increasing values of Area and,

(iii.) that no conditions were favorable for ACS when Area = 10 while the opposite was true when Area = 30.

Also from the same analysis, the importance of Area in explaining RE_{HH} is seen in Figures 53 and 54. Although there was no significant Critical*Sample interaction effect, an examination of the plots for these two variables shows the most important result to be that no conditions are favorable for ACS when Area is set to its smallest size while all conditions are favorable for ACS when Area is set to its largest size.

The Parents*Area interaction effect was significant for both responses. Figure 20 from the centerpoint FCA suggests conflicting results for RE^*_{HT}. There was an improvement in RE^*_{HT} with decreasing values of Area given the smallest setting for Parents (1). On the other hand, the opposite was true given the largest setting for Parents (5) with this linear effect being of larger magnitude. While all conditions were favorable for ACS given Parents = 1, the minimum RE^*_{HT} occurred for Area = 30 when Parents = 5. From the same analysis, for the case of RE_{HH}, there was an improvement regardless of the number of Parents, with increasing values of Area (Figure 31). The effect was of larger magnitude for the largest number of Parents. Regardless of the Parents level, optimum RE_{HH} was obtained when Area = 30. As was the case with the Area*Kids interaction effect, curvature in the response surface was due to a significant but practically meaningless quadratic Area effect.

To discuss the significant Spread*Area interaction effect, we first examine the response surface plot for RE^*_{HT} from the centerpoint FCA (Figure 18). An improvement in RE^*_{HT} was noted for decreasing size of the study area given the largest setting for Spread ($\sigma_x = .5$). However, a larger linear effect of an opposite nature was seen for the smallest setting for Spread ($\sigma_x = .1$). Despite this, all conditions were favorable for ACS. The plot for RE_{HH} (Figure 27) shows an improvement in RE_{HH} with increasing-sized study area regardless of the level of Spread, with the magnitude of the improvement being greatest when $\sigma_x = .5$. Regardless of the Spread level, optimum RE_{HH} was obtained when Area = 30. As we have seen with other interaction effects, slight curvature in the response surface was due to a significant but practically meaningless quadratic Area effect.

Figure 36 shows a two-dimensional plot for the significant quadratic Area effect from the centerpoint FCA for the RE_{HH} response. Curvature is slight and practically meaningless. Results are in line with previous results: there is an improvement in RE_{HH} as the size of the study area increases. A maximum RE_{HH} value was obtained for the largest setting of Area.

In conclusion, Tables 13 and 14 and an examination of the response surfaces from both the centerpoint FCA and the conditional FCA show that while Area is an important factor in explaining RE^*_{HT} it may very well be the most important factor in explaining RE_{HH}. The magnitudes of the comparable linear effects were in all cases greater for RE_{HH}. Defining the size of a study area is critical from a scaling

perspective because even a regularly distributed point pattern (Buzas and Hayek 1997) can be made to appear clustered by simply increasing the size of the margins of the study area. In terms of optimizing RE_{HH}, the conclusion regarding Area is evident: conduct the study in a large-sized study area. However, the results for RE^*_{HT} make conclusions less clear. In general, there is mild support for a large-sized study area with respect to optimizing RE^*_{HT}. As was the case when we examined Sample, caution must be exercised with this conclusion for conditions exist where a small study area would yield a desirable RE^*_{HT}. Such a population would have a small number of clusters with large cluster spread in conjunction with the selection of a large initial sample size. Once again, this scenario would in all likelihood lead to the formation of hypernetworks.

Cluster Morphology

To facilitate discussion, the following trio of fixed covariates will be used in the discussion of the morphology (or structure of clusters) and their levels will be given identifying descriptive labels.

(i.) Shape ($\frac{\sigma_y^2}{\sigma_x^2}$) with levels "circle" (-1 coded, 1 uncoded) and "cigar" (1 coded, 4 uncoded)

(ii.) Spread (σ_x) with levels "tight" (-1 coded, 0.1 uncoded) and "broad" (1 coded, 0.5 uncoded)

(iii.) Direct (ρ) with levels "not tilted" (-1 coded, 0.0 uncoded) and "tilted" (1 coded, 0.7 uncoded)

For RE^*_{HT}, the linear effects for Shape and Direct were not significant. It is interesting to note that for RE_{HH}, there was no significant linear effect for Direct nor was it involved in any significant interaction effect. Although cluster morphology is important in determining relative efficiency, an examination of ANOVA results in Tables 21 and 22, specifically the t values and significant interaction effects, could lead one to the conclusion that other variables are more important. We will see that this conjecture does in fact appear to be true. It is also noted that cluster morphology will only be discussed for the centerpoint FCA. Recall these variables were the fixed covariates whose levels were set for the subsequent conditional FCA.

Shape

In terms of the Shape variable, there were only two significant interaction effects, with one being held in common by both models. Looking first at the centerpoint FCA for the Shape*Spread interaction effect for RE^*_{HT}, the optimum and worst settings were Shape = circle, while it was Shape = cigar for both settings for RE_{HH}. Thus, there were cases from the centerpoint FCA where a setting did not match the setting from the global grid search. Recall, from the global grid search, that the optimum RE_{HH} setting was a circle and the worst RE_{HT} setting was a cigar.

Another discrepancy occured for the Shape*Critical interaction effect. The optimum RE^*_{HT} setting, a cigar, also did not match the setting, a circle, from the global

grid search. The response surface range for the Shape*Spread interaction effect was greater for RE_{HH} than for RE^*_{HT}. It is noted that for both the two significant interaction effects in the model for RE^*_{HT}, all conditions were favorable for ACS. Such was not the case for RE_{HH}.

There was a significant Shape*Spread interaction effect for both responses (Figures 14 and 25). RE^*_{HT} performance improved as the cluster became cigar-shaped given the cluster was tight. On the other hand, there was a slightly even bigger improvement as the cluster became circular given the cluster was broad. Regardless, all conditions were favorable for ACS. RE_{HH} performance improved as the cluster became cigar-shaped given the cluster was tight with all settings for Shape yielding RE_{HH} values greater than or equal to 1. Conversely, there was a much bigger improvement as the cluster became circular given the cluster was broad. Given the cluster spread was broad, the best RE_{HH} performance was when the cluster shape was circular. The magnitudes of the comparable linear effects were roughly equivalent.

The Shape*Critical interaction effect was unique to the RE^*_{HT} response (Figure 15) and was previously discussed in the section on controllable factors. RE^*_{HT} improved slightly as the cluster became more circular given $c_a = 2$. However, when $c_a = 1$, a smaller improvement occurred as the cluster became more cigar-shaped. Once again, all conditions were favorable for ACS.

In conclusion, the Shape variable does not appear to be important in explaining RE^*_{HT} or RE_{HH} relative to the other variables in the FCA, at least for the levels

chosen here. Consequently, these results do not appear to offer much support for the notion that the researcher should select a neighborhood that reflects the shape and direction of the clusters ((Christman 1996b, 2000)). For RE^*_{HT}, only weak support exists in favor of circular shaped clusters with respect to relative efficiency while, for RE_{HH}, perhaps a bit more evidence exists.

Direct

Results from the centerpoint FCA showed that tilted clusters are associated with both the optimum and worst RE^*_{HT} values. These settings matched those from the global grid search.

The Direct*Critical interaction was a significant effect unique to the RE^*_{HT} response (Figure 16) and was previously discussed in the section on controllable factors. RE^*_{HT} improved slightly as the cluster became less tilted given $c_a = 2$. However, when $c_a = 1$, a greater improvement, again slight, occurred as the cluster became more tilted. All conditions were favorable for ACS.

Christman (1997) stated that symmetry of the dispersion of the offspring around the parents was probably very unlikely to be observed in practice. Nonetheless, Direct appears to be the least important of all the variables in the FCA. Relative to the other variables and for these levels, within-cluster correlation was inconsequential in explaining RE_{HH}. There was only a weak preference towards tilted clusters for RE^*_{HT}.

Spread

We now look at Spread, the remaining variable under the cluster morphology grouping. For all four significant interaction effects, the optimum RE^*_{HT} setting from the centerpoint FCA was a broad cluster, unchanged from the global grid search. However, there were two instances where the worst RE^*_{HT} setting from the centerpoint FCA went from the worst case global grid search setting of a broad cluster to a tight one. These instances were for Spread*Shape and Spread*Area. It's interesting to note that for all six significant interaction effects for RE_{HH}, the optimum setting from the centerpoint FCA was tight, a change from the optimum global grid search setting of broad. On the other hand, the worst RE_{HH} setting from the centerpoint FCA was broad for all the significant interaction effects, unchanged from the global grid search. For three of the four significant interaction effects held in common by both response surface models of the FCA, the range of the minimum and maximum predicted responses was greater for RE_{HH} than for RE^*_{HT} (Tables 16 and 17). Three of the four significant interaction effects for the RE^*_{HT} model yielded only conditions favorable for ACS, whereas the six significant interaction effects for the RE_{HH} model yielded conditions that were both favorable and not favorable for ACS.

There was a significant Shape*Spread interaction effect for both responses and it was previously discussed with respect to Shape. With respect to Spread, RE^*_{HT} appeared to improve with increasing Spread only when the clusters were circles (Figure 14). All conditions were favorable for ACS. RE_{HH} improved with decreasing Spread,

with this linear effect being very evident when the clusters were cigars (Figure 25). Given the clusters were circular, all conditions were favorable for ACS. When the clusters were cigar-shaped, the optimum setting was for a cluster that was tight.

The Spread*Parents interaction effect was significant for both responses. An examination of Figure 17 shows an improvement in RE^*_{HT} with increasing Spread and that this linear effect occurred only for the lowest setting for Parents (1). All conditions were favorable for ACS. Figure 26 shows an improvement in RE_{HH} with decreasing Spread regardless of the number of parents. This linear effect was far more pronounced for the highest setting for Parents (5). Given there were few clusters, all conditions were favorable for ACS. When the clusters were most numerous, the optimum setting was for a cluster that was tight.

The Spread*Sample interaction effect was significant for both responses (Figures 19 and 30). RE^*_{HT} performance improved as the cluster became tight given the initial sample size was at its smallest level (10). Here, the optimum RE^*_{HT} setting was for a tight cluster. When the initial sample size was fixed to its largest level (50), the direction of the linear effect switched in favor of broad clusters and the effect itself was of much greater magnitude. Here, all conditions were favorable for ACS. For RE_{HH}, the direction of the linear effect was in favor of a tight cluster regardless of the initial sample size with the magnitude of this linear effect being somewhat greater for the smallest value of Sample. At the smallest value for Sample, the optimum setting was

for a tight cluster, while at the largest value for Sample, all conditions were favorable for ACS.

The Spread*Area interaction effect was significant for both responses. Figure 18 displays a major improvement in RE^*_{HT} with increasing Spread for the lowest setting of Area (10). All conditions were favorable for ACS. The figure for the Spread*Area RE_{HH} response surface (Figure 27) has curvature due to a significant quadratic effect for Area. Nonetheless, the practical significance of this curvature appears to be meaningless. RE_{HH} improved with decreasing Spread with this linear effect being much more pronounced for the lowest setting of Area. Once again we see the major role Area has in explaining RE_{HH}. None of the conditions are favorable for ACS given the smallest study area while all conditions are favorable for ACS given the largest study area.

The Spread*Kids interaction effect was significant only for the RE_{HH} response. An improvement in RE_{HH} occurred as Spread decreased and this linear effect was more pronounced for the highest setting of Kids (50) (Figure 29). The optimum setting was for tight clusters regardless of the number of offspring.

The Spread*Critical interaction effect was significant only for the RE_{HH} response. An improvement in RE_{HH} occurred as Spread decreased and this linear effect was more pronounced for the lowest setting for Critical (1) (Figure 28). The optimum setting was for tight clusters regardless of the value for c_a.

Spread obviously plays a significant role in determining the relative efficiency of both the Horvitz-Thompson and Hansen-Hurwitz estimators. Clearly, of the three variables discussed in this section that determine the cluster morphology, Spread is the most important in terms of the numbers and presence of significant interaction effects and the magnitude of their t values. There is no obvious generalized conclusion to be drawn based on what we see here, although once again the results demonstrate that some differences exist between the two estimators. Specifically,

(i.) Spread appears to be more important in explaining RE_{HH} than RE^*_{HT},

(ii.) the magnitudes of the comparable linear effects were roughly equivalent, and

(iii.) for RE^*_{HT}, there appears to be weak evidence in favor of a broad-shaped cluster.

Result (iii.) did not appear in the work of Brown (1996), but the methodology for that study was markedly different than the methodology used here. Most noticeably, our methodology utilized a designed experiment. Additionally, although a similar unit scaling was used, Brown (1996) utilized a different radial distance function from the cluster center along with a different domain input of distances. This in turn could attenuate the variable Pervarw, which the literature has shown, and this research will show, is an important explanatory variable for relative efficiency. The conclusions drawn for RE_{HH} are even more inconclusive. Although the global grid search suggested performance would be best for broad-shaped clusters, we have found a population scenario, i.e. the centerpoint condition, where there was mild to strong

support in favor of a tight cluster. There were several instances where RE^{*}_{HT} performance improved with increasing σ_x whereas for RE_{HH} this was never true. We have seen that the conclusions concerning Spread reinforce the necessity to exercise caution, a point we have strived to convey to the reader. As is often the case in an analysis involving significant interaction effects, care must be taken not to generalize conclusions across all conditions. The reader has to consider population settings on a case-by-case basis for the specific estimation type used.

Cluster spread and its subsequent impact on cluster aggregation is a variable warranting further research. In her research, Christman (1996b) found that RE_{HH} and RE_{HT} values were high for tightly clustered points noting further that as cluster spread increased, the units within networks became more homogeneous and within-network variability decreased. However, there are many major differences between her research and ours, most notably the selected domains of cluster spread and other variables common to both studies followed by the areas of replication, simulation, and the addressing of interaction. Nonetheless, we will discuss the relationship between Spread and Pervarw at further length later in this discussion.

Number of Parents and Children

The number of parents, or clusters, and the number of children, or offspring, were important explanatory variables in both response surface models from the FCA (Tables 13 and 14). Both variables will only be discussed for the centerpoint FCA

because they were fixed covariates whose levels were set for the subsequent conditional FCA.

Results from the centerpoint FCA yielded values of 1 and 5 parents for the optimum and worst settings respectively for both responses. These settings were unchanged from the global grid search results. Response surface ranges from the same analysis appeared to be substantially greater for RE_{HH} for two of the three commonly-held interactions. In one case, the Spread*Parents interaction, all conditions were favorable for RE^*_{HT} .

The Spread*Parents interaction effect was significant for both responses and was discussed in detail in the context of cluster morphology. In terms of the number of parents, Figures 17 and 26 show improvements in both RE^*_{HT} and RE_{HH} performance respectively as the number of parents decreases across the range of values for Spread with the improvement being greatest for the largest value of Spread ($\sigma_x = .5$). As mentioned, all conditions were favorable for RE^*_{HT}. Given the smallest value for Spread ($\sigma_x = .1$), all conditions were favorable for ACS for RE_{HH}. When Spread was set to its largest value, the optimum setting for RE_{HH} was one parent.

The Parents*Area interaction effect was significant for both responses and was discussed in detail in the context of controllable factors. In terms of the number of parents, Figures 20 and 31 show improvements in both RE^*_{HT} and RE_{HH} performance as the number of parents decreases across the range of values for Area with the improvement being greatest for the smallest value of Area (10). Looking at RE^*_{HT},

when Area was at its smallest value, the optimum setting was one parent while all conditions were favorable for ACS when Area was set to its largest value. Note once again the importance of the Area variable in terms of explaining RE_{HH}. No conditions were favorable for ACS regardless of the number of parents when Area was at its smallest value while all conditions were favorable for ACS regardless of the number of parents when Area was at its largest value (30).

The Parents*Kids interaction effect was significant for both responses. The performance of both RE^*_{HT} and RE_{HH} improved with a decreasing number of parents regardless of the number of offspring (Figures 21 and 32). The improvement was of greatest magnitude for the largest number of offspring (Kids = 50). Looking at RE^*_{HT}, when Kids was at its smallest value (10), the optimum setting was one parent while all conditions were favorable for ACS when Kids was set to its largest value. The optimal RE_{HH} response was obtained for one parent for both the smallest and largest values for Kids.

The Parents*Sample interaction effect was significant for the RE^*_{HT} response only and was discussed in the context of controllable factors (Figure 22). In terms of the number of parents, we once again see an improvement in RE^*_{HT} performance regardless of the setting for Sample as the number of parents decreases. The improvement was greatest for the smallest initial sample size ($n_1 = 10$). When Sample was at its smallest value, the optimum setting was one parent while all conditions were favorable for ACS given the largest initial sample size (50).

Both responses behaved similarly with respect to Parents. However, the magnitudes of the comparable linear effects and magnitudes of the t values from the three commonly-held interaction were in all cases greater for RE_{HH}. This suggests that the Parents variable might be more important in explaining RE_{HH} as opposed to RE^{*}_{HT}. In any event, in what appears to be a clearcut and key conclusion, we see that for both responses one would desire to have a scenario where there are few clusters. This result has been corroborated by both Brown (1996) and Christman (1996a) for RE_{HT} and by Christman (1996b) for RE_{HH}.

