Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

Measuring Industry Fragmentation: HHI, CR4 and Census Data

> Build a more reviewable acquisition pipeline with Docket. Our deal-origination software and managed research help private equity teams screen companies against a mandate and understand the evidence behind each finding.

Published by Docket · 2026-10-29 · industry fragmentation measurement · fragmented industry private equity · hhi herfindahl · concentration ratio census

Original article: https://docket.capital/articles/industry-fragmentation-hhi-concentration-census



In this article

1. What "fragmented" should mean in a thesis
2. HHI: the arithmetic
3. Thresholds from the 2023 Merger Guidelines
4. Concentration ratios: top 4, 8, 20, 50
5. Where to find the tables and what they exclude
6. Pitfalls: market definition, geography and ownership
7. From a fragmentation number to a target list
8. Worksheet

What "fragmented" should mean in a thesis

Almost every buy-and-build memo calls its target industry fragmented. The word usually means "there are many small companies and no dominant one", and it is usually asserted rather than measured. That matters, because the roll-up logic depends on it: the thesis assumes there are enough independent, acquirable companies for a platform to buy, and that no incumbent already controls pricing.

A defensible claim of fragmentation names three things:

1. **The market**, defined by product or service and by geography (national, regional, metro).
2. **The measure**: a concentration ratio, the Herfindahl-Hirschman Index (HHI), or a firm count by size band.
3. **The source and year**, with what the source excludes.

The rest of this article covers the two standard measures, where the U.S. government publishes them, and the ways they mislead. It does not repeat firm and establishment counting, which is covered in CBP vs QCEW vs SUSB market sizing.

HHI: the arithmetic

The 2023 Merger Guidelines define the HHI as "the sum of the squares of the market shares"; it "is small when there are many small firms and grows larger as the market becomes more concentrated, reaching 10,000 in a market with a single firm" (DOJ/FTC 2023 Merger Guidelines). Shares are expressed in percentage points, so a monopolist's 100% squares to 10,000. The DOJ's HHI page gives the same definition with an example: four firms with 30, 30, 20 and 20 percent shares have an HHI of 2,600 (DOJ Antitrust Division).

The HHI rises both when the number of firms falls and when the size gap between them widens (DOJ Antitrust Division). Equal-sized firms give a simple rule: with n equal firms, HHI = 10,000 / n. The Guidelines' own illustration is five equal firms at 2,000 and six at about 1,667 (DOJ/FTC 2023 Merger Guidelines).

Illustrative examples (fictional)

Fictional market	Shares (%)	Calculation	HHI	CR4
A: five firms, unequal	35, 25, 20, 12, 8	$35^2 + 25^2 + 20^2 + 12^2 + 8^2 = 1{,}225 + 625 + 400 + 144 + 64$	2,458	92%
B: 50 equal firms	2 each	50×2^2	200	8%
C: long tail	Top four 4, 3, 2.5, 2; next 46 at 0.4 each; remaining 70.1% spread over 1,400 firms (about 0.05 each)	Top 50: 42.61; tail: $1{,}400 \times 0.05^2 \approx 3.51$	≈ 46	11.5%

Market C is what a fragmented services industry often looks like: a long tail of owner-operated companies. Two details of the arithmetic matter in practice:

- **Small firms barely move the HHI.** In market C the 1,400 smallest firms hold 70% of revenue but contribute about 3.5 points. Squaring makes the index almost entirely a statement about the largest firms.
- **A merger's effect is 2 × a × b.** The Guidelines note that the change in HHI from combining firms with shares *a* and *b* equals 2ab (DOJ/FTC 2023 Merger Guidelines). Combining market C's two largest firms (4% and 3%) adds 24 points. Combining two 0.05% firms adds 0.005.

When CR4 and HHI disagree

The two measures can tell different stories about the same market. Consider two fictional markets in which the four largest firms hold 40% of revenue and the remaining 60% is spread across 600 small firms at 0.1% each:

Fictional market	Top four shares (%)	CR4	HHI from top four	HHI from tail	Total HHI
D: four mid-sized peers	10, 10, 10, 10	40%	400	6	406
E: one leader	37, 1, 1, 1	40%	1,372	6	1,378

CR4 is identical. HHI is more than three times higher in market E, because one firm holds 37% of revenue. For a roll-up thesis the difference is material: in market D a new platform competes with three peers of similar size; in market E it competes with an incumbent that may set prices, recruit from the same labor pool and bid for the same targets. Record both measures, and when they diverge, find out which firm is driving the HHI.

The reverse case also exists. A market with a CR4 of 60% split evenly (15% each) gets 900 HHI points from the top four. If the remaining 40% is spread thinly, the total stays below the 1,000 line the agencies once used for "moderately concentrated", even though four firms control most of the revenue. A single number is a summary; the share distribution is the evidence.

Thresholds from the 2023 Merger Guidelines

The best-known HHI thresholds come from antitrust law, not from investment practice. Under the 2023 Guidelines, "markets with an HHI greater than 1,800 are highly concentrated, and a change of more than 100 points is a significant increase" (DOJ/FTC 2023 Merger Guidelines). A merger is presumed to substantially lessen competition if it produces:

Indicator	Threshold for the structural presumption
Post-merger HHI	Greater than 1,800 **and** change greater than 100
Merged firm's market share	Greater than 30% **and** change in HHI greater than 100

Source: DOJ/FTC 2023 Merger Guidelines. The presumption "can be rebutted or disproved" (same source).

