SCHOOL OF OPERATIONS RESEARCH AND INFORMATION ENGINEERING
COLLEGE OF ENGINEERING
CORNELL FINANCIAL ENGINEERING MANHATTAN
CORNELL UNIVERSITY

Financial Engineering Project

BUILDING A PRIVATE CREDIT INDEX FOR ALTERNATIVE LENDING MARKETS

Presented to the Faculty of the Graduate School of Cornell University
in partial fulfillment of the
requirements for the Master of Engineering Degree
and the Financial Engineering Concentration

By:
Abhinav Chhabra
Ananya Bajaj
Chia-Yi Chien
Masis Akdemir
Mohamed Azahriou

Faculty Advisor:
Prof. Johannes Stefan Wissel

Sponsor:
Marathon Asset Management – Jason James Foster

December 2025

Faculty Advisor's Signature ______________________ Date ______________________

Contents

LIST OF ACRONYMS

BDC	Business Development Company
CFEM	Cornell Financial Engineering Manhattan
CDLI	Cliffwater Direct Lending Index
SOI	Schedule of Investments
iXBRL	Inline eXtensible Business Reporting Language
EDGAR	Electronic Data Gathering, Analysis, and Retrieval
PIC	Paid-In-Cash
PIK	Paid-In-Kind
FV	Fair Value
SOFR	Secured Overnight Financing Rate
AUM	Assets Under Management
CIK	Central Index Key
API	Application Programming Interface
EDA	Exploratory Data Analysis
LLM	Large Language Model

1 Introduction

Private credit has grown into one of the fastest-expanding segments of alternative finance, surpassing $1.5 trillion in AUM and serving as a critical source of capital for middle-market firms. Unlike public debt markets, private credit operates with limited transparency, infrequent valuations, and heterogeneous disclosure standards, making systematic performance measurement difficult. Investors, regulators, and asset managers lack a unified benchmark to evaluate risk, returns, and exposure across the private lending landscape. The absence of standardized performance data remains a central barrier to effective monitoring and market-wide comparability.

Business Development Companies (BDCs) offer a unique opportunity to address this opacity. As publicly regulated vehicles under the Investment Company Act of 1940, BDCs disclose loan-level information—including fair value, cost, coupon structures, PIK/PIC features, industry classifications, and maturity profiles - through quarterly 10-Q and annual 10-K filings. These disclosures provide rare transparency into the private credit ecosystem, enabling empirical measurement of risk and performance dynamics across thousands of underlying loans.

This project leverages that regulatory transparency. Between 2023Q1 and 2025Q2, the team extracted over 5,400 filings (approximately 5GB of raw data) and processed more than 300,000 investment records, including a substantial subset of iXBRL filings that offer machine-readable financial tagging. The availability of iXBRL significantly improved data consistency and reduced noise, forming the foundation for benchmarking private credit performance

2 Problem Statement

Despite rapid industry growth, private credit lacks a standardized, data-driven index capable of tracking performance at the asset level. Existing evaluations often rely on proprietary fund disclosures or high-level yield commentary rather than transparent, reproducible benchmarks. Building such an index presents several challenges.

First, reporting practices among BDCs are highly inconsistent. Filings differ in structure, terminology, and the completeness of key fields such as interest rates, spreads, and payment structures. Variations in taxonomy, including the use of custom tags and misclassified rate components, complicate aggregation across firms. Second, the underlying data appears in multiple formats, including HTML, TXT, and iXBRL. While iXBRL provides structured financial tags, only a fraction of filings utilize this format; other documents require extensive extraction, normalization, and interpretation before they can be incorporated into a unified dataset.

Third, loan-level disclosures frequently contain missing or ambiguous information. Interest-rate fields may be absent or incorrectly populated, and some reported instruments are not debt securities at all, necessitating additional classification before performance can be measured accurately.

Finally, there is no established methodology for aggregating BDC loan-level data into a market-wide benchmark.

As a result, one of the central questions motivating this capstone is: How can we construct a transparent, repeatable private credit index using publicly available BDC filings to enhance visibility into alternative lending markets?

3 Data Pipeline: From BDC Universe to Clean Investment-Level Panel Data

3.1 Identifying BDCs via N-54A / N-54C Filings

Our private-credit index is designed to track the U.S. direct-lending market through lenders that are legally structured as BDCs. The first step is therefore to build an accurate, up-to-date list of BDCs and their CIKs, along with the periods during which each firm was legally a BDC.

3.1.1 Regulatory Basis: N-54A and N-54C

BDC status is governed by two statutory forms:

- Form N-54A – Election to be Regulated as a BDC. Filed when a company elects to be regulated as a BDC under Section 54 of the Investment Company Act.
- Form N-54C – Withdrawal of Election. Filed when a company terminates its BDC status.

These forms are mandatory and timestamped, creating a clean regulatory record of when each firm entered or exited BDC status. Because our index includes all entities that have ever reported BDC portfolio holdings, we systematically extract all N-54A/N-54C filings from the SEC's EDGAR full index.

3.1.2 Fetching BDC Filings from EDGAR (2001–Present)

We implemented a crawler over the EDGAR full-index archives from 2001 to 2025. For each year and quarter:

- Download the quarterly `master.idx`.
- Scan rows where the Form Type is exactly N-54A or N-54C (amendments are excluded).
- Extract:
  - CIK
  - Company name
  - Filing date
  - Filing path (used to build a direct SEC link)

This produces a raw panel of N-54A and N-54C filings over ∼25 years. Some CIKs appear multiple times, reflecting firms that: elect BDC status, later withdraw it, and sometimes re-elect again. To convert these sequences into usable BDC status intervals, we apply the following logic for each CIK:

1. An N-54A starts a new BDC period (start_date).
2. The next N-54C closes that period (end_date).
3. If an N-54C appears without a prior N-54A, we record an interval with a missing start_date (status begins before our sample).
4. If an N-54A has no subsequent N-54C, the company is treated as still active (end_date = NaT).

3.1.3 Output Data Structure

The resulting dataset contains one row per continuous BDC status interval:

Column	**Meaning**
CIK	SEC Central Index Key (company identifier)
Company	Registered company name
start_date	Date of election as a BDC (N-54A)
end_date	Date of withdrawal (N-54C); NaT means "active as of sample end"
Link_A	URL to the N-54A filing that started the BDC period
Link_C	URL to the N-54C filing that ended the BDC period (if any)

3.2 Filing Downloader

3.2.1 Objective

Given the BDC status intervals, the next step is to collect the 10-K and 10-Q filings that contain portfolio holdings. The goals are:

- For each BDC, download all 10-K / 10-Q filings that fall within its BDC status periods.
- Build a central metadata table that fully indexes these filings and can be reused for parsing, EDA and future incremental updates, without repeatedly hitting the SEC APIs.

3.2.2 From BDC Intervals to SEC Submission Records

For each CIK in the BDC interval table:

- Query the SEC submissions API, which exposes a JSON record at `CIK##########.json`.
- Parse:
  - a recent block (latest ∼1,000 filings), and
  - any additional JSON fragments (i.e. `CIK##########-submission1.json`) listed under `files` for older records.
- Convert the JSON into a DataFrame.

This yields a full submission history per CIK.

3.2.3 Constructing a Reusable Metadata Table

We transform these raw submissions into a structured metadata table via:

Date filtering and form selection

- Restrict to specific form types. By default, we retain only BDC-relevant filings, "10-K" and "10-Q".
- Restrict by BDC active window, using `start_date` and `end_date` from Section 2.1. If `start_date` is missing (BDC status predates our sample), we use a default early date (e.g. 2001-01-01).

- Filtering is applied on `filingDate`, ensuring we only keep filings submitted while the firm was legally a BDC.

Selecting core fields We retain only the columns needed downstream: `accessionNumber`, `filingDate`, `reportDate`, `form`, `isXBRL`, `isInlineXBRL`, `primaryDocument`.

Constructing permanent EDGAR URLs Using the CIK and accession number we construct:

- `url` – direct link to the primary HTML filing (used for downloading).
- `index_url` – link to the SEC filing index page (used later to discover XBRL instance documents).

This table is reused as:

- The input list for the XBRL/HTML parsing pipeline,
- A coverage tracker and error log,
- A filter for EDA (by time, BDC, form),
- The basis for future incremental updates as new filings arrive.

3.2.4 Download Workflow

A centralized driver function manages the end-to-end filing download process in an incremental and reusable manner.

For each BDC, we query the SEC submissions API and filter 10-K and 10-Q filings that fall within the firm's active BDC status window. A persistent metadata table (`cache/all_metadata.csv`) is used to track previously downloaded filings, allowing the pipeline to identify and retrieve only new accession numbers while skipping files that already exist locally.

All filings are downloaded into a structured local cache organized by CIK, with rate limiting applied to comply with SEC access policies. After each run, newly discovered filings are appended to the metadata table, which serves as the single source of truth for downstream parsing, coverage tracking, and future incremental updates.

3.2.5 Outputs and Reusability

The filing downloader produces two key artefacts:

Local HTML filing cache

```
cache/
|-- all_metadata.csv
|-- {CIK_1}/  # all 10-K / 10-Q filings for BDC 1
|-- {CIK_2}/  # all 10-K / 10-Q filings for BDC 2
|-- ...
```

3.3 Extracting Schedule of Investments (SOI) Tables

Our goal is to recover machine-readable Schedule of Investments (SOI) data for each BDC filing. We explored two approaches:

- A generic HTML + table-of-contents (ToC) heuristic that works on any HTML filing.
- A more structured iXBRL-based pipeline that leverages "extracted XBRL instance documents".

