Discussion Paper

Deutsche Bundesbank
No 15/2026

Bivariate sudden stop analysis of equity and bond fund flows to emerging markets using isolation forest

Susanne Griebsch
(University of Applied Sciences Mainz)

Andreas Röthig
(Deutsche Bundesbank)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank, Press and Public Relations Division, at the above address or via email: www.bundesbank.de/contact

Internet http://www.bundesbank.de

DOI https://doi.org/10.71734/DP-2026-15
ISBN 978-3-98848-072-9
ISSN 2941-7503

Deutsche Bundesbank Discussion Paper No 15/2026

Bivariate Sudden Stop Analysis of Equity and Bond Fund Flows to Emerging Markets using Isolation Forest*

Susanne Griebsch
University of Applied Sciences Mainz

Andreas Röthig
Deutsche Bundesbank

Abstract

This paper applies machine learning methods and anomaly detection to sudden stop analysis of portfolio flows. Using the isolation forest methodology, univariate as well as bivariate sudden stops of equity and bond fund flows to emerging markets are generated. An anomaly score and an anomaly classification are provided. The results point to an increase in anomalous portfolio flows to emerging markets in recent years. In addition, the isolation forest methodology appears to yield better results than the traditional approach to sudden stop analysis in classifying anomalies connected with the recent capital flow volatility related to the outbreak of the COVID-19 pandemic as well as the interest rate reversal in advanced economies in recent years. The bivariate approach to anomaly detection is better able to identify anomalous episodes of financial stress, where both equity and bond markets are simultaneously affected. Most of the classified anomalies are related to fund flow stops (i.e. simultaneous stops to both equity and bond flows) or surges (i.e. surges in both equity and bond flows). In general, univariate and bivariate anomaly detection using machine learning techniques can play an important part and lead to a better understanding of sudden stops and surges.

Keywords: Capital Flows, Portfolio Flows, Sudden Stops, Emerging Markets, Machine Learning, Isolation Forest, Anomaly Detection

JEL classification: E32, F30, F32, G15.

*Contact: Susanne Griebsch, University of Applied Sciences Mainz, susanne.griebsch@hs-mainz.de. Andreas Röthig, Deutsche Bundesbank, andreas.roethig@bundesbank.de. We would like to thank Christoph Memmel, Christian Proaño and participants at the Bundesbank's research seminar for helpful comments.

1 Introduction

Extreme waves of capital flows can pose a significant risk to macroeconomic and financial stability in emerging markets. Analysing and monitoring of swings in capital flows have therefore long been a key concern for market participants, researchers and policymakers alike. According to the Committee on the Global Financial System (2021), the most significant risks are related to sudden stops of capital flows and the growing importance of portfolio investors. In addition to sudden stops[1], large capital inflows (i.e. surges) can pose significant problems, especially if they are very large when compared to the size of the domestic economy or the domestic financial sector. Agosin and Huaita (2011, p. 663) argue that capital flow surges have the potential to “sow the seeds for the ensuing sudden stops”.[2] Especially in emerging market economies, capital flow waves are associated with an increased likelihood of banking and economic crises.[3] It remains crucial that policymakers try to monitor and better understand sudden stops as well as the potential damage these waves of capital flows can cause.

The rising importance of portfolio investors with regard to international capital flows is another key topic. While historically, capital flow analysis tended to focus on foreign direct investment, more recently bank flows and in particular portfolio flows gained significantly in importance.[4] Sudden stop analysis is particularly suited for portfolio flows, such as equity and bond fund flows, because flows into equity and bond markets can be prone to sudden reversals. In addition, Eichengreen and Gupta (2016) argue that the growing size of international financial markets and financial transactions can lead to countries being even more exposed to capital flow reversals than in the past. Emerging markets have also become more fully integrated into international financial markets over the past decades.[5] Sudden stop analysis is therefore an appropriate tool for understanding and monitoring portfolio flows. However, traditional sudden stop analysis has three main shortcomings. First, it is a univariate time series analysis. That is, regarding portfolio flows, sudden stops for equity flows and bond flows are typically investigated in isolation. Second, the outcome of sudden stop analysis is generally a variable with only three states: sudden stops (i.e. significant outflows), normal flows, and surges (i.e. significant inflows). Since the resulting variable is not continuous, detailed interpretation and further work based on this information is difficult. Third, these three states of capital flows are set rather arbitrarily, based on the empirical distribution. Therefore, whether a sudden stop or surge is identified depends on pre-set thresholds, for example on rolling standard deviations.

[1] Sudden stops can be caused by both the retreat of international investors and by the sudden flight of local investors, as discussed in Rothenberg and Warnock (2011).

[2] See also Agosin and Huaita (2012). Tillmann (2013) argues that capital inflow shocks can have a significant impact on house prices and equity prices, potentially leading to price bubbles. For detailed information on the potential linkages between surges and sudden stops, see Sula (2010) and Efremidze, Kim, Sula, and Willett (2017).

[3] See e.g. Reinhart and Reinhart (2008), Hutchinson, Noy, and Wang (2010), and Guidotti, Sturzenegger, Villar, de Gregorio, and Goldfajn (2004). Calvo (1998, p. 47) calls sudden stops “dangerous”, resulting potentially in “bankruptcies, and destruction of human capital and local credit channels.” Hutchinson and Noy (2006) find that sudden stop crises can affect economies more than currency crises.

[4] See e.g. Byrne and Fiess (2016).

[5] See e.g. Broto, Díaz-Cassou, and Erce (2011). For more information on the potential implications of global financial integration, see also the recent research on the global financial cycle. See e.g. Beutel, Emter, Metiu, Prieto, and Schüler (2025) and Proaño, Virla, and Strohsal (2025) for more details.

This paper endeavours to correct these three shortcomings. The contributions of this paper are therefore threefold. First, we build a continuous sudden stop variable using anomaly scores from isolation forest analysis. This sudden stop anomaly score allows for a detailed analysis and interpretation of the results. It can also act as a continuous stress variable for further analysis of capital flows. Second, in addition to univariate sudden stop analysis, we perform a bivariate sudden stop analysis for equity and bond flows to emerging markets using isolation forest. This allows for a more comprehensive analysis of the key portfolio flows to emerging markets. Third, sudden stops and surges are identified by employing the depth of the isolation tree required to isolate each individual data point. This path length represents the decision function. In general, anomalies are less frequent than regular observations and they lie further away from the more regular observations in the feature space. A more regular data point should therefore require more partitions to be identified, and hence a longer path length. Accordingly, the isolation forest algorithm leads to significantly shorter path lengths for anomalies. Using this approach, sudden stops are identified without exogenously applying standard deviation bands and without relying on the empirical distribution. In addition to these three main contributions, this paper applies sudden stop analysis to weekly EPFR data - instead of monthly or quarterly data - and uses machine learning techniques (i.e. isolation forest) for anomaly detection in capital flows. We hence seek to contribute to a better understanding and alignment between traditional sudden stop analysis and modern anomaly detection methods using machine learning.

Applying the univariate isolation forest and comparing its results with those of the traditional sudden stop analysis yields the following findings. First, one of the key advantages of the isolation forest approach is, that it provides a continuous anomaly score series, allowing for a detailed analysis of the magnitude of anomalies. The results show, that both the frequency of the occurrence (i.e. the anomaly classification) and the intensity of anomalies (i.e. the anomaly score) have significantly increased over the observed time period. This is the case for both equity and bond markets. Second, it is useful to differentiate between anomalous data points and anomalous data patterns. Hence, while the anomaly classifications for specific data points may differ between the two alternative methods, the comparison of anomalous episodes might point to more similarities. For example, while the classification of anomalous equity flows yields different results using the two different methods, the isolation forest algorithm identifies all the sudden stop and surge episodes, also identified by the traditional method (although the durations of these episodes are not equal). In addition, episodes where the traditional sudden stop analysis points to normal flows are also sensibly classified as normal by the isolation forest.

The bivariate analysis of anomalies using isolation forest leads to several interesting results. First, the bivariate representation of the anomaly classification shows that the anomalies are on the edges of the data space. In contrast, the sudden stops and surges identified by the traditional method are often closer to the center of the data space. In addition, most of these anomalies that the traditional method detects are surrounded by other data points, and therefore not isolated. These traditional sudden stops and surges would therefore generally not be regarded as belonging to an anomalous pattern within the data space. Hence, it can be concluded, that a large fraction of anomalies detected by the traditional method are not anomalous but more likely normal data points. In contrast, the anomalies classified by the bivariate isolation forest are clearly on the edges of the data space. Second, the anomalies classified by the bivariate isolation forest mainly correspond to overall fund inflows (i.e. simultaneous inflows into equity and bond funds) or outflows (i.e. simultaneous outflows from equity and bond funds), and therefore to bivariate fund surges or bivariate fund sudden stops. A smaller number of identified anomalies represent a mixture of equity outflows and bond inflows which could be labelled 'safe haven' or 'risk-off' flows. And a small number of anomalies represent a mixture of equity inflows and bond outflows, which could be interpreted as 'risk-on' flows from bond into equity markets. Again, the anomalies identified by the traditional method cannot be interpreted in this way since they are largely in the center of the data space.

In addition, and given the recent history of financial markets and crises since the outbreak of the COVID-19 pandemic, as well as the significant interest rate reversal in advanced economies in recent years, the isolation forest methodology appears to yield better results in classifying anomalies than the traditional approach. The bivariate approach to anomaly detection is better able to identify anomalous episodes of volatility and turmoil, where both equity and bond markets are simultaneously affected. In general, univariate and bivariate anomaly detection using machine learning techniques can play an important part and lead to a better understanding of sudden stops and surges.