The optimum setting from the centerpoint FCA was 50 offspring for both responses, thus matching the global grid search results. However, for RE^{*}_{HT}, the worst setting from the analysis of 10 offspring did not match the global grid search results for either significant interaction involved. Likewise, for RE_{HH}, the worst setting from the centerpoint FCA of 10 offspring did not match the global grid search results for three of the significant interactions involved. Response surface ranges from the centerpoint FCA were greater for RE^{*}_{HT} than RE_{HH} for the two interactions common to both models.

The Area*Kids interaction effect was significant for only the RE_{HH} response and was discussed in the context of controllable factors (Figure 34). In terms of the number of offspring, RE_{HH} performance improved substantially with an increasing number of offspring when Area was at its largest level. Furthermore, all conditions are favorable for ACS which emphasizes the importance of the Area variable in explaining

RE_{HH}. However, given Area was at its smallest level, we actually see a very slight improvement in RE_{HH} performance with a decreasing number of offspring with no conditions being favorable for ACS.

We now look at the Spread*Kids interaction effect, significant only for the RE_{HH} response (Figure 29). In terms of the number of offspring, we see an improvement in RE_{HH} performance with an increasing number of offspring regardless of the level of Spread. The magnitude of this linear effect was greatest for the smallest setting for Spread ($\sigma_x = .1$) where all conditions were favorable for ACS. When Spread was at its largest setting, no conditions were favorable for ACS.

The Kids*Sample interaction effect was significant for both responses and was discussed in the context of controllable factors (Figures 24 and 35). In terms of the number of children, there is an improvement in performance of both responses with an increasing number of offspring regardless of the initial sample size. The magnitude of the linear effect is greatest when n_1 was at its largest size (50) for both responses. Recall the importance of the Sample variable for RE^*_{HT}. No conditions were favorable for ACS when n_1 was set to its smallest size (10) while all conditions were favorable for ACS when n_1 was its largest size. On the other hand, for RE_{HH}, the optimum setting was for the largest number of offspring regardless of the two extreme Sample levels.

The Parents*Kids interaction effect was significant for both responses. RE^*_{HT} performance improved with an increasing number of offspring regardless of the number

of parents. The improvement was of greatest magnitude for the smallest number of parents (1). All conditions were favorable for ACS when there was but one parent while the optimum setting was 50 offspring given there was 5 parents. The very same conclusions were seen with RE_{HH} (Figure 32) with respect to magnitude and direction of the linear effects. Note that no conditions were favorable for ACS when the number of parents was at its largest setting (5) while all conditions were favorable for ACS when the number of parents was at its smallest setting.

Both responses behaved similarly with respect to Kids. The magnitudes of the comparable linear effects were in all cases greater for RE^{*}_{HT}. Judging by the number of significant interactions and the magnitude of the t values, the number of offspring perhaps might be more important in explaining RE_{HH} than RE^{*}_{HT}. For our study, as with Parents, we once again see what appears to be a clearcut conclusion. For both responses one would desire to have a scenario where there are clusters containing many offspring. This result did not appear in the work of Brown (1996) for RE_{HT} but the methodology for that study was markedly different than the methodology used here. Most noticeably, our methodology utilized a designed experiment. Perhaps our results could be explained simply by the fact that for our experimental setup, more offspring per cluster may have desirably increased the levels of within-network variability.

The Random Covariates - Initial Results

We consider the behavior of the four random covariates (Pervarw, Predp, Numnets, and Rarity) by first examining the results from the initial study (see Table 6, Figure 8, Figure 9, and Figure 10). A portion of Table 10 is reproduced in Table 36.

Table 36. Summary of medians and ranges from simulation for initial study, $N = 500$.

	Nominal		Extreme1		Extreme2	
Parameter	Median	Range	Median	Range	Median	Range
$\frac{\sigma_W^2}{\sigma^2}$	0.198	0.615	0.537	0.652	0.038	0.553
Number of Networks	96.0	15.0	60.0	76.0	899.0	5.0
$\widehat{\lambda}_p$	4.9	8.3	4.9	10.4	0.9	3.5
$\frac{E[\nu]}{n_1}$	1.512	2.500	2.604	2.135	1.015	0.076

For the proportion of within-network variability, or Pervarw, the shape of the distribution was skewed right for the nominal setting, very skewed right for the extreme 2 setting and was relatively normal for the extreme 1 setting. Proceeding in the same order, the median went from 0.198 to 0.038 to 0.537. We thus conclude that the behavior of Pervarw changes considerably between population settings. Also, for all settings, the approximate spread or range of the distribution was a large 0.600, indicating considerable Pervarw variation within population settings as well.

The shape of the distribution for the number of study area networks, or Numnets, was skewed left, highly skewed left, and approximately normal for the nominal, extreme 2 and extreme 1 settings, respectively. As would intuitively be expected due to its strong association with Area, we note the median values of 96.0 for the nominal

setting and 899.0 for the extreme 2 setting. The reduced median value of 60.0 for the extreme 1 setting would be due to less fragmentation of the study area into networks in a majority of the populations due to the high population density. As was the case with Pervarw, we thus conclude major differences for Numnets exist between population settings. The ranges of the distributions were 15.0, 5.0 and 76.0 for the nominal, extreme 2 and extreme 1 settings respectively with the value of 76.0 being due to the just stated reason of less study area fragmentation and the occasional formation of hypernetworks.

For the predicted number of parents, or Predp, the distribution was roughly normal for both the nominal and extreme 1 settings but was highly skewed right for the extreme 2 setting. The median was 4.9 for the nominal and extreme 1 settings and 0.9 for the extreme 2 setting. These results aren't surprising given the actual number of parents, λ_p, for each case. Range values were 8.3, 3.5, and 10.4 for nominal, extreme 2, and extreme 1 settings, respectively, and indicate substantial variation within a population setting.

Next, we examine the rarity index $\frac{E[\nu]}{n_1}$, or Rarity. With Rarity, we again see substantive differences between population settings. In terms of the shape, the distributions are skewed right, highly skewed right, and skewed left for the nominal, extreme 2, and extreme 1 settings, respectively. In the same order the median values were 1.512, 1.015 and 2.604. The range values were 2.5 and approximately 2.1 for the nominal and extreme 1 settings, respectively, and shrunk to a surprising 0.0760

for the rarest population setting, Extreme 2. The implication here might be that if a population truly can be characterized as rare, then the rarity index will be very close to 1.000. We also note these results are in line with what has been seen in the literature. For example, variation in final sample size was greatest for the least patchy populations, or those with large and many clusters (Brown 1994). The results of Smith et al. (1995) clearly showed the lowest rarity values occurred for the species (green-winged teal) with the lowest density, or the most patchy distribution, while high rarity values occurred for those species (ring-necked ducks, for example) with a relatively higher density, or less patchy distribution.

Finally, we focus our attention on the 72 trial results from the study (see Figure 12 and Table 12). The distribution for Pervarw was somewhat skewed right with median = 0.216. There was a large amount of variability among the individual trials as can be evidenced by the large approximate range of 0.547. The distributions for Numnets and Predp were not enlightening. For Numnets, the trimodal distribution, with median and range values of 397.0 and 827.0, respectively, reflects the strong confounding with the levels of Area. For Predp, the distribution was bathtub-shaped with median and range values of 2.7 and 4.5 respectively. Intuitively, these values aren't surprising given Predp's relationship with Parents. Next, the sampling distribution for Rarity was very skewed right and had median and range values of 1.132 and an approximate 3.5 respectively. Although the range was large, the result isn't surprising given the wide array of population settings used in the study.

Within-Network Variability

When considering the proportion of total variability attributed to within-network variability, the optimum and worst-case settings from the global grid search were 0.547 and 0.000, respectively, for both responses (Table 23). These were also the highest and lowest values obtained for Pervarw. These results were matched in both the centerpoint RCA and the conditional RCA. There was one interaction, Pervarw*Numnets, held in common by both RCA models that allowed for a direct comparison of the response surface ranges. The RE^*_{HT} range was not only very large (1.449), it was much greater than the RE_{HH} range (0.705). In fact, as an aside, it was noticed that the response surface ranges for the RCA were generally much larger than those from the FCA. For both interactions in the model for RE_{HH}, there were no conditions found that were favorable for ACS in the centerpoint RCA.

The Pervarw*Numnets interaction effect was significant for both responses. For the centerpoint RCA, improvement in RE^*_{HT} occurred with increasing Pervarw with the effect being more pronounced for the smallest value of Numnets (73.0) (Figure 55). The optimum setting was 0.547 given 73 networks while no conditions were favorable for ACS given 900 networks, the largest value of Numnets. For RE_{HH}, improvement occurred with increasing Pervarw but the effect was more pronounced for the largest value of Numnets (Figure 58). There were no conditions favorable for ACS for this interaction. Slight but negligible curvature in the response surfaces for both responses

is attributed to significant quadratic effects and had no practical bearing on the conclusions. For the conditional RCA, the interpretation of the response surfaces for RE^*_{HT} across levels of Rarity would be the same as before (Figures 69 through 71). When Rarity was at its lowest level (1.009) and there were 73 networks, the Pervarw value of 0.547 was the optimum setting while all conditions were favorable for ACS when the number of networks was fixed to 900. When Rarity was set to its highest level (4.521), no conditions were favorable for ACS. The interpretation of the response surfaces for RE_{HH} across levels of Rarity and Predp would also be the same as before (Figures 76 through 78). When Rarity was at its lowest level and Predp was set to its lowest level (0.8), the Pervarw value of 0.547 was the optimal setting for both extreme levels of Numnets. Modulating the settings to values of 3.819 and 5.3 for Rarity and Predp, respectively, resulted in no conditions favorable for ACS.

There was a significant Pervarw quadratic effect for RE^*_{HT}. This curvature appears to be slight and practically meaningless. For the centerpoint RCA, performance improved with increasing Pervarw value. The minimum RE^*_{HT} was obtained when Pervarw was at its highest setting, but from Figure 56, we see there were almost no conditions favorable for ACS. Results from the conditional RCA contrasted sharply. Figure 72 shows that when both Numnets and Rarity are at their lowest settings, almost all conditions are favorable for ACS with Pervarw equal to 0.547 being the optimum setting. However, when Rarity is increased to its highest setting, there are no conditions favorable for ACS (Figure 73).

Finally, there was a significant Pervarw*Predp interaction effect for RE_{HH}. It appears to be of minor practical consequence. For the centerpoint RCA, Figure 59 shows an improvement in RE_{HH} with increasing Pervarw with the linear effect being slightly more pronounced for the lowest setting of Predp. There were no conditions favorable for ACS. The conditional RCA (Figures 79 through 81) showed the same type of response surface. For the highest setting of Numnets and the lowest setting of Rarity, maximum RE_{HH} was obtained at the highest Pervarw setting for both extreme levels of Predp. When Numnets was set to its lowest setting and Rarity set to a high level (3.819), there were no conditions favorable for ACS.

Judging by the t values and significant effects, within-network variability (Pervarw) appears to be an equally important major variable in explaining RE^*_{HT} and RE_{HH}. The results from this paper do not support the claim of Smith et al. (1995) that Pervarw was the most important variable in explaining RE^*_{HT}, although certainly it was important. One might assert, as we will soon see, that Rarity could be the most important explanatory variable for RE^*_{HT}. Solid evidence suggests that one would want the proportion of within-network variability to be high for both responses which corroborates the results from ACS research (see the section in the introduction, "Within-network Variability and Related Issues"). For RE^*_{HT}, additional evidence supporting this conclusion is seen in the plots for Pervarw*Rarity (not significant) where Numnets = 73.0 (Figures 63 through 64) and Numnets*Rarity (not significant)

across the range of settings for Pervarw (Figures 65 through 68). For RE_{HH}, additional evidence supporting this conclusion is seen in the plots for Pervarw*Rarity (not significant) across different Numnets and Predp combinations (Figures 82 through 84).

Careful consideration needs to be given to the proportion of within-network variability when designing any survey. The reason is based on the potential associations that could exist with other variables depending on their domains. We have discussed already that variables such as Spread and Critical can make the units within networks more or less homogeneous and thus attenuate Pervarw. In other words, the observed association between Pervarw and relative efficiency is due to common response whose lurking variables are the aforementioned. There are other variables that would affect Pervarw, such as the number of offspring within a cluster (Kids). Certainly, the scaling of the unit size, a variable not included in this study, would affect the proportion of within-network variability (Christman 1997).

Number of Study Area Networks

An examination of the ANOVA results in Tables 21 and 22 show that the random covariate Numnets is an important variable in explaining RE^{*}_{HT} and may be the most important variable in explaining RE_{HH}, especially for the RCA. This should not come as a surprise to the reader since Numnets is highly positively correlated with Area ($\mathrm{P}[R > |0.9953|) < .0001$). Numnets was involved in at least one significant

interaction effect for both models of the RCA. Furthermore, for RE_{HH}, there was a significant quadratic effect which is not surprising in light of the aforementioned correlation with Area. The reader is reminded that the theoretical maximum for Numnets is 900.0, the number of units in the 30 by 30 unit study area.

For the Pervarw*Numnets interaction, the only interaction both models had in common, the centerpoint RCA yielded Numnet values of 73.0 and 900.0 for the optimal RE^*_{HT} and RE_{HH}, respectively. This result is predictable considering the result obtained for Area from the FCA (Table 15). The worst Numnets setting was 73.0 for both responses. These settings matched the results from the global grid search. For the quadratic Numnets effect and the Numnets*Predp interaction, the optimum Numnets setting for RE_{HH} dropped slightly to 817.3 and 755.3 respectively, a minor drop of no substantive practical consequence. When populations from the conditional RCA were examined there were several deviations from the global grid search for RE_{HH}. For the Pervarw*Numnets interaction, the population where Predp was at its lowest level (0.8) and Rarity was at its lowest level (1.009) had a worst case setting of 900.0 networks. For the quadratic Numnets effect, the optimum Numnets setting dropped to 569.2 when Pervarw, Predp and Rarity were set to 0.000 (the lowest level), 5.3 (the highest level) and 3.819 respectively. Finally, for the Numnets*Predp interaction, the optimum Numnets setting dropped to 445.2 when Pervarw and Rarity were set to 0.000 and 3.819, respectively. In looking at the corresponding figures and the magnitude of the response values, these settings are again of no practical concern. Because

the interaction was common to both models, an examination of Pervarw*Numnets allowed for a direct comparison of the response surface ranges. The RE^*_{HT} range was not only very large (1.449), it was much greater than the RE_{HH} range (0.705). The magnitude of the linear effects at both extreme settings of Pervarw was also greater for RE^*_{HT} than for RE_{HH}. For the centerpoint RCA, Hansen-Hurwitz estimation was a poor performer for all three relevant terms because there were no conditions favorable for ACS. This type of extremeness was also seen in the conditional RCA. We see for the Pervarw*Numnets interaction, one constructed population for each response where no conditions are favorable for ACS. Continuing, for RE_{HH}, we see two populations for the quadratic Numnets effect and two for the Numnets*Predp interaction effect, where either all or none of the conditions are favorable for ACS (Tables 26 and 27).

The Pervarw*Numnets interaction effect, significant for both responses, was already discussed in the section of this discussion on within-network variability. We will now discuss it in the context of Numnets. Starting with the centerpoint RCA, for RE^*_{HT} (Figure 55), when Pervarw was at its highest level (0.547), we see an improvement in performance with a decreasing number of networks. The optimum setting for obtaining minimum RE^*_{HT} was 73.0 networks. However, the direction of the linear effect was reversed given Pervarw was set to its lowest level, a setting at which conditions were all very unfavorable for ACS. The magnitude of the effect was much greater when Pervarw = 0.547. For RE_{HH} (Figure 58), performance improved with

an increasing number of networks given that Pervarw was set to its highest level. When Pervarw was set to its lowest level, the curvature in the response surface made drawing a generalized conclusion impossible. Despite the curvature, the RE_{HH} values were extremely low. For the conditional RCA, the conclusions regarding magnitude and direction of the linear effects at the two opposing settings for Pervarw remain the same for RE^{*}_{HT} (Figures 69 through 71) across the extreme settings for Rarity. When Rarity = 1.009, all conditions were favorable for ACS given Pervarw was set to 0.547. When Pervarw was set to 0.000, the optimum setting was 900.0 networks. However, when Rarity = 4.521, there were no conditions favorable for ACS regardless of the Pervarw setting. As for RE_{HH} (Figures 76 through 78), the conclusions regarding magnitude and direction of the linear effects at the two opposing settings for Pervarw remain the same. When Predp = 0.8 and Rarity = 1.009, no conditions were favorable for ACS when Pervarw = 0.000, while the opposite was true when Pervarw = 0.547. When Predp = 5.3 and Rarity = 3.819, there were no conditions favorable for ACS regardless of the Pervarw setting.

As mentioned, there was a significant Numnets quadratic effect for the RE_{HH} response (Figure 61). Although this made interpretation of related response surfaces somewhat less clear than those we have already discussed throughout this text, there is no practical significance of this quadratic effect. For the centerpoint RCA, there is an improvement in performance with an increasing number of networks. Note that no conditions are favorable for ACS. When Pervarw = 0.547, Predp = 0.8, and Rarity

= 1.009 (Figure 93), all conditions were favorable for ACS with an improvement in RE_{HH} performance as the number of networks increased. On the other hand, when Pervarw = 0.000, Predp = 5.3, and Rarity = 3.819 (Figure 94), a setting that corresponded to the minimum RE_{HH} from the global grid search, the conditions were extremely unfavorable for ACS even though the direction of the linear effect is still in favor of an increasing number of networks. Additionally, the relationship between RE_{HH} and Numnets was more curvilinear.