A footnote in the Guidelines records the history: the 1982 guidelines called HHI between 1,000 and 1,800 "moderately concentrated" and above 1,800 "highly concentrated", the 2010 guidelines raised the thresholds, and the 2023 Guidelines returned to the original levels (DOJ/FTC 2023 Merger Guidelines). The DOJ's HHI page still describes markets between 1,000 and 1,800 as "moderately concentrated" (DOJ Antitrust Division).

FTC and DOJ leadership announced on February 18, 2025 that the 2023 Guidelines remain in effect (FTC, February 18, 2025), and on October 28, 2026 the DOJ's merger guidelines page still listed them, issued December 18, 2023, with no notice of withdrawal or revision (DOJ, Merger Guidelines).

Use these numbers as a reference scale, not a fragmentation test. They were written to identify mergers that may harm competition in a properly defined antitrust market. An investment thesis uses them for two narrower purposes:

1. **Context.** An industry far below 1,000 on a credible market definition is unconcentrated by any of these conventions.
2. **Risk flag.** A roll-up that becomes large in a narrowly defined local market can approach the levels the agencies examine, even if national figures look fragmented. Antitrust questions belong with counsel.

Concentration ratios: top 4, 8, 20, 50

A concentration ratio (CR) is the share of an industry's revenue held by its largest *k* firms. CR4 is the top four, CR8 the top eight, and so on. It is easier to explain than HHI and easier to check against trade claims, but it ignores how revenue is split inside the top group and says nothing about the rest.

The Census Bureau publishes both measures from the Economic Census. For 2017, "Firm Concentration Ratios are given for the 4, 8, 20, and 50 largest firms", HHI "is calculated using the top 50 firms", and "2017 is the first time HHI is available for sectors other than Manufacturing" (Census Bureau, What's New for 2017). The same page notes that 2017 was also the first time concentration ratios were available for the Mining (NAICS 21) and Construction (NAICS 23) sectors, and that the size and concentration reports were consolidated and standardized across sectors beginning in 2017 (Census Bureau, What's New for 2017).

Because Census computes HHI from the top 50 firms only, it is close to the full-market HHI in fragmented industries (in fictional market C, the tail beyond 50 firms added about 3.5 points) and should be read as the top-50 index, not a census of every firm.

Where to find the tables and what they exclude

2017. The 2017 tables are published as EC1700SIZECONCEN ("Selected Sectors: Concentration of Largest Firms for the U.S.: 2017") (Census Bureau, What's New for 2017) and through the Census Data API dataset `2017/ecnsize`, which "includes only establishments and firms with payroll" (Census Data API, 2017 ecnsize).

2022. The Census Data API lists a 2022 edition of the same dataset, described as covering "concentration of largest firms" for selected industries and again including "only establishments and firms with payroll" (Census Data API, 2022 ecnsize). Its variable list includes a firm concentration code, the share of revenue held by the largest firms, and an HHI variable, within a group named EC2200SIZECONCEN, "Concentration of Largest Firms for the U.S.: 2022" (Census Data API, 2022 ecnsize variables). The published table can be read without an API key on data.census.gov. For NAICS 238220, Plumbing, Heating, and Air-Conditioning Contractors, the 2022 table reports 112,088 employer firms with $297.6 billion in revenue; the 4, 8, 20 and 50 largest firms hold 4.7%, 6.7%, 10.2% and 14.4% of revenue, and the top-50 HHI is 9.3 (Census Bureau, EC2200SIZECONCEN, NAICS 238220). By any convention above, that industry is unconcentrated nationally.

Steps to pull a row:

1. Find the six-digit NAICS code for the market. If your thesis is written in SIC terms or trade language, map it first; see the NAICS-SIC crosswalk.
2. Search data.census.gov for the concentration table for the year (EC1700SIZECONCEN or EC2200SIZECONCEN) and filter to the code.
3. Record CR4, CR8, CR20, CR50 and HHI, the measure they are based on (sales, shipments or revenue), the year and any suppression flags.
4. Write down what is excluded: nonemployer businesses (no payroll), industries outside the "selected sectors", and anything below national geography.

Pitfalls: market definition, geography and ownership

National tables, local markets

Economic Census concentration figures are national. Many services that sponsors roll up (HVAC, plumbing, veterinary clinics, dental practices, waste hauling) compete locally. A national CR4 of a few percent is compatible with metro markets where four firms hold most of the revenue. In a fictional metro with five firms at 30, 25, 20, 15 and 10 percent, the local HHI is 2,250, above the Guidelines' highly concentrated line, while the same firms are invisible in the national figure.

If the thesis depends on local pricing power or local consolidation, build a local share estimate from establishment counts and firm evidence. National concentration only shows that no national consolidator dominates.

The holding-company problem

Concentration ratios are only as good as the definition of "firm". If brands are owned by a common parent and counted separately, measured concentration understates real concentration. The reverse error is common in trade claims.