The remainder of the paper is structured as follows. Section 2 describes the data and presents the traditional approach to sudden stop analysis. Section 3 introduces the new approach to sudden stop analysis using isolation forest. Section 4 compares the results of the univariate anomaly detection to the traditional sudden stop analysis. Section 5 discusses the findings of the bivariate anomaly detection for equity and bond fund flows. Finally, section 6 provides conclusions.

2 Data and traditional sudden stop analysis

We study sudden stops and surges of international equity and bond fund flows to emerging markets using weekly data from the Emerging Portfolio Fund Research (EPFR) Global Database. The weekly data cover equity fund and bond fund investing in emerging markets from 17 November 2004 to 2 October 2024 ($N = 1038$ observations). The methodology for the traditional calculation of sudden stops and surges follows the approach used by Li, de Haan, and Scholtens (2019) and Forbes and Warnock (2012).[6] However, while Li et al. (2019) and Forbes and Warnock (2012) employ monthly and quarterly data, we apply this method to weekly EPFR data.

We first compute the 52-week moving sum C_t of fund flows P_t, for equity and bond flows respectively, and afterwards year-over-year changes $\Delta(C_t)$.[7]

$$C_t = \sum_{i=0}^{51} P_{t-i}, \qquad t = 52, 53, ..., N. \tag{1}$$

[6]For more information on EPFR data, as well as an application to fund flow surges, see Li, de Haan, and Scholtens (2018).

[7]See e.g. Forbes and Warnock (2012) for a similar approach to sudden stop analysis. Calvo, Izquierdo, and Talvi (2006, p. 405) note that a "sudden stop is a sharp fall in capital inflows relative to their past trajectory." In addition, the use of year-on-year changes and rolling means and standard deviation in the calculation of sudden stops fits well to the characterization of capital flow dynamics as waves or swings, as described in Forbes and Warnock (2021).

Table 1: Summary statistics

Series	Obs.	Mean	Max.	Min.	Std. Dev.	Skewness	Kurtosis
EPFR flows							
- Equity	1038	667.29	21012.41	-10108.68	2678.98	0.8139	8.6099
- Bonds	1038	137.88	4671.40	-20991.38	1585.60	-4.3555	53.2567
$\Delta(C_t)$							
- Equity	832	5802.90	202692.08	-121753.62	70229.92	0.4994	2,9216
- Bonds	832	-1601.39	104827.94	-149572.09	51803.78	-0.5542	3.2560

Notes: Emerging Portfolio Fund Research (EPFR) Global Database. Weekly data from 17 November 2004 till 2 October 2024. EPFR flows in Mil. USD.

$$\Delta(C_t) = C_t - C_{t-52}, \qquad t = 104, 105, ..., N. \tag{2}$$

Table 1 presents summary statistics for the raw EPFR data as well as for the $\Delta(C_t)$ series, for equity and bond flows, respectively.[8]

In addition, a 104-week rolling mean as well as a 104-week rolling standard deviation of $\Delta(C_t)$ are calculated. If $\Delta(C_t)$ is below [above] the rolling mean minus [plus] twice the rolling standard deviation, then a sudden stop [surge] is identified. Figure 1 shows the results of this traditional sudden stop analysis with the blue and rose areas being the 1-standard deviation and 2-standard deviation bands around the rolling 104-week mean of $\Delta(C_t)$. A sudden stop occurs when $\Delta(C_t)$ crosses the lower 2-standard deviation band from above, as is for example the case for equity flows around 2018. A surge occurs when $\Delta(C_t)$ crosses the upper 2-standard deviation band from below, as can be seen for equity flows around 2021.

Figure 1 presents the key result of the traditional sudden stop analysis for weekly data and the most important information for policy makers and market participants with the step function representing the three possible classifications: normal flows when $\Delta(C_t)$ is in-between the 2-standard deviation bands, sudden stops when $\Delta(C_t)$ crosses the lower 2-standard deviation line from above, and surges when $\Delta(C_t)$ crosses the upper 2-standard deviation line from below. A more detailed analysis of the intensity of sudden stops or surges is generally not possible using this approach. Although the standard deviation bands are set in a rather arbitrary fashion, and could therefore be adjusted to the respective data, this is generally not done in the literature. But even if these bands were adjusted, the result would again be a step function with a limited number of classifications, where all observations falling into the same category would be treated as equal. Hence, this approach, irrespective of the setting of standard deviation bands, does not differentiate between different sudden stops, but treats them all as equal. Trying to fix this shortcoming, the following section will introduce a new approach to sudden stop

[8] The series have been tested for stationarity using augmented Dickey Fuller (ADF) tests with different lag structures as well as Phillips-Perron (PP) tests and KPSS tests. The results overall point to the series being stationary. In addition, the analysis has also been carried out with the series being adjusted for inflation. The results using raw data are very similar to the ones employing inflation-adjusted data. Hence, the findings appear to be robust with regard to a potential impact of inflation. The remainder of the paper therefore focusses on the analysis using the raw data.

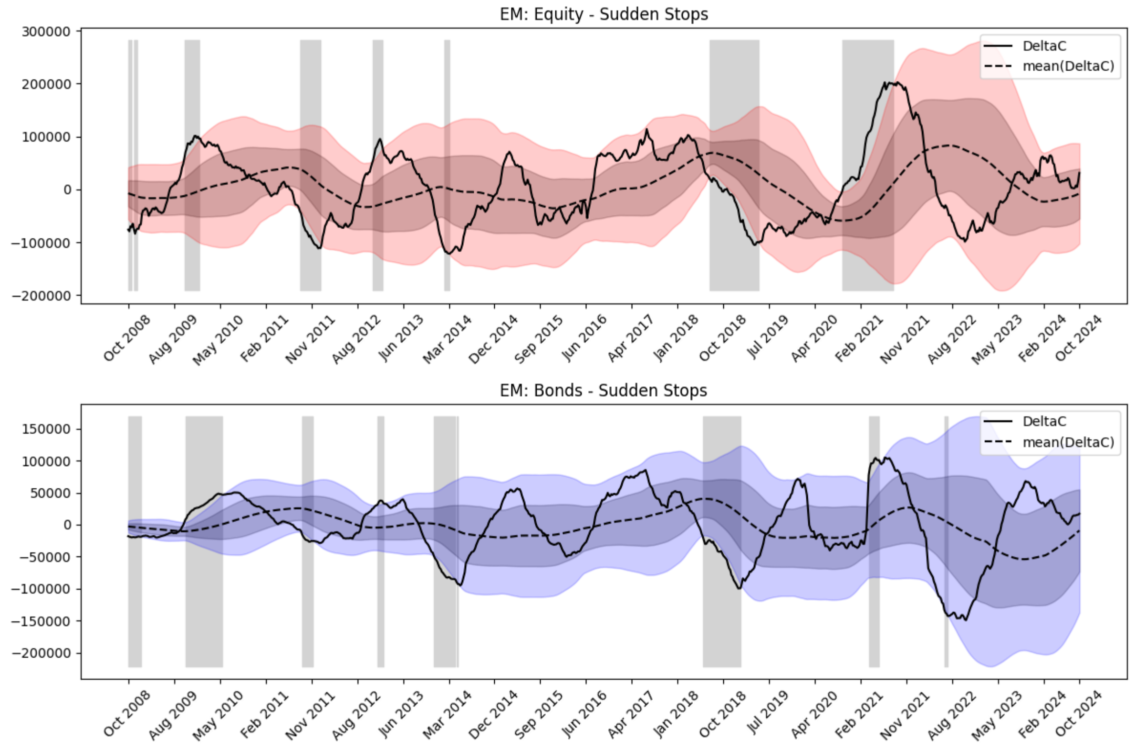

Figure 1: Traditional sudden stop analysis

Notes: One standard deviation and two standard deviation bands are the filled areas in blue and rose, respectively. Darker [lighter] colors correspond to the one [two] standard deviation bands. Gray areas show the occurrence of a sudden stop or surge.

analysis, where each individual data point will be given a unique anomaly score.

A further shortcoming of the traditional sudden stop analysis is the univariate approach, investigating bond and equity flows in isolation. Figure 2 shows the bond and equity signals (i.e. stops and surges) in a combined representation, either employing the 'AND' operator (i.e. simultaneous signals in both bond and equity flows) or the 'OR' operator (i.e. signals in bond or equity flows). Figure 2 shows that, if combined, the signals are often not on the edge of the data space, but rather in the center. Regarding this bivariate representation, it would be difficult to argue, why certain points in the center of the distribution are considered anomalies, i.e. stops or surges. This finding as presented in Figure 2 serves as a further motivation for the introduction of bivariate sudden stop analysis, in addition to univariate sudden stop analysis.