The Numnets*Predp interaction effect was significant only for the RE_{HH} response (Figure 60). Since Numnets is highly correlated with Area, the conclusions for this interaction will be similar to the conclusions reached for the Area*Parents interaction from the FCA. From the centerpoint RCA, RE_{HH} performance improved with an increasing number of networks regardless of the predicted number of parents. The magnitude of this linear effect was greatest for the highest level of Predp. No conditions are favorable for ACS. With Pervarw at its highest level and Rarity at its lowest (Figure 85), all conditions were favorable for ACS and there is still a marked improved performance with an increasing number of networks. When Pervarw was set to its lowest level and Rarity was set to 3.819 (Figure 86), all conditions became very unfavorable for ACS. The response surface was mildly curved with the direction of the linear effect still in favor of an increasing number of networks, albeit barely.

In summary, there is an improvement in RE_{HH} performance as the number of networks increases regardless of other factors. This improvement becomes stronger

as the proportion of within-network variability increases. Any possibility that this improvement actually can be attributed to changes in Pervarw (confounding) can be ruled out because there is no evidence to suggest a significant linear association exists between Numnets and Pervarw ($P[R > |-0.0570|] = .6345$). Although there is mild support for a large number of networks (or as mentioned before, a large area) based only on the direction of the linear effects, for this model we note that whether or not the RE_{HH} value actually exceeds 1.000 will depend on the settings of the other factors. For RE^*_{HT}, we only see weak to even inconclusive evidence for a large number of networks because this is a different model with other variables and the presence of hypernetworks. The evidence here is less than that accrued in favor of a large-sized study area. Recall when we examined the Area variable that we witnessed differences in the direction of linear effects between levels of a factor within an interaction (see the corresponding response surfaces for the Area*Sample, Spread*Area and Parents*Area interaction) and the same thing occurs here with the Pervarw*Numnets interaction. We see that the conditions for RE^*_{HT} which would be favorable for ACS would depend also on the settings of the other variables. Additional support for these claims are seen in the response surface plots for Numnets*Rarity (not significant) for both RE_{HH} (Figures 87 through 89) and RE^*_{HT} (Figures 65 through 68).

Predicted Number of Parents

Many times throughout this discussion we have pointed out differences in the behavior associated with RE^*_{HT} and RE_{HH}. In this section, we again note differences involving Predp. Predp was significant only for the RE_{HH} model from part two of the analysis. Results from the preliminary multicollinearity analysis conducted before the variable selection process when generating our model for RE^*_{HT} had suggested that Predp need not even be considered as a candidate variable for the starting set.

Predp was involved in the significant Pervarw*Predp and Numnets*Predp interaction effects in the RE_{HH} model (which have already been discussed). The centerpoint RCA yielded values of 0.8 and 5.3 for the optimum and worst case settings, respectively. These results matched the settings from the global grid search. Similar results were seen with several populations under the conditional RCA with one minor exception. For the Pervarw*Predp interaction where Numnets was at its highest setting (900.0) and Rarity at its lowest setting (1.009), the Predp setting dropped to 0.8 for the worst case. Once again, unlike the models discussed in the FCA, we see major shifts in the response surfaces depending on the population. For both the centerpoint RCA and the conditional RCA, either all conditions were favorable for ACS or they were all not.

Figure 59 is a plot of the Pervarw*Predp response surface from the centerpoint RCA. There was an improvement in RE_{HH} performance as the number of predicted parents declined regardless of the Pervarw setting. This linear effect was much stronger in magnitude when Pervarw was at its highest setting (0.547). No conditions

were favorable for ACS regardless of the value for Pervarw. For the conditional RCA, we see that given Numnets was at its highest setting and Rarity at its lowest setting (Figure 79 and 80), virtually no significant linear Predp effect exists when Pervarw was at its lowest setting (0.000). No conditions were favorable for ACS. However, when Pervarw was at its highest setting, there was an improvement in RE_{HH} performance as Predp decreased. Here, all conditions were favorable for ACS. When Numnets was reduced to its lowest setting (73.0) and Rarity was increased to 3.819 (Figure 81), we still reach the same general conclusions regarding magnitude and direction of the linear effects. No conditions were favorable for ACS regardless of the value for Pervarw because initial conditions were most unfavorable for ACS.

In the Numnets*Predp plot (Figure 60) from the centerpoint RCA, an improvement in RE_{HH} can be seen with decreasing values for Predp regardless of the number of networks. The improvement was relatively greatest in magnitude for the lowest value of Numnets (73.0). No conditions were favorable for ACS regardless of the value for Numnets. For the conditional RCA, we first look at the condition where Pervarw was at its highest setting and Rarity at its lowest setting (Figure 85). Again we see the improvement in RE_{HH} across the values for Numnets as Predp values decreased. All conditions were favorable for ACS regardless of the value for Numnets. The same conclusions regarding magnitude and direction of the linear effects are reached but are more subtle in nature when Pervarw was at its lowest level and Rarity was set to 3.819 (Figure 86). We return to a scenario where no conditions were favorable for

ACS regardless of the value for Numnets. It should be noted that our results will ultimately show that the initial given conditions for Pervarw and Rarity were already most unfavorable for ACS.

Predp appears to be the least important variable in the RCA. It is obvious that there would exist a strong significant linear association between Predp and Parents ($P[R > |0.9830|] < .0001$). Thus, we observe the same general overall conclusion for RE_{HH} discussed in the section on number of parents. For RE_{HH}, a sampling situation would be ideal if few clusters were present in the study area. There are differences, however, between the FCA and the RCA. The evidence supporting the conclusion regarding the ideal sampling condition for RE_{HH} is mild compared to the evidence we saw regarding the Parents variable from the FCA. Similarly, whereas Parents was a significant explanatory variable in the FCA, we note again the lack of significance of Predp for RE^*_{HT}.

Rarity

An inspection of the RCA ANOVA results in Tables 21 and 22 reveals several interesting issues. First, Rarity was not involved in a significant interaction effect for either RCA model. Second, there was a highly significant quadratic effect in the RE_{HH} model. Third, by inspection of the tables, Rarity is an important explanatory variable for both models, arguably being the most important one for RE^*_{HT}. For the RE^*_{HT} model, we estimate an increase of approximately 1.03 in the average RE^*_{HT}

median for every one unit increase in the rarity index given that all other explanatory variables are already in the model. The presence of the significant quadratic effect makes a similar linear interpretation for the RE_{HH} model much less useful.

Results from the centerpoint RCA yielded an optimum setting of 1.009 for each response, or the lowest setting for Rarity, for each term in the respective models. This setting was the same as the optimum obtained under the global grid search. For RE^{*}_{HT}, the worst setting of Rarity was 4.521, which was also the highest setting for Rarity, and was the same result obtained under the global grid search. For the quadratic Rarity term for RE_{HH}, the worst setting was 3.555, a result not dissimilar from the value of 3.819 obtained under the global grid search. This discrepancy is all but meaningless for the range of RE_{HH} values associated with this high of a Rarity value. When we examine the worst case settings for RE_{HH} for the two scenarios under the conditional RCA, the Rarity value decreased slightly further to 3.467, still a practically meaningless discrepancy. For RE^{*}_{HT}, the response surface range for the three types of populations examined with respect to the Rarity term was huge (2.066). For RE_{HH}, the response surface range for the three types of populations examined with respect to the quadratic Rarity term was not nearly as large but, at a value of 0.818, was larger than any other range encountered in the FCA and RCA.

The linear Rarity effect for RE^{*}_{HT} under the centerpoint RCA can be seen in Figure 57. The direction of the effect is in favor of decreasing Rarity values culminating

in an optimum at a Rarity value of 1.009. For the two populations under the conditional RCA (Figures 74 and 75), we still see the improvement in RE^*_{HT} performance with decreasing levels of Rarity. Note that when the number of networks is 73.0 and Pervarw is set to 0.547, the optimum Rarity setting is still 1.009. However, when Pervarw is then set to 0.000, the result is that no conditions are favorable for ACS.

The plot in Figure 62 shows the significant Rarity quadratic effect for RE_{HH} from the centerpoint RCA. The maximum RE_{HH} was obtained at the lowest setting for Rarity. Although there is significant curvature, there are no practical implications because the neighborhood of critical values associated with the global minimum already yielded predicted RE_{HH} values extremely unfavorable for ACS. Figure 95, from the conditional RCA, shows an upward shift of the graph when Pervarw = 0.547, Numnets = 900.0, and Predp = 0.8. This result is not surprising in light of our previous discussion because those are settings which are favorable for ACS. Maximum RE_{HH} was again obtained at the lowest setting for Rarity. Figure 96 shows a downward shift of the graph when Pervarw = 0.000, Numnets = 73.0, and Predp = 5.3. Once again, this result is not surprising because these conditional settings have been shown to be unfavorable for ACS. In fact, the plot shows that no favorable condition exists for ACS.

We now look at plots from the conditional RCA for the remainder of our discussion on Rarity despite the absence of significant interaction effects in either model.

Figure 63 is a RE^*_{HT} response surface plot for Pervarw*Rarity given Numnets was at its lowest level (73.0). There was a dramatic improvement in RE^*_{HT} performance with decreasing values for Rarity regardless of the setting for Pervarw. Figure 64 shows the region of favorability, given a small study area, to be associated with proportionally high within-network variability and low rarity index values. Given a large study area (Numnets = 900.0) and the smallest setting for Predp (0.8), these conclusions were also obtained for the RE_{HH} model (Figure 82 and 83) given Rarity was approximately less than 3.000. This restriction on the domain was due to the significant Rarity quadratic effect. When conditions were set to a small study area (Numnets = 73.0) and the highest setting for Predp (5.3), the overall conclusions remained the same but the entire response surface ended up being essentially shifted substantially downwards (Figure 84).

We now examine RE^*_{HT} response surface plots for Numnets*Rarity (Figures 65 through 68). For the condition where Pervarw = 0.000, we again see the dramatic improvement of RE^*_{HT} as the Rarity values decreased regardless of the value for Numnets. Only a very small region of favorability exists for the highest settings for Numnets (around 900.0) and lowest setting for Rarity. When Pervarw was then fixed to 0.547, the same general overall conclusion concerning Rarity is reached with the response surface now being shifted downwards and tilted slightly to favor a smaller number of networks, a consequence of the significant Pervarw*Numnets interaction effect. Consequently, the region of favorability is now much larger with RE^*_{HT} values

less than 1.000 now being obtained for the lowest settings of Rarity, say less than approximately 2.500, regardless of the number of networks. Figures 87 through 89 correspond to the RE_{HH} response. Given Pervarw = 0.547 and Predp = 0.8, the response surface indicates an improvement in RE_{HH} performance as the rarity index value decreased given Rarity was approximately less than 3.000. The tilt of the response surface was in favor of large numbers of networks. When Pervarw = 0.000 and Predp = 5.3, the response surface was simply thrust substantially downwards so an improvement in RE_{HH} performance with decreasing values for Rarity was still seen within the previous domain of interest. The region of favorability has all but disappeared.

The importance of Rarity in explaining RE^*_{HT} can be seen clearly in Figures 69 through 73. When Rarity was switched from its lowest to highest (4.521) setting, the response surface for Pervarw*Numnets shifted dramatically upwards. The shift was so dramatic that we go from having virtually all conditions favorable for ACS, save for those where there were low numbers of networks and proportionally low within-network variability, to one where no conditions are favorable for ACS. When Numnets was subsequently fixed at its lowest setting, the graphs of the curve for Pervarw yielded the very same results.

Figures 90 through 92 demonstrate the importance of Rarity in explaining RE_{HH}, via response surface plots for Predp*Rarity, although not as strikingly as was the case with RE^*_{HT}. Within the region of favorability, RE_{HH} performance improved with

decreasing levels of Rarity, without regard to the levels for Predp, for the condition where both Pervarw and Numnets were at their highest settings. When the condition was switched such that both Pervarw and Numnets were set to their lowest levels, the response surface shifted downward to the extent that there was no longer a region of favorability.

In summary, we arrive at the obvious solid conclusion that for either response, one would desire to have the lowest level of Rarity possible. Additionally, it is noted that, generally speaking, this conclusion is one that is robust to changes in the other variables. This corroborates findings in the history of ACS literature. Christman (1996b) stated that Hansen-Hurwitz estimation exhibited higher relative efficiency when the final and initial sample sizes were close in value. Smith et al. (1995) claimed that for RE_{HT} a small difference between the final sampling fraction and the initial sampling fraction is characteristic of an efficient design. However, for the same response, Christman (1996a,b) used results from her research and the previous study and showed that this was a sufficient but not a necessary characteristic due to those populations containing hypernetworks.

As with Pervarw, careful consideration needs to be given to the rarity index when setting up any survey. The reason is simple. In the introductory material we mentioned in the section "Sample Size and Cost - Theory" that $\frac{E[\nu]}{n_1}$ is decreasing in n_1 for both types of estimation. In other words, the observed association between Rarity and relative efficiency is due to common response where the lurking variable is

Sample. Hence, given that Rarity is significantly negatively associated with Sample and given the results up to date, we would want as large an initial sample as possible. However, one is again faced with the tradeoff between gain in relative efficiency and the unit cost of sampling.

Discussion of Confirmation Runs

An important component of the experimental cycle is to validate response surface models generated via the designed experiment through the use of confirmation trials. Table 37 shows the results of the confirmation runs as was described in the Methods section. The 95% prediction interval for y_0 was calculated using

$$\hat{y}(\mathbf{x}_0) \pm t^* SE[\hat{y}(\mathbf{x}_0)]$$

where

$$SE[\hat{y}(\mathbf{x}_0)] = \sqrt{\hat{\sigma}^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}$$

and t^* is the 97.5th percentile of the t distribution with $n-p-1$ where p is the number of predictor variables. Checkpoint settings used, in addition to the centerpoint, for the prediction intervals from the FCA are again from the results of the global grid search (Table 15).

We note a couple of minor issues. The two prediction intervals for RE^*_{HT} under the optimum RE^*_{HT} setting had lower limits that were negative and consequently set to zero. Also, the worstcase RE_{HH} setting yielded a value for Rarity that was outside

Table 37. Results of confirmation runs. The FCA PIs immediately to the right of the RE columns are the 95% prediction intervals from the FCA. The RCA PIs, analagously, are the 95% prediction intervals from the RCA with levels of the four random covariates listed in the Setting column. Observations that were not captured by their respective prediction intervals are marked with an asterisk.

Setting	RE^*_{HT}	FCA PI	RCA PI	RE_{HH}	FCA PI	RCA PI
centerpoint Pervarw = 0.286 Numnets = 392.0 Predp = 3.3 Rarity = 1.224	0.835	[0.600, 1.169]	[0.407, 1.004]	1.087	[0.893, 1.155]	[0.936, 1.216]
min RE^*_{HT} Pervarw = 0.455 Numnets = 95.0 Predp = 1.4 Rarity = 1.157	0.099	[0.000, 0.307]	[0.000, 0.394]	1.214	[0.830, 1.116]*	[0.951, 1.249]
max RE^*_{HT} Pervarw = 0.388 Numnets = 73.0 Predp = 4.9 Rarity = 4.192	1.905	[1.309, 1.954]	[1.705, 2.410]	0.238	[0.161, 0.446]	[0.061, 0.375]
min RE_{HH} Pervarw = 0.537 Numnets = 61.0 Predp = 4.9 Rarity = 6.274	1.842	[1.104, 1.753]*		0.141	[0.063, 0.351]	
max RE_{HH} Pervarw = 0.468 Numnets = 893.0 Predp = 0.9 Rarity = 1.107	0.532	[0.147, 0.789]	[0.157, 0.809]	1.609	[1.341, 1.625]	[1.384, 1.686]

the domain of Rarity values used to construct the model for RE_{HH} from part two of the analysis. Prediction done under this circumstance would require extrapolation and thus no RCA prediction intervals were constructed.

The ability of the models to predict new response observations seems reasonably sound. Two out of the eighteen cases had prediction intervals failing to capture their

respective observations. Those were for the model for RE_{HH} from the FCA when input with the optimum RE^*_{HT} setting and the model for RE^*_{HT} from the FCA when input with the worst case RE_{HH} setting. We also note, not surprisingly, the narrower interval widths for RE_{HH} as opposed to RE^*_{HT}.

Other Factors and Issues With Ideas Towards Future Research

A second trial of the initial study that replicated $N = 250$ populations for all three settings was run in order to assess the consistency of the program results. When the results from this second trial were compared to those from the original initial study, the reproducibility was excellent. We should also note that further work will be done to make the coding more efficient in an attempt to reduce processing time.

We mention in this section several other factors that were held fixed for the purposes of this study for a variety of reasons. Discussion of a factor in this section does not mean the factor is unimportant with respect to relative efficiency nor does it imply the factor is less important than those examined in this study. Rather, these factors are all fertile ground for future similar research using response surface methodology. It should be noted that we did not consider three-way interaction effects for possible inclusion in our models. Although such effects may occasionally be statistically significant, the emphases were placed on the principle of parsimony and on the ease of interpretation.

