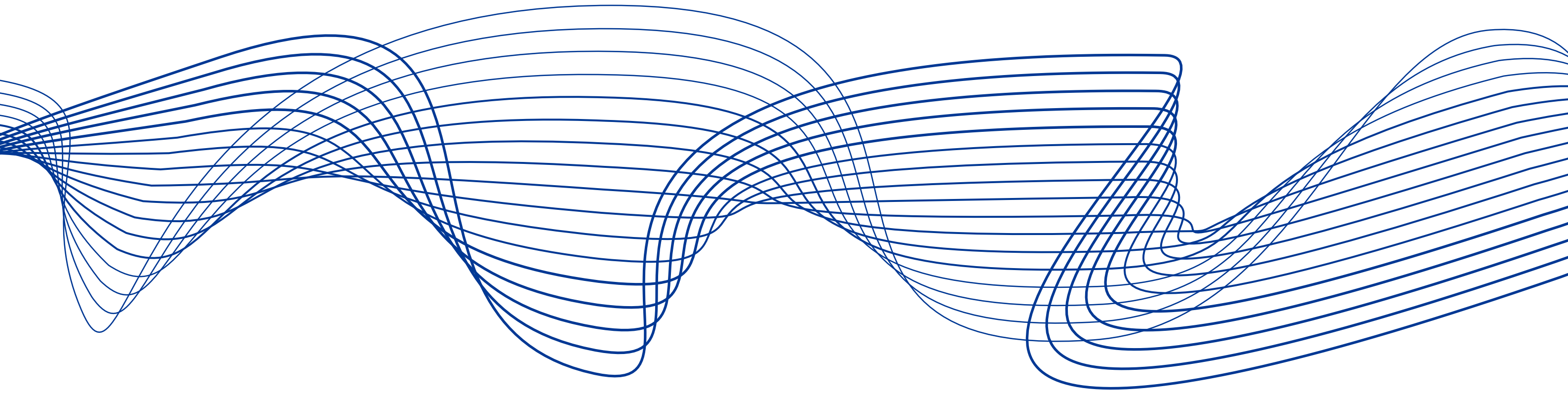
ESRB
European Systemic Risk Board
European System of Financial Supervision

Occasional Paper 30

No labels, no problem: Identifying investment fund cohorts through clustering



Sophie Almekinders, Antoine Bouveret, Massimo Ferrari, Michael Grill, Mattia Pividori, Roberto Proietti, Daniel Jonas Schmidt

Contents

Abstract

This paper applies a tailored clustering approach to identify cohorts of investment funds for financial stability assessment. To define clusters, we use regulatory data on asset class exposures reported under the Alternative Investment Fund Managers Directive (AIFMD). We applied our approach to more than 10,000 alternative investment funds (AIFs) holding €3.7 trillion in assets, revealing 12 economically interpretable fund cohorts, including traditional bond and equity funds, liability-driven investment (LDI) funds, and private asset funds. Our cluster-based approach substantially outperforms traditional AIFMD categories in explaining fund return variance and points to material vulnerabilities, notably concentrated leverage in GBP-denominated LDI funds and widespread liquidity mismatches. The dispersion of these cohorts across EU jurisdictions underscores the need for oversight and cross-border coordination in ensuring the macroprudential oversight of the investment fund sector. This framework provides regulators, supervisors and macroprudential authorities with a practical framework for identifying funds whose collective behaviour could amplify systemic risks during periods of market stress.

Keywords: Investment funds, clustering, financial stability, macroprudential policy, portfolio similarity, collective behaviour

JEL codes: G01, G11, G18, G23, C38

Executive summary

The growth of the asset management sector has created new channels for systemic risk transmission within the financial system, particularly through the collective behaviour of funds with similar investment strategies. Recent crisis episodes have underscored the relevance of this issue. During the 2022 UK gilt crisis, liability-driven investment (LDI) funds exacerbated the turmoil in the gilt markets through collective selloffs (see Dunne et al., 2023). Similarly, during the March 2020 “dash for cash” episode, significant sales of Treasuries by leveraged hedge funds worsened strains in the US Treasuries market (see Financial Stability Board, 2020). These examples illustrate how groups of funds with similar exposures can amplify market stress through synchronised asset liquidations, thus magnifying instability across financial markets. While regulatory frameworks increasingly recognise the importance of monitoring fund groups rather than individual entities, existing classification systems remain insufficient for reliably identifying cohorts that may behave collectively during periods of market stress.

This paper further develops a clustering methodology for supervisory use on investment fund data at scale to identify cohorts of funds with similar exposures. We apply our clustering approach to data reported under the Alternative Investment Fund Managers Directive (AIFMD), focusing on funds classified as “other” or “none” – broad residual categories that encompass €3.7 trillion in assets across more than 10,000 funds in Q4 2024. Our methodology incorporates detailed exposure information across 70 asset categories and eight geographic regions and accounts for the hierarchical structure of asset classes. We identify 12 distinct fund cohorts with clear economic interpretations, including traditional bond and equity funds, GBP-denominated LDI funds, private asset funds, and mixed investment vehicles. The identified fund cohorts outperform traditional AIFMD classifications by 16 percentage points in explaining fund return variance.

Our cluster-based analysis identifies groups of funds posing financial stability risks that traditional regulatory classifications based on fund types alone may fail to capture. For example, our approach allows us to identify those GBP-denominated LDI funds that exhibit the highest leverage levels (an average gross leverage level of 355% of NAV), while eight out of 12 clusters show substantial liquidity mismatches between portfolio assets and investor redemption terms. Financial institutions, particularly insurance companies and pension funds, dominate the investor base across all clusters, creating direct channels for systemic risk transmission. The geographic dispersion of similar fund cohorts across EU jurisdictions highlights the critical importance of cross-border coordination in macroprudential oversight.

This paper provides authorities with tools to move beyond individual fund analysis towards a systematic evaluation of collective risks. In practice, clustering helps surface funds within broad categories that share similar exposures and risk profiles – such as GBP-denominated LDI funds with very high leverage – and groups with pronounced liquidity mismatches, enabling more targeted monitoring, prioritisation,

and stress-testing than when relying on coarse labels alone. At the same time, our clustering approach is more a diagnostic screening tool than a prescriptive categorisation: it indicates where collective behaviours may arise, although any supervisory or regulatory response should be based on a case-by-case assessment, supported by additional evidence and existing frameworks, thereby avoiding a mechanical translation of clusters into rules. The identification of fund cohorts with distinct risk profiles in AIFMD data thus demonstrates the practical value of our clustering approach for regulators, supervisors and macroprudential authorities.

1 Introduction

The growth of the asset management sector over the last two decades has provided investors with a range of investment options and increased the funding available to the real economy, but it has also introduced new potential financial stability risks. Open-ended funds, which allow investors to redeem shares frequently, often hold a mix of assets with varying levels of liquidity, creating inherent liquidity transformation risks. Many funds also employ leverage to enhance returns, amplifying both potential gains and losses while making them more vulnerable to market shocks. In periods of market stress, forced asset sales – whether triggered by investor redemptions, margin calls or portfolio rebalancing – can exacerbate downward pressure on market prices and destabilise broader financial markets.

The financial stability risks posed by investment funds often stem more from collective behaviour than from the individual actions of asset managers. Most individual funds are relatively small and therefore appear to pose limited systemic risk on their own. The greater concern arises when a large number of funds hold similar portfolios and face comparable incentives, potentially prompting correlated responses during periods of market stress.[1] In such episodes, funds can react to shocks in broadly related ways – for example, by selling similar assets at once, drawing liquidity from overlapping sources, and adjusting leverage through comparable mechanisms. Such coordinated responses can transform what might otherwise be manageable market pressures into severe disruptions. Coordination failures and strategic complementarities among investors (see Chen et al., 2010) amplify these effects, as procyclical asset sales by similar funds create powerful feedback loops capable of destabilising entire market segments.

Past stress episodes have laid bare the risks posed by the collective behaviour of similar funds. In the case of GBP-denominated liability-driven investment (LDI) funds, no individual fund was large enough to create systemic stress. However, these funds had very similar exposures to long-term gilts and used repos and derivatives to leverage their positions. When faced with simultaneous margin calls during the mini-budget crisis of September 2022, their collective selling of sovereign bonds amplified stress in core markets, particularly given their large market footprint in certain segments of the UK government bond market (see Dunne et al., 2023; ESRB, 2023). Similarly, during the March 2020 "dash for cash" episode, heavy investor redemptions forced asset sales, which in turn amplified the initial shocks across multiple asset classes (see FSB, 2020). These episodes also demonstrate that the propagation of risk throughout the system can occur without institutional failure; coordinated asset sales alone may be sufficient to trigger significant negative effects.

The regulatory and policy framework emphasises the importance of assessing risks to financial stability that may stem from groups of similar funds. In the

---

[1] That said, there have been notable instances where a single fund caused severe market disruptions—such as Long-Term Capital Management in 1998 and Archegos Capital in 2021.

EU, all alternative investment funds (AIFs) undergo an annual risk assessment by the national competent authorities (NCAs), which must consider how groups of funds could threaten financial stability. The Guidelines on Article 25 of the AIFMD explicitly provide that: “Competent authorities should consider risks posed by common exposures” (ESMA, 2021). Meanwhile, the FSB recommendations on non-bank leverage direct authorities to consider: “the combined positions of non-bank financial entities with similar business models or investment strategies” (FSB, 2024). This regulatory focus reflects the growing recognition that effective prudential oversight means understanding not just individual fund risks, but also the systemic implications of collective behaviour among similar entities.

This paper’s key contribution is a data-driven method that rigorously identifies economically meaningful groups of similar funds, complementing – and helping to address limitations in – existing regulatory classifications. While regulators are becoming increasingly cognisant of the importance of the collective behaviour of funds, little work has been done to identify groups of similar funds using data-driven methods. Most work relies instead on regulatory classifications, which are applied inconsistently and feature broad residual categories that provide little insight into the underlying investments. This results in regulatory labels that often fail to capture which funds hold similar portfolios in practice. We demonstrate how clustering can be used to objectively identify funds with similar exposures and risk profiles, filling a critical gap in the toolkit of financial stability authorities.

Our method builds on clustering algorithms that enable a systematic, data-driven approach to identifying groups of entities with similar exposures. Such algorithms measure pairwise similarity (or distance) between observations and aggregate those that are closest under a chosen metric. As an unsupervised machine learning technique, clustering infers natural groupings directly from the input data rather than relying on predefined categories. Applied to investment fund exposures, clustering can reveal groups with comparable portfolio allocations that may, in turn, display similar behaviour during periods of market stress.

We apply a clustering methodology specifically designed for high-dimensional fund exposure data. Our methodology builds on the K-medoids algorithm proposed by Kaufman and Rousseeuw (1987), which partitions observations into K clusters by minimising within-cluster distances and representing each cluster by a representative data point. We develop a custom Manhattan distance metric that accounts for the hierarchical structure of asset classes, weighting broad asset class distinctions (such as government versus corporate bonds) more heavily than granular sub-classifications. The primary inputs are exposures to 138 asset classes, complemented by exposures to eight geographical regions and one leverage metric.

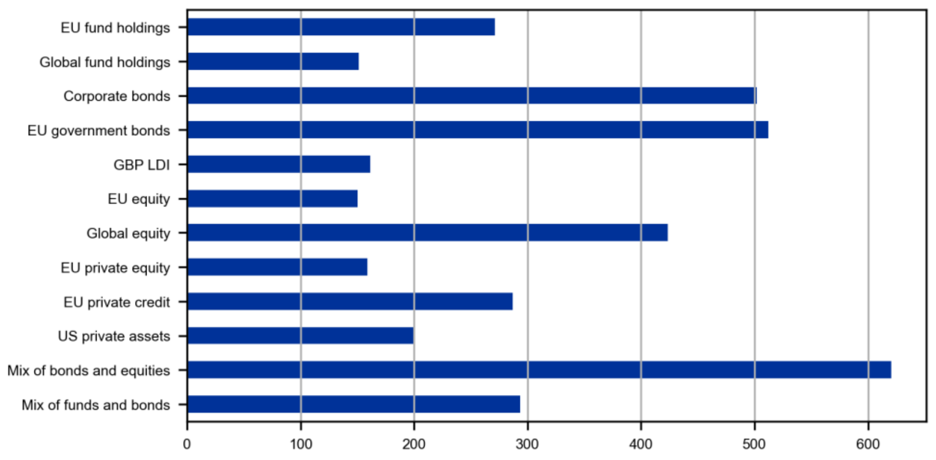
We use regulatory data on AIFs to demonstrate the benefits of our clustering methodology. Our analysis is based on AIFMD data, which provide comprehensive reporting on fund exposures, leverage, liquidity profiles and investor composition. At the end of 2024, the dataset contained information on more than 26,000 EU-domiciled AIFs with a net asset value (NAV) of roughly €7.2 trillion. While the AIFMD classification provides a useful starting point, a large share of funds are placed in broad residual categories that offer limited insight into the underlying portfolios, namely “other” AIFs (about 50% of total NAV) and funds with no predominant type (“none”; around 4%), which together provide a suitable application context for our methodology.

Our clustering of AIFs labelled “other” or “none” reveals 12 distinct fund cohorts with clear economic interpretations. Our sample of AIFs grouped into these residual categories represents €3.7 trillion in assets under management across 10,191 funds. Figure 1 below shows the 12 clusters: equity funds, government and corporate bond funds, liability-driven investment (LDI) funds, private asset funds, funds investing primarily in other fund holdings, mixed bond-equity funds, and funds combining bonds with fund holdings. Most of these groups are further differentiated by geographical focus.

Figure 1

Net asset value of the identified AIF cohorts

(EUR billions)



Sources: AIFMD data (2024 Q4), own calculations.
Notes: The figure displays the 12 clusters identified in Section 4.1 based on a sample of EU-domiciled AIFs of the types “other” or “none”. The cluster labels refer to the primary asset type and geographical exposures in each cluster.

A comparison with self-reported AIFMD strategies reveals the advantages of our data-driven approach over the AIFMD classification. Our approach separates the heterogeneous group of funds in the “other – other” strategy into more homogeneous groups and provides greater granularity for the somewhat vague “equity” and “fixed income” strategies within the “other” AIF type. Moreover, we classify funds that resemble funds of funds and private equity funds within the “other” category, despite these being established AIFMD categories, highlighting the inconsistent application of regulatory classifications.

Multiple validation exercises confirm the economic meaningfulness and robustness of our identified clusters. A comparison with a hand-collected list of LDI funds shows that over half of the funds in our GBP LDI cluster were independently identified as LDI funds, with only one known GBP LDI fund falling outside our cluster. Our clustering explains more than twice as much variance in fund returns compared with self-reported AIFMD strategies, demonstrating superior predictive power in terms of risk profiles. An analysis of data from 2021 to 2023 shows that clusters are relatively stable, with 86% of funds maintaining the same cluster assignment in 2023 compared with 2024, declining to 77% when comparing the 2021 and 2024 clusters.

Our cluster-based analysis reveals potential financial stability risks arising from similar exposures within cohorts of funds that the regulatory classification may fail to detect. For example, GBP-denominated LDI funds exhibit by far the highest leverage levels, averaging 355% in gross terms and 257% after netting and hedging, yet these risks may be obscured when such funds are grouped with other AIFs under AIFMD reporting. By identifying clusters based on actual portfolio composition, our approach allows for a more accurate assessment of structural vulnerabilities. The analysis also highlights widespread liquidity mismatches, particularly in clusters where funds invest in less liquid assets but offer frequent redemption terms. We find substantial mismatches between portfolio liquidity and investor redemption terms in eight out of 12 clusters, especially over short time horizons. Although LDI and private asset funds appear to be more structurally protected due to their redemption restrictions, reported liquidity metrics generally reflect normal market conditions and may underestimate risks under stress. Investor composition further heightens the risk of contagion. Financial institutions – especially insurance companies and pension funds – dominate the investor base across nearly all clusters, creating direct channels through which AIF losses could propagate into the broader financial system. The presence of other investment funds as investors in private asset clusters also introduces intra-sector linkages that could amplify stress via cross-holdings and correlated redemptions.

