LANGER RESEARCH ASSOCIATES
SURVEY RESEARCH DESIGN • MANAGEMENT • ANALYSIS

Adding Cell Phone Sampling to a Long-Term Trend Study: The Bloomberg Consumer Comfort Index

Julie E. Phelan, Ph.D. & Gary Langer

Paper presented at the annual conference of the
American Association for Public Opinion Research
May 2013

Introduction

At least three in 10 Americans live in cell-phone only households, raising the question of whether traditional landline (LL) telephone surveys represent an adequately robust survey methodology. The consensus among most serious researchers is that they do not.

Two questions define the issue: First, whether LL samples sufficiently represent the national population in terms of the attitudinal measurements of interest; second, if yes, whether they do so in a way that allows for sufficiently comprehensive data analysis.

The first is unresolved. Weighting to Census demographic norms clearly corrects some discrepancies in LL samples, such that many differences in LL vs. landline-plus-cell estimates have been shown to be minimal in terms of the attitudes commonly measured in opinion surveys. But larger biases have been shown to exist in less commonly measured variables. And other, untested differences may exist or may develop over time.

The answer to the second question is less equivocal. LL samples are marred by disproportionately small sample sizes in undercovered groups and by higher design effects due to weighting. Both constrain data analysis, by reducing some subgroup sizes below acceptable levels and by increasing sampling error, thus limiting statistical power in both marginal and cross-sectional comparisons. On a more perceptional than empirical basis, moreover, 30 percent noncoverage raises questions of face validity.

For these reasons, as the cell-phone only population has grown, most reputable national public opinion polls have moved in recent years to include cell-phone respondents in their samples. This paper reports on a test of such a shift in the sampling approach of the ongoing Bloomberg Consumer Comfort Index.

Spanish-language

In addition to cell phones, another element of this study tests the inclusion of Spanish-language interviews in the CCI methodology. The share of the nation's adult population

that is Hispanic increased from 6.3 percent at the start of the CCI in late 1985 to 14.1 percent today. An estimated 22 percent of this population is conversant in Spanish only.

The noncoverage of Spanish speakers in English-only questionnaires is far smaller than the noncoverage of cell-phone respondents, and as such is insufficient to influence most estimates of the national population overall. Research however suggests it may compromise within-group evaluations of Hispanics themselves. Including Spanish interviews, therefore, while not essential for general purposes, can add analytical insight, and may be of particular value as the Hispanic population continues to grow. For this reason the field work vendor for the CCI survey, Social Science Research Solutions, offers the inclusion of Spanish-language interviews in its sample design.

Design Principles

Any change in survey research methodology must be undertaken cautiously, particularly when the project relies heavily on tracking attitudinal changes over time, e.g. in ongoing measurements of consumer sentiment. Careful testing and evaluation are required before such a move can be contemplated.

Still, while caution is warranted, inaction carries its own risks. Changes in the accessibility of the target population themselves represent a de-facto methodological departure that may influence trend data. An adjustment to achieve fuller coverage can be seen as a move to maintain a methodology given changing communication methods, rather than as a change in methodology itself.

The University of Michigan has reported on one test of relevance to this paper, adding cell-phone interviews to the customary LL sample in its ongoing Survey of Consumer Sentiment. Its 2009 evaluation, available online, reports: "After controlling for differences in the demographic characteristics of the two samples no statistically significant differences in the Sentiment Index were found."[i]

The Michigan researchers may have relied upon this finding as a rationale for not switching to mixed sampling; their online disclosure makes no reference to the inclusion of cell phones in their samples. Of course, the opposite conclusion could be equally defensible: If no significant differences are apparent, including cell phones provides other benefits, in terms of larger subgroup sample sizes, lower sampling error and face validity.

The CCI Test

The present test evaluates the impact of the use of an overlapping dual frame ("mixed") sample[ii] including cell-phone and Spanish-language interviews, as well as traditional landline interviews, on the Bloomberg CCI, a weekly survey of the public's economic attitudes. The CCI asks a random national sample of 250 respondents weekly their views of the national economy, their personal finances and the buying climate. Results are reported in a four-week rolling average with a total sample size of 1,000 adults, weighted to population norms.