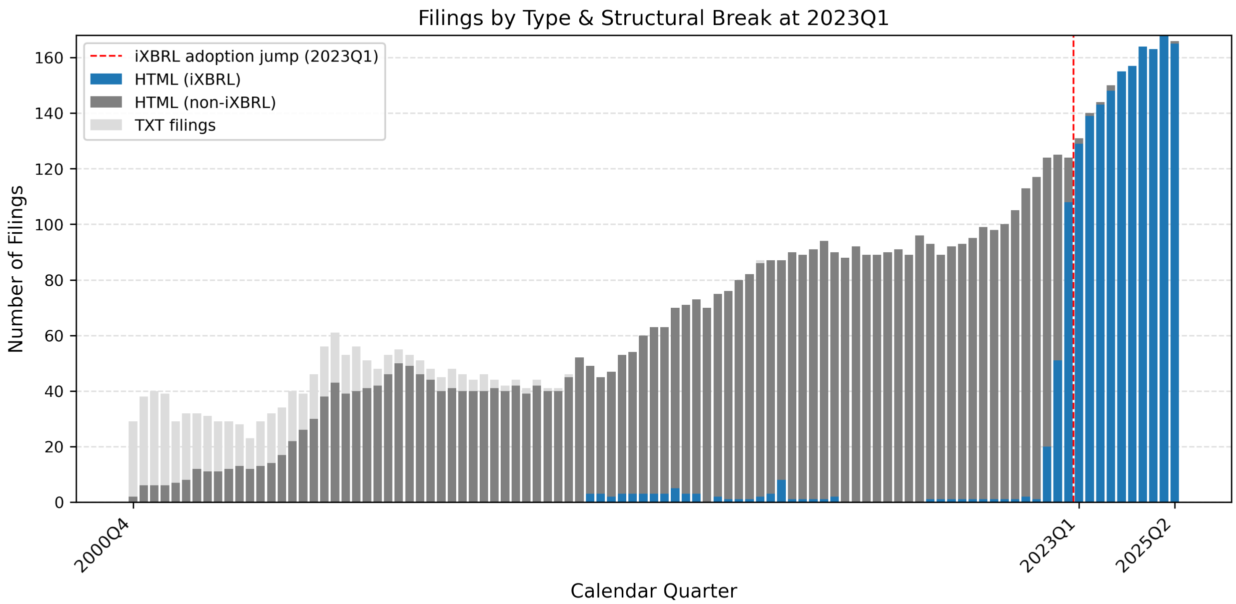


Figure 1: Filings By Type and Structural Break Around 2023Q1

The evolution of filing formats (Figure 1) shows a shift from TXT to HTML and then a sharp increase in iXBRL usage around 2023Q1. In recent quarters, almost all filings are HTML with iXBRL, which makes an XBRL-centric approach both feasible and scalable.

3.3.1 HTML + Table-of-Contents Heuristic

Our initial approach was deliberately format-agnostic:

Parse ToC and locate SOI anchors

- Parse each file with BeautifulSoup.
- Treat every `<a href="#...">` as a potential ToC entry.
- Normalize the link text and match it against a list of SOI-related keywords, e.g.
  - "Schedule of Investments"
  - "Portfolio of Investments"
  - "Loans and Investments"
- The first match is taken as the SOI start; the next anchor with a different target is treated as the end.

Clean fragment and extract tables

- Slice the raw HTML between the start and end anchors.
- Pass the cleaned fragment into `pd.read_html` to get a list of tables.
- Save each table as a CSV and record a manifest with:
  - CIK, file path, anchor names
  - number of tables extracted
  - status (e.g. `no_toc_keyword_found`, `no_parsable_tables`, `error`)

This method worked on many filings but had two critical drawbacks:

- The resulting DataFrames were heterogeneous and noisy (multi-row headers, footnotes, subtotals, inconsistent layouts).
- It was hard to design one cleaning rule set that handled all historical layout idiosyncrasies.

We therefore retained this as an exploratory fallback and transitioned to an iXBRL-first pipeline.

3.3.2 iXBRL-Based Extraction of SOI Facts

Inline XBRL embeds structured tags inside human-readable HTML. Most recent BDC filings provide an "extracted XBRL instance document" on the SEC index page, exposing the same facts in clean XML. Our pipeline uses this as the primary source of SOI data.

Identify the iXBRL universe

- From `all_metadata.csv`, select filings where `isXBRL = 1` and `isInlineXBRL = 1`.
- As shown in the Figure 2, both the number of iXBRL filings and our parsing success rate increase over time, exceeding ∼95% in recent quarters.

Find and download the extracted instance

- For each filing, visit `index_url` and scan table rows whose description contains "extracted xbrl instance document" (with fallbacks to generic "xbrl instance document" / "inline").
- Construct an absolute URL and download the XML into `extract_ixbrl/raw_xml/`.

Understand contexts, dimensions, and members XBRL facts are grouped by contexts (time + dimensional slice). Dimensions attach members via:

- `explicitMember`: chosen from a pre-defined taxonomy list.
- `typedMember`: a custom XML element whose text is filer-defined.

In SOI tables, many BDCs define an `InvestmentIdentifierAxis` and then encode each position as a typed member (the text being a custom investment identifier). Some filers use explicit members instead, but these are less standardized.

In practice, typed members are more common and consistent. The usage and coverage increase over time, in line with iXBRL adoption. Therefore, our extraction focuses on typedMember to map contexts to investment identifiers, while still recording explicitMember for diagnostics or future improvement.

Extract contexts and members For each XML file:

- Build a namespace map from the root.
- Iterate over all <xbrli:context> nodes and collect:
  - context_id
  - period (instant date)
  - all explicit and typed members with member_kind ∈ {"explicit", "typed"}

This yields a contexts table: context_id, dimension, value, period, member_kind, stored per filing and later appended into contexts_master.csv.

Typed-member pipeline for SOI facts The core SOI extraction uses only typedMember:

- Filter contexts to member_kind == "typed".
- Restrict to contexts whose period matches the filing's reportDate to avoid mixing periods.
- For each context, treat the first typed member value as its investment_identifier.

Then:

- Traverse the XML tree and collect all elements with contextRef.
- Exclude structural namespaces (e.g. xbrli, xbrldi, link).
- For each fact, build a column name from the QName (e.g. us-gaap:InvestmentOwnedAtCost), record its value and unitRef.

Assemble a wide table keyed by context_id:

- one row per context,
- columns for each concept, plus ConceptName–unitRef columns,
- include period and investment_identifier.

The per-filing wide table is saved as: extract_ixbrl/per_file/{CIK}/{accession}_tree_extracted_with_al

Finally, we create a long format master_long.csv: each row is one fact: (cik, form, accession, reportDate, context_id, period, investment_identifier, concept, value, unitRef).

The Parsing coverage figure summarizes performance by calendar quarter: success rates stay above 80% and rise above 95% in recent quarters while the iXBRL universe expands.

3.3.3 Output Formats from the iXBRL Pipeline

The iXBRL pipeline produces four main output types:

Per-file wide CSVs
Path: extract_ixbrl_with/per_file/{CIK}/{accession}_tree_extracted_with_allfacts.csv

Contents:

- identifiers: cik, accession, context_id, period, investment_identifier
- a wide set of SOI facts (cost, fair value, principal, rates, spreads, etc.)
- for each concept X, a companion X-unitRef column

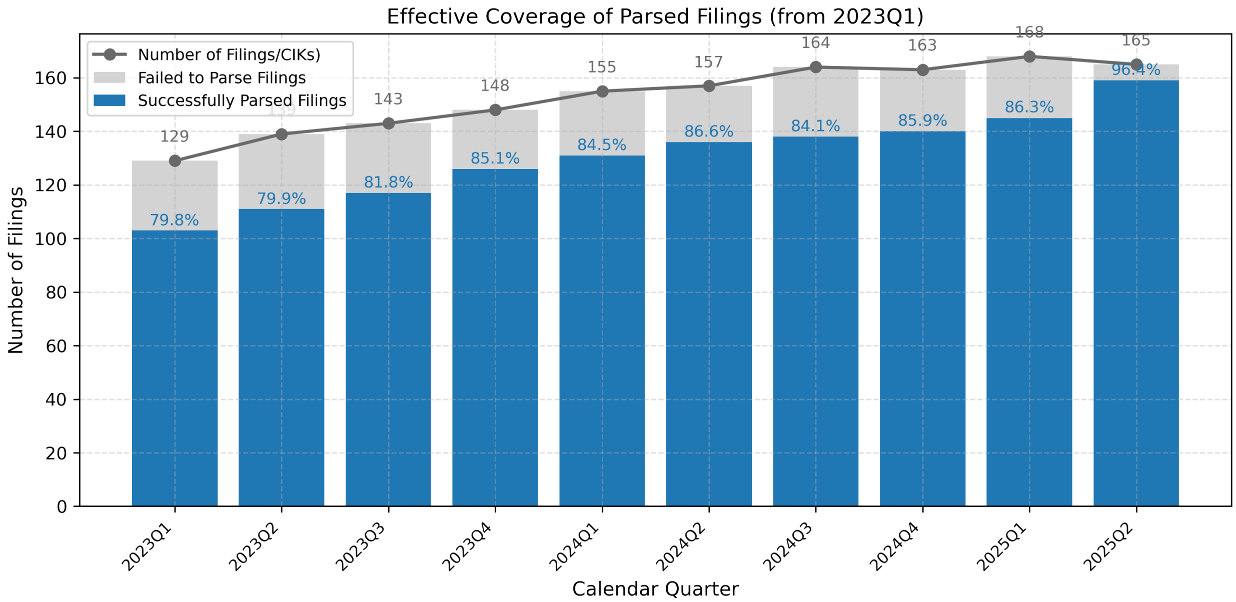


Figure 2: Effective Coverage of Parsed Filings (from 2023Q1)

Per-file context CSVs (diagnostics)
Path: `extract_ixbrl/per_file/{CIK}/{accession}_contexts.csv`
Columns: `context_id`, `dimension`, `value`, `period`, `member_kind`
Used to study how filers encode investment identifiers and other axes.

Global long-format master table
Path: `extract_ixbrl/master_long.csv`
One row per fact; useful for tag-level statistics and reshaping.

Global contexts master
Path: `extract_ixbrl/contexts_master.csv`
Aggregates all context–dimension memberships across filings.

