

State-Level Weighting Update:
Why the Consumer Expenditure Surveys Program
Is No Longer Providing State Weights for New Jersey

August 2023

Sharon Krieger and David Swanson

Consumer Expenditure Surveys Program Report Series

Introduction

The Bureau of Labor Statistics (BLS) Consumer Expenditure Surveys (CE) program recently began providing state weights to its public use microdata users for five states. It began providing them for California, Florida, and New Jersey for data collected in 2016; and for New York and Texas for data collected in 2017. The goal was to expand the universe of geographic areas for which statistically valid expenditure estimates could be computed. Before then, statistically valid expenditure estimates could be computed at the national, regional, and divisional levels, but not at the state level.¹

Then in 2016, the CE program developed a methodology for calculating state weights that allowed statistically valid expenditure estimates to be made at the state level. The methodology was applied to five of the more populous states in the country, including New Jersey, because the CE program collected a large amount of data in them.²

Being a new product, the CE program considered the state weights to be “experimental,” and posted them on its “Research Products” webpage while it monitored their long-term viability. Unfortunately, over time, the number of respondent households in New Jersey decreased to the point where producing weights for it was no longer viable. That led the CE program to discontinue producing state weights for New Jersey beginning with data collected in 2021. This paper explains the technical difficulties that led to that decision.

Background

The CE is a nationally representative household survey that collects expenditure data from a random sample of consumer units (CUs) across the U.S. CUs are basically the same thing as households, and the terms are often used interchangeably. However, there is a slight difference between them. Households are groups of people who live together, while CUs are groups of people who live together and share their living expenses. Households and CUs are usually the same thing, but occasionally a household will have more than one CU. Thus, CUs are slightly smaller than households on average.³

A random sample of CUs is selected for the CE survey in a two-step process. First, a random sample of geographic areas is selected from across the U.S. Those geographic areas are small clusters of counties

¹ They could also be computed for the 23 largest metropolitan statistical areas in the country (New York, Chicago, Los Angeles, etc.). “Regions” and “divisions” are broad geographic areas defined by the Census Bureau. The Census Bureau partitions the U.S. into four regions (Northeast, Midwest, South, and West), and it partitions each region into two divisions (e.g., it partitions the Northeast region into the New England and Middle Atlantic divisions), except the South, which it partitions into three divisions because its population is so much larger than that of the other regions.

² For more information on how the state weights are computed, see <https://www.bls.gov/cex/stateweights-documentation.pdf>. The methodology is also described in Susan King’s 2017 JSM paper, “Calculating State Weights for the Consumer Expenditure Survey,” at https://www.bls.gov/cex/research_papers/pdf/calculating-state-weights-for-the-consumer-expenditure-survey.pdf.

³ To be more precise, a consumer unit comprises either: (1) all members of a particular household who are related by blood, marriage, adoption, or other legal arrangements; (2) a person living alone or sharing a household with others or living as a roomer in a private home or lodging house or in permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more persons living together who use their income to make joint expenditure decisions. Financial independence is determined by the three major expense categories: Housing, food, and other living expenses. To be considered financially independent, at least two of the three major expense categories have to be provided entirely, or in part, by the respondent.

called Primary Sampling Units (PSUs), and they range in size from one county to twenty-nine counties. To be precise, they are the “Core-Based Statistical Areas” (CBSAs) defined by the Office of Management and Budget. Second, within those geographic areas, a random sample of households is selected. Then after the households are selected, every CU in those households is interviewed.

The two PSUs that will be of most concern in this paper are the New York and Philadelphia PSUs, because most of New Jersey is in them.

The CE program uses the data collected from the CUs in its sample to produce expenditure estimates. However, before producing those expenditure estimates, the CE program assigns weights to the CUs. Weights are needed to properly reflect the CUs’ probability of being selected for the sample, and their likelihood of participating in the survey.

The CUs’ weights are calculated in a three-step process. First, the CUs in the sample are assigned a base weight equal to the number of CUs in the population that they represent. If one out of 15,000 CUs in the population is selected for the sample, then a CU in the sample is assigned a base weight of 15,000. Second, the base weights are increased to account for nonresponse. If only one-third of the sampled CUs respond to the survey, then the base weights of the respondent CUs are multiplied by a factor of 3.0 to account for the nonrespondents. That makes the nonresponse-adjusted weights equal to 45,000. And third, the nonresponse-adjusted weights are adjusted by a “calibration” process to make the respondents’ weights add up to the Current Population Survey’s (CPS’s) estimates of the total number of people and households in the population. These calibration-adjusted weights are called final weights. If a respondent’s final weight equals 50,000, that means it represents 50,000 CUs in the population.