3 A new approach to sudden stop analysis using isolation forest

In contrast to the traditional sudden stop analysis, we implement a machine learning approach without exogenously applying standard deviation bands for identifying sudden stops or surges. The machine learning methodology will itself identify sudden stops, surges, and normal flows based on the ΔCM_t time series for equity and bond flows, where

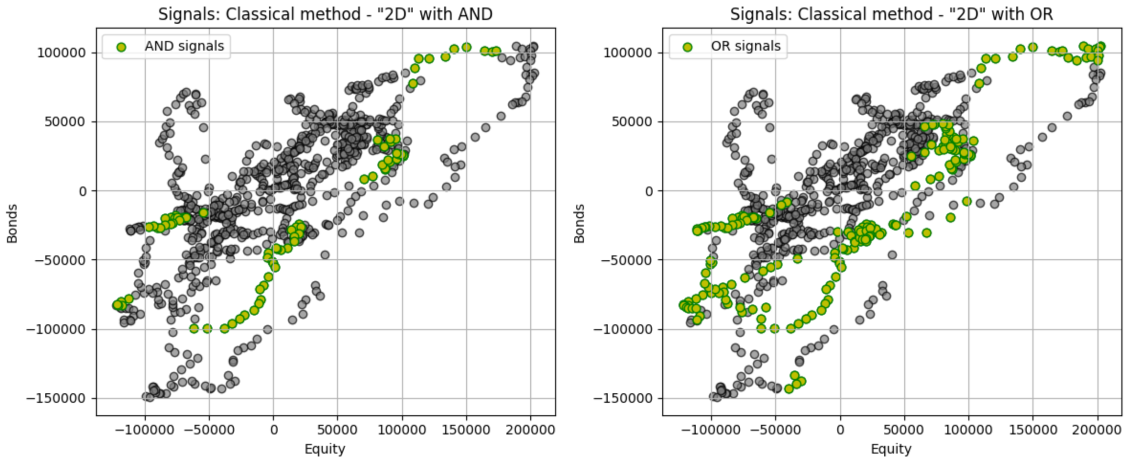

Figure 2: Traditional sudden stop analysis: Bivariate representation

ΔCM_t is computed as the difference between $\Delta(C_t)$ and a 104-week rolling mean of $\Delta(C_t)$.

$$\Delta CM_t = \Delta(C_t) - mean(\Delta(C_t)). \tag{3}$$

This new approach to sudden stop analysis has several advantages when compared to the traditional approach. First, the result of sudden stop analysis will be a continuous variable, the anomaly score. Based on this anomaly score, the flows will be classified as normal or anomalous (i.e. sudden stops and surges). The anomaly score gives researchers and policy makers more detailed and nuanced information regarding the dynamics of the flows, when compared to the traditional approach. Second, an anomaly score and classification are computed for the univariate case (i.e. equity and bond flows separately) as well as the bivariate case (i.e. equity and bond flows together).

An anomaly detection problem can generally be defined by finding data points or patterns that are distinct from the rest of the data. In general, anomaly detection analysis using machine learning can be divided into three main categories: (i) Supervised; (ii) Semi-supervised; or (iii) Unsupervised anomaly detection.[9] Supervised detection can be used when anomalies and normal data are previously known, and are classified accordingly. This classification is then generally used to compare new data with the classified one. The semi-supervised detection is generally employed when only the normal data are known. Detection algorithms would then train with these data to assess the presence of anomalies in new data. Unsupervised anomaly detection is used on unlabeled data, where normal data and anomalies are not previously known. These algorithms assume that there are only few anomalies in the data and that these anomalies have different characteristics from the rest of the data. Since our capital flow data are unlabeled (i.e. anomalies are not previously known), we choose the unsupervised anomaly detection approach.[10]

[9] For more information, see Mendes, Cardoso, Monteiro, and Raposo (2023). For a more detailed survey of anomaly detection approaches, see e.g. Feng, Cai, Yue, Xu, Lin, Chen, and Hu (2022).

[10] Kaufman and Iaremenko (2022, p. 3) argue that "hand labeling anomalies rarely works well." The question of when a data point is really an anomaly is often not easy to answer. In addition, "hand labeling" results in a binary label, indicating whether the data point is normal or not. A continuous anomaly score, based on machine learning techniques, allows to facilitate the early detection of "weak

The machine learning methodology we employ for anomaly detection is the isolation forest.[11] Traditional anomaly detection methods, such as statistical methods, usually build a distribution of normal behavior and then choose a threshold to identify periods that do not conform to the normal behavior. Hence, these approaches first of all detect normal instances, and are not optimized to detect anomalies. This can lead to false alarms or too few anomalies being detected.[12] Isolation forest takes a different approach that explicitly isolates anomalies without profiling normal instances. The key advantage of this approach is that it is based purely on the concept of isolation, without using any measure for density or distance. The isolation forest algorithm, for example, is able to detect 'inlying outliers', which are anomalies surrounded by normal data points. In addition, Liu et al. (2012) argue that the isolation algorithm can detect both clustered and scattered anomalies, whereas density and distance measures mainly detect scattered anomalies.

The isolation forest algorithm, like other tree ensemble methods, is built using decision trees. The algorithm isolates observations by randomly selecting a feature and then selecting a split value between the minimum and the maximum values of the selected feature. The number of splits required to isolate an observation is equivalent to the length of the path from the root node to the terminal node.[13] The algorithm therefore detects anomalies by the depth of the tree required to isolate each individual data point. This path length, i.e. the number of edges an observation must pass in the tree going from the root to the terminal node, represents the decision function. Anomalies are less frequent than regular observations and they generally lie further away from the normal observations in the feature space. They should therefore be identified closer to the root of the tree, using random partitioning.[14] A more regular data point requires more partitions to be identified than an anomaly. The algorithm therefore leads to significantly shorter path lengths for anomalies.

signals" of potential anomalies, as discussed in Staerman, Adjakossa, Mozharovskyi, Hofer, and Gupta (2023).

[11]See Liu, Ting, and Zhou (2008), Liu, Ting, and Zhou (2012), Mensi, Tax, and Bicego (2023), Tao, Peng, Zhao, Zhao, and Wang (2018), Palekar, Kharade, Zade, Ali, Kamble, and Ambatkar (2020), Jiang, Guo, and Yang (2021), Shao, Du, Yu, and Chen (2022) for more information on the isolation forest methodology. In a recent study, analyzing 33 unsupervised anomaly detection algorithms, Bouman, Bukhsh, and Heskes (2024) find isolation forest algorithms to be the adequate approach when dealing with global anomalies, that is anomalies regarding the entire dataset, in contrast to local anomalies in subsamples.

[12]See Liu et al. (2008) and Debener, Heinke, and Kriebel (2023) for more information.

[13]An isolation tree generally has internal nodes and external nodes, where nodes without child nodes are called external nodes. The internal node has an attribute q, a split value p of the attribute, and two child nodes (T_l, T_r). The attribute and split value are randomly chosen to split the data set until each node contains only one sample or a number of samples with identical values. The data set is sampled n times and the resulting different isolation trees constitute the isolation forest. Hence, the isolation forest algorithm partitions the data until each observation is isolated and assigned an anomaly score, which in turn corresponds to the path length. If the path length is less than a certain threshold, then the data point is considered an anomaly. For more information on the algorithm, see e.g. Staerman, Mozharovskyi, Clémençon, and d'Alché-Buc (2019), McKinnon, Carroll, McDonald, Koukoura, and Plumley (2021) and Zhang, Yao, Lv, and Wang (2022). Chen and Wu (2019) discuss the algorithm using a schematic diagram of an isolation tree.

[14]For more information on isolation partitioning strategies on a two-dimensional dataset, see Cao, Xiang, Zhang, Zhu, and Ting (2024).

There are numerous implementations of the isolation forest algorithm across various programming languages and libraries, including Python, R, MATLAB, Julia, C++ and Java. This paper primarily focuses on the use of the respective libraries in Python and R, as these are widely used programming languages in data analysis and machine learning. Both R and Python are excellent at integrating into existing data analysis pipelines. This facilitates the combination of anomaly detection with other analysis or preprocessing steps. The implementations in R and Python allow for the adjustment of various parameters, making it easy to tailor the method to our specific use case.

In Python, the isolation forest method is provided through the *sklearn.ensemble.IsolationForest* class, which is part of the *Scikit-Learn* library that offers a wide array of machine learning algorithms and tools. The parameters in the *Scikit-Learn* implementation of the isolation forest method include parameters such as the number of base estimators in the ensemble, the number of samples drawn to train each base estimator and the proportion of outliers in a dataset, which indicates the level of contamination.[15] Each of these parameters influences the algorithm's behavior.

For assessment of the results of the applied unsupervised learning method, we require several criteria in order to ensure the quality and usefulness of the results. First, it is evaluated if the identified patterns and structures in the data align with existing knowledge or theoretical assumptions, i.e. they should show a realistic behavior over time and past known sudden stops. Second, the results should show some stability, i.e. remain consistent across different runs of the same algorithm with slightly different parameters. Hence, the parameters were initially selected using a grid search, with the results of the traditional sudden stop analysis serving as a benchmark. Following this, a sensitivity analysis of the parameters was conducted. The findings demonstrate that the results from the *Scikit-Learn* isolation forest method exhibit a high degree of robustness. As a result, the parameters for the isolation forest method were set as follows: the number of isolation trees to be created in the ensemble was set to 500, the proportion of the dataset to be used for building each individual isolation tree was set to 0.1, the expected proportion of outliers in the dataset was set to the results of the traditional sudden stop analysis for the equity data and the bond data, respectively. These parameter settings were chosen to ensure a reliable anomaly detection in the dataset.

In R, the isolation forest method is typically implemented in the *isolation.forest* function from the *isotree* package. While the fundamental concepts behind the isolation forest method are the same in both implementations, *Scikit-Learn* and R, they slightly differ in parameters. There may be specific options that are not present in the Python implementation or vice versa, e.g. the contamination factor is not available in the *isotree* package. However, the R documentation provides guidance on selecting parameters in order to mimic the Python library.[16] A very interesting advantage of the R implementation of isolation forests is that it also offers an extended isolation forest method, which enhances the standard isolation forest approach in several ways. According to the R documentation, one enhancement in the extended isolation forest methods involves the careful selection of split points based on criteria related to standard deviations, density or other criteria such as

[15] For more details, see https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html.

