Sample Source Auditors

CONSISTENCY ANALYSIS AND PANEL BLENDING

Mktg Inc.
200 Carleton Avenue, East Islip, NY 11730
(631) 277-7000
www.mktginc.com

CONSISTENCY ANALYSIS AND PANEL BLENDING

Abstract

In recent years online panels and other online sample sources have multiplied, grown and morphed, with new sources born and old sources merged at a breathtaking rate. In the meantime, strikingly little has been written about the variability between sources and within sources over time. The impact of this variability on critical survey data has been almost completely ignored. Sample differences appear not just across panels, but within the same panels over time. Blending of panels has been proposed as a means to help stabilize the underlying samples. The dilemma is that sample purchasers rarely have access to sufficient information to evaluate the samples sufficiently, let alone how to blend them. Until now, too little analysis has been done and too few tools have been developed to provide guidance to researchers who wish to properly understand and manage online sample source bias. We develop the use of optimization theory to provide stable sample blends. We propose and analyze possible optimization targets and the means for achieving such blends.

This paper is an attempt to bring clarity to the online sample universe and provide an anchor for quality measurement. We do this by using combinations of quantifiable metrics, segmentation, and comparisons to the *Grand Mean* of the sample universe. Our research indicates we can make sensible comparisons and, by virtue of structural segmentations, capture the nature of the underlying sampling population. We can compare panels to one another, evaluate them through time (*Consistency Analysis*) and make statements about their consistency, predictability, and reliability.

We cannot stabilize online data unless we create reference points as anchors. Our goal is to use segmentation to create a fingerprint that can be consistently maintained by blending panels. To define these references, we have launched a data collection effort in forty countries around the globe. In each country, we have been collecting data from all willing data sources using a standard questionnaire. Metrics were standardized and independent. The reference or basis of comparison for all metrics in these tests was the average value across all samples in a region. This we define as the *Grand Mean*.

Three segmentation schemes are used in this evaluation focusing on buyer behavior, sociographic factors, and media use factors. While statistical cluster analysis methods are very robust and almost any variables set could be used to identify segments; not all will meet the stringent requirements needed for these applications. As such, this process was iterative where groups of variables are tested until a satisfactory set was identified.

The optimum mix of sample sources is determined by varying the weight of each of the panels in a set. The optimum is the point where the disparity of the distribution of segments from the *Grand Mean* is at a minimum. Optimum panel mixes are not created equal. The choice of panels for an optimum mix must therefore depend on the results and other conditions. These include not only the impact on optimizing on one set of segments, but potential impact on other segmentation schemes.

It is the measure of sampling frames consisting of the online panels and data-sources through time that we believe is critical to establish a consistent, predictable and reliable sample frame.

CONSISTENCY ANALYSIS AND PANEL BLENDING

Introduction

In recent years online panels and other online sample sources have multiplied, grown and morphed, with new sources born and old sources merged at a breathtaking rate. While little has been written about the effects each online sample source and combinations have on critical survey data, there have been indications of significant difference existing between sample sources. At the 2006 ESOMAR World Research Conference in Barcelona, De Gaudemar spoke of the need for blending through the use of a dynamic routing system. "…Despite a complex and demanding set of nested quotas, quota sampling alone does not totally neutralize the inherent biases of each sample source. The results of a study fielded with one sample source could have been different with another sample source.

Clouding the issue further is the fact that sample differences appear not just across panels, but within the same panels over time. This demonstrable fact should be evident to anyone who has followed the rapid evolution of the online sample industry. Our own research, presented at the 2009 CASRO Panel Conference in New Orleans, shows that differences in purchase intent and other key metrics exist across sample sources ostensibly representing the same market. We also presented a preliminary framework for online sample source optimization.

Blending panels has been proposed as means of stabilizing data sources. De Gaudemar stated: "It is the sample buyers' responsibility to use different variables…to combine potential sample partners into an ideal sample mix." These facts present researchers with a dilemma. Sample purchasers rarely have access to sufficient information to blend samples. Until now, too little analysis has been done and too few tools have been developed to provide guidance to researchers who wish to do right by their clients and properly understand and manage online sample source bias. Accordingly, all too often marketing researchers turn a blind eye to the process, with the hope that their sample blend will somehow meet clients' needs.

Survey research has historically relied on a probabilistic model to underlie its sampling frame. However, few if any sampling frames are probabilistic. To ensure that non-probabilistic sampling frames exhibit appropriate levels of continuity, predictability and reliability, they must be measured and calibrated over time.

In this paper we further develop the use of optimization theory to try to provide stable sample blends. We propose and analyze possible optimization targets and potential methods for achieving such blends. However, all optimization efforts must rely on effective and consistent measurements.

Obtaining Sample Data

As panels' and other sample sources' recruitment models continue to shift, panels will age and shift with them; we need a reliable anchor that rises above these problems. It is essential that we explore tools to measure these changes. Without a means of comparison, we cannot expect to measure drift nor can we expect to have a platform for predicting the future. To define this reference point, we have launched a data collection effort in forty countries around the globe. In each country, we have been collecting data from all willing respondent databases using a standard questionnaire.

The questionnaire has been translated into multiple local languages for global execution. Each sample source provides a minimum of 400 interviews with quotas on income, age, and gender appropriate for that country. We use native language translators to assist us in modifying our instrument for local use. Currently (August 2009) we have completed some 65,000 interviews across 35 countries. In the United States alone we have collected data from 150 sources.

