Technical Document: An Approach to Geographic Aggregations of County Statistics

Brian Quistorff
Brian.Quistorff@bea.gov
February 2026

BEA first announced on its public website in 2025 and subsequently in its February 5, 2026, news release "Gross Domestic Product by County and Personal Income by County, 2024" that it will no longer produce gross domestic product (GDP) and personal income estimates for Metropolitan Statistical Areas (*MSAs*), Micropolitan Statistical Areas (*MICs*), Metropolitan Divisions (*DIVs*), Combined Statistical Areas (*CSAs*), and metropolitan and nonmetropolitan portions (*PORTs*). Publishing estimates for these geographic aggregates is time-intensive due to (a) identifying estimates for suppression to ensure data privacy, and (b) re-aggregating county statistics across the entire time series to account for the changing definitions of these geographic aggregates as determined by the Office of Management and Budget (OMB). Additionally, discontinuing these geographic aggregates allows for more county statistics to be unsuppressed.

These "geographic aggregates" can be useful for users. This document provides information about how users can derive, on their own, statistical approximations for metro areas and other geographic aggregates via an aggregator tool that uses available county statistics and national industry-level price data. These derived estimates for geographic aggregates, however, are of lesser quality than the official statistics produced by BEA. The quality differs due to many factors: the published industry price and GDP data for the derived estimates are less detailed than the data used for official calculations; the public county data for current dollar GDP contains suppressions; there is no workaround for cases where standard price indexing fails; and the use of rounded published data for counties and national prices. This document notes when this procedure cannot derive estimates and when provided estimates may differ from BEA-published estimates. Accompanying source code is also provided.

2. Methodology

In this section, the structure of geographic aggregation, the separate estimation procedures for GDP and personal income, and, finally, some common numerical processing standards are addressed.

Geographic aggregation structure

Each aggregation type defines a sequence of aggregation layers where components of the previous layer are aggregated to the next layer. The sequence of layers is:

- MSA: Counties (in MSAs) → MSAs → US Metro portion

- MIC: Counties (in MICs) → MICs
- CSA: Counties (in CSAs) → CSAs
- DIV: Counties (in DIVs) → DIVs
- PORT: Counties (all) → State by Metro/Nonmetro portions → BEA Region by Metro/Nonmetro portions → US Metro/Nonmetro portions.

OMB defines MSAs, MICs, CSAs, and DIVs[1]. BEA uses a slightly modified definition of counties and has custom region definitions[2].

Gross Domestic Product

There is a "canonical" process (following essentially what BEA does for official statistics) that would be used when the county-level data on GDP in current dollars does not contain suppressions. For suppressions, the tool takes a conservative approach, so that any calculation that uses a suppressed value results in a suppressed value. The canonical process starts with the fine-grained data, county by "leaf industry" (industries that have no children in the published tables), and aggregates up to all other estimates. While this is the most accurate way to calculate "real" estimates (quantity indexes for real GDP, real dollar GDP, and contributions to percent change in real GDP), when following this process with published estimates that contain suppressions, these suppressions propagate through the calculations causing many higher-level estimates to be suppressed as well. BEA's county estimates, though, often provide unsuppressed estimates at higher-levels of industrial classification even when leaf-industry estimates are suppressed. Therefore, a modified process is used that takes advantage of the availability of these higher-level industry estimates in county data to reduce the number of propagated suppressions.

Current dollar GDP for geographic aggregates is calculated as the sum of those values for the same industry in the component counties (BEA Regional table CAGDP2).

Canonical Process:
Step 1: Start with the following source data:

- BEA's Regional "CAGDP2 GDP by industry in current dollars" (annual) table for counties
- BEA's GDP-by-Industry "Chain-Type Price Indexes for Value Added by Industry" annual table.

They are then linked by industry and year.

