

Koepke, Robin; Paetzold, Simon

**Article — Published Version**

## Capital flow data—A guide for empirical analysis and real-time tracking

International Journal of Finance & Economics

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Koepke, Robin; Paetzold, Simon (2022) : Capital flow data—A guide for empirical analysis and real-time tracking, International Journal of Finance & Economics, ISSN 1099-1158, John Wiley & Sons, Ltd., Chichester, UK, Vol. 29, Iss. 1, pp. 311-331, <https://doi.org/10.1002/ijfe.2687>

This Version is available at:

<https://hdl.handle.net/10419/288216>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

## RESEARCH ARTICLE

# Capital flow data—A guide for empirical analysis and real-time tracking

Robin Koepke<sup>1</sup>  | Simon Paetzold<sup>2</sup> 

<sup>1</sup>International Monetary Fund,  
Washington, District of Columbia, USA

<sup>2</sup>Goethe-Universität Frankfurt am Main,  
Frankfurt, Germany

## Correspondence

Simon Paetzold, Goethe-Universität  
Frankfurt am Main, Theodor-W.-Adorno-  
Platz 4, 60323 Frankfurt, Germany.  
Email: [simon.paetzold@stud.uni-frankfurt.de](mailto:simon.paetzold@stud.uni-frankfurt.de)

## Abstract

This paper provides an analytical overview of the most widely used capital flow datasets. The paper is written as a guide for academics who embark on empirical research projects and for policymakers who need timely information on capital flow developments to inform their decisions. We address common misconceptions about capital flow data and discuss differences between high-frequency proxies for portfolio flows. In a nowcasting “horse race” we show that high-frequency proxies have significant predictive content for portfolio flows from the balance of payments (BoP). We also construct a new dataset for academic use, consisting of monthly portfolio flows broadly consistent with BoP data.

## KEYWORDS

balance of payments, coincident indicators, fund flows, portfolio flows, capital flow guide, real-time tracking

## 1 | INTRODUCTION

In early 2020, the COVID-19 pandemic triggered one of the sharpest reversals in portfolio flows to emerging markets (EMs) on record. Timely capital flow proxies served as early warning indicators, alerting policymakers to the severity of the shock. While traditional quarterly balance of payments (BoP) data would only become available several months after 2020:Q1, proxy data on monthly, weekly and even daily flows from Emerging Portfolio Fund Research (EPFR) and the Institute of International Finance (IIF) showed as early as March 2020 that the reversal was extremely severe in speed and magnitude (IIF, 2020; IMF, 2020).

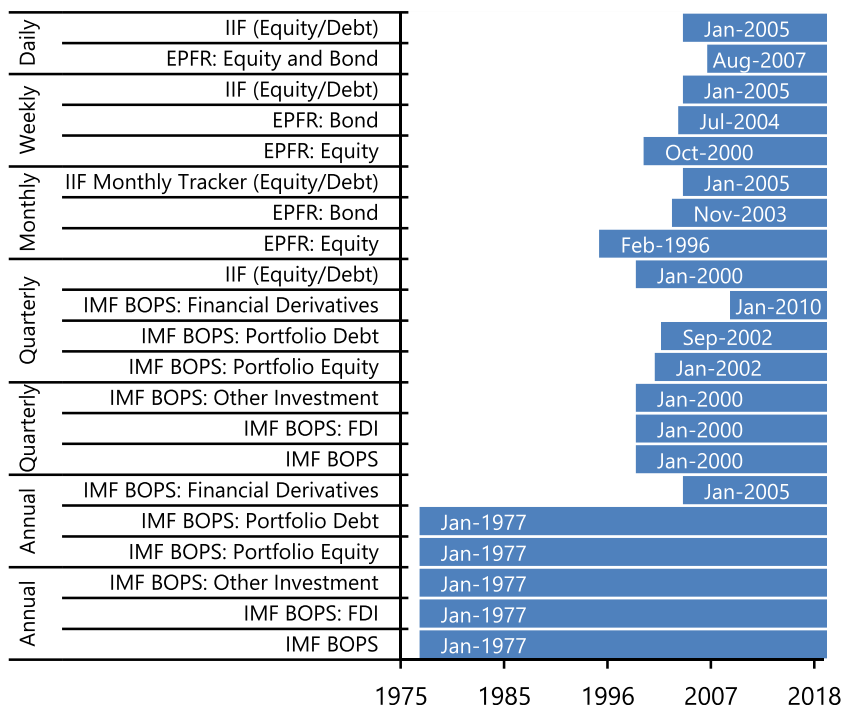
Data availability of international capital flows has improved dramatically over the past two decades, particularly of high-frequency proxies for portfolio flows (Figure 1). New data sources provide valuable

opportunities for advancing academic research and informing policy decisions in real time. However, these data sources also pose new challenges to researchers and policymakers. There are numerous conceptual and measurement issues surrounding these datasets, which are compounded by long-standing misconceptions about capital flow data in general.

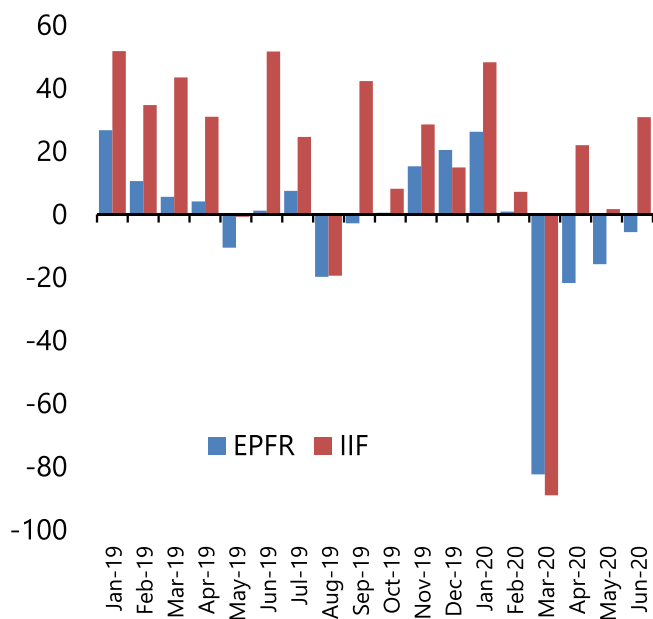
Many of these conceptual and measurement issues are reflected in the differences between EPFR and IIF proxies for portfolio flows to EMs. While both datasets generally signal turning points correctly, they often provide widely different estimates of EM portfolio flows (Figure 2). For example, from 2010:Q1–2019:Q2 average quarterly fund flows to EMs as a group were about \$11 billion according to EPFR fund flow data, compared to \$68 billion for the IIF Portfolio Flows Tracker and \$71 billion for IMF BoP portfolio flow data.<sup>1</sup>

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *International Journal of Finance & Economics* published by John Wiley & Sons Ltd.