Augmented metadata
Path: `extract_ixbrl/metadata_augmented.csv`
Extends `all_metadata.csv` with parsing flags and enables incremental runs.

3.3.4 Combining Per-File Outputs into an EDA-Ready Dataset

For EDA and cleaning, we consolidate per-file wide tables:

Combine per-file wide CSVs

- Scan `per_file/` for all `*_tree_extracted_with_allfacts.csv`.
- Infer `cik` from the folder and `accession` from the filename.
- Concatenate all DataFrames into `per_file_combined_wide.csv`, optionally de-duplicating rows and sorting by `cik`, `accession`, `context_id`.

Columns pruning in per_file_combined_wide.csv

- Add calendar-quarter labels (`cal_q`, `cal_qe`) from `period`.
- Strip namespace prefixes from column names (e.g. `us-gaap:InvestmentOwnedAtCost` → `InvestmentOwnedAtCost`).
- Group columns with the same local name and merge them via `combine_first`.

This "prefix merge" reduces dimensionality from 1,507 to 1,055 columns (–452) without losing information.

The top-frequency concepts (Figure 3) are exactly those expected from SOI tables.

column	non_null_count	percentage
cik	553594	100.000000
investment_identifier	553594	100.000000
cal_qe	553594	100.000000
accession	553594	100.000000
cal_q	553594	100.000000
period	553594	100.000000
context_id	553594	100.000000
InvestmentOwnedAtFairValue	466906	84.340871
InvestmentOwnedAtCost	447852	80.898998
InvestmentOwnedBalancePrincipalAmount	376893	68.081121
InvestmentBasisSpreadVariableRate	334217	60.372222
InvestmentInterestRate	280078	50.592673
InvestmentOwnedPercentOfNetAssets	255781	46.203716
InvestmentMaturityDate	132931	24.012363
InvestmentInterestRateFloor	78593	14.196866
InvestmentCompanyFinancialCommitmentToInvesteeFutureAmount	71629	12.938905
InvestmentVariableInterestRateTypeExtensibleEnumeration	57757	10.433097
InvestmentOwnedBalanceShares	55660	10.054300
InvestmentInterestRatePaidInKind	40330	7.285122
InvestmentIndustrySectorExtensibleEnumeration	21595	3.900873

Figure 3: Column-wise Data Coverage

3.3.5 Select SOI-relevant fields and restrict the sample

From the merged table, we retain the following categories of variables:

- **Identifiers**
  - `cik`, `accession`, `investment_identifier`
  - `context_id`, `period`, `cal_qe`, `cal_q`
- **Amount variables**
  - `InvestmentOwnedAtCost`
  - `InvestmentOwnedAtFairValue`
  - `InvestmentOwnedBalancePrincipalAmount`
- **Contractual terms**
  - `InvestmentBasisSpreadVariableRate`
  - `InvestmentInterestRate`
  - `InvestmentInterestRatePaidInCash`
  - `InvestmentInterestRatePaidInKind`
  - `InvestmentVariableInterestRateTypeExtensibleEnumeration`

- InvestmentMaturityDate

- **Unit references**
  - unitRef fields corresponding to all monetary amount columns
- **Share balance**
  - InvestmentOwnedBalanceShares

3.3.6 Context classification

For each (cik, accession, context_id) row, we assign a context_type based on the presence of term and amount fields:

- terms_only – only term variables (spread / rate / maturity) present
- amounts_only – only amounts (cost / fair value / principal) present
- mixed – both terms and amounts present
- empty – neither present

In subsequent interest-rate work, we focus on contexts that:

- have no share count (InvestmentOwnedBalanceShares is missing), and
- contain interest-related information (context_type $\neq$ amounts_only, empty).

3.3.7 Interest-rate normalization

Raw spread and rate fields use heterogeneous conventions (decimals, %, bps, sign errors). We therefore:

Define plausible ranges

- spreads / floors: 0%–20%
- coupon rates (cash / PIK): 0%–50%

Normalize values For each value:

- If $|value|$ already lies in range, keep.
- Else, try dividing by 100 or 10,000; if this yields a plausible value, adopt that scale.
- If no scale works, keep the original value and flag as unresolved.

For cash, PIK and floor fields, if the normalized value remains negative, flip the sign and append |sign_error to the scale flag.

The resulting _normalized columns and _scale_flags greatly reduce outliers and make the distributions suitable for EDA and index construction.

Currency normalization Finally, to make value-weighted analyses comparable across BDCs, we convert all monetary fields in `amount_cols` into USD. This proceeds in three steps:

- Normalize currency strings: We map noisy `unitRef` strings (e.g. "usd$", "US dollars", "CAD – Canadian Dollar") to clean ISO codes (USD, CAD, EUR, GBP, ...) using a dictionary-based matcher.
- Merge with quarterly FX rates: Using `FX.csv`, we build a table of average exchange rates per currency per calendar quarter (`cal_q`). USD is assigned a fixed rate of 1.0.
- Convert values and store the result in new variables such as `InvestmentOwnedAtFairValue_normalized`.

This dataset is the basis for all subsequent EDA and the rule-based cleaning described next.

3.4 Rule-Based Cleaning of Interest-Rate Field Misuse

Even after prefix merging and numeric normalization, we observe systematic misuse of interest-rate fields in SOI tables. This arises from how filers design their tables, not from our parsing logic.

Conceptually:

- Fixed-rate loans should report their coupon in `InvestmentInterestRate`.
- Floating-rate loans should report a spread, and sometimes also the total coupon.

In practice, filers often:

- put spreads in `InvestmentInterestRate`,
- put total coupon in PIC,
- put coupon values in the spread field,
- or leave some fields blank when the underlying information is implied (e.g. spread + base).

Because four columns (rate, PIC, PIK, spread) can carry overlapping information, the space of mis-tagging patterns is large. No general approach can reliably reconstruct the intended mapping. If we leave the data as-is, the resulting index displays unrealistic level shifts and volatility. We therefore implement a rule-based cleaning layer grounded in economic logic, internal consistency, and empirical ranges.

3.4.1 Principles Guiding the Rules

Our rules are built around four ideas:

Economic identities as consistency checks For most loans:

$$\text{Total coupon} = \text{PIC} + \text{PIK}$$

$$\text{Total coupon} \approx \text{spread} + \text{base rate}$$

When these identities fail in obvious ways (e.g. coupon $<$ spread, or PIC + PIK implausible), the row likely suffers from column misuse, and then we implement rules to reassign values.

Values must lie within plausible ranges We use percentile-based ranges estimated from the dataset. For example, a "spread" of 15–20% is almost certainly a misreported coupon. Thresholds are computed via 5–95% and 10–90% percentiles rather than hard-coded numbers, making the system data-driven and robust to level shifts.

Estimate missing coupon when spread exists Floating-rate loans are the majority in private credit. When spread is present and coupon is missing, we estimate:

$$\text{coupon} = \text{spread} + \text{SOFR}_{\text{quarter}}$$

For USD loans we use 3-month SOFR by calendar quarter. This materially improves coverage.

Cross-field consistency and prioritisation If exactly one field has a plausible value, we treat it as the intended one. When multiple fields exist but contradict (e.g. coupon < spread), we apply a broad priority:

- PIC + PIK (if available)
- InterestRate (if within coupon range)
- spread + base
- PIK-only (special cases)
- Otherwise: leave unresolved and tag for manual review

The concrete implementation is more nuanced, but follows these priorities.

3.4.2 Representative Rules

As an illustration, consider a common case where `InterestRate` exists but spread is missing (`rate` $\neq$ NaN, `spread` = NaN). We branch on the presence and size of PIC / PIK.

Sub-rule 1 – Standalone rate that behaves like a spread
Conditions:

- rate exists, spread missing
- PIC and PIK both missing
- rate lies in the strict spread range

Action:

- Move rate → spread
- Set rate = NaN

Interpretation: a single small number, with no PIC/PIK, is almost certainly a spread mis-placed into the interest-rate column.

Sub-rule 2 – Rate and PIK exist (PIC missing)
Here spread is missing, PIK is present, and PIC is missing. We distinguish three sub-cases:

2.1: rate > PIK   Action: PIC = rate – PIK
Interpretation: rate is the total coupon; PIK is correctly recorded; PIC is the residual.

2.2: rate < PIK but (rate + PIK) is a plausible coupon   Action:

- interpret rate as cash: PIC = rate
- set rate = PIC + PIK

Interpretation: the reported "rate" is just the cash portion; we reconstruct the full coupon.

2.3: rate < PIK and (rate + PIK) is not plausible   Action:

- treat rate as spread: spread = rate
- treat PIK as coupon: rate = PIK
- set PIC = 0

Interpretation: coupon was entirely placed in PIK, while spread was mis-tagged in rate.

Sub-rule 3 – Rate and PIC exist (PIK missing)
Now PIK is missing, spread is missing, PIC and rate exist:

- 3.1: rate ≈ PIC. Both columns represent the same coupon; we keep PIC as primary.
- 3.2: rate > PIC. Interpret PIC as spread: spread = PIC, PIC = rate (coupon).
- 3.3: rate < PIC. Interpret rate as spread: spread = rate, rate = PIC (coupon).

All rows processed by this rule block are tagged (e.g. `ir_no_spread_case`) so that downstream users can identify and audit corrected records.

Overall, this rule-based layer is essential: without it, the raw iXBRL data produces an index with poor signal quality and unstable dynamics. With it, interest-rate and spread distributions become economically reasonable, and the resulting BDC-based private-credit index behaves in line with market intuition and aligns with the CDLI index.