[16] See https://www.rdocumentation.org/packages/isotree/versions/0.6.1-1/topics/isolation.forest.

ranges, kurtosis etc. In isolation forest methods, the split point is often chosen randomly within a certain range of the selected feature values. While this can work well in many cases, it may not always yield the most effective splits, particularly when the data has varying distributions over time, as is the case for our data. Another enhancement in the extended isolation forest methods is their ability to effectively manage high-dimensional data, which is particularly beneficial when analysing two-dimensional datasets, as is the case here, or even higher-dimensional data. In the basic isolation forest method splits are typically made along the axes of the feature space, resulting in axis-parallel splits. While this approach is straightforward, it can be limiting in high-dimensional spaces where the distribution of data points may not align neatly with the axes. In contrast, extended isolation forest methods can utilize randomly-chosen hyperplanes for making splits. By randomly selecting a hyperplane that is not restricted to being axis-aligned, the algorithm may capture more complex relationships within the data. This may lead to effectively isolate anomalies and achieve a more nuanced understanding of the data structure.

In terms of computational times, we found no significant difference between the two implementations. Even when analyzing the two-dimensional data, the computational times remained under 0.1 seconds, suggesting that the method is well-suited for use with higher-dimensional datasets (e.g. researching fund flows country wise) or for data that shows regime switching behavior requiring an analysis across multiple time periods rather than just a single one.

4 Univariate sudden stop analysis using isolation forests

In this section we apply univariate isolation forest analysis to equity and bond fund flows separately, using the ΔCM series for equity and bond flows. Since each partition in the isolation forest algorithm is randomly generated, the individual isolation trees are generated with different sets of partitions. We use 500 isolation trees to construct the isolation forest and the average path length. The path length of a specific data point is measured by the number of edges it traverses from the root node of the isolation tree until the traversal ends at an external node. The average path length is then used to compute the anomaly score. Data points are sorted according to their path length or anomaly scores, with anomalies being data points at the top of this list.

It is important to note, that the anomaly score does not explicitly differentiate between sudden stops and surges. Anomalies include both sudden stops and surges. Data points are generally classified as anomalies if the anomaly score is below the cutoff point of -0.5. The lower the anomaly score, the more anomalous is a data point. However, in this paper we will set different cutoff points for equity and bond flows. In order to compare the results of the isolation forest to those of the traditional sudden stop analysis, the percentage number of anomalies detected by the traditional method should be similar to the number identified by the isolation forest. The share of sudden stops and surges identified by the traditional method is 17.30% for equity and 15.62% for bond flows. Accordingly, the contamination rates for the isolation forest will be set to 0.17 for equity and 0.16 for bonds. These contamination rates lead to cutoff points for the classification of anomalies of -0.5284 for equity and -0.5622 for bonds. Figure 3 shows the absolute frequencies of the anomaly scores for equity and bond flows with the respective cutoff points of -0.5284 for equity and -0.5622 for bonds in the upper panel. The data points with anomaly

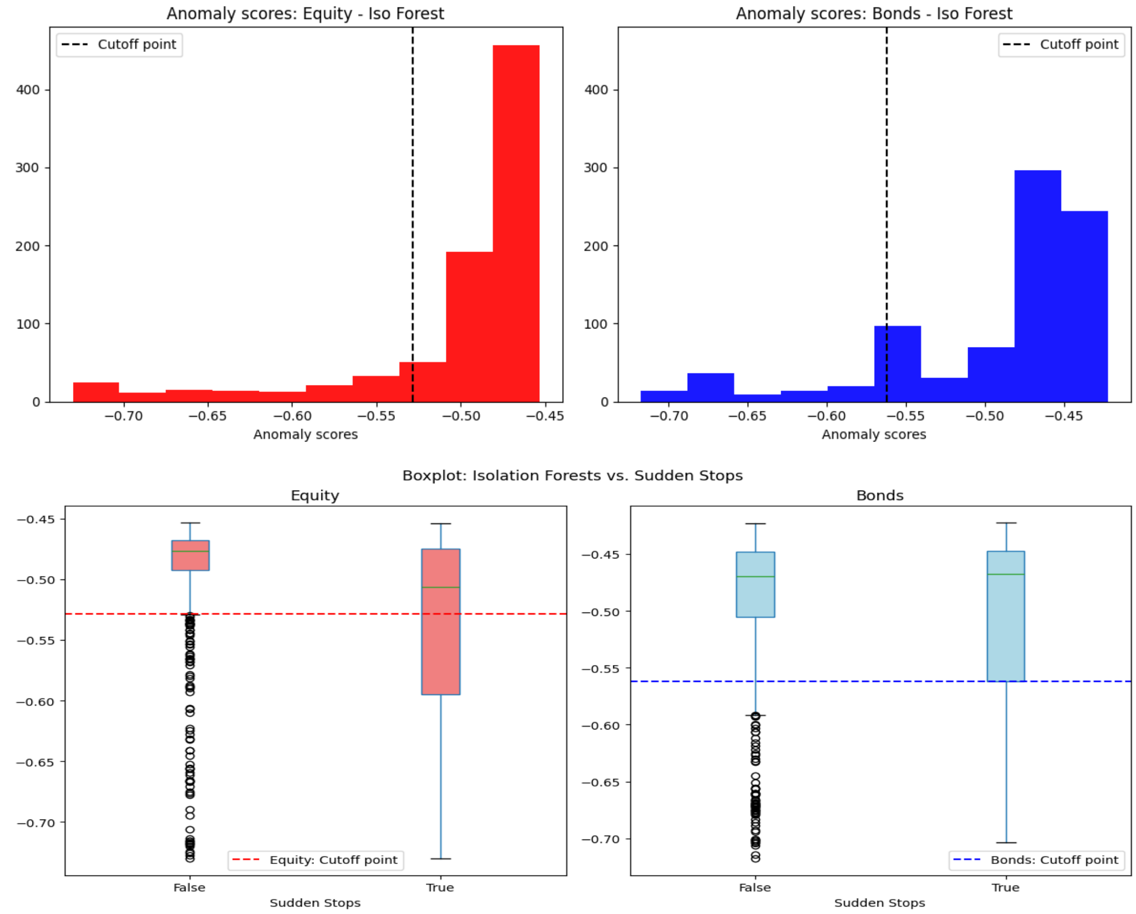

Figure 3: Univariate isolation forest: Cutoff points and traditional sudden stops

scores left of the cutoff points (i.e. with numbers smaller than the cutoff threshold) will be classified as anomalies, while the data points with scores right of the cutoff point will be classified as normal. Setting the cutoff point to the standard (i.e. 'auto') level of -0.5 would result in more data points being classified as anomalies. In contrast, shifting the cutoff line to the left would lead to the identification of less anomalies. Again, in this paper we choose to set the cutoff points corresponding to the results of the traditional sudden stop analysis in order to better be able to compare these different approaches.

The first comparison between the traditional method and the isolation forest is presented in the lower panel of Figure 3. Shown is the distribution of the anomaly scores for equity flows on the left and for bond flows on the right. In each case, on the right hand side, is the distribution for the events which the traditional method marks as anomalies (i.e. sudden stops and surges; marked in the figure as 'True'), and on the left hand side, the distribution for the cases which the traditional method identifies as normal (labelled as 'False' in the figure).[17] Most events, that are identified as normal (i.e. 'False') by

[17]It is important to stress that, while the results of the traditional analysis are labelled 'True' and 'False' in Figure 3, these are not 'True' or 'False' observations of sudden stops and surges, since sudden stops and surges are not observable, but generated using the traditional statistical method. Therefore it is not the task of this paper to replicate the findings of the traditional method, but to provide and discuss an alternative method for sudden stop analysis, using isolation forest. The discussion of contamination rates,

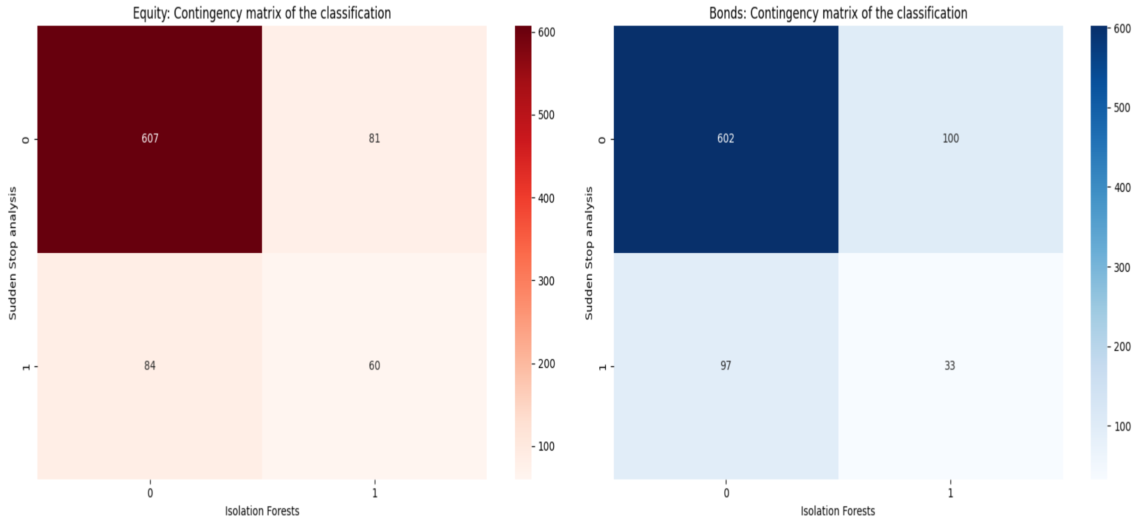

Figure 4: Univariate isolation forest: Contingency matrix of the classification

the traditional method, are classified as normal as well by the isolation forest. These are represented by the boxplots being mostly above the cutoff points on the respective left hand sides of the figures. However, there are a few data points classified as anomalies by the isolation forest, but not by the traditional method (i.e. observations below the cutoff points). The results are more mixed regarding the data points the traditional method classifies as 'True' anomalies. Here, the traditional method and the isolation forest yield significantly different results. A large proportion of the sudden stops and surges identified by the traditional method are not classified as anomalies by the isolation forest. For equity flows more than 50% and for bond flows about 75% of 'True' sudden stops and surges, according to the traditional method, are not classified as anomalies, as presented in the contingency matrix in Figure 4. Changing the contamination rates and therefore the cut-off points could lead to results more comparable with the traditional method. However, while lowering the cutoff points in Figure 3 (i.e. lowering the dotted red and blue lines in the figure) would lead to more sudden stops and surges being classified as anomalies, it would also result in more normal data points being labelled anomalies. Therefore we will continue the analysis with the set contamination rates at 0.17 for equity and 0.16 for bonds. Next, the development of anomaly scores and classifications over time will be discussed and compared to the traditional approach.