For this test the CCI's current LL/English-only methodology was augmented with a sample of cell-phone respondents, and with Spanish interviews (both landline and cell), during an 11-week period from January to March, 2012, producing eight weeks of CCI estimates. The LL sample was composed of a total of 2,750 LL respondents (250 per week). The cell-phone test was conducted among 825 cell-phone respondents in the same period (75 per week). An additional 82 interviews were conducted in Spanish (43 on landlines, 39 on cell phones). These cell phone and Spanish language interviews were combined with 1,882 of the LL interviews to produce a mixed sample for comparison with the LL sample.

For appropriate comparison, the LL samples for each four-week rolling average (*n* = 1,000) were weighted using existing CCI weighting protocols. The mixed sample used a weighting design that accounts for dual-frame respondents (weighted down by half) and adjusts for phone usage in post-stratification, with existing CCI weights then applied. An additional weight was created to evaluate results in the mixed sample without the Spanish interviews.

Comparing CCI Estimates By Sample Type

Table 1 compares responses to the three CCI questions by sample types averaged across the full period. None of the differences is statistically significant.[iii]

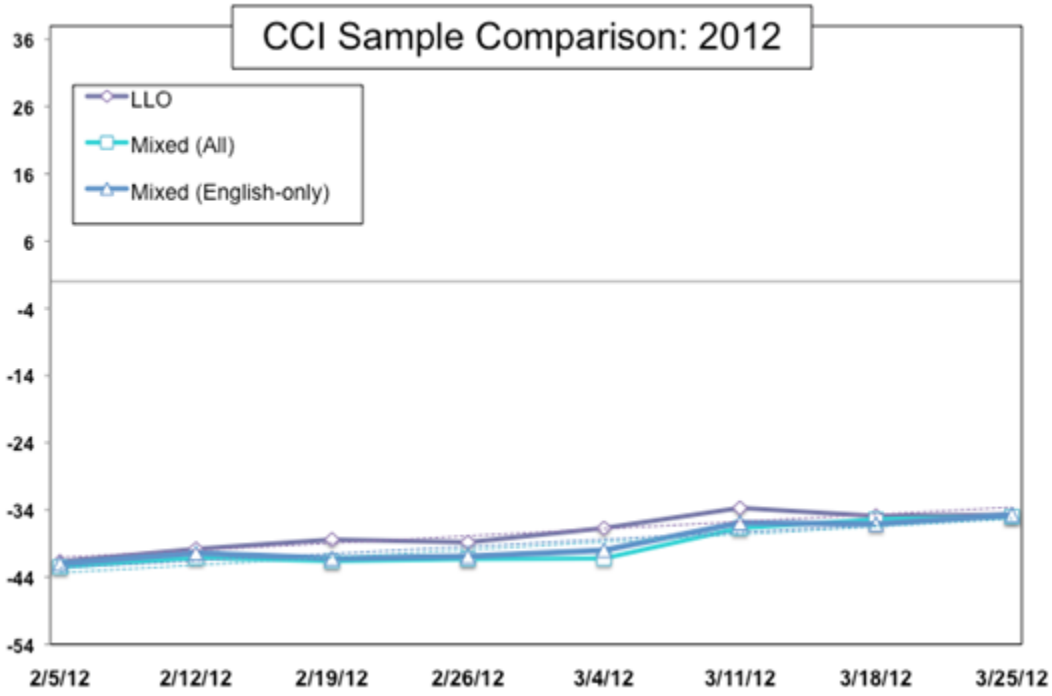
Table 1: Aggregated Responses to CCI Questions by Sample Type

	LL sample	**Mixed sample (All interviews)**	**Mixed sample (English interviews)**
NATIONAL ECONOMY			
Positive NET	14.8%	13.4%	13.3%
Excellent	0.8%	1.0%	1.0%
Good	14.0%	12.4%	12.3%
Negative NET	85.2%	86.6%	86.6%
Not so good	45.7%	46.8%	46.5%
Poor	39.5%	39.8%	40.1%
PERSONAL FINANCES			
Positive NET	49.9%	49.1%	49.8%
Excellent	4.7%	5.0%	4.9%
Good	45.2%	44.1%	44.9%
Negative NET	50.1%	50.9%	50.2%
Not so good	34.0%	33.5%	32.3%
Poor	16.1%	17.4%	17.9%
BUYING CLIMATE			
Positive NET	29.3%	28.1%	28.5%
Excellent	2.2%	2.1%	2.2%
Good	27.1%	26.0%	26.3%
Negative NET	70.7%	71.9%	71.5%
Not so good	46.3%	46.8%	46.6%
Poor	24.4%	25.1%	24.9%

