

Counting a trade in one metropolitan area: union membership for inside electrical construction from federal filings, 2010 to 2025

Version 1.0. 20 September 2026. This paper is deposited under its own concept DOI, 10.5281/zenodo.22861996. The companion research note, which reports the data-center application as its own line of work, is deposited under its own concept DOI, 10.5281/zenodo.22866646, and reports the same estimates under a different framing. The earlier working paper those estimates were first published in remains separately deposited under concept DOI 10.5281/zenodo.22310705. Each of the three is its own record; none is a version of another.

Tommy N. Turner. Independent researcher, Williamsburg, Virginia. tommy@gofast.pro.

Data availability. The panel, the codebook, the code and the retrieved sources are deposited as a separate citable dataset under concept DOI 10.5281/zenodo.22862472, cited by this paper and by the companion research note alike. That record carries everything required to rebuild every figure here, together with `LM2_CLASSIFICATION_ERROR_2026-09-20.md`, and is the object to cite when the data rather than either manuscript is what is being used. The panel, codebook and code are released under CC BY 4.0. Retrieved third-party sources are redistributed only where their own terms permit it, and the dataset's README names the terms for each. The Union Membership and Coverage Database figures used in sections 1 and 4 are not redistributed; they are cited to their publisher and the extraction scripts reproduce them from the publisher's own tables.

Disclosure. The author serves on the advisory board of an Alfred P. Sloan Foundation-funded study at Emory University and the University of Tennessee on data centers and energy in Virginia, Georgia and Tennessee. This paper was produced independently and is not a product of that study. The author has no financial interest in any data center operator, electrical contractor or labor organisation.

Use of generative artificial intelligence. The author used a large language model in preparing this work and is responsible for its content.

Abstract

Union density is published for metropolitan areas and, within them, for private construction. It is not published annually for a single trade inside a metropolitan area, and the survey cannot supply that series. The Current Population Survey codes construction as one undifferentiated industry in the three classification vintages spanning these years, so the industry variable does not reach a trade at any sample size. The occupation variable does, but no table family the Union Membership and Coverage Database publishes crosses it with metropolitan area, so the trade is reached only nationally: the CPS sample behind the published electrician estimate runs from 695 to 940 observations a year for the entire United States, which caps the mean across the 253 to 260 published metropolitan areas below four in every year of the panel. The sector-by-metro cells one level up already run to a median of ten observations. Pooling years relaxes the cell constraint and is treated as the serious alternative it is: it buys a usable cross-sectional level at the cost of the annual variation that is this paper's product. This paper builds a membership count from administrative records instead. It is a count and not a density. An employment denominator is available at this resolution from administrative sources and the paper uses one, but the membership count is not a subset of any such denominator, so the ratio is reported as a diagnostic rather than as a rate.

Every labor organisation covered by federal labor-reporting statutes must file an annual report with the Department of Labor stating its membership, and the International Brotherhood of Electrical Workers publishes maps of which of its locals holds jurisdiction over inside construction, the wiring of buildings, in each area. Joining the two yields a sixteen-year panel of 11,645 local-years covering 786 filing organisations, which is every IBEW local union that filed a report of any kind in any year, with membership, apprentice counts, headquarters geography and two vintages of the jurisdictional classification. The panel is released with the paper and set against two external series. Against the survey estimate of construction union membership in the same metros the comparison fails: the panel's count exceeds the survey's count for all of construction in 18.2 percent of metro-years, which is possible only because the two count different populations, and within a metro the survey's figure moves eleven and a half times as much from year to year as the panel's, and about twice as much in the smallest sample cells as in the largest, though not monotonically across them. Against Quarterly Census of Employment and Wages counts of electrical-contractor employment, which carry no sampling error, the two sources correlate in log levels in every year of the panel, at 0.428 to 0.494 on the broader industry code and 0.457 to 0.504 on the narrower, and at 0.55 in growth measured over the panel's full fifteen-year span, against 0.23 in year-over-year change. They agree about where the trade is and how it moved over a decade, and disagree about what it did last year. Known defects are quantified. Absorption of one local by another, which the filing cannot distinguish from organic growth, is bounded at 17.2 percent of aggregate growth in the treated group and 19.7 percent in the nearest comparison group, counting disappearing locals of 50 members or more; counting disappearances of every size raises both shares by about one and a half points. No such ceiling is defined for the third group, whose aggregate membership fell, so nothing bounds absorption on the contrast against it. One application is reported, and not as a causal estimate. Among inside locals of at least 1,000 members, those in metropolitan areas where any of four large hyperscale operators published a data center campus grew 27.4 percentage points faster over 2010 to 2025 than those elsewhere. That estimate is stable in sign and magnitude and fragile in significance. Two of the nine specifications in Table 9 clear conventional significance, both of them randomization tests; six do not, including the clustered regression itself; and the leave-one-out range carries no p-value and never crosses zero. Separately, the contrast against Group C, whose states neither hold a published site nor are named in a treated area's title, clears at $p = 0.005$. Membership per electrical-contractor job nonetheless fell in those metros, by 17.0 percent against 21.0 percent elsewhere on the industry code closest to the work, and by 11.8 against 21.5 percent on the broader one.

Keywords: union membership; measurement; Form LM-2; QCEW; building trades; metropolitan labor markets

1. Introduction

The standard measure of union membership in the United States is survey-based. The Current Population Survey asks employed wage and salary workers whether they belong to a union, and the Union Membership and Coverage Database that Hirsch, Macpherson and Even maintain aggregates the answers into published tables.[1] Those tables reach further than a reader might assume. They cover metropolitan areas annually from 1986, and within a metropolitan area they separate the public sector, the private sector, private manufacturing and private construction. Metropolitan union density is a published quantity, and this paper does not claim otherwise.

What the tables do not reach is a trade within a metropolitan area in a single year, and the first obstacle is classification rather than sample size. The CPS codes construction as a single industry, Census code 0770, corresponding to NAICS 23, with no specialty-trade detail in the three classification vintages covering the years of this panel.[2] Electrical contracting has no industry code of its own in those vintages, so no sample size would yield it through the industry variable. Detailed occupation codes do exist, and an occupational route to a trade is available in principle in the national tables, but none of the table families the database publishes crosses occupation with metropolitan area.[3]

The obvious rejoinder is that a researcher would go to the occupation variable rather than the industry variable, and it can be answered with the database’s own published tables. The national occupation tables report the CPS sample size behind every occupation estimate. For electricians, Census occupation code 6350 in the 2010 vintage and 6355 from 2011, published as one continuous series, that sample runs from 695 observations to 940 across 2010 to 2025, with a median of 809.5.[4] Eight hundred respondents, for the entire United States, in a year.

That figure caps every metropolitan cell, because a quantity summed over metropolitan areas cannot exceed its national total. The metropolitan tables publish 253 areas from 2010 to 2014 and 260 from 2015. The binding year is the one with the largest sample relative to the areas published, 2017, where 940 observations across 260 areas cap the mean metropolitan cell at 3.62; the cap is below four in every year of the panel and is 2.67 in 2025. No assumption is involved: it is the national count divided by the number of areas, and the true figure is lower still because electricians outside published metros consume part of the national total.

Table 1. The CPS cell for one trade, nationally and as a cap on the metropolitan cell[4]

	2010	2015	2020	2025
national CPS observations, electricians	811	904	778	695
metropolitan areas published	253	260	260	260
mean per metropolitan area, upper bound	3.2	3.5	3.0	2.7
national union density estimated from that sample	34.2%	32.7%	31.1%	34.3%

Allocating the national count across metros in proportion to each metro’s CPS sample is not what the survey does, but it indicates the order of magnitude. On that allocation the largest metropolitan sample in the country, New York, would carry about 37 electrician observations in 2025, Chicago about 20, Atlanta about 11, and the median published metro about one. Two hundred and twenty-four of the 260 metros would fall below five.[4]

So the occupation route reaches a trade nationally, on a sample small enough that the database’s own authors flag it, and it does not reach a trade within a metropolitan area annually at all. That is the gap this paper fills, and it is a measured gap rather than an assumed one.

The serious objection to all of this is that a researcher would pool years, which is the standard remedy for a thin cell and is a better alternative than anything the preceding paragraphs address, because every figure above is annual. It deserves the same arithmetic rather than a dismissal. Pooling works. Summed across 2010 to 2025 the national electrician sample is 13,315 observations, and on the same proportional allocation a sixteen-year pool would put roughly 714 observations in New York and about 20 in the median published metro. That is a usable cell.

What it buys is a level, and what it spends is the time variation. A sixteen-year pooled cell yields one number per metropolitan area for the whole period: it cannot give 2010 against 2025, cannot trace a year-by-year path, and cannot place an observation on each side of a dated event. Shortening the pool to recover two periods halves the cell, and a five-year pool returns the median metro to about six observations on that allocation, 5.8, which is the order of magnitude this section set out to escape. The panel built below is 11,645 local-years; the pooled survey alternative is one cross-section.

Three limits survive pooling entirely. The occupation route counts electricians wherever they work, including in utilities and manufacturing, and misses IBEW inside members not coded to the occupation, so it is not the same population as an inside local's roll in either direction. Metropolitan boundaries are redefined inside the window, so a pooled cell is not a constant geography. And none of the table families this database publishes crosses occupation with metropolitan area at any degree of pooling, so a pooled estimate requires microdata and is a different object from the published tables this section is about. The claim this paper is entitled to make is accordingly the narrow one: the survey cannot supply an annual trade-by-metropolitan-area series, and a pooled survey estimate answers a different question.[4]

Sample size binds as a second constraint, and it binds on the whole sector before a trade is asked for. Across the 4,069 metropolitan-year cells carrying a private-construction row from 2010 to 2025, excluding the consolidated areas that aggregate them, the median CPS sample size is ten observations and 95 percent rest on fewer than a hundred. Restricted to the metros where this panel places an inside local, which are larger than average, the median is forty and 83 percent still rest on fewer than a hundred.[5] The database's own documentation is explicit about what follows. BLS "does not provide estimates at such disaggregated levels as does UnionStats due to concern about small sample sizes," and "the chief reliability concern at UnionStats continues to be sampling and reporting error for estimates within narrow categories based on small sample sizes." [3] Hirsch, Macpherson and Vroman had made the same argument about geography two decades earlier, writing that the case against publishing official state estimates "applies with even greater force to density estimates for metropolitan areas and detailed industries and occupations." [6] A count of inside electricians in the Phoenix metropolitan area in a given year sits two steps past where that caution begins.