Figure 5 shows the development of the anomaly scores over time. The first result regarding Figure 5 is, that the occurrence and magnitude of anomalies increases over time. This is the case for both equity and bond markets. In addition, Figure 5 presents different decision thresholds for the classification of anomalies, given different contamination rates. The standard setting of the threshold as generally employed in the literature is at -0.5, labelled as 'auto' in Figure 5. The lower the contamination rate is set, the fewer data points are classified as anomalies. Hence, lowering the cutoff points to -0.5284 (i.e. contamination rate of 17%) for equity and -0.5622 (i.e. contamination rate of 16%) for

cutoff points, and contingency matrices simply serves to better compare the two different approaches to sudden stop analysis.

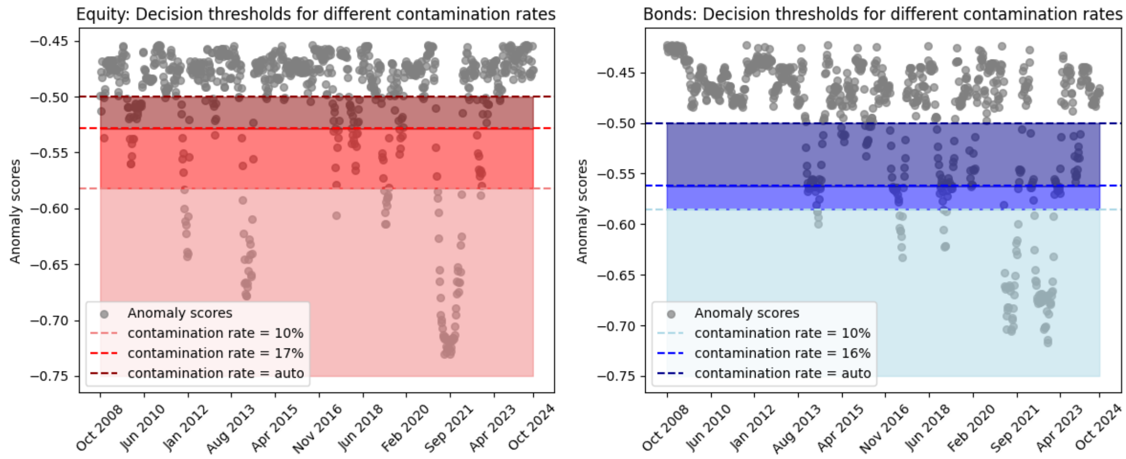

Figure 5: Univariate isolation forest: Anomalies and contamination rates

bonds leads to fewer anomalies being classified, when compared to the 'auto' setting. In addition, Figure 5 also shows the cutoff points for a contamination rate of 10%, as a further illustration of the relation between contamination rates and cutoff points. Interestingly, changes to the contamination rate lead to different reactions of the cutoff points for equity and bonds, respectively. The move from the 'auto' setting of the cutoff point to the respective contamination rate for equity (17%) and bonds (16%) is larger for bonds (cutoff point from −0.5 to −0.5622) than for equity (cutoff point from −0.5 to −0.5284). However, moving the contamination rate from 17% and 16% to 10% for both bonds and equity, as shown in Figure 5, leads to a stronger reaction of the cutoff point for equity (from −0.5284 to −0.5867%), when compared to bonds (from −0.5622 to −0.5853%).

Figure 6 presents the isolation forest anomaly classifications for equity and bond flows as well as the traditional sudden stop and surge classifications over time. The results as presented in Figure 6 are mixed, and generally correspond well to the discussion of Figures 3 and 4. However, in contrast to Figures 3 and 4, Figure 6 allows for the discussion of anomalous episodes in addition to anomalous data points. Hence, it provides information on potential patterns of anomalous data points in certain time periods.[18] Therefore, while the results for specific data points may point to the two alternative methods yielding significantly different results (see Figures 3 and 4), the analysis of anomalous episodes might point to more similarities. Regarding equity flows, the isolation forest algorithm identifies all the sudden stop and surge episodes, also identified by the traditional method. In addition to that, sudden stop [surge] episodes are identified in 2022 [2017/2018] by the isolation forest. For bond flows, there are larger differences between the traditional analysis and the isolation forest results. The sudden stops and surges identified by the traditional analysis in the time period before 2014 are not classified as anomalies by the isolation forest. In contrast, the sudden stop and surge around 2021/2022 are much more pronounced

[18] For more information on different types of anomalies, especially on anomalous points and patterns (e.g., contextual anomalies and collective anomalies), see for example Siddiqui, Krishna, and Kalra (2024) and Kaufman and Iaremenko (2022). There are applications of the isolation forest algorithm, where only patterns of a certain number of consecutive detections are considered an anomaly, while single detections are not classified as an anomaly, see e.g. McKinnon et al. (2021). However, in our approach, we will analyse anomalous points as well as patterns.

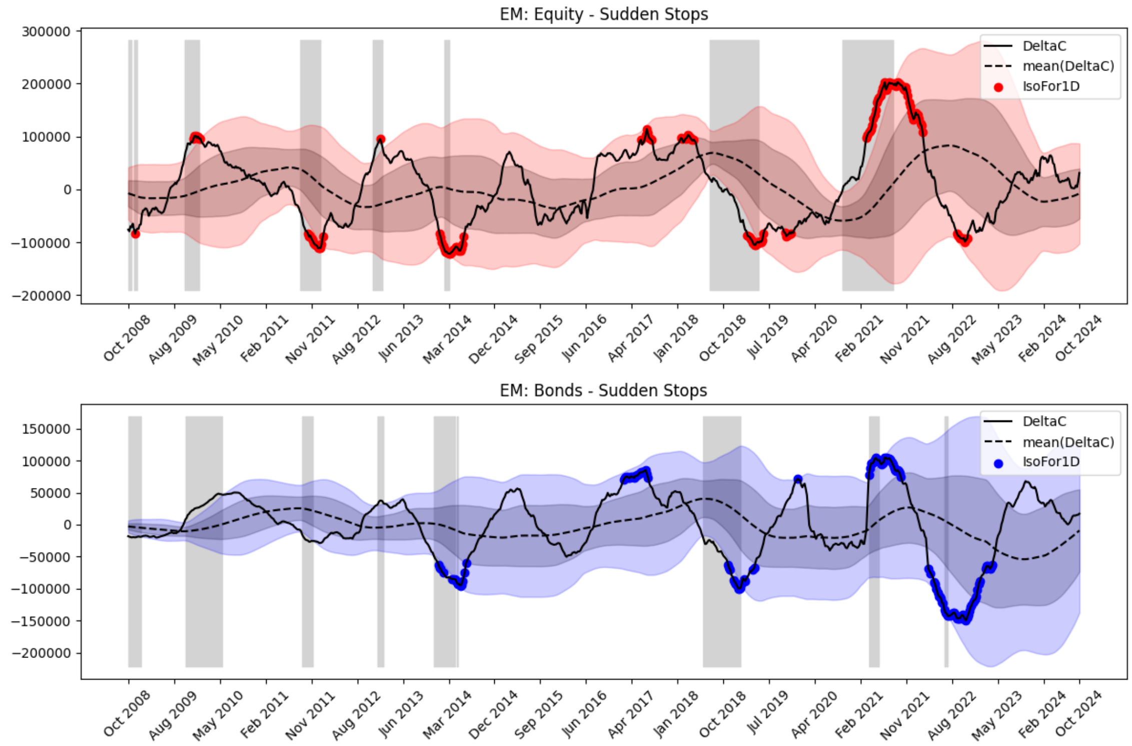

Figure 6: Univariate isolation forest: Anomalies and traditional sudden stops

when using the isolation forest methodology. Overall, however, the two approaches yield comparable results. Especially, episodes where the traditional sudden stop analysis points to normal flows are also classified as normal by the isolation forest. Comparing Figures 6 and 3, the discrepancy between the traditional results and the isolation forest results, as shown in the boxplots in Figure 3, becomes more clear. While the isolation forest classifies a large number of data points as anomalies where the traditional method does not find sudden stops or surges, the anomaly periods nevertheless often correspond to episodes where the traditional method also finds sudden stops or surges (albeit shorter episodes), or is close to identifying the data points as sudden stops or surges, i.e. close to the two standard deviation threshold. This can be seen, for example, for bond flows around 2021 and 2022, where both methods point to anomalies, with the difference, that the episode of classified anomalies is significantly longer when using the isolation forest. In addition, the isolation forest identifies a period of anomalous bond flows in 2017, where the traditional method is very close to the two standard deviation threshold, but does not pass this threshold. Summarising, it is important to emphasize, that even if the traditional method and the isolation forest yield different results for specific data points, as shown in the boxplots in Figure 3, the results are more homogeneous with regard to anomalous episodes, as presented in Figures 6.