While the differences were not statistically significant, most assessments in the mixed sample were slightly less positive than those in the LL sample, resulting in an average CCI value of -39.6 for the full mixed sample and -38.9 for the mixed sample excluding Spanish interviews, compared with -37.3 for the traditional LL sample. Of the total 2.3 points difference between the full mixed sample and the LL sample, then, 1.6 points was attributable to the inclusion of cell phones, 0.7 points to Spanish interviews.

We investigated this difference by looking at weekly results by sample type, including looking at the mixed sample both with and without Spanish interviews. The same general pattern was observed regardless of sample type or whether or not Spanish interviews were included; both mixed samples correlated with the LL CCI at .9 (see Figure 1).

Figure 1: Week-to-Week Comparison of CCI by Sample Type[iv]



The mixed sample without Spanish interviews averaged 1.6 points lower than the LL, ranging from 3.3 points lower the week of March 4 to no difference the week of March 25. The full mixed sample (including Spanish interviews) was lower by an average 2.0 points,[v] ranging from 4.5 points the week of March 4 to 0.3 points the week of March 25.

We conducted individual z-tests to compare responses to the three questions that make up the CCI by sample type for each week's data, in order to test whether these differences were due to sample type or simply due to chance variation. These and all subsequent analyses compared the full mixed sample (including Spanish interviews) to the LL sample. We used the full mixed sample for two reasons: one, comparisons involving the full mixed sample (hereafter referred to simply as the "mixed sample") provide a more

stringent test of the impact of the sampling shift[vi] and two, the Spanish-inclusive estimates are preferable on theoretical grounds.

Results, depicted in Table 2, show that weekly results on all three questions did not significantly differ by sample type.[vii] In other words, in each week, sample type differences in responses to the three CCI questions were no greater than would be expected by chance sampling variation alone.

Table 2: Positive Ratings By Sample Type and Week

	Economy			Finances			Buying Climate		
	LL	Mixed	Sig?	LL	Mixed	Sig?	LL	Mixed	Sig?
2/5	12%	11%	N	50%	48%	N	26%	27%	N
2/12	14%	13%	N	50%	48%	N	26%	27%	N
2/19	14%	13%	N	51%	49%	N	29%	27%	N
2/26	14%	13%	N	50%	47%	N	28%	28%	N
3/4	15%	13%	N	50%	48%	N	30%	27%	N
3/11	17%	15%	N	50%	50%	N	33%	31%	N
3/18	18%	15%	N	49%	51%	N	31%	30%	N
3/25	17%	16%	N	50%	52%	N	32%	29%	N

Further, the variability from week-to-week within each of the samples was greater than the average discrepancy between the two samples. The standard deviations across the eight weekly CCI estimates within the LL sample (2.8 points) and in the mixed sample (2.9 points) are greater than the amount the two samples deviated from one another each week averaged across the 8-week period (2.1 points). This, in tandem with the lack of statistically significant differences in the three CCI components, suggests that a switch in sampling methods is unlikely to significantly impact estimates or trend for the overall CCI.

CCI Among Groups

Table 3 shows the average difference in weekly CCI estimates by sample type among demographic groups. In most cases, these differences are greater than the difference between the full samples, an expected result since smaller sample sizes carry a higher margin of sampling error. The differences are not consistently unidirectional; mixed-sample values averaged numerically higher than LL values in 12 of 36 cases, and lower in 24.

The third and fourth columns of Table 3 show the standard deviations (*SDs*) of the weekly CCI within each of the samples during the test period. In 31 of the 36 subgroups, the average difference between the samples is either smaller than the average variability within one or both samples (in 26 cases) or smaller than the average variability within the two samples combined (in five cases).

That leaves five groups with differences greater than customary week-to-week variation. One of these groups (18- to 34-year-olds), has an average sample size of less than 100 in

the LL sample. We typically do not report estimates for sample sizes smaller than 100 because they can be unstable; therefore the discrepancy in this group is not surprising.