Fund cohorts are geographically dispersed across multiple EU jurisdictions, highlighting the importance of cross-border supervision and coordination in risk assessment. While some clusters are concentrated in specific countries – such as mixed funds in Germany and LDI funds in Ireland and Luxembourg – most are distributed across several jurisdictions. As a result of this geographical distribution, individual NCAs may have only a partial view of EU-wide risks. The European Securities and Markets Authority (ESMA) and the European Systemic Risk Board (ESRB) therefore play a vital coordinating role in ensuring that collective behaviour and cross-border spillovers are effectively monitored and addressed.

We also look at how our methodology could be applied to the entire universe of EU-domiciled AIFs. Our results for this extended sample of more than 20,000 AIFs (€6.6 trillion in NAV) show strong alignment between the identified clusters and well-defined AIF types such as funds of funds, private equity, and real estate funds. Our analysis confirms the robustness of the clusters based on the sample of “other” funds, while also highlighting funds with exposures that appear to be inconsistent with their reported AIF type. These funds account for around €400 billion in NAV and may merit supervisory scrutiny.

Our work builds on prior analyses of AIFMD data at ESMA, ESRB and the ECB. The most important related work is the chapter titled “AIFMD fund classification – shedding light on ‘other AIFs’” in ESMA (2020), which documents the issue of the large, opaque “other” category in AIFMD data. ESMA attempts to gain a better understanding of these funds by using the ECB fund classification, but with limited success due to the lack of Legal Entity Identifiers (LEIs) and International Securities Identification Numbers (ISINs) in AIFMD data and because the alternative regulatory classification, while potentially clearer in some respects, still faces similar challenges in capturing the diversity of fund strategies. Recent work by Haquin and Proietti (2024, 2025) and Bouveret et al. (2025A, 2025B) focuses on risks posed by highly leveraged AIFs, while ESRB (2023) looks specifically at LDI funds.

Our work is related to the existing academic literature on clustering financial institutions and portfolio similarity. Girardi et al. (2021) provide crucial empirical evidence by examining portfolio similarity and asset liquidation in the insurance industry, demonstrating that entities with similar portfolios tend to sell in unison during stress periods and that such similarity is often linked to collective behaviour. Their work supports the assumption that funds with similar exposures may act collectively during periods of market stress. Additional academic work includes Farnè and Vouldis (2017), who apply clustering to bank balance sheets to identify business models within the euro area, as well as Lucas et al. (2019) and Custodio João et al. (2023, 2024), who develop dynamic clustering algorithms and apply them to banks and insurance companies. Barucca et al. (2021) use holdings data from UK banks, insurance companies and funds to define clusters of entities with similar exposures.

To the best of our knowledge, this paper represents the first time clustering has been applied to regulatory investment fund data to assess financial stability risks. For example, Sakakibara et al. (2015) apply clustering to mutual fund data for portfolio optimisation. Satone et al. (2021) focus on commercial uses such as recommender systems and marketing, while Vozlyublennaia and Wu (2016) discuss the relationship between fund uniqueness and performance. Our work extends beyond existing approaches because our methodology has been specifically designed for the hierarchical structure of asset class data and can be effectively applied to risk assessment.

2 Data on investment funds reported under the AIFMD

This section introduces the regulatory dataset on AIFs that we use throughout our analysis. The first part outlines the legal foundation for the data collection. The second part examines how investment funds are categorised in the dataset and highlights the key limitations of the existing classification. These limitations warrant the development of a data-driven alternative. Lastly, we present an overview of the key variables that we use in our empirical analysis.

2.1 The Alternative Investment Fund Managers Directive

The Alternative Investment Fund Managers Directive (AIFMD) provides the regulatory framework governing the management and oversight of AIFs in the EU. Unlike UCITS funds, which are designed for retail investors and are subject to strict rules on liquidity and portfolio diversification, AIFs – such as hedge funds, real estate funds and private equity funds – are typically geared towards professional investors and face fewer product-level constraints, including on leverage. AIFMD applies to fund managers that are either domiciled in the EU or that manage or market alternative investment funds within the EU, creating a harmonised regulatory and supervisory environment for AIFMs operating across Member States. Article 25 of the AIFMD enables national competent authorities (NCAs) to impose leverage limits when leveraged AIFs pose systemic risks. So far, this macroprudential tool has been applied to Irish property funds (CBI, 2022) and GBP-denominated LDI funds (CBI, 2024; CSSF, 2024).

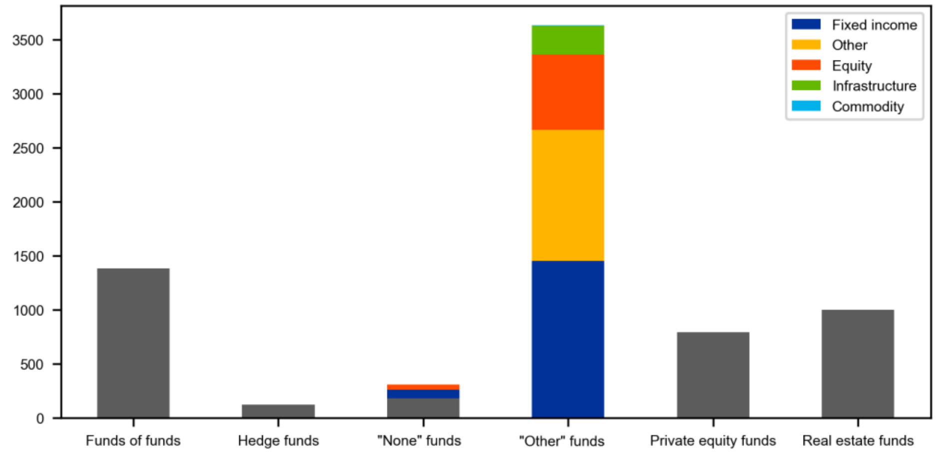
The AIFMD introduces extensive and regular reporting obligations that enable authorities to monitor the activities of AIFMs and assess risks to financial stability. Fund managers must report granular data on each fund at a quarterly, semi-annual or annual frequency, with the specific reporting frequency and the reported variables depending on characteristics such as fund size, use of leverage and domicile. While certain smaller, sub-threshold funds are exempt from regular reporting, end-of-year data are available for the full population of non-exempt AIFs, allowing for a broad-based analysis of the sector. AIFMD data serve as a foundational input for NCAs’ annual leverage risk assessments under Article 25 (see Haquin and Proietti, 2024, 2025) and support broader macroprudential surveillance by ESMA and national authorities.

2.2 Regulatory classification of investment funds under AIFMD reporting

The AIFMD requires fund managers to classify AIFs based on their predominant AIF type and primary strategy. The AIF types available for reporting purposes include “hedge fund”, “private equity fund”, “real estate fund”, “fund of funds” and “other”. If a fund spans multiple AIF types, managers can select “none” to indicate the absence of a single predominant type. In addition to the type, managers must provide a breakdown of the fund’s investment strategies and identify the primary one. For instance, private equity strategies include “venture capital” and “growth capital”, while real estate strategies include “commercial” and “residential”. A complete list of strategies is available in ESMA (2014).

Figure 2
Net asset value by type and primary strategy reported under the AIFMD

(EUR billions)



Source: AIFMD data (2024 Q4).
Notes: The labels on the x-axis are AIF types. The colours indicate strategies that belong to the “other” type. Grey bars indicate strategies that belong to funds of funds, hedge funds, private equity funds or real estate funds. “None” funds, i.e. funds without a predominant type, use strategies from different AIF types. We refer to all “other” and “none” funds as “broad residual categories” because AIF types are mutually exclusive, whereas strategies are not. For example, if an AIF classified under the type “other” reports two strategies – let’s say 60% equity and 40% other – it is still reasonable to consider it part of a broad residual category, even though equity is its primary strategy. This approach ensures consistency in how we interpret and group these funds.

The residual “other” category is by far the largest AIF type, with an aggregate NAV of €3.6 trillion, accounting for 50% of the total NAV of EU-domiciled AIFs (see Figure 2). This group includes all funds that do not follow fund of funds, hedge fund, private equity or real estate strategies. Within the “other” category, the most common primary strategies are “fixed income” (€1.5 trillion), “other” (€1.2 trillion) and “equity” (€0.7 trillion). Infrastructure and commodity funds make up only a small share, with NAVs of €266 billion and €5 billion, respectively.

The second major residual group of funds labelled with “none” as their predominant type represents an aggregate NAV of €306 billion, accounting for 4% of the total, and cannot be clearly separated from “other” AIFs in practice. According to the AIFMD, the “none” designation applies when a fund combines strategies across multiple AIF types, whereas “other” AIFs report combinations within a limited set of strategies, namely fixed income, equity, other, infrastructure and commodity. However, the data show that almost none of the funds classified as “none” pursue hedge fund, private equity or real estate strategies. Instead, they typically combine fund of funds strategies with strategies associated with “other” AIFs. Furthermore, our analysis finds that many “other” AIFs hold substantial positions in other funds. Because of this overlap, a practical distinction between “none” and “other” AIFs is difficult to establish.

The lack of clear definitions and the predominance of broad residual categories in the AIFMD classification significantly limit the usefulness of the data. The two largest type-strategy combinations, namely “other – fixed income” and “other – other”, together account for 37% of total NAV, yet both categories lack specificity. For example, “fixed income” could refer to government bonds, corporate bonds or loans, and does not distinguish between traditional bond funds and more specialised vehicles such as LDI funds. Similarly, the “other – other” category is a virtual black box, providing little to no clarity about the nature of the underlying investments. It remains unclear whether such funds invest in a mix of fixed income and listed equities, or broader multi-asset portfolios, or are specialised vehicles such as private LDI funds or private credit funds. Moreover, the AIFMD classification lacks precise thresholds for identifying funds of funds and fails to provide a clear definition of hedge funds. As a result, entities that resemble funds of funds or hedge funds may be misclassified as “other”. Our findings in Sections 4.2 and 6.2 add to this concern, showing that funds with similar asset allocations are dispersed across different AIF types and strategies. This points to inconsistencies and confusion in how the AIFMD classification system is applied.

2.3 Variables reported under the AIFMD

Fund managers report their fund’s exposure to different asset classes, which we then use as the basis for an alternative, data-driven classification of AIFs. Exposures are divided into five “macro asset types”: securities, derivatives, physical assets, collective investment undertakings, and other. These are further subdivided into “asset types” and “sub-asset types”, the latter of which has 70 categories and forms the basis for our clustering approach.[2] For example, “listed equities” is an asset type that belongs to the macro asset type “securities” and is further sub-divided depending on whether the issuer is a financial or non-financial company. See ESMA (2014) for a complete list of asset types and sub-asset types. Long and short positions are reported for almost all sub-asset types. The only exceptions are foreign exchange derivatives and interest rate derivatives, for which gross positions are reported. Fund managers also report assets under management (AuM), defined as the sum of absolute values of all positions. Individual position information is available for the five largest holdings of an investment fund, but not for the entire portfolio.

The geographic breakdown of exposures is also included in AIFMD reporting. These exposures are categorised by continents or economic regions: Africa, Asia

---

[2] The AIFMD reporting template provides full details of the 70 asset classes.

and Pacific, European Economic Area (EEA), non-EEA Europe, Middle East, North America, South America and “Supranational / Multiple Regions”. While the sub-asset types of the “sovereign bonds” asset type contain some information on the geographical breakdown (EU, non-EU G10, non-G10) for this particular asset type, AIFMD does not otherwise provide statistics on exposures to combinations of asset types and geographic regions.

Fund managers are required to report on their use of leverage, liquidity mismatches in funds, and other variables that are valuable for our assessment of financial stability risks.[3] Leverage must be reported both using the gross method and the commitment method, along with a breakdown by source of leverage (unsecured borrowing, collateralised borrowing and derivatives). Other useful inputs for our analysis include information on redemption terms, the portfolio liquidity profile (fraction of the portfolio capable of being liquidated within a given time period), the investor liquidity profile (fraction of investor shares that can be redeemed within a given time period), information on the investor base, and the jurisdiction of the fund manager. We use these variables in Section 5 for a financial stability assessment based on the identified clusters. Fund managers also report monthly returns and we use the return time series in Section 4.3 to validate the identified clusters of funds.

---

[3] See ESMA (2014) for a detailed description of data reported under the AIFMD.

3 A clustering approach for AIFMD data

Clustering techniques offer a robust, data-driven approach to classifying investment funds by grouping entities with similar characteristics. As an unsupervised learning technique, clustering identifies groups of similar data points based solely on input features, without using predefined labels. This approach is particularly useful in the context of AIFMD data, as the existing AIFMD classification provides limited insight into the exposures and strategies of investment funds, making it less effective for identifying cohorts of funds that are significant from a financial stability perspective. By leveraging clustering, we can group together funds with similar exposures, enabling a more meaningful and accurate classification of funds.

In this section, we present our clustering algorithm specifically designed for fund exposure data and apply it to the AIFMD dataset. First, we describe the custom measure of portfolio similarity across funds. Second, we outline K-medoids, our preferred clustering algorithm for identifying groups of funds based on their similarity, given its robustness and stability compared with alternative approaches. Third, we describe additional methodological choices, such as the construction of the input data and the choice of the number of clusters.

3.1 A distance metric for fund exposure data

The choice of the distance metric is critical in clustering, as it determines which observations are considered close or distant, thereby directly influencing the formation of clusters. A distance metric is mathematically defined as a function $d(x, y)$ that takes two points, x and y, as inputs and satisfies four standard conditions: identity (the distance between a point and itself is zero), non-negativity, symmetry, and the triangle inequality. Since clustering algorithms form groups based on relative proximity, the choice of the metric can significantly affect outcomes, particularly in high-dimensional spaces. An ill-suited metric may cluster dissimilar funds together or split similar ones apart, resulting in misleading patterns and poor interpretability. For this reason, selecting a metric that reflects the nature of the input data is essential to obtain meaningful results (see Kumar et al, 2014).

When using normalised portfolio exposures as input data, the distance metric should reflect how much the fund portfolios overlap in terms of asset allocation. Because the absolute size of a fund should not influence the clustering outcome, we normalise overall exposures by dividing each fund's nominal exposure to an asset class by the fund's total exposure. This yields a vector of normalised exposures, x, where $\sum_i x_i = 1$. For such data, an appropriate distance metric should satisfy two key properties:

1. Funds with no overlapping exposures should be assigned the maximum possible distance.

2. Funds with overlapping exposures should be assigned smaller distances than those without any overlap.

Asset classes typically follow a hierarchical structure, with broad categories subdivided into more specific ones, and this structure should be reflected in the distance metric. For example, let's consider government bonds, financial sector equities, and non-financial sector equities. The two equity categories share exposure to overall equity market risk despite being from different sectors. This makes them more similar to each other than either is to government bonds, where interest rates are the primary risk factor. Consequently, equity funds that focus on different sectors exhibit more similar risk-return profiles to each other than they do to government bond funds. We therefore add a third requirement for the distance metric:

3. The distance metric should reflect relationships between asset classes, assigning smaller distances to funds investing in similar asset classes than to those investing in very different asset classes.

Table 1
Distance matrices for hypothetical fund portfolios

a) Manhattan distance

	Fund 1	Fund 2	Fund 3	Fund 4
Fund 1	0	2	2	2
Fund 2		0	2	1.8
Fund 3			0	1
Fund 4				0

b) Manhattan distance for hierarchically-structured asset classes

	Fund 1	Fund 2	Fund 3	Fund 4
Fund 1	0	1	2	1.9
Fund 2		0	2	1.8
Fund 3			0	0.6
Fund 4				0

Source: Own calculations.
Notes: Distance matrices are based on the following hypothetical fund portfolios: Fund 1: 100% Asset A; Fund 2: 100% Asset B; Fund 3: 50% Asset C, 50% Asset D; Fund 4: 10% Asset B, 90% Asset C. For the Manhattan distance for hierarchically-structured asset classes, we use a weight $w = 0.5$ for broad asset classes compared with detailed asset classes and we assume that assets A and B belong to the same broad asset class, and that assets C and D belong to another.

Manhattan distance, defined as $d(x,y) = \sum_i |x_i - y_i|$, accurately captures the degree of portfolio overlap. Consider the following hypothetical fund portfolios:

- Fund 1: 100% Asset A
- Fund 2: 100% Asset B
- Fund 3: 50% Asset C, 50% Asset D
- Fund 4: 10% Asset B, 90% Asset C

As illustrated in Table 1, panel a) above, Manhattan distance satisfies the first two desired properties stated above: it assigns the maximum possible distance to portfolios with no overlap and assigns smaller distances when overlap increases. In Annex B, we show that Euclidean distance, the default choice for many clustering algorithms, fails to satisfy these two properties.