Union-reported membership is the older source. Before the CPS collected union status regularly, from 1973, state density estimates came from the Bureau of Labor Statistics *Directory of National Unions and Employee Associations*. The Directory collected membership by mail questionnaire to national unions, employee associations and AFL-CIO state bodies, and BLS aggregated the responses to state totals.[7] Those figures were reported by unions about themselves rather than sampled from workers, which is the property the present paper relies on. The Directory and the CPS overlap for 1974, 1976 and 1978. Across those three years the median of the state-specific ratios of CPS to Directory density was 1.02, with a range from 0.72 in Missouri to 1.41 in South Dakota.[8] Agreement between a union-reported count and a household survey is therefore good at the centre of the distribution and loose at its edges.

The filings this paper uses are a different union-reported source and a stricter one, because they are compelled rather than solicited. The Labor-Management Reporting and Disclosure Act has required an annual report of every covered labor organisation since 1959, and the reports are published. This paper returns to them one level below the state, for a single branch of a single trade.

What this paper adds to what is already there. Three literatures touch it and none occupies the same ground. The measurement of union membership from the CPS is the work of Hirsch, Macpherson and Even, whose database this paper uses as its comparison and whose published caution about small cells it quantifies rather than disputes.[9] The metropolitan question is older than that database: Kokkelenberg and Sockell reported union membership for selected states and standard metropolitan statistical areas from the May CPS for 1973 to 1981, so the metro cell has been attempted before. Its abstract describes data “by occupation, industry, selected states and SMSAs, race, gender, and educational attainment”, and no published table located by the literature searches described below crosses that geography with a single trade. Only the abstract and reference list were retrieved for this article; how its tables cross those dimensions is not established here.[10] The union financial filings themselves have been used as a research source, though at the level of the national union rather than the local: Masters, Gibney and Albright analyse the finances of 53 major national unions over 2006 to 2019.[11] What is new here is the local union as the unit, the trade as the classification, and the metropolitan area as the geography.

The application in section 6 sits in the place-based literature on large plant openings and local labour markets, which supplies its design template and its principal alternative explanation.[12] A search for quantitative measurement of data center construction employment at any geography returned nothing, so section 6 is an early entry rather than a contribution to a settled debate. That search covered two OpenAlex queries and one publisher pass in September 2026, roughly sixty records, and did not reach EconLit or JSTOR. It is not an exhaustive claim, and the same database failed to surface one of the canonical works in the place-based literature cited above, which is a fair indication of its limits.[13]

Section 2 sets out what the filing measures and what it does not. Section 3 gives the construction rule. Section 4 sets the panel against a survey estimate of a closely related quantity and against an independent employment count, and says what each comparison can establish. Section 5 quantifies the defects. Section 6 reports one application, and section 6.4 states what that application cannot settle.

2. The source and what it measures

2.1 The filing

Every labor organisation subject to the Labor-Management Reporting and Disclosure Act of 1959, the Civil Service Reform Act or the Foreign Service Act must file an annual financial report with the Office of Labor-Management Standards of the Department of Labor. What receipts determine is which report is owed, not whether one is owed. Organisations with total annual receipts of \$250,000 or more file Form LM-2, the most detailed form, and smaller organisations file the LM-3 or LM-4 unless they are in trusteeship, in which case they file the LM-2 whatever their receipts. Coverage reaches labor organisations representing employees in private industry, the Postal Service and most of the federal government, and organisations representing only state, county or municipal employees are not required to file at all. Item 20 of the form carries the number of members.[14]

On the LM-2, Item 20 is computed rather than entered. Schedule 13, Membership Status Information, asks the filer to list “the categories of membership tracked by the reporting labor organization” in one column and the number in each in the next. The software then sums that column into Item 20.[15] Agency fee payers, who pay in lieu of dues without joining, are entered on a separate line and

excluded from the total. Work permit holders are defined elsewhere on the form as non-members working within the local’s jurisdiction, and Schedule 13 counts members, so they fall outside it.

Three properties of the count follow from this, and they govern what the panel can be used for. Apprentices are counted only where the local tracks them as a category, which between 22.8 and 26.0 percent of IBEW LM-2 filings do in a given year. The share is lower among larger locals without falling monotonically: 24.9 percent across all LM-2 local-years, 20.5 percent among those reporting at least a thousand members, 18.7 percent among those reporting at least two thousand, and 19.8 percent among those reporting at least five thousand.[16] Travellers, members of one IBEW local working on a permit in the jurisdiction of another, are not counted by the local whose jurisdiction they are working in; their home local counts them, so they sit in the panel in the wrong place rather than being absent from it. And what “member” means is whatever the filing local says it means in Item 69, where Schedule 13 requires each category to be defined.

2.2 What it is not

It is not employment. The form reports the union’s own finances, including the salaries and hours of its officers and staff. Nothing in the LM-2 sections examined for this paper asks about the employment of the membership at large, and no field in any panel row carries such a figure; the LM-3 and LM-4 instructions were not retrieved, so the claim is bounded by what those sections and the released fields show. A local’s roll moves with organising, with jurisdictional transfers, with the absorption of a neighbouring local, and with the local’s own decisions about whom to count. None of those has a field in any released panel row, and none is asked for in the LM-2 sections examined here.

It is not confined to the local’s own metropolitan area. A member may work anywhere, including outside the trade. Retired members are counted where the local tracks them as a category, which 10 percent of the IBEW LM-2 filings in this panel do on a search of their Schedule 13 labels for “retire” or “pension”; locals that fold retirees into a broader category are not detectable this way, so 10 percent is a floor. The share is computed by `lm2_prose_figures.py`, which exists because a series of figures in this paper, including this one, were stated in prose with no deposited script behind them until audits on 20 September 2026 found them. Its blocks are keyed to the sections that cite them.

It is not a sample. Filing is compulsory at every size for a covered organisation, with the form determined as in section 2.1, so the panel carries a census of the filings of covered organisations rather than an estimate with sampling error. It is not a census of IBEW locals, because locals representing only state, county or municipal employees are outside the statutes and file nothing. Its errors are reporting errors, classification errors and the ambiguity of absorption, which behave differently and are quantified in section 5. Whether every organisation that owed a report filed one is a separate question and is not testable from the extracts, which record what was filed and not what was owed.

2.3 The extracts

The Office of Labor-Management Standards publishes each year’s filings as pipe-delimited relational tables. The `lm_data` table carries one row per filing with affiliation, designation, headquarters address, fiscal period, membership and an amendment flag; the `ar_membership` table carries the Schedule 13 rows, joined on a report identifier.[17] Extracts for 2010 through 2025 were retrieved in September 2026. The column layout is stable across the whole span: the layout file shipped with

each of the sixteen annual archives is byte-identical to every other. The gate log records that check for the eight archives added in 2026 against the 2018 versions; the comparison across all sixteen is not separately recorded.

The download links carry session tokens that rotate, so a saved URL expires and the download page has to be re-read rather than cited.[18]

3. Construction

Three rules turn a filing into a row in the panel, and each rests on a published external source rather than on the author’s judgement alone.

3.1 Which organisations

The panel holds every filing whose affiliation is IBEW and whose designation is LU, which is the union’s own label for a local union. District councils, state conferences and the international body are excluded. Where a local-year carries more than one filing, the original is used; where only amendments exist, the highest-numbered amendment is used. A membership field that is blank or not a whole number is treated as missing and never as zero. No screen on size or on movement is applied at this stage; the released panel is the whole filing population, and the estimation sample of section 6 is a subset of it.

The result is 11,645 local-years covering 786 filing organisations over sixteen years, carrying 785 distinct local-union numbers: two local numbers each appear under two file numbers, and one file number carries two different local numbers over the panel’s span. By form, 6,738 filings are LM-2, 4,747 LM-3 and 160 LM-4.[16] No LM-3 or LM-4 row in the panel carries a Schedule 13 entry, so the apprentice and category fields are empty for every row on a smaller form rather than for a scattered subset of them, which is the pattern a form-design difference would produce and not the pattern missing data would. The LM-3 and LM-4 instructions were not retrieved, so that reading is inferred from the filings rather than confirmed against the forms. The released file carries the form type on every row, and a reader should filter on it before reading any Schedule 13 field.

3.2 Which locals do inside construction

IBEW locals do not all do the same work, and the union distinguishes inside construction, the wiring of buildings, from its other branches. For each state it publishes a map captioned “Inside Jurisdiction” showing which local holds that jurisdiction for each area, and it is those maps, not a local’s name or anything on its filing, that determine the classification used here.[19] The retrieved maps establish which local holds the inside jurisdiction where. They do not establish the union’s full branch taxonomy, or whether a local holding inside jurisdiction also holds another, and nothing here turns on either.

The 53 map PDFs linked from the union’s index page were retrieved and parsed. Fifty-one cover United States jurisdictions and two cover the Caribbean and Panama; only the 51 are used. The maps carry two kinds of purely numeric text, highway shields and local-union labels, and the two separate cleanly by size: shields sit at 12 points and below, labels at 16.5 and above, and nothing falls between. Non-numeric text runs as high as 26.4 and overlaps both ranges, so a label is a text item that is both numeric and set at 16 points or more.

The maps also carry 12-point notes naming a local in another state’s jurisdiction, such as “LU 1925: See Tennessee map for” on the Kentucky map and “LU 654 (Chester, PA)” on the Delaware map. The parser captures these separately. They add three locals to jurisdictions where no label names them, and none of the three changes which locals the panel flags as inside: the classification is applied as a set of local numbers, and every local named in a note anywhere on the 51 maps also carries a label of its own on some state’s map.[20]

The maps carry revision dates and they are not uniform. Fifty of the 53 are dated February 2018; Colorado is July 2019, Alabama January 2021 and Iowa November 2023. Carrying that classification back to 2010 is an assumption, and section 5.4 tests it against the union’s own earlier maps.

3.3 Which metropolitan area

Each local’s headquarters address as filed was geocoded to a county through the Census Bureau’s Geographies API, and the county mapped to a core based statistical area through the Office of Management and Budget’s July 2023 delineations.[21]

Geography was resolved for the estimation sample and not for the whole panel, and a user of the released file should know it. The crosswalk holds 158 filing organisations, of which 156 carry a CBSA; the two that do not are in West Frankfort, Illinois and Hazen, North Dakota, whose counties fall in no metropolitan area, so for those two an empty CBSA means non-metropolitan rather than unattempted. The 628 organisations outside the crosswalk carry empty geography fields because the geocoder was never run on them. Those are the locals that do not enter the section 6 sample frame. 9,153 of the 11,645 released rows, 78.6 percent, carry no metropolitan identifier. The missing addresses are present in the file, so the gap is closable by anyone willing to run the geocoder over them, and closing it is the single most useful extension to the deposit.