**FIGURE 1** Start-date of common portfolio flow data sources. For IMF BOPS the start date is determined by the first time more than 50% of current EMDEs report their data. *Source:* IMF BOPS, EPFR, IIF [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 2** IIF portfolio flows versus EPFR fund flows. Monthly data (IIF data released in July 2020), USD billion. *Source:* IIF and EPFR [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

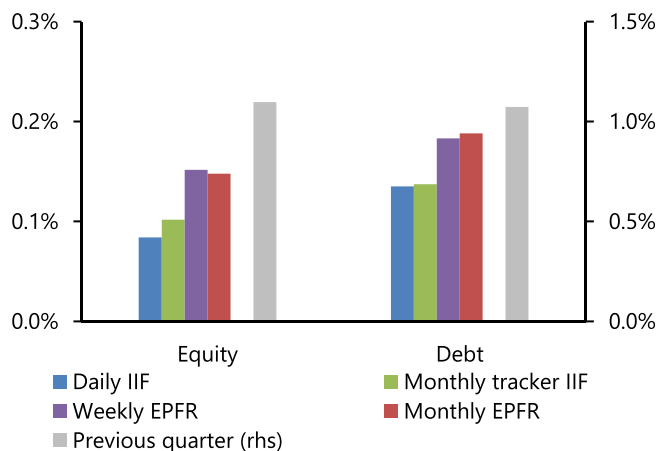
These issues are exacerbated by the fact that private sector data providers often make available only limited documentation of the methodological principles underlying the construction of their datasets. In particular, data providers (and the empirical literature) devote little attention to discussing how these principles differ from standard balance of payments accounting conventions.<sup>2</sup> Moreover, only very limited work has been done comparing and

contrasting different datasets in order to assess what data are most appropriate to answer which research questions.

The objective of this paper is to provide clarity on these issues. The paper is written as a guide for academics who embark on empirical research projects and for policy-makers and market participants who need timely information on capital flow developments to inform their decisions. From a policy perspective, the data discussed in this paper are an important ingredient for understanding the policy challenges created by international capital flows, documented in an extensive literature (some surveys include Dooley, 1996; BIS, 2009; Milne, 2014).

This paper makes three main contributions. First, we provide an overview of the most widely used datasets available today, along with a discussion of the strengths and weaknesses of various data sources. Special attention is given to high frequency measures of financial transactions like portfolio flow and fund flow data. We also address common misconceptions about capital flow data in the context of the BoP accounting framework and discuss recent research highlighting the limitations of that framework (e.g., Coppola et al., 2020; Lane & Milesi-Ferretti, 2018).

Second, we address a gap in the availability of portfolio flow data for academic research. Our meta-study shows that academic research has made little use of BoP-consistent portfolio flow data at high data frequencies. This is in part due to subscription requirements and issues with the construction of country samples. We make a contribution towards closing the gap in data availability by providing a free online dataset with



**FIGURE 3** Equity and debt, RMSFE in % of GDP. Using data released after 1.5 months in a quarter with data from 2005-2019. Using all countries for which equity and debt flows are available from all data sources. *Source:* Authors' calculations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

monthly emerging market portfolio flow data, specifically geared towards academic use.

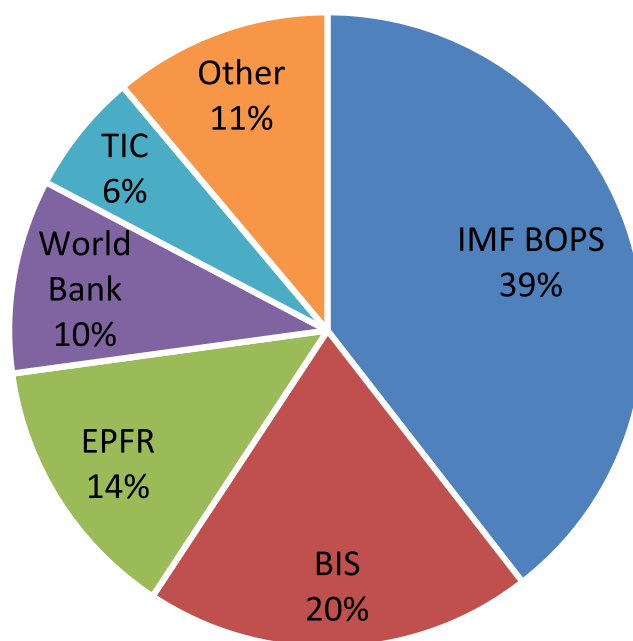
Third, we provide a quantitative assessment of how well widely used portfolio flow proxies track portfolio flow developments in real-time. Results of a nowcasting “horse race” suggest that IIF and EPFR data have significant predictive content for BoP-based portfolio flows, reducing forecast errors by 80%–90% relative to an autoregressive model (Figure 3). Portfolio flow proxies at the daily and weekly data frequencies outperform monthly data in the first half of the current quarter, and IIF data generally outperform EPFR data.

The rest of this paper is structured as follows. Section 2 provides an overview of capital flow data sources and addresses common misconceptions. Section 3 introduces a new monthly portfolio flow data set and Section 4 assesses the predictive content of several portfolio flow proxies. Section 5 concludes the paper.