4 Exploratory Data Analysis

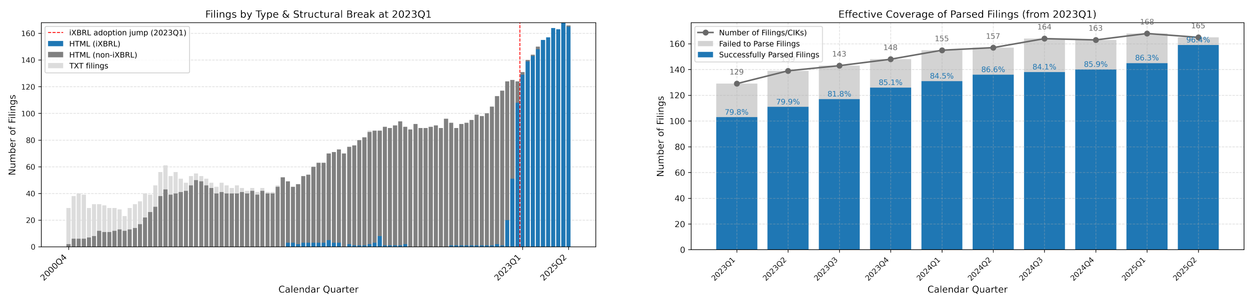
This section provides an overview of the cleaned iXBRL dataset, covering filing coverage, investment counts, PIK prevalence, currency composition, and the behaviour of interest-rate components over time.

4.1 Filing Coverage and Evolution of the BDC Universe

We begin with a long-horizon view of all 10-K/10-Q filings submitted by BDCs since 2001.

In the first chart, each bar corresponds to one calendar quarter and shows:

- the number of filings, and
- the number of BDCs active in that quarter.



(a) Filings by Type

(b) Parsing Coverage

Figure 4: Filing Format Evolution and Parsing Coverage

In 2001, only ∼40 BDCs filed reports; by 2025, this number has grown to ∼170 BDCs, reflecting the rapid expansion of the private-credit sector. The expansion speed accelerates from 2022.

Because SEC reporting standards and technology have evolved, the format quality of filings varies substantially across the sample. The bars are decomposed into: plain text filings, traditional HTML, iXBRL filings.

Only iXBRL filings our extraction algorithm can reliably process. The share of iXBRL filings rises sharply beginning in 2023Q1, after which over 90% of all filings are iXBRL, making automated extraction feasible at scale. For this reason, our final analysis dataset focuses on the 10 quarters from 2023Q1 to 2025Q2, during which both coverage and data quality are high and stable.

The second chart zooms into this window and shows, for each quarter:

- the total number of filings and BDCs, and
- the subset of filings that our pipeline successfully parsed into structured data.

The parsing success rate remains consistently high and gradually improving, indicating that our dataset does not suffer from systematic missingness that could bias downstream analysis. The steady increase in the number of BDCs over this window also confirms the accelerating growth of the BDC universe.

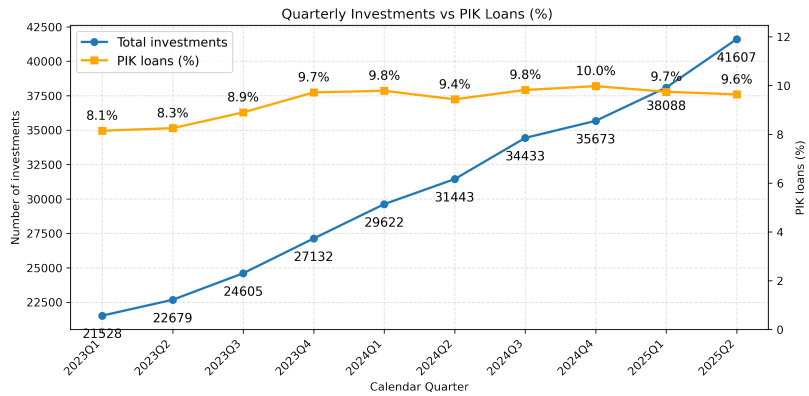


Figure 5: Quarterly Investments v/s PIK Loans

4.2 Sample Size and Prevalence of PIK Loans

Across 2023Q1–2025Q2, our cleaned per-filing dataset contains:

- over 300,000 investment-level observations, and
- ∼30,000 PIK-bearing loans (≈10%).

Two patterns stand out:

Rapid expansion of total investments. The number of investment rows increases from ∼22k in 2023Q1 to over 41k in 2025Q3, again reflecting the continued expansion of BDC portfolios.

Stable share of PIK loans. Despite the rapid growth in total observations, the PIK share remains stable in an 8–10% range. This indicates:

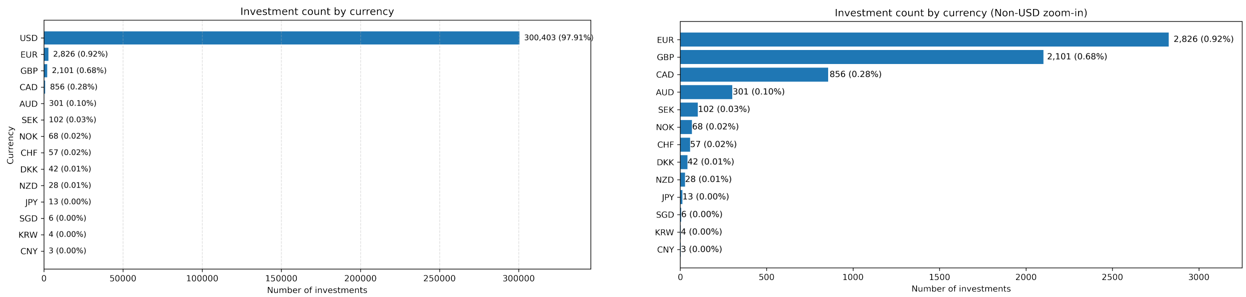
- PIK is a persistent but minority feature of BDC lending,
- PIK usage does not exhibit strong cyclicality in this window, and
- the index is not dominated by PIK loans, but they remain important enough to model explicitly.

4.3 Currency Composition of BDC Portfolios

We examine the reporting currency of investment amounts, based on cleaned unitRef fields and quarterly FX mapping. We found that USD overwhelmingly dominates the loan book. Roughly 98% of all observations are denominated in USD (-300k rows). Non-USD currencies form a long tail, each contributing well below 1% of the sample individually (e.g., EUR, GBP, CAD), with many appearing only a few times. The BDC market is effectively a USD private-credit market, so FX effects play a negligible aggregate role. Our FX normalization step is necessary for consistency, but it has minimal impact on index-level results. For non-USD floating-rate loans, base rates may be more difficult to source, but given their tiny share, this does not materially affect data quality when estimating total coupon via spread + base.

4.4 Distribution of Interest-Rate Components Over Time

Figure 7 summarises quarterly distributions of PIC, PIK, total interest rate, and spread.



(a) Investment Count by Currency

(b) Investment Count by Currency (Non-USD Zoom-In)

Figure 6: Investment Count by Currency

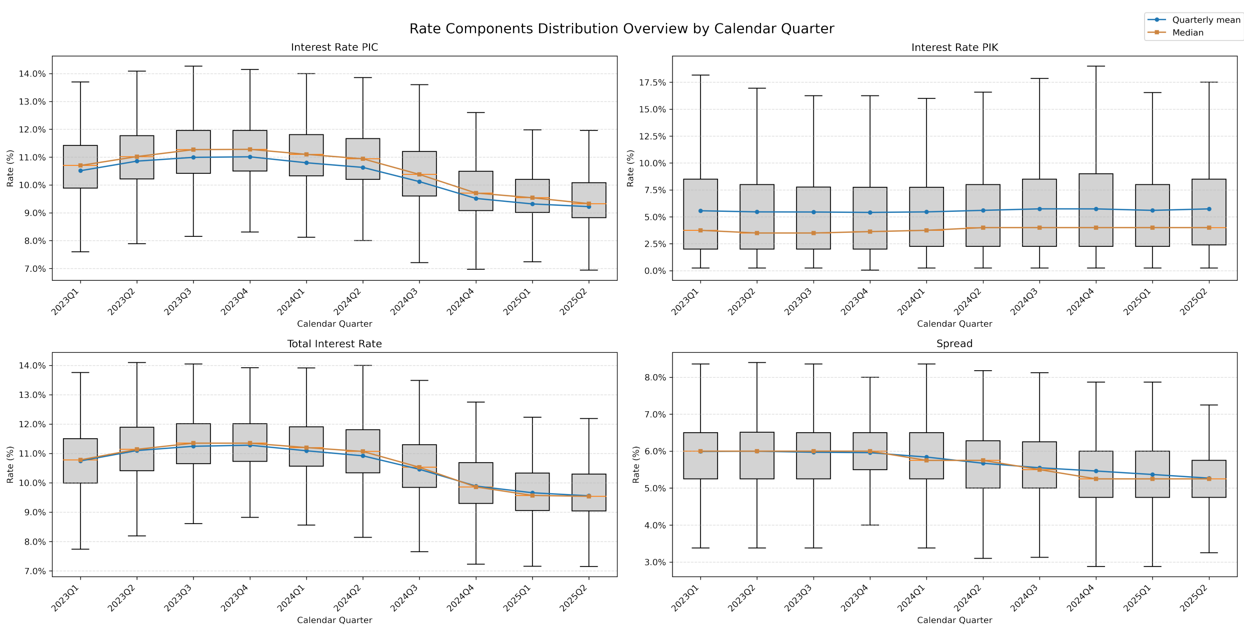


Figure 7: Rate Components Distribution

Main patterns.

PIC and total coupon track each other closely. Cash-pay coupons and total interest rates follow nearly identical trajectories, reflecting the fact that most loans do not feature PIK and that cash interest constitutes the dominant component of total coupon payments. Total interest rates peak around 2023Q3–2023Q4 and decline steadily through 2024–2025, mirroring the broader easing cycle in U.S. short-term interest rates.

PIK rates exhibit greater dispersion. The PIK distribution displays a pronounced right tail, with mean values exceeding medians and a mild upward drift over time, indicating that a small subset of loans carries elevated PIK rates. This pattern is consistent with the use of PIK in higher-risk or stressed financing structures.

Spreads show mild compression. Median spreads decline from approximately 6% in 2023 to around 5% by 2025, with stable interquartile ranges and no abnormal outliers after cleaning.

Across all four rate components, the cleaned dataset exhibits:

- smooth quarter-to-quarter behaviour,
- economically consistent levels,
- no artificial jumps caused by data artefacts, and
- distributions aligned with macro conditions.