The testing procedures were based on the execution of a standard questionnaire instrument. The questions were selected to allow a consistent standard and independent assessment of the panel or data source. The questions were selected to provide the development of two types of data source metrics based on: (1) **respondent behavioral characteristics** and (2) **structural information.** Respondent behavior metrics measure the ability of the panel to reflect benchmarks (i.e. demographics) and to indicate respondent characteristics that are currently the target of a quality debate that rages in the industry. These include demographics, hyperactive respondents (professionals), speeders, straight-liners, and faults. As previously noted quotas were imposed for age, income and gender, while suppliers were requested to use “best practices” to balance region. No instructions were given regarding employment, education or other demographics. These then acted as respondent behavior characteristics for the panels.

Structural Segmentation

The structural information is based on respondents’ behavior and attitudes we use to form segmentation structures. These structural segments provide measures that reflect the underlying population rather than being sensitive to changes in individual characteristics. They reflect the potential bias or hopefully the lack of it in the panel. Ultimately, the goal would be to balance panels to better reflect the target population distribution of these segments. Our goal is to use segmentation to create a fingerprint that can be consistently maintained by blending panels. The process of identifying structural segments can be thought of as having four steps, going from the selection of the variables through identifying segments and developing a regression model.

Select Variables → Cluster Analysis → Logit Regression Model → Test Results

The details of the analysis procedure are shown in the appendix. This analysis is done with a substantial set of data, within a single country in order to provide a stable structure. The resulting regression model is then used to assign segments for all other datasets. The resulting regression model represents the segmentation scheme and is a critical focus of the analysis and testing.

Three segmentation schemes are being used in this evaluation focusing on buyer behavior, sociographic, and media use factors. These are not the only segmentation schemes that can be developed for this process. However, these were well supported by the test questionnaire and fulfilled the requirements as structural segments. Studies involving purchases will depend on the distribution of “buyer behavior” segments. It is for this reason that it is the primary focus in this analysis. However, there are other pragmatic approaches that are of value. For example media preferences could be the driver for a media company and sociographic analysis can be critical in public opinion data.

Buyer Behavior Segments

The buyer behavior segments are intended to capture the variability in the attitudes and actions regarding the purchase of a broad range of products. The standardized profiles are shown in Figure 1 (see Appendix) and reflect the response to 37 input variables. The titles of the segments reflect the strongest loading variables making up the segment. The purpose of this scheme is to reflect differences between sources of data and the *Grand Mean* representing that region. The segments vary widely between different countries, as shown in Figure 2. These are likely due to cultural variations. However, we expect the distribution of these segments among panel and sources of data within regions to be far more consistent. Figure 3 compares buyer behavior segment distributions for a typical online access panel in the United States as compared to the U.S. *Grand Mean*.

Sociographic Segments

The sociographic segments are intended to capture the variability in behavior and attitudes regarding a broad range of lifestyle decisions. The standardized profiles are shown in Figure 4 and reflect the response to 31 input variables. As in the case of the buyer behavior segments, the titles of the segments reflect the strongest loading variables making up the segment. It is important to note that the distributions of these segments are expected to vary widely between different countries and regions as shown in Figure 5. However, we expect the distribution of these segments among panel and sources of data within regions to be consistent as before. Figure 6 is the comparison between the sociographic segment distributions for the same typical online panel in the United States compared to the U.S. Grand Total.

Media Segments

The media segments are intended to capture the variability in the use of various sources of communications and activities. The standardized profiles are shown in Figure 7, and reflect the response to 31 input variables. The variables used were combinations of those also used for the buyer behavior and sociographic segmentation but focused on media issues only. As in the case of the other segmentation schemes, the titles of the segments reflect the strongest loading variables making up the segment.

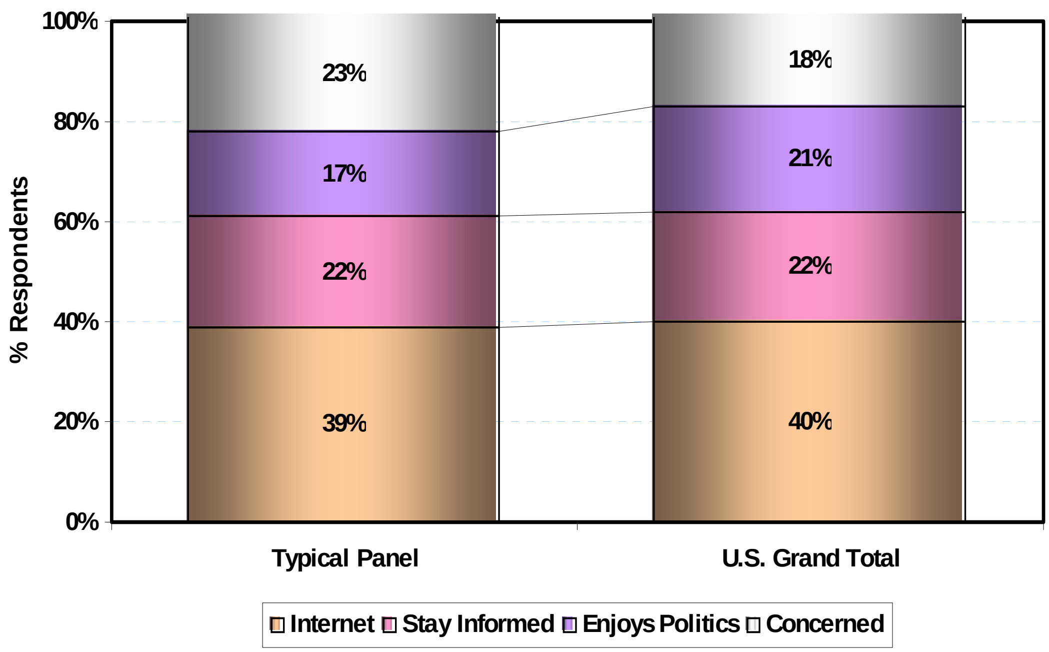

Figure 8, is the comparison between the media segment distributions for the same panel as in the above sections compared to the U.S. Grand Total.