Headquarters location is a proxy for jurisdiction and not the thing itself. An IBEW inside jurisdiction is defined by the union’s map, covers multiple counties, and is not coterminous with any statistical area, so a local headquartered in one metro may hold jurisdiction reaching into another. This approximation touches every row that carries a geography, which is 21 percent of the released file and every row used in section 6. Fixing it would need jurisdiction boundaries, and the union’s published maps are raster and vector PDFs without coordinates. No machine-readable boundary file was found on the union’s own site, which is the only source searched.

3.4 What is in the released panel

A paper whose contribution is a dataset should say what the dataset contains before it argues from it.

Table 2. The released panel, selected years[22]

	2010	2015	2020	2025
filing organisations	777	741	712	673
of which inside on the current roster	260	261	261	261
LM-2	410	415	425	433
LM-3	351	315	280	233
LM-4	16	11	7	7
membership blank	3	4	3	3

	2010	2015	2020	2025
total membership	591,675	579,813	605,520	667,230
median local	306	315	324	392
ninetieth percentile	1,815	1,931	2,068	2,594
largest local	28,565	27,817	25,686	27,835
reporting an apprentice category	94	105	104	110

Three features of that table govern what the panel can be used for.

It is a panel of small organisations with a long right tail. The concentration and form-transition figures in this section come from `lm2_prose_figures.py`, block [5]; the size distribution comes from `lm2_panel_description.py`. Across all 11,593 local-years carrying a membership figure the median local has 320 members and the mean 828. A quarter of local-years fall below 100 members and 1.7 percent above 5,000. Only 22 organisations ever exceed 5,000 members, and those 22 carry 26.8 percent of total membership in 2010 and 28.0 percent in 2025; the locals standing above 5,000 in a given year, 13 in 2010 and 18 in 2025, carry 21.4 and 25.7 percent. Either way, weighting by membership shifts about a quarter of the weight onto roughly two dozen locals. That is why section 6 reports the weighted estimate separately rather than as the headline.

The filing population shrinks while membership grows. Organisations fall from 777 to 673 over the panel while total membership rises from 591,675 to 667,230 and the median local grows from 306 to 392. Locals are consolidating. That is the absorption problem of section 5.3 visible in the aggregate, and it is why the sample rule in section 6 conditions on survival.

The mix of forms shifts toward the LM-2, mostly by exit rather than by promotion. LM-3 filings fall from 351 to 233. Of the 351 organisations filing an LM-3 in 2010, 225 still file one in 2025, 87 have left the panel altogether, 35 have moved up to the LM-2 and 4 have moved down to the LM-4; the four categories sum to 351. Of the 233 LM-3 filers in 2025, 8 were not LM-3 filers in 2010. So of the 126 organisations that moved out of the LM-3, 87 disappeared and 39 crossed a threshold in one direction or the other: 69 percent and 31 percent. Either way the LM-2 share of the panel rises, from 52.8 percent to 64.3 percent, and with it the share carrying Schedule 13 and therefore an apprentice count. A reader treating the apprentice series as a time series should hold that constant.

4. Validation

A series assembled from union filings is useful only if it tracks something counted elsewhere. Two comparisons are available and neither measures the same quantity as the panel, so neither yields a correlation that can be read as a pass mark. Saying what each can establish is therefore part of the exercise rather than a preliminary to it.

The first comparison is against a survey estimate of union membership in the same places. It shares the panel’s subject and not its method. What it can establish is which of the two series carries the noise, because the survey’s precision varies with a quantity the administrative count knows nothing about, namely how many households happened to fall in the cell. The second comparison is against an administrative count of employment. It shares the panel’s method and not its subject, since it counts jobs rather than members and counts non-union jobs alongside union ones. What

it can establish is whether the panel locates the trade in the right places and moves with it over time, subject to a qualification section 4.2 records: both series place a worker at an employer's or a local's reporting address rather than at the job, so their location errors are correlated and the agreement between them is to that extent flattered.

4.1 Against a survey estimate of a closely related quantity

The Union Membership and Coverage Database publishes, for each metropolitan area, the number of employed wage and salary workers in private construction who report union membership.[1] That is the closest published measure to the panel's subject: union members, in construction, in a named metropolitan area. It is one level coarser, covering the whole sector rather than one trade, and it is a survey estimate rather than a count.

Setting the two side by side does not validate the panel, and the reason is worth separating into two parts. The levels do not correspond, because the universes differ. The year-to-year movements do not correspond either, and that second fact has a structure the first does not: it varies with a property of the survey and with nothing on the panel's side. The second is the finding. The first is arithmetic and is reported so the reader can see how far apart the two definitions put the levels. Across 958 metropolitan-years in 69 metros where both sources report, the panel's membership is a median of 0.45 of the survey's count of construction union members, but the ratio exceeds 1.0 in 174 of those metro-years, 18.2 percent of them, and exceeds 2.0 in 42.[23] A ratio above 1 says the panel finds more IBEW inside members in a metro than the survey finds union members across all of construction there. That is not a contradiction, because the two universes are not nested: the two count different populations, and the next paragraph sets out three differences in level, two of which put members in the panel that the survey excludes and one of which is a geographic mismatch running both ways, together with two further differences that bear on movement rather than on level. What the ratio establishes is that the two series are not measuring the same population, which bears on any attempt to use one to check the other, and which the rest of this section takes as its starting point rather than its conclusion.

Three things contribute to the difference in level, and only one of them is a defect in the panel. The panel counts members, including the retired, the unemployed, and those working outside construction or outside the area, while the survey counts people currently employed in construction in that area, and by its own sample definition excludes the self-employed. Some IBEW inside members work in industries the survey codes to utilities or manufacturing. An IBEW jurisdiction is not coterminous with a statistical area, which is section 3.3's defect. Two further differences bear on how the two series move rather than on the gap between their levels. The panel reports a stock, the membership roll on a single date, while the survey reports employment-conditioned status in a reference week, and a stock is mechanically smoother than a flow before any question of sampling arises; nothing below separates that effect from the sampling one. And the survey estimates rest on the sector-by-metropolitan-area cell sizes of section 1, a median of ten observations, not on the national trade cell that section leads with.

Which series is moving can be settled, and it is the survey. Within a metro, across 844 consecutive year pairs, the panel's membership count moves by a median of 3.3 percent in absolute log terms and the survey's construction count moves by 37.7 percent, eleven and a half times as much, and the survey moves more in 92 percent of pairs. The gap widens as the cell shrinks.

Figure 1. Year-to-year movement of each series against the size of the CPS cell.[24] The survey's movement rises as its cell shrinks; the panel's does not. The paragraph below Table

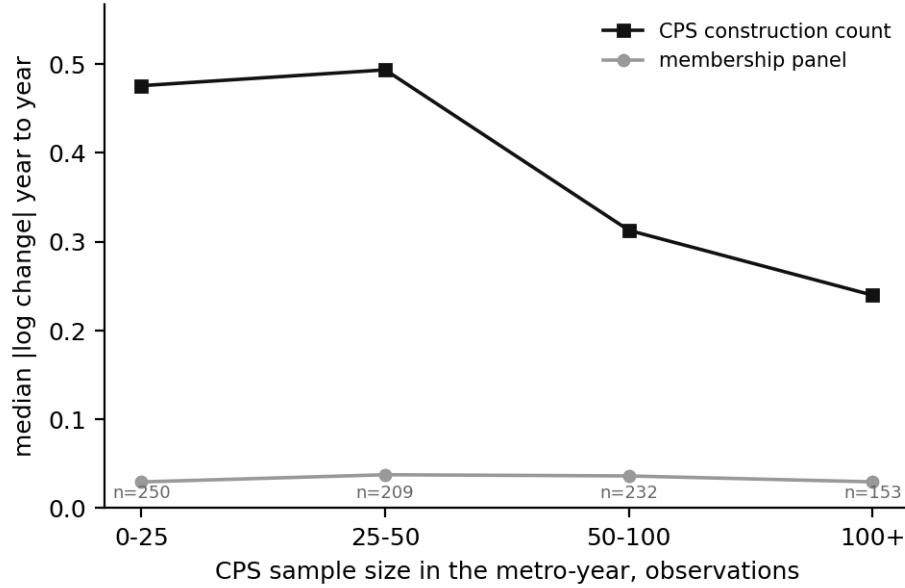


Figure 1: Year-to-year movement against CPS cell size

3 states the gradient and its caveats.

Table 3. Year-to-year movement within a metro, by CPS cell size[23] The pair counts and the movement columns are computed on consecutive metro-year pairs; the median union rate is computed on the metro-year cells themselves, so the two rest on different denominators within the same band.

CPS observations	metro-year pairs	median union rate p	panel	survey	ratio
under 25	250	0.199	0.030	0.476	16.1
25 to 50	209	0.114	0.038	0.493	13.1
50 to 100	232	0.146	0.036	0.313	8.6
100 or more	153	0.103	0.030	0.240	8.1

Cell size is a property of the survey and bears no mechanical relation to what a local union reports to the Department of Labor. The panel’s movement is flat across the bands, between 0.030 and 0.038. The survey’s is about twice as large in the smallest band as in the largest, though the step falls between the second band and the third rather than running monotonically: the two smallest bands sit at 0.476 and 0.493, a difference no test reported here evaluates. The median union rate is reported because the sampling benchmark below is a function of it as well as of n . The ratio column is computed on unrounded medians, so recomputing it from the three-decimal columns printed here returns a slightly different figure.

The source documents non-calendar windows for 2014, 2015 and 2025, which makes four year pairings draw on such a window on at least one side: 2013-2014, 2014-2015, 2015-2016 and 2024-2025. The robustness cut drops every metro pair drawn from all four, removing 227 of the 844 and widening the overall ratio to 12.4, with the band ratios at 16.4, 15.1, 9.9 and 8.2. An earlier version of this cut dropped only three, omitting 2014-2015, which is the pairing with a non-calendar

window on both sides; that looser cut gives 12.1 and is reported in the log beside the one used here.[23]

Two checks bound that reading. The panel is not smooth because locals copy the prior year’s figure: across 3,875 inside local-year pairs only 128, or 3.3 percent, repeat a figure exactly, the median absolute log change at the local level is 0.037, and 19 percent of pairs move by more than ten percent.[23] And the survey’s movement is about the size sampling variation alone would produce. For a cell of n construction workers with union membership rate p , the member count carries relative standard error $\sqrt{(1-p)/(np)}$, and two independent years differ by $\sqrt{2}$ times that in standard deviation. The observed quantity is a median absolute log change rather than a standard deviation, and for a normal deviate the median absolute value is 0.674 of the standard deviation, so the benchmark must be scaled by that factor before the two are comparable. Scaled, the observed movement is 0.79, 1.07, 1.09 and 1.08 times the benchmark across the four bands: at parity in three of them.[23] Two further effects push the observed figure below the benchmark and are not corrected for. The outgoing rotation group design returns a household twelve months after its first outgoing month, so consecutive annual files share households and are not independent. And the calculation ignores the survey’s design effect, which for a clustered sample exceeds one. The claim is that the movement is of the order sampling variation produces, not that it is exactly sampling error.

