RECENT DEVELOPMENTS IN ADDRESS BASED SAMPLING

Overview

Survey and market researchers are increasingly reconsidering address-based sampling (ABS) methodologies for survey administrations as well as commercial applications. Specific enhancements provided by Marketing Systems Group (MSG) play an integral role in establishing the ABS methodology as a practical solution for survey and market research applications. In particular, such enhancements include amelioration of some of the known coverage problems of the ABS frames as well as their augmentations with geodemographic and other ancillary data items. These enhancements allow researchers to develop more efficient sample designs and broaden the analytical possibilities through an expanded set of covariates for hypothesis testing and related inferential tasks.

The Computerized Delivery Sequence File (CDSF) is derived from the U.S. postal frame used for mail delivery. The raw CDSF needs refinements in several aspects before it can qualify as a credible tool for survey sampling. First and foremost, this database does not include geodemographic indicators for effective sample stratification, an issue of critical importance for complex designs. Moreover, for surveys where data is to be collected on-site, reliance on delivery information may not be adequate as the exact location of all sample dwellings must be known. This is of particular concern when a P.O. Box is the only means of delivery for a household. On the other hand, there are households that have both residential addresses as well as P.O. Boxes. Ignoring this problem leads to frame multiplicity since such households will have multiple chances of selection. In what follows, we provide a brief overview of how MSG manages many of these challenges and the ability to append a wealth of ancillary data.

Data Improvement in Residential Addresses – the MSG Difference

Advances in database technologies along with improvements in coverage of residential addresses have provided a promising alternative for surveys that require representative samples of households. The CDSF, with over 159 million residential and business delivery points on file, is the most complete address database available in the US. Many commercially available consumer databases lack the coverage and address accuracy that the CDSF offers. The CDSF is continuously updated with feedback compiled from individual Post Offices and letter carriers. This continuous address hygiene applied to the CDSF results in fewer undeliverable addresses and increases the speed of delivery.

The following table provides distinct summary counts for the main types of delivery points.

June 2020 ABS Frame

| | CDS | |
| | Residential | Business |
Address Type	Count	Count
Total Addresses	146,467,273	12,623,736
City Style	130,730,203	9,469,300
Drop Point	700,424	112,732
Drop Units	1,947,756	2,491,795
Drop Unit Augments (non-CDS)	41,894	19,979
Seasonal	805,715	3,914
Educational	100,361	6,240
Vacant	3,142,045	1,113,364
Throwback	191,605	137,918
Augments (non-CDS)	63,432	0
Rural Route/Highway Contract	46,756	1,368
PO Box	15,690,314	3,153,068
PO Box (Traditional)	14,260,389	3,024,992
PO Box (OWGM)	1,428,986	128,076
PO Box Augments (non-CDS)	939	0

There is not a one-to-one correspondence between Postal geography and Census geography. As such, it is not possible to append many of the ancillary data items available from various government and commercial sources directly to the raw CDSF. However, by geo-coding each CDSF address to a unique Census block, this divide is bridged and allows the appending of valuable information to all US addresses. This is the crossroad where basic list suppliers – those that can simply offer raw extracts from the CDSF – are differentiated from MSG. By providing the following suite of enhancements, MSG can help advance the CDSF from a basic delivery database into a bonafide sampling frame suitable for complex surveys.

Detailed Geodemographic Information

By mapping each address to Census localities and accessing several public and commercial databases, MSG can append a host of geographic and household attributes to each address. Starting from the ZIP+4 level, we can roll up to higher levels of aggregation including all Census geographic domains (Block, Block Group, Tract, County, MSA, and State); marketing domains (Media Markets and DMAs); and custom defined areas (retail trading areas and specific geographies based on distance or radius).

Contact Information Retrieval

Relying on multiple database sources, MSG makes it possible to append name, telephone number and/or email address to individual CDSF addresses where possible. On average more than 90% of our sample addresses can be name-matched, 50%-60% can be linked to a landline and cellular telephone number, and 30%-40% can be linked to email address. Obviously, match rates can vary and depend on several factors including which address types are ultimately selected.