We modify the Manhattan distance to incorporate asset class relationships by computing a weighted average across different levels of the asset class hierarchy. The hierarchical structure of asset classes can be formalised by defining sets of asset classes I_j that belong to the same broad asset class j. A vector of broad asset class exposures denoted $\tilde{x}$ can then be constructed as $\tilde{x}_j = \sum_{i \in I_j} x_i$. The Manhattan distance between aggregate asset class exposures is given by $d(\tilde{x}, \tilde{y}) = \sum_j \left|\tilde{x}_j - \tilde{y}_j\right|$. The overall distance is then computed as a weighted average of the Manhattan distances at the detailed and broad levels:

$$d_1\left(x, y; w, \{I_j\}\right) = w d(\tilde{x}, \tilde{y}) + (1 - w) d(x, y)$$

where w denotes the weight assigned to the broad asset class level relative to the detailed asset class level. Table 1, panel b) applies this custom distance metric to the illustrative example above with $w = 0.5$ and a hierarchical grouping of $I_1 = \{A, B\}$, $I_2 = \{C, D\}$, reflecting the interpretation of A and B as different types of bonds, and C and D as different types of listed equity. The resulting distance matrix captures the hierarchical structure of the data without discarding granular information. Using only the Manhattan distance at the broad level would obscure meaningful differences between individual asset class exposures, while relying solely on the detailed level (as shown in Table 1, panel a) would ignore economically important structural relationships. The weighted average formulation helps mitigate issues associated with the curse of dimensionality in high-dimensional datasets, balancing the trade-off between granularity and aggregation when performing portfolio analysis.

When data on asset class exposures are complemented by information on geographical exposures and other relevant variables, the overall distance metric can be constructed as the sum of component-wise distances. Geographical exposures offer an alternative categorisation of portfolio positions and can provide valuable complementary insights. Like asset class exposures, they should be normalised to avoid placing excessive weight on the clustering of larger funds. Additional variables can also be incorporated, depending on the purpose of the clustering. For example, when we apply our clustering approach to AIFMD data, we include a leverage proxy because it provides information not reflected in normalised exposures and may help generate more economically relevant clusters. The resulting composite distance metric is defined as:

$$d_{123}\left(x, y; w, \{I_j\}\right) = d_1\left(x_1, y_1; w, \{I_j\}\right) + d_2(x_2, y_2) + d_3(x_3, y_3)$$

Here, x and y are stacked vectors of the sub-vectors x_1, x_2, x_3 and y_1, y_2, y_3, respectively. x_1 and y_1 are the vectors of detailed asset class exposures (previously referred to as x and y). x_2 and y_2 denote geographical exposures, and x_3 and y_3 represent additional variables such as leverage. The functions d_1, d_2, and d_3 are the associated distance metrics for each pair of sub-vectors. For d_2, the Manhattan distance is the natural choice, based on the same rationale as for asset class exposures. In our application, we also define d_3 as Manhattan distance for consistency.

3.2 The k-medoids clustering algorithm

Centroid-based clustering algorithms aim to minimise the sum of within-cluster distances. This is typically expressed by the objective function $\sum_{i=1}^{k} \sum_{x \in C_i} d(x, c_i)$, where k is the number of clusters, C_i represents the set of points in cluster i, c_i is the cluster centre, and $d(x, c_i)$ is the distance between a data point x and its cluster centre (see Shalev-Shwartz and Ben-David, 2014). As discussed in Section 3.1, we use a custom version of the Manhattan distance, modified to reflect the hierarchical structure of asset classes. Centroid-based algorithms differ in how cluster centres are defined. For instance, the k-means algorithm selects the arithmetic mean of points as the cluster centre. The optimal clusters minimise the above objective function, ensuring that points within each cluster are as close as possible to their respective centres.

We choose k-medoids as our clustering algorithm due to its robustness and stability. While k-means is popular for its simplicity and efficiency, it is sensitive to outliers, as it defines cluster centres as the mean of all points in a cluster. In contrast, k-medoids selects actual data points (medoids) as centres, making it more robust to extreme values – an advantage in noisy, high-dimensional datasets like ours (see Park and Jun, 2009). Using real observations also improves interpretability, since k-means centres may not resemble any actual fund profile. Another benefit is that k-medoids relies solely on pairwise distances, allowing the distance matrix to be precomputed, which can improve performance with custom distance metrics.

K-medoids is an iterative clustering algorithm that alternates between assigning observations to clusters and updating medoids. The algorithm begins by choosing k data points to act as cluster centres (the medoids). From this initial choice, the algorithm proceeds in two main steps that repeat until convergence:

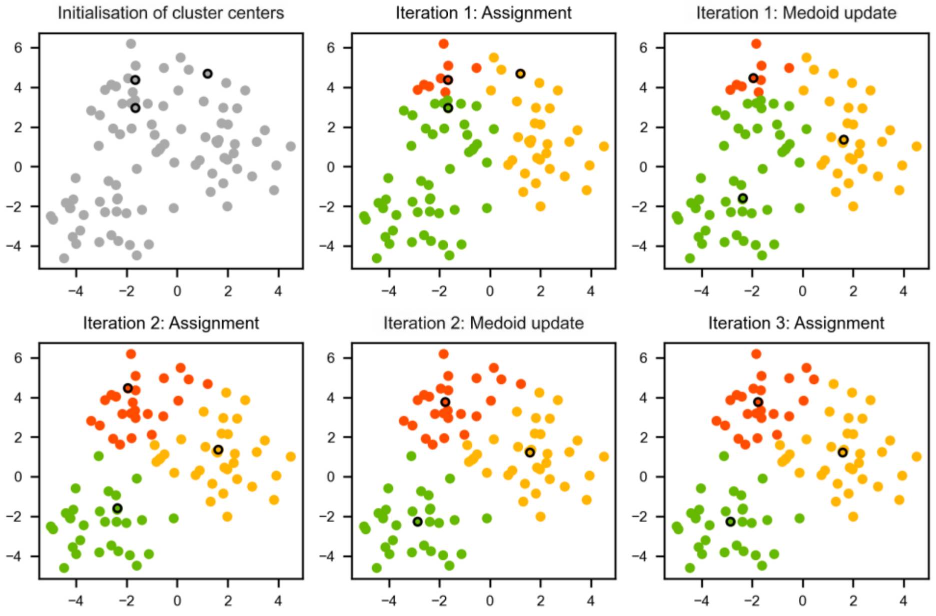
- Assignment: each data point is assigned to the nearest medoid based on the chosen distance metric.
- Medoid update: for each cluster, the algorithm then selects a new medoid from among the data points assigned to that cluster. The medoid is the point that minimises the total dissimilarity (i.e. the sum of pairwise distances) to all other points in the cluster.

The algorithm typically stops once medoids and cluster assignments no longer change. Figure 3 below illustrates how medoids and cluster assignments evolve over successive iterations.

We use the k-medoids++ method to initialise cluster centres. The choice of initial medoids is crucial in order to speed up convergence and reduce the likelihood of getting stuck in a poor local minimum. K-medoids++ is therefore an initialisation method designed to select initial cluster centres in an efficient and strategic manner. The algorithm begins by selecting the first medoid with uniform probability and then chooses subsequent ones with probability proportional to the squared distance to the nearest selected medoid (see Arthur and Vassilvitskii, 2007). This approach ensures well-separated starting points that better represent the data structure, improving the likelihood of finding a near-optimal clustering and accelerating convergence.

Figure 3

Illustration of the k-medoids algorithm



Source: Own calculations.
Notes: See Section 3.2 for a description of the two steps of the k-medoids algorithm. The k-medoids algorithm in the illustrative example uses Manhattan distance, just like the k-medoids algorithm applied to AIFMD data. The medoid update in iteration 3 does not change the chosen medoids, which means that the algorithm converges after the assignment step in iteration 3, as shown in the bottom right panel.

3.3 Applying the clustering algorithm to AIFMD data

For the main part of the analysis, we apply our clustering algorithm to end-2024 data on “other” AIFs and AIFs without a predominant type, domiciled in the EU. We focus on these two residual categories in the AIFMD classification because a data-driven approach adds the most value for funds lacking informative labels. Moreover, an analysis of asset class exposures by AIF type (see ESMA, 2020) reveals significant heterogeneity among AIFs in these two categories, while the asset class exposures of private equity funds, real estate funds and funds of funds are much more homogeneous. As explained in Section 2.2, “other” AIFs and those labelled “none” tend to follow similar strategies and exhibit comparable asset allocations, making it appropriate to cluster them jointly. Excluding funds labelled as hedge funds, private equity funds, real estate funds, and funds of funds also helps keep the number of clusters manageable and the analysis concise. In Section 4.4, we extend the analysis to historical data from 2021 to 2023, and in Section 6 to all AIFs.

Our clustering input includes detailed asset class exposures, geographical exposures and a leverage proxy. The primary input consists of normalised exposures to the 70 sub-asset types reported under AIFMD (see Section 2.3 and ESMA, 2014), excluding FX derivatives used solely for hedging. Long and short positions are treated separately, and gross positions are used for interest rate and FX derivatives held for speculative purposes, where directional exposures are unavailable. The resulting vector x_1 of normalized exposures to sub-asset types has a length 138. We also include a vector x_2 of normalised exposures to eight geographical regions (see Section 2.3), and a scalar measure of leverage defined as the ratio of NAV to total exposures. This ratio, constructed analogously to the normalised exposures, lies between 0 and 1 and offers a bounded proxy for leverage – unlike its inverse, which is unbounded. Consequently, the vector of input data for one AIF has a length of 147.

We exclude FX derivatives used for hedging from the input data to produce more economically meaningful clusters. Such derivatives are widely used by European funds to hedge currency risk when investing outside the euro area and can dominate reported exposures despite not reflecting the fund’s core investment strategy. Including them leads the clustering algorithm to group together funds with large FX derivative positions but otherwise diverse portfolios (ranging from bonds to equities), resulting in misleading clusters. Fortunately, AIFMD data distinguishes between FX derivatives used for hedging and those used for speculation, allowing us to filter out the former. While gross exposures to interest rate derivatives can also inflate their perceived importance, we chose to retain these exposures due to the lack of a hedging/speculation breakdown in AIFMD and because they have less of an influence on clustering outcomes.

Our sample of “other” and “none” AIFs includes 10,191 individual funds with a combined NAV of €3.7 trillion, after excluding low-quality observations representing less than €0.2 trillion. A total of 776 AIFs with poor data quality, corresponding to less than 2% of the NAV in the sample, are excluded based on the criteria outlined in Annex A. Meanwhile, an additional 932 AIFs (about 3% of NAV) are excluded due to missing or incomplete geographical exposure data, where reported exposures do not sum to 100%.

We use the custom Manhattan distance from Section 3.1, assigning a weight of $w = 0.75$ to broad asset classes relative to detailed ones. This effectively mitigates the curse of dimensionality and reflects our view that broad distinctions (e.g. government vs corporate bonds) are more important, although the information provided by detailed categories (e.g. bond ratings) still adds relevant nuance. The weighting ensures that finer classifications do not disproportionately influence the clustering. Section C in the Annex presents a robustness check with $w = 0.5$. While a few AIFs switch clusters, the general characteristics of the identified fund cohorts are robust to such a reweighting of input variables.

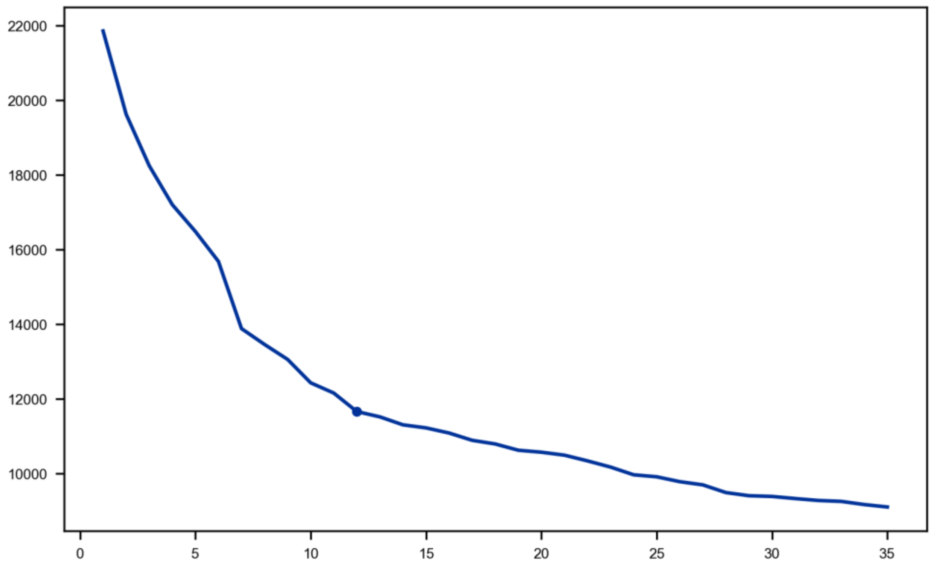
We set the number of clusters to $k = 12$ to obtain interpretable and stable clusters, guided by the elbow method. We use this method to narrow the feasible range of k, and then select the specification that maximises economic interpretability and stability. We considered multiple values of k. In our setting, k = 12 produced the most coherent and stable cohorts. As shown in Figure 4 below, the objective function declines as k increases, since more clusters allow for a closer fit to the data. However, the rate of improvement drops sharply at around k = 12, indicating diminishing returns from adding more clusters. This suggests that using 12 clusters strikes a good balance between model complexity and explanatory power. While the elbow method provides a useful guideline, inspecting the resulting clusters also helps ensure that the chosen number yields the most meaningful groupings.

We evaluate 1,000 different random initialisations of the k-medoids algorithm and select the one with the lowest value of the objective function. Since the algorithm involves randomness in the initialisation of cluster medoids and can converge to local minima, outcomes vary across runs. Exploring a broad range of initialisations ensures that we find a solution close to the global minimum. The configuration with the lowest total distance between points and their assigned medoids is used for all subsequent analyses.

Figure 4

Informing the choice of the number of clusters using the Elbow method

(x-axis: number of clusters; y-axis: value of objective function)



Sources: AIFMD data (2024 Q4), own calculations.
Notes: The objective function is the sum of distances of each point to the assigned medoid. The decrease in the objective function from increasing the number of clusters beyond $k = 12$, as highlighted by the dot, is small, justifying our choice of the number of clusters.

4 Cohorts of AIFs with similar exposures

This section describes the fund cohorts obtained from applying the k-medoids algorithm to “other” AIFs and to AIFs without a predominant type. As described in detail in Section 3, we apply the clustering algorithm to detailed asset class exposures, geographical exposures and a leverage proxy. We use a modified Manhattan distance metric that takes the hierarchical structure of asset classes into account. The final selection of 12 cohorts reflects expert judgement about economic interpretability and stability. These cohorts are presented in Table 2 and Figure 1.

We use various approaches to characterise and validate the clustering results. In Section 4.1, we describe average asset class and geographical exposures to better understand the composition of each cohort. In Section 4.2, we compare our clustering results to the self-reported AIF strategies and a list of hand-collected LDI funds. Moreover, we examine whether the identified cohorts exhibit similar returns in Section 4.3 and test the stability of clusters over time in Section 4.4. These analyses provide additional evidence of the robustness of our results.

Table 2

Size of the identified AIF cohorts

Cluster	NAV in EUR billions	AuM in EUR billions	Number of funds
EU fund holdings	271	355	758
Global fund holdings	151	192	538
Corporate bonds	502	677	1447
EU government bonds	512	776	1259
GBP LDI	162	575	416
EU equity	150	177	729
Global equity	424	496	824
EU private equity	159	189	421
EU private credit	287	392	1136
US private assets	201	287	593
Mix of bonds and equities	620	1018	1415
Mix of funds and bonds	294	403	655

Sources: AIFMD data (2024 Q4), own calculations.
Notes: The table displays the 12 AIF clusters identified based on a sample of EU-domiciled AIFs with the types “other” or “none”. The cluster labels refer to the primary asset type and geographical exposures in each cluster.