The calibration adjustment process will be of concern in this paper because the decreasing number of respondents in New Jersey has been causing that step of the weighting process to fail in recent years.

Calculating the State-Level Weights – Overview

At the state level, the CUs’ weights are calculated in the same three-step process. First, the CUs in the sample are assigned a base weight equal to the number of CUs in the population that they represent. Second, the base weights are increased to account for nonresponse. And third, the nonresponse-adjusted weights are adjusted by a calibration process to make the respondents’ weights add up to CPS’s estimates of the total number of people and households in the population. Some minor changes were made to the first two steps to make the CUs’ base weights and nonresponse-adjusted weights reflect the state in which they are located, rather than the region and division of the country. However, the third step, calibration, will be the focus of this paper, because it is where the difficulties for New Jersey occurred.

Calibration is a mathematical procedure that modifies the nonresponse-adjusted weights of the respondents to make them add up to CPS’s estimates of the total number of households in the population. There are infinitely many sets of weights that add up to CPS’s population counts, and calibration uses nonlinear programming to find the set of weights that is as close as possible to the nonresponse-adjusted weights while also making them add up to CPS’s population counts. It is an iterative procedure which makes a sequence of guesses at the weights, with each guess being closer to the nonresponse-adjusted weights than

the previous guess. When the guesses stop improving, the procedure is said to have “converged,” and the final guess is the best set of weights.

Calibration is a quarterly process that is performed on quarterly data. At the national level it makes the respondents’ weights add up to 35 CPS population counts, which includes households in the nation, people in the nine Census divisions (plus nine more counts for those in the urban parts of those divisions), plus seventeen demographic subgroups for age, race, ethnicity (Hispanic, Non-Hispanic), and tenure (homeowner, renter). At the state level it makes the respondents’ weights add up to 11 CPS population counts, which includes households in the state, people in the state, plus nine demographic subgroups for age, ethnicity, and tenure. The number of demographic subgroups was reduced at the state level to make sure there were enough respondents for the calibration process to converge. With 11 population counts, it was empirically observed that the calibration process needs approximately 100 respondents per quarter to converge.

Unfortunately, in recent years, the calibration step in the weighting process began to fail for New Jersey, even with its reduced number of population counts. The next section describes what led to its failure.

The Problem with New Jersey Data – Decreasing Participation

In recent years, calculating state weights for CUs in New Jersey became difficult because of its decreasing number of respondents. Between 2010 and 2021 the number of respondents in New Jersey decreased 50 percent in both the Interview and Diary Surveys. This decrease can be attributed to two things: (1) the changing definitions of the New York and Philadelphia PSUs; and (2) decreasing response rates. Both of these things decreased the number of respondents in New Jersey.

The changing definitions of the New York and Philadelphia PSUs decreased the number of respondents in New Jersey by decreasing the initial number of CUs in its sample. The CE Survey’s sample size is fixed at the national level due to budget constraints, and the initial number of CUs in the sample at the national level is allocated to the individual PSUs in the sample proportional to their populations. That means if a PSU’s definition changes in a way that causes it to have fewer people, then the PSU will be assigned fewer CUs for its sample. That is what happened in New Jersey.

New Jersey is right next to New York and Philadelphia, so most of its counties are in those two PSUs. In fact, before 2015, every county in New Jersey was in the New York and Philadelphia PSUs. That meant all 21 New Jersey counties were in the CE’s sample. However, in 2015, the definitions of the New York and Philadelphia PSUs changed. That caused five New Jersey counties to be dropped from the sample. Those five counties were moved into other newly created PSUs that were not selected for the sample. When that happened, the number of counties in New Jersey in the CE’s sample decreased from 21 to 16, and that loss of 5 counties decreased the number of CUs assigned to its sample by 10 percent.

In addition, decreasing response rates resulted in fewer respondents in the New Jersey counties that remained in the New York and Philadelphia PSUs. That decreased the number of respondents by 40 percent.

Decreasing response rates is a universal problem in the survey world. Its exact cause is not known, but for our purposes we simply note that it is a universal problem.⁴

So, the changing PSU definitions reduced the number of CUs in the sample by 10 percent, and the decreasing response rate reduced the number of respondents in the sample by 40 percent. Together, they decreased the number of respondents in New Jersey by 50 percent.