The causes of the disparity between the segment distributions of a panel or data source and that of the collective total for a country are probably the product of concentrations of various behavioral groups. These are in turn created by different methods of recruiting participants, the incentive process and panel management. These disparities are going to exist for any specific sample source, but particularly for non-probabilistic online sources. The issue is how these disparities can be reduced from a sampling frame.

Obtaining the *Grand Mean*

We cannot stabilize online data unless we create a reference point as an anchor. The reference or basis of comparison for all metrics for this test was the average value across all samples in a region. This we define as the *Grand Mean*. The objective is to test a minimum of five data sources in each country. We anticipate that the first wave of data collection will be completed in fall of 2009. These data are the basis for our blending models and consistency testing. We use only commercially available online access panels to form the *Grand Mean*. Alternative sources of data such as random phone dialing were not included in their formulation. It is important to also note that the *Grand Mean* is specifically regional. That is, it reflects the samples for a specified country and are not global.

Typically, panels and lists are filtered to balance demographics against some external standard such as the known general population. However, this still does not assure that the source reflects the targeted group of respondents or even the larger population. We have collected data on over thirty standard reference points to assist us in evaluating the *Grand Mean.*

Optimization

Mixing of panels has been suggested as an obvious means to stabilize data sources. The resultant variability of an average is most often less than that of a single panel. The old adage of not putting all ones eggs in a single basket appears to apply. On the one hand, there is no assurance that simply combining of panels will produce the desired result. On the other, having obtained metrics on the distribution of segments in the panel community allows optimum mixes of panels to be obtained. Thus the collection of *Grand Mean* data is critical to understanding the potential for optimum blends.

The optimum mix is determined by varying the weight of each of the panels in a set. While any number of panels can be used for the set, usually for practical reasons it is desirable to keep the number small. For Buyer Behavior segment distributions, there are four segments but only three are independent values, since the total weights must equal one. In this case, three panel sets are used. Note, however, that the eventual optimum may have less than three if any of the weights are close to zero.

The optimum is the point where the disparity of the distribution of segments from the *Grand Mean* is at a minimum. This disparity is the objective of the optimization and is assumed to be the Root Mean Squared Distance[1]. The size of the segments of the resulting mixed panel is the weighted average of the component panel values. The optimum can be found graphically as shown in Figure 9 for a particular case. The bottom-most point on the surface represents the optimum value. This point can be seen in Figure 10, which is a contour map. The red zone area represents the lowest point, the center of which would be the optimum.

[1] This is the square root of the average of the squares of the difference between the panels' segment sizes minus that of the *Grand Mean*.

An optimum solution was obtained from each of the sets of three panels or sources from our database of 17 US panels. The *Grand Mean* was estimated from these data sources. This resulted in an ensemble of 680 possible optimum sets[2]. The distribution of the optima is shown in Figure 11. Over 25% of these optima indicated less than a one percent deviation from the *Grand Mean*. However, in some cases the deviations were quite high. Not all mixes of panels can develop improved optima. In some cases two or even one panel remained in the optimum solution.

Optimum panel mixes are, therefore, not created equal. The choice of panels for an optimum mix must therefore depend on the results and other conditions. These include not only the impact on balancing one set of segments, in this case the Buyer Behavior segments, but potential on other segmentation schemes. The chart in Figure 12 shows the relationship between the residuals for the Buyer Behavior segment distributions for optimum panel sets compared to the residuals for the Sociographic segment distributions.

The optimization was based on the Buyer Behavior and therefore, we would expect that the residuals will be much smaller for them than for the Sociographic segments. In general, that is true. What is important to note here is that this relationship is not monotonic, though there is a significant relationship between them. Optimizing on one set of segmentation schemes does not necessarily assure low deviations in others.

Conclusion

Previously, panel sources have been regarded as enigmatic masses of respondents that float in their own world, undefined and mysterious. This paper is an attempt to bring clarity to the online sample universe and provide an anchor for quality measurement. We do this by using combinations of quantifiable metrics, segmentation, and comparisons to the *Grand Mean* of the sample universe. Our research indicates we can make sensible comparisons and, by virtue of structural segmentations, capture the nature of the underlying sampling population. We can compare panels to one another, evaluate them through time and make statements about their consistency, predictability, and reliability.

It is the measure of sampling frames consisting of the online panels and data-sources through time that we believe is critical to establish a consistent, predictable and reliable sample frame.

We call this new framework *Consistency Analysis*. We seek to anchor online research in this new framework. By measuring panel samples we can better understand how they may drift through time in a world of ever-changing sources. By frequent measurement and re-calibration, we believe consistency measures are possible.

[2] This is the results of all combinations of 17 items taken 3 at a time.

Appendix - Segmentation Analysis Process

As such, this process can be iterative where groups of variables are tested until a satisfactory set are identified.

- **Selecting Variables** – Traditional cluster analysis abhors missing data and as such questions that either lack data or contain "don't know" responses are usually excluded. Metric variables are preferred.

- **Adjusting the Data** – It is useful to have all variables monotonic[3] and balanced. That is, preferred values going in the same direction. Furthermore, it is also useful to transform categorical data into combined metric variables[4]. For use with traditional cluster analysis methods, the data is normalized across the respondents. Otherwise, the techniques tend to focus on total average values rather than strength of particular variables.

