STATISTICS IN TRANSITION-new series, December 2010
Vol. 11, No. 3, pp. 539—554

MODELS IN SURVEY SAMPLING

Carl-Erik Särndal[1]

ABSTRACT

Models, especially in the form of assumed relationships between study variables and auxiliary variables, have influenced survey sampling theory and practice over the last four decades. Some of the early debates between the design-based school and the model-based school are revisited. In their pure forms, they offer two fundamentally different outlooks and approaches to inference in sample surveys. Complete reconciliation and agreement cannot be expected. But the tendency today is that each of the two approaches recognizes and profits from important elements in the other. We see an often fruitful interaction, as discussed in this article.

1. A polarization occurs

The objective in this article is to reflect on a topic that has received much attention, in discussions at conferences and seminars and, more formally, in articles and books. I am referring to the role of models in survey sampling theory and practice. My remarks do not attempt to paint a complete picture. They represent a few personal impressions and conclusions about a development in which I participated. Many important developments go unmentioned in the text that follows.

Behind the topic (in its modern aspect) lies a roughly forty year old split, or scientific conflict if one prefers. To assign a time span of forty years is of course rather arbitrary, but not without some good justification.

Among opinions that have been expressed: Why should survey samplers be different? Just about every other branch of statistics is built around modeling, why should survey samplers resist? Those who persist in the established tradition of survey sampling (the design-based framework), are they not keeping in step with the times? But recent years have brought a change: models are now well engrained in the design-based philosophy and practice as well.

Thus, a polarization occurred around four decades ago: design-based inference became contrasted with model-based (or model dependent) inference. These terms were not in common use before 1970. Today they are standard usage,

[1] Statistics Sweden, S-70189 Örebro, Sweden; carl.sarndal@scb.se.

not only among specialists in survey sampling, but among other categories of scientists as well. After all these years, the issue is not settled; neither side is a winner.

Once an awareness had been created, and the basic differences between the two approaches had been made clear, many were attracted to the topic. There was a period of intellectual curiosity. What are, more precisely, the differences? Which estimators are favored (have better accuracy) under one or the other approach? Some took a categorical stand in favour of one approach, convinced that the other was wrong. Others were more neutral, content to understand and appreciate each approach for what it is, without taking any demonstrative position in favour of one or the other.

The early debates in the 1970's and 1980's took place for the most part in the arena of those "pure survey conditions" that I mention later. The survey practitioner, who seldom encounters pure conditions, feels uneasy at times about the incapacity of the standard theory (the design-based one) to address all the contingencies – nonresponse and many others – that affect surveys today.

Much has been written on the concept of models, and on their function, by authors with a philosophy of science perspective on economics, sociology or other disciplines. Serious model building is a maturing process, involving, over a long time, a thorough interplay between a theoretical content and empirical content. But in survey sampling, "the model" is often "a default statistical model", a specification, suitable for the moment, of a formula for a hypothetical relationship between a study variable y (one of the usually many in a survey) and those other variables called auxiliary, because something is known about them at a level beyond the sample itself. When the survey methodologist says "should we model it", he/she is wondering whether to bring in some assumptions of relationship, unverifiable but not entirely out of place, to save survey resources or to bypass other practical difficulties.

2. The early debates

Two papers that made an impression on me as a young man, and on many others, were Brewer (1963) and Royall (1970). The first is titled *Ratio estimation and finite population: some results deductible from the assumption of an underlying stochastic process*. The second is titled *On finite population sampling under certain linear regression models*. (I cite full titles, because they in themselves convey a message.) While the second is traditionally cited, I personally assign much value also to Brewer's 1963 contribution, a feeling underscored also by Brewer's relentless and insightful efforts later to combine, and to seek the positive side of, both approaches, as I discuss later.

In those articles, Brewer and Royall examine a basic situation: The estimation of a population total $Y = \sum_U y_k$, where y_k is observed for all units $k \in s$, where s is a probability sample from $U = \{1, 2,..., N\}$, and where population totals are known for one or more auxiliary variables. (I focus in this article on the estimation of finite population parameters called descriptive, such as totals, means and functions of totals. Inference about super-population parameters, called analytic, is a different story. Also, I do not address issues arising in the important area of longitudinal surveys.)