4.1 Description of the identified cohorts of funds

We characterise the identified groups of AIFs using two key visualisations: treemap charts depicting asset class exposures and a stacked bar chart showing geographical exposures. Figure 5 below shows the average exposure to various asset classes defined in AIFMD reporting, weighted by NAV. Each asset class exposure is colour-coded according to its broader macro asset class, with distinct shades used to differentiate asset classes within the broad category. The FX derivative exposures displayed in the treemaps also include FX derivatives used for hedging, even though these exposures were not included in the clustering input data.[4] To ensure readability, we do not display the most detailed sub-asset classes in the treemaps. This approach is also justified because our clustering algorithm places a stronger weight on asset classes (75%) than on sub-asset classes. Figure 6 shows the average geographical exposure of each fund cohort, weighted by NAV.

Clusters 1 and 2 contain funds that primarily invest in other funds and collectively represent over €400 billion of NAV. Even though the "fund of funds" AIF type typically invests heavily in collective investment undertakings (CIUs), these funds have been assigned to the "other" or "none" category by their managers. The first and largest cluster of the two, with a NAV of €270 billion, is more focused on the EU, while the second cluster, with a NAV of €150 billion, has a predominant geographical exposure to the category "supranational/ multiple regions", presumably indicating investment in CIUs with a global focus. In both fund clusters, the bulk of the exposure is concentrated in the "other CIUs" sub-asset class, with money market funds and exchange-traded funds playing only minor roles. Aside from CIU shares, AIFs in these two clusters also hold cash and other securities, as well as interest rate derivatives and FX derivatives.

Clusters 3 and 4 represent traditional fixed income funds, with a combined NAV exceeding €1 trillion. Cluster 3 is predominantly focused on EU corporate bonds, with some additional investments in US and UK corporate bonds, while funds in Cluster 4 predominantly hold EU sovereign bonds. The AIFs included in these two clusters show varying degrees of diversification, with some focusing solely on one bond type while others maintain mixed portfolios of corporate and government bonds. Besides bond holdings, these funds also maintain cash positions and use interest rate and FX derivatives.

Cluster 5 consists mostly of GBP-denominated LDI funds, which cater to UK pension funds and insurance companies as investors. To match the liabilities of pension funds and insurance companies, these funds typically hold long positions in UK sovereign bonds and have exposures to interest rate derivatives. Moreover, GBP-denominated LDI funds often borrow in the repo market to increase leverage, which is reported as short positions in loans under AIFMD reporting. Even though this is one of the smaller fund cohorts, with only €162 billion of NAV, forced sales by GBP-denominated LDI funds triggered financial stability issues in 2022 during the UK gilt market turmoil (see ESRB, 2023).

Clusters 6 and 7 group together equity funds with an EU and global focus, respectively. The AIFs included in these two clusters are almost entirely focused on equity investments and together account for around €570 billion of NAV. Cluster 7 with a more global focus, is the larger of the two, with a combined NAV of more than €420 billion. North America accounts for 40% of the total exposure in this cluster,

---

[4] FX derivative exposures represent the combined total of FX derivatives used for both hedging and investment purposes. In our dataset, FX derivatives are predominantly employed for hedging, with the average firm reporting a 5% exposure to FX derivatives for hedging purposes compared with only 0.7% exposure for investment purposes.

with Asia and the EU each accounting for around 15%. The AIFs in Cluster 6 have a clear geographical focus on the EU and have a NAV of €150 billion.

Clusters 8, 9 and 10 bring together private asset funds with different geographic focuses. Clusters 8 and 9 consist of private equity and private credit funds, respectively, with a strong focus on EU markets. Meanwhile, Cluster 10 includes private equity and credit funds with primarily North American – and therefore likely US – investments. Many funds within these clusters also use a significant amount of foreign exchange derivatives for hedging purposes. Collectively, these funds have a NAV of almost €650 billion. The EU private credit cluster is the largest, accounting for €290 billion, while the US private assets cluster accounts for €200 billion and the EU private equity cluster for €160 billion.

Cluster 11 encompasses funds that invest in a mix of bonds and listed equity and represents the largest cluster, with an aggregate NAV of €620 billion. Due to their diversification across asset classes, such funds are often referred to as mixed funds or balanced funds. While a mix of listed equity, corporate bonds and sovereign bonds forms the core of their strategy, most funds also have some exposure to derivatives and CIUs. While the primary geographic focus of the funds in the cluster is the EU, there are also significant exposures to the United States and smaller exposures to Asia and the United Kingdom.

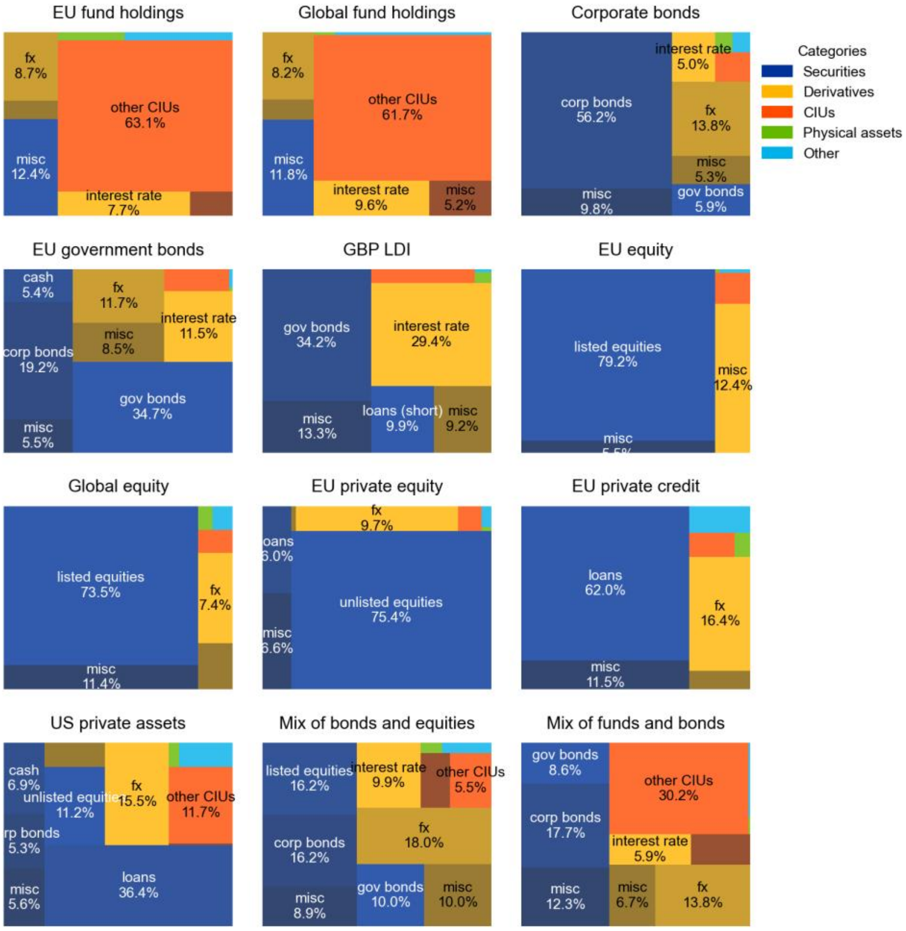
Cluster 12 comprises AIFs that invest in both CIUs and fixed income instruments. These funds have a total NAV of €290 billion. AIFs in this cluster have an average exposure of about 30% to both bonds and “other CIUs”, with smaller exposures to other securities, exchange-traded funds and derivatives. The geographic exposures of this cluster are roughly equally split between the EU and the “supranational/ multiple region” category, which likely indicates investment in geographically diverse CIUs.

While the cluster labels that we assign based on average exposures effectively describe most funds within each cluster, they may not accurately characterise all funds in each group. This limitation arises from the heterogeneity of the AIF universe, where 12 clusters are insufficient to fully represent the diverse range of investment styles and strategies employed by these funds. Consequently, funds with rare combinations of asset class and geographical exposure may be grouped with larger, more homogeneous clusters. For instance, if we look at Cluster 10, which we label as “US private assets”, some of the AIFs included here are heavily invested in “other CIUs” and are therefore more similar to AIFs in Clusters 1 and 2 than to private credit and private equity funds in terms of their asset class exposures. However, they are grouped together with US private asset funds because of their geographical exposure to North America. This reveals a limitation in using clustering algorithms to categorise funds.

Figure 5

Asset class exposures of the identified fund cohorts

(share of total exposures in percent)

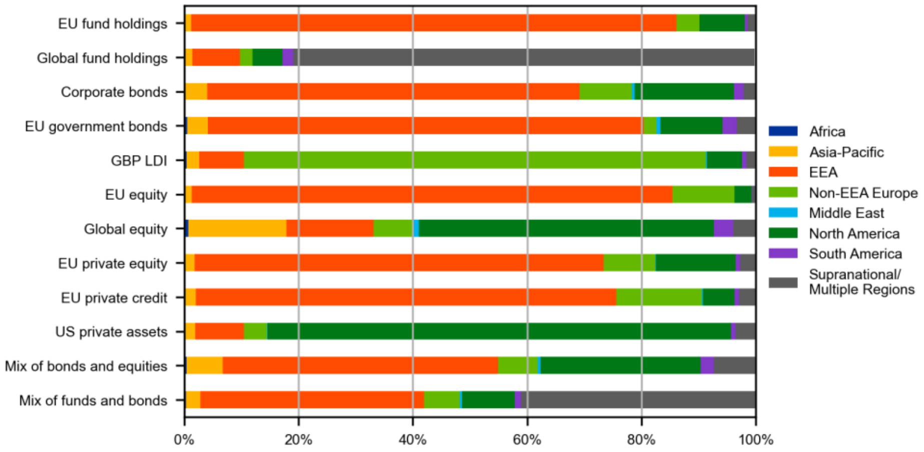


Sources: AIFMD data (2024 Q4), own calculations.
Notes: Asset classes with less than 5% average exposure within a cluster are aggregated into their respective macro asset class. If this aggregated exposure exceeds 5%, it is labelled as “misc” to indicate miscellaneous exposures within that category.

Figure 6

Geographical exposures of the identified fund cohorts

(share of total exposures in percent)



Sources: AIFMD data (2024 Q4), own calculations.
Note: Geographical exposures may not sum to 100% because certain data are withheld for confidentiality reasons.

4.2 Comparison with AIFMD classification and list of LDI funds

A comparison of our clustering results with self-reported AIFMD strategies illustrates the advantages of our data-driven approach. As described in Section 2.2, fund managers choose the primary strategy of the AIF from a restricted list of options. Figure 7, panel a) below compares this self-reported classification with the identified fund cohorts. While some of the clusters align well with the AIFMD strategies, our clustering approach offers several improvements over the self-reported classification. First, our method opens the black box of the “other-other” category: this uninformative label is used to refer to funds with specialised investment strategies, such as LDI funds or private credit funds that are not listed in the AIFMD classification, though also to funds that invest in a mix of asset classes. The clustering algorithm disentangles these specialised fund types and mixed funds in a data-driven way. Second, for more defined strategies such as “fixed income” and “equity”, our approach provides greater granularity by identifying subgroups based on geographic and asset class exposures. Third, our method identifies a significant number of “other” and “none” funds that resemble funds of funds and private equity funds, even though funds of funds and private equity funds are distinct AIF types under AIFMD reporting.

While AIFMD strategies leave room for different interpretations, clustering creates fund cohorts based on objective mathematical rules. Fund strategies reported under AIFMD do not have precise definitions, as shown by the fact that funds with different strategy labels end up being grouped within the same cluster. For example, while most fund managers denote LDI funds as “other” funds, some label them as fixed income funds. The same is true for private credit funds.

Clustering overcomes these issues since the clusters are formed according to objective mathematical rules using the funds’ actual exposures and are therefore not influenced by differing interpretations.

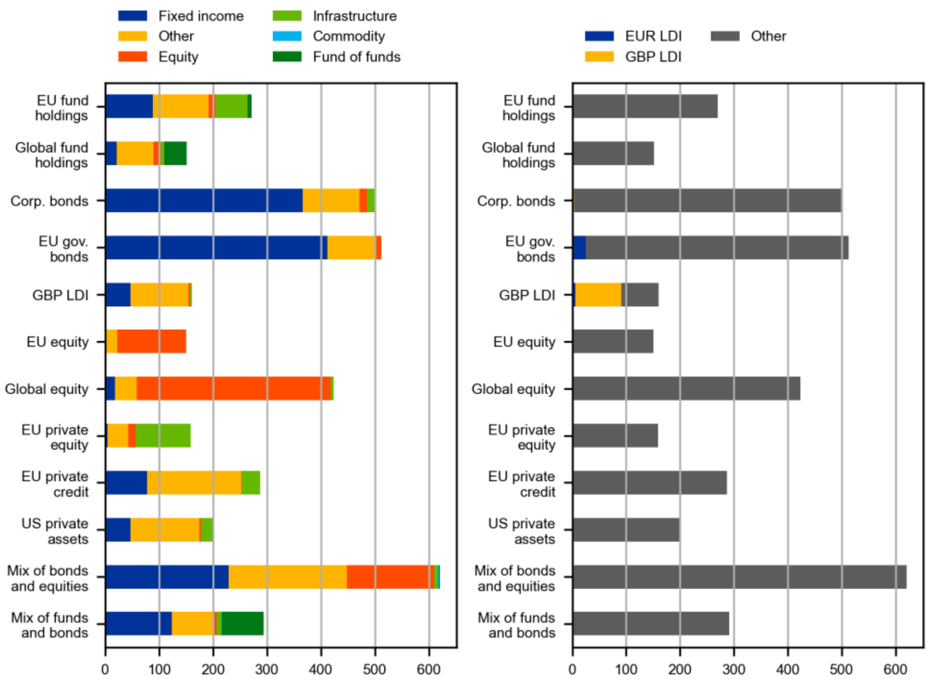
Comparing our clusters with a hand-collected list of LDI funds validates the GBP-denominated LDI cluster and underlines the benefits of clustering compared with manually collecting funds. For the special feature on LDI funds in ESRB (2023), the authors hand-collected a list of LDI funds based on fund names and corroborated the list with national supervisors. Figure 7, panel b) below shows a great degree of overlap between this list and the cluster of GBP-denominated LDI funds. All hand-collected GBP-denominated LDI funds except one are included in the GBP LDI cluster, and the AIFs on the list make up more than half of the NAV of the cluster. The remaining funds in that cluster may be LDI funds missed in the data collection for ESRB (2023), newly created LDI funds, or traditional sovereign bond funds with a focus on the United Kingdom. Because there are only a few EUR-denominated LDI funds, these are grouped into the EU government bond funds cluster.

Figure 7

Comparison of identified fund cohorts with other data sources

(NAV in EUR billions)

a) AIF strategies reported under AIFMD **b) LDI funds collected for ESRB (2023)**



Sources: AIFMD data (2024 Q4), ESRB (2023), own calculations.
Notes: Panel a): the first five strategies listed in the legend are possible “other” AIF strategies. AIFs without a predominant type are not restricted to these five “other” strategies and sometimes indicate a fund of funds strategy. Very uncommon cluster-AIF strategy combinations are not displayed for reasons of confidentiality. Panel b): the colours highlight EUR-denominated and GBP-denominated LDI funds that were hand-collected for the special feature in ESRB (2023). A total of 13 LDI funds from the ESRB (2023) list are scattered across atypical clusters (i.e. clusters without a focus on fixed income products) and therefore cannot be displayed for reasons of confidentiality.

We apply our clustering approach to the whole universe of AIFs and find a strong correspondence between our clusters and predefined AIF types. As a robustness check, we apply the same approach to all AIFs beyond “other” AIFs, as detailed in Annex D. Hence, our extended sample of more than 21,000 funds includes funds of funds, private equity funds, real estate funds and hedge funds. Using the elbow method, we choose 20 clusters to account for the greater diversity of funds. Clusters 1 to 4 largely comprise funds of funds with different geographical exposures. Clusters 5 and 6 relate to private equity funds with an EU and/or a US/global focus. Cluster 7 consists of real estate funds. Clusters 8 to 17 correspond mostly to “other” AIFs, as identified in Section 4.1. The three remaining clusters represent small groups of funds which are not well covered by regulatory types or the clusters identified earlier for “other” AIFs. Interestingly, hedge funds are scattered across clusters. This is because there is no regulatory definition of hedge funds in AIFMD and the strategies they pursue are carried out across asset classes using mostly synthetic leverage. The largest concentrations appear in clusters that predominantly contain traditional equity, bond and mixed funds. This pattern is not surprising: equity hedge funds and fixed income hedge funds have more similar exposures to traditional funds targeting the same asset classes than to each other. Overall, our identified clusters align well with AIF types, providing further support for our approach.