The location of the offspring within a cluster was assumed to follow a bivariate normal distribution for this study and almost all of the simulations in the literature. Other distributions could be used. For example, Brown (1994, 1996) assumed that each individual was located at a radial distance from the center of the cluster selected from an exponential distribution and at a random angle uniformly distributed between 0° and 360°. The program used in this study could easily be augmented to permit simulation for additional distributions. A review of the simulation history in the literature shows that only simple and compound symmetric covariances structures have been used for offspring location. Thus, our covariance matrix was deliberately unstructured not only for the purpose of running this experiment, but also to maintain generality for other types of covariance structures one might be interested in using.

Another area of potential future research addresses the restriction that the location of the clusters was uniformly distributed throughout the study area. In practice, particularly in the area of biology where socio-behaviors come into play, the clusters themselves may actually either display a tendency towards "flocking" or display an avoidance mechanism, such as "territoriality". Our computer program could be augmented to simulate these types of behaviors.

It is important to note that the results discussed in this paper apply only to the situation where the initial design is a SRSWOR. It is not known whether or not these results would extend to other types of initial designs, such as stratified random sampling. Thus, initial sample design would represent another area of research worth

exploring. Currently, the R simulation program is set up only for SRSWOR, but a future revision could incorporate other initial design choices.

The type of neighborhood was set to 1st order, a common choice that we saw no reason to deviate from. Given the ACS research history, we did not consider other types like 2nd order, truncated 1st order and 4-away. Our program could be augmented to allow for these different neighborhood types.

The shape of the study area might be a consideration. The current *genpop* function is set up for square study areas. However, if one is willing to accept rectangular sized units as modulated via the xscale and yscale inputs, then the number of rows and columns in the final population matrix can differ. The next version of the *genpop* function could easily be revised to allow for rectangular study areas containing either square or rectangular sized units.

Finally, in previous studies, the study area was fixed and the unit size was varied (see Smith et al. (1995), for example). As we have discussed, in our research the units were fixed to size 1 x 1 and the study area size was varied. Perhaps the former scenario might be more likely to be encountered in practice and consequently will be addressed in future research. The choice of the unit size would affect the degree of cluster aggregation or, moreover, the number and size of networks. As was discussed previously, this in turn will effect the proportion of within-network variability. Christman (1996b, 2000) suggested that if some prior knowledge of the size of the cluster (or value for Spread) could be obtained, then the efficiency of ACS could be

increased by scaling the unit size accordingly. The choice of the unit size would have to be carefully balanced so that a cluster of elements would be contained in several, but not many, units and, yet, more than one unit. Obviously, this choice is a difficult one. Buzas and Hayek (1997) state that a precision-optimal approach is to err on the side of more, smaller units than fewer, larger units when dealing with spatially aggregated populations. By picking a smaller unit size, a larger number of smaller networks of low density will be formed. This is beneficial because it would minimize the chance of forming hypernetworks. If the unit size is too small, however, the mean density will become low enough that even a clustered population can "appear" randomly distributed with the end result being a reduction in the value for Pervarw, not to mention the addition of many useless edge units. Conversely, let's consider a larger unit size. By picking a larger unit size, fewer, larger networks of high density will be formed. One advantage would be a reduction in the number of edge units with the possible formation of hypernetworks being an obvious disadvantage. If, however, the unit size is too large, a scenario can exist where too few networks are formed with again a reduced value for Pervarw. Buzas and Hayek (1997) also showed that for spatially clustered populations, the higher the mean unit density, more units would need to be sampled in order to ensure a desired level of precision for an estimate of the mean density. They further state that if a good deal of spatial aggregation exists and if the unit size can be adjusted, then selecting a unit size to ensure the mean density is less than 10 is a good choice to maximize precision. Some prior knowledge

of the size of the clusters at hand, be it through historical data or a pilot study, would be necessary in order to construct a Pervarw-optimal unit size.

It is also highly probable that unit size would interact with other factors. For example, Smith et al. (1995) detected what appeared to be a significant unit size and critical value interaction effect, specifically in the ring-necked duck data, although they never explicitly acknowledged this.

Finally, it could be reasoned that the choice of the unit *shape* would also affect the number and size of networks and might also be an area of future research. For example, perhaps we might choose to partition the study area into equally-sized hexagons.

One issue that can be confusing when developing an understanding of ACS is that we are dealing with a study population of units, not the target population of elements! Recall, a population is considered rare and clustered if only a few units contain elements, i.e. most of the y-values are zero or close to zero, and those units that do contain elements are spatially clustered so that nonzero units are contiguous (Christman 2000). The Rarity index is indirectly measuring the number of occupied units while Pervarw is a measure of the aggregate clustering, or patchiness, of the population of units. These two variables are independent of one another and do not consider the elements themselves. Few studies consider not only the rareness and the clustering of the units, but also consider the rareness and the clustering of the elements as well. Christman (2000) had noted that if one were estimating a population in which

the number of elements as well as the number of units containing those elements was small, ACS would not be recommended. In any event, as it turns out, one can easily contrive hypothetical populations for scenarios that differ markedly from an element-wise perspective, although the populations exhibit the same type of rareness and clustering from a unit-wise perspective. We attempted in this study to look at the rareness of elements (via the Kids and Parents variables) and the clustering of elements (via the Spread, Direct and Shape variables), but this is a fertile area for future research, particularly if the size of the unit is also considered.

From an implementation point of view, it is comforting to note that the two (arguably) most important variables, Area for RE_{HH} and Sample for RE^*_{HT}, are controllable factors. They are factors whose settings the researcher or designer of the survey can actually control and modify. It has been suggested that one possible mechanism for selecting appropriate settings would be to use the idea of a pilot study as discussed by Brown (1996). One possible modification and implementation proposed here would be to first census a small section of the study area. We will assume the use of the traditional first-order neighborhood. If the researcher wanted to include Critical as a pilot study variable, then he or she would be interested in selecting some possible values for c_a. One possible approach which has actually been alluded to in Brown (1994); Felix-Medina and Thompson (1999) would be to calculate Pervarw for the section that was censused. Analagous to the gage reproducibility and repeatability studies in the field of quality control, if the Pervarw value was high, the

recommendation would be to sample many units within a few clusters. Thus, c_a would be set low. On the other hand, if the Pervarw value was low, the recommendation would be to sample few units within many clusters and, thus, c_a would be set high. The author feels this is missing the point for a couple of reasons. First, recall that Critical was shown not to be important relative to the other explanatory variables. Furthermore, the evidence for a low value of c_a is weak to mild, depending on the response and across a broad spectrum of populations. Second, one might argue that if the value for Pervarw was low, ACS should not be done in the first place per the results from this paper. Christman (2000) had noted that increasing c_a caused a decrease in within-network variability and thus reduced the efficiency of ACS.

Area and Sample will therefore be the major role players to consider in any pilot study. For the pilot study, we will have the flexibility to vary the unit size within the fixed-size section of the study area. One approach to selecting the unit size would be to initially sample with a small unit size. By constructing increasingly bigger unit sizes and assessing the effect on within-network variability, we would be able to select a unit size per the detailed guidance presented several pages ago. We know the results from this paper suggested the use of a large study area, particularly if one were to be using Hansen-Hurwitz estimation. Alternatively, this suggests using a small unit size if the study area size was fixed. In terms of the initial sample size, our study, in general, suggested taking a large initial sample size. However, one would have to proceed with caution in the case of Horvitz-Thompson estimation (see the discussion

on the Area*Sample interaction in the Controllable Factors section). An equally important consideration would be the cost of sampling and this in and of itself might determine the initial sample size. Another approach to select both the unit size and the initial sample size might entail selecting various unit sizes and partitioning the study area section into networks. Sampling would be done at various initial sample sizes. Much as we have done here, after calculating the relative efficiency for the treatments, response surface techniques would be used to find the optimal unit size and initial sample size. The initial sample size would simply be the optimal sample size for the study area section multiplied by the inverse of the proportion of the study area the section area occupied. One assumption, perhaps unrealistic, made in regard to the pilot study, is that the section surveyed would be representative of the study area as a whole, thereby justifying the use of conclusions from the pilot study in the design of the actual survey. Nonetheless, the design and setup of an ACS survey through the use of prior information collected via a pilot study is a fertile area of research.

One of the problems associated with this study was the number of explanatory variables in the FCA and the RCA. For the sake of parsimony, we will discuss one idea that incorporates hindsight and the merging of the FCA and RCA. In the methods, multicollinearity problems were detected with the inclusion of a large number of explanatory variables. For example, we have already discussed the notion of common response regarding Pervarw. Consequently, after review of our results, we now know

how to eliminate some of the variables. In this discussion, we have mentioned two strong linear associations between a random covariate and a fixed covariate. Thus, for the sake of implementation of a designed experiment, we would eliminate the random covariate in each pair. Specifically, we would eliminate Predp in favor of Parents and eliminate Numnets in favor of Area. Whereas the study area size was allowed to vary in this case, it might be more useful and insightful to consider eliminating both the Area and Sample variable and use the initial sampling fraction as a variable instead.

Upon reflection on the results of this study, a question naturally arises: could we remove Pervarw? It too is a random covariate and not amenable to use in a designed experiment. We also know that it would have a significant linear association with at least one other variable from the FCA, i.e. a negative one with Critical. To answer this question, we used the simulation program. 100 populations were generated for each of the Nominal, Extreme 1, and Extreme 2 study populations from the initial study at the Spread values of 0.1 through 0.5 by 0.1. All other settings from the initial study were held constant (Table 6). The median Pervarw value was recorded, and two fitted models were generated using simple linear and quadratic regression of Pervarw on Spread. The results are summarized in the Table 38.

Consequently, we could have reasonably eliminated the Pervarw variable and used Spread and Spread^2 in its stead. We note that this quadratic association holds for this domain of Spread values and a fixed 1 by 1 unit size. However, in practice and as we mentioned in our discussion on unit size, the existence, strength, and form of

Table 38. Pervarw vs. Spread - simulation study. Values in the table are the median Pervarw values for 100 populations, the coefficients of (multiple) determination, and P-values for the test for the significance of the linear (quadratic) association.

	Spread, σ_x					Simple		Quadratic	
Population	0.1	0.2	0.3	0.4	0.5	r^2	Prob $> F$	r^2	Prob $> F$
Nominal	0.177	0.343	0.432	0.439	0.422	0.70	0.08	0.99	0.01
Extreme 1	0.390	0.496	0.519	0.563	0.537	0.73	0.06	0.97	0.03
Extreme 2	0.026	0.229	0.266	0.257	0.230	0.48	0.19	0.94	0.06

an association would need to be checked for the various unit sizes, for it is the value for Spread that will then be fixed. For this fixed unit size and these fixed population parameters, the quadratic form of this association tells us that we were overzealous in our choice of the upper limit of the domain for the Spread variable. In other words, the size of the cluster became too diffuse across an increasing number of units to the point where the within-network variability decreased.

We finish with a discussion of the random covariate Rarity. We should first clarify what constitutes a "low" rarity value. For the type of population that we would most likely apply ACS, i.e. a patchy or rare and clustered population, we consider the Extreme 2 setting from the initial study. From Table 10, we see the maximum Rarity value among the 500 populations was 1.078. In essence, "low" means very close to the lower bound of 1! Even though Smith et al. (1995) claimed in their research that the final sampling fraction was close to the initial sampling fraction, we claim upon inspection of their Table 1 that we would consider the Rarity index for only two out of six cases to be sufficiently low and still have good RE_{HT} performance

associated with them. Since their experimental setup was markedly different and their initial sampling fraction was low, it is not surprising that the Pervarw variable might therefore appear to be the most important explanatory variable overall. We also note that these two cases were instances where the quadrat size was reduced which, from our standpoint, would have the same effect as increasing the size of the study area. The remaining third species also showed a reduction in Rarity for the smaller unit size.

Consider the Extreme 2 example population (Figure 4). Clearly, the population can be characterized as being rare and clustered. Our results have told us that whether the initial sample size was small or large, a low rarity index value would occur and RE_{HT} performance would be good. Now consider the Extreme 1 example population (Figure 3) where a very large hypernetwork has formed. In this case, the population could not be considered rare by any stretch of the imagination. Because this is a small study area, our results tell us that if the initial sample size was small, a high rarity index would occur and RE_{HT} performance would be poor. The danger lies when the initial sample size is large. Depending on how large the initial sample size is, and although one will obtain a relatively lower rarity index value than the case where the initial sample size was small, the rarity value will still be high yet RE_{HT} performance will be great! Thus, we summarize and re-emphasize as follows. Obtaining a low Rarity index value is a sufficient but not a necessary condition for good RE_{HT} performance for if the population is not rare and clustered and the

initial sample size is high enough, then one can obtain a high rarity index value in conjunction with good RE_{HT} performance.

We look at some examples contrasting Hansen-Hurwitz and Hurwitz-Thompson estimation. Looking first at the Extreme 1 example population from the initial study (Table 9), the rarity index value of 3.072 would be considered high. We see the population is not rare. The RE_{HH} value of 0.107 is well below 1.00, while the RE_{HT} value of 2.652, is well above 1.00. The initial sample comprises a large 30% of the population units. Now, if we look at the 14th population of Trial 50 from the experimental results (Table 31), the initial sample comprises 50% of the population. The rarity index value of 1.560 has dropped accordingly but would still be considered high. Again, the population would certainly not be considered rare. The RE_{HH} value of 0.489 is again well below 1.00 while the RE_{HT} value is extremely high and well above 1.00. The value of RE_{HH}, less than 1.00 in both cases, more accurately reflects the true underlying rarity status of the population and the fact that, from a sample size and cost standpoint, one should not be using ACS in the first place. However, with Horvitz-Thompson estimation, we obtain high RE_{HT} values associated with high Rarity index values. In other words, even though one can obtain good performance with Horvitz-Thompson estimation in these previous types of example populations, it still comes at the expense of a large increase in the final ACS sample size.

In any event, achieving good RE_{HT} performance in conjunction with a high Rarity value at the cost of a huge effective final sample size provides the researcher a

budgetary post mortem. He or she has conducted ACS for a population that is not rare or not clustered. This demonstrates why one should always do a pilot study! We want the value for Rarity to be low, for a low Rarity value will not only imply good relative efficiency performance for both estimation types but reduced sampling costs as well. One could eliminate Rarity, concentrating instead on the initial sample size because we know that minimum Rarity is obtained for maximum initial sample size. There are other associations that exist as well that could help us minimize the Rarity index. We know that a smaller unit size (or larger study area) will produce smaller Rarity values ((Smith et al. 1995) and equation 3.1). However, we would still have the relative efficiency versus sampling cost tradeoff issue to contend with. Smith et al. (1995) stated that comparisons between ACS and SRS based on cost equations would tend to favor ACS more than comparisons based solely on expected effective sample size as was done in their research and the research conducted here. At this point, the reader is encouraged to review the section in Chapter 2 on Limiting Sample Effort and Cost in ACS in light of the results from this research.

It is unrealistic to ignore real world cost issues associated with a large sample size. There must be a balance between relative efficiency and sampling costs. Thus, we propose future research towards a simultaneous optimization of two response surfaces, both formed from the now reduced set of candidate variables. The first response would be relative efficiency while the second would be the rarity index value. Overlaying two-dimensional contour plots is impractical in light of the number of explanatory

variables at hand, even after removing variables recommended in our previous discussion. The traditional approach of Khuri and Conlon (1981) would not work either because their approach requires the responses to have no linear functional relationships. Obviously, that is not the case here. Two other approaches are worthy of investigation. One would be the dual response approach of Myers and Carter (1973). For our purposes, the secondary, or yield, variable would be the relative efficiency while the primary, or cost, variable would be the rarity index. We would attempt to establish and determine the conditions on the design variables that produce a minimum rarity index subject to the constraint that the relative efficiency was optimal (above or below 1.000). A most promising approach would be to use the desirability function of Derringer and Suich (1980). Modified desirability functions exist which allow for the use of more efficient gradient-based optimization methods. Additionally, the researcher can assign different weights or priorities among the responses so that, for example, having an optimal relative efficiency would be twice as important as having a minimal rarity index value ((Del Castillo et al. 1996)). Finally, it turns out that the final solution from this latter approach is hardly affected by correlations among the responses (Douglas Montgomery, personal communication, February 28, 2003).

REFERENCES CITED

R. Bivand and A. Gebhardt. Implementing functions for spatial statistical analysis using the R language. *Journal of Geographical Systems*, 2:307–317, 2000.

David Blackwell. Conditional expectation and unbiased sequential estimation. *The Annals of Mathematical Statistics*, 18:105–110, 1947.

John J. Borkowski. Network inclusion probabilities and Horvitz-Thompson estimation for adaptive simple Latin square sampling. *Environmental and Ecological Statistics*, 6:291–311, 1999.

J. A. Brown. The application of adaptive cluster sampling to ecological studies. In *Statistics in Ecology and Environmental Monitoring*, pages 86–97, 1994.

J. A. Brown. *The relative efficiency of adaptive cluster sampling for ecological surveys.* Mathematical and Information Sciences Report Series B:96/08, Massey University, Palmerton North, Department of Statistics, 1996.