5 Index Construction and Results

5.1 Methodology

The objective of the index is to provide a transparent, market-level measure of private credit performance using publicly disclosed BDC data. Given the opacity and infrequent pricing inherent in private credit markets, the index is constructed using a quarterly, flow-based, market-value-weighted framework that closely mirrors how investors experience returns in practice.

The underlying dataset consists of loan-level observations extracted from SEC Schedule of Investments (SOI) disclosures in iXBRL format. A robust data-cleaning pipeline standardizes fair value, cost, par amount, currency, and interest rate fields, corrects filing inconsistencies, and imputes missing rate components using contemporaneous 3-month SOFR when necessary.

Index construction proceeds in five systematic steps.

Data standardization and filtering. Only valid, active loan positions are retained. All monetary values are converted to USD, and interest-rate fields are harmonized across filings to ensure consistency.

Yield reconstruction. Missing values in the index are reconstructed using the following identities whenever necessary:

$$\text{Total Interest Rate} = \text{SOFR} + \text{Spread}, \tag{1}$$

$$\text{Total Interest Rate} = \text{PIC} + \text{PIK}. \tag{2}$$

This reconstruction procedure relies on the assumption that the data cleaning steps described in Section 3 have already corrected cases of column misuse or mislabeling. Under this assumption, we can then use Equation (2) to reconstruct the coupon components. For example, whenever the total interest rate and the PIK component are available but the PIC component is missing, we recover the cash-pay coupon as PIC = Total Interest Rate − PIK. If the available information does not allow the coupon components to be reconstructed using Equation (2), we instead estimate the total interest rate using Equation (1), combining the contemporaneous reference rate (SOFR) with the reported credit spread. This hierarchical approach ensures internal consistency across yield components while maximizing data coverage for downstream index construction.

PIC	Interest Rate	PIK	Count	Percentage
TRUE	TRUE	TRUE	18,100	6.08%
TRUE	TRUE	FALSE	23,533	7.91%
TRUE	FALSE	TRUE	3,922	1.32%
TRUE	FALSE	FALSE	543	0.18%
FALSE	TRUE	TRUE	850	0.29%
FALSE	TRUE	FALSE	244,398	82.13%
FALSE	FALSE	TRUE	4,906	1.65%
FALSE	FALSE	FALSE	1,340	0.45%

Missing Fair Values	0.00%
Missing Cost Values	1.90%
Missing PAR Values	2.59%
Missing PIC Values	84.51%
Missing Interest Rate	3.60%
Missing PIK Values	90.67%

Figure 8: Data Coverage

Position-level return computation. Quarterly returns for each investment are calculated as the change in fair value plus accrued cash and PIK income over the quarter. This formulation captures both income generation and valuation effects.

Market-value weighting. Individual loan returns are weighted by their prior-quarter fair value, ensuring that larger exposures exert proportionally greater influence on index performance.

Index aggregation. The weighted returns are aggregated to produce a quarterly index return. The index level is then computed as cumulative growth from a base value of 100.

This methodology produces a replicable and fully transparent index, avoiding black-box assumptions while remaining closely aligned with how private credit portfolios are managed and reported.

5.2 Index Results

The resulting index exhibits consistent and stable performance throughout the sample period. Quarterly returns range from approximately 1.9% to 3.3%, with the strongest returns observed during mid-2023 when base rates peaked. As monetary conditions eased through 2024 and into 2025, returns moderated slightly but remained firmly positive, typically within the 2.3%–3.0% range.

Over the full horizon, the index level increases from 101.9 in 2023Q1 to 130.5 by 2025Q2, highlighting the strong cumulative impact of steady income generation. The trajectory is notably smooth, reflecting the low volatility and limited mark-to-market noise characteristic of private credit portfolios.

Quarter	Quarter Return	Index Value
2023Q1	1.90%	101.90
2023Q2	2.91%	104.87
2023Q3	3.13%	108.15
2023Q4	3.30%	111.72
2024Q1	2.98%	115.05
2024Q2	2.81%	118.28
2024Q3	2.61%	121.38
2024Q4	2.41%	124.30
2025Q1	2.30%	127.16
2025Q2	2.62%	130.49

Figure 9: Index Returns and Index Levels

Comparison with the Cliffwater Direct Lending Index (CDLI) shows a high degree of alignment. The constructed index achieves an annualized return of approximately 11.23%, with a correlation of 86.95% to CDLI and a tracking error of just 0.33%, validating the accuracy and representativeness of the methodology.

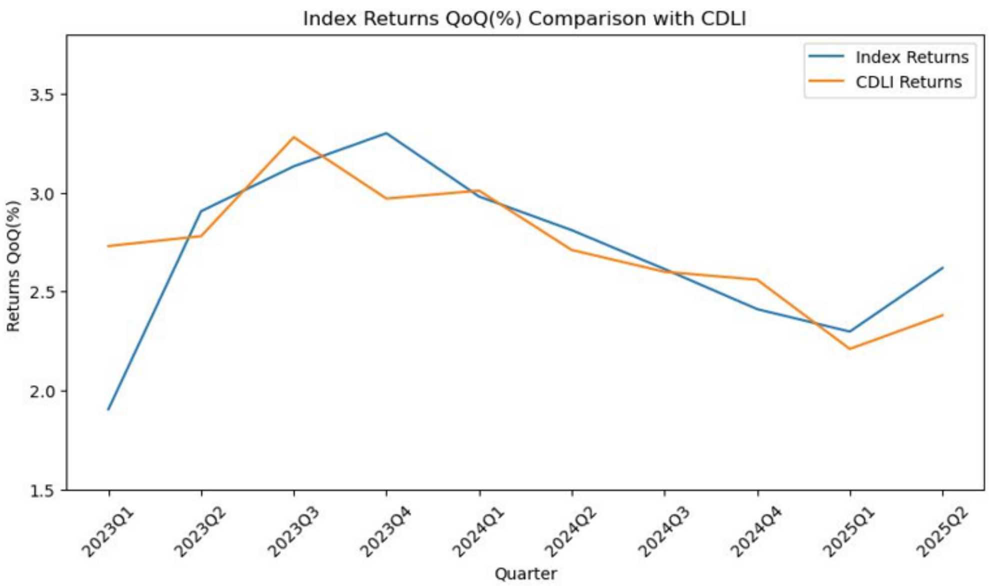


Figure 10: Comparing Index Returns v/s CDLI

Return decomposition further reveals that cash interest income is the dominant driver of performance, contributing roughly 2.3%–2.8% per quarter. PIK income adds a smaller but stable uplift, while price-driven valuation changes are minimal and occasionally negative. Importantly, income consistently offsets valuation noise, resulting in positive total returns across all observed quarters. This confirms that private credit performance is overwhelmingly yield-driven and resilient, even as interest rates normalize.

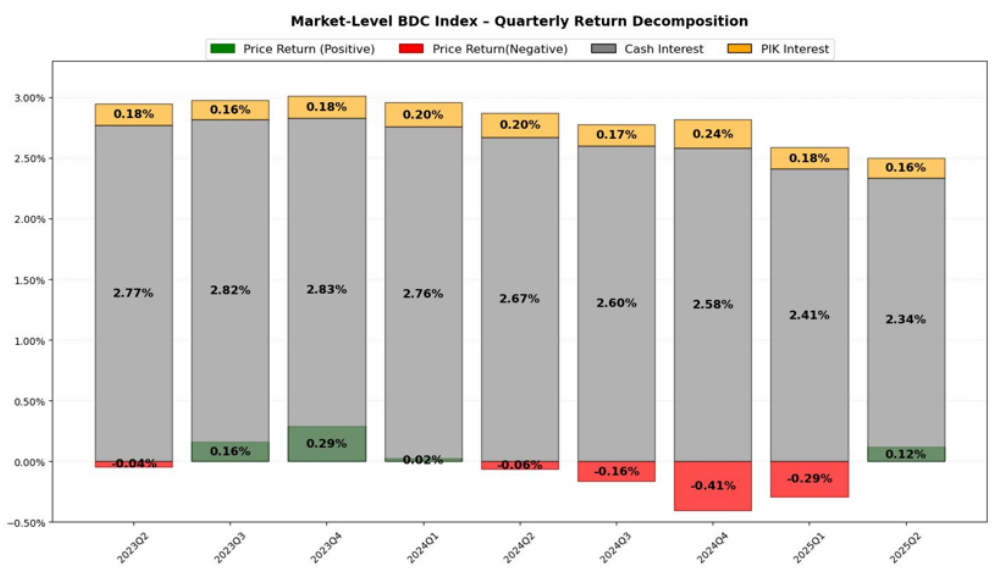


Figure 11: Quarterly Return Decompisition in Cash, Price, and PIK Components

6 Contribution

This paper makes three primary contributions to the measurement and analysis of private credit markets, addressing key limitations of existing indices such as CDLI.

Transparent and reproducible measurement. We extract and process every Schedule of Investments (SOI) line item directly from SEC filings, rather than relying on proprietary or aggregated disclosures. All data-cleaning, valuation, and index-construction steps are fully documented and reproducible. This eliminates black-box methodology and allows users to trace index movements back to individual investment-level inputs.

Granular and flexible index construction. Our approach enables loan-level visibility across the entire BDC universe, supporting valuation comparisons across lenders and borrowers over time. The resulting dataset allows the construction of sector-specific, instrument-based, and custom thematic indices, as well as detailed analysis of PIK usage, spread dynamics, and fair-value versus cost movements. This level of granularity is not available in paywalled or aggregate private-credit benchmarks.

Real-time early-warning capabilities. Because the pipeline operates directly on newly filed SEC reports, it supports near–real-time monitoring of borrower-level stress. The framework can detect early signals such as fair-value markdowns, rising PIK usage, and spread changes before they materially affect aggregate indices. This makes the index suitable not only for performance measurement but also for risk monitoring and forward-looking analysis.