None of this impugns the database, which states the limitation itself: its documentation records that BLS “does not provide estimates at such disaggregated levels as does UnionStats due to concern about small sample sizes,” and that “the chief reliability concern at UnionStats continues to be sampling and reporting error for estimates within narrow categories based on small sample sizes.”[3] The comparison measures a caution its authors publish.

4.2 Against an administrative employment count

The Quarterly Census of Employment and Wages, the Bureau of Labor Statistics census of employment derived from employer unemployment-insurance tax filings, counts the same places by a wholly separate route and takes no input from the union filings. Because it is a census of employer filings rather than a sample of households, it carries no sampling error, which is what makes it usable where section 4.1’s comparison is not. It counts jobs rather than members, and non-union jobs alongside union ones, so agreement between the two is bounded above by however much the union share varies across metros. No correlation near one should be expected, and none is found. Two industry codes are used: 23821, electrical contractors and other wiring installation contractors, and 238212, nonresidential electrical contractors.

QCEW county employment was assembled for every year 2010 through 2025 from two routes. The Bureau’s Open Data Access industry files return HTTP 404 for every year probed from 1990 to 2013 and resolve for 2014, the first year probed that does. Earlier years therefore come from the annual singlefile bulk archives.[25] A metro’s employment in a year is summed over the counties disclosed in every year of the panel. A county suppressed in one year is thus dropped from all years for that metro, and movement in disclosure cannot manufacture a trend.

Figure 2. Correlation of growth against horizon, both industry codes.[24] The horizontal axis is years from the 2010 base. The profile rises unevenly and reverses at five of the fourteen steps on the broader code and four on the narrower.

Table 4. Agreement between the membership panel and QCEW employment[26]

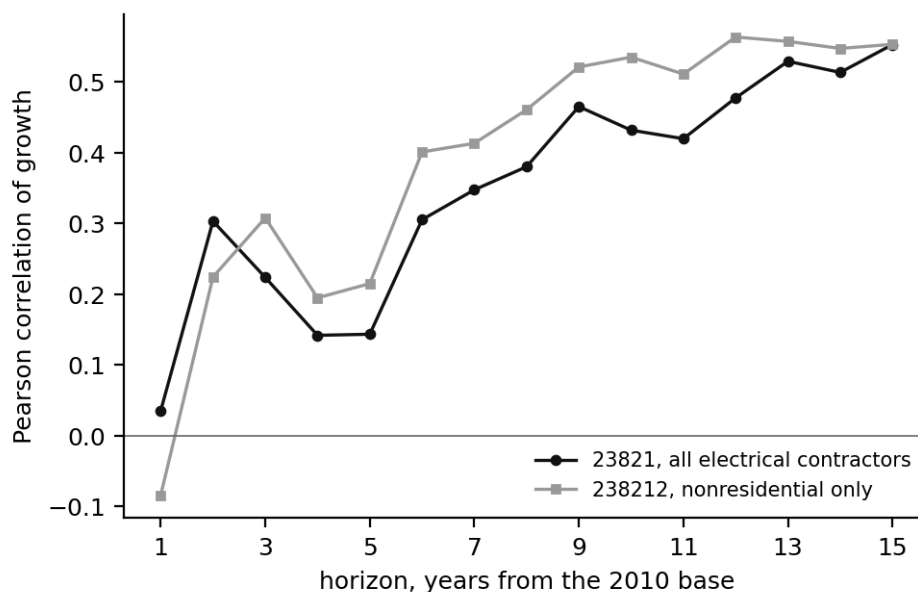


Figure 2: Growth correlation against horizon

	23821	238212
locals with a complete series and a usable metro	77 in 61 metros	80 in 64 metros
cross-sectional Pearson in log levels, 2010 and 2025	0.494, 0.428	0.504, 0.465
range of the log-levels Pearson across all sixteen years	0.428 to 0.494	0.457 to 0.504
range of the log-levels Spearman across all sixteen years	0.378 to 0.504	0.433 to 0.530
growth correlation, one-year horizon	0.036	−0.084
growth correlation, fifteen-year horizon	0.553	0.553
year-over-year change, pooled Pearson	+0.232	+0.227
members per job, 2010	0.562	0.793
members per job, 2025	0.437	0.597

The growth rows carry the result. Correlation of growth measured from the base year rises with horizon, and it rises unevenly: it falls back at five of the fourteen steps on the broader code and at four on the narrower. Those reversals shrink as the horizon lengthens, the largest being 0.11 at the four-year horizon and none past a seven-year horizon exceeding 0.04. Both codes reach 0.55 at the full fifteen-year horizon, and neither is informative at one year. Pooled year-over-year change correlates at 0.23, so agreement at the full horizon is more than twice what it is at one year on that

measure, and an order of magnitude greater on the base-year growth measure.

That profile does not say which series carries the short-run noise, and it should not be read as saying so. Correlation of a long difference rises with horizon whenever signal variance accumulates faster than error variance, and that holds whether the year-specific error sits in the panel, in the employment count, or in both. Section 4.1 could attribute noise because cell size varies the survey’s precision and nothing else’s; section 4.2 has no such lever and makes no such claim. What the profile establishes is joint: over a decade the two series agree, and from one year to the next they do not. A reader wanting the panel for annual inference in a single place should treat that as a reason not to, whichever series is responsible.

Agreement in log levels drifts down on both codes, without moving monotonically, which is consistent with union coverage becoming a smaller and more variable share of the industry. The rank correlations move with the Pearson figures and are reported beside them, so the drift is not an artefact of a few large metros. Membership per job also falls across the panel, with one interruption in 2020 on both codes and a second in 2025 on the broader one.

That last ratio is a diagnostic and not a density. QCEW counts jobs at establishments in the area, union and non-union, and it assigns them to the establishment’s reporting address rather than to the site where the work is performed. For construction that distinction is not incidental: a contractor headquartered in one county and wiring a building in the next reports its payroll at the headquarters. Multi-establishment employers file a breakout by physical location, but the assignment still follows the reporting unit, and no adjustment for it is attempted here. The effect runs in the same direction as the panel’s own geography defect of section 3.3, since both place a worker at an employer’s or a local’s address rather than at the job, so the two series’ errors are correlated and the correlations in Table 4 are, to that extent, flattered. The membership count includes members who are retired, unemployed, or working outside the industry and outside the area. The numerator is not a subset of the denominator, so the ratio cannot be read as a rate. What its decline supports is that the two series move apart slowly and, the two interruptions noted above aside, in one direction, which is the behaviour a coverage measure would show and a measurement artefact would not. Two constructions of the ratio appear in this paper and they are not comparable in level. Table 4 counts each metro’s employment once, across the 77 and 80 locals with a complete series. Section 6.3 sums employment for each local in the estimation sample, so a metro holding four locals contributes its employment four times.[26] The first is a panel-wide diagnostic and the second a within-comparison one.

4.3 Where the sources disagree

The metros where the two counts diverge most over the panel’s fifteen-year growth window are informative about what each is picking up. At one end, the Davenport-Moline-Rock Island area of Iowa and Illinois reports 13.2 percent membership growth against 138.6 percent employment growth, and Minneapolis-St. Paul 5.8 against 100.0. At the other, Dallas-Fort Worth reports 133.8 against 62.2, and Louisville 105.2 against 38.2.[26]

Treated metros appear on both sides of the divergence. Three of the five largest positive gaps are treated, and so is one of the five largest negative gaps on the broader industry code, Minneapolis, and two of five on the narrower. That pattern is what a mismatch of geographic denominators would produce: a membership series responds to construction activity in the local’s own jurisdiction, while the employment count responds to establishment payrolls across a whole statistical area, and the two areas are not the same shape. A composition story in which treated metros differ in the mix

of work would produce a similar pattern. The panel cannot separate the two, and the divergence list is an observation about the measures rather than a test of either account.

5. Defects, quantified

5.1 Reporting defects

Table 5. Reporting defects in the released panel[16]

defect	extent
membership blank, all forms	52 of 11,645 local-years
membership blank, LM-2 only	33 of 6,738
Schedule 13 absent on an LM-2	77 of 6,738
Schedule 13 sum does not equal Item 20	795 of 6,738
only an amendment on file for the local-year	670
more than one filing for the local-year	1
distinct free-text Schedule 13 category labels	2,883

Two of these constrain use. Category labels are free text, and 2,883 distinct strings appear across the panel, so an exact-string join on that field returns silent zeros. One local’s 2023 filing lists “APPRENTICES AND TRAINEES.”[27] And the Schedule 13 sum fails to equal Item 20 in 12 percent of LM-2 local-years, which means the two membership figures on the same form disagree, and the panel carries both rather than choosing.

Free-text categories set two traps for anyone extracting an apprentice count from Schedule 13. Locals routinely report more than one apprentice category, so a rule keeping only one of them understates the total. Fifteen of the 28 locals that section 6 treats report an apprentice category in at least one year, and on seven of those fifteen series a keep-one rule understates the count; the magnitude of the worst case is not reported by any deposited script and is therefore not quoted here. Both counts come from `lm2_v3_0_tables67.py`. And a substring test for “apprentice” scores negations as positives: one local collapsed its whole schedule into a single line reading “ALL ACTIVE MEMBERS WITH EXCEPTION OF APPRENTICES AND CONSTRUCTION ELECTRICIANS,” which such a test reads as 1,015 apprentices in a local of 1,015 members. The released rule sums every apprentice category and rejects negations.[28]

5.2 The apprentice series, and its coverage

Schedule 13 yields a second series. Where a local tracks apprentices as a membership category, the panel carries the count, and that is a measure of the training pipeline rather than of the membership stock.

Coverage constrains what the apprentice series can support. In a given year, between 22.8 and 26.0 percent of IBEW LM-2 filings report an apprentice category, a share rising by 0.13 points a year over the sixteen years, or two points end to end. Among locals on the current inside roster the rate is 39.5 percent of LM-2 local-years, or 38.5 percent counting every form. The coverage shares in this paragraph come from `lm2_prose_figures.py`, block [2]. Of the 84 inside locals in

the estimation sample defined in section 6.1, which `lm2_v3_0.py` selects, 25 report an apprentice category in both 2010 and 2025, from `lm2_apprentice_groups.py`.