## 2 | OVERVIEW OF CAPITAL FLOW DATA

This section provides a concise overview of the most commonly used capital flow data sources. We construct two sets of metadata to assess what data sources academics and policymakers have used the most. The first dataset represents the use of capital flow data in the empirical literature (“academic dataset”), while the second dataset represents the use of timely portfolio flow proxies by policymakers (“policy dataset”).<sup>3</sup>

In the 88 studies that constitute the academic dataset, IMF balance of payments data are the most common data



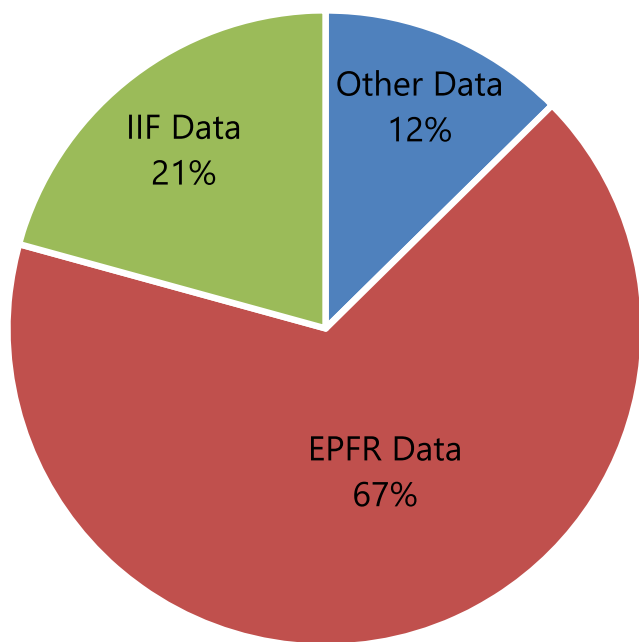
**FIGURE 4** Data sources in academic dataset. The sample comprises 88 papers published since 1993. If a data source was used less than 5 times it is grouped under “Other.” Seven studies did not use capital flow data or the data source was ambiguous (not included in this figure). *Source:* Authors' calculations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

source (used in 39% of all studies), followed by BIS data (20%) and EPFR data (14%; Figure 4). Less commonly used data sources include the World Bank’s World Development Indicators and the U.S. Treasury International Capital (TIC) data.

In the 111 studies that constitute the policy dataset, high-frequency portfolio flow proxies are used in about 50% of studies since 2010 (and about 60% of those published in 2017–2019). EPFR accounts for about two thirds of those studies (and about 50% in 2017–2019), followed by IIF data (21% in the full sample, 34% in 2017–2019; Figure 5). The remaining 12% of studies make use of other high-frequency data sources, mostly national or regional stock exchange data.

It is worth noting that data sources used in academic studies have shifted towards higher frequencies over the course of the last 15 years (Figure 6). In the 15 years before the Global Financial Crisis, about half of 88 the studies in our sample used annual data, while that share was only a quarter in the years since the GFC.<sup>4</sup> Consistent with the shift towards higher frequency data, studies have also tended to make greater use of data on (fast-moving) portfolio and bank-related flows rather than (slow-moving) FDI (Figure 7).

Beyond the data sources shown in Figures 4 and 5, there are a number of additional useful capital flow data



**FIGURE 5** High frequency capital flow data in policy dataset, 2010–2020. The chart shows data usage in reports by G20 EMs, the IMF, World Bank, and BIS since 2010. “Other Data” includes national stock exchange and Bloomberg data. *Source:* Authors’ calculations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

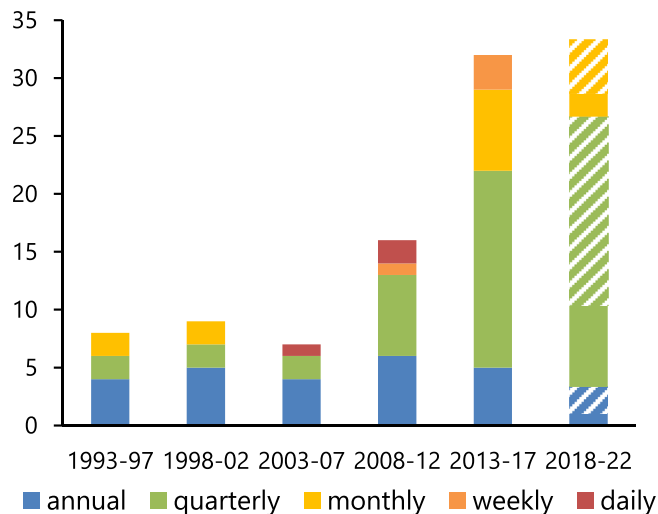
sources that deserve attention even though they have seen little or no use in the studies included in our academic and policy datasets. Some of these data sources are relatively new and likely to receive future attention in the literature, notably “big data” initiatives such as SWIFT’s data on transactions facilitated by banks. Another area where data coverage has expanded significantly in recent years is bilateral capital flow data (“from-whom-to-whom” data), which are discussed in Annex I (Data S1). Additional notable datasets include the BIS debt securities statistics, the ECB’s Securities Holdings Statistics, Morningstar’s data on investment funds, UNCTAD data on FDI, and Bureau van Dijk’s Zephyr on M&A and equity investments.

## 2.1 | Key capital flow data sources

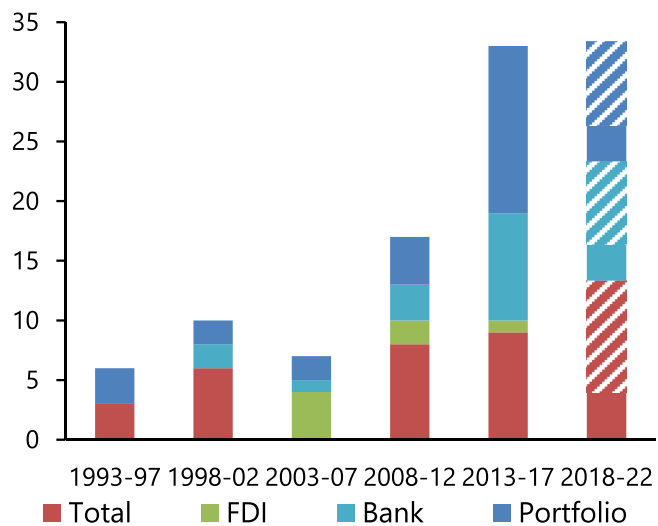
Table 1 presents detailed information on data characteristics as well as advantages and caveats relating to most commonly used data sources among scholars and practitioners.