- **Hierarchical Clustering[5] of Normalized Sample** – This is used to help identify the number of clusters that would be needed[6].

- **K-Means[7] Clustering of Normalized Data** – All the normalized data is then clustered based on the number of selected clusters. This generates segment assignments to each respondent record.

- **Profile Segments (Non-Normalized Data) -** Summary statistics for the total database and for each of the segments is then generated to determine the distinctness of the segments and determine a description of them. This is done based on the original altered data (not normalized). These profiles are used to assess the quality of the segmentation.

- **Determine Multinomial Logit Model** – Logit (Logistic) Regression is a non-linear curve fitting technique used with a categorical dependent variable. In its full form, it generates a probability or likelihood of a respondent being in a specific group. It is used here to develop a "progressive model" which will be used to assign respondents to segments for datasets not used for the original segmentation.

- **Test the Regression Model** – Because of the planned projected use of the Logit segment model, it needs to be extremely accurate in assigning segments. This reflects the distinctiveness of the segmentation scheme and its reliability. For this use, we required almost a 100%[8] recapture of the original assignments by the model.

---

[3] Changing the direction of questions (best to worst) is often useful for testing consistency and can be required for specific methods of data collection.

[4] For example, we combine identified purchased products into a single variable of number of such products purchased.

[5] Hierarchical Clustering used Wards linkage with Euclidean distances.

[6] The tools associated with this approach gives insight into the impact of the number of cluster, including cluster trees.

[7] K-Means clustering is used here due to the large dataset.

[8] For the three segmentation schemes, the resulting models were totally (100%) able to reproduce the assignments.