Table 3: CCI Subgroup Differences

Demographic Group	**Average mixed-LL difference**	**8-week LL *SD***	**8-week mixed *SD***
Men	-2.6 pts	2.6 pts	2.6 pts
Women	-1.7 pts	3.9 pts	4.5 pts
Age: 18-34*	**-8.2 pts**	**2.4 pts**	**5.1 pts**
Age: 35-44	-4.4 pts	9.5 pts	2.9 pts
Age: 45-54	+1.7 pts	4.7 pts	6.2 pts
Age: 55-64	+0.8 pts	4.3 pts	4.1 pts
Age: 65+	+3.4 pts	4.3 pts	2.4 pts
Income: <$15K	-0.2 pts	4.7 pts	6.5 pts
Income: $15-25K	-3.8 pts	14.7 pts	4.7 pts
Income: $25-40K	-0.4 pts	4.6 pts	3.4 pts
Income: $40-50K*	+1.0 pts	7.5 pts	6.5 pts
Income: $50K+	+1.8 pts	3.5 pts	3.2 pts
Income: $50-75K	-1.4 pts	7.2 pts	4.5 pts
Income: $75-100K*	-2.5 pts	5.4 pts	3.7 pts
Income: $100K+	**+10.2 pts**	**4.2 pts**	**4.9 pts**
Region: Northeast	+0.9 pts	10.1 pts	4.4 pts
Region: Midwest	+1.9 pts	3.3 pts	3.6 pts
Region: South	-1.6 pts	5.4 pts	4.6 pts
Region: West	**-9.2 pts**	**4.4 pts**	**2.5 pts**
Race: White	-2.4 pts	3.2 pts	2.8 pts
Race: Black*	-9.6 pts	10.3 pts	4.5 pts
Race: Hispanic*	-12.6 pts	7.6 pts	6.8 pts
Party ID: Republican	-7.8 pts	6.3 pts	3.6 pts
Party ID: Democrat	+6.4 pts	5.1 pts	5.0 pts
Party ID: Independent	**-6.2 pts**	**2.9 pts**	**1.8 pts**
Education: Less than H.S.*	-11.7 pts	7.8 pts	5.7 pts
Education: H.S. graduate	**-7.2 pts**	**2.4 pts**	**2.3 pts**
Education: Some college+	+3.3 pts	3.7 pts	5.0 pts
Home: Owned	-2.5 pts	3.3 pts	3.7 pts
Home: Rented	+1.1 pts	3.8 pts	3.1 pts
Relationship status: Single	-2.4 pts	2.5 pts	5.2 pts
Relationship status: Married	-4.4 pts	2.4 pts	2.3 pts
Relationship status: Sep./Wid./Div.	+2.1 pts	6.0 pts	4.2 pts
Employment: Full-time	-1.8 pts	5.5 pts	1.9 pts
Employment: Part-time	-4.3 pts	5.4 pts	3.1 pts
Employment: Not at all	-2.0 pts	3.0 pts	4.5 pts
Average:	**4.0 pts**	**5.1 pts**	**4.1 pts**

* indicates that the average weekly sample size in at least one of the sample types is <100

For the four remaining groups (highest income earners, Westerners, those with a high school diploma and political independents), we conducted z-tests to determine whether responses to any of the three components of the CCI significantly differed by sample type

during each week of the test period (the same approach shown in Table 2). We found only one statistically significant difference in all 96 comparisons – on March 18, 2012, significantly more political independents in the LL sample rated the national economy positively than in the mixed sample.

No other comparison was statistically significant. For wealthy respondents, Westerners and those with a high-school diploma, responses to the three CCI questions in the mixed sample did not significantly differ from responses in the LL sample in any of the eight weeks tested. For independents, this was true in seven of the eight weeks, and in that one remaining week, just one of the three questions showed a significant difference.

Additionally, aggregating across all subgroups, the average discrepancy between the CCI estimates in the two samples (4.0 points) was no higher than the average week-to-week within-sample deviation averaged across subgroups. Thus, taken as a whole, the difference in methodology did not produce any changes in subgroup CCI values above what would be expected in normal week-to-week variation.

Finally, the average subgroup standard deviation in the mixed sample (4.1 points) is smaller than the standard deviation in the LL sample (5.1 points). This means that on average the subgroup estimates produced by the mixed sample were more stable than the estimates produced by the LL sample. This is likely due to the fact that the sample sizes of several subgroups (e.g., 18- to 34-year-olds, blacks, Hispanics and those with less than a high school education) are larger, and therefore more stable, in the mixed sample than in the LL sample.