J. A. Brown and B. J. F. Manly. Restricted adaptive cluster sampling. *Environmental and Ecological Statistics*, 5:49–63, 1998.

M. A. Buzas and L. C. Hayek. *Surveying Natural Populations.* Columbia University Press, 1997.

Claes-Magnus Cassel, Carl-Erik Särndal, and Jan Haakan Wretman. *Foundations of inference in survey sampling.* John Wiley & Sons, 1977.

A. Chaudhuri and J. W. E. Vos. *Unified theory and strategies of survey sampling.* Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], 1988.

Mary C. Christman. Comparison of efficiency of adaptive sampling in some spatially clustered populations. In *ASA Proceedings of the Section on Statistics and the Environment*, pages 122–126, 1996a.

Mary C. Christman. *Efficiency of adaptive sampling designs for spatially clustered populations.* Technical Report, Stanford University, Department of Statistics, 1996b.

Mary C. Christman. Efficiency of some sampling designs for spatially clustered populations. *Environmetrics*, 8:145–166, 1997.

Mary C. Christman. A review of quadrat-based sampling of rare, geographically clustered populations. *Journal of Agricultural, Biological, and Environmental Statistics*, 5(2):168–201, 2000.

Mary C. Christman and Feng Lan. Sequential adaptive sampling designs to estimate abundance in rare populations. In *ASA Proceedings of the Section on Statistics and the Environment*, pages 87–96, 1998.

Mary C. Christman and Jeffrey S. Pontius. Bootstrap confidence intervals for adaptive cluster sampling. *Biometrics*, 56(2):503–510, 2000.

Noel A. C. Cressie. *Statistics for spatial data (Revised edition)*. Wiley-Interscience, 1993.

Enrique Del Castillo, Douglas C. Montgomery, and Daniel R. McCarville. Modified desirability functions for multiple response optimization. *Journal of Quality Technology*, 28:337–345, 1996.

George Derringer and Ronald Suich. Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12:214–219, 1980.

Tonio Di Battista and Domenico Di Spalatro. A bootstrap method for adaptive cluster sampling. In *Classification and Data Analysis: Theory and Application – Proceedings of the Biannual Meeting of the Classification Group of the Societ Italiana di Statistica (SIS)*, pages 19–26, 1999.

Arthur L. Dryver and Steven K. Thompson. Improving unbiased estimators in adaptive cluster sampling. In *ASA Proceedings of the Section on Survey Research Methods*, pages 727–731, 1998.

Martin H. Felix-Medina and Steven K. Thompson. Adaptive cluster double sampling. In *ASA Proceedings of the Section on Survey Research Methods*, pages 339–344, 1999.

M. M. Hansen and W. N. Hurwitz. On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14:333–362, 1943.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685, 1952.

A. I. Khuri and M. Conlon. Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics*, 23:363–375, 1981.

E. L. Lehmann. *Theory of point estimation.* John Wiley & Sons, 1983.

P. J. McCarthy and C. B. Snowden. The bootstrap and finite population sampling. In *Vital and Health Statistics, Series 2, Volume 95, Public Health Service Publication 85-1369. Washington, D.C.: Department of Health and Human Services*, 1985.

M. N. Murthy. Ordered and unordered estimators in sampling without replacement. *Sankhyā*, 18:379–390, 1957.

Raymond H. Myers and Jr Carter, Walter H. Response surface techniques for dual response systems. *Technometrics*, 15:301–317, 1973.

J. S. Pontius and M. C. Christman. Bootstrap confidence intervals from adaptive sampling of an insect population. In *Proceedings of the Ninth Annual Kansas State University Conference on Applied Statistics in Agriculture, 1997*, pages 191–203, 1997.

Jeffrey S. Pontius. *Probability proportional to size selection in strip adaptive sampling.* Technical Report Series II, Theory and Methods, Kansas State University, Department of Statistics, 1996.

Jeffrey S. Pontius. Strip adaptive cluster sampling: Probability proportional to size selection of primary units. *Biometrics*, 53:1092–1096, 1997.

D. Raj. Some estimators in sampling with varying probabilities without replacement. *Journal of the American Statistical Association*, 51:269–284, 1956.

C. R. Rao. Information and accuracy attainable in estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

Jr. Roesch, F. A. Adaptive cluster sampling for forest inventories. *Forest Science*, 39: 655–669, 1993.

B. S. Rowlingson and P. J. Diggle. Splancs: Spatial point pattern analysis code in S-plus (STMA V34 4780). *Computers & Geosciences*, 19:627–655, 1993.

B. S. Rowlingson and P. J. Diggle. *Splancs supplement - spatial and spatial-temporal analysis.* Lancaster University, Lancaster, U.K., 1996.

M. M. Salehi. Rao-Blackwell versions of the Horvitz-Thompson and Hansen-Hurwitz in adaptive cluster sampling. *Environmental and Ecological Statistics*, 6:183–195, 1999.

M. M. Salehi and George A. F. Seber. Adaptive cluster sampling with networks selected without replacement. *Biometrika*, 84:209–219, 1997a.

M. M. Salehi and George A. F. Seber. Two-stage adaptive cluster sampling. *Biometrics*, 53:959–970, 1997b.

Richard L. Scheaffer, William Mendenhall, and Lyman Ott. *Elementary survey sampling (4th ed.).* PWS-Kent Publishing Co., 1990.

G. A. F. Seber and S. K. Thompson. Environmental adaptive sampling. In *Handbook of statistics. Volume 12: Environmental statistics*, pages 201–220, 1994.

R. R. Sitter. A resampling procedure for complex survey data. *Journal of the American Statistical Association*, 87:755–765, 1992a.

R. R. Sitter. Comparing three bootstrap methods for survey data. *The Canadian Journal of Statistics*, 20:135–154, 1992b.

David R. Smith, Michael J. Conroy, and David H. Brakhage. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, 51: 777–788, 1995.

Steven K. Thompson. Adaptive cluster sampling. *Journal of the American Statistical Association*, 85:1050–1059, 1990.

Steven K. Thompson. Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics*, 47:1103–1115, 1991a.

Steven K. Thompson. Stratified adaptive cluster sampling. *Biometrika*, 78:389–397, 1991b.

Steven K. Thompson. *Sampling.* John Wiley & Sons, 1992.

Steven K. Thompson. *Factors influencing the efficiency of adaptive cluster sampling.* Technical Report Number 94-0301, Technical Reports and Reprints Series, Pennsylvania State University, Department of Statistics, 1994.

Steven K. Thompson. Adaptive cluster sampling based on order statistics. *Environ-Metrics*, 7:123–133, 1996.

Steven K. Thompson and George A. F. Seber. Detectability in conventional and adaptive sampling. *Biometrics*, 50:712–724, 1994.

Steven K. Thompson and George A. F. Seber. *Adaptive sampling.* John Wiley & Sons, 1996.