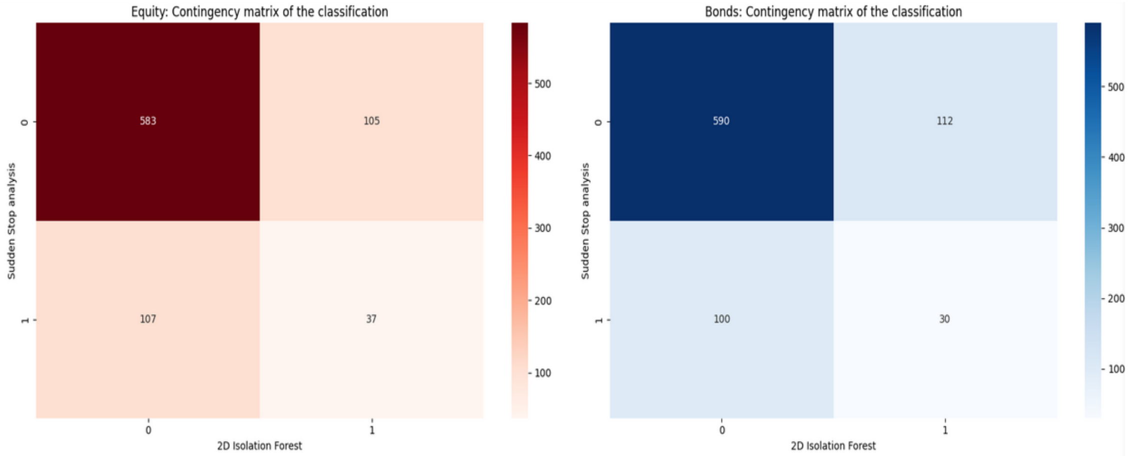

Figure 7: Bivariate isolation forests: Contingency matrix of the classification

5 Bivariate sudden stop analysis using isolation forests

In addition to univariate anomaly detection, we apply the isolation forest algorithm to the bivariate dataset of equity and bond flows to emerging markets. As pointed out by Kaufman and Iaremenko (2022), while a single feature in the dataset may not appear to be an outlier, the dataset as a whole can be anomalous. Analysing the collective anomalies of equity and bond flows, as the two key categories of portfolio flows, allows for a more comprehensive assessment of potential anomalies in portfolio flows. In the analysis, the bivariate isolation forests were implemented using the extended isolation forest algorithm to make use of its advantages in anomaly detection for multivariate data. Given that Python does not currently offer an implementation of extended isolation forests, the implementation of the extended isolation forests in R was used.[19]

Again, the traditional method and the bivariate isolation forest yield significantly different results, as shown in the contingency matrix in Figure 7. As in the univariate case, a large proportion of the sudden stops and surges detected by the traditional method are not classified as anomalies by the isolation forest. For equity flows, the differences are more significant than in the univariate case with 74% of traditional sudden stops and surges not classified as anomalies by the isolation forest. For bond flows, the results are comparable to the univariate isolation forest, with again about 77% of traditional sudden stops and surges not classified as anomalies by the isolation forest.

The left hand illustration in Figure 8 shows a scatterplot for equity and bond flows, with anomaly classifications of the bivariate isolation forest being presented in different colors: gray for normal flows and green for anomalies. There is a clear positive correlation between the fund flow series, with most observations either pointing to fund flow surges (i.e. inflows into both equity and bond funds, see the upper right corner of the figure) or to fund flow stops (i.e. outflows out of both equity and bond funds, see the lower left

[19] Several different models of extended isolation forests were tested, and based on the previously established quality criteria, the EIF model detailed in the R documentation yielded the best results. The following parameters were used: *ndim*=2, *sample_size*=512, *maxdepth*=7, *ntrees*=125, with *missing_action* set to 'fail', *coefs* defined as 'uniform' and *standardize_data* set to 'true'. A threshold of 0.6025 regarding the anomaly score was used to make the results comparable to the traditional sudden stop analysis.

corner of the figure). A minority of observations are related to the combinations of equity stops and bond surges or equity surges and bond stops, respectively. Interestingly, most anomalies are located in the 'fund flow surge' and 'fund flow stop' areas in Figure 8.

In addition, the illustration on the right hand side of Figure 8 shows the results of the traditional approach to sudden stop analysis, where data points are classified as anomalies if either the traditional univariate analysis of equity or the univariate analysis of bond flows detect a sudden stop or a surge. Comparing the results of the bivariate isolation forest to the results of the traditional method, as presented in Figure 8, offers some interesting findings. First, the anomalies identified by the bivariate isolation forest are on the edges of the data space. In contrast, a large proportion of sudden stops and surges detected by the traditional method are closer to the center of the data space. In addition, most of these data points are surrounded by other data points, and therefore not isolated. They can therefore most likely not be classified as 'inlying outliers' or as belonging to an anomalous pattern within the data space. Therefore, by visual inspection, it could be concluded, that a large fraction of anomalies detected by the traditional method are not anomalous but normal data points. In contrast, the anomalies identified by the isolation forest are clearly on the edges of the data space and can hence be classified as anomalies. Second, as mentioned earlier, the anomalies identified by the bivariate isolation forest are mostly in the upper right or lower left corner of the figure. Hence, they correspond to overall fund inflows (i.e. simultaneous inflows into equity and bond funds) or outflows (i.e. simultaneous outflows from equity and bond funds), and therefore to bivariate fund surges or bivariate fund sudden stops. A smaller number of identified anomalies represent a mixture of equity outflows and bond inflows (upper left corner in Figure 8) which could be labelled 'safe haven' or 'risk-off' flows. And, in the lower right corner of the figure a small number of anomalies represent a mixture of equity inflows and bond outflows, which could be interpreted as 'risk-on' flows from bond into equity markets. Again, the anomalies identified by the traditional method, as shown on the right hand side of Figure 8, cannot be interpreted in this way since they are largely in the center of the data space, and therefore most likely would be classified as normal data points by visual inspection.

As mentioned earlier, one key contribution of the isolation forest approach to sudden stop analysis, is that it provides a continuous anomaly score variable. Based on these anomaly scores, data points can be classified as anomalous or as normal. Figure 9 presents, on the left hand side, the results of the bivariate isolation forest (as already shown in Figure 8) plus decision boundaries for the anomaly classification. The right hand side of Figure 9 shows the anomaly score values for each individual data point.

Next, we want to use the findings from the bivariate investigation to get a better idea of the univariate time series representation, as shown in Figure 10. Figure 10 presents the results of the univariate traditional sudden stop analysis as well as the univariate and bivariate isolation forest anomaly classifications over time. Again, gray areas in the figure show the occurrence of a sudden stop or surge, according to the traditional method. The red and blue dots represent the anomalies in equity and bond flows as classified by the bivariate isolation forest, including the cases where the univariate isolation forest points to anomalies as well. The green dots show the data points classified as anomalies by the univariate but not by the bivariate isolation forest (see also Figure 6 for a comparison). While there is a large overlap between the classifications by the univariate and bivariate isolation forests, there is at least one clear distinction. Although the univariate isolation forest

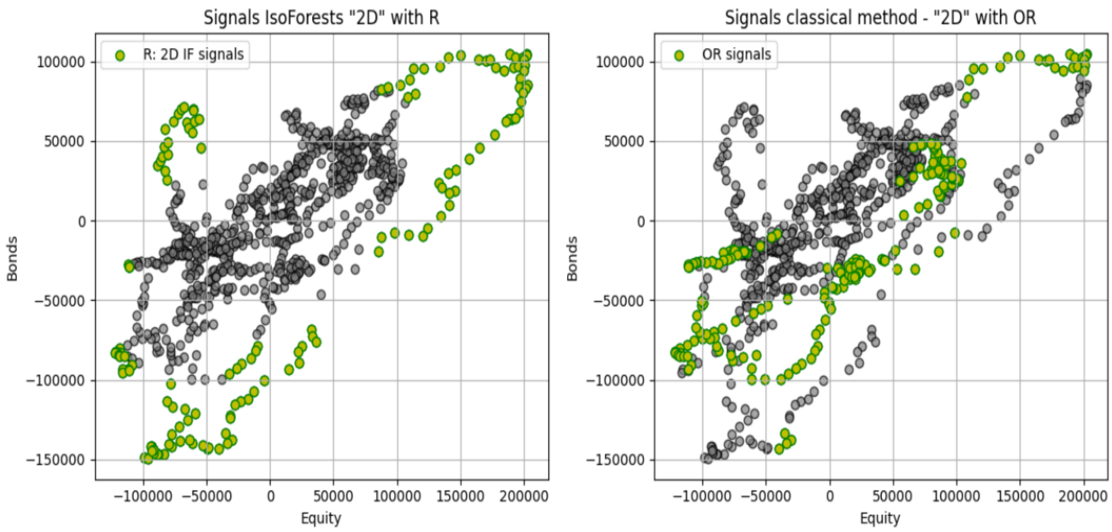

Figure 8: Bivariate isolation forest vs. traditional method: Anomalies for equity and bond flows

already classified the surge and sudden stop episodes around 2021 and 2022 as anomalies, these anomalous patterns are even more pronounced by the bivariate isolation forest. The results point to a long time period of volatile fund flows from about the beginning of 2021 to the end of 2022, with a small interruption of the anomaly classification around the turn of the year 2021/2022. Interestingly, the surge and stop episodes have about the same duration, with the volatility receding significantly after 2022, especially regarding equity flows. In and around this time frame fall several global and EME-specific crises. The key global crisis that rapidly spread across 213 countries[20], affecting both advanced and emerging economies, and that lasted a considerable time period was the COVID-19 pandemic.[21] During this global health crisis many countries around the world instituted lockdowns, travel restrictions and social distancing with profound consequences for global supply chains, production, and financial stability. Recent studies point to the increased transmission of stress and uncertainty between advanced and emerging markets during the pandemic period, potentially contributing to increased fund flow volatility.[22] In addition to this global crisis, there were several local crises in different emerging markets in this time period. The most important being the real estate crisis in China, potentially coupled with the severe COVID-lockdown in China, and the beginning of Russia's invasion of Ukraine, leading to severe disruptions in global supply chains, turmoil in energy markets, and the implementation of significant financial sanctions.