### 2.1.1 | IMF balance of payments statistics

Capital flow data as reported in the IMF’s balance of payment statistics (BOPS) provide the most comprehensive



**FIGURE 6** Data frequency in meta-dataset. Shaded bars show extrapolation for future years by scaling the number of papers written from 2018 until 2019H1 to the entire five-year period. Totals per time period are not equal in both charts as information for capital component type or data frequency are unavailable or ambiguous for a limited number of papers. *Source:* Authors’ calculations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 7** Capital flow type in meta-dataset. Shaded bars show extrapolation for future years by scaling the number of papers written from 2018 until 2019H1 to the entire five-year period. Totals per time period are not equal in both charts as information for capital component type or data frequency are unavailable or ambiguous for a limited number of papers. *Source:* Authors’ calculations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

country coverage and detailed methodology, reflecting its central role in defining standards for the compilation of capital flow statistics. The data are reported to the IMF

TABLE 1 Overview of capital flow data sources

Data source	Data frequency	Typical release lag	Country coverage	Scope	Advantages	Caveats	Public (P) or subscription-based (S)	Start date
International Monetary Fund (IMF): Balance of Payments Statistics (BOPS)	Quarterly  Annual	2–4 months	Almost all EMS	All major capital flow components	Comprehensive coverage of cross-border transactions; well-defined and established methodology	Statistical break in the 2005–2008 period due to shift from BPM5 to BPM6 (the IMF BOPS adjusts and publishes historical data in the BPM6 format)	P	Jan-2002 (equity); Jul-2002 (debt) 1/1977 (equity and debt) 1/
Bank of International Settlements (BIS): Locational Banking Statistics (LBS)	Quarterly	4 months	Based on counterparty reporting for up to 200 countries/jurisdictions; around 95% of cross-border banking activity captured	Cross-border banking activity: Loans and deposits, debt securities, other instruments	Comprehensive coverage of cross-border banking activity; currency composition and counterpart sector available; data on both immediate and ultimate counterparty basis	No direct mapping between BIS data and standard BoP components of capital flows; flows are constructed based on stock data, adjusted for FX valuation effects	P	Oct-1977
BIS Consolidated Banking Statistics (CBS)	Quarterly (since 1999Q4); bi-annual (since 1983Q4)		Based on counterparty reporting for up to 200 countries/jurisdictions		Additional data on guarantor (since early 2005) and consolidated basis	No data on constructed flows		Jan-1984
Institute of International Finance (IIF): Portfolio Flows Trackers and Capital Flows Databases	Daily/weekly	1–3 days	15 countries for equity flows, 9 for debt flows, 2 for equity and debt combined	Portfolio equity and debt flows (only non-resident flows to EMS)	Good proxy for quarterly BoP data, especially for equity flows; high data frequencies and short release lag	For some countries, debt flows cover only local currency and/or sovereign bonds; does not capture bond purchases in the primary market and maturing bonds	S	Jan-2005 (equity and debt)

(Continues)

TABLE 1 (Continued)

Data source	Data frequency	Typical release lag	Country coverage	Scope	Advantages	Caveats	Public (P) or subscription-based (S)	Start date
	Monthly	1–5 days	35 countries, mostly EMs and some frontier countries; some have only equity/debt flows	Portfolio equity and debt flows (only non-resident flows to EMs)	Good proxy for quarterly BoP data; short release lag	Data for the most recent 2–3 months get revised; for some countries, debt flows cover only local currency and/or sovereign bond		Jan-2005 (equity and debt)
	Quarterly	2–4 months	25 EMs	FDI, portfolio equity/debt flows, other investment (only non-resident flows to EMs)	Includes forecasts for the current year and year ahead	Limited country sample relative to IMF BOPS		Jan-2005 (equity and debt)
	Annual	2–4 months	25 EMs	FDI, portfolio equity/debt flows, other investment	Comprehensive cross-country data; includes forecasts for current year	Low data frequency; limited country sample relative to IMF BOPS		1978 (equity and debt)
Emerging Portfolio Fund Research (EPFR): Fund Flows Data	Daily	1 day	Almost all EMs, some frontier markets	Inflows into EM-dedicated investment funds (equity, debt)	High data frequencies and short release lags; provides insights into the behaviour of ultimate investors of investment funds; granular data available by currency denomination, fund domicile, and investor type, among others	Conceptually different from BoP data; based on large sample of reporting funds; institutional investors are underrepresented; country-level flows are estimated top-down	S	May-2007 (equity and debt) Oct-2000 (equity); Jul-2004 (debt) Feb-1996 (equity); Nov-2003 (debt)
	Weekly	1 day						
	Monthly	15 days						

Note: 1/The start date is determined by the first time more than 50% of current EMDEs report their data in the IMF BOPS. Source: BIS, BOPS, EPFR, IIF, and IMF.

by country authorities and cover the various capital flow components (presented as subcomponents of the financial account) as well as all other major international transactions, including current and capital transfers. Capital flow data are generally available both on a gross and net basis for each major component (such as “net FDI”). For the majority of countries, data are available on a quarterly and annual basis and are typically released with a lag of two to four months.

The IMF's balance of payments accounting principles serve as an important anchor for the analysis of capital flow data. Now in its sixth edition, the IMF's Balance of Payments and International Investment Position Manual (BPM6) and its companion Compilation Guide provide detailed guidance on methodological standards.<sup>5</sup> These internationally recognized standards ensure that capital flow data are comparable across countries and time, which is an important precondition for empirical research and policy analysis.

Some of the key accounting principles include:

1. *Residency*: Capital flows arise from the acquisition and disposal of financial assets and liabilities between residents of different countries.<sup>6,7</sup>
2. *Quadruple entry bookkeeping*: Each transaction is recorded twice in each of the two countries' BoP, reflecting the source of funds and the use of funds in each country.<sup>8</sup>
3. *Transactions at market value*: To the extent possible, capital flows are recorded using the market value at the time of the transaction. Changes in valuations (“valuation effects”) do not affect BoP data.<sup>9</sup>

### 2.1.2 | BIS locational and consolidated banking statistics

The cross-border banking data from the BIS provide comprehensive coverage of international banking flows and positions, including currency composition, instrument type, and sector and residency of counterparty (BIS, 2019). The data are reported to the BIS by national central banks and compiled following methods that are broadly consistent with BoP accounting principles. The Locational Banking Statistics (LBS) data are available on a quarterly basis since December 1977, with instrument breakdown available from December 1995. The LBS data are most closely related to the “other investment” component in the BoP, which includes cross-border bank loans and deposits, but the LBS data also capture banks' holdings of debt securities, which in the BoP are classified as portfolio investment.<sup>10</sup> The reported currency breakdown and break-in-series allow the BIS to derive “FX and

break-adjusted changes” and unlike BoP flows, the adjusted changes are technically not “flows” in a narrow sense.

In the Consolidated Banking Statistics (CBS), intra-banking group claims are excluded, in contrast to the LBS.<sup>11</sup> The CBS data on an immediate counterparty basis are available at the semi-annual frequency from December 1983 for 16 years and on a quarterly basis from December 1999 onwards. In addition, the CBS provide data both on an immediate counterparty basis and on a guarantor basis, where the guarantor is the entity assuming contractual responsibilities if the immediate counterparty defaults (BIS, 2019). The quarterly CBS data on a guarantor basis are publicly available since early 2005. Guarantor basis data record a cross-border banking flow between country A and the country of residence of the counterparty's guarantor, rather than the country of residence of the counterparty itself.