APPENDICES

APPENDIX A

R Code

```
> genpop
function(lambdap, lambdao, var1, var2, cov, studymin, studymax,
         xscale, yscale)

{

##### function genpop()
#
# Generates a simulated population.
# Arguments: lambdap: Poisson lambda for mean number of parents
#                     in the study area,
#            lambdao: Poisson lambda for mean number of offspring
#                     per cluster,
#            var1: normal variance for offspring
#                  dispersal in x-direction,
#            var2: normal variance for offspring
#                  dispersal in y-direction,
#            cov: covariance for offspring dispersal,
#            studymin: minimum value coordinate for study area,
#            studymax: maximum value coordinate for study area,
#            xscale: scaling factor for unit width,
#            yscale: scaling factor for unit length.
#
#-----------------------------------------------------------------
                  require(splancs)
                  require(MASS)
                  # library(help = splancs)
                  # library(help = MASS)

                  areaA <- c(studymin, studymin,
                             studymax, studymin,
                             studymax, studymax,
                             studymin, studymax,
                             studymin, studymin)
                  Apts <- spoints(areaA)
                  areaB <- c(studymin - 1, studymin - 1,
                             studymax + 1, studymin - 1,
                             studymax + 1, studymax + 1,
                             studymin - 1, studymax + 1,
                             studymin - 1, studymin - 1)
                  Bpts <- spoints(areaB)

                  rows <- studymax/yscale
                  cols <- studymax/xscale

                  simpop <- matrix(data = 0, nrow = rows,
                                   ncol = cols)

                  mup <- lambdap/(rows*cols)
                  # cat("Intensity parameter per unit=", mup, "\n")

                  tlambdap <- mup*(rows + 2)*(cols + 2)
```

```
# cat("Area B mean=", tlambdap, "\n")

SDp <- sqrt((tlambdap^2 + tlambdap)/
            (1 - exp(-tlambdap))
            - (tlambdap/(1 - exp(-tlambdap)))^2)
parents <- sample(1:(1000*SDp), 1, prob = dpois(
           1:(1000*SDp),tlambdap), replace = T)
# cat("Area B number of parents=", parents, "\n")
tmup <- parents/((rows + 2)*(cols + 2))
predlambdap <- round(tmup*rows*cols,1)
# cat("predicted study area number of parents=",
#      predlambdap, "\n")

SDo <- sqrt((lambdao^2 + lambdao)/
            (1 - exp(-lambdao))
            - (lambdao/(1 - exp(-lambdao)))^2)
offspring <- sample(1:(1000*SDo), parents,
             prob = dpois(
             1:(1000*SDo),lambdao), replace = T)
# cat("Numbers of offspring=", offspring, "\n")

poptotal <- sum(offspring)
# cat("Original population total=", poptotal, "\n")
np <- rep(1:parents, offspring)
# print(np)

centers <- matrix(data = csr(Bpts, parents), nrow =
           parents, ncol = 2)
# print("Cluster center locations")
# print(centers)
VarCovar <- matrix(c(var1, cov, cov, var2), 2, 2)
xydis <- mvrnorm(n = poptotal, rep(0, 2), VarCovar)
points <- matrix(centers[np, ] + xydis,
                 nrow = poptotal, ncol = 2)
# cat("points dimensions", dim(points), "\n")
# print("Original offspring locations")
# print(points)

if (length(pip(points, Apts))/2 >= 1)
   {

   newpoints <- pip(points, Apts)
   # cat("Safety check=",
   #     (length(newpoints)/2), "\n")
   # print("Truncated offspring locations")
   # print(newpoints)

   plot(Bpts, asp = 1, type = "n")
   pointmap(points, add = T)
   pointmap(centers, pch = '+', add = T)
   polymap(Bpts, add = T)
   polymap(Apts, add = T)
```

```
                    # dev.print()

                    for(i in 1:(length(newpoints)/2))
                       {

                       simpop[ceiling(newpoints[i,2]/yscale),
                              ceiling(newpoints[i,1]/xscale)] <-
                       simpop[ceiling(newpoints[i,2]/yscale),
                              ceiling(newpoints[i,1]/xscale)] + 1

                       }

                    # newpoptotal <- sum(simpop)
                    # cat("Truncated population total=",
                          newpoptotal, "\n")

                    simpop <- simpop[rows:1,]
                    # print(simpop)

                    # detach("package:splancs")

                    returnval <- list(POP = simpop,
                                      predparents = predlambdap)

                    class(returnval) <- "genpop"

                    returnval

                    }

                 else
                    {

                    print("Try again!")

                    genpop(lambdap, lambdao, var1, var2, cov,
                           studymin, studymax, xscale, yscale)

                    }

}


> TraceNetwork
function(i, j, cn, C, POP)

{

##### function TraceNetwork()
#
# A recursive function used in the Netid function.
# Arguments: see Netid function.
```

```
#
#-------------------------------------------------
                        # print(paste(i, j, cn))

                        NET[i,j] <<- cn

                        row <- i
                        col <- j + 1

                        if (POP[row, col] >= C && NET[row, col] == 0)
                           {TraceNetwork(row, col, cn, C, POP)}

                           row <- i
                           col <- j - 1

                        if (POP[row, col] >= C && NET[row, col] == 0)
                           {TraceNetwork(row, col, cn, C, POP)}

                           row <- i + 1
                           col <- j

                        if (POP[row, col] >= C && NET[row, col] == 0)
                           {TraceNetwork(row, col, cn, C, POP)}

                           row <- i - 1
                           col <- j

                        if (POP[row, col] >= C && NET[row, col] == 0)
                           {TraceNetwork(row, col, cn, C, POP)}

}


> Netid
function(C, POP)

{

##### function Netid()
#
# Partitions the population into the set of mutually exclusive
# and exhaustive network labels.
# Arguments: r and c: population size is rxc; the padded population
#                     size will be r2xc2,
#            C: the critical value,
#            POP: the matrix of population y-values from genpop()
#
#-------------------------------------------------------------------
                    r2 <- length(POP[,1]) + 2
                    c2 <- length(POP[1,]) + 2

                    NET <<- matrix(data = 0, nrow = r2, ncol = c2)
```

```
                    POP <- cbind(0, rbind(0, POP, 0), 0)

                    cn <- 0

                    for(j in 2:(c2-1))
                       {

                       for(i in 2:(r2-1))
                          {

                          cn <- cn + 1

                          if(NET[i,j] == 0)
                            {

                          # cn <- cn + 1

                          # print(paste("Location:",i,",", j,"
                          # Network Number", cn))

                          if(POP[i,j] >= C)
                            {TraceNetwork(i, j, cn, C, POP)}

                          else
                            {NET[i,j] <<- cn}

                            }

                          }

                       }

                       NET[(2:(r2-1)),(2:(c2-1))]

}


> netstats
function(NET,POP,n)

{

##### function netstats()
#
# Computes necessary statistics associated with the set of
# networks -- sizes, sums, means, marginal initial intersection
# probabilities, Z_HH's, Z_HT's, a_i's and pi_i's.
# Arguments: NET: output from the Netid() function; the output
#                 is the set of network labels,
#            POP: the matrix of population y-values,
#            n: size of the initial sample.
#
```

```
#-------------------------------------------------------------

numrows <- length(POP[,1])
numcols <- length(POP[1,])
popsize <- numcols*numrows
netsum <- matrix(data=NA, nrow=numrows,ncol=numcols)
netm <- matrix(data=NA, nrow=numrows,ncol=numcols)
miip <- matrix(data=NA, nrow=numrows,ncol=numcols)

for(i in 1:length(POP))
   {

   netm[i] <- length(POP[NET == NET[i]])
   netsum[i] <-  sum(POP[NET == NET[i]])

   if(netm[i] == 1)

   miip[i] <- n/popsize

   if(netm[i] > 1)

   miip[i] <- 1 - (prod((popsize - n - netm[i] + 1):
              (popsize - n))/ prod((popsize - netm[i]
              + 1):popsize))

   }

netw <- netsum/netm
netzhh <- popsize*netw
netzht <- netsum/miip

r2 <- numrows + 2
c2 <- numcols + 2
padPOP <- cbind(0, rbind(0, POP, 0), 0)
padnetm <- cbind(0, rbind(0, netm, 0), 0)
padNET <- cbind(0, rbind(0, NET, 0), 0)
temp <- matrix(data = 0, nrow = r2, ncol = c2)

for(j in 2:(c2 - 1))
   {

   for(i in 2:(r2 - 1))
      {

      # cat("i and j are", i, j, "\n")

      if(padPOP[i,j] >= C)
        {temp[i,j] <- 0}

      else
        {
```

```
        identifier <- matrix(c(
                        padNET[i - 1, j],
                        padNET[i + 1, j],
                        padNET[i, j + 1],
                        padNET[i, j - 1],
                        padnetm[i - 1, j],
                        padnetm[i + 1, j],
                        padnetm[i, j + 1],
                        padnetm[i, j - 1],
                        padPOP[i - 1, j],
                        padPOP[i + 1, j],
                        padPOP[i, j + 1],
                        padPOP[i, j - 1]),
                        ncol=3)

        # print("identifier is")
        # print(identifier)

        identifier2 <-
        identifier[!duplicated(identifier[,1]),
                    , drop = F]

        # print("identifier2 is")
        # print(identifier2)

        if(sum(identifier2[,3]) == 0)
          {temp[i,j] <- 0}

        else
          {

          identifier3 <-
          identifier2[identifier2[,3] >= C, , drop = F]

          temp[i,j] <- sum(identifier3[,2])

          }

        }

      }

   }

neta <- temp[(2:(r2-1)),(2:(c2-1))]

pi <- matrix(data = NA, nrow = numrows, ncol = numcols)

for(i in 1:length(POP))
   {

   pi[i] <- 1 - (nchoosex((popsize - netm[i]
```

```
                        - neta[i]), n)/nchoosex(popsize, n))
                }

            returnval <- list(popsize = popsize, netm = netm,
                              netsum = netsum, netw = netw,
                              netzhh = netzhh, miip = miip,
                              netzht = netzht, neta = neta, pi = pi)

            class(returnval) <- "netstats"

            returnval

}


> print.netstats
function(NS)

{

##### function print.netstats()
#
# Prints labels for matrices from netstats function.
# Arguments: NS: all the network information output
#                from the netstats function
#
#-------------------------------------------------

cat("Size of population=", "\n")
print(NS$popsize)

cat("Matrix of network sizes m_i's=", "\n")
print(NS$netm)

cat("Matrix of network sums=", "\n")
print(NS$netsum)

cat("Matrix of network means=", "\n")
print(NS$netw)

cat("Matrix of network Z_HH's=", "\n")
print(NS$netzhh)

cat("Matrix of m.i.i.p's=", "\n")
print(NS$miip)

cat("Matrix of network Z_HT's=", "\n")
print(NS$netzht)

cat("Matrix of network a_i's=", "\n")
print(NS$neta)
```

```
cat("Matrix of marginal inclusion probs.=", "\n")
print(NS$pi)

}


> getsamples
function(NET, POP, numsamples, n, popsize)

{

##### function getsamples()
#
# Select adaptive cluster samples using an initial srswor.
# Arguments: NET: list of matrices of network labels,
#            POP: the matrix of population y-values,
#            numsamples: number of samples to select,
#            n: size of initial srswor sample,
#            popsize: population size from the netstats
#                     function.
#
#-------------------------------------------------------

                     sid <- matrix(data = NA, nrow =
                                   numsamples, ncol = n)
                     unitid <- matrix(data = NA, nrow =
                                   numsamples, ncol = n)
                     unityvalues <- matrix(data = NA, nrow =
                                   numsamples, ncol = n)

                     counter <- 1.

                     while(counter <= numsamples)
                          {

                          indexsample <- sort(sample
                                         (popsize, n))

                          unitid[counter, ] <- indexsample
                          unityvalues[counter, ] <- POP[indexsample]
                          sid[counter,  ] <- NET[indexsample]

                          # if(max(netzhh[indexsample] > 0.) &
                          #   max(netzht[indexsample] > 0.))
                            {counter <- counter + 1.}

                          }

                     cbind(unitid, unityvalues, sid)

}
```

```
> TraceSample
function(i, j, C, POP)

{

##### function TraceSample()
#
# A recursive function used in the finalACS.sample function.
# Arguments: i: row location in padded population of a
#               unit from the initial sample,
#            j: column location in padded population of a
#               unit from the initial sample,
#            C: the critical value,
#            POP: the matrix of population y-values.
#
#-------------------------------------------------------------

                        # print(paste(i, j, cn))

                        if(POP[i,j] < C)
                          {finalACS[i,j] <<- 1}

                        else
                          {

                          finalACS[i,j] <<- 1

                          row <- i
                          col <- j + 1

                          if(POP[row, col] >= C &&
                             finalACS[row, col] == 0)
                            {TraceSample(row, col, C, POP)}

                          if(POP[row, col] < C &&
                             finalACS[row, col] == 0)
                            {finalACS[row, col] <<- 1}

                          row <- i
                          col <- j - 1

                          if(POP[row, col] >= C &&
                             finalACS[row, col] == 0)
                            {TraceSample(row, col, C, POP)}

                          if(POP[row, col] < C &&
                             finalACS[row, col] == 0)
                            {finalACS[row, col] <<- 1}

                          row <- i + 1
```

```
                        col <- j

                        if(POP[row, col] >= C &&
                           finalACS[row, col] == 0)
                          {TraceSample(row, col, C, POP)}

                        if(POP[row, col] < C &&
                           finalACS[row, col] == 0)
                          {finalACS[row, col] <<- 1}

                        row <- i - 1
                        col <- j

                        if(POP[row, col] >= C &&
                           finalACS[row, col] == 0)
                          {TraceSample(row, col, C, POP)}

                        if(POP[row, col] < C &&
                           finalACS[row, col] == 0)
                          {finalACS[row, col] <<- 1}

                        }

}


> finalACS.sample
function(sample,C,POP)

{

##### function finalACS.sample()
#
# Generates a matrix of indicator values corresponding to
# the final ACS sample.
# Arguments: sample: the initial SRSWOR,
#            C: the critical value,
#            POP: the matrix of population y-values.
#
#-------------------------------------------------------

    finalACS <<- matrix(data = 0, nrow = length(POP[,1]) + 2,
                                  ncol = length(POP[1,]) + 2)
    padPOP <- cbind(0, rbind(0, POP, 0), 0)
    # print(padPOP)

     for(index in sample)
        {

        i <- (((index - 1) %% 2) + 1) + 1
        j <- (((index - 1) %/% 2) + 1) + 1
```

```
        # cat("index,i and j are", index, i, j, "\n")

        TraceSample(i,j,C,padPOP)

        }

    finalACS <- finalACS[(2:(length(POP[,1]) + 1)),
                         (2:(length(POP[1,]) + 1))]
    # cat("Matrix of indicator values for final ACS=", "\n")
    # print(finalACS)
    ESS <- sum(finalACS)
    # cat("Effective sample size=", ESS, "\n")
    ESS

}


> nchoosex
function(num, den)

{

##### function nchoosex()
#
# iteratively computes combination coefficient.
#
#-------------------------------------------

                    temp <- 1.

                    for(i in 1.:den)
                       {

                       temp <- (temp * (num - i + 1.))/i

                       }

                       return(temp)

}


> compute.estimates
function(NET, netm, sid, miip, n, popsize,
         netzht, netzhh, numsamples)

{

##### function compute.estimates()
#
# Estimate total and var(total estimate) using
# Horvitz-Thompson (HT) & Hansen-Hurwitz (HH)approaches.
```

```
# Arguments: NET: list with matrices containing network
#                    estimates,
#            netm: matrix of the number of units belonging to each
#                network including singletons,
#            sid: list of srswor samples,
#            miip: matrix of network marginal initial intersection
#                  probabilities,
#            n: size of initial sample,
#            popsize: population size
#            netzht: output from the netstats() function; the
#                    output is the set of network sum_i / alpha_i,
#            netzhh: output from the netstats() function; the
#                    output is the set of N * network mean_i,
#            numsamples: number of samples to select,
#
#-----------------------------------------------------------------

# cat("popsize=", popsize, "n=", n, "\n")
tauhatHT <- matrix(data = NA, nrow = numsamples, ncol = 1)
tauhatHH <- matrix(data = NA, nrow = numsamples, ncol = 1)
varhatHT <- matrix(data = NA, nrow = numsamples, ncol = 1)
varhatHH <- matrix(data = NA, nrow = numsamples, ncol = 1)

##### LOOP SUBROUTINES FOR COMPUTING THE ESTIMATED VARIANCE of TAU_HT

##### MAIN LOOP BEGINS

  var.loop1 <- function(sidred, NET, netm, miip, popsize, n)
                       {

  ##### SUB LOOP BEGINS

    var.loop2 <- function(sidred, NET, netm, miip, tempvar,
                          popsize, n, h)
                         {

    for(k in sidred)
       {

       if(netzht[k] != 0.)
         {

         if(h == k)
           {

           miiph <- miip[h]
           miipterm <- 1 - miip[h]
           # cat("miip(h,term)=", miiph, miipterm, "\n")
           # cat("h and k are...", h, k, "\n")

           }
```

```
        if(h != k)
          {

        miiph <- miip[h]
        miipk <- miip[k]
        phk <- nchoosex((popsize - netm[h] - netm[k]), n)/
          nchoosex(popsize, n)
        # cat("phk=", phk, "\n")
        miiphk <- miiph + miipk - (1 - phk)
        # cat("miip(h,k,hk)=", miiph, miipk, miiphk, "\n")
        miipterm <- 1. - (miiph *miipk)/miiphk
        # cat("miipterm=", miipterm, "\n")
        # cat("h and k are...", h, k, "\n")

          }

        tempvar <- tempvar + netzht[h]*netzht[k]*miipterm

        }

      }

    return(tempvar)

                          }

  ##### SUB LOOP ENDS

  tempvar <- 0.

  for(h in sidred)
     {

     # cat("if h=", h, "then tempvar=", tempvar, "\n")
     if(netzht[h] != 0.)
       {

       tempvar <- var.loop2(sidred, NET, netm, miip, tempvar,
                            popsize, n, h)

       }

     }

  return(tempvar)

                          }

  ##### MAIN LOOP ENDS

for(asample in 1.:numsamples)
   {
```

```
    # cat("sample no.=", asample, "\n ")
    # cat("network ids=", sid[asample, ], "\n")
    sidred <- unique(as.numeric(sid[asample,  ]))
    # cat("unique network ids", sidred, "\n")
    tauhatHT[asample,] <- sum(netzht[sidred])
    varhatHT[asample,] <- var.loop1(sidred, NET, netm, miip,
                                    popsize, n)
    tauhatHH[asample,] <- mean(netzhh[sid[asample,  ]])
    varhatHH[asample,] <- ((popsize - n)/(popsize * n * (n - 1.))) *
    sum((netzhh[sid[asample, ]] - tauhatHH[asample])^2.)

    }

cbind(tauhatHT, varhatHT, tauhatHH, varhatHH)

}


> HT.variance
function(NET, netzht, netm, miip, n, popsize)

{

##### function HT.variance()
#
# Calculates the variance of tauhat_HT for a population where:
#
# Var[tauhat_HT] = sum_j^capkappa sum_k^capkappa Z_HT_j*Z_HT_k *
#                  (alpha_jk - ( alpha_j * alpha_k ))
# Arguments: NET: output from the Netid() function; the output
#                 is the set of network labels,
#            netzht: output from the netstats() function; the
#                    output is the set of network sum_i /
#                    alpha_i;
#            netm: output from the netstats() function; the
#                  output is the matrix of the number of units
#                  belonging to each network including
#                  singletons;
#            miip: output from the netstats() function; the output
#                  is the matrix of network marginal initial
#                  intersection probabilities,
#            n: size of initial sample,
#            popsize: population size.
#
#-------------------------------------------------------------------

index <- 0

for(j in 1:length(unique(as.numeric(NET))))
   {
```

```
    if((netzht[unique(as.numeric(NET))])[j] != 0)
      {

      for(k in 1:length(unique(as.numeric(NET))))
         {

         if((netzht[unique(as.numeric(NET))])[k] != 0.)
           {

           if(j == k)
             {

             miipj <- (miip[unique(as.numeric(NET))])[j]
             miipterm <- miipj - (miipj * miipj)

             }

           else
             {

             miipj <- (miip[unique(as.numeric(NET))])[j]
             miipk <- (miip[unique(as.numeric(NET))])[k]
             miipjk <- miipj + miipk - 1. +
                       nchoosex((popsize -
                       (netm[unique(as.numeric(NET))])[j] -
                       (netm[unique(as.numeric(NET))])[k]), n)/
                       nchoosex(popsize, n)
             miipterm <- miipjk - (miipj * miipk)

             }

           index <- index + ((netzht[unique(as.numeric(NET))])[j] *
                    (netzht[unique(as.numeric(NET))])[k] * miipterm)

           }

         }

      }

   }

index

}


> Realdeal
function(C, n, numsamples, lambdap, lambdao,
         var1, var2, cov, studymin,
         studymax, xscale, yscale)
```

```
{

##### funtion Realdeal()
#
# Calculates a list of ACS outputs for a simulated population.
# Arguments: see headers of functions used herein.
#
#-------------------------------------------------------------

                    shape <- var2/var1
                    rho <-  cov/(sqrt(var1)*sqrt(var2))
                    spread <- sqrt(var1)

                    POPinfo <- genpop(lambdap, lambdao, var1, var2,
                               cov, studymin,
                               studymax, xscale, yscale)
                    attach(POPinfo)
                    # cat("Population matrix=", "\n")
                    print(POPinfo$POP)
                    # cat("Matrix of network labels=", "\n")
                    NET <- Netid(C, POPinfo$POP)
                    # print(NET)
                    Networkinfo <- netstats(NET, POPinfo$POP, n)
                    attach(Networkinfo)
                    # print.netstats(Networkinfo)

                    # sampleinfo <-  getsamples(NET, POPinfo$POP,
                    #                           numsamples, n,
                    #                           Networkinfo$popsize)
                    # cat("Sample unit ids y values network labels=",
                    #     "\n")
                    # print(sampleinfo)
                    # ESSvector <- matrix(apply(sampleinfo[,1:n], 1,
                    #                     finalACS.sample, C=C,
                    #                     POPinfo$POP=POPinfo$POP))
                    # cat("Vector of effective sample sizes=", "\n")
                    # print(ESSvector)
                    # avgESS <- mean(ESSvector)
                    # cat("Average ESS from simulation=", avgESS,
                    #     "\n")
                    # output <- compute.estimates(NET,
                    #                             Networkinfo$netm,
                    #                             sampleinfo[,(2*n +
                    #                             1):(3*n)],
                    #                             Networkinfo$miip,
                    #                             n,
                    #                             Networkinfo$popsize,
                    #                             Networkinfo$netzht,
                    #                             Networkinfo$netzhh,
                    #                             numsamples)
                    # cat("tauHT varhatHT tauHH varhatHH=", "\n")
                    # print(output)
```

```
tau <- sum(POPinfo$POP)
# cat("tau=", tau, "\n")
sigmasq <- sum((POPinfo$POP -
              mean(POPinfo$POP))^2.)/
              (Networkinfo$popsize - 1)
# cat("population variance=", sigmasq,"\n")

varB <- sum((Networkinfo$netw -
           mean(POPinfo$POP))^2.)/
           (Networkinfo$popsize - 1)

varW <- sum((POPinfo$POP -
           Networkinfo$netw)^2.)/
           (Networkinfo$popsize - 1)

pervarW <- varW/sigmasq

# avg.tauht <- mean(output[,1])
# cat("average tauhat_HT=", avg.tauht, "\n")
# avg.tauhh <- mean(output[,3])
# cat("average tauhat_HH=", avg.tauhh, "\n")

var.tauhatHH <- ((Networkinfo$popsize - n)/
                 (Networkinfo$popsize * n *
                 (Networkinfo$popsize - 1.))) *
                 sum((Networkinfo$netzhh -
                      tau)^2.)
# cat("variance of tauhat_HH=", var.tauhatHH,
      "\n")

# avg.varhatHH <- mean(output[,4])
# cat("average estimated variance of tauhat_HH=",
#     avg.varhatHH, "\n")

var.tauhatHT <- HT.variance(NET,
                Networkinfo$netzht,
                Networkinfo$netm,
                Networkinfo$miip,
                n, Networkinfo$popsize)

# cat("variance of tauhat_HT=", var.tauhatHT,
      "\n")

# avg.varhatHT <- mean(output[,2])
# cat("average estimated variance of tauhat_HT=",
#     avg.varhatHT, "\n")

Numnets <- length(unique(as.numeric(NET)))
predparents <- POPinfo$predparents
ExpESS <- sum(Networkinfo$pi)
# cat("Expected effective sample size=", ExpESS,
```

```
      "\n")
rarity <- ExpESS/n

var.tauhatSRS <- (Networkinfo$popsize*
                  (Networkinfo$popsize - ExpESS)*
                  (sigmasq/ExpESS))
# cat("Modified variance of tauhat_SRS=",
#     var.tauhatSRS, "\n")

reHT <- var.tauhatSRS/var.tauhatHT

reHH <- var.tauhatSRS/var.tauhatHH

list(shape = shape, direct = rho,
     spread = spread, parents = lambdap,
     area = studymax, critical = C,
     kids = lambdao, sample = n,
     tau = tau, sigmasq = sigmasq,
     varB = varB, varW = varW, pervarW = pervarW,
     Numnets = Numnets,
     predparents = predparents,
     ExpESS = ExpESS,
     rarity = rarity,
     var.tauhatHH = var.tauhatHH,
     var.tauhatHT = var.tauhatHT,
     var.tauhatSRS = var.tauhatSRS,
     reHT = reHT,
     reHH = reHH)

}


> Fireaway
function(simulations, C, n, numsamples, lambdap, lambdao,
         var1, var2, cov, studymin, studymax, xscale,
         yscale)

{

##### function Fireaway()
#
# Calls the Realdeal function for however many simulated
# populations are desired.
# Arguments: see Realdeal function.
#
#-------------------------------------------------------

initial.time <- proc.time()

results <- matrix(data = 0, nrow = simulations, ncol = 22)

for(i in 1:simulations)
```

```
    {

    cat("Population", i, "\n")

    results[i,] <- unlist(Realdeal(C, n, numsamples, lambdap, lambdao,
                   var1, var2, cov, studymin,
                   studymax, xscale, yscale))

    }

pervarW.median <- median(results[,13])
numnets.median <- median(results[,14])
predparents.median <- median(results[,15])
rarity.median <- median(results[,17])
reHT.median <- median(results[,21])
reHH.median <- median(results[,22])

results.names <- c("shape", "direct", "spread", "parents",
                   "area", "critical", "kids", "sample",
                   "tau", "sigmasq", "varB", "varW",
                   "pervarW", "Numnets", "predparents", "ExpESS",
                   "rarity", "vartauhatHH", "vartauhatHT",
                   "vartauhatSRS", "reHT", "reHH")
dimnames(results) <- list(NULL, results.names)
print(results)

results2 <- matrix(c((round(pervarW.median, digits = 3)),
                   (round(numnets.median, digits = 1)),
                   (round(predparents.median, digits = 2)),
                   (round(rarity.median, digits = 3)),
                   (round(reHT.median, digits = 3)),
                   (round(reHH.median, digits = 3))),
                    nrow = 1, ncol = 6, byrow = T)

results2.names <- c("pervarW.median",
                  "numnets.median",
                  "predparents.median",
                  "rarity.median",
                  "reHT.median",
                  "reHH.median")

dimnames(results2) <- list(NULL, results2.names)

write.table(results2, "thesis/data/sastable", append = T,
            quote = F, row.names = F, col.names = F)

cat("Total execution time in seconds =", signif((proc.time() -
    initial.time), 3)[3], "\n")

return(results2)

}
```