4.3 Validation with returns data

We examine fund returns to further validate our identified clusters. If the clustering algorithm produces economically meaningful groupings, we expect returns to co-move within clusters of funds with similar exposures. To check this, we compile time series of monthly gross returns from 2016 to 2024 for each AIF from AIFMD reporting. We use two metrics to quantify how much variation in returns is explained by the identified clusters: one scalar metric that can be interpreted as an extension of R^2 to panel data, and a matrix of pairwise similarity scores between fund cohorts.

The quality of the returns data varies significantly, requiring data cleaning procedures. To address this, we exclude funds whose time series contain zero or missing entries more than once per year on average. This step resulted in the removal of approximately 30% of the funds from the returns dataset, representing 27% of the total NAV. We also winsorise returns at the 2.5th and 97.5th percentiles separately for each month to reduce the impact of outliers.

First, we consider a metric that quantifies the fraction of variance explained by fund groupings using the residuals from two linear regressions. As a first step, we run a regression of winsorised returns on group × time fixed effects:

$$returns_{i,t} = \sum_{j=1}^{12} 1_{group_{i,t}=j}\, \beta_{j,t} + \epsilon_{i,t}$$

where i denotes the fund, t the month, and the coefficient $\beta_{j,t}$ represents the group × time fixed effect. From this regression, we compute the sum of squared residuals $SS_{time\times gr}$. Next, we run a regression on time fixed effects only:

$$returns_{i,t} = \delta_t + \omega_{i,t}$$

where δ_t is the time fixed effect. We refer to the sum of squared residuals from that regression as SS_{time}. Finally, we define the fraction of the variation in returns explained by group membership as

$$1 - \frac{SS_{time \times gr}}{SS_{time}}$$

This metric can be interpreted as a version of R^2 for panel data. The better cluster membership explains returns relative to a regression with time fixed effects only, the smaller the ratio of the sum of squared residuals, and the closer the metric is to 1.

The fund cohorts identified through clustering perform substantially better at explaining return variance than fund strategies reported by managers under the AIFMD (see Table 3). While the AIFMD strategies explain only 12% of the variance in returns, the fund clusters explain close to 28% of this variance. This result shows that our clustering methodology results in economically meaningful groups that are subject to similar risk factors. The regulatory categorisation of AIFs, on the other hand, is much less useful because funds with similar exposures and returns are sometimes placed in different categories, while dissimilar funds are placed in the same category.

Table 3

Fraction of return variance explained

Grouping	Fraction explained
Fund cohort identified through clustering	27.9%
Primary strategy reported under the AIFMD	11.6%

Sources: AIFMD data (2016-2024 Q4), own calculations.
Note: See the main text for a description of the underlying methodology.

Second, we compute average pairwise distances between return time series both within and across clusters. As a first step, we compute pairwise cosine similarities between time series of returns of all AIFs. Given two times series vectors, A and B, their cosine similarity is defined as:

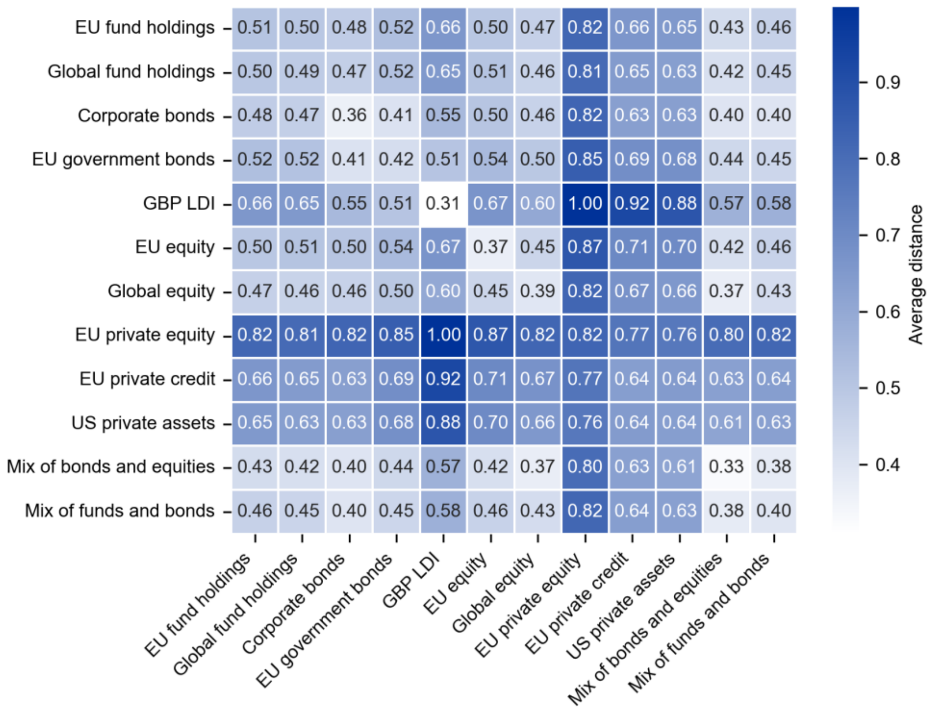
$$cosine\ similarity\ (A, B) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}$$

This measure reflects the angle between the vectors defined by the time series in a high-dimensional space. The cosine similarity is bounded between -1 and 1, where a result of -1 indicates that the two vectors are pointing in exactly opposite directions and 1 indicates that they point in precisely the same direction. Next, we average the pairwise cosine similarities between funds both within the same cluster and across different clusters. These averages are then subtracted from 1 to obtain a distance metric, where values closer to 0 indicate that funds' return profiles in the two groups being compared are more similar.

This granular analysis reveals insights on the co-movement of returns within and across fund cohorts not captured by the scalar metric (see Figure 8). The diagonal matrix elements represent the average distance between funds within the same cluster, while the off-diagonal elements show the average distances between funds across different clusters. This detailed breakdown helps us understand which clusters exhibit the highest internal similarity in returns and contribute most to the percentage of variance explained. For example, LDI funds show the highest internal similarity combined with a high level of dissimilarity from other types of funds, indicating that they have uniquely correlated and specific returns trends. The equity, fixed income and mixed clusters also show a high degree of internal similarity. The three clusters containing private asset funds show the least internal and external similarity in returns. This is plausible given the idiosyncratic nature of these funds' investments and the fact that valuations are undertaken less frequently for private asset funds, given the nature of their underlying investments.[5]

Figure 8

Average pairwise distances between return time series



	EU fund holdings	Global fund holdings	Corporate bonds	EU government bonds	GBP LDI	EU equity	Global equity	EU private equity	EU private credit	US private assets	Mix of bonds and equities	Mix of funds and bonds
EU fund holdings	0.51	0.50	0.48	0.52	0.66	0.50	0.47	0.82	0.66	0.65	0.43	0.46
Global fund holdings	0.50	0.49	0.47	0.52	0.65	0.51	0.46	0.81	0.65	0.63	0.42	0.45
Corporate bonds	0.48	0.47	0.36	0.41	0.55	0.50	0.46	0.82	0.63	0.63	0.40	0.40
EU government bonds	0.52	0.52	0.41	0.42	0.51	0.54	0.50	0.85	0.69	0.68	0.44	0.45
GBP LDI	0.66	0.65	0.55	0.51	0.31	0.67	0.60	1.00	0.92	0.88	0.57	0.58
EU equity	0.50	0.51	0.50	0.54	0.67	0.37	0.45	0.87	0.71	0.70	0.42	0.46
Global equity	0.47	0.46	0.46	0.50	0.60	0.45	0.39	0.82	0.67	0.66	0.37	0.43
EU private equity	0.82	0.81	0.82	0.85	1.00	0.87	0.82	0.82	0.77	0.76	0.80	0.82
EU private credit	0.66	0.65	0.63	0.69	0.92	0.71	0.67	0.77	0.64	0.64	0.63	0.64
US private assets	0.65	0.63	0.63	0.68	0.88	0.70	0.66	0.76	0.64	0.64	0.61	0.63
Mix of bonds and equities	0.43	0.42	0.40	0.44	0.57	0.42	0.37	0.80	0.63	0.61	0.33	0.38
Mix of funds and bonds	0.46	0.45	0.40	0.45	0.58	0.46	0.43	0.82	0.64	0.63	0.38	0.40

Sources: AIFMD data (2016-2024 Q4), own calculations.
Notes: See the main text for a description of the underlying methodology. Higher values indicate a lower level of co-movement of returns.

4.4 Cohorts of funds over time

To better understand whether the identified clusters remain stable over time and how their size has changed, we now extend our analysis beyond the

---

[5] In some cases, the distance is higher within the cluster than across clusters. This can be explained by common exposures across clusters (e.g. similar bond or equity exposures) or by the role played by global factors in driving returns. In some asset classes, including private credit and bond markets, global and US factors might play a significant role, given the size of those markets compared with European ones.

cross-section data from end-of-2024. Specifically, we examine changes in the NAV of the identified AIF clusters from 2021 to 2024 and evaluate the consistency of cluster assignments throughout this period.

Our analysis of fund panel data maintains consistency with the cross-sectional results presented earlier in this section. Multiple approaches exist for applying clustering algorithms to panel data: algorithms designed for cross-sectional data can be applied to individual time periods, to pooled data across all periods, or to averaged input data from different periods. Alternatively, dynamic clustering algorithms specifically designed for panel data allow for time-varying cluster centres and transitions between clusters. However, these approaches would generate clusters that differ at least slightly from those obtained previously, significantly complicating the presentation of the results. Therefore, we adopt a simpler methodology: we keep the medoids identified from end-2024 data fixed and classify fund exposures from previous years based on their proximity to these established cluster centres.

Figure 9 shows the evolution of NAV by cluster from 2021[6] to 2024. The most notable finding is the rapid growth of the three private asset clusters, which doubled in size between 2021 and 2024. This trend aligns with what other financial market observers have noted about the rapid expansion of the private credit and private equity sectors in recent years (see Cera et al., 2024). Most other clusters show moderate or no growth over this time period; the NAV of the “Global equity” and “GBP LDI” clusters even decreases.[7]

---

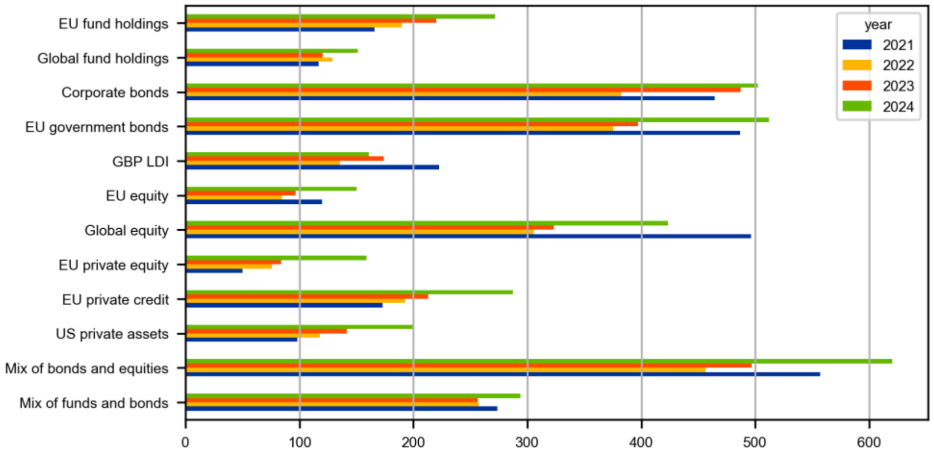
[6] We do not use data from 2020 or earlier because exposures to UK assets were categorised as geographical exposures to the EU before Brexit. Therefore, GBP-denominated LDI funds would not be correctly identified if we used medoids estimated with 2024 data.

[7] When interpreting the growth of the private equity clusters and fund holdings clusters, it is important to keep in mind that we only consider AIFs categorised as “other” or as “none” here. AIFs that are categorised as funds of funds or private equity funds in the regulatory reporting are not included in this analysis.

Figure 9

Net asset value of the identified AIF cohorts over time

(NAV in EUR billions)



Sources: AIFMD data (2021 – 2024), own calculations.
Notes: Cluster membership in the years 2021 to 2023 is determined based on the location of the 2024 medoids.

We assess the stability of the clusters obtained by tracking the cluster membership of individual funds over time. If we keep the medoids fixed and assign funds to clusters using data from different years, cluster membership can change for one of the following three reasons:

- Fund exposures can change substantially due to tactical asset allocation decisions made by fund managers. Such changes are typically short-lived, as the main asset classes and overall investment strategy are defined in the funds' constitutive documents (e.g. prospectuses, memoranda and other fund documents) and tend to remain broadly stable over time.

- Funds positioned near the midpoint between cluster centres may experience membership changes following minor exposure adjustments. This phenomenon particularly affects clusters forming natural continuums, such as equity funds, mixed equity-bond funds and government/corporate bond funds.

- Fund exposures are reported inconsistently across years. For example, infrastructure investments may be classified as unlisted equity or within the residual "other-other" asset class. Similarly, managers of real estate fund of funds might report exposures to collective investment vehicles in some years, while reporting real estate exposures in others.

Our analysis reveals that the clusters remain relatively stable over time, with 85% of AIFs assigned to the same cluster in 2024 as in 2023. Predictably, the likelihood of funds switching clusters increases over time: while 86.1% of funds retain their 2024 cluster assignment when compared against 2023 data, this proportion drops to 77.0% when drawing comparisons with 2021 data. This gradual decline in cluster stability suggests that changes in AIF exposures accumulate over time, underscoring the importance of periodically re-analysing the AIF landscape to account for long-term shifts. When funds switch clusters, these transitions often involve the “Mix of bonds and equities” cluster. This is to be expected, as this cluster borders on the equity fund clusters and the bond fund clusters. As explained earlier, for funds positioned at cluster boundaries, such as a mixed fund that leans more towards equity than the average mixed fund, small changes in exposures can easily result in a change in cluster membership.

If greater stability of the identified clusters over time is desired, the algorithm can be modified to reduce the frequency of transitions. One possibility would be to apply the algorithm to exposure data averaged over specified periods in order to smooth short-term fluctuations. Alternatively, introducing a penalty term for cluster transitions could help to increase stability, while still maintaining responsiveness to genuine strategic shifts.

Our classification over time relies on a constant set of cluster centres. We keep the end-2024 medoids fixed when assigning funds in earlier years in order to maximise comparability of clusters over time in this illustrative exercise. Beyond the fund-level transitions documented above, we have also re-estimated the clustering on earlier cross-sections (e.g. using only 2021 or 2022 data). The resulting medoids are very similar to those obtained from the 2024 sample, both in terms of asset class exposure profiles and economic interpretation. This indicates that the main “types” of funds we identify are broadly unchanged over the sample period; changes over time primarily reflect the entry and exit of funds and shifts in the relative size of the clusters, rather than the appearance or disappearance of entirely new fund business models.

The clustering approach could be extended to prospective applications and real-time supervision. In prospective or real-time applications, there is a trade-off between keeping clusters unchanged for consistent monitoring and updating them to reflect structural changes in the fund universe (such as the gradual emergence of new investment strategies). While fully dynamic clustering methods that allow centres to move and clusters to appear or disappear over time would address this, they are typically designed for lower-dimensional problems and are computationally demanding at the scale of AIFMD asset-class exposures. A practical supervisory strategy with static clustering tools would be to: (i) estimate clusters on the cross-section at a reference date (t); (ii) in subsequent periods, compare the results obtained by keeping these medoids fixed with the results obtained from re-estimating clusters on the new data; and (iii) use diagnostics (such as changes in the objective function or within-cluster dispersion), together with supervisory judgement and knowledge of market developments, to decide whether to update the medoids. Given that fund strategies and asset class exposures typically evolve gradually, such updates would likely be required only intermittently (for example, every three to five years).

5 Assessing financial stability risks from the identified cohorts of funds

In this section, we analyse financial stability risks associated with “other” and “none” AIFs, using the identified funds cohorts. Our approach follows the framework set out in ESMA’s Statistical Reports on Alternative Investment Funds, focusing on leverage, liquidity mismatches and interconnectedness with the broader financial system. While ESMA’s analyses rely on the AIFMD classification, our clustering method allows for a more granular assessment of risks by grouping funds based on actual exposures, yielding more precise insights. For instance, the classification embedded in AIFMD reporting would obscure the concentrated leverage risks present in LDI funds and inappropriately group illiquid private credit funds with more liquid investment vehicles. In contrast, our cluster-based analysis can help identify specific risk concentrations within groups of AIFs with similar profiles. In the final subsection, we examine the geographic distribution of fund managers and find that similar funds are often spread across multiple jurisdictions. This underscores the need for effective information sharing via ESMA and the ESRB to ensure the accurate assessment of risks stemming from collective behaviour.