Among those that do, aggregate counts rose 99.9 percent in treated metros and 39.1 percent elsewhere, and per-local medians rose 124.5 and 85.8 percent.[29] Those figures rest on 12 treated and 13 untreated locals and should not carry weight on their own. A local reporting no apprentice category is excluded rather than counted as zero, because the categories are free text and a local may fold apprentices into another line.

5.3 Absorption

The filing reports a count, never a reason. A local that absorbs a neighbour gains its members in one step, and its filing is identical to that of a local that organised the same number. The bound below is circumstantial by construction, because the extracts record that a local terminated and record nothing about where its members went.

A screen bounds the size of the problem.[30] Over the panel, 37 IBEW local unions of at least 50 members file in some year and in no later year, together holding 8,646 members. Each of the comparison groups defined in section 6.1 is then credited with every one of those disappearing locals in every state any of its own locals occupies. The resulting figure is a ceiling rather than an estimate, because it counts locals that dissolved with no successor, and because a state holding locals from two groups has its disappearances credited to both. The ceiling is 17.2 percent of aggregate growth for Group A, the treated group, and 19.7 percent for Group B, the untreated group whose state holds a campus. Counting disappearances of every size, not only those at or above the 50-member floor, raises both shares by about one and a half points. Group C is also untreated and its aggregate membership fell, so no such ratio is defined for it. Absorption is therefore material to levels. The two ceilings are close, but they are ceilings rather than estimates, and two upper bounds being close does not establish that the quantities beneath them are. No ceiling is defined for Group C, so nothing here bounds absorption on the comparison against that group.

At the top of the distribution the bound is tighter. Six treated locals more than doubled. Three of them are headquartered in Phoenix, Des Moines and Louisville, and in Arizona, Iowa and Kentucky no local above the 50-member floor stopped filing at any point in the sixteen years. For the other three the same-state ceiling is 7 to 14 percent of the local's own gain.

Below that floor the three states are not identical, and the sub-floor counts come from `lm2_prose_figures.py` rather than from the merger screen, whose own floor is 50 members. Arizona records no disappearance of any size. Iowa records one, a 21-member local in 2017, and Kentucky three, of 15 members and two of zero. Together those four could account for 36 of the 3,860 members the Des Moines and Louisville locals gained, and for none of the Phoenix local's.

5.4 Classification vintage

The inside classification comes from maps revised in 2018 and later. The Internet Archive holds the union's earlier maps at a different address, and they were retrieved and parsed by the same rule.[31] The captures used are dated June 2010 for 318 of the 340 state-local pairs, June 2006 for 14 and December 2013 for 8. Those files carry their own preparation dates of 2003 to 2007, so they are the maps in force at the panel's opening year. The comparison runs over the same 51 United States jurisdictions used to build the current roster.

Across the 51, the archived set carries 340 state-local pairs against 329 in the current set. Seventeen

pairs differ, in ten jurisdictions; because the classification is applied as a set of local numbers, the operative difference is twelve numbers. Eight of the twelve file no report in 2018 at all and three more report between 263 and 851 members. One differing local is large: Local 111 in Denver, on the current Colorado map and not the archived one, which carries Local 969 in that role. Local 969 is recorded as terminated on 1 July 2019 and the current Colorado map is the July 2019 revision, so the two maps are consistent with each other and each is correct for its own date. Four captures of the Colorado map, dated June 2006, June 2010, June 2014 and September 2014, are identical byte for byte and all show Local 969, so the change falls somewhere between September 2014 and the July 2019 revision and is not dated more closely here. The panel ships both classifications as separate flags for this reason.

A second test does not depend on the maps. It uses a sample built on a 2018 base rather than the 2010 base of section 6, and is larger for that reason: 89 inside locals rather than 84. All 89 file a report in every one of the nine years from 2010 through 2018, and 87 of the 89 also carry a non-blank membership figure in every one of those years. Among IBEW locals at or above 1,000 members, the count moves from 154 in 2010 to 156 in 2018. One of the 154 leaves the filing population altogether and one of the 156 enters it; the remaining movement is threshold-crossing among locals that file in both years, with 18 rising above 1,000 members and 16 falling below; those two counts come from `lm2_prose_figures.py`, while the 154 and 156 come from the churn check.[32] Entry and exit from the filing population are therefore negligible over the window, while movement across the size threshold is not, which is the reason section 6 conditions on the base year rather than on size in each year.

5.5 Coverage against the union’s own count

The panel is built from the filings of local unions. The international office files its own Form LM-2 under designation NHQ, and its Item 20 is an independent statement of total membership by the organisation those locals belong to. Summing the locals and dividing gives a coverage check that needs no outside source.

Table 6. Sum of local filings as a share of the international’s reported membership[33]

year	international, Item 20	sum of locals	share
2010	685,498	591,675	86.3%
2015	662,175	579,813	87.6%
2020	683,287	605,520	88.6%
2025	740,182	667,230	90.1%

Across all sixteen years the share stays between 85.6 and 90.1 percent and rises gently. The international’s Item 20 is itself the sum of two Schedule 13 categories, A Members and BA Members, with agency fee payers listed separately and excluded, so the two sides of the comparison are built the same way in all sixteen years.

A parsing defect in this project produced a false claim about two of those years, and it is recorded here because the source is on disk and checkable. Until 20 September 2026 this section stated that the 2020 and 2021 extracts carry only a single “A MEMBERS” row for the international and that Schedule 13 does not reconcile to Item 20 in those years. Both years carry three rows. The international’s category labels are literally “A” MEMBERS and “BA” MEMBERS, and

in 2020 and 2021 the closing quotation mark of the second is mis-typed as an apostrophe. The parser treated those quotation marks as field quoting and merged rows, so the BA row disappeared. Read with quoting disabled, the two member categories sum to Item 20 exactly in both years: 401,511 plus 281,776 in 2020, and 408,388 plus 269,979 in 2021. The years in which Schedule 13 does not reconcile are 2011 and 2012, by 4,600 and 4,804, where the agency-fee row carries a blank label and is therefore not excluded by a test on the label. The year-by-year reconciliation is in `lm2_prose_figures.py`, block [6a]. The comparison uses Item 20 throughout and is unaffected either way.

Part of the gap has a known cause and part does not. The IBEW is an international union and its Canadian locals do not file with the United States Department of Labor, so they cannot appear in the sum. The union’s First District reported more than 71,000 members in August 2026 and growth of more than 9,000 over the four preceding years.[34] Both figures are lower bounds, so their difference fixes no level for 2022; what they establish is that Canadian membership in the early 2020s was of the same order as the gap in those years, which was 74,450 in 2022. They establish nothing about 2010, when the gap was 93,823. No Canadian figure before 2022 was retrieved, and the retrieval covered the First District’s own website only, so the share of the gap that Canada accounts for in the panel’s earlier years is not determined here. Three further candidates are untested here. Locals representing only state, county or municipal employees are outside the reporting statutes, as section 2.1 records, so any such local is in the international’s count and cannot be in the panel; the IBEW represents public-power and municipal utility workers, so this is not a negligible category, and its size was not established. Members may be carried at the international rather than in a local. And the international’s fiscal year ends 30 June while most locals use 31 December, which smears the comparison; the panel does carry each filing’s fiscal period, so the distribution of year ends is testable even though the resulting correction is not.

The check bounds what the panel is. It is a census of the filings of United States locals, covering between 85.6 and 90.1 percent of what the union reports for itself, with a share that does not drift downward over sixteen years.

5.6 A geography defect found in this panel, and what it cost

The defects above are properties of the source. This one was a property of the code that reads it, and it is reported because a paper about the discoverable faults of an administrative measure should report the fault it discovered in its own handling of one.

Metropolitan assignment in the application below originally matched a city-name prefix against a crosswalk of treated areas. That match returns nothing for a local headquartered in a suburb absent from the prefix list, even where the address geocodes to a treated area. The panel also carries a county geocode for every organisation in the estimation frame, built independently, and comparing the two instruments across that frame finds three organisations where the geocode places the local in a treated metropolitan area and the prefix match does not. One of the three holds inside jurisdiction: Local 150 of Libertyville, Illinois, whose county sits in the Chicago-Naperville-Elgin area, where every other local in the panel carries the treated label. It held 1,089 members in 2010, contracted 21.5 percent over the panel, and was counted as a control.

The correction reassigns it. `metro_of` now falls back to the county geocode where the prefix match returns nothing and the area maps to exactly one treated metropolitan area, and the superseded classification is reproducible inside `metro_of` by setting `LM2_NO_GEOFALLBACK=1`. That switch alone does not restore it everywhere: six deposited scripts read `metro_omb` out of the repaired crosswalk

rather than calling `metro_of`, so reproducing the superseded run end to end also means rebuilding `lm2_local_metro_omb.csv` without `lm2_crosswalk_repair.py`. The cost is stated here rather than left to the reader to infer from the tables: the headline contrast falls from +30.73 to +27.35 and stops clearing conventional significance in the clustered regression, while the contrast against Group C and both randomization tests continue to clear it and the leave-one-out range, which carries no p-value, continues not to cross zero. The joint pre-trend test moves from $p = 0.186$ to $p = 0.225$.

Two things generalise past this panel. A crosswalk keyed on place names fails silently rather than loudly, because a miss returns the same value as a true negative, and only a second independent instrument distinguishes them. And the note that first used these two instruments treated their agreement as validation without examining the rows where they disagree. Two derived files agreeing is evidence about the files, not about the thing they describe. The sweep, the quantified effect on every specification, and the preserved superseded artifacts are recorded in `LM2_CLASSIFICATION_ERROR_2026-09-20.md`, deposited in the dataset at 10.5281/zenodo.22862472.

6. An application: hyperscale data center metropolitan areas

6.1 Treatment and groups

A local is treated if its base-year headquarters city lies in a metropolitan area for which Google, Meta, Microsoft or Amazon Web Services, four large hyperscale operators, has published an owned campus. The campus file holds 124 sites, of which 95 fall in the 48 metropolitan areas that treatment is assigned on; the remaining 29 are labelled rural or carry no area. It is assembled from the operators' own location pages and investment announcements. Amazon published no consolidated site list on any of the four pages examined, so coverage of its sites is incomplete, and the treated group is therefore a subset of the metros that hold a campus rather than the whole of it.[35] The consequence runs in both directions and is not signed here. Group A is missing metros that belong in it, and those metros sit in Groups B and C instead. Because both comparison groups are defined by the absence of a published campus, every omission in the file enters them as a possible misclassification, making them over-inclusive rather than incomplete. Group C is sixteen locals in seven states and is the baseline for the largest coefficient reported below; if any of those states holds an unpublished campus, that coefficient is attenuated, and if the unpublished sites sit in states already treated, it is not. Which obtains is not determinable from a file the operators control.