Together, these contributions expand the scope of private-credit indexing beyond existing benchmarks by combining transparency, flexibility, and timeliness with an economically grounded and scalable methodology.

6.1 Sub-index generation: Seniority

One of the main advantages of building a loan-level dataset is that we can create transparent, rules-based sub-indices that match how investors actually segment private credit risk. A first application is a seniority sub-index, which allows us to track senior and unitranche performance using our extracted and cleaned loan data, and to compare it to an existing market benchmark (CDLI-S).

6.1.1 CDLI-S

Cliffwater publishes CDLI sub-indices to benchmark specific lending strategies. CDLI-S (CDLI-Senior) is defined as being comprised primarily (95%+) of senior and unitranche loans held within BDCs. We use CDLI-S as a reference benchmark to evaluate how closely our seniority-based sub-index tracks a widely used senior private-credit measure.

6.1.2 Methodology

We construct seniority labels using a rules-based engine that maps common private-credit terminology found in filings (instrument descriptions, security names, debt-type fields, etc.) into a standardized seniority taxonomy.

Step 1: Normalize text fields. Deal- and instrument-level text fields are cleaned and standardized, including case normalization, punctuation removal, and harmonization of common abbreviations such as "1L", "2L", and "TL".

Step 2: Apply keyword-mapping rules. Examples of rule triggers include:

- `SENIOR_SECURED_1L`: "first lien", "1st lien", "1L", "senior secured", "revolving credit facility" (when clearly senior secured)
- `SENIOR_SECURED_2L`: "second lien", "2nd lien", "2L"
- `UNITRANCHE`: "unitranche", "first-out/last-out"
- `SUBORDINATED_MEZZ`: "mezzanine", "subordinated", "junior"
- `SENIOR_UNSECURED`: "senior unsecured", "unsecured notes"

When no rule is triggered with sufficient confidence, the observation is labeled `OTHER_UNKNOWN` to avoid forcing noisy classifications.

Step 3: Merge labels back to the loan panel. Seniority labels are merged back into the final cleaned dataset using a unique key, ensuring full traceability from each label to its originating filing context.

6.1.3 Coverage and distribution

After applying the seniority rules, coverage is as follows:

- Classified: 200,457 (64.07%)
- Unclassified: 112,427 (35.93%)

The distribution across seniority buckets is:

Table 1: Distribution of Loans by Seniority Classification

Seniority Category	**Count**	**Share (%)**
Senior Secured 1L	153,394	49.03
Senior Secured 2L	31,235	9.98
Senior Secured Unspecified	7,274	2.32
Unitranche	4,025	1.29
Subordinated / Mezzanine	3,385	1.08
Senior Unsecured	1,008	0.32
CLO Structured Equity	136	0.04
Unclassified	112,427	35.93

The classified set is dominated by senior secured first-lien loans, consistent with expectations when isolating senior private-credit exposures. The unclassified bucket is intentionally conservative and reflects cases where filing language is insufficiently specific to classify without introducing noise.

6.1.4 Sub-index construction and validation against CDLI-S

Using the senior-only subset (senior secured and unitranche loans), we compute quarterly returns for our senior sub-index and compare them to CDLI-S.

The results are:

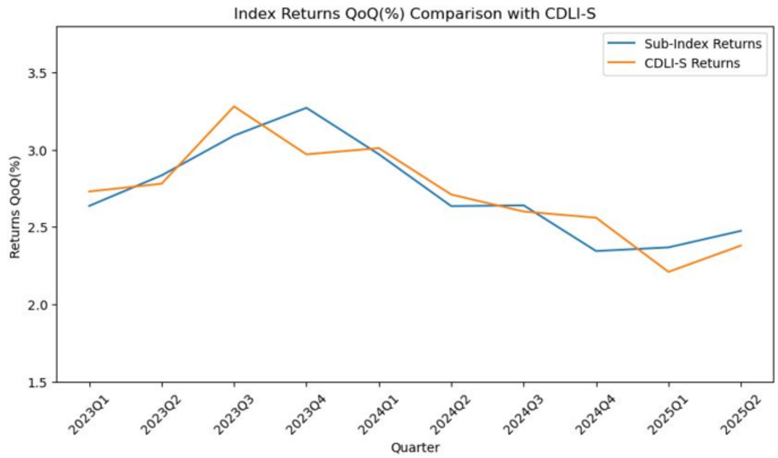


Figure 12: Senior-Index Returns v/s CDLI-S

- Correlation: 87.5%
- Tracking Error: 0.32%

These statistics indicate that our seniority-based sub-index captures similar directional dynamics as CDLI-S, with only modest dispersion in returns. This supports the conclusion that a transparent, loan-level seniority mapping can produce a credible senior private-credit signal comparable to an established benchmark.

6.2 Sub-index generation: Sectors

Beyond seniority, the pipeline extends naturally to more complex attributes that do not map cleanly to deterministic keyword rules. Sector classification is a prominent example: sector labels are often implicit, inconsistently described, or missing entirely in raw filings. Our approach combines extracted loan-level data with an LLM-based classifier and explicit confidence thresholding.

6.2.1 Methodology

For each loan, an LLM is prompted to infer the borrower or asset sector using available text fields. The model returns both a sector label and a confidence score. We then apply the following rule:

- If confidence $\geq$ threshold: accept the sector label.
- If confidence $<$ threshold: assign `Other / Unknown / Unresolved`.

This design makes the precision–coverage tradeoff explicit and auditable. Higher thresholds improve precision but reduce coverage, while lower thresholds increase coverage at the cost of greater uncertainty.

A known limitation is that LLMs may hallucinate or overcommit when input text is sparse. Confidence thresholding mitigates this risk, though further improvements are possible via longer context windows, retrieval augmentation, or consensus-based prompting.

The sector taxonomy used in this analysis is as follows: Communication & Media; Consumer; Energy & Utilities; Financials & Insurance; Healthcare; Materials; Real Estate; Technology; Transportation; and Unknown.

6.2.2 Coverage results at different thresholds

Sector coverage under two confidence thresholds is reported below:

- 70% threshold: 222,953 classified (71.26%), 89,931 unclassified (28.74%)
- 90% threshold: 175,066 classified (55.95%), 137,818 unclassified (44.05%)

At the 90% threshold, fewer observations are labeled, but those retained are higher confidence and better suited for formal reporting. At the 70% threshold, coverage increases and is more appropriate for exploratory analysis.

The top three sectors at the 90% threshold are:

- Technology: 54,904 (17.55%)
- Healthcare: 39,145 (12.51%)
- Financials & Insurance: 26,281 (8.40%)

6.2.3 Example sector sub-index: Technology

Using the classified sector labels, we construct sector-specific return series. Sector returns can further be decomposed into their underlying components, allowing us to identify the drivers of performance quarter by quarter.

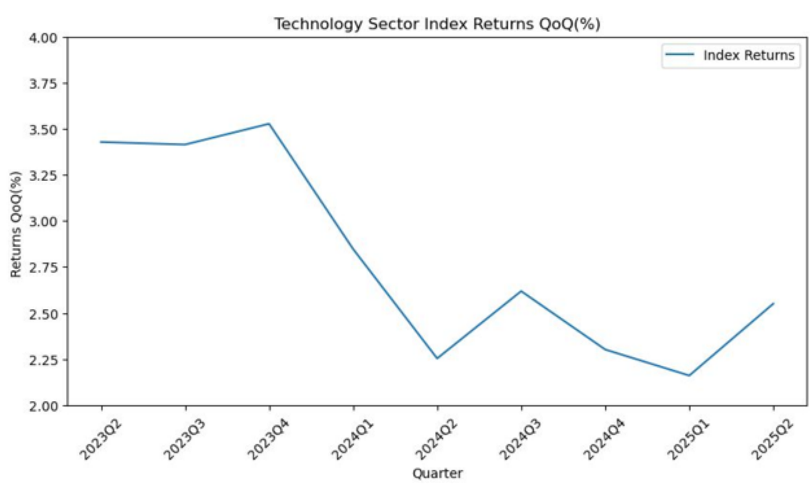


Figure 13: Technology Sector Index

This type of decomposition highlights the value of a richer dataset: rather than producing only headline index numbers, we can explain performance dynamics and compare behavior across sectors.

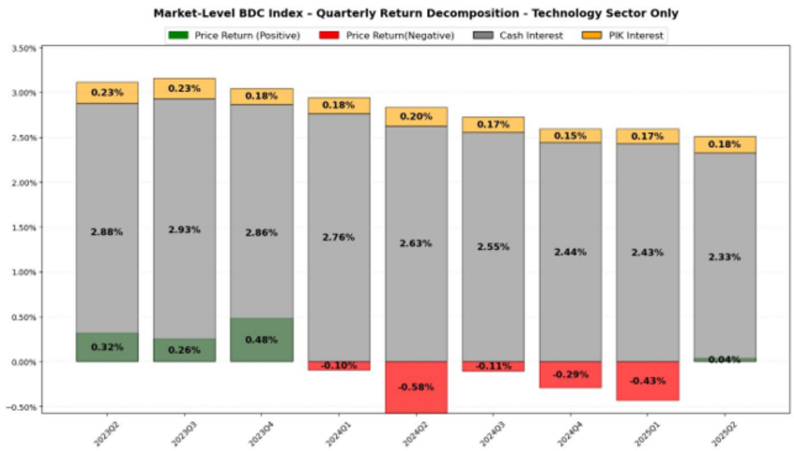


Figure 14: Technology Sector Index - Return Decomposition

In this section, we illustrate this flexibility by constructing two representative sub-indices: a seniority-based sub-index and a sector-based sub-index. Importantly, these examples are not meant to be exhaustive. Because our underlying panel is loan-level and standardized across the BDC universe, the same framework can generate customized, on-demand indices along virtually any dimension available in the data—such as instrument type, yield/spread ranges, PIK features, maturity profiles, currency, or user-defined thematic screens.

6.3 Advanced analysis example: Good vs. Bad PIK loans

Once loan attributes are extracted and standardized, the dataset supports deeper diagnostics and early-warning signals. The Good versus Bad PIK analysis illustrates this capability.