5.1 Leverage

The only cluster of AIFs in our sample with a substantial average level of leverage consists of GBP-denominated LDI funds. Figure 10 below presents both net leverage, computed according to the commitment method, and a proxy of gross leverage, measured as the ratio of AuM to NAV.[8] A gross or net leverage ratio of 100% indicates that the AIF does not use derivatives or borrowing to amplify its exposures beyond its NAV. The key distinction between the commitment method and the AuM/NAV measure lies in the treatment of netting and hedging: the commitment approach accounts for these adjustments, which typically reduce net exposures and therefore lower the measured leverage. GBP-denominated LDI funds exhibit by far the highest average levels of leverage, with a gross leverage of 355% and a commitment leverage of 257%. All remaining clusters of “other” AIFs show notably lower levels of leverage, with average gross leverage ranging from 117% to 164% and average net leverage of 105% to 126%, indicating limited use of effective leverage.[9] The difference between gross and net leverage reflects the widespread use of FX hedging and offsetting interest rate swap positions, which inflate gross exposures but are neutralised in net terms.

---

[8] Leverage according to the commitment and gross methods is frequently misreported under AIFMD (see ESMA, 2019, p.40-44). AuM/NAV is an alternative leverage measure that is less error-prone and closely approximates leverage according to the gross method.

[9] Note that our sample of AIFs focuses on “other” AIFs and therefore does not include hedge funds, which are a separate category under AIFMD. Hedge funds have the highest levels of leverage among AIFs in the data.

AIFs with gross leverage above 300% are common among LDI funds but rare in other clusters of AIFs. Figure 11 further below breaks down fund clusters into three categories: funds with gross leverage below 125%; those with moderate leverage of between 125% and 300%; and those with high leverage above 300%.[10] Notably, 41% of GBP-denominated LDI funds fall into the highly leveraged category, which also includes some EUR-denominated LDI funds, which are grouped with traditional EU government bond exposures due to their relatively small numbers. The remaining highly leveraged funds with gross leverage above 300% are dispersed across clusters, with the most significant concentration found in the "Mix of bonds and equities" category.[11]

From a financial stability perspective, the degree and concentration of leverage in the LDI cluster raise concerns. First, high gross and net leverage levels imply that these funds are not merely hedging but are structurally leveraged, making them more sensitive to market movements and margin shocks. Second, as shown during the UK gilt market turmoil in 2022, LDI strategies are particularly vulnerable to interest rate volatility and liquidity stress, which can trigger large-scale collateral calls and forced asset sales. Such procyclical behaviour risks amplifying price dynamics in core government bond markets and can transmit stress to other financial institutions, including pension funds and banks exposed to sovereign debt or derivatives counterparties. In addition, our LDI cluster identifies a very high degree of asset class concentration: GBP-denominated LDI funds are all exposed to the same instruments (UK sovereign bonds and interest rate derivatives), meaning that sales by some of those funds can have a high impact on funds within the same cohort, amplifying shocks.

Following the gilt market stress event, the two NCAs where EU GBP-denominated LDI funds were domiciled at that time took action to ensure the resilience of GBP-denominated LDI funds. In November 2022, the Commission de Surveillance du Secteur Financier (CSSF) and the Central Bank of Ireland (CBI) provided supervisory guidance to GBP-denominated LDI managers, by asking them to maintain an appropriate level of resilience (measured by a fund's ability to withstand an interest rate shock of at least 300 to 400 basis points) and to ensure that their funds have a reduced risk profile (see ESMA, 2022b). In April 2024, both NCAs imposed macroprudential measures under Article 25(3) of AIFMD. Article 25 of AIFMD gives NCAs the power to impose leverage limits or any other restrictions on AIFMs with respect to the AIFs they manage. This power was used for the first time in 2022 by the Central Bank of Ireland for commercial real estate funds (see ESMA, 2022a). In the case of GBP-denominated LDI funds domiciled in Ireland and Luxembourg, both NCAs chose to impose a yield buffer requirement rather than

---

[10] Article 111(1) of Commission Delegated Regulation (EU) No 231/2013 under AIFMD defines AIFs that employ leverage on a substantial basis as AIFs with commitment leverage above three times its net asset value. Due to data quality issues with the commitment leverage variable, we proxy highly-leveraged AIFs as those with an AuM/NAV ratio above three, following Bouveret et al. (2025A, 2025B).

[11] See Bouveret et al. (2025A, 2025B) for a more detailed analysis of the risks from leverage in both hedge funds and "other" AIFs with gross leverage above three times their net asset value, including a decomposition of leverage into synthetic and financial leverage, an analysis of the effect of interest rate shocks on these funds, and a discussion of policy measures.

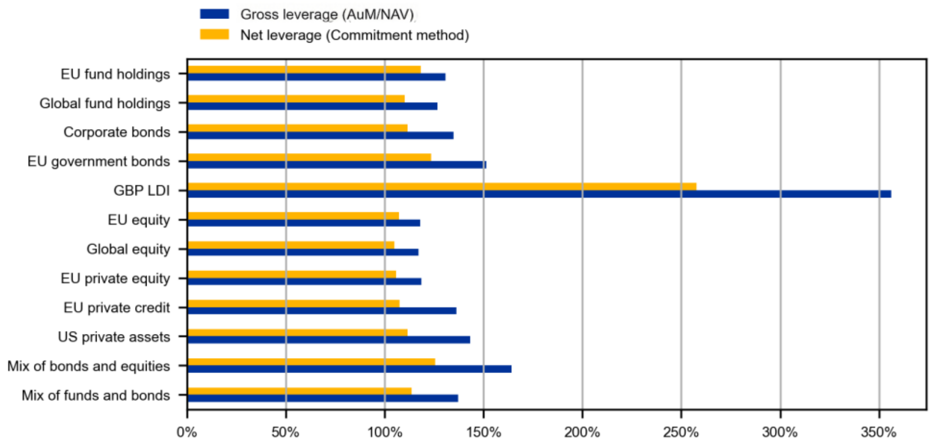
direct limits on leverage. The measure required such funds to be resilient to a shock of at least 300 basis points, effective from end-July 2024 (see ESMA, 2024).

Overall, the heterogeneity in leverage across AIF clusters reinforces the need for granular, exposure-based risk assessments rather than reliance on broad fund labels. It also supports the case for targeted macroprudential tools, data sharing, and supervisory coordination, especially in areas where leverage, liquidity mismatch and interconnectedness co-exist.

Figure 10

Average leverage by fund cohort

(Leverage in percent)

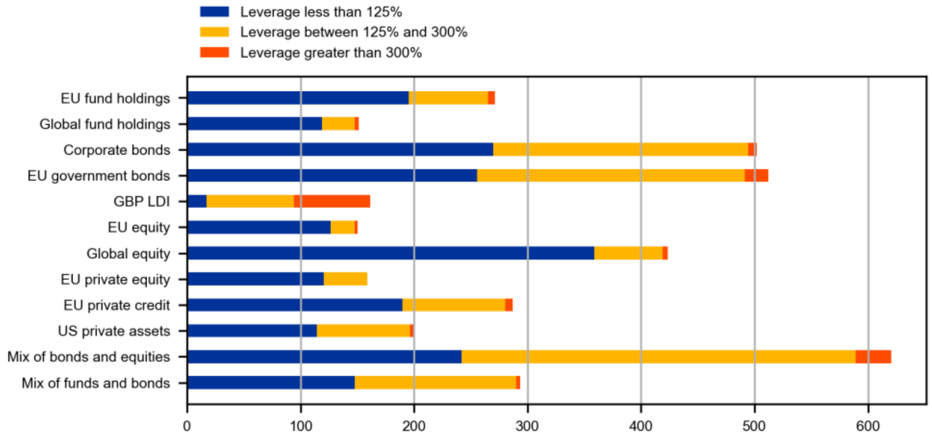


Sources: AIFMD data (2024 Q4), own calculations.
Note: In contrast to gross leverage, net leverage takes netting and hedging arrangements into account and therefore tends to be lower than gross leverage.

Figure 11

Prevalence of AIFs with high leverage

(NAV in EUR billions)



Sources: AIFMD data (2024 Q4), own calculations.
Note: To define the leverage buckets, we consider gross leverage (AuM/NAV).

5.2 Liquidity mismatch

Most private credit and private equity funds are closed-ended, while funds with more liquid portfolios typically offer daily redemptions. Figure 12 below shows the investor redemption frequencies of funds by cluster, weighted by fund NAV. The vast majority of "other" AIFs that primarily invest in bonds and listed equities offer daily redemptions, reflecting the liquid nature of their underlying assets. The same holds for fund clusters that are characterised by significant holdings of shares in other funds, although assessing the effective liquidity of these investments requires information on the redemption terms of the underlying funds.[12] In contrast, private asset funds are typically closed-ended or offer redemptions only at lower frequencies, such as quarterly intervals. These structural features of closed-ended funds act as a safeguard against liquidity mismatches, reducing the likelihood of forced asset sales during periods of investor stress.

To assess liquidity mismatches, we compare self-reported portfolio and investor liquidity at different time horizons. Under AIFMD reporting requirements, fund managers must disclose the proportion of their portfolio that can be liquidated within specific time frames, as well as the percentage of investor capital that can be redeemed within those same time periods. Figure 13 further below shows the average liquidity profiles (weighted by NAV) for the "EU private credit" cluster (panel a) and the "Mix of funds and bonds" cluster (panel b). A comparison of the liquidity profiles reveals several key insights. First, mixed-asset funds exhibit high investor liquidity in the short term, with a large share of capital redeemable within the shortest time horizon. In contrast, private credit funds allow redemptions primarily over longer periods, reflecting more limited short-term investor liquidity. This suggests that mixed funds may face greater liquidity strain if the underlying assets cannot be liquidated rapidly, increasing the risk of forced sales or liquidity pressures. Private credit funds, by allowing staggered redemptions, mitigate this risk by providing more time to realise assets. Second, investor liquidity and portfolio liquidity tend to move in parallel across time horizons in both clusters. This alignment suggests that redemption terms broadly match asset liquidity, reducing the risk of liquidity shortfalls or fire sales. Lastly, Figure 13, panel a) shows that portfolio liquidity exceeds investor liquidity, meaning that assets can be liquidated faster than investors can redeem them, providing evidence of conservative liquidity management. By contrast, Figure 13, panel b) reveals the opposite: investor liquidity exceeds portfolio liquidity, and this gap is particularly pronounced over the one-day time horizon. Due to this mismatch, funds may struggle to meet redemptions without resorting to delayed payments, suspensions or distressed asset sales.

We identify significant liquidity mismatches in eight out of the 12 clusters. Average cluster-level mismatches can mask mismatches in individual funds, as excess portfolio liquidity in some funds may offset the lack of portfolio liquidity in others. Figure 14 shows funds with significant liquidity mismatches, defined as portfolio liquidity being at least 20 percentage points below redeemable shares. We evaluate this over one-day and one-year horizons. In eight out of the 12 clusters, at

---

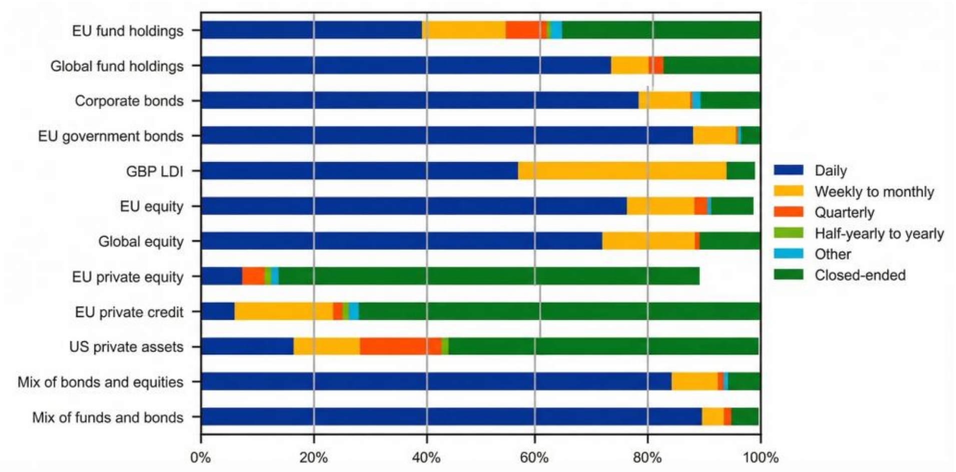
[12] The redemption terms of indirect fund holdings are not available in AIFMD data.

least 15% of the AIFs exhibit a significant liquidity mismatch of 20 percentage points or more over the one-day time horizon. The only clusters in which portfolio liquidity consistently meets or exceeds investor liquidity are private asset funds and LDI funds, whose closed-ended or low-frequency redemption structures reduce mismatch risks. Although liquidity mismatches tend to decrease over longer time horizons, it is notable that even at the one-year horizon more than 15% of funds in the "Global fund holdings" and "Mix of funds and bonds" clusters still show substantial mismatches, indicating persistent structural vulnerabilities in some strategies.

Our findings underline the need for robust liquidity management and supervisory scrutiny of redemption terms, especially where leverage or interconnected exposures may amplify the impact of liquidity shocks. It is important to note that the reported liquidity profiles generally reflect assumptions under normal market conditions. In times of market stress, actual liquidity may deteriorate significantly, not only for fixed income instruments but also for fund of funds exposures. If market conditions deteriorate and investors seek to redeem shares rapidly, funds with insufficient liquid assets may be forced to sell holdings under pressure, potentially at discounted prices. This can amplify asset price volatility and transmit stress to interconnected parts of the financial system, including underlying funds or fixed income markets. The clustering approach could help inform policy discussions on whether further policy steps are needed to address risks related to liquidity mismatch for specific cohorts of funds, also drawing on existing recommendations and guidelines on policy tools to address liquidity mismatch.[13]

Figure 12
Investor redemption frequency

(share of NAV in percent)



Sources: AIFMD data (2024 Q4), own calculations.
Note: Cluster NAVs may not sum to 100% because certain data are withheld for confidentiality reasons.

---

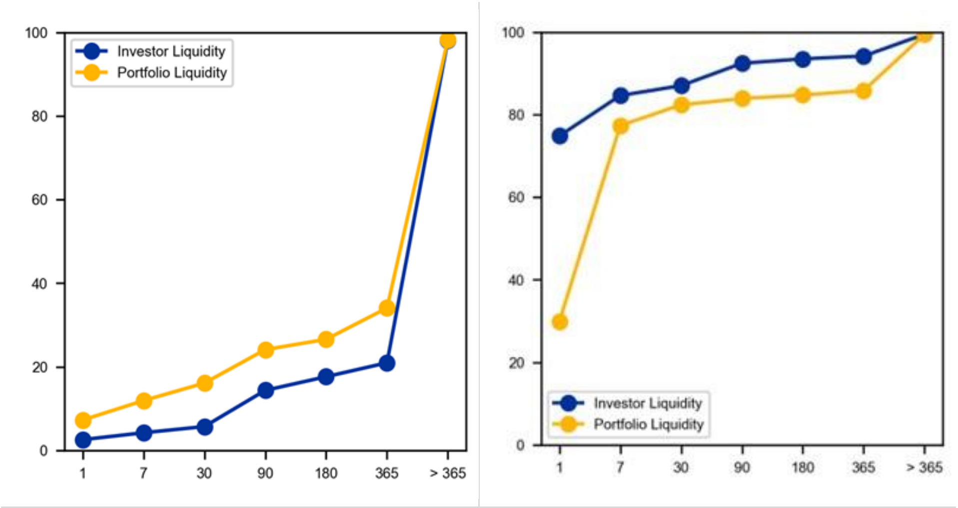
[13] See, for example, FSB (2024) and ESMA (2024).