Real estimates are most accurate when aggregated up from the smallest geography-by-industry cell, which would be county-by-"leaf industry". Loop over each aggregation (A) needed to

[1] See the latest definition in bulletin no. 23-01 issued July 21, 2023.

[2] See https://apps.bea.gov/regional/geography.htm for definitions.

calculate (each industry for each geographic aggregation layer above counties) and perform the following, using the data for component child cells (all 'leaf industries-county' $i \in A$):
Step 2: Sum current dollar GDP (e_{it}) in the cells for each time period, to get the aggregate current dollar GDP e_{At}.
Step 3: For each cell i and time period t, calculate $q_{it} = e_{it}/p_{ind,t}$ where $p_{ind,t}$ is the national price index in time t for industry ind. Then calculate the four quantities: $p_t q_{it}$, $p_{t-1} q_{it}$, $p_t q_{it-1}$, $p_{t-1} q_{it-1}$. Sum these four quantities for the aggregate as E_{Att}, $E_{A,t-1,t}$, $E_{A,t,t-1}$, $E_{A,t-1,t-1}$. Then construct the relative quantity indexes for the aggregate, $Q_A^L = E_{Att}/E_{A,t,t-1}$, $Q_A^P = E_{A,t-1,t}/E_{A,t-1,t-1}$, and the Fischer $Q_{At} = \sqrt{Q_A^L \times Q_A^P}$. The chain-type quantity index I_{At} is set as (a) 100 if t is the reference year (currently 2017), (b) $I_{At} = I_{At-1} * Q_{At}$ for later years, and (c) $I_{At} = I_{At+1}/Q_{At+1}$ for earlier years. If current dollar GDP for the aggregate is 0 in the reference year, then the whole quantity index is missing.
Step 4: Real GDP in chained dollars is $r_{At} = I_{At} * e_{A,2017}$. For true-leaf industries, if this results in a missing value (e.g., due to non-positive values in the reference year), it is instead calculated as $r_{At} = e_{At}/p_{ind,t}$
Step 5: Calculate this aggregate's contributions towards growth in "All industry total" for the same geography (call this A'). This is done using the following formula, starting with the cell data:

$$C\%\Delta_{it} = 100 \times \frac{\frac{E_{A'tt}}{E_{A',t-1,t-1}} \cdot \left(E_{i,t-1,t} - E_{i,t-1,t-1}\right) + Q_{A't} \cdot \left(E_{itt} - E_{i,t,t-1}\right)}{E_{A'tt} + Q_{A't} E_{A't,t-1}}$$

Then calculate $C\%\Delta_{At} = \sum_{i \in A} C\%\Delta_{it}$.

Modification: When leaf data is suppressed, unsuppressed data may exist at higher level industries. Use this data when possible, to increase the number of statistics provided. For each year, county, industry aggregate combination, calculate "pseudo-leaves" that may be used: Start with an initial selection of just using the industry aggregate, and then recursively look at the set of selected elements and if it is an industry aggregate and all child industries have non-suppressed data for the county in that year and the year prior, replace that selected element with its children.

Discrepancies with BEA-published statistics: Estimates from this procedure would deviate from BEA-published estimates for three reasons:

1. Suppressions: Some estimates may result in a suppressed value due to propagation from a suppressed nominal county value, that might not have been otherwise suppressed by BEA.
2. "Fisher failures": A zero nominal value for an industry for a geographic aggregate will result in a Fisher quantity index of 0 that will then propagate zeros away from the reference period even when other periods have positive nominal values. In the BEA, various methods are used to deal with these failures.

3. Prices and nominal GDP at published industrial levels are at a coarser level of industrial granularity than would be used internally in BEA as they use more detailed unpublished price indexes.
    a. Additionally, the modification to find pseudo-leaves will use prices at even more aggregated level causing additional discrepancies.
4. Intermediate rounding: Rounding at an intermediate stage will introduce noise into the final estimates that would not be present in BEA-published estimates. These will typically be quite small for current dollar as it is a simple addition, but larger for the other calculations that involve more complex and longer chains of calculation.