The design-based approach (although not referred to by that name until later) had become the ruling methodology, following Neyman (1934) and contributions in its footsteps during the 1940's and 1950's. In this approach, the probability structure for inferences comes from the randomization distribution, from the probabilities with which different samples are potentially drawn (although one and only one is realized in a survey). The statistical properties (mean, variance and so on) of an estimate are evaluated by averaging over all possible samples under the given sampling design. It is an unconditional distribution.

By contrast, Royall (1970) and his followers present a serious attempt to construct inferences from an alternative source, the model alone, conditionally on the set, s, of units sampled from U in some way, not necessarily by probability sampling. Although Royall did not use the term, "model-based inference" quickly became a recognized concept and part of the standard terminology. This approach revived the old controversy dating back to the beginning of the 20th century. Forms of purposive (non-probability) sampling and balanced sampling had already been practiced in the 1890's; the Norwegian experience attributed to Kiaer is frequently cited.

3. The literature in a context of pure conditions

A sometimes heated debate took place 1970-1990, principally among sampling theoreticians. The arena was one of "pure conditions", or one may call it "debate on the foundations".

Pure conditions address (with minor variations) the following situation: A probability sample s is drawn from the finite population $U = \{1, 2,..., k,..., N\}$. The known design weight for unit k is $d_k = 1/\pi_k$, where $\pi_k = \Pr(k \in s) > 0$ is the inclusion probability (whether to use the π_k or not in inference is a divisive question). The value y_k of the study variable y is recorded for all $k \in s$. The objective is to estimate the population total $Y = \sum_U y_k$. Auxiliary information will be used, usually so that $\mathbf{x}_k$ is an auxiliary vector value known for $k \in U$ (or, at a minimum, so that the total $\sum_U \mathbf{x}_k$ is known, imported from an accurate source). Nonresponse, measurement error, frame error and other non-sampling errors are absent. There is only *one* study variable y, although in practice most surveys have many. No particular country, no particular survey is addressed; we are concerned with bare basics, the foundations.

Following the early debates in the 1970's, many articles have been written on some aspect of estimation under pure conditions. They are valuable contributions to the literature, especially in the light of recent advances such as multilevel modeling, non-parametric regression modeling and others. But pure conditions have little to do with today's harsh survey conditions with high nonresponse, frame errors and other imperfections. Still it is appropriate today to place a piece of research within the context of the pure conditions, and develop one's topic within either of the two paradigms, the design-based one or the model-based (model dependent) one, and to get one's theoretical article accepted in the best of journals.

In the decades following 1970 one took great interest in comparing estimators generated by one or the other approach, and, more importantly, in comparing the properties (bias, variance, and so on) under one or the other of the two frameworks. As a typical example of that period, I was concerned in Särndal (1978), *Design-based and model-based inference in survey sampling*, with exploring the possibilities under the two modes. I was not alone in this type of endeavor.

4. Passionate stands, and modeling as an act of taking responsibility

In their article *An evaluation of model-dependent and probability sampling inferences in sample surveys*, Hansen, Madow and Tepping (1983) took a spirited stand in favour of probability sampling inference (that is, design-based inference), as opposed to a model-dependent (model-based) inference. It was an attack perhaps not so much on the use of models as a concern about the lack of robustness of the model-based estimation. (Hansen and collaborators were not opposed to the idea of "giving models their just and appropriate place"; they had shown the rich possibilities of model oriented reasoning in their work on total survey error models.) As a discussant of that paper, Royall passionately defended his model-based view.

An influential participant in the early debates, Smith (1976a) also strongly favoured the model-based outlook: "The basic question to ask is why should finite population inference be different from inferences made in the rest of statistics? … My view is that survey statisticians should accept their responsibility for providing stochastic models for finite populations in the same way as statisticians in the experimental sciences". Later, Smith's position was to change dramatically, no doubt a result of careful inspection.

As Brewer (1999) notes, "Royall argued that survey sampling was out of step with statistics as a whole. Statisticians working in other fields used their data to build models and analyzed them in those terms … but survey statisticians were using an entirely irrelevant source of probability structure not related to the data themselves but only to the manner in which they had been collected."

The thought that design-based survey statisticians might fail to "take responsibility" for their design-based estimates is somewhat of an antithesis. Morris Hansen and other influential proponents of the design-based mode, from the 1940's and on, were the opposite of "not responsible", conscientious as they were to provide policy makers and authorities in government with impartial, defendable statistics. In fact, why should survey statisticians not be different from the experimentalists? Experiments are typically small, with relatively few units undergoing treatment, and focus is on hypothesis testing of treatment differences. By contrast, government surveys are large, multipurpose, and they cater to a different type of user.