Figure 13

Comparison of portfolio liquidity profile with investor liquidity profile

a) Funds in "EU private credit" cluster **b) Funds in "Mix of funds and bonds" cluster**

(x-axis: days; y-axis: percent)

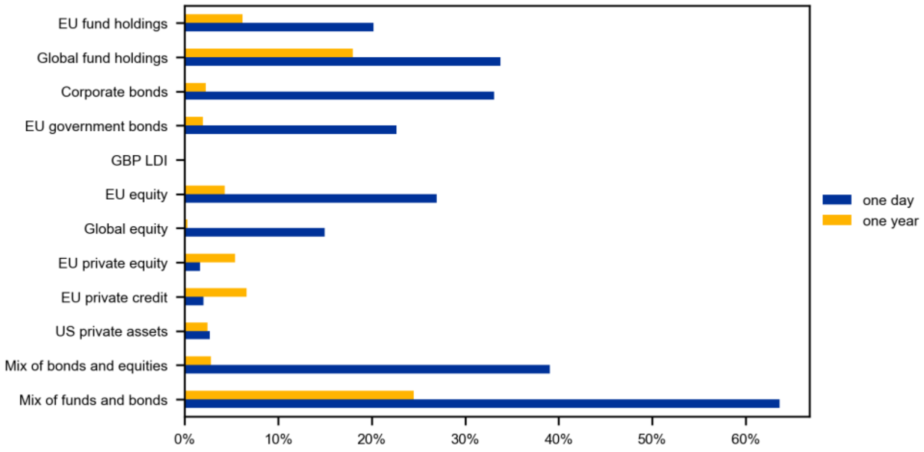


Sources: AIFMD data (2024 Q4), own calculations.
Notes: Portfolio liquidity denotes the fraction of assets that can be liquidated within 1 day, 7 days, 30 days etc. Investor liquidity denotes the fraction of funds shares that investors can redeem in the same time periods. We compute averages over all funds in a cluster, weighted by NAV.

Figure 14

Prevalence of AIFs with substantial liquidity mismatch

(share of NAV in percent)



Sources: AIFMD data (2024 Q4), own calculations.
Notes: The figure shows the fraction of AIFs (in terms of NAV) for which the difference between investor and portfolio liquidity is larger than 20 percentage points at the given time horizon. The share of GBP-denominated LDI funds with a substantial liquidity mismatch is set to zero due to data confidentiality reasons.

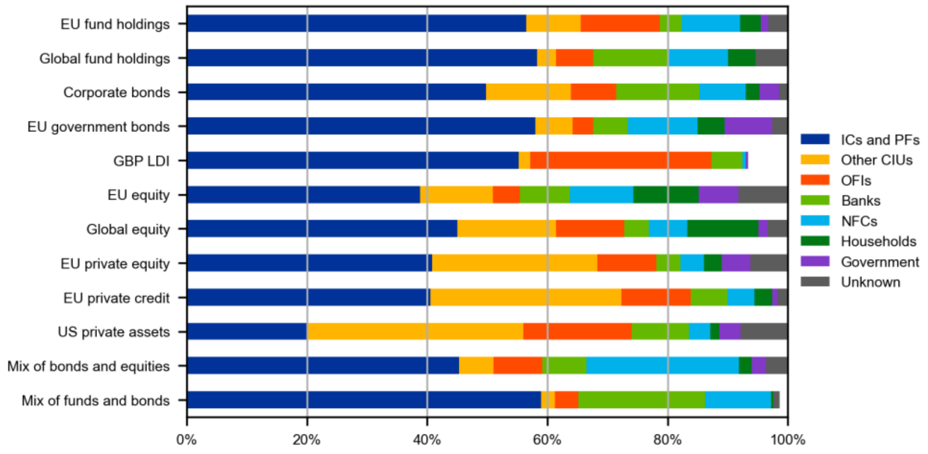
5.3 Interconnectedness through the investor base

Financial institutions dominate the investor base across all AIF clusters, typically accounting for over 80% of NAV (see Figure 15). Insurance companies and pension funds (ICs/PFs) are consistently the largest investor category, with the notable exception of the “US private assets” cluster, where investments are more evenly distributed across different types of financial institutions. Another important investor type are other collective investment undertakings (CIUs), which account for approximately 30% of the investor base in the three private asset clusters. The only clusters featuring a significant participation of non-financial companies and households are equity funds and mixed bond-equity funds, though even here their combined share remains below 40%. The prevalence of financial institutions in the investor base of “other” AIFs reflects not only institutional search-for-yield strategies but also significant cross-holdings within the investment fund sector.

The interconnectedness with other financial institutions through the investor base illustrates the potential role of “other” AIFs in the transmission of stress within the financial sector. High IC/PF concentration exposes these entities to mark-to-market losses during AIF distress, potentially triggering regulatory capital pressures and solvency breaches that could prompt procyclical deleveraging and withdrawal from illiquid exposures. The prominent role of CIUs in private asset clusters creates additional horizontal contagion risk through redemption pressure[14] and valuation shocks across fund structures – dynamics that are particularly difficult to detect given the opacity of multi-layered fund relationships. While household and corporate exposures are limited overall, their presence in selected clusters means that reputational effects could still arise during significant AIF losses or suspensions. This concentrated and interlinked investor base positions “other” AIFs as potential transmission nodes, amplifying stress through both asset channels (valuation effects) and liability channels (redemptions and capital calls).

Figure 15
Investor base

(share of NAV in percent)



Sources: AIFMD data (2024 Q4), own calculations.
Notes: Cluster NAVs may not sum to 100% because certain data are withheld for confidentiality reasons. ICs: insurance companies; PFs: pension funds; CIUs: collective investment undertakings; OFIs: other financial institutions; NFCs: non-financial companies.

---

[14] Within investor types, funds tend to redeem more than other investors (see Allaire et al., 2023).

5.4 Cross-border spillover risks

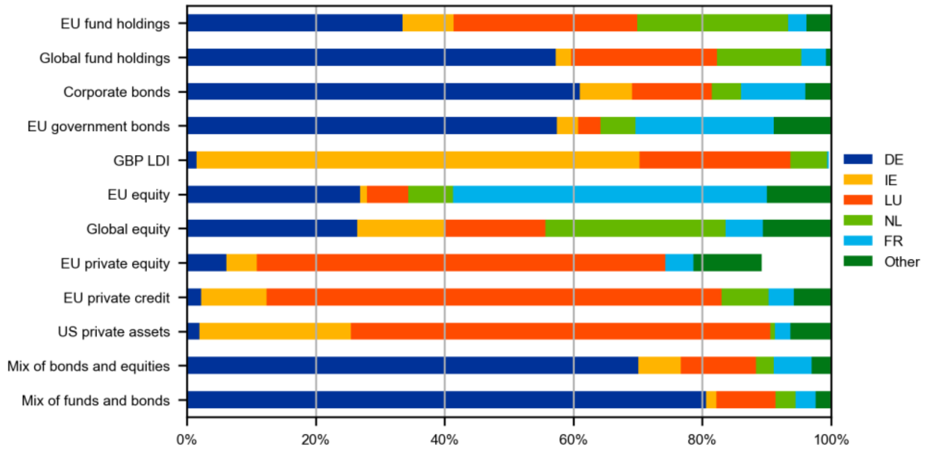
Fund cohorts are dispersed across multiple jurisdictions. Figure 16 shows, for each cluster, the distribution of NAV by fund manager location. The most relevant countries for “other” and “none” AIFs are Germany, Ireland, Luxembourg, the Netherlands and France. Some clusters are broadly distributed across jurisdictions: the “global equity” and “EU fund holdings” clusters show a relatively balanced geographical footprint, with no single country holding more than 50% of the market share. Meanwhile, other clusters are more geographically concentrated: mixed funds, bond funds and “Global fund holdings” funds are predominantly managed in Germany, GBP-denominated LDI funds and private asset funds are typically managed in Ireland or Luxembourg, and EU equity funds are heavily concentrated in France.

The geographic distribution of fund clusters across EU jurisdictions creates distinct challenges depending on whether similar funds are dispersed or concentrated. When similar AIFs are spread across multiple jurisdictions, shocks affecting a cluster may originate in one location but generate spillover effects through shared exposures in others. Consequently, financial stability risks are not confined to where they first materialise, which complicates cross-border oversight. Conversely, geographic concentration raises other concerns: Luxembourg and Ireland manage the vast majority of private credit and private equity funds, thus concentrating potential risks from illiquid, opaque exposures in just two locations. Similarly, Germany hosts the bulk of bond-focused and mixed-strategy AIFs, creating exposure to concentrated interest rate and duration risks. These concentrations may create operational bottlenecks during stress episodes if multiple high-risk funds require simultaneous monitoring or intervention.

Effective oversight of these cross-jurisdictional fund clusters requires coordinated pan-European supervision and risk assessment frameworks. Since funds with similar exposures operate across multiple EU jurisdictions, meaningful monitoring depends on coordination among NCAs rather than isolated national oversight. ESMA and the ESRB play crucial roles in identifying risks from collective behaviour within fund clusters, assessing cluster-level risk signals and facilitating information-sharing between NCAs. The introduction of a common yield buffer requirement for GBP-denominated LDI funds in Ireland and Luxembourg, coordinated with ESMA, illustrates how coordination among authorities can help mitigate cross-border risks and avoid regulatory arbitrage. Our cluster-based analysis provides diagnostic tools to help NCAs evaluate whether their domestic fund sector represents a small or significant portion of EU-wide funds with specific portfolio allocations, offering essential context for the accurate assessment of the systemic risks posed by AIFs.

Figure 16
Location of fund manager

(share of NAV in percent)



Sources: AIFMD data (2024 Q4), own calculations.
Notes: Cluster NAVs may not sum to 100% because certain data are withheld for confidentiality reasons. DE: Germany; IE: Ireland; LU: Luxembourg; NL: Netherlands; FR: France.

6 Conclusion

We demonstrate how unsupervised clustering algorithms can systematically identify groups of investment funds with similar strategies and exposures. Our analysis of AIFMD data reveals 12 distinct fund cohorts among “other” and “none” AIFs, representing €3.7 trillion in assets under management. These clusters exhibit clear economic interpretations – from traditional bond and equity funds to specialised vehicles like GBP-denominated LDI funds and private asset funds – that are often obscured by regulatory categories. Our clustering methodology outperforms traditional AIFMD classifications in explaining fund return variance and successfully identifies funds with similar risk profiles that regulatory labels fail to capture.

Our clustering approach provides authorities with an improved framework for assessing collective risks. For example, our cluster-based analysis is able to reveal concentrated leverage risks in GBP-denominated LDI funds and widespread liquidity mismatches across fund types; risks that would remain hidden under traditional regulatory classifications, which group specialised vehicles like LDI funds and private credit funds alongside conventional bond funds. The extensive interconnectedness of all fund cohorts with insurance companies and pension funds creates channels for systemic risk transmission. The geographic dispersion of similar fund cohorts across EU jurisdictions underscores the importance of cross-border coordination through ESMA and the ESRB for effective macroprudential oversight of collective behaviour.

Future research could build on our methodology to deepen the understanding of collective fund behaviour. One promising direction would be to examine the conditions under which funds with similar exposures simultaneously sell assets, drawing on the substantial anecdotal evidence from past crisis episodes. While our work uses fund exposures to identify potentially problematic fund cohorts, the specific catalysts and mechanisms that lead to coordinated behaviour remain unexplored. Such research would benefit from incorporating additional factors beyond portfolio exposures, such as the composition of the investor base.

Our clustering approach can be readily applied to other datasets and refined further. The algorithm could be extended to granular holdings data available to NCAs, commercial datasets with a similar structure, or combined datasets that include both UCITS funds and AIFs in order to identify similar investment patterns across different regulatory frameworks. EMIR data could enrich the analysis by providing more granular information on derivatives exposures. Various refinements of the clustering algorithm are possible, including dynamic clustering with time-varying parameters or hierarchical clustering approaches.

As the asset management sector continues to grow and evolve, there is a pressing need for sophisticated tools to identify and monitor collective risks. Our work provides financial stability authorities with a practical methodology for moving beyond individual fund analysis towards the systematic assessment of fund groups. This approach enables more effective prudential oversight in an era where collective behaviour may increasingly pose systemic risks.

References

Allaire, N., Breckenfelder, J. and Hoerova, M. (2023), “Fund fragility: the role of the investor base”, *ECB Working Paper*, No 2874.

Arthur, D. and Vassilvitskii, S. (2007), “K-Means++: The Advantages of Careful Seeding”, *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, 7-9 January 2007, pp. 1027-1035.

Barucca, P., Mahmood, T. and Silvestri, L. (2021), “Common asset holdings and systemic vulnerability across multiple types of financial institution”, *Journal of Financial Stability*, Vol. 52.

Bouveret, A., Ferrari, M., Grill, M., Molestina Vivar, L., Schmidt, D.J. and Weistroffer, C. (2025A), “Leveraged investment funds: A framework for assessing risks and designing policies”, *Macroprudential Bulletin*, 26.

Bouveret, A., Darpeix, P.-E., Ferrari, M., Grill, M., Molestina Vivar, L., Okseniuk, D., Raillon, F., Schäfer, A., Schmidt, D.J. and Weistroffer, C. (2025B), “Containing risks from leverage in alternative investment funds”, *ESRB Occasional Paper Series* 28.

Central Bank of Ireland (2022), *The Central Bank’s macroprudential policy framework for Irish property funds*, November.

Central Bank of Ireland (2024), *The Central Bank’s macroprudential policy framework for Irish authorised GBP-denominated LDI funds*, April.

Cera, K., Daly, P., Hermans, L., Molitor, P., Schwartz Blicke, O., Sowiński, A. and Telesca, E. (2024), “Private markets, public risk? Financial stability implications of alternative funding sources” *Financial Stability Review*, May.

Chen, Q., Goldstein, I. and Jiang, W. (2010), “Payoff complementarities and financial fragility: Evidence from mutual fund outflows”, *Journal of Financial Economics*, Vol. 97(2), pp. 239-262.

Commission de Surveillance du Secteur Financier (2024), *Macroprudential measures for GBP Liability Driven Investment Funds*.

Custodio João, I., Lucas, A., Schaumburg, J. and Schwaab, B. (2023), “Dynamic clustering of multivariate panel data”, *Journal of Econometrics*, Vol. 237(2), 105281.

Custodio João, I., Schaumburg, J., Lucas, A. and Schwaab, B. (2024), “Dynamic nonparametric clustering of multivariate panel data”, *Journal of Financial Econometrics*, Vol. 22(2), pp. 335-374.

Dunne, P., Ghiselli, A., Ledoux, F. and McCarthy, B. (2023), “Irish-Resident LDI Funds and the 2022 Gilt Market Crisis”, *Central Bank of Ireland Financial Stability Notes*, Vol. 2023, No 7.

European Securities and Markets Authority (2014*), Guidelines on reporting obligations under Articles 3(3)(d) and 24(1), (2) and (4) of the AIFMD*.

European Securities and Markets Authority (2019), *EU Alternative Investment Funds - 2019 Statistical Report*.

European Securities and Markets Authority (2020), *EU Alternative Investment Funds - 2020 Statistical Report*.

European Securities and Markets Authority (2021), *Guidelines on Article 25 of Directive 2011/61/EU*.

European Securities and Markets Authority (2022a), "Advice of the European Securities and Markets Authority of 23 November 2022 on a proposed measure by the Central Bank of Ireland under Article 25 of Directive 2011/61/EU", *advice*.

European Securities and Markets Authority (2022b), "ESMA welcomes NCAs' work to maintain resilience of liability driven investment funds", *press release*.

European Securities and Markets Authority (2024), "Advice of the European Securities and Markets Authority of 26 April 2024 on a proposed measure by the Commission de Surveillance du Secteur Financier under Article 25 of Directive 2011/61/EU", *advice*.

ESMA (2025), *Final Report on the Guidelines on LMTs of UCITS and open-ended AIFs*.

European Systemic Risk Board (2023), "EU Non-bank Financial Intermediation Risk Monitor 2023", *NBFI Monitor*, No 8, June.

Farnè, M. and Vouldis, A. (2017), "Business models of the banks in the euro area", *ECB Working Paper*, No 2070, May.

Financial Stability Board (2020), *Holistic Review of the March Market Turmoil*, November.

Financial Stability Board (2024), *Leverage in Non-bank Financial Intermediation - Consultation report*.

FSB (2024), *Liquidity Preparedness for Margin and Collateral Calls: Final report*, December.

Girardi, G., Hanley, K.W., Nikolova, S., Pelizzon, L. and Sherman, M.G. (2021), "Portfolio similarity and asset liquidation in the insurance industry", *Journal of Financial Economics*, Vol. 142(1), pp. 69-96.