6.3.1 Definitions

Bad PIK loans are defined as loans that either (i) initially had no PIK feature but later introduced PIK, or (ii) experienced an increase in the PIK component over time. Both patterns are commonly interpreted as signals of deteriorating credit conditions or rising borrower stress, as lenders shift compensation from cash-pay to deferred interest. Good PIK loans are those where PIK is present from inception and remains broadly stable over time, or where no meaningful increase in the PIK component is observed throughout the loan's life.

6.3.2 Findings

The first result is a quarterly trend in the share of Bad PIK loans. This share rises from 9.8% in 2023Q1 to 33.6% in 2025Q2 (as a fraction of total PIK loans). This suggests that, within our sample, the incidence of PIK being added later has grown meaningfully, consistent with a tougher credit environment.

The second result compares PIK rate distributions for Good and Bad PIK loans using quarterly boxplots. Good PIK loans display relatively stable distributions with a gradual upward drift in mean and median. Bad PIK loans exhibit greater dispersion and shifting distributions, consistent with transitions into more stressed credit structures.

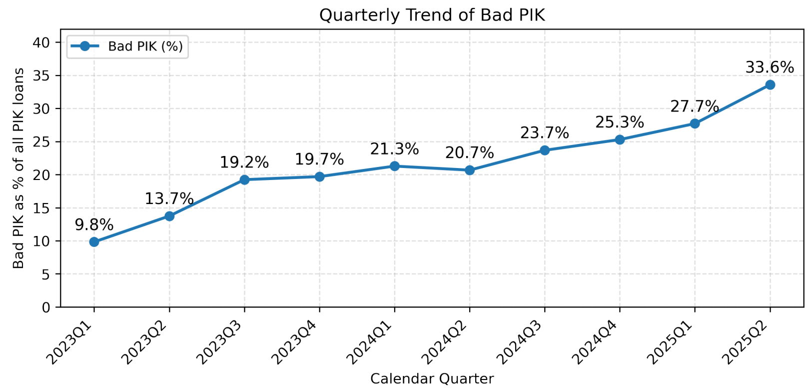
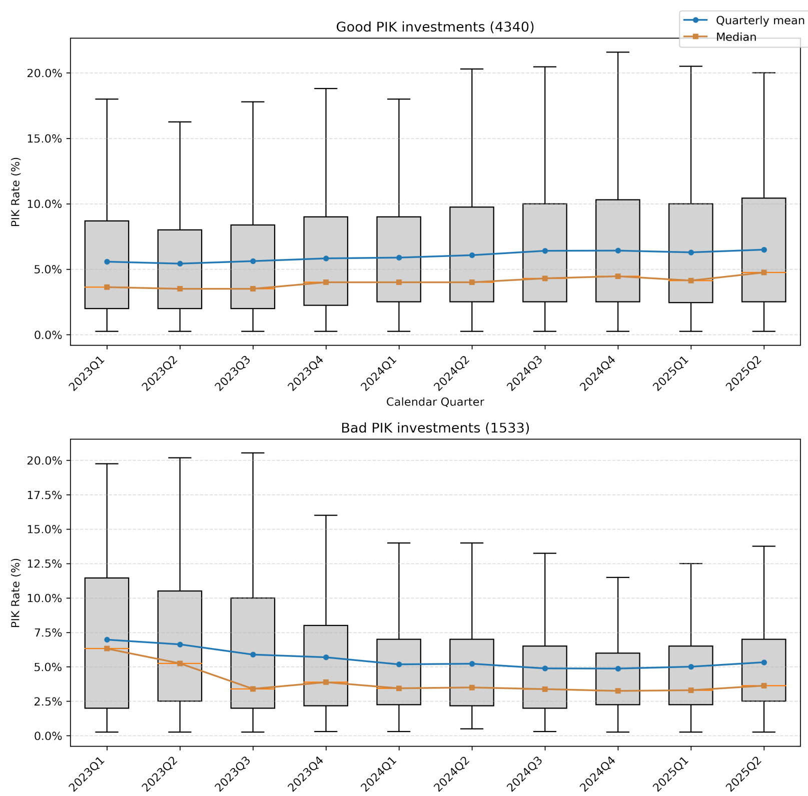


Figure 15: Prevalence of Bad PIK Loans over the Years



6.3.3 Why this matters

This analysis is possible because the dataset enables tracking of contractual features over time at the loan level. The Good/Bad PIK split is only one example. The same framework can be extended to other transition signals, including covenant changes, spread resets, reporting-language shifts, or valuation-methodology changes.

Overall, richer and standardized private-credit data increases transparency, allowing analysis to move beyond aggregate index summaries toward explainable, loan-level insights into risk buildup and return drivers.

7 Case Study: First Brands Group

This section presents a detailed case study of First Brands Group to illustrate how the loan-level dataset and analytics developed in this paper can be used to identify early warning signals, track valuation dynamics, and assess lender-level exposure concentration in private credit markets.

7.1 Background and Motivation

First Brands Group filed for Chapter 11 bankruptcy in September 2025 after disclosing more than $10 billion in liabilities, including significant off-balance-sheet financing exposure. The firm was a widely held borrower across private credit markets, with material exposure spread across multiple BDCs.

The case is particularly instructive because it demonstrates how credit deterioration can be detected well before a formal default event. Broad lender participation and growing exposure meant that valuation changes at First Brands had the potential to affect a meaningful fraction of the BDC universe. This highlights the importance of systematic monitoring tools that operate at the loan level and allow for cross-BDC comparability.

Our dataset enables such analysis by providing consistent extraction of Schedule of Investments data across BDCs, improving transparency on loan terms, valuations, and their evolution over time.

7.2 BDC Exposure Dynamics

Using the cleaned and standardized loan-level dataset, we track the evolution of BDC exposure to First Brands Group over time.

Between 2023Q1 and 2025Q3:

- The number of exposed BDCs increased from 4 to 15, representing approximately 10% of the BDC universe.
- The number of loan positions grew from 6 to 25.
- Total cost exposure expanded from approximately $70 million to over $250 million.

These dynamics confirm the broader market narrative: First Brands transitioned from a relatively contained borrower to a systemically important exposure within private credit markets. Increasing exposure combined with rising lender participation implied growing systemic vulnerability.

Notably, First Brands loans did not involve PIK interest, underscoring the importance of valuation-based signals—rather than contractual features alone—as early indicators of stress.

7.3 Exposure Concentration Across BDCs

Exposure to First Brands was highly concentrated among a small subset of lenders. The top five BDCs collectively accounted for approximately 78% of total exposure, with the largest single lender holding 36%.

This concentration implies that bankruptcy-related losses would disproportionately affect a small number of institutions. For publicly traded BDCs (two of the top 5 BDCs are publicly traded), such concentration translated directly into net asset value volatility and stock price pressure.

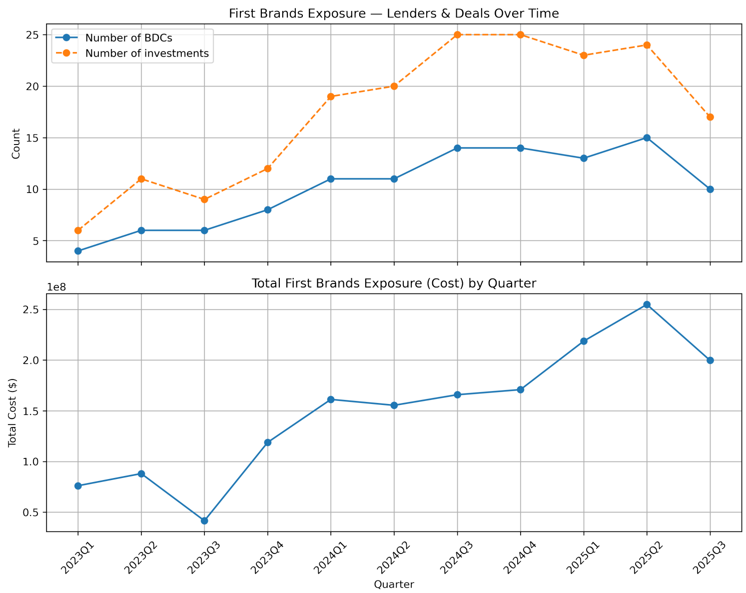


Figure 16: BDC Exposure Dynamics to First Brands Group

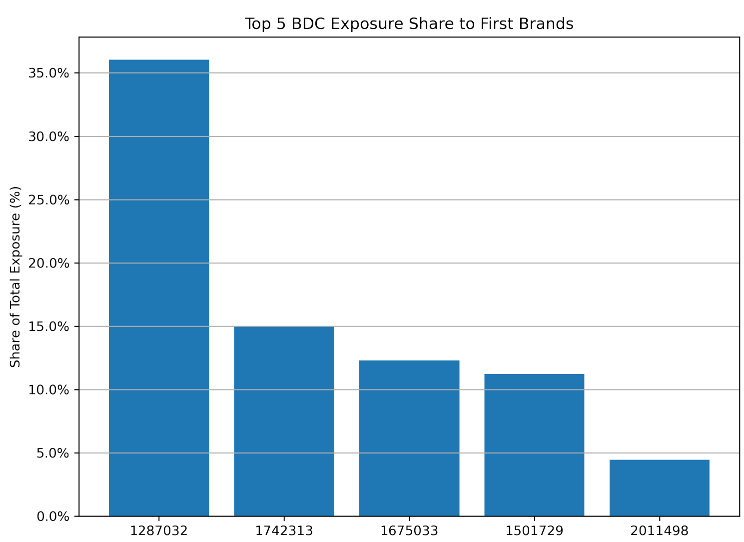


Figure 17: Top Five BDC Exposure Share to First Brands Group

7.4 Valuation Trajectory: FV/Cost Ratio

The evolution of the fair value-to-cost (FV/Cost) ratio provides a clear, interpretable signal of credit deterioration. The valuation trajectory can be divided into three distinct phases:

- **Stable valuation (2023–2024Q2):** FV/Cost remained tightly clustered around 0.99–1.01, indicating broad consensus that the loan was performing.
- **Early warning phase (2024Q3–2024Q4):** FV/Cost declined to approximately 0.95, signaling emerging stress well before the bankruptcy filing.
- **Major impairment (2025Q3):** FV/Cost collapsed to approximately 0.35 as bankruptcy or restructuring became fully priced in.