5. Is reconciliation possible?

Now, if indeed differences exist, are they of crucial importance? Did the debate exaggerate them? Can the two approaches be reconciled? The verb "to reconcile" means both (a) to return to friendly relations and harmony, and (b) to make consistent or congruous or compatible. Both meanings are relevant here. Smith (1994) titles his Hansen lecture *Sample surveys 1975-1990; An age of reconciliation*? A hint of "not reconciled but coming fairly close" lies in the title of Brewer, Hanif and Tam (1988), *How nearly can model-based prediction and design-based estimation be reconciled*?

Smith (1994) notes "In the absence of models for the underlying social processes which are generally held to be true, model-based inferences lose all their desirable properties." A discussant at that occasion also, Royall (1994) persisted: "Professor Smith's paper … is an announcement of a dramatic change in his own thinking. … I will simply try to show that there *must* be errors in his reasoning, because his conclusion is wrong … He goes to the shocking extreme of advising … what is clearly *wrong*, namely quoting the unconditional standard error."

The underlying statistical principles differ fundamentally; to make the two thought processes consistent and compatible is simply not possible: Design-based inference is "unconditional, referring to all possible probability samples under the given sampling design"; model-based inference "conditional, under the model, for the one and only realized sample". That is an irreconcilable difference. Now, in a given situation, under comparable conditions, the estimators delivered by the two approaches may agree. But the variances and the mean squared errors do not generally agree. They refer to different conceptions of "long run repetitions", over all possible samples in the design-based case, over all finite populations generated under the model in the other case. Nevertheless one can, in some situations at least, take action to also make measures of precision acceptable from both vantage points, as for example Särndal, Swensson and Wretman (1989) show in

The weighted residual technique for estimating the variance of the general regression estimator of the finite population total.

Not surprisingly, the literature has examined the compromise of "let us average over both", that is, over both the distribution determined by the assumed model and the distribution determined by the randomized sample selection. One is led to consider the concept of anticipated variance (model-expected design-expected mean squared error).

Brewer in particular has contributed much to a vision of making the inferences palatable from both angles, the design-based one and the model-based one; Brewer (1999) argues "there is something substantial to be gained by using them in combination", in a single estimator defendable from both sides. This theme is also developed in Brewer (1995), *Combining design-based and model-based inference* and in his book, Brewer (2002), *Combined survey sampling inference; weighing Basu's elephants*.

The survey sampling literature has thus come to harbour two streams with fundamentally different starting points. But how does an individual statistician react? Must he/she choose sides? Can a self-respecting survey sampling statistician embrace both lines of thought, defending sometimes one, sometimes the other approach? I have observed colleagues and friends adopt different attitudes in regard to those questions. One option is to switch sides, depending on the practical problem at hand, as for example when one defends model-dependent inference in regard to the small area problem, while advocating, in other instances, a design-based approach. As for myself, I experience some difficulty with subscribing to a "double-natured ethic". An adaptable attitude is however suggested in Smith's (1994) remark: "My overall conclusion is that there is no single paradigm for statistical inference and that different classes of problems require different solutions. Instead of looking for unity we should concentrate on identifying the differences and enjoy the diversity of our subject. Complete reconciliation is neither possible nor desirable. Vive la différence." But Smith's own preference is stated in these words: "The case I am making is for procedural inference, and this refers to the unconditional randomization distribution. I now find the case for hard-line randomization inference based on the unconditional distribution to be acceptable … I now think that the framework for descriptive inference should be the unconditional distribution relating to the original sampling procedure." This corresponds well with my own thinking since the early 1980's.

6. Awareness not without risk of confusion.

The opposition design-based vs. model based may be perfectly clear and simple to experts in survey sampling, those who have regularly monitored the theoretical literature over the past 30 years. But have the discussions led by survey theoreticians created undesirable confusion in other fields, among scientists who rely on statistics and understand a great deal about survey design and estimation, without being specialists? Some signs suggest that this has indeed occurred. Design-based inference is unconditional, model-based inference is conditional, that sounds simple, but explaining the implications to specialists in applied fields may not be a trivial task. In any case, it is interesting to note that the debate on the foundations among survey sampling specialists has reached beyond the science of statistics itself.