Appendix - Charts

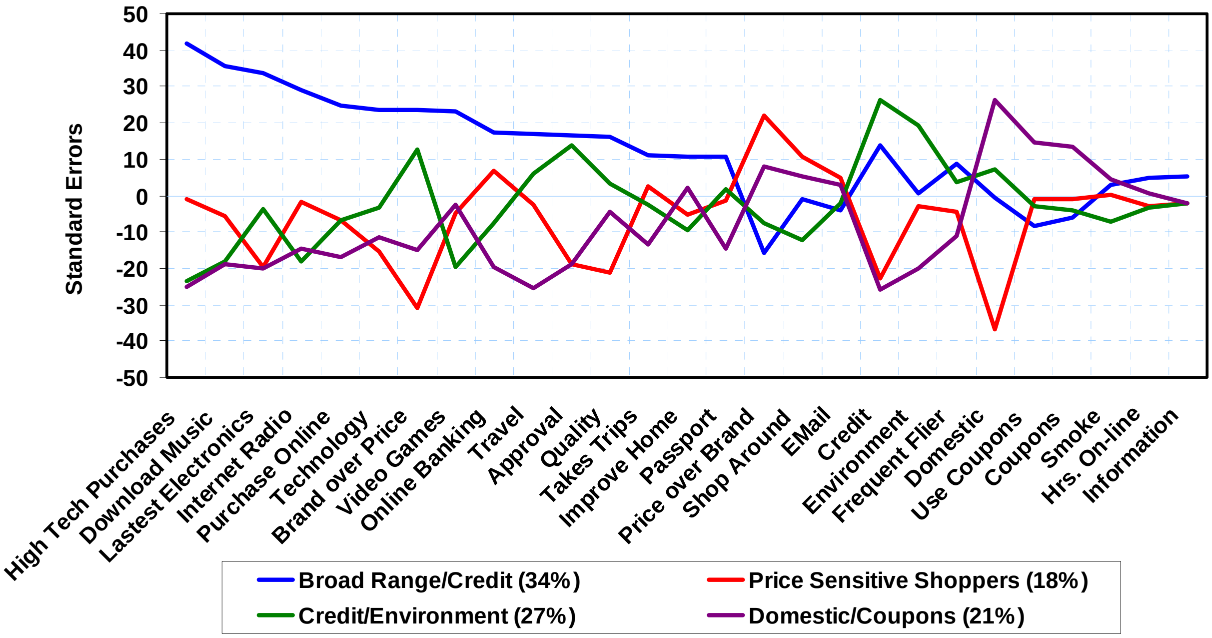

Figure 1, Standard Profiles for Buyer Behavior Segments

Figure 2, Country Average Buyer Behavior Segment Distribution

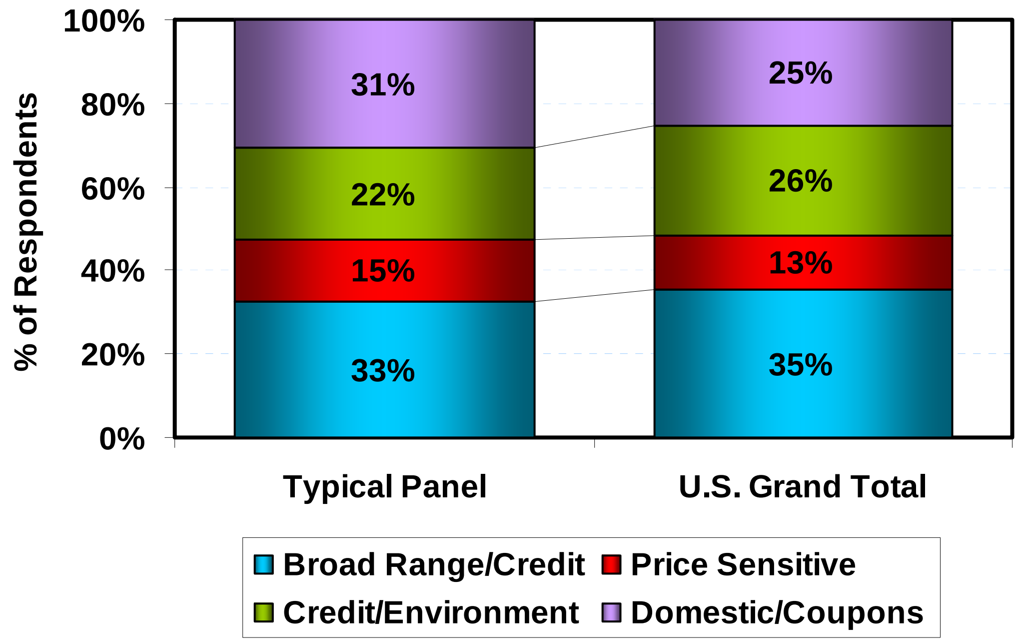

Figure 3, Comparison of a Panel to Grand Total for Buyer Behavior Segments

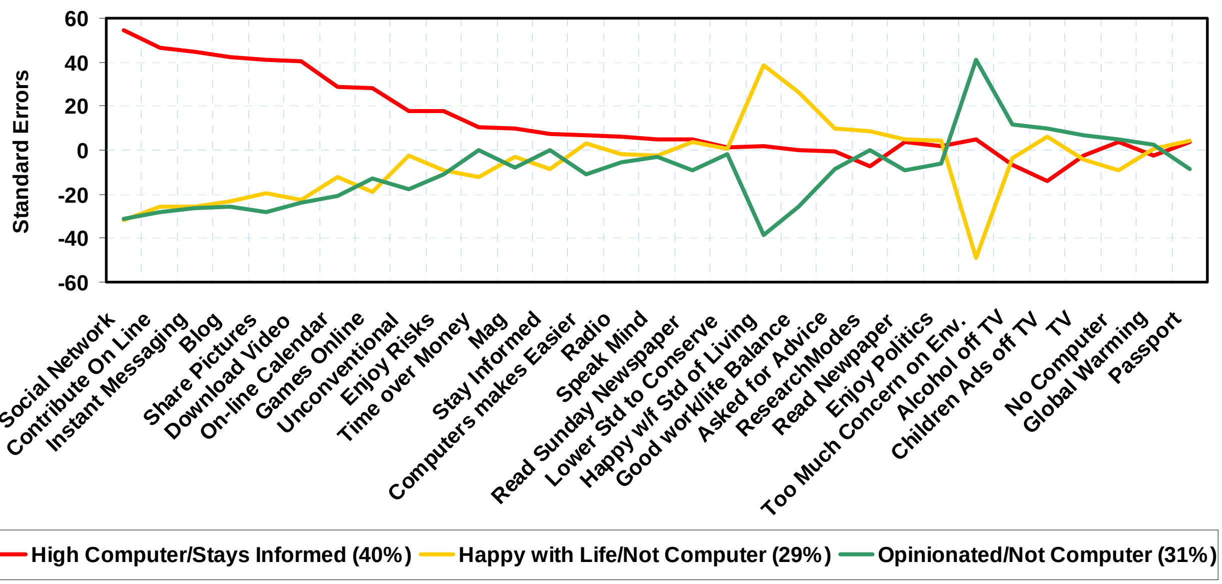

Figure 4, Standard Profiles for Sociographic Segments

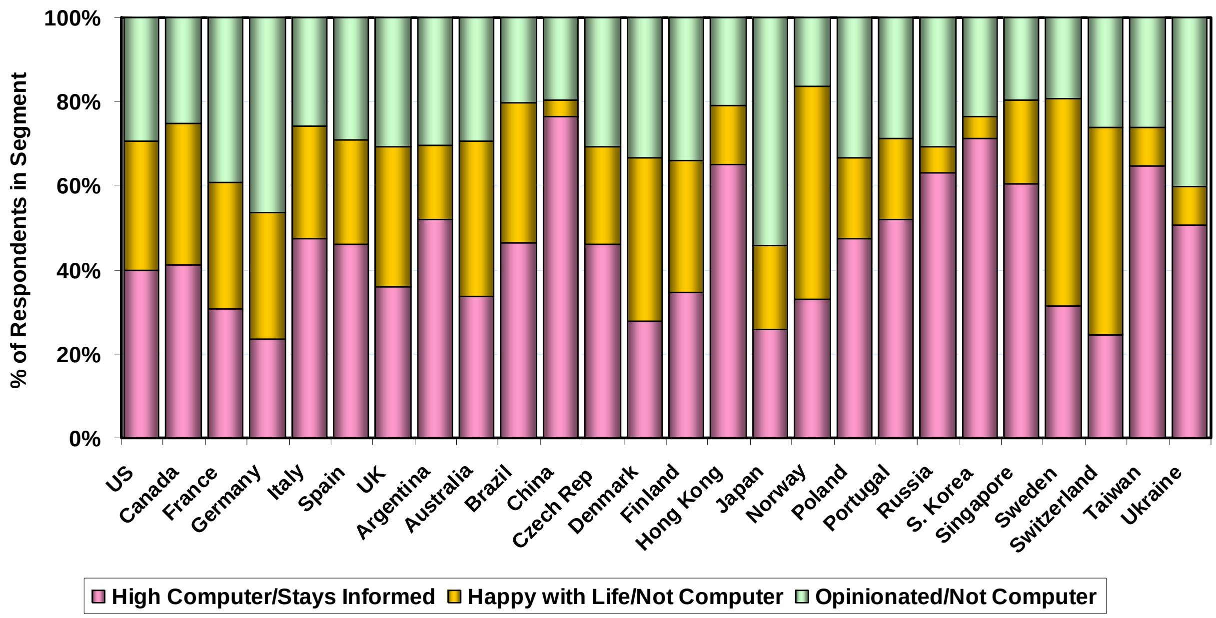