This pattern demonstrates that valuation-based signals embedded in SOI disclosures provide meaningful early warnings ahead of discrete default events.

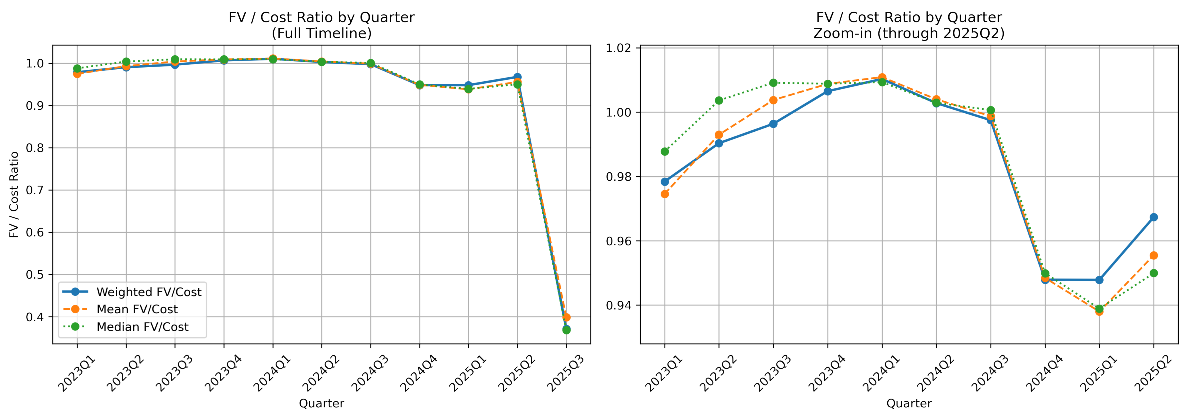


Figure 18: FV/Cost Ratio Trajectory for First Brands Group

7.5 Valuation Dispersion as a Stress Indicator

Beyond the level of valuations, dispersion across lenders provides additional insight. From 2023Q1 to 2024Q3, fair value markdowns were extremely tight, reflecting strong agreement across BDCs.

From 2024Q4 to 2025Q2, valuation dispersion widened meaningfully as some lenders began marking down earlier than others. By 2025Q3, markdowns ranged from approximately 10% to over 90%, consistent with contemporaneous media reports highlighting large valuation disagreements.

Such dispersion likely reflects differences in internal valuation models, recovery assumptions, and lien positions. Importantly, dispersion itself emerges as a leading indicator of credit stress.

7.6 Identifying Vulnerable Lenders

Finally, we examine valuation trajectories at the individual BDC level. The results reveal substantial heterogeneity in credit risk management practices:

- Some BDCs initiated gradual markdowns as early as 2024Q4, suggesting more conservative valuation approaches.
- All BDCs marked down by 2025Q3, but the severity of markdowns varied dramatically.

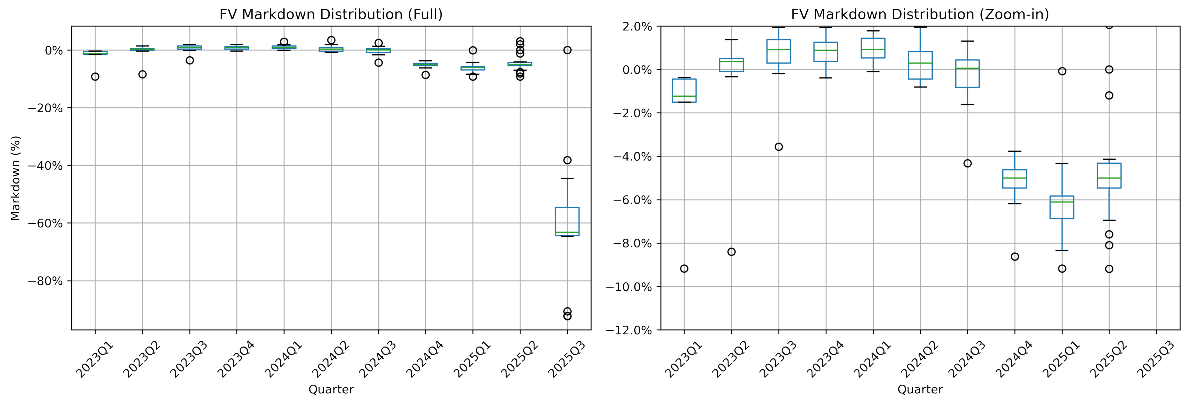


Figure 19: Valuation Dispersion and Fair Value Markdowns

- Five BDCs exited their positions before 2025Q2, avoiding the most severe losses.

These trajectories allow us to identify which lenders tend to mark conservatively versus aggressively and which systematically lag broader market signals. The analysis underscores the importance of examining lender-level behavior rather than relying solely on aggregate indices.

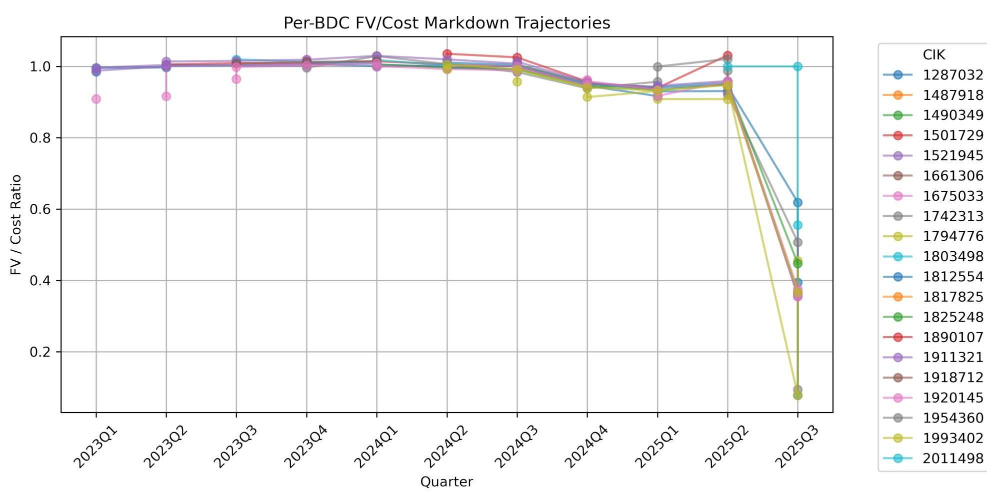


Figure 20: Per-BDC FV/Cost Markdown Trajectories

7.7 Implications

The First Brands case study demonstrates the analytical power of a clean, loan-level private credit dataset. By tracking exposure growth, valuation levels, dispersion, and lender-specific behavior, the framework enables early detection of credit stress and supports more transparent, explainable risk assessment. These capabilities extend well beyond headline index construction and provide a foundation for real-time monitoring and stress diagnostics in private credit markets.

8 Future Scope

The framework developed in this paper naturally extends to several directions that can further enhance the timeliness, coverage, and analytical power of private credit indices.

Higher-frequency index construction and nowcasting. An important extension is the development of higher-frequency private credit indices with reduced reporting lag. By incorporating interim disclosures, rolling updates from newly filed SOI tables, and partial-quarter aggregation, the pipeline can support nowcasting of private credit returns. This would materially improve the usefulness of the index for real-time monitoring, particularly during periods of rapidly changing market conditions.

Return smoothing using expanded data coverage. As data coverage expands across filers, time, and instrument types, the index can be further desmoothed by leveraging a larger cross-section of observations. Increased coverage allows idiosyncratic valuation noise to average out more effectively, yielding cleaner signals of underlying market dynamics while preserving economically meaningful variation.

Expanded sub-index ecosystem. Beyond seniority and sector sub-indices, the same loan-level infrastructure can be used to construct indices along additional dimensions such as industry, market size, yield bucket, or risk profile. This opens the door to transparent benchmarks designed to track existing market products (e.g., CDLI-V, CDLI-P, CDLI-UMM, CDLI-LMM) using fully documented and reproducible methodologies.

Improved coverage and signal strength. Finally, continued growth in iXBRL adoption and improvements in extraction techniques will increase both the breadth and depth of the dataset. Greater coverage enhances statistical power, strengthens early-warning capabilities, and enables more robust cross-sectional and time-series analysis of private credit risk and performance.

Together, these extensions highlight the flexibility of a loan-level, rules-based private credit dataset and underscore its potential to evolve from a quarterly benchmark into a comprehensive, real-time monitoring and research platform.

9 Conclusion

This project demonstrates that meaningful transparency in private credit markets can be achieved using publicly available regulatory disclosures, when paired with systematic data engineering and rigorous index construction. By leveraging loan-level SOI data from BDC filings, the project overcomes many limitations of traditional private credit benchmarks, including opacity, delayed reporting, and restricted access.

The constructed index provides a reliable, market-representative measure of private credit performance, closely tracking established benchmarks such as CDLI while offering materially greater granularity. Beyond headline returns, the framework enables decomposition of income and valuation effects, construction of customized sub-indices by sector or seniority, and identification of emerging credit stress through borrower-level and lender-level analytics.

Empirical results reinforce the view that private credit delivers stable, income-dominated returns with low volatility, supporting its role as a core allocation in diversified portfolios. The First Brands Group case study further illustrates the practical value of the dataset, showing how valuation dispersion and exposure concentration can act as early warning signals well before widely reported distress events.

Overall, the project establishes a scalable and extensible foundation for future private credit research and product development. With improved data accessibility, transparent methodology, and enhanced analytical capabilities, this framework positions investors, researchers, and regulators to make more informed decisions and gain deeper insight into the evolving private credit landscape