Forest research is an area where sampling has a long tradition, for example for the estimation of tree volume. A thorough review destined to researchers in that discipline is Gregoire (1998), *Design-based and model-based inference in survey sampling: appreciating the difference.* Although the article is essentially on the side of the design-based tradition, it is noted that recent times have led to a debate also in that discipline: "… current literature in forestry and ecology indicate much ongoing confusion about the distinction between these two modes of inference. A failure to appreciate the underpinnings of one or the other mode of inference could lead to needless abandonment of a survey design that might otherwise be ideal for purposes of scientific inquiry."

Other examples of "uncertainty and confusion" could be mentioned. From the field of neuromorphology comes an article by Geuna (2000) titled *Appreciating the difference between design-based and model-based sampling strategies in quantitative morphology of the nervous system*. This author states "New technical procedures have been devised and applied successfully to neuromorphological research. However … one element of confusion is related to uncertainty about the meaning, implications, and advantages of the design-based sampling strategy that characterize the new techniques." In this scientific discipline, the design-based alternative appears to be the novelty, a challenger to an earlier established model-based view. This confirms the impression that in any given science, one of the two views is the well entrenched one, the other being an intruder that must fight a battle to be recognized.

To explain (to readers other than the experts) the basic difference between the unconditional (design-based) and the conditional (model-based) mode may be comparatively easy. However, the task of explaining is further complicated by the rise in the last 25 years of mixture modes: Design-based inference can be (and is usually) model assisted, while, on the other side, the model-based inference can or should account for the randomized sample selection. The next few sections examine those proliferations.

7. They cannot live without one another

How did models influence the classical design-based survey statisticians, those who built the design-based tradition, by books and influential articles, in the 1940's to 1960's? By all signs, models had a place in their thinking, but they were "not explicit". These writers did not abhor models, but apparently they abstained, except in rare occasions, from stating them, as if there was no need to be explicit. For example, models were used in a variety of ways to design samples, in modeling (or, less pretentiously, just guessing) stratum variances to determine strata sampling fractions.

Now today, given that the two fundamentally different lines of thought exist, can the proponents of the design-based vision proceed effectively without reference to models? Can proponents of the model-based approach work without any reference to the features of the randomized sample selection? The answer is "no" in both cases. It is very hard to maintain that "elements of the other" should not be present; only the most "hard-core defendants" (if they still exist at this time) of either approach would pretend otherwise.

In each approach, the need is felt to "account for the other", to integrate aspects of the other. For design-based people, the challenge was, from the 1980's and on, to make models explicit in the formal presentation. For the model-based people the challenge was, and still is, to make the formal presentation (notably the model statements) reflect and incorporate the randomized sample selection.

Brewer (1999) describes how design-based theory underwent a change from implicit to explicit usage of models: "The ratio estimator … provides a good example of the way in which models of the population were long used in an implicit fashion by design-oriented survey statisticians. The essential difference between ordinary design-based inference and model assisted survey sampling is that the latter brings such implicit assumptions into the open … Early examples may be found in Cochran (1953, 1963, 1977), Brewer (1963) and Foreman and Brewer (1971), but it was definitely established as the dominant version of the design-based approach with the publication of Särndal, Swensson and Wretman (1992)." It was apparent to my co-authors and me, in the mid-1980's when essential parts of *Model Assisted Survey Sampling* were written, that "the marriage had to take place". We opted for the alternative "design-based assisted by models". A broader perspective opened up for models inside the design-based framework; "the dominant version" it became, perhaps, but I like to see it as a "necessary version", considering how those of us in design-based sampling theory experienced the field in the 1980's.

Another example of design-based accommodation to model features becomes apparent when we look at the sample selection stage. The appealing idea of balanced sampling resided in the tradition of purposive, non-randomized selection, that is, the idea that sample means of auxiliary variables should, roughly at least, equal their known population counterparts. The challenge is: Find a proper probability sampling device (thus placed within randomization theory) that is certain to deliver a balanced sample. An answer is the cube method of Deville and Tillé (2004); by ingenious randomization, we select only among samples guaranteed to be (almost) balanced; the randomization device seeks only among "good, balanced samples" and selects one of those.