Growth is measured for each local as its 2025 membership over its 2010 membership, less one, in percentage points. Table 7's medians and the cross-sectional coefficient in section 6.2 both use that variable. The annual path in Table 8 uses log membership by year, and section 6.3 uses members per job; each says so where it appears.

The estimation sample holds every IBEW local union reporting 1,000 or more members in 2010 and a non-blank figure in 2025, after a screen for implausible movement. That screen drops any local whose membership falls from 1,000 or more to under 100, or rises more than tenfold, between any consecutive pair of years in the panel. It runs over every filing organisation in the extracts, not only IBEW locals, and flags 685 of them. Among IBEW locals that would otherwise have met the sample rule it removes exactly two, and in both the local reported a literal zero for one or more

years and then resumed near its former level. The rule admits 149 locals, of which 84 hold inside jurisdiction and 65 do not; the estimates below run on the 84 except where a specification says otherwise. Of the 84, 28 are treated. The 56 untreated divide into 40 whose state holds a campus, Group B, and 16 whose state neither holds a site nor appears in a treated area’s title, Group C. The label matters: under the rule below, a state with no campus of its own is still a campus state if a treated metropolitan area’s title names it, so Group C is narrower than “no campus in the state”.[36]

Requiring a 2025 figure selects on survival, and the direction is known. Section 5.3 establishes that 37 locals of at least 50 members file in some year and in no later year, and that the filing records nothing about where their members went. A local that absorbs a neighbour remains in the sample and carries the gain; a local that is absorbed leaves it. The sample rule therefore retains absorbing locals and drops absorbed ones by construction. This is sample selection rather than measurement error, and section 5.3’s ceiling does not bound it, because that ceiling is computed on aggregate growth while the estimates below are computed on each local’s own growth rate. A single absorption moves one local’s growth rate a great deal and an aggregate very little.

Treatment is defined on the metropolitan area, and the B-versus-C split is defined consistently with it: a state counts as a campus state if it holds a site or is named in a treated area’s title. The second clause matters for any multi-state statistical area in which at least one component state holds no site of its own. The Louisville KY-IN area is treated on a campus in Jeffersonville, Indiana, so its Kentucky local is treated. Without the second clause Kentucky would at the same time count as a state with no published campus, putting the local in Paducah, at the other end of Kentucky, into Group C. New Jersey raises the same problem through the Philadelphia PA-NJ-DE-MD area. Multi-state areas in which every component state holds a site of its own, such as Chicago-Naperville-Elgin IL-IN, are unaffected.

Table 7. Median membership growth, inside locals, 2010 to 2025, in percent[36]

group	n	median
A, campus published in the metro	28	+42.13
B, campus in the state, metro untreated	40	+24.45
C, state neither holds a site nor is named in a treated area’s title	16	+3.96

6.2 Estimates

The cross-sectional estimate regresses each local’s growth on a treatment indicator,

$$\text{growth}_i = a + b * \text{treated}_i + e_i,$$

with standard errors clustered on the local’s headquarters state. Across the 84 inside locals in 33 states, b is +27.35 percentage points, 95 percent interval -1.35 to $+56.05$, $p = 0.062$. [37] Table 7’s medians describe the same variable without the regression; the coefficient is a difference in means and the table reports medians, which is why the two do not coincide arithmetically. Against Group C alone the coefficient is +40.89 at $p = 0.005$, and Group B carries +18.95 in its own right. Group

B therefore sits between the treated group and the group with no campus anywhere in the state. The A minus B difference in means is +21.94 at $p = 0.140$, which does not clear conventional significance and does not settle whether Group B is partially exposed or is a second control.[38] Differences in means and differences in the medians in Table 7 are not the same quantity and are labelled as such throughout: the A minus B median difference is +17.68 and the A minus C median difference +38.17, against coefficients of +21.94 and +40.89. Group B’s intermediate position is consistent with a state-level association, with a state-level confounder, and with sampling variation. The design does not choose among the three.

The headline coefficient does not clear conventional significance and both randomization tests do. The clustered regression above rests on 33 state clusters, which is where a cluster-robust standard error is least reliable. Randomization inference permuting whole metropolitan areas returns $p = 0.0136$ and permuting whole states $p = 0.0102$; the contrast against Group C clears at $p = 0.005$; and the leave-one-out range runs +23.66 to +30.03 without crossing zero. The permutation units are the correct ones here because treatment is assigned to a metropolitan area and several locals share one; the paragraph below Table 9 sets that out. The estimate is therefore stable in sign and magnitude and fragile in conventional significance, and this paper reports it that way rather than selecting the inference route that reads best.

The annual path is estimated as an event study and should not be interpreted as one. The specification is the standard one — local and year fixed effects, treatment interacted with every year, one year omitted — and the companion research note refers to it by that name. What it lacks is an event. Treatment here is time-invariant: a local is treated if its metropolitan area carries a published campus at any point, and while 22 of the 28 treated locals sit in a metro carrying at least one dated site, and 21 in a metro whose earliest date falls inside the panel — Indianapolis is dated 2026 — the dates themselves are publication dates for individual campuses rather than a moment at which a metro becomes treated, and 78 of the 124 sites carry no date at all. What the specification estimates is how the gap between treated and untreated locals evolves around an arbitrary normalisation,

$$\log(\text{members}_{it}) = a_i + d_t + \sum_s g_s * (\text{treated}_i * 1[t = s]) + e_{it},$$

with 2017 omitted, local fixed effects a_i , year fixed effects d_t , and standard errors clustered on state. The balanced panel holds 82 of the 84 locals, two dropped for a blank membership figure, in 32 states.

Figure 3. The treated-untreated gap in log membership, by year, with 95 percent intervals.[24] The open circle at 2017 is the omitted year. It is a normalisation and not a treatment date, for the reasons given below.

Table 8. The treated-untreated gap in log membership, by year, 2017 omitted[39]

year	coefficient	95 percent interval, x100	p
2010	+0.0122	-7.33 to +9.77	0.779
2011	-0.0060	-9.01 to +7.81	0.889
2012	-0.0087	-7.78 to +6.04	0.805
2013	-0.0177	-6.98 to +3.43	0.505
2014	+0.0269	-1.06 to +6.43	0.159
2015	+0.0119	-3.57 to +5.96	0.623
2016	-0.0045	-2.72 to +1.83	0.699

year	coefficient	95 percent interval, x100	p
2018	+0.0117	-3.48 to +5.82	0.623
2019	+0.0189	-4.41 to +8.19	0.556
2020	+0.0242	-5.66 to +10.49	0.558
2021	+0.0407	-4.62 to +12.75	0.359
2022	+0.0687	-3.92 to +17.66	0.212
2023	+0.1096	-1.39 to +23.31	0.082
2024	+0.1384	-0.54 to +28.23	0.059
2025	+0.1729	+0.36 to +34.21	0.045

The gap is flat and indistinguishable from zero through 2016, jointly as well as individually, with a chi-square on seven restrictions of 9.40 and $p = 0.225$. It then widens monotonically and reaches +0.173 log points by 2025, about 19 percent, clearing conventional significance only in the final year.

What that joint test is and is not. It establishes that the treated-untreated gap did not trend over 2010 to 2016. It is not a parallel-trends test for a treatment beginning in 2017, because no treatment begins in 2017. Four of the 23 treated metros in the sample carry a dated site before that year, the earliest Des Moines in 2013, so for the treated locals in those four metros the years before 2017 include years after a campus was published in their metro. Those four sit inside the window the test covers. A flat pre-2017 gap is therefore evidence that the two groups were not already diverging, which is worth having, and not evidence that the design satisfies an identifying assumption.

Table 9. Robustness of the treated coefficient[38]

specification	result
Inside only, treated	+27.35 (p 0.062)
with state fixed effects, treated $\times$ inside	+21.56 (p 0.332)
drop metros hosting a large semiconductor or battery plant	+18.37 (p 0.209)
drop metros whose published sites all carry a date of 2025 or later	+31.02 (p 0.055)
weighted by base-year membership	+20.35 (p 0.241)
leave-one-out range	+23.66 to +30.03
randomization inference, permuting metropolitan areas	p 0.0136
randomization inference, permuting states	p 0.0102
inside defined by the 2003-2007 map vintage instead of the current one	+27.82 (p 0.058)

The second row is not a like-for-like substitution and should not be read as one. It changes two things at once: it adds state fixed effects, and it reports the treated-by-inside interaction rather than the treated coefficient the first row reports. The drop from +27.35 to +21.56 therefore mixes the absorption of between-state variation with a change of estimand, and the share attributable to each is not identified by the two rows as they stand. What the row does

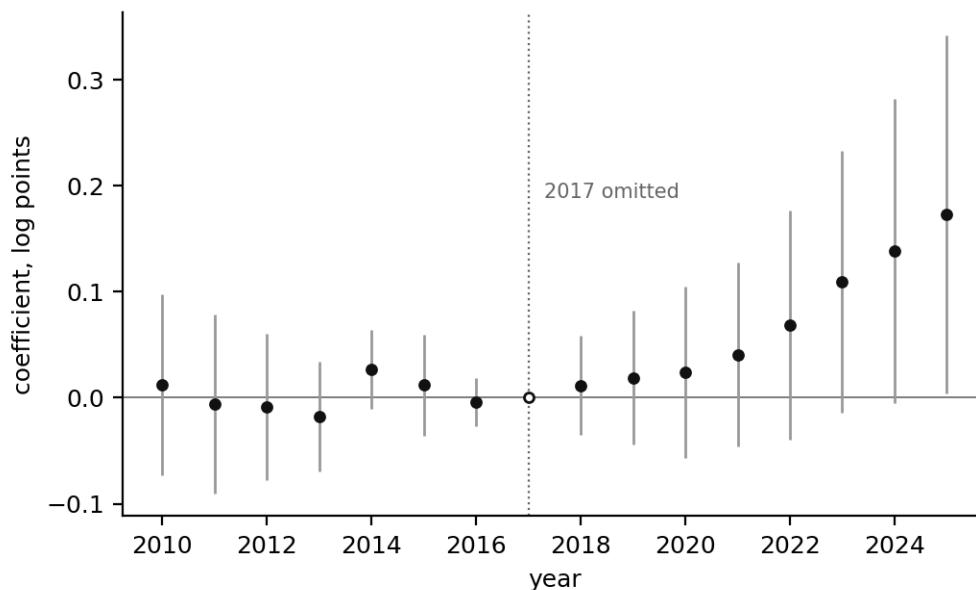


Figure 3: The treated-untreated gap by year

establish is the weaker claim made in section 6.4: once state fixed effects are imposed the contrast no longer clears conventional significance. It does not establish that state fixed effects alone cost the 5.8 points the two rows differ by.