Figure 1 below shows the decreasing annual response rates for New Jersey, from 2010 through 2021.

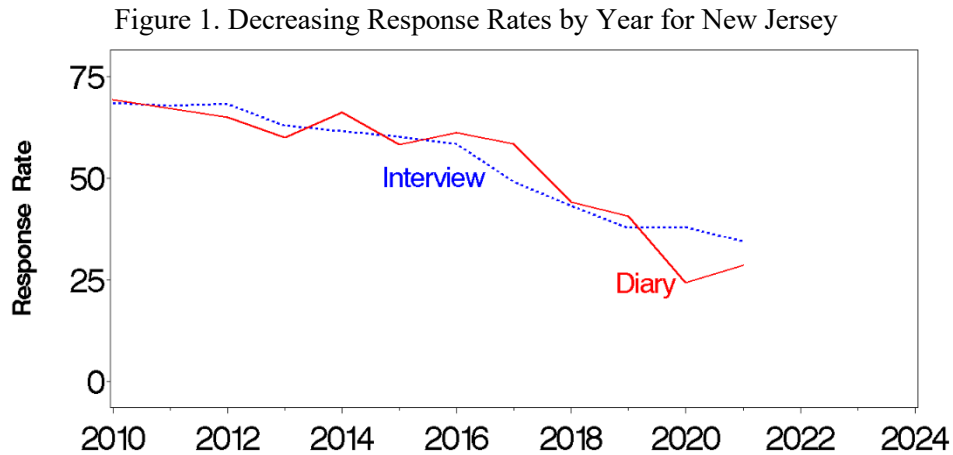
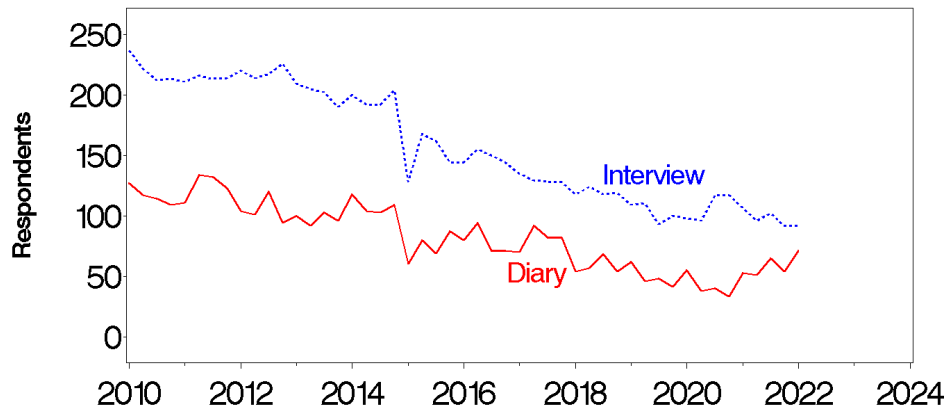


Figure 2 below shows the decreasing number of quarterly respondents in New Jersey, from the first quarter of 2010 through the first quarter of 2022. This graph shows quarterly numbers instead of annual numbers because the problem in state weights occurred in the calibration adjustment process, and that is a quarterly process.⁵

⁴ Its universality in the CE Survey can be seen in Table 1 at the end of this report, where the number of respondents for the two surveys (Interview, Diary), and the five states displayed (California, Florida, New Jersey, New York, Texas), all show significant decreases between 2010 and 2021. New Jersey's decreases are greater than other states, but they all show significant decreases in the number of respondents.

⁵ For more details, and to compare New Jersey's numbers to other states, see Table 1 at the end of the report for the quarterly counts of respondents in New Jersey since 2010, along with those in the other four states.

Figure 2. Decreasing Number of Respondents by Quarter for New Jersey



Procedural Changes to Account for the Decreasing Number of Respondents in New Jersey

As Figure 2 above shows, between 2010 and 2021 the number of respondents in New Jersey decreased from 200 CUs per quarter to 100 CUs per quarter in the Interview Survey, and from 100 CUs per quarter to 50 CUs per quarter in the Diary Survey. And as mentioned above, our empirical experience shows the calibration process needs around 100 CUs per quarter to converge. That is the reason the calibration process had difficulty converging in the Diary Survey in the last few years, and the reason it was starting to have difficulty converging in the Interview Survey too.