So it was almost by necessity that the scientific evolution brought, as one possibility, the model assisted design-based perspective and, as a second possibility, a model-based (model dependent) design-influenced perspective. The latter calls for integrating the features of the randomized sample selection (stratified sampling, two-stage sampling and so on) in the model statement. For example, Smith (1976b), in his model-based period, challenges the design-based model assisted GREG (for generalized regression estimator) in Cassel, Särndal and Wretman (1976): "Why should the selection probabilities, p, take any precedence over the model ξ? … The design p is at the choice of the statistician and would usually be based on prior information about the population which should already be embodied in ξ".

The second possibility is also explicit in Kott (2005), *Randomization-assisted model-based survey sampling*. His position, "long espoused in public", is that "the dominant model-assisted (randomization-based) survey sampling paradigm, although fruitful in many ways, should be supplanted by a randomization-assisted model-based one. That is because inference should be based on the sample actually observed rather than averaged over all potential samples."

Also in a model-based vein, Little (2003) states, in *To model or not to model? Competing modes of inference for finite population sampling*: "Models need to properly reflect features of the sample design such as weighing, stratification and clustering, or (model-based) inferences are likely to be distorted" . He points out that models of high complexity can now be entertained: "Computational power has expanded dramatically since the days of early model versus randomization debates, and much can be accomplished using software for mixed models in the major statistical packages." But needless to say, these advanced models are now also incorporated in the design-based model assisted framework.

8. Models as assistants

The model assisted design-based approach is thus a means of bringing model features into the open without upsetting the design-based basis for inference. The characteristics of the model are not crucial to the validity of the design-based inferences.

In the randomization/model marriage that produced the model assisted outlook, the models play an obedient role. A relationship between y and $\mathbf{x}$ is taken into account, but "its truth" is not essential; it remains in the background, lacking the influential role it would have in model-based inference.

Science thrives by arguments and counterarguments. As one can expect, criticism has been levied of this seemingly subservient role of the model: The design-based inferences are recognized as "robust", but may be less efficient than they could be; it is argued that the full potential of "a correct model" is not realized, that the model denied its justified, more influential role, and so on.

Smith (1994) points out that "randomizers should not make concessions towards predictive inference". The advent of the model assisted design-based approach granted randomizers this freedom. Nevertheless, this approach still resides essentially within “pure conditions”. It requires, ideally at least, an absence of nonresponse and other non-sampling errors. I return later to the question of survey nonresponse.

9. Looking beyond the pure conditions: The practice of survey sampling

Three articles by Kalton (1981, 1983, 2002) are titled *Models in the practice of survey sampling*, the third one with the addition “revisited”. I examine these contributions with particular interest; they are highly relevant in a discussion on the role of models in survey sampling, because they examine the confrontation of theory with thorny practical matters, leaving aside the theoretician’s predilection for “pure conditions” and “foundations”.

Kalton (2002) notes that: “Models are widely used within the design-based mode of inference, both in sample design and in estimation, but in a “model-assisted” manner so that the validity of the survey estimates does not depend on the validity of the model assumptions.” However, as this author goes on to say, “Valid design-based inferences require that nonresponse and other non-sampling errors be of next-to-negligible extent … The design-based approach … for descriptive analysis of large scale surveys … cannot fully address all problems of making inferences … Although design-based inference is the standard form of inference with large-scale sample surveys, in practice some reliance on model-dependent inference is necessary.” These phrases reflect a regret that the standard theory (the design-based one), although preferred and held in high regard by practitioners, does not provide answers for all circumstances. The words “cannot fully assess address all the problems” and “some reliance on models” are crucial. What role, more precisely, should then be attributed to models?

10. Models as crutches

Kalton (2002) states: “My general approach to the use of model-dependent methods for descriptive estimation is to treat the model as a crutch, to be used only to the extent that the survey data cannot fully support the desired estimates. If the sample is strong enough, and if there is no weakness from missing data, then design-based inferences alone will serve well.”

“To treat the model as a crutch” is a colorful image. It is not entirely misplaced to view the design-based theory as a handicapped guide for practice, one that is in need of support from artificial devices (models) to meet the needs. A certain gap exists between the practical reality and the pure design-based theory that is supposed to back up that practice.