Figure 5, Country Average Sociographic Segment Distribution

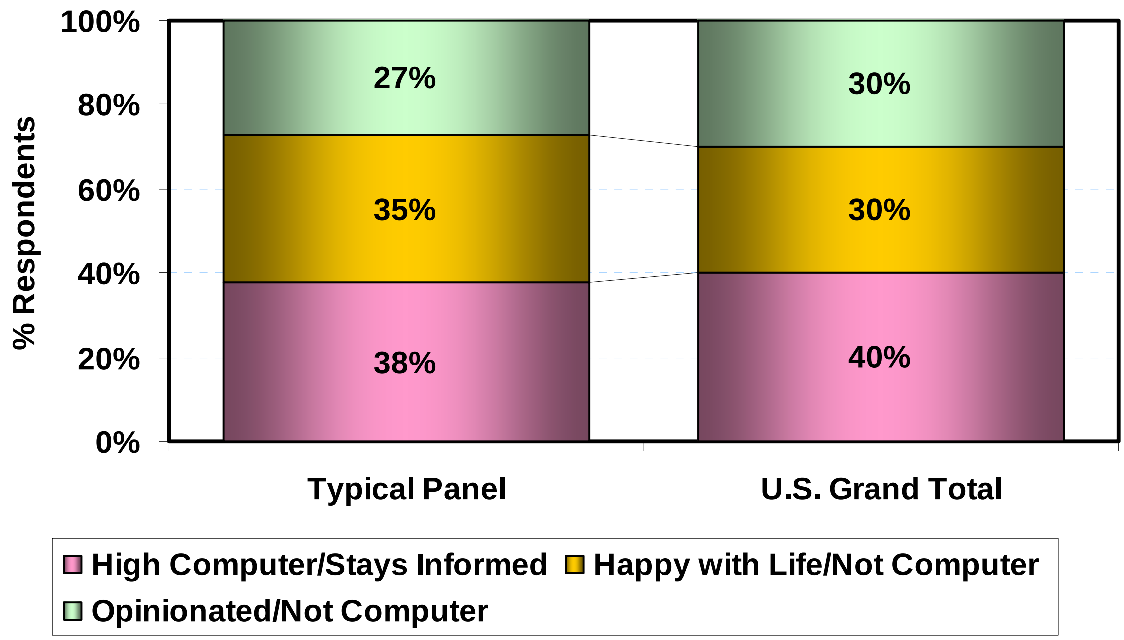

Figure 6, Comparison of a Panel to Grand Total for Sociographic Segments

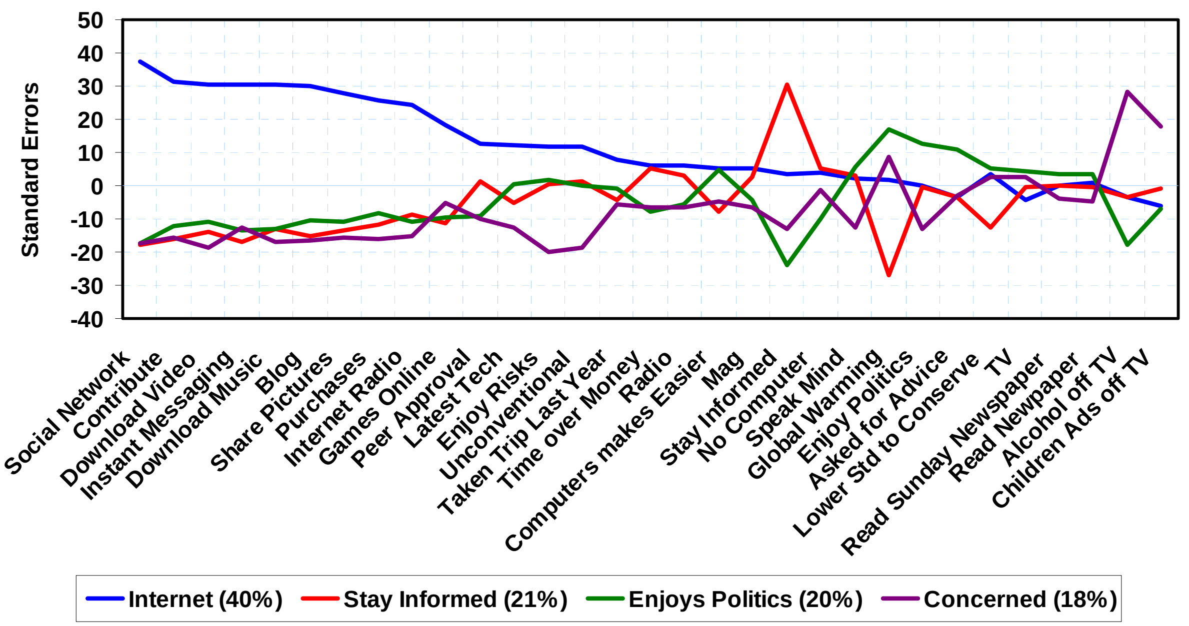

Figure 7, Standardized Profiles of Media Segments

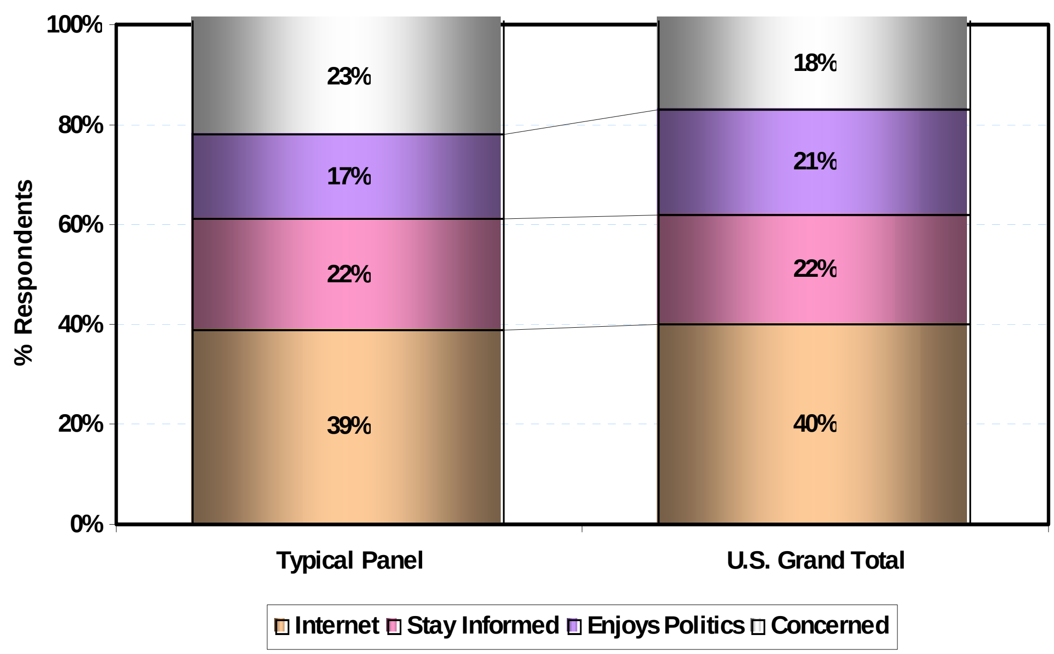

Figure 8, Comparison of a Panel to Grand Total for Media Segments

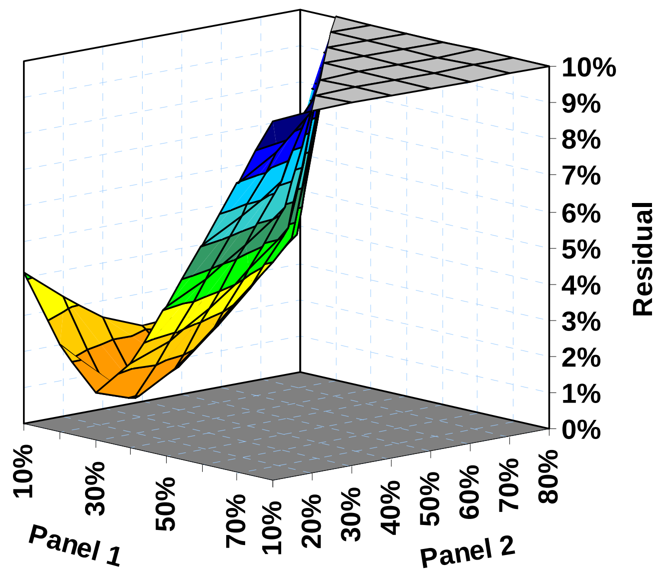

Figure 9, Seeking the Optimum, Surface Map

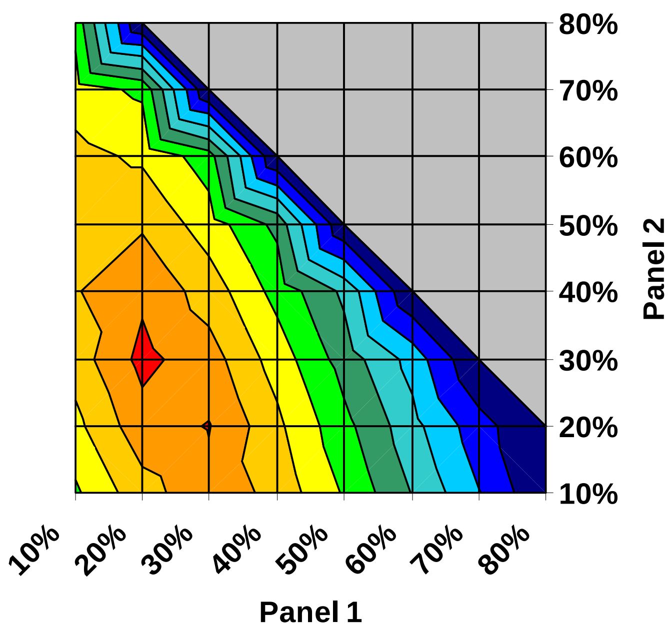