We explore the quality of the two samples in further detail in the next section.

Comparing the Quality of the LL and Mixed Samples

Our finding that variation in the CCI by sample type is no more than would be expected by random chance can support a conclusion to switch, or not to switch, to a mixed sample. No significant impact on the data means it doesn't matter either way.

However, as noted, LL samples suffer from considerable and growing noncoverage, especially of groups that are particularly likely to lack landline service, such as young adults, renters, lower-income respondents, minorities and those with less education. LL samples can account for this non-coverage through weighting, but that can lead to a substantially higher design effect.

In this section, we compare the unweighted demographic composition of the two samples, finding that that the quality of the mixed sample indeed is superior to that of the LL sample. Table 4 shows the differences between the unweighted LL and unweighted mixed sample demographics, identifying which are statistically significant. Tables 5 and 6 compare those for which there are significant differences with available benchmarks.

Table 4: Comparison of Unweighted Demographics

	LL sample	Mixed sample	Difference
Men	43.8%	48.6%	+4.8%*
Women	56.2%	51.4%	-4.8%*
Age: 18-34	9.1%	17.1%	+8.0%*
Age: 35-44	10.5%	11.6%	+1.1%
Age: 45-54	17.7%	19.3%	+1.6%
Age: 55-64	24.4%	21.4%	-3.0%*
Age: 65+	35.3%	29.1%	-6.2%*
Income: <$15,000	11.0%	12.3%	+1.3%
Income: $15,000-25,000	11.1%	11.7%	+0.6%
Income: $25,000-40,000	15.1%	15.3%	+0.2%
Income: $40,000-50,000	8.4%	7.8%	-0.6%
Income: $50,000+	40.3%	39.8%	-0.5%
Income: $50,000-75,000	13.7%	13.9%	+0.2%
Income: $75,000-$100,000	9.6%	9.6%	--
Income: $100,000+	16.2%	15.2%	-1.0%
Region: North East	20.8%	19.8%	-1.0%
Region: Midwest	24.3%	23.4%	-0.9%
Region: South	37.9%	37.6%	-0.3%
Region: West	17.0%	19.2%	+2.2%
Race: White (non-Hispanic)	80.1%	75.0%	-5.1%*
Race: Black (on-Hispanic)	9.3%	9.6%	+0.3%
Race: Hispanic	4.9%	9.1%	+4.2%*
Party ID: Republican	28.1%	26.9%	-1.2%
Party ID: Democrat	32.0%	31.8%	-0.2%
Party ID: Independent	33.7%	34.3%	+0.6%
Education: Less than high school	7.7%	9.5%	+1.8%
Education: High school graduate	27.4%	27.0%	-0.4%
Education: Some college+	64.1%	62.8%	-1.3%
Education: College grad+	38.1%	35.9%	-2.2%
Home: Owned	79.8%	73.2%	-6.6%*
Home: Rented	19.3%	25.5%	+6.2%*
Relationship status: Never married	15.2%	21.2%	+6.0%*
Relationship status: Married	55.6%	52.6%	-3.0%
Relationship status: Sep./Div.	13.0%	12.6%	-0.4%
Relationship status: Widowed	14.9%	12.2%	-2.7%*
Employment: Full-time	35.4%	39.6%	+4.2%*
Employment: Part-time	12.1%	13.4%	+1.3%
Employment: Not at all	52.1%	46.7%	-5.4%*

Note. Bolded variables are weighting variables. *Indicates statistical significance at $p < .05$.

There are statistically significant differences between the unweighted LL and unweighted mixed sample demographics in 13 of the 38 variables studied: sex, age (18-34, 55-64, and 65+), race (white and Hispanic), home ownership, relationship status (never married and widowed) and employment (excluding part-time).

Weighting variables. For the variables that are used in weighting (sex, age, and race) we can compare unweighted data with the Census-targets used in weighting the data to determine which sample better represents the national population.[viii] In every case, the mixed sample more closely approximates the Census targets than does the LL sample.