The last row answers a question section 3.2 raises and section 5.4 only half answers. Carrying the current inside classification back to 2010 is an assumption, and the panel ships the 2003-2007 vintage as a second flag. Rebuilding the sample on that vintage drops one local from the inside set altogether, Local 111 in Denver, leaves Groups A and B identical at 28 and 40, and gives a treated coefficient of +27.82 against +27.35. The A minus B median difference is unchanged at +17.68. Group C falls to fifteen and its median growth moves from +3.96 to -2.32, so the A minus C median difference widens from +38.17 to +44.46. The headline contrast is insensitive to the choice between the two map vintages held on disk; the contrast against the smallest group is not. No intervening vintage was retrieved, so this is a comparison of two classifications rather than a survey of all of them.[36]

The two randomization rows permute whole metros and whole states rather than individual locals: 68 metropolitan areas and 33 states across the 84 locals of the estimation sample, on 5,000 draws each, reassigning at each level the number of units treated at that level. Table 9 reports point estimates and p-values and carries no intervals; the intervals given elsewhere in this section are analytic and cluster-robust, and the 999-draw cluster bootstrap seeded at 20260919 in `lm2_identification_v3.py` computes intervals for the Callaway and Sant’Anna cells that this paper does not report. The randomization uses `numpy.random.default_rng(1)`, re-seeded at the start of each permutation level; `lm2_referee_checks_v3.py` records that, because changing the seed, the draw count or the re-seeding point moves every p-value here. Treatment is assigned to a metropolitan area and several locals share one, so permuting locals independently destroys the clustering in the assignment mechanism and produces a null distribution that is too tight by a factor of three: the same data give $p = 0.0040$ under local-level permutation, against 0.0136 and 0.0102 under the two correct units.

6.3 Membership against employment

Membership per electrical-contractor job fell on both sides of the treatment split. On the broader industry code it fell in treated metros from 0.338 to 0.298, a decline of 11.8 percent, against a fall from 0.331 to 0.260, or 21.5 percent, elsewhere. On the narrower code the treated decline is 17.0 percent against 21.0 percent elsewhere. These levels are computed on the estimation sample and sum employment per local, so they are not comparable with the panel-wide ratios in Table 4.[40]

What the comparison supports is a differential and not a level. Membership per job fell in both groups, and fell less where the campuses were published. Both figures are group aggregates; no per-metro distribution is reported, so “in both groups” is the claim and not a statement about every metro. On the broader code that difference is 9.7 points and on the narrower code 4.0 points, and the narrower code is the one that maps more closely to the work a data center campus generates. A reading on which union coverage rose in treated metros is not supported by either code.

6.4 The limits of the design

The estimates above are reported as a description of the panel and not as a causal estimate. Seven things stand in the way. Table 9 bears on two of them, dropping the affected metros for each, and its state-fixed-effects row speaks to a third, subject to the caveat section 6.2 attaches to that row. The remaining four are untested here.

Treatment is built from the defect section 3.3 identifies. A local is assigned to a metropolitan area by its headquarters address, and section 3.3 states that an IBEW inside jurisdiction is not coterminous with a statistical area. Section 4.3 shows that mismatch operating in the validation, with treated metros on both sides of the largest divergences between the membership and employment counts. Nothing here corrects for it, and no machine-readable jurisdiction boundary file was found on the union’s own site, which is the only source searched.

There is no instrument and no comparison of near-misses. Greenstone, Hornbeck and Moretti compare counties that received a large new plant with the counties that were finalists for it and did not.[41] The operators published no site-selection finalists on any page examined in the retrieval behind the campus file, so that design cannot be run here.

Treatment is undated for most sites, so the annual path in Table 8 is normalised on a convention rather than on an event, as that table’s own discussion sets out.

State fixed effects absorb the treated-by-inside interaction into insignificance. That row changes the estimand as well as the specification, as section 6.2 sets out, so it does not by itself measure how much of the contrast is between states rather than within them. What it establishes is that the contrast does not survive the imposition of state fixed effects.

Group C is seven states, which is too few for its composition to be incidental, and it is defined by the absence of a published campus in a file section 6.1 records as incomplete.

Four metros in the campus file also host a large semiconductor or battery plant, three of which enter this sample, of the kind the CHIPS and Science Act and the Inflation Reduction Act, both enacted in August 2022, were written to subsidise. No award list was retrieved, so whether each plant received an award is not established here, and four is a floor rather than a census. Removing the three that enter the sample costs nine points and conventional significance.[38]

Travellers place part of the treatment in the comparison group. Section 2.1 records that a

member working on a permit in another local’s jurisdiction is counted by the home local, not by the local whose territory holds the work. Large construction projects are exactly the occasion for that movement. A campus in a treated metropolitan area drawing electricians from locals elsewhere raises employment in the treated metro and raises membership, if it raises it at all, on the rolls of the untreated locals that supply the travellers. The comparison group is therefore partially exposed, which would bias the estimated gap toward zero to the extent it operates. How much it operates is not established here, so the direction is stated as a property of this channel and not as a correction to the estimate. No field in any released panel row records the jurisdiction worked as against the member’s home local, and none of the LM-2 sections examined for this paper asks for one, so the magnitude is not recoverable from this source and is not bounded here.

An eighth consideration is not a defect in the design but a rival explanation for what it finds. The strongest alternative account is agglomeration, on which a trade concentrates where it is already thick. Ellison and Glaeser stated the identification problem in their own terms, writing that “we cannot distinguish comparative advantage from spillover theories based only on the mean levels of geographic concentration.”[42] Mean levels are what this panel has.

7. Conclusion

The measurement gap this paper fills is narrow and it is real. Union density is published for metropolitan areas and for private construction within them. It stops there for an annual trade-within-a-metro figure, and the arithmetic of why, together with what multi-year pooling would and would not recover, is in section 1. The survey carries no industry code for a building trade in any vintage examined here. Its occupation variable does carry one, and the national sample behind it runs from 695 to 940 observations a year, which caps the average metropolitan cell below four in every year of the panel; allocating that national count across metros in proportion to their CPS samples, which is an illustration rather than a measurement, would leave the median published metro at about one. The sector-by-metro cells one level up carry a median of ten. Pooling years buys a usable cross-sectional cell and forfeits the annual variation, which section 1 states with the same arithmetic. Compelled filings reach the local union, and the union’s own published jurisdiction maps carry the local to the trade. This paper builds sixteen years of them for inside electrical construction, sets the result against two external series, and quantifies what is wrong with it. Neither comparison is a validation in the sense of a pass mark, because neither external series measures what the panel measures, and section 4 says so before reporting either.

The limit is the more useful of the validation’s two findings. The panel correlates with an administrative employment count across the sixty-odd metros where both series are complete, at Pearson correlations in log levels of 0.428 to 0.494 on one industry code and 0.457 to 0.504 on the other in every year, and at 0.55 in growth over the full span, against 0.23 from one year to the next. It is a decade measure. Work that needs annual precision in a single place should not use it, and work that needs a consistent long-run measure of one trade’s strength across many places now has one.

The application shows what the series can and cannot settle. Among inside locals of at least 1,000 members, those in metros where one of the four operators named in section 6.1 published a campus grew about 27 points faster than those elsewhere from 2010 to 2025. Two of the nine specifications in Table 9 clear conventional significance, both of them randomization tests; six do not, including the clustered regression itself; and the leave-one-out range carries no p-value and never crosses zero. Separately, the contrast against Group C, whose states neither hold a published site nor

are named in a treated area’s title, clears at $p = 0.005$. The clustered regression rests on 33 state clusters, which is where that standard error is least reliable, and the paper reports the disagreement between the inference routes rather than choosing the one that reads best. Membership per job nonetheless fell in those metros, more slowly than elsewhere. Whether campus construction caused the difference is not answerable from this data.

Whether this is a method or a dataset turns on whether other unions publish what the IBEW publishes, and they partly do. The rate-limiting input is not the filing, which exists for every covered organisation, but the jurisdiction map, which supplies the trade. Of five building-trades internationals examined in September 2026, the Iron Workers publish a geospatial application whose documented layers include local union jurisdictions distinct from local union offices, with address search; the Carpenters publish a map at regional-council granularity, too coarse to identify a local; SMART publishes an area label for each sheet metal local in text; and the United Association and the Operating Engineers publish locators keyed to office address and proximity, which is a different thing.[43] **What that establishes is weaker than transfer.** The Iron Workers case shows that the input the method needs may exist for another trade. It does not show that the method runs, because nothing was parsed, nothing was joined to a filing, and the one property that makes the IBEW case work is explicitly open: whether the Iron Workers layer separates structural, ornamental, reinforcing and rigging work, as the IBEW’s maps separate inside construction from its other branches, is not stated in the documentation and could not be read without running the application. Since section 3.2 defines the classification by that branch separation, the open item is the whole question. The honest statement is that one other international publishes jurisdiction geography at the local level, which is a necessary condition for the method and not a sufficient one.

One of the five, the Operating Engineers’ locator, loads by JavaScript and did not render; two retrieval routes failed, so whether its records carry a territory or branch field is not determined. The sheet metal directory PDF was not fetched and whether the United Association holds a non-public master jurisdiction document was not determined either. Those three characterisations sit under a limit; the Iron Workers and Carpenters pages rendered and were read. No union constitution, printed directory or member-only area was examined for any of the five, and absence of a map on the pages examined is not evidence that none exists.

One observation runs against section 3.3. The Iron Workers’ help document describes local-union jurisdictions published as GIS layers rather than as PDFs. If that description holds, the boundary problem this paper treats as a defect would be a property of the IBEW’s chosen publication format rather than of the approach. The layer’s contents were not opened, so whether it separates the trade’s branches as the IBEW’s maps do is not established.

Three limits would need new retrieval. No Colorado inside map between September 2014 and the July 2019 revision was located, so the year Local 111 took the Denver jurisdiction is not established. No project list was retrieved that would bound the semiconductor and battery confound in section 6.4. And no machine-readable jurisdiction boundary file was located in the sources searched, so headquarters location stands in for jurisdiction throughout.