### 2.1.3 | IIF capital flow data and portfolio flow trackers

The Institute of International Finance (IIF) has provided independent, private sector estimates of capital flows to and from emerging market economies since the 1990s (IIF, 2007). Up until 2013, IIF data was exclusively at the annual frequency, covering all major capital flow components, and estimated using a proprietary methodology to ensure consistency with stock data on external debt. In 2014 and 2015, the IIF began making available datasets on monthly and daily portfolio flow data that are now widely used (Koepke & Kunii, 2015; Koepke & Mohammed, 2014a, 2014b). These datasets serve as timely proxies for BoP-based portfolio flow data and are constructed using country-level data from individual national sources.<sup>12</sup> These national sources are often used as input data for authorities' official balance of payments data on portfolio flows (typically published at the quarterly or annual frequency). Separately, the IIF also provides monthly estimates of net capital flows (i.e., the financial account balance) for a group of 23 emerging market economies (Fortun, 2020).

The coverage of IIF monthly and especially daily portfolio flows data varies at the country level, depending on the information provided by the underlying source. For equity flows, coverage is quite comprehensive, reflected in a close match with BoP portfolio equity flows. For bond flows, some country data do not include certain types of securities such as hard currency debt or corporate debt (Farnham & Koepke, 2015; see also Tables A2.1 and A2.2 in Annex II for detailed country information in Data S1).



*Monthly flows:* The IIF's monthly tracker for overall emerging market flows uses an econometric model in which underlying portfolio flow data are supplemented with financial variables and bond issuance data to minimize the statistical deviation from quarterly BoP portfolio flow data. The calibration of the model has evolved over the years and saw its fourth update in 2018 (Farnham & Tiftik, 2018).

*Daily flows:* In contrast to the monthly dataset, the aggregate daily data are simply the sum of all reported country data. Some countries are excluded from the aggregate daily tracker because the release lag is longer than for the other countries (e.g., for Mexico, the release lag is typically 10 days). In terms of the underlying data, it is worth noting that for some countries, daily flows are estimated based on stock data, which are sometimes published in local currency terms and thus are being converted into U.S. dollar terms. Purchases of newly issued bonds and redemptions of maturing bonds are generally not included.

#### 2.1.4 | Fund flow data

Another popular high-frequency proxy for portfolio flows is fund flow data. Several private sector data providers make available subscription-based data on flows to investment funds (i.e., mutual funds and exchange-traded funds). The most widely used data provider is EPFR, which makes available monthly, weekly and daily estimates of fund flows for EMs as a group and at the country level, based on a large sample of reporting funds. For monthly data, more than 18,000 reporting equity funds and more than 9,000 reporting debt funds cover around 96% of assets under management (AUM) of the global investment fund industry, using the Investment Company Institute's global AUM measure (as of September 2019, Informa Financial Intelligence, 2020a). Weekly and daily data comprise more than 15,000 reporting equity funds and 7,000 reporting debt funds with about 65% of the AUM of monthly EPFR data (Informa Financial Intelligence, 2020b).

EPFR also publishes estimates of country-level fund flows, which are constructed by applying the monthly average of a fund group's country specific portfolio allocation share to the flows reported by that fund group. These estimates rely on several simplifying assumptions. For example, valuation changes affecting the change in country allocations from one period to the next are assumed to be zero. Moreover, not all funds make available the country-level portfolio allocations needed to estimate country flows at the fund level, so EPFR applies the average country allocation of one fund group to all funds in

this group. Therefore, EPFR's data on country-level flows are considerably less robust than its data on flows to emerging markets as a group and typically differ greatly from country-level portfolio flows data.

A major benefit of EPFR data is that fund flow data are available by fund types, such as ETFs versus mutual funds, funds oriented towards retail versus institutional investors, and active versus passive funds. Moreover, the data can be disaggregated by the fund domicile, investment benchmark, thematic fund category, and by local versus hard currency denomination. Additionally, the data can be further disaggregated by specific categories for both equity and bond funds (e.g., corporate/sovereign, duration, capitalization, sector, etc.).

An important caveat is that EPFR data do not cover all types of emerging market investors, only those investing via mutual funds and exchange traded funds. For example, large institutional investors like sovereign wealth funds, pension funds, hedge funds, and banks' proprietary trading desks typically purchase EM securities directly and are generally not reflected in EPFR data.

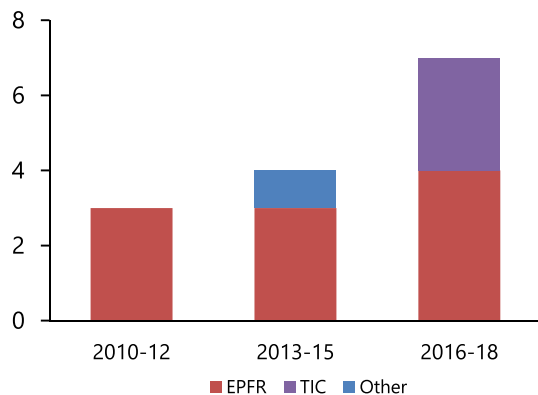
Additional providers of fund flow data include Lipper Fund Flows and Trounceflow, State Street and BNY Mellon offer data products based on their custody holdings.

## 2.2 | The rise of portfolio flow proxies: What accounts for the differences between EPFR, IIF and BoP data?

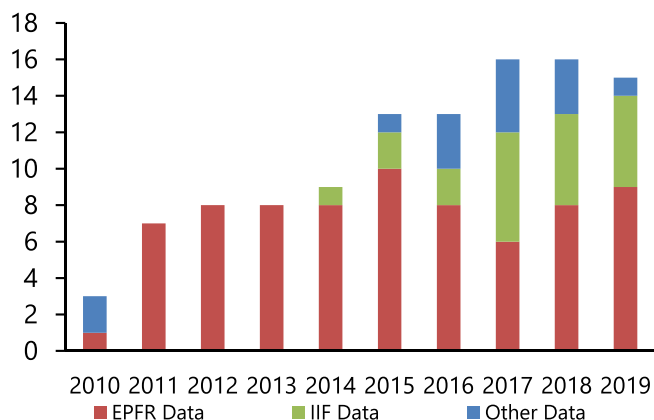
In recent years, portfolio flows have received substantial attention from academics, policymakers and market participants. One reason is that portfolio flows have grown rapidly after the global financial crisis (notably debt flows) and have been the most volatile component of capital flows. Moreover, portfolio debt and equity flows are most closely tied to asset price and exchange rate fluctuations (Bergant & Schmitz, 2018), making them highly relevant for central bank policy decisions. In addition, data availability on portfolio flows is far better than for any other component of capital flows, since there are no comparable monthly, weekly or daily data on banking flows, FDI, or "other investment" flows.