The largest difference is the number of 18- to 34-year-olds in each sample. For LL, 9.1 percent of the unweighted sample was 18-34, vs. 17.1 percent in the mixed sample. These estimates are 21.4 and 13.3 points below Census estimates, respectively. This makes sense: young adults are difficult to reach, especially via landline. Therefore, while still underestimating the true number of young adults in the national population, the mixed sample provides a much better estimate because it includes cell phones. Indeed, as noted, the number of 18- to 34-year-olds interviewed using the LL method now frequently dips below the customary $n = 100$ threshold for reporting results. The sample size of young adults consistently rises above $n = 100$ in the mixed-sample method.

The LL sample also underestimates men and the number of Hispanics, while overestimating women, people age 55 and up and whites. The mixed sample also underestimates Hispanics, and overestimates adults age 55 and up and whites, but does so to a much smaller degree (see Table 5). In contrast to the LL sample, the gender composition of the unweighted mixed sample exactly matches the Census targets.

Table 5: Differences Between Unweighted Estimates and Census Targets

	LL sample		**Mixed sample**	
	Census targets	**Difference from unwgtd.**	**Census targets**	**Difference from unwgtd.**
Men	48.6%	4.8% pts.	48.6%	0 pts.
Women	51.4%	4.8% pts.	51.4%	0 pts.
Age: 18-34	30. 5%	21.4% pts.	30.4%	13.3% pts.
Age: 55-64	15.8%	8.6% pts.	15.8%	5.6% pts.
Age: 65+	16.7%	18.6% pts.	16.7%	12.4% pts.
Race: White (non-Hispanic)	66.9%	13.2% pts.	66.9%	8.1% pts.
Race: Hispanic	13.9%	9.0% pts.	13.9%	4.8% pts.

Non-weighting variables. For non-weighting variables on which there were significant differences between the two unweighted samples (home ownership, relationship status and employment status) we can also compare with Census data, albeit less precisely.

The March 2011 Current Population Survey (CPS) measures marital status, home ownership and employment status. Its questions on the first two are roughly equivalent to the CCI's. For employment status, however, the CPS does not distinguish between full-

time and part-time employment, so the only directly comparable category is "not at all employed" These comparisons are shown in Table 6.

Table 6: Differences Between CPS and Unweighted Estimates

	LL sample		Mixed sample	
	CPS target	**Diff. from unwgtd.**	**CPS target**	**Diff. from unwgtd.**
Home: Owned	68.9%	10.9 pts.	68.6%	4.6 pts.
Home: Rented	30.3%	11.0 pts.	30.1%	4.6 pts.
Relationship status: Never married	27.1%	11.9 pts.	27.0%	5.8 pts.
Relationship status: Widowed	6.1%	8.8 pts.	6.1%	6.1 pts.
Employed: Not at all	40.1%	12.0 pts.	40.1%	6.6 pts.

For each of the five categories that significantly differ by sample type, the unweighted mixed sample provided a better estimate of the national population than did the LL sample. Home ownership rates were overestimated in both samples, but by more than twice as much in the LL as in the mixed sample. Both unweighted samples underrepresented single Americans and overestimated not-employed Americans, but the LL sample did so by roughly twice as much as the mixed sample.

In summary, the unweighted sample obtained using the mixed-frame design is far superior to the sample obtained using the LL design. In every case in which demographic estimates differed by sample type, the mixed estimate more closely matched true population parameters than the LL estimate.

Design Effects

A design effect is an estimate of the impact on sampling error caused by a survey's departure from simple random sampling. A significant contributor to design effect is the weighting used to adjust for bias caused by non-response and noncoverage. The larger the bias, the more work is done by weighting adjustments, the larger the design effect, and thus the greater the margin of sampling error.

Given the greater bias in its estimates compared with Census benchmarks, the LL sample in this study has an average design effect of 2.3, while the mixed sample has an average design effect of 1.5. Assuming a 50-50 division in opinion, this means that the margin for sampling error for 1,000 interviews in the LL sample would be plus or minus 4.7 percentage points, vs. 3.8 points in the mixed sample. For a sample of 500, it would 6.6 points LL vs. 5.4 points mixed. And for a sample of 100, it would be 14.9 points LL vs. 12.0 points mixed.

Spanish Interviews

As noted, we completed 82 Spanish-language interviews in this study. The CCI value for this group was -62.2, compared with a CCI of -43.0 using English-only Hispanic interviews.[ix] Including Spanish interviews moves the CCI to -53.0 among Hispanic

respondents. This suggests that estimates of consumer confidence among Hispanics may be overstated using only English-language interviews.