Endnotes

[1] Hirsch, Barry T., David A. Macpherson, and William E. Even (2026). Union Membership, Coverage, and Earnings from the CPS. <https://unionstats.com>. Metropolitan tables downloaded

19 September 2026 and held at `unionstats_src/`. The source states that use of the data requires citation.

[2] CPS industry classifications published by the same source for 2009 to 2013, 2014 to 2019, and 2020 forward. Construction is Census industry code 0770 with NAICS equivalent 23 in each, with no subsector entries. Scope: those three documents, which are the vintages covering 2010 to 2025. They do not state whether specialty trades are separated anywhere else in the classification.

[3] Union Membership and Coverage Database, cited in full above. The two quotations are from its documentation page on the correspondence between its own estimates and those in the BLS news release, https://unionstats.com/cps/BLS_vs_UnionStats.htm, retrieved 19 September 2026 and quoted verbatim in `unionstats_src/UNIONSTATS_DOCUMENTATION_2026-09-19.md`. The same file records the table families the database publishes, which is the basis for the statement that none crosses occupation with metropolitan area; the scope of that negative is the published table families of this database, not every published table from any source.

[4] `lm2_occupation_cell.log`, computed from the national occupation tables and the metropolitan tables of the Union Membership and Coverage Database, extracted to `unionstats_src/OCC_construction_trades`. The extract covers all sixteen years and five further building trades besides electricians. Two cautions recorded there travel with it: the electrician occupation code is 6350 in 2010 and 6355 from 2011, and the plumbing row is a combined pipelayers-and-plumbers series through 2019 and a narrower one from 2020, so that series is not continuous across the panel. No page carries an all-occupations national total, so the metropolitan share of the national sample is not computed.

[5] `lm2_cps_comparison.py`, sections 1 and 2 of its log, computed from `unionstats_msa_long.csv`. Restricted to the metros where the panel places an inside local, which are larger than average, the median is 40 and 82.6 percent of matched metro-years still rest on fewer than 100 observations.

[6] Hirsch, Macpherson and Vroman, “State-Level Estimates of Union Density, 1964 to Present,” author manuscript, footnote 2. The same work appears in condensed form as “Estimates of Union Density by State,” *Monthly Labor Review*, July 2001, which does not carry the footnote. Both versions were compared and the footnote appears only in the manuscript.

[7] Same manuscript, on the *Directory of National Unions and Employee Associations* and the mail questionnaire behind it.

[8] Same manuscript, comparing survey and Directory estimates for the three overlapping years.

[9] Hirsch, Barry T. and David A. Macpherson, “Union Membership and Coverage Database from the Current Population Survey: Note,” *ILR Review* 56, no. 2 (2003): 349-354.

[10] Kokkelenberg, Edward C. and Donna R. Sockell, “Union Membership in the United States, 1973-1981,” *ILR Review* 38, no. 4 (1985): 497-543.

[11] Masters, Marick F., Raymond F. Gibney and Robert Albright, “The financial status of national unions,” *Industrial Relations Journal* 53, no. 4 (2022): 303-335.

[12] Greenstone, Hornbeck and Moretti (2010) for the design template; Ellison and Glaeser (1997) for the alternative explanation. Both are cited in full in section 6.4’s notes.

[13] `LM2_Literature_Position_v0_1.md`, 4 September 2026, which records the queries run, the approximate record count, and that EconLit and JSTOR were not searched.

- [14] Form LM-2 instructions, revised 2020, sections I, II and Item 20, retrieved to `olms_src/LM-2_Instructions_Re`. Item 20 reads in full: “After Schedule 13 is completed and the ‘Save & Calculate’ button is clicked, the software will enter the total number of members into Item 20.” The instructions for the LM-3 and LM-4 were not retrieved, so section 2.1’s account of Item 20 is verified for the LM-2 only.
- [15] Same source, Schedule 13 and Item 21(e) on work permit holders. Item 69’s own instructions were not captured in the retrieval; the requirement to define each category there is stated within Schedule 13’s instructions, which were.
- [16] `lm2_panel_release.py`, which writes `IBEW_Local_Membership_Panel_2010_2025.csv` and its codebook, and generates Table 5. The counts of file numbers and local-union numbers are derived from that CSV; the script prints the row and organisation totals but not the breakdown of the two duplicated local numbers.
- [17] OLMS yearly data extracts, `lm_data` and `ar_membership` tables.
- [18] Three distinct token stems were observed across fourteen days in September 2026; the stems and dates are recorded in `LM2_Extension_Gate_Log_2026-09-05.md`, Gate 1.
- [19] IBEW inside jurisdictional maps, retrieved from the union’s own index page in September 2026.
- [20] `lm2_ibew_roster.py`, which writes `lm2_ibew_roster_rebuilt.csv` with a `basis` column recording whether each local entered by label or by note. The size histogram supporting the rule is in `ibew_src/IBEW_Inside_maps_textitems_2026-09-05.txt`.
- [21] `lm2_omb_crosswalk.py` and `lm2_omb_crosswalk_extend.py`. Census Bureau Geocoder Geographies API, benchmark Public_AR_Current, vintage Current_Current, layers=Counties, one request per address.
- [22] `lm2_panel_description.log`, computed from `IBEW_Local_Membership_Panel_2010_2025.csv`. The log carries all sixteen years, the full size distribution and the geography coverage figures.
- [23] `lm2_cps_comparison.py`, which reads `unionstats_msa_long.csv` and the released panel and writes `lm2_cps_comparison.csv` and `lm2_cps_comparison.log`. The log carries the per-band figures, the window-overlap robustness, the sampling benchmark and the local-level stability check in full.
- [24] `lm2_figures.py`, which draws each figure from the JSON and CSV outputs of the scripts named above and prints the underlying series to `lm2_figures.log`, so every figure can be checked against numbers rather than read off an image.
- [25] Probed 19 September 2026: 1990, 2000, 2005, 2007, 2010, 2012 and 2013 all return HTTP 404 at `data.bls.gov/cew/data/api/{year}/a/industry/`; 2014 returns 3,453 rows. Years between those probes were not tested individually. Provenance for each year is in the header of its file under `qcew_src/`.
- [26] `lm2_validation.py`. Table 4 is assembled from its passes 1 through 3 and the divergence list in section 4.3 from pass 4. Its members-per-job figures count each metro’s employment once; the figures in section 6.3 come from `lm2_qcew_benchmark_v3.py`, which sums employment for each local in the estimation sample, so a metro with several locals contributes its employment several times. The two are not comparable in level.
- [27] Local 26, fiscal 2023, report 873507, at line 2640 of `olms_src/extracts/2023/ar_membership_data_2023.txt`

[28] `lm2_v3_0_tables67.py` carries the released rule and `lm2_apprentice_flag_check.py` opened the filings behind both pitfalls. The quoted category label is Local 1579, fiscal 2023, report 884989, at line 7987 of `olms_src/extracts/2023/ar_membership_data_2023.txt`; internal spacing is normalised to single spaces here, and the filing carries double spaces before “AND” and before “ELECTRICIANS.”

[29] `lm2_apprentice_groups.py`. Aggregate counts among locals reporting in both endpoint years.

[30] `lm2_merger_screen.py`.

[31] `ibew_src/C0_Inside_map_captures_2026-09-19.txt` and `lm2_ibew_roster_archived.py`.

[32] `lm2_churn_2010_2018.py`.

[33] `lm2_nhq_coverage.log`, computed from the OLMS extracts and `IBEW_Local_Membership_Panel_2010_2025`. The table shows four years; all sixteen are in the log.

[34] IBEW First District (Canada), “IBEW Canada Reaches Over 71,000 Members,” 24 August 2026, retrieved to `ibew_src/IBEW_Canada_membership_2026-09-19.txt`. A union’s own announcement rather than an audited count, and not reported on the same basis or fiscal calendar as Form LM-2 Item 20.

[35] `Hyperscaler_Campus_Locations.csv`. 78 of the 124 sites carry no announcement year. A Callaway and Sant’Anna estimator runs in `lm2_identification_v3.py` on the 21 treated locals whose metro carries a date, and its output ships in the deposited log. Neither this paper nor the companion research note reports it as a finding, for the same reason: a cohort structure built on the 46 dated sites of 124 will not carry a claim about timing. The note records the aggregate of +13.8 percent and its range across five passes in its robustness section, says why seven treated locals drop out, and rests nothing on it. Both documents run the same estimator on the same 21 locals and get the same numbers.

[36] `lm2_v3_0.py`, which applies the sample rule, the movement screen and the group definitions, and generates Table 7. Set `LM2_ROSTER=archived` to rebuild the sample on the 2003-2007 map vintage; the group medians in that run come from this script and the coefficient of +27.82 from `lm2_identification_v3.py` under the same setting. The deposited log for that run is `lm2_roster_drop111_base2010.log`, whose command line reads `LM2_ROSTER_DROP=111`. The two switches are equivalent here and were verified to be so on 20 September 2026: the vintages differ on twelve local numbers and Local 111 is the only one whose reclassification changes membership of the inside set for this sample, so both produce the same 28-local Group A, the same 15-local Group C and the same coefficient.

[37] `lm2_identification_v3.py`, specifications R3 and R4.

[38] `lm2_referee_checks_v3.py`. Table 9’s first eight rows, the A minus B Wald test, the fab and battery cut and the randomization inference all come from this script. Its final row, the classification-vintage cut, comes from `lm2_v3_0.py` and `lm2_identification_v3.py` run under `LM2_ROSTER=archived`. The deposited log for that row is `lm2_roster_drop111_base2010.log`, whose command line uses the equivalent `LM2_ROSTER_DROP=111`; endnote [36] records why the two switches produce the same sample here.

[39] `lm2_identification_v3.py`, event study section. The joint test is a chi-square Wald statistic on seven restrictions. Coefficients are in log points; the script’s percent column is the linear approximation, which understates the 2025 coefficient by about 1.6 points.

[40] `lm2_qcew_benchmark_v3.py`, which reports members per job by treatment group on the estimation sample. The note to Table 4 gives the two constructions and why their levels are not comparable.

[41] Greenstone, Michael, Richard Hornbeck and Enrico Moretti, “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy* 118, no. 3 (2010): 536-598.

[42] Ellison, Glenn and Edward L. Glaeser, “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach,” *Journal of Political Economy* 105, no. 5 (1997): 889-927. The quoted sentence is from § 2.2.2 of the working-paper version, NBER 4840, which is the text held on disk for this project.

[43] `ibew_src/OTHER_TRADES_jurisdiction_publication_2026-09-19.md`, which records what was retrieved from each union’s own site, the URLs, and what could not be determined. The Iron Workers layer structure is quoted from their application’s own help document.