In addition to these global and local crises during this time period, financial markets in advanced economies underwent a significant turnaround from a low yield environment with quantitative easing and official asset purchase programs to a more restrictive monetary policy and significant increases in interest rates in many countries. The turnaround from a

[20] See e.g. Harjoto and Rossi (2023).

[21] The European Systemic Risk Board (ESRB) categorizes the COVID-19 pandemic as a systemic crisis, starting in March 2020 and ending in April 2021, see e.g. the ESRB homepage at *www.esrb.europa.eu/pub/financial − crises/html/index.de.html*.

[22] See e.g. Belaid, Amar, Goutte, and Guesmi (2023) and Candelon and Moura (2023).

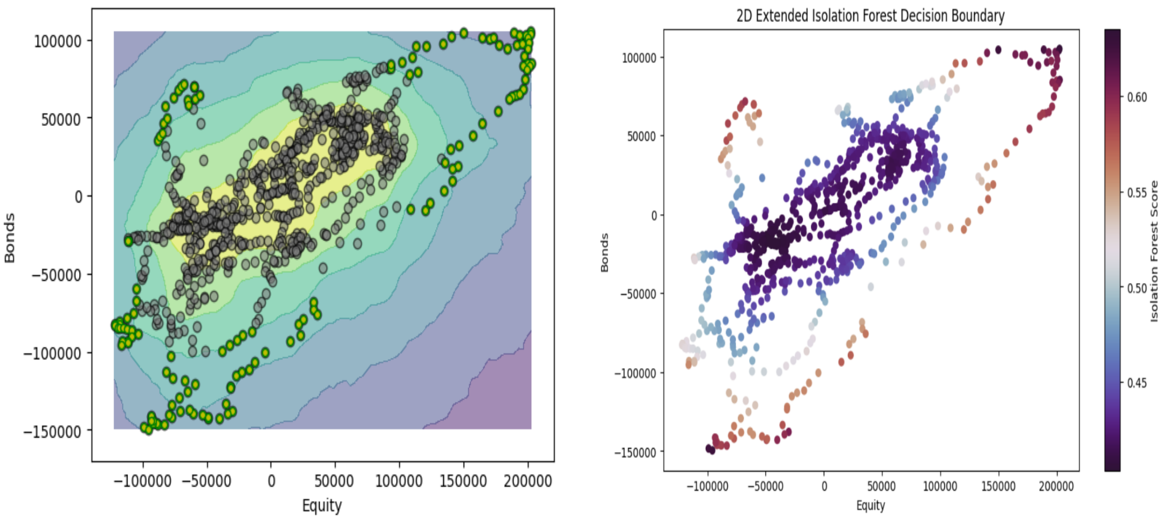

Figure 9: Bivariate isolation forest: Anomaly classifications and scores

low yield environment to an increase of interest rates in advanced economies took place in the spring and summer of 2022.[23] Interestingly, this time period of monetary tightening in advanced economies corresponds well with the sudden stop period identified by the bivariate isolation forest. In contrast, the surge period during 2021, as classified by the isolation forest, could be related to the low yield environment before the interest reversal in 2022.[24] This low yield environment in advanced economies which added to a 'search for yield'[25] among investors potentially contributed to a surge of inflows into emerging markets. Considering, the significant upheaval in global financial markets in this time period, it is astonishing, that the traditional method for the analysis of sudden stops only point to two brief episodes of anomalies in bond market and does not identify the sudden stop around 2022 in equity markets. Given the recent history of financial markets and crises, the isolation forest methodology appears to be the better alternative for the analysis of anomalous patterns in portfolio flows. In general, univariate and bivariate anomaly detection using machine learning techniques can play an important part and lead to a better understanding of sudden stops and surges.

[23] The Federal Reserve started increasing interest rates in March 2022 and the ECB in July 2022.

[24] Interestingly, the Bank for International Settlement identifies a phase from 2009 to 2021 (after the Global Financial Crisis), which is to some degree characterized as being dominated by a 'shift to bond credit' (i.e. bank retrenchment), in particular to borrowers in emerging market economies. For more information, see Hardy and von Peter (2023).

[25] The 'search for yield' effect refers to the behaviour of investors, searching for higher returns in other asset classes when yields on government bonds decrease or are at a low level, see e.g. Arora and Kashiramka (2023). See also Boonman (2023) on the 'search for yield' effect and the surge in portfolio flows to emerging markets after the global financial crisis.

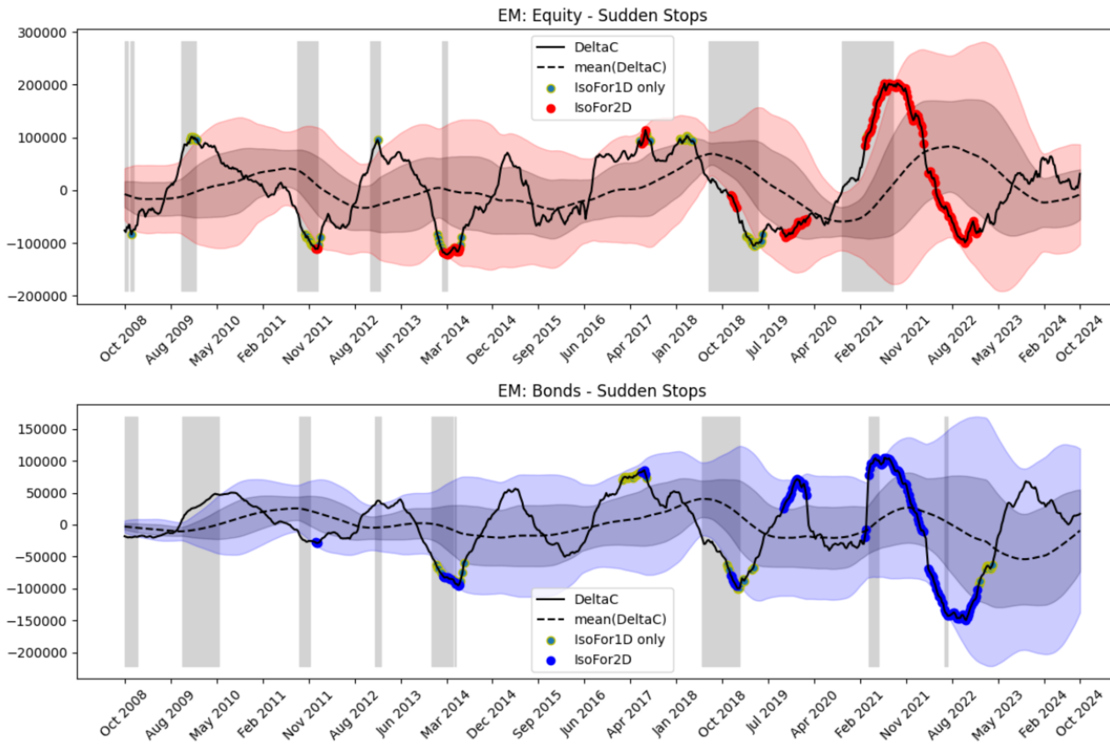

Figure 10: Traditional sudden stops and bivariate isolation forests: Anomaly classifications

6 Conclusion

This study introduces a machine learning methodology, and in particular the isolation forest algorithm for anomaly detection, to sudden stop analysis. First, weekly data on equity and bond flows to emerging markets are analysed using the univariate isolation forest, and the results are compared to those from traditional sudden stop analysis. Second, bivariate isolation forest is employed for anomaly detection of equity and bond flows combined. The results point to portfolio flow anomalies mostly being related to fund flow stops (i.e. stops to both equity and bond flows) or surges (i.e. surges in both equity and bond flows). Both the frequency of occurrence and the intensity of fund flow anomalies have increased significantly in recent years.

Especially given the recent history of financial markets and crises since the outbreak of the COVID-19 pandemic, as well as the significant interest rate reversal in advanced economies in this time period, the isolation forest methodology appears to yield better results in classifying anomalies than the traditional approach. The bivariate approach to anomaly detection is better able to identify anomalous episodes of volatility and turmoil, where both equity and bond markets are simultaneously affected. In general, univariate and bivariate anomaly detection using machine learning techniques can play an important part and lead to a better understanding of sudden stops and surges.