The growing use of high-frequency portfolio flow proxies is evident in both the academic and policy meta-datasets compiled for this paper (Figures 8 and 9). EPFR data in particular have been used frequently in recent academic studies (e.g., Ananchotikul & Zhang, 2014; Bonizzi, 2017; Converse et al., 2020; Fratzscher, 2012; Puy, 2016).

Available datasets on high-frequency portfolio flows differ widely in scope. There are three main sources of discrepancies between EPFR data, IIF data, and quarterly balance of payments data. They relate to conceptual



**FIGURE 8** High frequency capital flow data in the academic dataset, 2010–2018. The chart shows 14 papers published since 2010. *Source:* Authors' calculations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 9** High frequency capital flow data in financial stability reports, 2010–2019. The chart shows 111 reports by G20 EMs, the IMF, World Bank, and BIS. *Source:* Authors' calculations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

differences, data coverage, and EM country samples (Farnham & Tiftik, 2018; Koepke & Mohammed, 2014a, 2014b).

*Conceptual differences:* The IIF's monthly and daily data aim to be consistent with balance of payments accounting principles, in which an inflow is recorded if there is a transaction between a non-resident and a resident. By contrast, flows into investment funds do not necessarily result in cross-border transactions. For example, an EM-dedicated mutual fund may see an inflow from ultimate investors and increase its cash buffer rather than buying EM securities. Similarly, if the mutual fund is domiciled in the U.S. and buys foreign securities from another U.S. investor this transaction would not be recorded as a capital flow in the balance of payments. This issue also applies to residents of emerging markets

who purchase shares of funds that invest domestically (which again would not be recorded in the balance of payments). In EPFR's country flow data, this issue is particularly relevant for Thailand and India, which have sizeable investment funds that are domiciled and invested locally.

*Data coverage:* The limitations to data coverage for both datasets (discussed in the prior section) account for some of the discrepancies between IIF and EPFR data. For the IIF data, these limitations relate to individual series (as they depend on coverage of the original data provider), while EPFR data across the board do not cover all types of emerging market investors, only mutual funds and exchange traded funds.

*Country sample:* There is no universally accepted definition of which countries are emerging markets. For EPFR data, the emerging market universe is guided by which countries are included in key benchmark indices and covers data for a total of 98 countries for equity and 113 countries for debt flows. For IIF data, the emerging market universe is additionally constrained by which countries make available timely portfolio flow data at the monthly (35 countries) and daily (21 countries) frequencies.

*Bottom line:* When to use EPFR versus IIF data? Both datasets are useful for monitoring directional shifts in investor interest in emerging market assets. Overall, EPFR data seem best suited for analysing questions relating to (fund) investor behaviour. In academic research, such questions are more likely to arise in the finance literature than in the international economics literature (e.g., Hu et al., 2014; Morris et al., 2017; Shek et al., 2018). Additionally, fund flow data may be useful to investment professionals for informing asset allocation decisions.

By contrast, IIF data seem most appropriate for analyzing portfolio flows in a macroeconomic and external financing context, which are more common in the international economics literature than in the finance literature (e.g., Cerdeiro & Komaromi, 2019; Forbes & Warnock, 2012; Koepke, 2018; Milesi-Ferretti & Tille, 2011). IIF data may also be preferable for answering most policy-related questions (unless they relate to (fund) investor behaviour). Finally, for country level analysis, IIF data are often preferable to EPFR data, given the simplifying assumptions used to construct the latter (see above, "Fund Flow Data").

### 2.3 | Balance of payments framework: Common misconceptions and measurement limitations

Navigating capital flow data is sometimes complicated by misconceptions that relate to the two-way nature of

capital flows. Moreover, measurement limitations of the balance of payments accounting framework can obscure the true nature and magnitude of flows. Three common misconceptions include:

1. *First, there are several relevant levels of netting, making the commonly used terms of “net” and “gross” capital flows ambiguous.* For example, an important measure of capital flows is the net change in liabilities in the financial account, which captures the net acquisition of a country’s assets by non-residents. This is a “net” measure of capital flows in the sense that different liability components (inward FDI, inward portfolio flows, etc.) are netted against each other. But it is a “gross” measure of capital flows in the sense that changes in liabilities are not netted against changes in assets. Indeed, much of the literature refer to this concept as “gross capital flows” (e.g., Chuhan et al., 1998; Broner et al., 2013; Forbes & Warnock, 2012; see also the discussion in Avdjiev et al., 2020).
2. *Second, long-standing sign conventions changed with the introduction of the sixth edition of the IMF’s Balance of Payments and International Investment Position Manual (BPM6) in 2009.* The most notable changes included that an increase in foreign assets (including reserves) is shown with a positive sign (and the same holds for a decrease in liabilities; see Table 2 below).
3. *Third, there are several different types (or “components”) of capital flows.* The main components are foreign direct investment (FDI), portfolio equity and debt flows, and other investment (which mainly captures banks’ deposit and lending transactions). These different types of capital flows differ greatly from each other in their empirical determinants and dynamic behaviour (Koepke, 2019). Yet, the umbrella term “capital flows” is commonly used to refer to a particular type of capital flows, notably the more volatile portfolio and bank-related flows.<sup>13</sup> Moreover, the term “capital flows” is sometimes used loosely to refer to net purchases of investment fund shares (more appropriately referred to as “fund flows”).