Given that Spanish language interviews make up less than 3 percent of the mixed sample, their inclusion does not significantly influence the full sample estimate of the weekly CCI. While the 0.7-point difference in the test described above is small, the latter Spanish-inclusive CCI is a preferable estimate on theoretical grounds.

Discussion

This review finds no material, consistent impact of including cell-phone and Spanish interviews in the Bloomberg Consumer Comfort Index. The overall week-to-week pattern of results remains similar, and any difference in the overall CCI by sample type is less than the normal weekly variation. Moreover, the demographic composition of the unweighted mixed sample provides a truer estimate of the national population.

Possible impacts among subgroups are more difficult to identify, given the challenges in differentiating between possible methodological influences and normal variability in smaller samples. But the fact that the CCI differed less on average between sample types than it did within samples suggests that most of the differences are attributable to customary sampling variation, not methodology.

One potential negative impact of a switch to mixed sampling is that cell-phone sampling tends to achieve a lower response rate than landline sampling. However, extensive academic literature has established that response rates in and of themselves are a poor indicator of data quality. Meanwhile there are clear positives in mixed-sample designs: Far fuller coverage of the target population, and the face validity that provides; and, through larger sample sizes of undercovered groups, a lower design effect due to weighting and thus a lower margin of sampling error.

Endnotes

[i] One may question whether controlling for demographic differences is the best approach in such a test. However an informed critique is not possible, as the publicly available materials on the Michigan test are insufficient for comprehensive analysis. See http://www.oecd.org/dataoecd/15/45/43852284.pdf.

[ii] In this approach two separate, overlapping sample frames are used to interview the population of interest - 1) a randomly generated sample of landline telephone households and 2) a randomly generated cell-phone sample. Dual-frame sampling provides greater coverage of the target population by including individuals who can only or mainly be reached on cell phones along with those reachable via landlines. The approach requires adjustment for the greater probability of selection of respondents who can be reached on both phone types, with dual landline/cell users weighted to 0.5. The sample also is weighted to reflect the known distribution of cell-phone-only, landline-only and dual-user adults, as well as to Census norms for demographic variables.

[iii] All assessments of statistical significance in this report are at the 95% confidence level. For all comparisons we computed both an independent samples z-test as well as a z-test that accounts for the overlap in samples. We report the more conservative of these estimates throughout the paper. In Tables 1 and 2, where we want to be confident that no differences by sample type exist, we only report no significant differences when none was found in either independent or dependent z-tests. In Table 4, where we want to

be confident that the differences we report do exist, we report differences only where they are statistically significant in both z-tests. All z-tests also accounted for design effects of weighting. In the aggregated samples the design effects are 2.3 for the LL sample, 1.5 for the full mixed sample and 1.8 for the English interview mixed sample.

[iv] Theoretically the CCI can range from -100 to +100. As endpoints Figure 1 uses its actual record low and record high in 26 years of weekly data, -54 and +38.

[v] The 1.6- and 2.3-point differences reported on p.4 are based on averaging across all cases during the eight-week period for each sample type, calculating the CCI value, and subtracting. The 1.6- and 2.0-point difference is based on calculating the CCI for each week for each sample, calculating the difference between samples each week, and then averaging those differences.

[vi] As noted, some of the difference between the full mixed-sample CCI and the LL CCI was due to the inclusion of Spanish language interviews. Therefore if analyses show no differences between the full mixed sample and the landline sample we can be sure there are no differences between the English-only mixed sample and the LL sample.

[vii] Z-tests again accounted for the design effect, which varied slightly from week-to-week. For the LL data, the design effect ranged from 2.1 to 2.6 across the eight-week period (M = 2.3) and for the mixed data the design effect ranged from 1.4 to 1.5 (M = 1.5).

[viii] There are slight differences in the Census targets for these variables by sample due to the fact that weighting targets are adjusted for missing data. Census targets were computed by simply multiplying the Census estimates by the percent of respondents who gave a response to the question (e.g., if 98.6 percent of respondents answered the age question, the CPS estimates for age were multiplied by .986). Slight differences in refusal rates for the two samples can result in slight differences in the final targets.

[ix] The estimate for the English-only Hispanic CCI value was calculated using the properly weighted English-only mixed sample dataset.