Undeliverable Address Resolutions

The CDSF does not contain address records that are void of delivery information, commonly referred to as simplified addresses. While the number of simplified addresses continues to decrease as they go through the on-going 911 address conversion, there are still just over 100,000 simplified addresses in the US. MSG can resolve the majority of the remaining simplified addresses and append all available ancillary data to the resulting addressed households. Additionally, MSG has an agreement in place to receive data on addresses identified by the USPS as No-Stat. Addresses on the No-Stat file include long-term vacant units and new/planned housing units, as well as delivery points within gated communities. Addresses contained on the No-Stat file are deemed inactive by the USPS and cannot be mailed to. However, the No-Stat file can be used in conjunction with the CDSF for determining universe estimates and is helpful in situations that require enumeration.

Identification of Areas with a Potential Coverage Problem

There are delivery points that are reachable only via P.O. Boxes. Also, in certain areas there are newly constructed dwellings that are currently not registered with the Postal Service. In such cases the physical location of the corresponding households may be unknown and are not included in the CDSF. Given the significant cost of on-site enumeration that can become necessary for such areas, MSG has developed procedures for predicting the quality of the CDSF coverage. This is based on specific characteristics of area segments obtained from commercial and public databases.

Frame Multiplicity Reduction

As mentioned above, there are residential delivery points in the CDSF that are reachable only through a P.O. Box. Known as "Only Way to Get Mail" (OWGM) P.O Boxes, there are over 1.4 million of these delivery points with 60% of them non-vacant. Aside from these, there are an additional 14 million traditional P.O. Boxes that are not the only means of delivery. In all likelihood, these delivery points correspond to households that are represented in the CDSF multiple times; once as a residential address and one or more times via traditional P.O. Boxes. By eliminating traditional P.O. Boxes and those that are the only means of delivery but vacant, it is possible to remove virtually all duplicate listings from the CDSF. Before selection

of samples, MSG can provide the counts of these and all other delivery types in the geography of interest so that one can determine the exact composition of the sampling frame for their survey.

Selection of Samples Based on Complex Designs

In addition to the above enhancements, MSG's team of in-house sampling statisticians and methodologists can assist in the designing and selection of samples based on complex sampling plans. We offer sampling design consultation to clients with a wide range of market and survey research needs. Proper selection of such samples often require advanced procedures to accommodate multi-stage sampling in conjunction with selection methods of proportional probabilities to various size measures.

Virtual Platform for Address Based Sampling

In order to accommodate our clients' sampling needs on a 24/7 basis, MSG is the only company that offers virtual sampling – both for RDD and ABS – through our secure client-server architecture. Virtual ABS makes it possible to select probability-based sample addresses from our enhanced address frame. Several selection parameters are available including various levels of geography and individual address attributes. As with custom orders, Virtual ABS clients have access to our advanced process for name, telephone number and auxiliary variable retrieval that result in the highest match rates possible.

Improvements for Increasing Response Rates

It is becoming progressively more difficult to secure respectable rates of response for all surveys, particularly those relying on single methods of contact. Given that nonresponse is highly differential and can vary significantly across different demographic subgroups, it is of great concern when a notable percentage of sample units opt not to respond to a survey. Even when sophisticated nonresponse adjustment procedures are employed to reduce the incurred bias, it may not be enough to reduce nonresponse bias to a tolerable and measurable level. Furthermore, reducing nonresponse bias via weighting is always exercised at the expense of reduced precision of survey estimates, as weighting inflates the variance of survey estimates. In addition to address, you can append land line and cellular telephone number as well as email address where available. Multiple modes of contact allow for the deployment of an efficient multi-mode design that can be implemented sequentially based on costs. Furthermore, the increase in costs for off-line data collection methods are driving more to seek on-line solutions for survey administration. To this end, ABS is the ideal frame for a mail push to web methodology.