This suggests an approach, rather informal and unstructured, in which models are brought to bear, whenever needed to supplement or mend the ailing theory. On the other hand, it might imply a somewhat uncontrolled “model interference” to overcome survey imperfections. Also, the pure concepts of design-based unbiasedness and design-based variance become compromised. For example, what value is there in computing the design-based variance of survey estimates when the unknown squared bias (arising for example when missing data are imputed) is likely to be the dominating component of the mean squared error?

The weakness "sample not strong enough" is invoked in practically all the literature on model dependent small area estimation, where it breeds the idea of "borrowing strength" from data outside the domain, something which necessarily brings a presence of models in the construction of a small area estimator.

11. Critical areas: Small area estimation and nonresponse

In design-based practice, the predicament of "sample not strong enough" occurs often. Kalton (2002) continues: "… many of the developments in survey sampling in the past quarter century have been concerned with the application of model-dependent methods to address such problems as missing data and small area estimation". To my mind also, missing data and small area estimation, have clearly stood out (and still do) as two significant and critical areas (or testing grounds) for the reliance on models in survey sampling. They do so for different reasons.

Design-based inference, the standard in practice, grew out of a theory for essentially pure conditions; it "cannot fully address" estimation with missing data or for small areas. The reasons for failure are different in the two cases: For very small areas or sub-populations, the total sample size is insufficient to deliver acceptable design-based precision. For the missing data (or nonresponse) problem, it is the unknown response mechanism and the unknown response probabilities that are the root of the problem.

In both cases, taking measures to make the standard theory functional is viewed not as a scientific necessity, but rather as being too expensive. Increasing the response rate, by follow-ups and other measures, to reach "decent levels of response" is deemed prohibitively expensive, as is an increase in total sample size up to a point where well-supported design-based estimates become possible not only for large sub-populations but also for small ones. But the statistician is expected to cope with the poor prospects.

Kalton (2002) notes: "From the design-based perspective, the approach to estimation for small areas is first to seek direct estimators of adequate precision, making full use of auxiliary information in a model-assisted way. When that approach fails, it becomes necessary to resort to indirect estimators that depend on statistical models." The approach fails, because in our time pressure is imminent to produce numbers on any tiny group in society, without a willingness to supply the funds needed to do that with adequate precision. Hence, model-dependent small area estimation theory becomes one of the crutches, to use that colorful term; that theory drops from the start any ambition to build inferences on the randomization distribution.

The missing data syndrome is one from which practically all surveys suffer today. Nonresponse rates are very high and all the time increasing. The dominating line of thought is to mend, or repair, the design-based inferences through imputation and/or nonresponse weighting adjustment. Both are viewed as forms of "model intervention". I return to the nonresponse question in Section 13.

12. The calibration perspective

Another angle of the questions of "models or not?" and "models to what extent?" becomes apparent with the increased popularity in the last fifteen years of calibration theory and practice. The procedure of attaching weights to the observed values of a study variable, and to sum the weighted values, has always appealed to survey samplers, especially the practitioners. The idea to make the weight system respect "control totals" is also old. Weighting by poststratification is among the simplest and oldest examples of calibration. Deming (1943), in the book *Statistical Adjustment of Data*, focused on making the weights conform to marginal counts of cross-classified tables; this became known as the raking ratio method. Long tradition and expertise in "controlled weighting" are evident at the US Bureau of the Census, as seen for example in Alexander (1987), and in other important survey organizations, such as I.N.S.E.E. in France. The modern term is *calibration* (on auxiliary information); Deville and Särndal (1992), in *Calibration estimators in survey sampling*, contributed a more general and more complete view of that methodology.

To survey practitioners, the prime motivation for calibration is not as much the potential for increased accuracy as rather the desire to "achieve consistency", within a design-based perspective, with known or estimated statistics from other sources or survey occasions. In that frame of mind, a modeling of y-to-$\mathbf{x}$ relationships is secondary. Nevertheless, the aims of calibration are twofold: (i) realizing consistency and (ii) improving accuracy, reducing both bias (due to nonresponse for example) and variance.

The practitioner understands perfectly well the motivation behind calibration and its results. The theoretician on the other hand is looking for models behind the procedure and when he/she has difficulty identifying them, a natural reaction is that "models are there somehow, but they are skillfully hidden". Thus estimation by calibration brings the question of "the role of models" to a critical test.