### 2.3.1 | Measurement limitations: Residency versus nationality based recording

The BoP records transactions based on legal residence, which has limitations in today’s integrated global economy. Many firms and investors operate across national boundaries, not just in their country of legal residence. Recent research finds that two important drivers of the growing disconnect between legal residence and

TABLE 2 Sign conventions under BPM5 versus BPM6

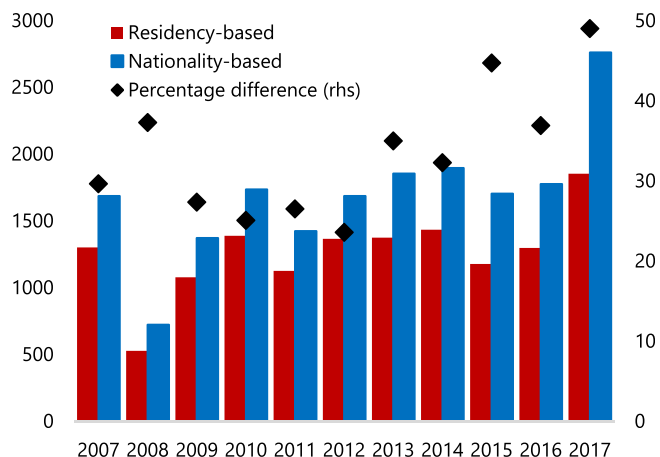
	BPM5	BPM6
Current and capital account	Credits with <i>positive</i> sign. Debits with <b>negative</b> sign.	Credits and debits with <i>positive</i> sign
Financial account	<i>Increase</i> in assets and <b>decrease</b> in liabilities with <b>negative</b> sign. <b>Decreases</b> in assets and <i>increase</i> in liabilities with <i>positive</i> sign.	<i>Increase</i> in assets or liabilities with <i>positive</i> sign. <b>Decrease</b> in assets or liabilities with <b>negative</b> sign.
Financial account balance	Change in assets <i>plus</i> change in liabilities ( <b>negative</b> sign <i>increases</i> IIP)	Change in assets <b>minus</b> change in liabilities ( <i>positive</i> sign <i>increases</i> IIP)

Source: Authors’ illustration adapted from IMF (2007) “FAQs on Conversion from BPM5 to BPM6.”

economic exposure are capital market access and tax considerations (Bertaut et al., 2019; Lane & Milesi-Ferretti, 2018). Common examples of issues that are not appropriately captured in the residency-based BoP framework include:

1. *Corporate offshore borrowing:* Multinational corporations may borrow outside their main domicile and recycle funds through various channels (Avdjiev et al., 2014; Lane & Milesi-Ferretti, 2018).
2. *Corporate tax optimization:* Foreign direct investment to tax havens is driven by tax optimization strategies of multinational corporations (Bertaut et al., 2019; Coppola et al., 2020; Damgaard et al., 2019).
3. *Portfolio allocation via financial centres:* Investment firms in financial centres manage portfolio assets on behalf of ultimate investors in third countries (Lane & Milesi-Ferretti, 2018).
4. *Portfolio allocation via investment funds:* The assets of investment funds are attributed to the country where the fund is registered (Bertaut et al., 2019; IMF, 2014). Moreover, under standard BoP accounting principles the holdings of investment fund shares are classified as equity investments (even for bond funds, for example).

Lane and Milesi-Ferretti (2018) highlight that intermediation through international financial centres obscures underlying capital movements in residency-based databases. Reporting on a nationality basis can complement residency-based reporting to capture underlying capital flow dynamics. Coppola et al. (2020) quantify this effect,



**FIGURE 10** Advanced economy holdings of EMDE equity, USD billion. *Source:* Data based on Coppola et al. (2020) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

finding that advanced economy (AE) holdings of emerging market and developing economy (EMDE) equities are around 50% higher in 2017 on a nationality basis than in the residency-based IMF Coordinated Portfolio Investment Survey (CPIS) and the Treasury International Capital System (TIC, see Figure 10).<sup>14</sup> Bertaut et al. (2019) estimate specifically for the United States that nationality-based reporting for 2017 would reclassify around USD 3.5 trillion or 30% of total US cross-border portfolio investments.

For foreign direct investment statistics, Damgaard et al. (2019) estimate that FDI in special purpose entities, predominantly located in financial centres, account for around 40% of global FDI. Reassigning the remaining “real” direct investments on a nationality basis, outward foreign direct investment stocks from AEs to EMDEs decrease by around USD 1.3 trillion, or around 7%, while outward investment stocks roughly double from USD 1.5 trillion to around USD 3.1 trillion for EMDEs compared to the immediate counterparty reporting in the IMF’s Coordinated Direct Investment Survey and the authors’ estimates.

### 3 | CONSTRUCTING A MONTHLY PORTFOLIO FLOW DATASET CONSISTENT WITH BOP PRINCIPLES

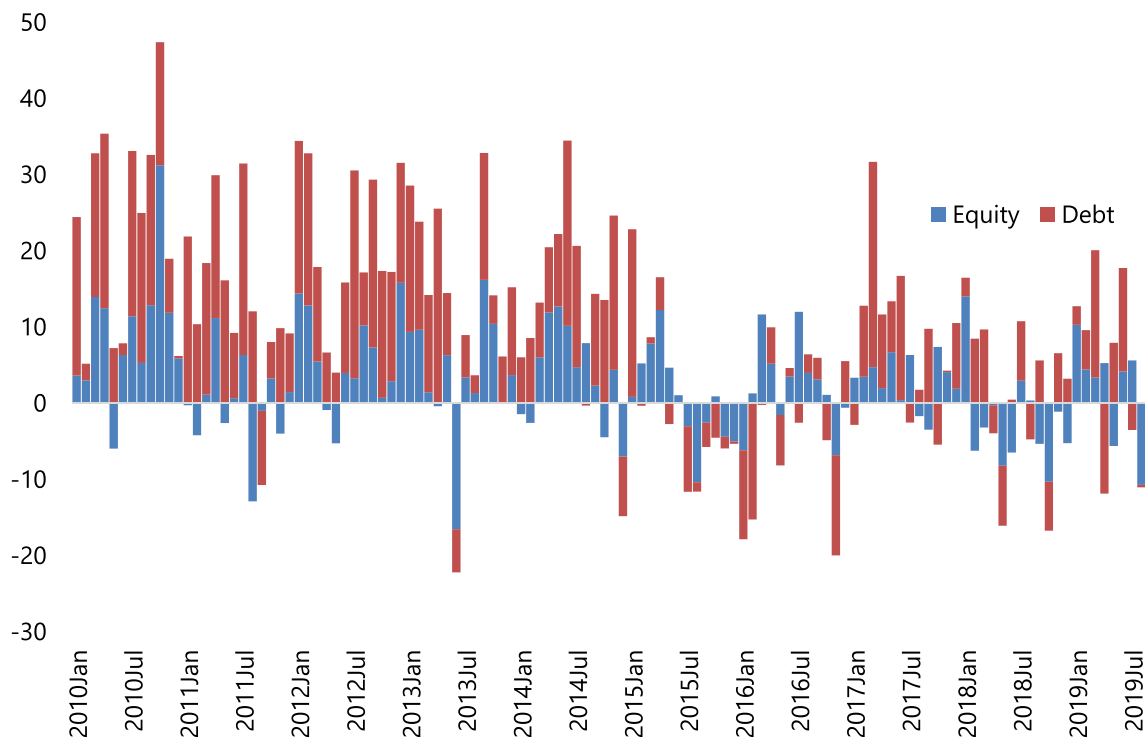
To fill a gap in the availability of portfolio flow data, we construct a dataset on monthly emerging market portfolio flows that is broadly consistent with balance of payments accounting principles and specifically geared towards academic use.<sup>15</sup> The dataset is similar to the monthly data on country-level portfolio flows compiled

by the IIF, but has the advantage that it is available online for free (whereas IIF data are not accessible to most academics because they are restricted to members, which are mostly financial institutions). The dataset is posted online along with this paper and will be updated periodically (<https://www.imf.org/en/Publications/WP/Issues/2020/08/21/Capital-Flow-Data-A-Guide-for-Empirical-Analysis-and-Real-time-Tracking-49646>).