Haquin, J.-H. and Proietti, R. (2024), "Assessing risks posed by leveraged AIFs in the EU", ESMA TRV Risk Analysis 60-1389274163-2572, January.

Haquin, J.-H. and Proietti, R. (2025), "Annual risk assessment of leveraged AIFs in the EU – 2024", ESMA TRV Risk Analysis 50-524821-3642, April.

Kaufman, L. and Rousseeuw, P.J. (1987), “Clustering by means of medoids”, Dodge, Y. (ed.), *Statistical Data Analysis Based on the L1-Norm and Related Methods*, Amsterdam: North-Holland, pp. 405-416.

Kumar, V., Chhabra, J.K. and Kumar, D. (2014), “Impact of distance measures on the performance of clustering algorithms”, in *Intelligent Computing, Networking, and Informatics: Proceedings of the International Conference on Advanced Computing, Networking, and Informatics*, India, June 2013, pp. 183-190, Springer India.

Lucas, A., Schaumburg, J. and Schwaab, B. (2017), “Bank business models at zero interest rates”, *Journal of Business & Economic Statistics*, Vol. 36(1), pp. 1-17.

Park, H.S. and Jun, C.H. (2009), “A simple and fast algorithm for K-medoids clustering”, *Expert Systems with Applications*, Vol. 36(2), pp. 3336-3341.

Sakakibara, T., Matsui, T., Mutoh, A. and Inuzuka, N. (2015), “Clustering mutual funds based on investment similarity”, *Procedia Computer Science*, Vol. 60, pp. 881-890.

Satone, V., Desai, D. and Mehta, D. (2021), “Fund2vec: mutual funds similarity using graph learning”, arXiv:2105.14889.

Shalev-Shwartz, S. & Ben-David, S. (2014), *Understanding machine learning: From theory to algorithms*, Cambridge University Press.

Vozlyublennaia, N. and Wu, Y. (2016), “Mutual Funds Apart From the Crowd”, *SEC Working Paper*, April.

Annex

A: Data quality filters

- We remove observations if at least one of the following conditions is satisfied:
- missing NAV, AuM or exposures data;
- NAV, AuM or total exposures are less than 1 million;
- the ratio of AuM to total exposures is significantly different from 1;
- the ratio of NAV to total exposures is greater than 1.05;
- the ratio of total exposures to NAV is greater than 100.

B: Euclidean distance applied to fund exposures

Euclidean distance, $d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$, the default measure of distance for clustering, can produce unintuitive results. In fact, Euclidean distance does not satisfy the two key properties that an effective distance measure for fund exposures should have (see Section 3.1). To illustrate these limitations, consider the following hypothetical fund portfolios:

- Fund 1: 100% Asset A
- Fund 2: 100% Asset B
- Fund 3: 50% Asset C, 50% Asset D
- Fund 4: 10% Asset B, 90% Asset C

Table 4 below shows the matrix of squared Euclidean distances between the four investment funds. Although funds 1, 2 and 3 are fully concentrated in different, non-overlapping assets, the Euclidean distance between fund 1 and fund 3 is smaller than that between fund 1 and fund 2. Similarly, despite funds 2 and 4 sharing exposure to Asset B, the Euclidean distance would suggest that fund 2 is more similar to fund 3 than to fund 4. This behaviour arises from the squaring of differences in each dimension. Squared terms reduce the impact of small deviations spread across many asset classes, causing diversified but non-overlapping portfolios to appear closer than they are. As a result, Euclidean distance tends to understate dissimilarity between diversified funds and overstate it between concentrated ones, even when there is no actual overlap. This makes it poorly suited to measuring portfolio similarity.

Table 4

Distance matrices for hypothetical fund portfolios: Squared Euclidean distance

	Fund 1	Fund 2	Fund 3	Fund 4
Fund 1	0	2	1.5	1.82
Fund 2		0	1.5	1.62
Fund 3			0	0.42
Fund 4				0

Source: Own calculations.
Note: Distance matrices are based on the following hypothetical fund portfolios: Fund 1: 100% Asset A; Fund 2: 100% Asset B; Fund 3: 50% Asset C, 50% Asset D; Fund 4: 10% Asset B, 90% Asset C.

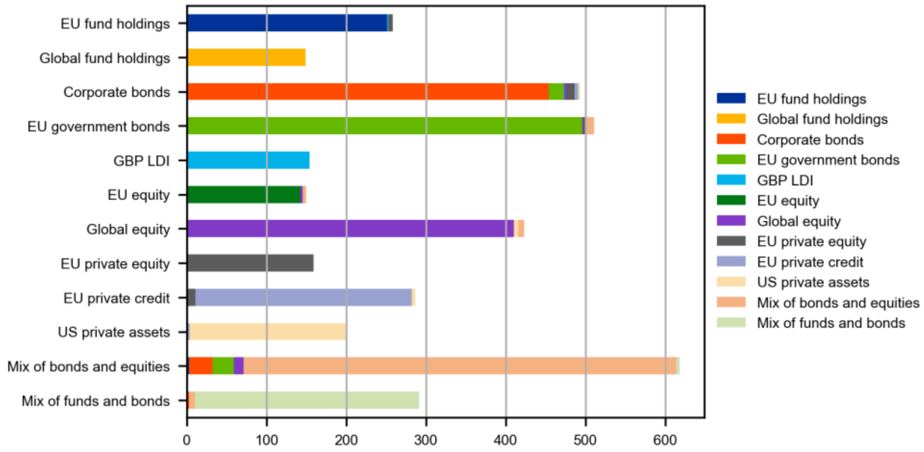
C: Robustness check with $w = 0.5$

To check robustness, we re-ran the clustering algorithm, adjusting the weights assigned to Manhattan distances calculated at both granular and aggregated levels of fund exposures. The resulting differences in cluster composition remain minimal (Figure 17).

Figure 17

Robustness check with equal weight $w = 0.5$ on detailed and broad asset classes

(x-axis: NAV in EUR billions; y-axis: original clusters with $w = 0.75$ presented in Section 4; legend: clusters with $w = 0.5$)



Sources: AIFMD data (2024 Q4), own calculations.
Notes: Very uncommon cluster membership combinations are not displayed due to confidentiality reasons. The NAV of these funds is approximately 1% of the total NAV of the sample.

D: Cohorts of funds based on the whole universe of AIFs

This section extends our clustering approach to the complete universe of alternative investment funds. While our previous focus on funds bearing the ambiguous labels “other” or “none” demonstrated the strengths of our method, applying it to clearly categorised funds – including funds of funds, hedge funds, real estate funds and private equity funds – provides additional validation and new insights into AIFMD data quality. If clusters align with established AIF types, this would validate both our methodology and the AIFMD categorisation system. However, if clusters contain heterogeneous fund types and strategies, this may indicate either imprecise AIF type definitions or inaccuracies in the exposure data reported. By systematically examining divergences between clustering outcomes and established classifications, we can better assess both the coherence of the AIFMD framework and the reliability of exposure reporting practices.

To perform clustering on the extended dataset, we adhere to the procedures outlined in Section 3.3. First, we remove those AIFs with low-quality exposure data, giving us a refined sample of 21,332 funds with a total NAV of €6.6 trillion. We then apply the elbow method to determine the optimal number of clusters and use the best initialisation from 1,000 random draws. Based on the elbow method, we proceed with 20 clusters for the comprehensive dataset. The larger cluster count reflects the greater diversity present in the complete AIF universe compared with our initial focused analysis of funds bearing the labels “other” or “none”. In contrast to the clusters presented in Section 4, we avoid naming clusters here so as to avoid confusion with the cluster labels used for our main results.

Clusters 1-7 mostly align with the self-reported AIF types. Figure 18 compares these types with the new clusters based on the entire universe of AIFs. Clusters 1-4 primarily comprise funds of funds with different geographical exposures, as well as funds labelled “other” or “none” that primarily invest in other funds. Similarly, Clusters 5 and 6 contain funds clearly labelled as private equity funds with an EU or US/global geographical focus, as well as funds labelled as “other” or “none” with similar exposures. Cluster 7 contains real estate funds. This strong alignment between clusters and AIF types validates the quality of the exposure data and the effectiveness of our clustering approach.

Clusters 8-17 are predominantly composed of “other” and “none” funds and often correspond to the clusters identified in Section 4. Figure 19 compares cluster assignments based on the sample of “other” and “none” funds with the new clusters based on the entire universe of AIFs. Most of the original clusters do not change much, such as the two bond fund clusters (Clusters 8 and 9), GBP-denominated LDI funds (Cluster 10), the two equity fund clusters (Clusters 11 and 12), the US private assets cluster (Cluster 15) and the two mixed fund clusters (Clusters 16 and 17). The original EU private credit cluster is split into two parts, with Cluster 13 capturing pure EU private credit funds and Cluster 14 grouping together private credit funds with more diversified exposures, both in terms of geography and asset class. Overall, the large overlap with clusters based on the sample of “other” and “none” funds illustrates the stability and robustness of our algorithm.

Clusters 18-20 represent small groups of funds that are not well-represented by self-reported AIF types or by the clusters from Section 4. These clusters predominantly contain funds with concentrated exposures to asset classes not represented in the other clusters: funds in Cluster 18 mostly hold cash, funds in Cluster 19 invest in assets in the “other-other” category, and funds in Cluster 20 are predominantly exposed to UK and Swiss private equity and fund holdings. Due to their small size, these groups of funds did not previously form distinct clusters.

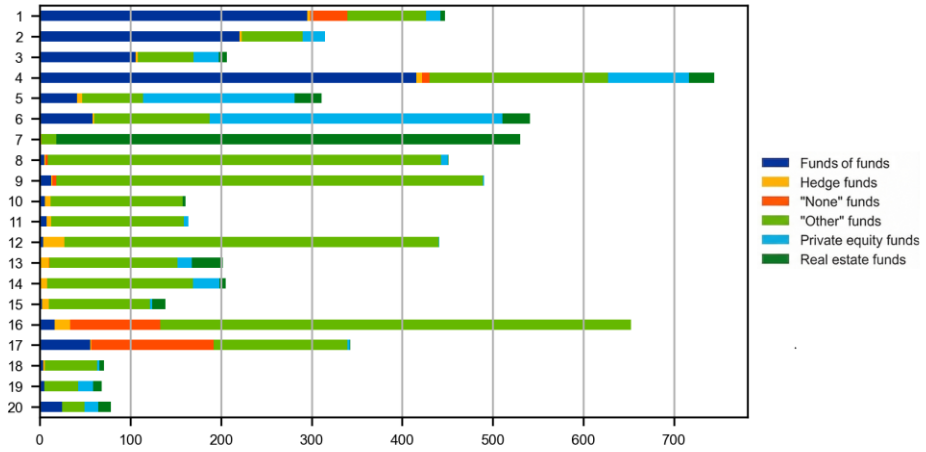
Hedge funds lack a precise definition based on their asset exposure and are consequently scattered across multiple clusters. Self-reported hedge funds are spread across different clusters because they rely on complex derivative strategies to build synthetic leverage rather than direct exposure to specific asset classes. The largest concentrations appear in clusters that predominantly contain traditional equity, bond and mixed funds. This pattern is not surprising: equity hedge funds and fixed income hedge funds have more similar exposures to traditional funds targeting the same asset classes than they do to each other. While meaningful differences exist between hedge funds and traditional funds within identical asset categories – with hedge funds typically employing short selling and derivatives – there are too few equity hedge funds and fixed income hedge funds to form separate clusters. Given these limitations, we recommend relying primarily on the detailed AIFMD sub-classification to identify hedge funds, using clustering only for cross-validation, since sophisticated trading strategies may not be adequately captured through standard exposure data.

Our approach identifies funds reporting inconsistent exposures and AIF types that merit supervisory investigation. Clusters 1-4, defined by their exposure to CIUs, contain significant proportions of funds labelled as private equity or real estate. Similarly, Clusters 5-6, defined by their exposure to unlisted equity, contain numerous funds of funds and real estate funds. These seemingly mislabelled funds account for €400 billion of NAV. However, these funds' exposures generally align with the average exposure profiles of their clusters, suggesting correct algorithmic placement rather than errors. For instance, private equity funds in Cluster 1 average 80% exposure to CIUs, thus justifying their grouping with funds of funds. This indicates potential data entry errors or misinterpretation of AIFMD guidelines – such as funds of private equity/real estate (PE/RE) funds being reported as PE/RE funds, or funds of PE/RE funds reporting their indirect exposures to PE/RE instead of their direct exposures to other CIUs.

Figure 18

Comparison of clusters based on the whole AIF universe with AIF types

(NAV in EUR billions)

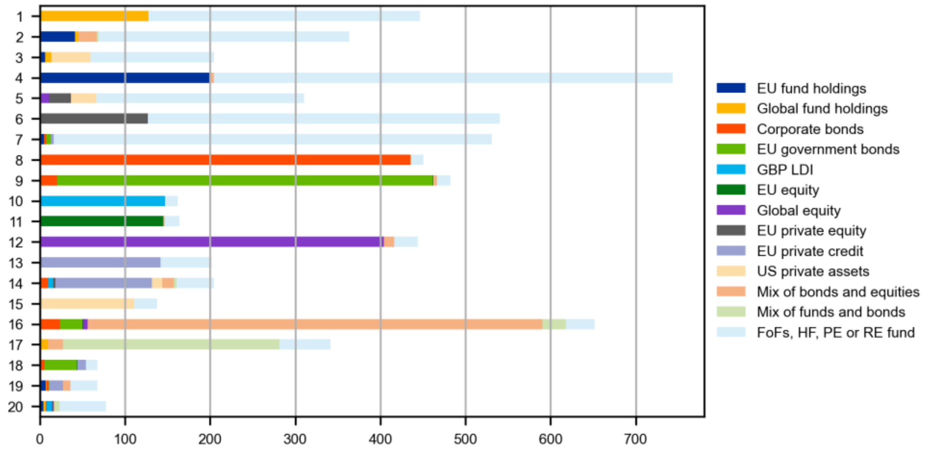


Sources: AIFMD data (2024 Q4), own calculations.
Notes: Comparison of clusters estimated using the whole universe of AIFs (y-axis) with AIF types reported by fund managers (colours). Very uncommon cluster-AIF type combinations are not displayed due to confidentiality reasons. The NAV of these funds is less than 1% of the total NAV of the sample.

Figure 19

Comparison of clusters based on whole AIF universe with clusters based on “other” and “none” AIFs

(NAV in EUR billions)



Sources: AIFMD data (2024 Q4), own calculations.
Notes: Comparison of the clusters estimated using the whole universe of AIFs (y-axis) with clusters of “other” and “none” funds from Section 4 (colours). Very uncommon cluster membership combinations are not displayed due to confidentiality reasons. The NAV of these funds is less than 1% of the total NAV of the sample.

Acknowledgements
We are grateful for the valuable comments made by members of the joint ATC-ASC Expert Group on Non-Bank Financial Intermediation (chaired by Steffen Kern and Richard Portes), by the editor of the ESRB Occasional Paper Series, by one anonymous reviewer, and by Bernd Schwaab. We would also like to thank Pawel Fiedor for his outstanding support and feedback throughout the project, as well as Marco D'Errico, Angel-Ivan Moreno and Dorota Okseniuk.
All remaining errors are ours.

Sophie Almekinders
London School of Economics and Political Science, London, United Kingdom; email: s.almekinders@lse.ac.uk

Antoine Bouveret
European Securities and Markets Authority, Paris, France; email: antoine.bouveret@esma.europa.eu

Massimo Ferrari
European Securities and Markets Authority, Paris, France; email: massimo.ferrari@esma.europa.eu

Michael Grill
European Central Bank, Frankfurt am Main, Germany; email: michael.grill@ecb.europa.eu

Mattia Pividori
Commission de Surveillance du Secteur Financier, Luxembourg, Luxembourg; email: mattia.pividori@cssf.lu

Roberto Proietti
European Securities and Markets Authority, Paris, France; email: roberto.proietti@esma.europa.eu

Daniel Jonas Schmidt
De Nederlandsche Bank, Amsterdam, The Netherlands; email: d.j.schmidt@dnb.nl

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.esrb.europa.eu

For specific terminology please refer to the ESRB glossary (available in English only).

PDF ISBN 978-92-9472-441-0, ISSN 2467-0669, doi:10.2849/9907233, DT-01-26-005-EN-N