Personal Income

For each of the Personal Income tables (CAINC1, CAINC4, CAINC5N, CAINC6N, CAINC30) start with data for counties. To calculate the estimates for aggregates, simply sum the corresponding values of the component counties. Table lines that are ratios are recalculated after the numerator and denominator have been aggregated. If data for a county are suppressed, then data for that line for the aggregate are suppressed as well.

Discrepancies with BEA-published statistics: These estimates may have discrepancies due to two sources:

1. Suppressions.
2. Intermediate rounding: These will typically be quite small for statics that involve simple summation and potentially slightly larger for those that involve ratios.

Common Computational Issues

Division by zero: By convention, estimates that result in division by zero are returned as missing and marked as "(NA)" (in the API they may be listed as 0 with a Note in the NoteRef field). For GDP tables, this includes the real estimates. For Personal Income tables this includes the ratio line-codes.

Breaks in county definitions: For counties with a break in their series (e.g., the Connecticut county-equivalent planning regions), calculations involving previous periods (the "canonical" GDP process), will be unavailable for the first year of the new series.

3. Deviations from published estimates

In this section, deviations of tool-calculated estimates from previously published official estimates are analyzed. Increases in the percentage of estimates suppressed/missing as well as the accuracy of the tool estimates in terms of how similar they are to the previous official estimates are examined.

For GDP estimates, the time-span 2017 (the last reference year) to 2023 (the last year of published results for geographic aggregates) is used.

Table 1 shows the percentage of estimates suppressed for each of the four variables for both official statistics and the tool-calculated estimates for each of the industries. Apart from the "All Industry total" and its two child industries, there were no suppressions. There was a moderate rise in suppressions in the estimates derived by the tool compared with official statistics. (There were no cases where the tool could calculate an unsuppressed estimate when the official published estimate was suppressed.)

Table 1 - Percent estimates suppressed for official published and tool calculated estiamtes

Line Code	Description	GDP in current dollars		Chain-type quantity-index		Real GDP in chained dollars		Contributions to percent change in real GDP	
		Pub.	Calc.	Pub.	Calc.	Pub.	Calc.	Pub.	Calc.
1	All industry total	0	0	0	0	0	0	0	0
2	Private industries	0	0	0	0	0	0	0	0
3	Agriculture, forestry, fishing and hunting	22.8	25.6	44.2	54.4	22.8	25.6	24	35.2
6	Mining, quarrying, and oil and gas extraction	15.4	16.7	16.6	22.8	15.4	16.7	16.1	19.9
10	Utilities	28	32.7	40.8	50.9	28	32.7	31	41.5
11	Construction	9	10.1	12.5	15.1	9	10.1	9.5	12
12	Manufacturing	8.2	9.6	11.4	13.7	8.2	13.7	8.5	11.2
13	Durable goods manufacturing	11.2	12.8	14.6	18.3	11.2	12.8	10.7	15.3
25	Nondurable goods manufacturing	11.8	13.2	15.6	18	11.8	13.2	9.4	15.3
34	Wholesale trade	25.8	33.4	33.4	41.7	25.8	33.4	18.1	37.1
35	Retail trade	2.6	2.6	4.4	5.2	2.6	2.6	2.2	3.4
36	Transportation and warehousing	26.4	39.3	34.4	47.1	26.4	39.3	22	42.8
45	Information	18.4	23.3	24	30.1	18.4	23.3	17.4	26.6
50	Finance, insurance, real estate, rental, and leasing	12.9	15.9	12.9	20.8	12.9	20.8	12.7	18.6
51	Finance and insurance	10.9	12.8	13.8	17.3	10.9	12.8	7.6	15
56	Real estate and rental and leasing	15.2	20.1	20.3	26.7	15.2	20.1	15.4	24.6
59	Professional and business services	16.4	21.9	24.7	37.5	16.4	37.5	18.2	29.6
60	Professional, scientific, and technical services	24.7	31.3	32	40.9	24.7	31.3	22.8	36.3
64	Management of companies and enterprises	33.3	42.3	40.6	50.4	33.3	42.3	24.1	46.5