Introducing machine learning techniques to sudden stop analysis adds to the traditional approach by bringing new possibilities and concepts of anomaly detection into this research area. The key contributions of this paper to the sudden stop literature are threefold. First, using isolation forest, we can analyse the occurrence and the intensity of anomalies. While traditional sudden stop analysis typically focusses on the occurrence of sudden stops and surges, and neglects the intensity by only differentiating between normal flows, sudden stops and surges, we provide the anomaly score of fund flows as a continuous time series. This allows for a more detailed and nuanced analysis of anomalous capital flows. Second, we perform a bivariate sudden stop analysis for both equity and bond flows to emerging markets simultaneously. This allows for a more comprehensive analysis of the two key components of portfolio flows to emerging markets. Third, anomalies are identified by using the depth of the isolation tree required to isolate each individual data point. Hence, sudden stops and surges are classified without exogenously applying standard deviation bands and without relying on the empirical distribution.

Especially the bivariate and multivariate analysis of anomalies in capital flows is an important topic for future research. A simultaneous sudden stop in both equity and bond markets poses more challenges to market functioning and financial stability than the occurrence of a sudden stop in one of these markets. The multivariate approach promises new insights and a better understanding of overall portfolio flow dynamics. A second important topic for future research is to try and better align traditional sudden stop analysis with modern anomaly detection methods, for example using machine learning. Especially the concept of isolation employed in this paper, versus the more established concepts of density and distance, can lead to interesting new insights into different types and approaches to anomalies, and therefore to a better understanding of sudden stops and surges.

References

Agosin, M. R. and F. Huaita (2011). Capital flows to emerging economies: Minsky in the tropics. *Cambridge Journal of Economics 35*, 663–683.

Agosin, M. R. and F. Huaita (2012). Overreaction in capital flows to emerging markets: Booms and sudden stops. *Journal of International Money and Finance 31*, 1140–1155.

Arora, D. and S. Kashiramka (2023). What drives the growth of shadow banks? Evidence from emerging markets. *Emerging Markets Review 54*, 1–21.

Belaid, F., A. B. Amar, S. Goutte, and K. Guesmi (2023). Emerging and advanced economies markets behaviour during the COVID-19 crisis era. *International Journal of Finance and Economics 28*, 1563–1581.

Beutel, J., L. Emter, N. Metiu, E. Prieto, and Y. Schüler (2025). The global financial cycle and macroeconomic tail risks. *Journal of International Money and Finance 156*, 1–24.

Boonman, T. M. (2023). Portfolio capital flows before and after the Global Financial Crisis. *Economic Modelling 127*, 1–17.

Bouman, R., Z. Bukhsh, and T. Heskes (2024). Unsupervised anomaly detection algorithms on real-world data: How many do we need? *Journal of Machine Learning Research 25*, 1–34.

Broto, C., J. Díaz-Cassou, and A. Erce (2011). Measuring and explaining the volatility of capital flows to emerging countries. *Journal of Banking and Finance 35*, 1941–1953.

Byrne, J. P. and N. Fiess (2016). International capital flows to emerging markets: National and global determinants. *Journal of International Money and Finance 61*, 82–100.

Calvo, G. A. (1998). Capital flows and capital-market crises: The simple economics of sudden stops. *Journal of Applied Economics 1*, 35–54.

Calvo, G. A., A. Izquierdo, and E. Talvi (2006). Sudden stops and phoenix miracles in emerging markets. *American Economic Review 96*, 405–410.

Candelon, B. and R. Moura (2023). Sovereign yield curves and the COVID-19 in emerging markets. *Economic Modelling 127*, 1–14.

Cao, Y., H. Xiang, H. Zhang, Y. Zhu, and K. M. Ting (2024). Anomaly detection based on isolation mechanisms: A survey. *arXiv:2403.10802*, 1–10.

Chen, Y. and W. Wu (2019). Isolation forest as an alternative data-driven mineral prospectivity mapping method with a higher data-processing efficiency. *Natural Resources Research 28*, 31–46.

Committee on the Global Financial System (2021). Changing patterns of capital flows. *CGFS Papers 66*.

Debener, J., V. Heinke, and J. Kriebel (2023). Detecting insurance fraud using supervised and unsupervised machine learning. *Journal of Risk and Insurance 90*, 743–768.

Efremidze, L., S. Kim, O. Sula, and T. D. Willett (2017). The relationships among capital flow surges, reversals and sudden stops. *Journal of Financial Economic Policy 9*, 393–413.

Eichengreen, B. and P. Gupta (2016). Managing sudden stops. *World Bank Group, Policy Research Working Paper 7639*.

Feng, Y., W. Cai, H. Yue, J. Xu, Y. Lin, J. Chen, and Z. Hu (2022). An improved x-means and isolation forest based methodology for network traffic anomaly detection. *PLoS ONE 17*, 1–18.

Forbes, K. J. and F. E. Warnock (2012). Capital flow waves: Surges, stops, flight, and retrenchment. *Journal of International Economics 88*, 235–251.

Forbes, K. J. and F. E. Warnock (2021). Capital flow waves-or ripples? extreme capital flow movements since the crisis. *Journal of International Money and Finance 116*, 1–24.

Guidotti, P. E., F. Sturzenegger, A. Villar, J. de Gregorio, and I. Goldfajn (2004). On the consequences of sudden stops. *Economía 4*, 171–214.

Hardy, B. and G. von Peter (2023). Global liquidity: a new phase? *BIS Quarterly Review, December*, 21–31.

Harjoto, M. A. and F. Rossi (2023). Market reaction to the COVID-19 pandemic: Evidence from emerging markets. *International Journal of Emerging Markets 18*, 173–199.

Hutchinson, M. M. and I. Noy (2006). Sudden stops and the Mexican wave: Currency crises, capital flow reversals and output loss in emerging markets. *Journal of Development Economics 79*, 225–248.

Hutchinson, M. M., I. Noy, and L. Wang (2010). Fiscal and monetary policies and the cost of sudden stops. *Journal of International Money and Finance 29*, 973–987.

Jiang, L., Y.-C. Guo, and J.-C. Yang (2021). Detecting anomalous quartic gauge couplings using the isolation forest machine learning algorithm. *Physical Review D 104*, 1–10.

Kaufman, E. and A. Iaremenko (2022). Anomaly detection for fraud in cryptocurrency time series. *arXiv:2207.11466*, 1–17.

Li, S., J. de Haan, and B. Scholtens (2018). Surges of international fund flows. *Journal of International Money and Finance 82*, 97–119.

Li, S., J. de Haan, and B. Scholtens (2019). Sudden stops of international fund flows: Occurrence and magnitude. *Review of International Economics 27*, 468–497.

Liu, F. T., K. M. Ting, and Z.-H. Zhou (2008). Isolation forest. *Eighth IEEE International Conference on Data Mining*.

Liu, F. T., K. M. Ting, and Z.-H. Zhou (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data 6*, 1–39.

McKinnon, C., J. Carroll, A. McDonald, S. Koukoura, and C. Plumley (2021). Investigation of isolation forest for wind turbine pitch system condition monitoring using SCADA data. *Energies 14*, 1–20.

Mendes, T., P. J. S. Cardoso, J. Monteiro, and J. Raposo (2023). Anomaly detection of consumption in hotel units: A case study comparing isolation forest and variational autoencoder algorithms. *Applied Sciences 13*, 1–25.

Mensi, A., D. M. J. Tax, and M. Bicego (2023). Detecting outliers from pairwise proximities: Proximity isolation forests. *Pattern Recognition 138*, 1–12.

Palekar, V., S. Kharade, H. Zade, S. Ali, K. Kamble, and S. Ambatkar (2020). Credit card fraud detection using isolation forest. *International Research Journal of Engineering and Technology 7*, 3538–3543.

Proaño, C. R., L. Q. Virla, and T. Strohsal (2025). How strong is the link between the global financial cycle and national macro-financial dynamics? A wavelet analysis. *Journal of International Money and Finance 159*, 1–15.

Reinhart, C. M. and V. R. Reinhart (2008). Capital flow bonanzas: An encompassing view of the past and present. *NBER Working Paper Series 14321*.

Rothenberg, A. D. and F. E. Warnock (2011). Sudden flight and true sudden stops. *Review of International Economics 19*, 509–524.

Shao, C., X. Du, J. Yu, and J. Chen (2022). Cluster-based improved isolation forest. *Entropy 24*, 1–17.

Siddiqui, M. A., C. R. Krishna, and M. Kalra (2024). Anomaly detection framework for IoT-enabled appliances using machine learning. *Cluster Computing 27*, 9811–9835.

Staerman, G., E. Adjakossa, P. Mozharovskyi, V. Hofer, and J. S. Gupta (2023). Functional anomaly detection: A benchmark study. *International Journal of Data Science and Analytics 16*, 101–117.

Staerman, G., P. Mozharovskyi, S. Clémençon, and F. d'Alché-Buc (2019). Functional isolation forest. *Proceedings of Machine Learning Research 101*, 332–347.

Sula, O. (2010). Surges and sudden stops of capital flows to emerging markets. *Open Economies Review 21*, 589–605.

Tao, X., Y. Peng, F. Zhao, P. Zhao, and Y. Wang (2018). A parallel algorithm for network traffic anomaly detection based on isolation forest. *International Journal of Distributed Sensor Networks 14*, 1–11.

Tillmann, P. (2013). Capital inflows and asset prices: Evidence from emerging Asia. *Journal of Banking and Finance 37*, 717–729.

Zhang, X., Y. Yao, C. Lv, and T. Wang (2022). Anomaly credit data detection based on enhanced isolation forest. *International Journal of Advanced Manufacturing Technology 122*, 185–192.