As mentioned above, the number of demographic subgroups in the calibration process was reduced from 35 at the national level to 11 at the state level to make sure they had enough CUs for the calibration process to converge. However, with only 50 CUs responding per quarter in the Diary Survey, it was no surprise that some demographic groups had no CUs, or very few CUs, in them. This caused the calibration process to fail to converge. It created a problem where the weights of the CUs did not add up to the CPS population counts for one or more demographic subgroups, hence an optimal set of weights was not obtained. And with only 100 CUs responding per quarter in the Interview Survey, it was no surprise that the same problem was starting to occur in the Interview Survey as well.

The CE program circumvented this problem in recent years by making ad hoc adjustments to the calibration process. Those ad hoc adjustments began by further reducing the number of demographic subgroups. That meant continuing to collapse demographic subgroups until the collapsed subgroups had enough CUs for the process to converge. The subgroups that were collapsed changed from quarter to quarter, depending on the characteristics of the survey respondents, which is what made the procedure ad hoc.

Unfortunately, even with this additional collapsing, the calibration process sometimes still failed to converge. That led the CE program to also change the lower bound or upper bound on the calibration adjustment factors. That was done in an ad hoc way too, depending on the diagnostics from the nonlinear programming software. Ideally, the same lower and upper bounds would be used for both national and state weights, but sometimes state weights required a smaller lower bound or a larger upper bound than national weights to converge. These two ad hoc procedures indicate that the calibration process was at the limits of its capabilities for New Jersey.

Summary of Adjustments Used for New Jersey for Data Collected in 2019 and 2020

So, two ad hoc adjustments were used to make the calibration process converge: (1) collapsing the 11 demographic subgroups down to a smaller number of demographic subgroups; and (2) changing the lower or upper bounds on the calibration adjustment factors. These adjustments were done in an iterative way until the calibration process converged.

Table 2 below summarizes the adjustments that were made in the Diary Survey on a quarter-by-quarter basis for data collected in 2019 and 2020. During that two-year period, demographic subgroups were collapsed in all eight quarters, and the lower bound was decreased or the upper bound was increased in three of the eight quarters.

Table 2. Ad Hoc Adjustments Made in the Diary Survey for New Jersey

Quarter	Collapsed Subgroups	Changed Bounds
2019Q1	✓	
2019Q2	✓	
2019Q3	✓	
2019Q4	✓	✓
2020Q1	✓	
2020Q2	✓	
2020Q3	✓	✓
2020Q4	✓	✓

The adjustments typically began by collapsing some of the age subgroups together. Then if more collapsing was needed, other subgroups were collapsed together, such as ethnicity or tenure. And if that was still not enough, then the lower or upper bounds were changed.

Ideally, the lower bound or upper bound should never be hit during the calibration process, but unfortunately, it happened for CUS in New Jersey. When it happened, the CE program decided to accept the results, provided the lower bound was not hit by more than 10 percent of all CUs, and the upper bound was not hit by more than 10 percent of all CUs. If a lower or upper bound was hit by more than 10 percent of all CUs, then the lower bound was decreased or the upper bound was increased, respectively.

For a while these ad hoc adjustments gave the calibration process the help it needed to converge, and the results were accepted. And sometimes there was more than one way of collapsing the demographic subgroups together to help the process converge. However, this led to the question – which set of weights were the right weights? The CE program observed that the results were sensitive to which demographic subgroups were collapsed together, and even small changes in the adjustments resulted in large changes to the weights. This added to the problem.

Overall, the number of problems kept increasing as the number of respondents kept decreasing. And with the long-term downward trend in response rates, it looked like the number of respondents would continue decreasing for the foreseeable future. This is when the CE program realized that producing the weights for New Jersey was no longer viable.

Conclusion

To summarize, calculating state weights for CUs in New Jersey became difficult in recent years because of the decreasing number of respondents. The decreasing number of respondents was primarily due to decreasing response rates, but the changing definitions of the New York and Philadelphia PSUs also contributed to it. That decrease caused some of the demographic subgroups used in the calibration process to have no CUs, or very few CUs, in them, which caused the calibration process to fail to converge.

For several years the CE program circumvented this problem by reducing the number of demographic subgroups at the state level, and sometimes by also changing the lower and upper bounds on the calibration factors. The number of demographic subgroups was initially reduced from 35 at the national level and down to 11 at the state level to make sure every demographic subgroup had enough CUs for the calibration process to work. However, due to the decreasing number of respondents, eventually the 11 demographic subgroups were too many for New Jersey, and the CE program had to reduce the number of them even further.