The model assisted design-based GREG estimator of the 1980's was an offspring of the idea that a relationship between a study variable y and the auxiliary vector $\mathbf{x}$ can be exploited to improve precision; that was the prime motivation. It was my view, too, and I arrived at calibration through the back door, so to speak. I recall my surprise, almost stupefaction, at my insight in the early 1980's that the model assisted design-based GREG has an expression as a weighted sum, $\sum_s w_k y_k$, with a weight system that has the intriguing property $\sum_s w_k \mathbf{x}_k = \mathbf{X}$, stating that the weights w_k for $k \in s$ are consistent with the auxiliary information $\mathbf{X}$ for the survey, where for example $\mathbf{X} = \sum_U \mathbf{x}_k$, a known vector total for the population. Today this calibration property is a self-evident and trivial consequence. I was surprised because that property seemed so remote from, or incongruent with, my vision at the time of the relationship y-to-$\mathbf{x}$ (and the prospect of a reduced variance) as the sole driving force. But one's vision can change, as exemplified earlier in this article.

I find the essence of calibration theory particularly striking when applied to estimation for surveys with nonresponse. There the question of "use of models or not" receives a different twist, contrasting with the common view, noted earlier, that nonresponse necessitates a reliance on modeling, notably a modeling of the unknown response mechanism.

13. Calibration for nonresponse adjustment

A probability sample s is drawn from the population $U = \{1, 2, ..., k, ..., N\}$. The known design weight is $d_k = 1/\pi_k$. Nonresponse occurs; the study variable value y_k is observed only for the response set r, where $r \subset s \subset U$. How r was generated is unknown. With suitably specified weights w_k, an estimator $\hat{Y} = \sum_r w_k y_k$ of $Y = \sum_U y_k$ is computed on the data y_k observed only for $k \in r$. In particular, we can use calibrated weights, made to satisfy $\sum_r w_k \mathbf{x}_k = \mathbf{X}$, where $\mathbf{X}$ summarizes the appropriate information for the auxiliary vector $\mathbf{x}_k$; such weights are $w_k = d_k \{1 + (\mathbf{X} - \sum_r d_k \mathbf{x}_k)'(\sum_r d_k \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k\}$. As presented in for example Särndal and Lundström (2005), *Estimation in surveys with nonresponse*, the information may be of two kinds: At the population level, transmitted by the vector $\mathbf{x}_k^*$, and at the sample level, transmitted by $\mathbf{x}_k^\circ$, both vector values known for all $k \in s$, that is, for respondents and for nonrespondents. The population total $\sum_U \mathbf{x}_k^*$ is known, consisting for example of (updated) census counts on groups based on age, sex and region. The total $\sum_U \mathbf{x}_k^\circ$ is unknown but is estimated without bias by $\sum_s d_k \mathbf{x}_k^\circ$, where $\mathbf{x}_k^\circ$ may express features of the data collection and other survey operations. Then we compute the calibrated weights so that $\sum_r w_k \mathbf{x}_k = \mathbf{X}$ holds with $\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_k^* \\ \mathbf{x}_k^\circ \end{pmatrix}$ and $\mathbf{X} = \begin{pmatrix} \sum_U \mathbf{x}_k^* \\ \sum_s d_k \mathbf{x}_k^\circ \end{pmatrix}$, implying $\sum_r w_k \mathbf{x}_k^* = \sum_U \mathbf{x}_k^*$ and $\sum_r w_k \mathbf{x}_k^\circ = \sum_s d_k \mathbf{x}_k^\circ$. In this procedure $\mathbf{X}$ represents the "clearly stated information" that one decides to use. That clear message is not enough for the modeler. He/she deplores a lack of a "clearly stated model assumptions", and may conclude that "calibration is a black box". In this and other applications, the calibration approach, with its emphasis on the information on which to calibrate rather than on modeled relationships, is received with suspicion by some.

14. Conclusion

The design-based perspective on survey sampling was, at least in its earlier years, hailed as "the scientific approach" to inquiries about human or other finite populations. The period 1940 to 1960 was a break-through, the heyday of the clean, trustworthy design-based theory. Today, judicious survey statisticians feel, with some justification, that survey sampling can be redeemed as a scientific field for future generations only by making it fully appreciated by "the other branches of statistics". There is a belief that this must necessarily happen, through a use of ever more advanced models. I am not convinced. Advances in modeling have taken place in the past few decades. Still, my impression is that while the models serve a useful purpose in the design-based model assisted theory, they must become much stronger to be "good enough" for a model-based theory, stronger not only in their mathematical formulation, but also in their capacity to grasp the real nature (the underlying social processes) of a relationship y-to-$\mathbf{x}$, and this not only for one isolated y-variable, but for every one of those many that partake in a large survey.