A Census working paper on U.S. advertising and marketing services is the clearest public example. Silk and King found that "at the holding company level, the four largest organizations account for about a quarter of the industry's total revenue", "a share lower by an order of magnitude than that frequently cited in the trade press" (Census CES-WP-09-16). That finding is specific to one industry and period and should not be generalized. The lesson is methodological: check how a concentration claim defines the firm, the industry and the denominator before repeating it.

Roll-ups create the same problem in real time. A platform that has acquired a dozen companies, each still trading under its own name, will appear as many small firms in directories and as one firm in Census enterprise data only if ownership has been consolidated in the records Census uses. When you count "independent" targets, resolve ownership first.

Other pitfalls

- **Revenue measure.** Census concentration is based on sales, value of shipments or revenue (Census Data API, 2022 ecnsize variables); a thesis about capacity or headcount may need a different basis.
- **Code boundaries.** A NAICS code can combine segments that do not compete (residential and commercial work, for example) or split one market across codes.
- **Age of data.** Economic Census data describe the reference year, not the year of your memo.

From a fragmentation number to a target list

A fragmentation measure supports a thesis; it does not produce targets. Census concentration tables contain no company names. They tell you that the long tail exists, not which companies in it are independent, owner-operated and in your size band.

A workable sequence for a geographic infill or roll-up thesis:

1. **State the market and measure** (NAICS code, geography, CR4/HHI, year).
2. **Size the tail** with firm and establishment counts by size class, from the sources compared in the CBP vs QCEW vs SUSB article.
3. **Check the local picture** in the metros that matter.
4. **Resolve ownership** at company level before counting a company as independent.
5. **Write the screen.** Fragmentation justifies the thesis; the screening criteria decide which companies enter the pipeline.

Worksheet

Field	Your entry	Illustrative entry (2022 Census values)
Market definition (product, geography)		Residential HVAC services, U.S. national
NAICS code and title		238220 Plumbing, Heating, and Air-Conditioning Contractors
Source table and year		EC2200SIZECONCEN, 2022
Basis of shares		Revenue
CR4 / CR8 / CR20 / CR50		4.7% / 6.7% / 10.2% / 14.4%
HHI (top 50)		9.3
Geography of competition		Metro
Local evidence collected?		Not yet
Exclusions noted		Nonemployers; non-selected industries
Ownership roll-up checked?		No; sponsor-backed platforms not yet reconciled
Conclusion and confidence		Nationally unconcentrated; local concentration unknown

The NAICS code is real; every number in the right-hand column is invented for illustration. Pull the actual values from Census before using them.

Docket's Scout agent answers a client's own criteria for each company on a list, with each value quoting a retained page, which is the company-level step that a concentration table cannot do.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.justice.gov/atr/media/1329301/dl?inline
2. https://www.justice.gov/atr/herfindahl-hirschman-index
3. https://www.ftc.gov/news-events/news/press-releases/2025/02/ftc-chairman-andrew-n-ferguson-announces-ftc-dojs-joint-2023-merger-guidelines-are-effect
4. https://www.justice.gov/atr/merger-guidelines
5. https://www.census.gov/programs-surveys/economic-census/year/2017/economic-census-2017/news-updates/whats-new.html
6. https://api.census.gov/data/2017/ecnsize.html
7. https://api.census.gov/data/2022/ecnsize.html
8. https://api.census.gov/data/2022/ecnsize/variables.html
9. https://data.census.gov/table/ECNSIZE2022.EC2200SIZECONCEN?n=238220
10. https://www.census.gov/library/working-papers/2009/adrm/ces-wp-09-16.html

THE PUBLISHER

About Docket

Docket provides deal-origination research software and managed research for private equity firms. We help investment teams investigate acquisition targets using structured screening criteria, retained sources and reviewable company evidence. Teams can work through a self-serve platform or use managed research, depending on how they want research delivered.

Research against a defined mandate

A useful target list needs more than company names. Docket focuses on the questions that determine whether a company fits an investment mandate, including the evidence needed to support or qualify each answer. Our research approach makes the connection between screening criteria, source material and conclusions visible to the team reviewing the work.

Triage, Scout and Audit

Docket's three named research agents perform complementary tasks. Triage resolves company identity and screens fit. Scout collects sourced answers against the mandate. Audit checks retained evidence, addresses contradictions and leaves unsupported answers visibly unresolved. This structure helps reviewers distinguish established findings from missing information and questions requiring further investigation.

Evidence that supports investment-team judgment

Our research library covers market mapping, screening criteria, private-company data, succession and ownership, source evaluation and evidence standards. These resources explain the methods and limitations behind origination research. Findings support a team's judgment; they do not establish that a company is for sale or guarantee a transaction or investment outcome.

Work with Docket

Visit Docket to explore the platform, managed research and the current contact path. Read about screening criteria, evidence standards, AI research agents and market mapping.

Public examples are illustrative unless explicitly identified otherwise. Research preparation and authorized outreach are separate activities; confidential target lists and customer outcomes should never be inferred from an educational example.

Website: https://docket.capital

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.