The reduction had to be done in an ad hoc way because the specific demographic subgroups that had no CUs, or very few CUs, changed from quarter to quarter. That lack of consistency in the process was undesirable from both a conceptual viewpoint and a computer programming maintenance viewpoint. Also, it slowed down the data generation process, which delayed the completion of “downstream” activities like data publication.

In addition, the CE program observed that the resulting weights for New Jersey’s respondents were sensitive to these procedural changes when the number of demographic subgroups went below 11. Even small changes in the demographic subgroups resulted in large changes to the weights, and there was no way of knowing which set of weights was right. So, in summary, although the ad hoc procedures worked for several years, the CE program ultimately decided that producing state weights for New Jersey is no longer viable. As a result, the CE program is discontinuing the production of state weights for New Jersey beginning with data collected in 2021.

Table 1. Number of Respondents in the CE Survey by State, 2010 – 2021

Quarter	Interview Survey					Diary Survey				
	CA	FL	NJ	NY	TX	CA	FL	NJ	NY	TX
2010Q1	780	398	237	470	522	385	220	127	283	239
2010Q2	810	421	222	446	531	381	189	117	247	311
2010Q3	781	414	212	442	545	373	206	114	253	251
2010Q4	743	420	213	451	527	338	213	109	237	263
2011Q1	739	429	211	455	519	384	224	111	224	252
2011Q2	715	398	216	451	528	399	203	134	231	256
2011Q3	720	375	213	428	509	335	268	132	267	258
2011Q4	737	379	214	454	534	337	217	123	244	250
2012Q1	733	385	220	449	538	367	202	104	228	248
2012Q2	739	397	214	427	535	394	180	101	221	263
2012Q3	729	396	217	435	517	329	233	120	241	237
2012Q4	729	406	226	405	515	375	208	94	230	261
2013Q1	755	403	209	386	536	328	203	100	219	253
2013Q2	766	434	205	374	533	316	206	92	225	244
2013Q3	734	421	202	339	517	269	217	103	166	200
2013Q4	708	387	190	328	473	303	159	96	176	216
2014Q1	779	428	200	360	517	383	205	118	214	260
2014Q2	777	404	192	379	481	341	231	104	232	264
2014Q3	757	410	192	363	504	360	198	103	232	258
2014Q4	740	410	204	381	517	370	209	109	217	234
2015Q1	464	270	128	278	333	335	163	60	181	207
2015Q2	708	406	168	436	483	337	133	80	212	193
2015Q3	689	387	162	423	463	331	171	69	195	169
2015Q4	711	365	144	402	449	309	200	87	205	181
2016Q1	723	354	144	398	512	323	152	80	227	234
2016Q2	675	381	155	409	477	353	153	94	223	206
2016Q3	660	391	150	410	507	346	147	71	210	236
2016Q4	652	394	144	382	468	282	164	71	211	171
2017Q1	652	396	135	388	469	286	153	70	175	210
2017Q2	643	396	129	371	463	315	170	92	203	232
2017Q3	639	339	128	381	432	283	141	82	186	196
2017Q4	601	329	128	372	449	285	171	82	190	205
2018Q1	600	331	118	386	429	322	147	54	183	191
2018Q2	630	332	124	369	402	318	160	57	165	225
2018Q3	619	343	118	338	403	316	197	68	162	214
2018Q4	624	317	119	330	378	323	132	54	158	188
2019Q1	589	312	109	347	413	307	167	62	144	198
2019Q2	582	331	110	325	394	304	167	46	139	176
2019Q3	577	324	93	304	388	306	167	48	145	200
2019Q4	543	333	100	297	356	289	189	41	111	149
2020Q1	548	308	98	320	321	366	227	55	144	243
2020Q2	551	246	96	307	272	274	89	38	119	101
2020Q3	508	264	117	329	286	299	118	40	144	143
2020Q4	537	326	117	326	307	353	174	33	145	186
2021Q1	546	345	107	254	284	302	193	53	133	190
2021Q2	574	362	96	236	305	418	220	51	145	209
2021Q3	598	327	102	263	318	346	175	65	178	191
2021Q4	538	305	92	280	287	305	190	54	160	195
2022Q1	606	296	92	288	341	407	203	71	212	232
Avg. 2010 (Q1 – Q4)	779	413	221	452	531	369	207	117	255	266
Avg. 2021 (Q1 – Q4)	564	335	99	258	299	343	195	56	154	196
% Change, 2010 - 2021	-28%	-19%	-55%	-43%	-44%	-7%	-6%	-52%	-40%	-26%