65	Administrative and support and waste management and remediation services	20.8	27	26.6	35.8	20.8	27	16.7	31.5
68	Educational services, health care, and social assistance	9.7	14.5	14.3	26.6	9.7	26.6	11.6	20.5
69	Educational services	24.2	34.6	29.9	41.2	24.2	34.6	19.3	37.8
70	Health care and social assistance	23.9	35	29.4	41.7	23.9	35	13.8	38.1
75	Arts, entertainment, recreation, accommodation, and food services	11	14	17	25	11	25	12.6	19.2
76	Arts, entertainment, and recreation	17.3	21.7	22.1	29.1	17.3	21.7	15.7	24.9
79	Accommodation and food services	15.8	20.4	20	26.9	15.8	20.4	10.5	23.4
82	Other services (except government and government enterprises)	11.3	12.8	15.5	19.3	11.3	12.8	10.6	16.2
83	Government and government enterprises	0	0	0	0	0	0	0	0
87	Natural resources and mining	37.3	41	48.3	55.5	37.3	55.5	30.8	48.2
88	Trade	26.7	34.2	35.5	43.2	26.7	43.2	19.2	38.2
89	Transportation and utilities	47.8	54.8	58.6	66.1	47.8	66.1	34.8	60.2
90	Manufacturing and information	21.1	26.1	27.3	35.3	21.1	35.3	20.9	31.3
91	Private goods-producing industries	30.6	34.6	41	49.3	30.6	49.3	24.9	41.6
92	Private services-providing industries	30.6	34.6	41	49.3	30.6	49.3	25.1	41.6

Table 2 notes the "accuracy" of the tool calculated estimates (in terms of how close it was to the official estimates) for the "All Industries" line for each of the four variables, for each of the years. For the first three variables that are in "levels", the accuracy metric is "mean absolute percent deviation" (MAPD). For the final variable that is percent change, the accuracy metric is "mean absolute deviation" (MAD). There is essentially no loss in accuracy for GDP in current dollars. For chain-type quantity indexes and real GDP in chained dollars, the deviations are very similar and grow over time (errors compound through chaining). For contributions to percent change in real GDP, there is a more muted trend of increasing errors over time.

Table 2 - Accuracy of GDP calculations over time

Year	GDP in current dollars (MAPD)	Chain-type quantity-index (MAPD)	Real GDP in chained dollars (MAPD)	Contributions to percent change in real GDP (MAD)
2017	0.00	0.00	0.00	0.48

2018	0.00	0.45	0.45	0.47
2019	0.00	0.35	0.35	0.53
2020	0.00	1.01	1.01	0.84
2021	0.00	0.98	0.98	1.28
2022	0.00	1.83	1.83	1.1
2023	0.00	1.67	1.67	1.27

For personal income, the focus was on table CAINC4. The percentage of statistics missing is essentially the same as published statistics. The accuracy was quite good, with any differences less than .01% in terms of mean absolute percent deviation. This is to be expected as the deviations are due to intermediate rounding and so quite small.

Discussion

The framework laid out in this document allows users to calculate, on their own, estimates of GDP and personal income by metro areas and other geographic aggregates. If a user needs only statistics prior to 2024, then the BEA's historical archive is available as a resource, although such archived data is not revised to reflect updated national prices by industry). Mixing estimates from the historical archive and this aggregator tool is not recommended because that could create discontinuities at the 'seam' between the two sources. The aggregator tool also can be used for (a) user-defined (or customized) geographic aggregates, and (b) estimates for some pre-defined geographic aggregates that had not been available for some BEA tables (not all tables included all aggregates).