Figure 10, Seeking the Optimum, Contour Map

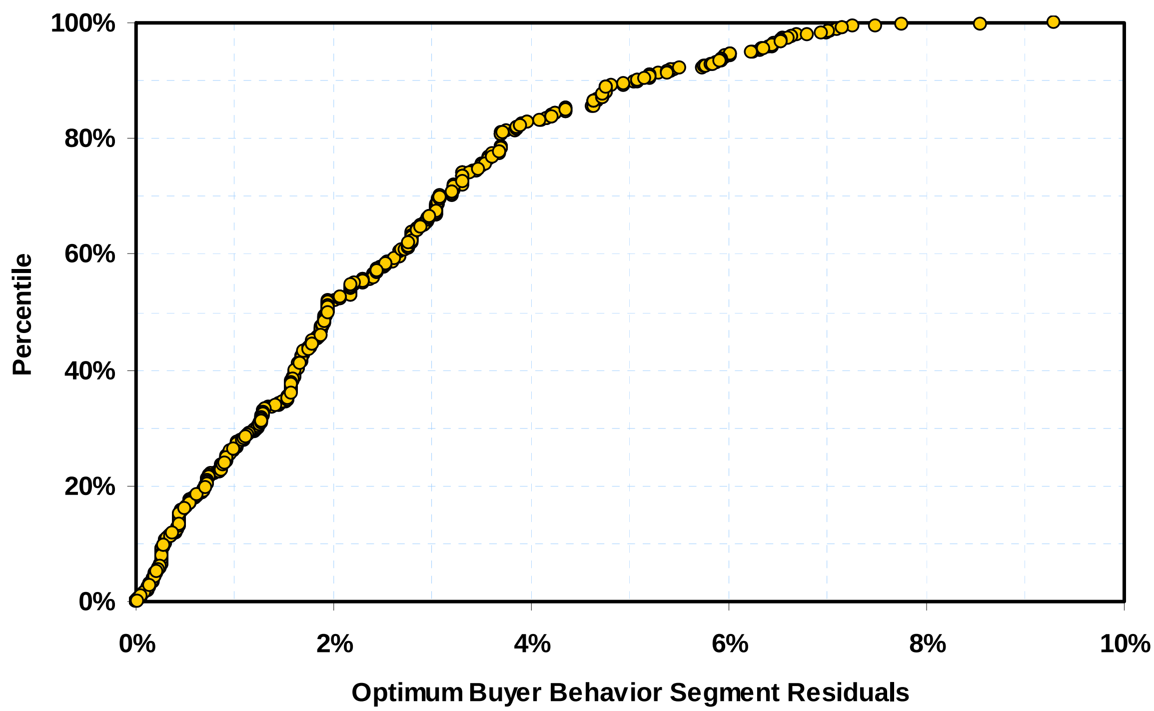

Figure 11, Distribution of the Optimum Buyer Behavior Segment Residuals

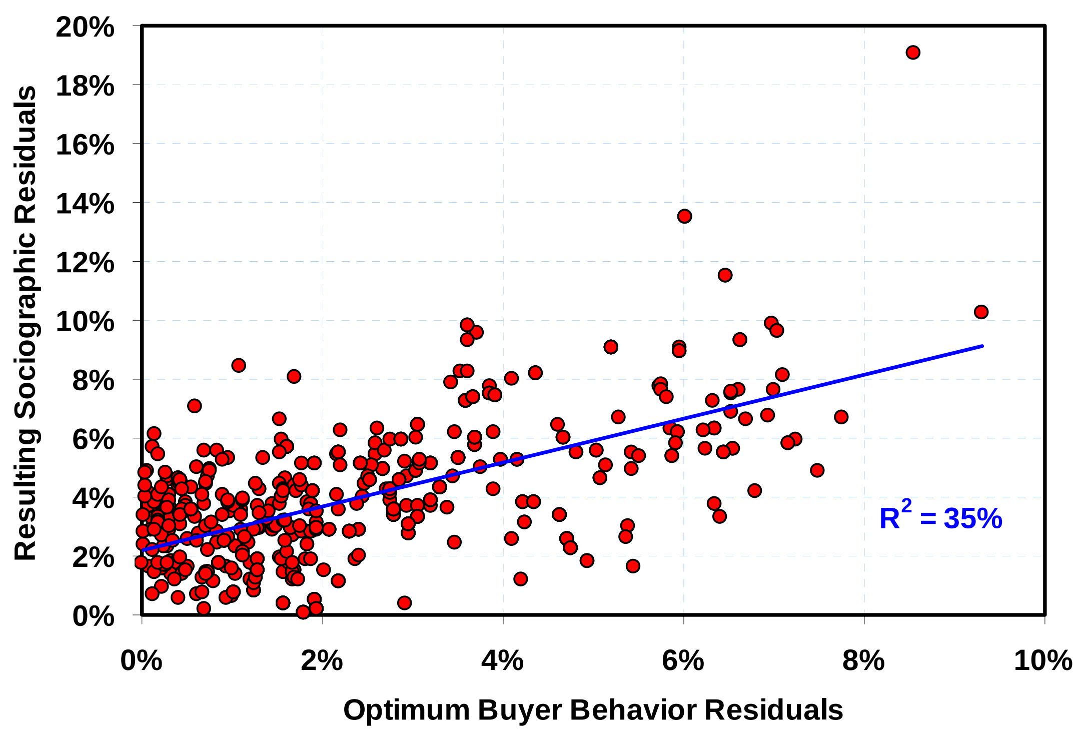

Figure 12, Residuals of the Sociographic and the Optimum Buyer Behavior Segments

Bibiliography

De Gaudemar, O. (2006). Benefits and challenges of multi-sourcing. Understanding differences between sample sources, Panel Research, 2006, ESOMAR World Research Conference, Barcelona.

Gittelman, S. and E. Trimarchi (2009). On the Road to Clarity: Differences between sample sources. CASRO Panel. New Orleans.

About the Author

Steven H. Gittelman, Ph.D. has been the president of Mktg, Inc. since 1979. He received his doctorate with a specialty in optimization theory in 1976 from the University of Connecticut. He has a background in statistics and sampling theory. He is the author of many technical papers and two books.