APPENDIX B

SAS Code

The following commented SAS program was used to generate the design.

```
options pageno = 1;

data in;

/*-------------------------------------------------------------*/
/*                                                             */
/* Revision - 4/30/02                                          */
/*                                                             */
/* Variables and Levels Key                                    */
/*                                                             */
/* Shape of Cluster; -1 = 1 and 1 = 4                          */
/* Direction of Cluster; -1 = 0 and 1 = .7                     */
/* Spread of Cluster; -1 = .1 and 1 = .5                       */
/* Number of Parents; -1 = 1, 0 = 3 and 1 = 5                  */
/* Study Area Dimension; -1 = 10, 0 = 20 and 1 = 30            */
/* Critical Value; -1 = 1 and 1 = 2                            */
/* Number of Offspring; -1 = 10, 0 = 30 and 1 = 50             */
/* Initial Sample Size; -1 = 10, 0 = 30 and 1 = 50             */
/*                                                             */
/*-------------------------------------------------------------*/

do shape = -1 to 1 by 2;
do direct = -1 to 1 by 2;
do spread = -1 to 1 by 2;
do parents = -1 to 1 by 1;
do area = -1 to 1 by 1;
do critical = -1 to 1 by 2;
do kids = -1 to 1 by 1;
do sample = -1 to 1 by 1;
output;
end;
end;
end;
end;
end;
end;
end;
end;
run;

proc optex data = in seed = 57162;
* examine design var;
model shape|direct|spread|parents|area|critical|kids|sample@2
parents*parents area*area kids*kids sample*sample;
generate method = detmax N = 72;
output out = dsgn1;
run;

* data one;
```

```
* set dsgn1 nobs = n;
* dummy = ceil(ranuni(12345)*n);
* run;

* proc print data = one;
* title 'Design 1';
* run;

* proc glm data = one;
* model dummy =
* shape|direct|spread|parents|area|critical|kids|sample@2
* parents*parents area*area kids*kids sample*sample / tolerance nouni;
* run;

data dsgn1a;
set dsgn1;
if shape = 1 then shape = 4;
if shape = -1 then shape = 1;
if direct = -1 then direct = 0;
if direct = 1 then direct = .7;
if spread = -1 then spread = .1;
if spread = 1 then spread = .5;
if parents = 1 then parents = 5;
if parents = 0 then parents = 3;
if parents = -1 then parents = 1;
if area = -1 then area = 10;
if area = 0 then area = 20;
if area = 1 then area = 30;
if critical = 1 then critical = 2;
if critical = -1 then critical = 1;
if kids = -1 then kids = 10;
if kids = 0 then kids = 30;
if kids = 1 then kids = 50;
if sample = -1 then sample = 10;
if sample = 0 then sample = 30;
if sample = 1 then sample = 50;
run;

proc print data = dsgn1;
run;
quit;
```

The following commented SAS program was used in the FCA for the response RE_{HT} and then subsequently edited for the response RE_{HH}.

```
options pageno = 1;
goptions reset = all device = win target = winprtm;

/*                                                          */
/* Key for variables:                                       */
/* X1 = shape of cluster                                    */
/* X2 = direction of cluster or rho                         */
/* X3 = absolute spread of cluster or sigma_x               */
/* X4 = number of parents or lambda_p                       */
/* X5 = study area                                          */
/* X6 = critical value                                      */
/* X7 = number of offspring or lambda_0                     */
/* X8 = initial sample size or n_1                          */
/*                                                          */

data sim1a;
input trial X1 X2 X3 X4 X5 X6 X7 X8;
cards;
 1 -1 -1 -1 -1 -1 -1  1 -1
 2 -1 -1 -1 -1  0  1  1 -1
 3 -1 -1 -1 -1  1 -1  1  0
 4 -1 -1 -1 -1  1  1  1  1
 5 -1 -1 -1  1 -1 -1  1  1
 6 -1 -1 -1  1 -1  1  1  0
 7 -1 -1 -1  1  0  1  0  1
 8 -1 -1 -1  1  1 -1  1 -1
 9 -1 -1  1 -1 -1 -1  1 -1
10 -1 -1  1 -1 -1  1  0  1
11 -1 -1  1 -1  1 -1  1  1
12 -1 -1  1  0  0  1  1  0
13 -1 -1  1  0  1  1  1 -1
14 -1 -1  1  1 -1 -1  1 -1
15 -1 -1  1  1 -1  1  1  1
16 -1 -1  1  1  1 -1  1  1
17 -1 -1  1  1  1  1  0 -1
18 -1  1 -1 -1 -1 -1  1  1
19 -1  1 -1 -1 -1  1  1  1
20 -1  1 -1 -1  1 -1  1 -1
21 -1  1 -1 -1  1  1  1  1
22 -1  1 -1  0 -1  1  0 -1
23 -1  1 -1  0  1 -1  0  1
24 -1  1 -1  1 -1  1  1  1
25 -1  1 -1  1  0 -1  1 -1
26 -1  1 -1  1  1  1  1 -1
27 -1  1  1 -1 -1 -1  1  1
28 -1  1  1 -1 -1  1  1 -1
```

```
29 -1  1  1 -1  1 -1  1 -1
30 -1  1  1 -1  1  1  0  0
31 -1  1  1  1 -1 -1  0  0
32 -1  1  1  1 -1  1  1 -1
33 -1  1  1  1  1 -1  1  1
34 -1  1  1  1  1  1  1  1
35  1 -1 -1 -1 -1 -1  1  1
36  1 -1 -1 -1 -1  1  1  0
37  1 -1 -1 -1  0 -1  1  1
38  1 -1 -1 -1  1 -1  0 -1
39  1 -1 -1  0 -1  1  1  1
40  1 -1 -1  0  1 -1  1 -1
41  1 -1 -1  1 -1 -1  1 -1
42  1 -1 -1  1 -1  1  1 -1
43  1 -1 -1  1  1 -1  1  1
44  1 -1 -1  1  1  1  1 -1
45  1 -1 -1  1  1  1  1  1
46  1 -1  1 -1 -1  1  1 -1
47  1 -1  1 -1  0 -1  1 -1
48  1 -1  1 -1  1  1  1  1
49  1 -1  1 -1  1  1  1 -1
50  1 -1  1  0 -1 -1  1  1
51  1 -1  1  1 -1  1  1  0
52  1 -1  1  1  0 -1  0  1
53  1 -1  1  1  1 -1  1 -1
54  1 -1  1  1  1  1  1  1
55  1  1 -1 -1 -1 -1  1 -1
56  1  1 -1 -1  0  1  1 -1
57  1  1 -1 -1  1  1  1 -1
58  1  1 -1 -1  1  1  1  1
59  1  1 -1  1 -1 -1  1  1
60  1  1 -1  1 -1  1  1  1
61  1  1 -1  1  0 -1  1 -1
62  1  1 -1  1  1 -1  1  1
63  1  1  1 -1 -1 -1  1 -1
64  1  1  1 -1 -1  1  1  0
65  1  1  1 -1 -1  1  1  1
66  1  1  1 -1  1 -1  1  1
67  1  1  1 -1  1 -1  1  1
68  1  1  1  0  0  1  1 -1
69  1  1  1  1 -1 -1  1  1
70  1  1  1  1 -1  1  1 -1
71  1  1  1  1  1 -1  0 -1
72  1  1  1  1  1  1  1  0
;

data sim1b;
input X9 X10 X11 X12 reHT reHH;
cards;
0.189 99 1.4 1.122 1.069 1.022
0 400 1.7 1.02 0.99 0.99
.
```

```
.
.
/* data follows */
.
.
.
0.384 863 5.3 1.918 0.877 0.844
0.302 871 5.3 1.536 1.015 0.921
;

data sim1c;
merge sim1a sim1b;
drop X9 X10 X11 X12;
reHT = 1/reHT;
X1X2 = (X1*X2);
X1X3 = (X1*X3);
X1X4 = (X1*X4);
X1X5 = (X1*X5);
X1X6 = (X1*X6);
X1X7 = (X1*X7);
X1X8 = (X1*X8);
X2X3 = (X2*X3);
X2X4 = (X2*X4);
X2X5 = (X2*X5);
X2X6 = (X2*X6);
X2X7 = (X2*X7);
X2X8 = (X2*X8);
X3X4 = (X3*X4);
X3X5 = (X3*X5);
X3X6 = (X3*X6);
X3X7 = (X3*X7);
X3X8 = (X3*X8);
X4X5 = (X4*X5);
X4X6 = (X4*X6);
X4X7 = (X4*X7);
X4X8 = (X4*X8);
X5X6 = (X5*X6);
X5X7 = (X5*X7);
X5X8 = (X5*X8);
X6X7 = (X6*X7);
X6X8 = (X6*X8);
X7X8 = (X7*X8);
X4X4 = (X4*X4);
X5X5 = (X5*X5);
X7X7 = (X7*X7);
X8X8 = (X8*X8);
run;

* proc print data = sim1c;
run;

/*                                                    */
```

```
/* The following code allows me to verify the G-efficiency */
/* of the design just as a precautionary measure.          */
/*                                                         */

* data dummy;
* do X1 = -1 to 1 by 2;
* do X2 = -1 to 1 by 2;
* do X3 = -1 to 1 by 2;
* do X4 = -1 to 1 by 1;
* do X5 = -1 to 1 by 1;
* do X6 = -1 to 1 by 2;
* do X7 = -1 to 1 by 1;
* do X8 = -1 to 1 by 1;
* output; * end; * end; * end; * end; * end; * end; * end; * end;
* run;

* proc optex data = dummy seed = 57162;
* model X1|X2|X3|X4|X5|X6|X7|X8@2
* X4*X4 X5*X5 X7*X7 X8*X8;
* generate method = detmax N = 72 augment = sim1c;
* run;

/*                                                         */
/* The following code is for the Box-Cox transformation.   */
/* It suggested to do the inverse transformation for reHT */
/* and none for reHH.                                      */
/*                                                         */

* %adxgen;
* %adxinit;
* %adxtrans(sim1c,tsim1c,reHH, X1 X2 X3 X4 X5 X6 X7 X8
            X1X2 X1X3 X1X4 X1X5 X1X6 X1X7 X1X8
            X2X3 X2X4 X2X5 X2X6 X2X7 X2X8 X3X4
            X3X5 X3X6 X3X7 X3X8 X4X5 X4X6 X4X7
            X4X8 X5X6 X5X7 X5X8 X6X7 X6X8 X7X8
            X4X4 X5X5 X7X7 X8X8);

/*                                  */
/* Checking for multicollinearity. */
/*                                  */

* proc glm data = sim1c;
* model reHT = X1|X2|X3|X4|X5|X6|X7|X8@2
              X4*X4 X5*X5 X7*X7 X8*X8 / tolerance nouni;
run;

/*                          */
/* Doing model selection. */
/*                          */

* proc reg data = sim1c;
* model reHT = X1 X2 X3 X4 X5 X6 X7 X8
```

```
              X1X2 X1X3 X1X4 X1X5 X1X6 X1X7 X1X8
              X2X3 X2X4 X2X5 X2X6 X2X7 X2X8 X3X4
              X3X5 X3X6 X3X7 X3X8 X4X5 X4X6 X4X7
              X4X8 X5X6 X5X7 X5X8 X6X7 X6X8 X7X8
              X4X4 X5X5 X7X7 X8X8 / selection = stepwise;
run;

* proc reg data = sim1c;
* model reHT = X1 X2 X3 X4 X5 X6 X7 X8
              X1X2 X1X3 X1X4 X1X5 X1X6 X1X7 X1X8
              X2X3 X2X4 X2X5 X2X6 X2X7 X2X8 X3X4
              X3X5 X3X6 X3X7 X3X8 X4X5 X4X6 X4X7
              X4X8 X5X6 X5X7 X5X8 X6X7 X6X8 X7X8
              X4X4 X5X5 X7X7 X8X8 / selection = backward;
run;

* proc reg data = sim1c graphics;
* model reHT = X1 X2 X3 X4 X5 X6 X7 X8
              X1X2 X1X3 X1X4 X1X5 X1X6 X1X7 X1X8
              X2X3 X2X4 X2X5 X2X6 X2X7 X2X8 X3X4
              X3X5 X3X6 X3X7 X3X8 X4X5 X4X6 X4X7
              X4X8 X5X6 X5X7 X5X8 X6X7 X6X8 X7X8
              X4X4 X5X5 X7X7 X8X8 /
              selection = cp best = 10 adjrsq aic mse;
* plot cp.*np. / cmallows = blue vaxis = 0 to 20 by 5;
run;

/*                                                  */
/* Checking multicollinearity of final model.       */
/*                                                  */

* proc glm data = sim1c;
* model reHT = X1 X2 X3 X4 X5 X6 X7 X8
              X1X3 X1X6 X2X6 X3X4 X3X5 X3X8
              X4X5 X4X7 X4X8 X5X8 X7X8
              / tolerance nouni;
run;

/*                                                  */
/* Using proc glm only to extract means for EDA.    */
/*                                                  */

* proc glm data = sim1c;
* class X1 X2 X3 X4 X5 X6 X7 X8
        X1X3 X1X6 X2X6 X3X4 X3X5
        X3X8 X4X5 X4X7 X4X8 X5X8 X7X8;
* model reHT = X1 X2 X3 X4 X5 X6 X7 X8
               X1X3 X1X6 X2X6 X3X4 X3X5
               X3X8 X4X5 X4X7 X4X8 X5X8 X7X8
               / nouni;
* means X1 X2 X3 X4 X5 X6 X7 X8
        X1X3 X1X6 X2X6 X3X4 X3X5
```

```
        X3X8 X4X5 X4X7 X4X8 X5X8 X7X8;
run;

proc reg data = sim1c;
model reHT = X1 X2 X3 X4 X5 X6 X7 X8
             X1X3 X1X6 X2X6 X3X4 X3X5 X3X8
             X4X5 X4X7 X4X8 X5X8 X7X8
             / dw influence spec cli;
output out = two rstudent = rstudent
                 residual = residual
                 predicted = pred;
run;

data two;
set two;
sabsr = sqrt(abs(residual));
run;

* proc print data = two;
run;

* proc gplot data = two;
* plot residual * pred / vref = 0;
* symbol i = sm99ps v = @ w = 1 h = 1 c = black;
run;

* proc gplot data = two;
* plot sabsr * pred;
* symbol i = sm99ps v = @ w = 1 h = 1 c = black;
run;

* proc univariate data = two normal;
* var residual;
run;

proc rank data = two normal = blom out = normset;
var residual;
ranks rankres;
run;

* proc gplot data = normset;
* plot residual*rankres;
run;

/*                                                          */
/* Following is a template to generate 3D centerpoint       */
/* plots for all the interaction terms.                     */
/*                                                          */

data in1;
set sim1c end = eof;
output;
```

```
if eof then do;
reHT = .;
do X1 = 0;
do X2 = 0;
do X3 = 0;
do X4 = 0;
do X5 = 0;
do X6 = 0;
do X7 = -1 to 1 by .05;
do X8 = -1 to 1 by .05;
do X1X3 = 0;
do X1X6 = 0;
do X2X6 = 0;
do X3X4 = 0;
do X3X5 = 0;
do X3X8 = 0;
do X4X5 = 0;
do X4X7 = 0;
do X4X8 = 0;
do X5X8 = 0;
X7X8 = X7*X8;
output;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
end;
run;

proc rsreg data = in1 out = three;
model reHT = X1 X2 X3 X4 X5 X6 X7 X8
             X1X3 X1X6 X2X6 X3X4 X3X5 X3X8
             X4X5 X4X7 X4X8 X5X8 X7X8 / covar = 19 predict noprint;
run;

data three;
set three;
if _N_ le 72 then delete;
```

```
run;

* proc print data = three;
run;

* proc g3d data = three;
* plot X7*X8 = reHT / rotate = 0 to 360 by 30;
* zmax = 1.43 zmin = .75;
run;

/*                                                  */
/* Following is code to extract the settings        */
/* that yield the reHT minimum for a given          */
/* pair of predictor variables to verify the        */
/* 3D centerpoint plot.                             */
/*                                                  */

proc means data = three noprint;
var reHT;
output out = junk min = min;
run;

data junk;
set junk;
drop _TYPE_ _FREQ_;
run;

data four;
retain temp 0;
merge three junk;
if min ne . then temp = min;
if reHT = temp;
keep reHT X7 X8;
run;

* proc print data = four;
run;

/*                                    */
/* Global grid search is started.     */
/*                                    */

data in2;
set sim1c end = eof;
keep X1 X2 X3 X4 X5 X6 X7 X8
     X1X3 X1X6 X2X6 X3X4 X3X5
     X3X8 X4X5 X4X7 X4X8 X5X8
     X7X8 reHT;
output;
if eof then do;
reHT = .;
do X1 = -1 to 1 by 2;
```

```
do X2 = -1 to 1 by 2;
do X3 = -1 to 1 by 2;
do X4 = -1 to 1 by 1;
do X5 = -1 to 1 by 1;
do X6 = -1 to 1 by 2;
do X7 = -1 to 1 by 1;
do X8 = -1 to 1 by 1;
X1X3 = X1*X3;
X1X6 = X1*X6;
X2X6 = X2*X6;
X3X4 = X3*X4;
X3X5 = X3*X5;
X3X8 = X3*X8;
X4X5 = X4*X5;
X4X7 = X4*X7;
X4X8 = X4*X8;
X5X8 = X5*X8;
X7X8 = X7*X8;
output;
end;
end;
end;
end;
end;
end;
end;
end;
end;
run;

* proc print data = in2;
run;

proc reg data = in2;
model reHT = X1 X2 X3 X4 X5 X6 X7 X8
             X1X3 X1X6 X2X6 X3X4 X3X5 X3X8
             X4X5 X4X7 X4X8 X5X8 X7X8 / noprint;
output out = five predicted = pred
                 stdp = stdp;
run;

* proc print data = five;
run;

proc means data = five noprint;
var pred;
output out = junk2 min = min max = max;
run;

data junk2;
set junk2;
drop _TYPE_ _FREQ_;
```

```
run;

data six;
retain temp 0;
merge five junk2;
if min ne . then temp = min;
if pred = temp;
drop temp min max;
run;

* proc print data = six;
run;

data seven;
retain temp 0;
merge five junk2;
if max ne . then temp = max;
if pred = temp;
drop temp min max;
run;

* proc print data = seven;
run;

/*                                                    */
/* Following is code for the conditional FCA.         */
/*                                                    */

data in3;
set sim1c end = eof;
keep X1 X2 X3 X4 X5 X6 X7 X8
     X1X3 X1X6 X2X6 X3X4 X3X5
     X3X8 X4X5 X4X7 X4X8 X5X8
     X7X8 reHT;
output;
if eof then do;
reHT = .;
do X1 = 1;
do X2 = -1;
do X3 = -1;
do X4 = 0;
do X5 = 1;
* do X5 = -1 to 1 by .1;
* do X6 = 1;
do X6 = -1 to 1 by 2;
do X7 = 0;
* do X8 = -1;
do X8 = -1 to 1 by .1;
X1X3 = X1*X3;
X1X6 = X1*X6;
X2X6 = X2*X6;
X3X4 = X3*X4;
```

```
X3X5 = X3*X5;
X3X8 = X3*X8;
X4X5 = X4*X5;
X4X7 = X4*X7;
X4X8 = X4*X8;
X5X8 = X5*X8;
X7X8 = X7*X8;
output;
end;
end;
end;
end;
end;
end;
end;
end;
end;
run;

* proc print data = in3;
run;

proc reg data = in3;
model reHT = X1 X2 X3 X4 X5 X6 X7 X8
             X1X3 X1X6 X2X6 X3X4 X3X5 X3X8
             X4X5 X4X7 X4X8 X5X8 X7X8 / noprint;
output out = eight predicted = pred
                   stdp = stdp;
run;

data eight;
set eight (firstobs = 73 obs = 114);
run;

* proc print data = eight;
run;

proc g3d data = eight;
plot X6*X8 = pred / rotate = 0 to 360 by 30; * zmax = 1.43 zmin = .75;
run;

proc gcontour data = eight;
plot X6*X8 = pred / levels = .8 to 1.2 by .1; * hreverse;
run;

proc means data = eight noprint;
var pred;
output out = junk3 min = min max = max;
run;

data junk3;
set junk3;
```

```
drop _TYPE_ _FREQ_;
run;

data nine;
retain temp 0;
merge eight junk3;
if min ne . then temp = min;
if pred = temp;
drop temp min max;
run;

proc print data = nine;
run;

data ten;
retain temp 0;
merge eight junk3;
if max ne . then temp = max;
if pred = temp;
drop temp min max;
run;

proc print data = ten;
run;
quit;
```