REFERENCES

ALEXANDER, C.H. (1987). A class of methods for using person controls in household weighting. *Survey Methodology*, 13, 183–188.

BREWER, K.R.W. (1963). Ratio estimation and finite population: some results deductible from the assumption of an underlying stochastic process. *Australian Journal of Statistics*, 5, 93–105.

BREWER, K.R.W. (1995). Combining design-based and model-based inference. In Cox, B.G. et al. (editors), *Business Survey Methods*. New York: Wiley, 589–606.

BREWER, K.R.W. (1999). Design-based or prediction inference? Stratified random vs stratified balanced sampling. *International Statistical Review*, 67, 35–47.

BREWER, K.R.W. (2002) *Combined Survey Sampling Inference.Weighing Basu's Elephants*. London: Arnold.

BREWER, K.R.W., HANIF, M. and TAM, S.M. (1988). How nearly can model-based prediction and design-based estimation be reconciled? *Journal of the American Statistical Association* 83, 128–32.

CASSEL, C.M, SÄRNDAL, C.E. and WRETMAN, J.H. (1976) Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63**, 615–620.

COCHRAN, W.G. (1953, 1963, 1977). *Sampling Techniques*. New York: Wiley; 1st, 2nd, 3rd edition.

DEMING, W.E. (1943). *Statistical Adjustment of Data*. New York: Wiley.

DEVILLE, J.C. and SÄRNDAL, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

DEVILLE, J.C. and TILLÉ, Y. (2004). Efficient balanced sampling: The Cube Method. *Biometrika*, 91, 893–912.

FOREMAN, E.K. and BREWER, K.R.W. (1971). The efficient use of supplementary information in standard sampling procedures. *Journal of the Royal Statistical Society, series B*, 33, 391–400.

GEUNA, S. (2000). Appreciating the difference between design-based and model-based sampling strategies in quantitative morphology of the nervous system. *The Journal of Comparative Neurology,* 427, 333–339.

GREGOIRE, T.G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28, 1429–1447.

HANSEN, M.H., MADOW, W.G. and TEPPING, B.J. (1983). An evaluation of model-dependent an probability-sampling inferences in sample surveys (with discussion). *Journal of the American Statistical Association*, 78, 776-807.

KALTON, G. (1981). Models in the practice of survey sampling. Invited paper, 43rd session of the International Statistical Institute, Buenos Aires.

KALTON, G. (1983). Models in the practice of survey sampling. *International Statistical Review*, 51, 175–188.

KALTON, G. (2002). Models in the practice of survey sampling (revisited). *Journal of Official Statistics*, 18, 129–154.

KOTT, P.S. (2005). Randomization-assisted model-based survey sampling. *Journal of Statistical Planning and Inference*, 129, 263–277.

LITTLE, R.J. (2003). To model or not to model? Competing modes of inference for finite population sampling. University of Michigan, Department of Biostatistics, working paper 2003:4.

NEYMAN, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-625.

ROYALL, R.M. (1970). On finite population sampling under certain linear regression models. *Biometrika*, 57, 377–387.

ROYALL, R.M. (1994). Discussion of Smith (1994). *International Statistical Review*, 62, 19–20.

SÄRNDAL, C.E. (1978). Design-based and model-based inference in survey sampling. *Scandinavian Journal of Statistics*, 5, 27–52.

SÄRNDAL, C.E., SWENSSON, B. and WRETMAN, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76, 527–537.

SÄRNDAL, C.E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SÄRNDAL, C.E. AND LUNDSTRÖM, S. (2005). *Estimation in surveys with nonresponse*. New York: Wiley.

SMITH, T.M.F. (1976a). The foundations of survey sampling: a review. *Journal of the Royal Statistical Society A*, 139, 183–204.

SMITH, T.M.F. (1976b). Discussion of Cassel, Särndal and Wretman (1976). *Biometrika* **63**, 620.

SMITH, T.M.F. (1994). Sample surveys 1975–1990; An age of reconciliation? *International Statistical Review*, 62, 5–19.