Chronological Summary of Research Related to the
Use of Address Lists in the CPI Housing Survey

October 23, 2006

Introduction

This memorandum summarizes the research that was conducted largely between 2002 and 2003 related to the decision to use address lists as a sampling frame for the CPI Housing Survey in place of listing. The overall motivation for making the decision was based both on cost and necessity, along with the belief, supported by the research summarized here, that such a change would not be detrimental to the CPI Housing Survey and in fact could be beneficial. It is expected that the use of address lists will lead to improvements in the survey, chiefly by more accurate and complete sampling frames leading to larger realized sample sizes in high owner occupied segments, which were exceptionally low in the 1998 revision.

Research Summary

In 2002, address lists were first considered for use as a replacement to the internal BLS listing process that was used in the 1998 revision. To evaluate this possibility, BLS contracted with Westat to perform research on private sector lists, evaluating them in terms of coverage and accuracy of the lists. A summary of this research was included in a May 2003 report to Ken Dalton, then Associate Commissioner of the Office of Prices and Living Conditions, from John Greenlees, then Assistant Commissioner of the Division of Consumer Prices and Price Indexes.

In addition to the Westat studies, other research by BLS was referenced in the report to Mr. Dalton. In March 2003, the two documents were disseminated. They were a document highlighting the some of this research, including some highlights from the Westat studies, and a high level sample design. In summary, the research showed that the tenure probability codes of 9 included on the address lists were highly indicative of an address being owner occupied, the coverage of the lists in terms of raw numbers of addresses were promising, especially at the block group level, and a sample drawn from the lists excluding the high probability owners had the potential to yield a much larger sample of renters than our previous sample.

Also in March 2003, staff met with an address list provider to get information about the construction of the lists, with respect to coverage concerns. At this meeting, the address list provider suggested using a list based on the US Postal Service delivery sequence file in addition to the marketing list CPI was considering. This was the ADVO list mentioned in the report to Mr. Dalton. In May 2003, we contracted with Westat to evaluate the coverage of the marketing list and the ADVO list for 542 block groups across the country corresponding to a random sample of 500 of the approximately 10,000 segments in our sample.

After reviewing the report, coverage rates for CPI listing results were compared to the coverage rates in the memo for the marketing list, the ADVO list, and the combination of the ADVO list and the marketing list. The results showed that in terms of raw address coverage, addresses produced from the CPI listing process were found on the private sector lists. This research has been replicated in the table below:

List	Region	Mean	Stand. dev.	Percentiles				
				90	75	50	25	10
InfoUSA	All	0.958	0.292	1.156	1.014	0.932	0.860	0.769
ADVO	All	1.102	0.384	1.373	1.159	1.062	0.993	0.889
InfoUSA or ADVO	All	1.252	0.507	1.569	1.296	1.175	1.076	1.004
BLS Listings	All	1.108	0.427	1.355	1.137	1.039	0.982	0.878

In addition to the Westat research, BLS performed three other pieces of research. The first two dealt with matching the purchased lists to the corresponding segments and housing units in the CPI Housing Sample. First, we compared the tenure results in the Housing survey with the tenure codes on the lists. The significant result drawn from this was that the tenure code of 9 on the address lists was a reliable indicator of an address belonging to an owner, with close to 99% of the addresses in the CPI sample identified as owner occupied were also coded as 9 on the marketing lists.

Second, we compared price relatives calculated using our complete sample in those segments with price relatives calculated using only the addresses that we were able to match to the lists and that did not have a tenure code of 9 from the lists in those same segments. The table below contains the results of a paired t-test of these price relatives showing no significant difference between them.

Variable	N	Mean	Std Error	T Value	Pr > |t|
Time period 1					
Economic rent relative difference	38	-0.0013885	0.0010564	-1.31	0.1968
Pure rent relative difference	38	-0.0003837	0.0006622	-0.58	0.5657
Time Period 2					
Economic rent relative difference	38	0.0009587	0.0009694	0.99	0.3294
Pure rent relative difference	38	-0.0001291	0.0012088	-0.12	0.9025

The third piece of research that was conducted with the data was a field test to determine the viability of the addresses, that is, whether the addresses actually corresponded to housing units. The field was given a sample of addresses of about 2080 addresses selected in block groups in 25 PSUs. With the exception of the addresses selected within each block group, the sample was one of convenience based on field resource availability but chosen to approximate the distribution of sample across city sizes. As a result, only summary results were generated from the sample. These have been condensed into the table below:

Status	All Addresses	ADVO only addresses	InfoUSA only addresses	Overlap addresses
Found Housing Units	1606	560	127	919
Housing Unit Found, ambiguous address easily corrected with procedures	106	17	75	14
Probable Housing Units, unable to personally verify	26	11	6	9
Possible Housing Units, ambiguous address, possibly corrected with procedures	121	26	58	37
Commercial Establishments	174	143	16	15
No Address Found	7	5	2	0
Total	2040	762	284	994
Percent of Locatable Addresses	84 - 85%	76 - 77%	71 - 73%	94 - 95%

Additionally, address characteristics highly indicative of commercial establishments were identified and if those addresses were eliminated ahead of time, the overall percent of addresses that could be found would increase to about 87 - 89%.

In late 2003 to early 2004, the research performed was reviewed and it was concluded that using address lists as a replacement for the listing process and taking advantage of the tenure codes on those lists to eliminate highly probable owners would have a positive effect on the CPI Program.