The following commented SAS program was used in the RCA for the response RE_{HT} and then subsequently edited for the response RE_{HH}.

```
options pageno = 1;
goptions reset = all device = win target = winprtm;

/*                                                              */
/* Key for variables:                                           */
/* X9 = covariate percentage within-network variability         */
/* X10 = covariate number of networks                           */
/* X11 = covariate predicted number of parents                  */
/* X12 = covariate rarity index                                 */
/*                                                              */

data sim1b;
input X9 X10 X11 X12 reHT reHH;
cards;
0.189 99 1.4 1.122 1.069 1.022
0 400 1.7 1.02 0.99 0.99
.
.
.
/* data follows */
.
.
.
0.384 863 5.3 1.918 0.877 0.844
0.302 871 5.3 1.536 1.015 0.921
;

proc means data = sim1b noprint;
output out = one range (X9 X10 X11 X12) = rX9 rX10 rX11 rX12
                 max (X9 X10 X11 X12) = maxX9 maxX10 maxX11 maxX12
                 min (X9 X10 X11 X12) = minX9 minX10 minX11 minX12;
run;

data temp;
retain rjunk1 0 rjunk2 0 rjunk3 0 rjunk4 0
       maxjunk1 0 maxjunk2 0 maxjunk3 0 maxjunk4 0
       minjunk1 0 minjunk2 0 minjunk3 0 minjunk4 0;
merge sim1b one;

if rX9 ne . then rjunk1 = rX9;
if rX10 ne . then rjunk2 = rX10;
if rX11 ne . then rjunk3 = rX11;
if rX12 ne . then rjunk4 = rX12;
if maxX9 ne . then maxjunk1 = maxX9;
if maxX10 ne . then maxjunk2 = maxX10;
if maxX11 ne . then maxjunk3 = maxX11;
if maxX12 ne . then maxjunk4 = maxX12;
```

```
if minX9 ne . then minjunk1 = minX9;
if minX10 ne . then minjunk2 = minX10;
if minX11 ne . then minjunk3 = minX11;
if minX12 ne . then minjunk4 = minX12;

if rX9 = . then rX9 = rjunk1;
if rX10 = . then rX10 = rjunk2;
if rX11 = . then rX11 = rjunk3;
if rX12 = . then rX12 = rjunk4;
if maxX9 = . then maxX9 = maxjunk1;
if maxX10 = . then maxX10 = maxjunk2;
if maxX11 = . then maxX11 = maxjunk3;
if maxX12 = . then maxX12 = maxjunk4;
if minX9 = . then minX9 = minjunk1;
if minX10 = . then minX10 = minjunk2;
if minX11 = . then minX11 = minjunk3;
if minX12 = . then minX12 = minjunk4;

drop rjunk1 rjunk2 rjunk3 rjunk4
     maxjunk1 maxjunk2 maxjunk3 maxjunk4
     minjunk1 minjunk2 minjunk3 minjunk4
     _TYPE_ _FREQ_;

midX9 = (minX9 + maxX9)/2;
midX10 = (minX10 + maxX10)/2;
midX11 = (minX11 + maxX11)/2;
midX12 = (minX12 + maxX12)/2;

X9 = (X9 - midX9)/(rX9/2);
X10 = (X10 - midX10)/(rX10/2);
X11 = (X11 - midX11)/(rX11/2);
X12 = (X12 - midX12)/(rX12/2);

drop rX9 rX10 rX11 rX12
     maxX9 maxX10 maxX11 maxX12
     minX9 minX10 minX11 minX12
     midX9 midX10 midX11 midX12;

run;

data sim1c;
set temp;
* if _n_ = 65 then delete;
* if _n_ = 69 then delete;
reHT = (1/reHT);
X9X10 = (X9*X10);
X9X11 = (X9*X11);
X9X12 = (X9*X12);
X10X11 = (X10*X11);
X10X12 = (X10*X12);
X11X12 = (X11*X12);
X9X9 = (X9*X9);
```

```
X10X10 = (X10*X10);
X11X11 = (X11*X11);
X12X12 = (X12*X12);
run;

* proc print data = sim1c;
run;

/*                                                              */
/* The following code is for the Box-Cox transformation.        */
/*                                                              */

* %adxgen;
* %adxinit;
* %adxtrans(sim1c, tsim1c, reHT, X9 X10 X11 X12
            X9X10 X9X11 X9X12 X10X11 X10X12 X11X12
            X9X9 X10X10 X11X11 X12X12);

/*                                                                    */
/* Checking for multicollinearity and reducing the number             */
/* of variables before embarking on model selection procedures.       */
/*                                                                    */

* proc corr data = sim1c;
* var X9 X10 X11 X12 X9X10 X9X11 X9X12 X10X11 X10X12 X11X12
  X9X9 X10X10 X11X11 X12X12;
run;

* proc reg data = sim1c graphics outest = pass1
*                 ridge = 0 to .6 by .05;
* model reHT = X9 X10 X11 X12
               X9X10 X9X11 X9X12 X10X11 X10X12 X11X12
               X9X9 X10X10 X11X11 X12X12 / vif;
* plot / ridgeplot nomodel nostat vref = 0;
run;

* proc print data = pass1;
run;

* proc reg data = sim1c graphics;
* model X12 = X10;
* plot X12 * X10;
run;

* proc reg data = sim1c graphics outest = pass2
*                 ridge = 0 to .6 by .05;
* model reHT = X9 X10 X11 X12
               X9X10 X9X11 X9X12 X10X11 X11X12
               X9X9 X10X10 X11X11 X12X12 / vif;
* plot / ridgeplot nomodel nostat vref = 0;
run;
```

```
* proc print data = pass2;
run;

* proc reg data = sim1c graphics outest = pass3
*                 ridge = 0 to .6 by .05;
* model reHT = X9 X10 X11 X12
               X9X10 X9X11 X9X12 X10X11
               X9X9 X10X10 X11X11 X12X12 / vif;
* plot / ridgeplot nomodel nostat vref = 0;
run;

* proc print data = pass3;
run;

* proc reg data = sim1c graphics outest = pass4
*                 ridge = 0 to .6 by .05;
* model reHT = X9 X10 X11 X12
               X9X10 X9X11 X10X11
               X9X9 X10X10 X11X11 X12X12 / vif;
* plot / ridgeplot nomodel nostat vref = 0;
run;

* proc print data = pass4;
run;

* proc reg data = sim1c graphics outest = pass1b
*                 ridge = 0 to .6 by .05;
* model reHT = X9 X10 X12
               X9X10 X9X12 X10X12
               X9X9 X10X10 X12X12 / vif;
* plot / ridgeplot nomodel nostat vref = 0;
run;

* proc print data = pass1b;
run;

* proc reg data = sim1c graphics outest = pass2b
*                 ridge = 0 to .6 by .05;
* model reHT = X9 X10 X12
               X9X10 X9X12
               X9X9 X10X10 X12X12 / vif;
* plot / ridgeplot nomodel nostat vref = 0;
run;

* proc print data = pass2b;
run;

* proc reg data = sim1c graphics outest = pass3b
*                 ridge = 0 to .6 by .05;
* model reHT = X9 X10 X12
               X9X10
               X9X9 X10X10 X12X12 / vif;
```

```
* plot / ridgeplot nomodel nostat vref = 0;
run;

* proc print data = pass3b;
run;

/*                          */
/* Doing model selection.   */
/*                          */

* proc reg data = sim1c;
* model reHT =  X9 X10 X12
                X9X10
                X9X9 X10X10 X12X12 / selection = stepwise;
run;

* proc reg data = sim1c;
* model reHT =  X9 X10 X12
                X9X10
                X9X9 X10X10 X12X12 / selection = backward;
run;

* proc reg data = sim1c graphics;
* model reHT =  X9 X10 X12
                X9X10
                X9X9 X10X10 X12X12 /
                selection = cp best = 10 adjrsq aic mse;
* plot cp.*np.;
run;

/*                                              */
/* Checking multicollinearity of final model.   */
/*                                              */

* proc glm data = sim1c;
* model reHT =  X9 X10 X12 X9X10 X9X9 / tolerance nouni;
run;

proc rsreg data = sim1c;
model reHT = X10 X12 X9X10 X9 / covar = 3;
run;

proc reg data = sim1c;
model reHT = X9 X10 X12 X9X10 X9X9 / dw influence spec cli;
output out = two rstudent = rstudent
                 residual = residual
                 predicted = pred;
run;

data two;
set two;
sabsr = sqrt(abs(residual));
```

```
run;

* proc gplot data = two;
* plot residual * pred / vref = 0;
* symbol i = sm99ps v = @ w = 1 h = 1;
run;

* proc gplot data = two;
* plot sabsr * pred / vref = 0;
* symbol i = sm99ps v = @ w = 1 h = 1;
run;

* proc univariate data = two normal;
* var residual;
run;

proc rank data = two normal = blom out = normset;
var residual;
ranks rankres;
run;

* proc gplot data = normset;
* plot residual * rankres;
run;

/*                                                    */
/* Following is code to generate 3D centerpoint       */
/* plots for all the interaction terms.               */
/*                                                    */

data in1;
set sim1c end = eof;
keep X9 X10 X12 X9X9 X9X10 reHT;
output;
if eof then do;
reHT =.;
do X9 = -1 to 1 by .05;
do X10 = 0;
do X12 = 0;
X9X9 = X9*X9;
X9X10 = X9*X10;
output;
end;
end;
end;
end;
run;

* proc print data = in1;
run;

proc rsreg data = in1 out = three;
```

```
model reHT = X10 X12 X9X10 X9 / covar = 3 predict noprint;
run;

data three;
set three;
if _N_ le 72 then delete;
run;

* proc print data = three;
run;

* proc g3d data = three;
* plot X9 * X10 = reHT / rotate = 0 to 360 by 30;
run;

* proc gplot data = three;
* plot reHT * X9;
* symbol v = @ w = 1 h = 1 c = black;
run;

/*                                                    */
/* Following is code to extract the settings that     */
/* yield the reHT minimum for a given pair of         */
/* predictor variables to verify the 3D centerpoint   */
/* plot.                                              */
/*                                                    */

proc means data = three noprint;
var reHT;
output out = junk min = min max = max;
run;

data junk;
set junk;
drop _TYPE_ _FREQ_;
run;

data four;
retain temp 0;
merge three junk;
if min ne . then temp = min;
if reHT = temp;
keep X9 X10;
run;

* proc print data = four;
run;

/*                                                    */
/* Following is code to extract the settings that     */
/* yield the reHT minimum on a global basis.          */
/*                                                    */
```

```
data in2;
set sim1c end = eof;
keep X9 X10 X12 X9X10 X9X9 reHT;
output;
if eof then do;
reHT = .;
do X9 = -1 to 1 by 2;
do X10 = -1 to 1 by 2;
do X12 = -1 to 1 by 2;
X9X10 = X9*X10;
X9X9 = X9*X9;
output;
end;
end;
end;
end;
run;

* proc print data = in2;
run;

proc reg data = in2;
model reHT = X9 X10 X12 X9X10 X9X9 / noprint;
output out = five predicted = pred stdp = stdp;
run;

* proc print data = five;
run;

proc means data = five noprint;
var pred;
output out = junk2 min = min max = max;
run;

data junk2;
set junk2;
drop _TYPE_ _FREQ_;
run;

data six;
retain temp 0;
merge five junk2;
if min ne . then temp = min;
if pred = temp;
drop temp min max;
run;

* proc print data = six;
run;

data seven;
```

```
retain temp 0;
merge five junk2;
if max ne . then temp = max;
if pred = temp;
drop temp min max;
run;

* proc print data = seven;
run;

/*                                                    */
/* Following is code for the conditional RCA.         */
/*                                                    */

data in3;
set sim1c end = eof;
keep X9 X10 X12 X9X10 X9X9 reHT;
output;
if eof then do;
reHT = .;
* do X9 = 1;
do X9 = -1 to 1 by .1;
* do X10 = -1;
do X10 = -1 to 1 by .1;
do X12 = 1;
* do X12 = -1 to 1 by .1;
X9X10 = X9*X10;
X9X9 = X9*X9;
output;
end;
end;
end;
end;
run;

* proc print data = in3;
run;

proc reg data = in3;
model reHT = X9 X10 X12 X9X10 X9X9 / noprint;
output out = eight predicted = pred stdp = stdp;
run;

* proc print data = eight;
run;

data eight;
set eight (firstobs = 73 obs = 513);
run;

proc g3d data = eight;
plot X9 * X10 = pred / rotate = 0 to 360 by 30;
```

```
run;

proc gcontour data = eight;
plot X9 * X10 = pred / levels = 0 to 3 by .5;
run;
quit;
```

APPENDIX C

Example Populations

The actual data for the 10 x 10 nominal population (Figure 2).

0	0	21	31	0	0	0	0	0	0
0	0	0	0	0	0	8	44	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
38	2	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	54	0	0	0	0	0	0	0	0

The actual data for the 10 x 10 extreme 1 population (Figure 3).

13	14	12	4	14	8	0	1	8	14
3	11	9	0	7	5	0	0	14	15
2	1	1	0	1	2	0	0	7	23
4	1	0	0	0	1	0	1	7	8
8	1	0	1	0	0	0	0	1	1
11	1	5	9	2	0	0	0	0	0
8	5	18	14	4	2	1	0	0	0
1	8	12	9	3	12	2	0	0	0
0	1	5	3	3	17	4	0	0	0
0	0	0	1	9	25	1	0	0	0

The actual data for the 30 x 30 extreme 2 population (Figure 4).

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0