In the rest of this paper, we refer to this dataset as the KP dataset (short for Koepke & Paetzold). The KP dataset is constructed using data from national sources in a set of 19 EMs<sup>16</sup>, with data for aggregate portfolio flows and debt and equity flows for 18 countries. The data are downloaded from national central banks, stock exchanges and statistical agencies for each country. We convert all data to USD where this has not been done by the national sources, using monthly average exchange rate data from the International Financial Statistics database. Country-specific adjustments to the data are explained in the read-me file provided with the dataset. For most countries, the dataset uses the data as published by national authorities. The data are not adjusted for any inconsistencies, methodological shortcomings or idiosyncrasies that may be present in source data from national authorities. But we hope that our paper will complement efforts by national authorities to provide capital flow data that is consistent with international standards. Data availability begins in 2010 for most countries. Figure 11 shows total monthly equity and debt flows, respectively.

Of the 18 countries, 13 report monthly data that sum up to the quarterly balance of payments data on portfolio flows published by the central bank or statistical authority of the respective country. For five countries, we collect monthly proxies for one or more portfolio flow type.<sup>17</sup> Table 3 provides additional details on the data used in the KP dataset. Regarding the release lags, it is worth noting that monthly data have an inherent advantage over quarterly data. For example, with a release lag of 2 months, data for the first month of a quarter will already be available at the end of any given quarter. Relying on quarterly data with a release lag of 2 months instead means that data for the entire quarter, including its first month, would only become available 2 months after the end of the quarter.

Figure 12 shows total EM equity and debt flows, respectively, for the KP data and two relevant comparators (IIF monthly tracker, EPFR monthly). The sum of total flows to EMs in the KP dataset from 2010:Q1 to 2019:Q2 covers around 46% of total flows in the BOPS, with somewhat higher coverage of equity flows (53%) and lower coverage for debt flows (44%). Equity data from the KP dataset and the IIF’s monthly portfolio flow



**FIGURE 11** Total monthly equity and debt flows from KP dataset, USD billion. *Source:* KP dataset [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

tracker are highly correlated with quarterly BoP-based equity flows (see Table 4). For the debt component, the KP data has a high correlation with BoP flows at the country level, but somewhat lower at an EM aggregate level. The lower correlation for debt flows can be attributed mainly to a few outlier periods in which the KP data do not capture the full magnitude of quarterly capital flows to emerging markets because not all emerging markets are included in the dataset. This discrepancy is also apparent in the right panel of Figure 12 (e.g., in 2017), while in earlier periods the KP data performed better than EPFR data in tracking debt flows. It is worth noting that the KP data is expanded to include newly available data, which should help reduce such discrepancies over time. On a country-by-country basis, the KP data also tracks BoP data very closely, although the correlation between BoP flows and KP data is higher for some countries than others (see Table 4).<sup>18</sup>

In contrast to investment fund data such as EPFR, the dataset provided with this paper more closely aligns with standard definitions in the BoP. Using this dataset may help researchers obtain empirical findings that are valid for “true” capital flow data, consistent with established BoP accounting principles. Compared to the IIF data, our dataset underperforms in tracking quarterly BoP flows, especially debt flows. While the underlying data sources for our dataset and IIF seem to align closely,

the IIF's econometric model by design minimizes the discrepancy vis-a-vis BOP flows.

#### 4 | REAL-TIME TRACKING OF PORTFOLIO FLOWS

In many emerging market countries, capital flows and exchange rate pressures have long been a key factor driving central banks' policy decisions (Calvo & Reinhart, 2002; Vegh & Vuletin, 2012). Policymakers need timely information on capital flows to inform decisions on monetary policy and foreign exchange intervention. Official BoP data are of little help, given their long release lag. Instead, policymakers often rely on a combination of national data on foreign exchange transactions involving their own currencies and private sector data providers on portfolio flows, notably EPFR and IIF data.

In order to assess which portfolio flow proxies are best suited for real-time tracking, we conduct a quantitative assessment of how well EPFR and IIF data predict official BoP data. We run a nowcasting “horse race,” in which quarterly BoP portfolio flow data are predicted using available portfolio flow proxies as of a given point in time.<sup>19</sup> To our knowledge, this is the first systematic assessment of the predictive content of commonly used portfolio flow proxies.

TABLE 3 Description of data in the KP monthly portfolio flow dataset

Country	Components	Data source	Release lag in months (Approx.)	Currency	Proxy data	Correlation with total BoP flows (all available data until 2019Q2)
Brazil	Debt/Equity	Central Bank of Brazil	1–2	USD		1.00
Bulgaria	Debt/Equity	Eurostat	2–3	EUR		1.00
Chile	Debt/Equity	Central Bank of Chile	2–3	USD		1.00
Czech Republic <sup>a</sup>	Debt/Equity	Czech National Bank	2–3	EUR		1.00
Hungary	Debt/Equity	Eurostat	2–3	EUR		1.00
India <sup>b</sup>	Debt/Equity	Securities & Exchange Board of India	0–1	INR	Y	0.92
Korea <sup>a</sup>	Debt/Equity	Bank of Korea	1–2	USD		1.00
Lebanon	Debt/Equity	Bank of Lebanon	10–11	USD		0.97
Mexico	Debt/Equity	Bank of Mexico	3–4	USD	Y	0.86
Pakistan	Debt/Equity	State Bank of Pakistan	1–2	USD		0.98
Philippines	Debt/Equity	Central Bank of the Philippines	4–5	USD		1.00
Poland	Debt/Equity	National Bank of Poland	2–3	EUR		1.00
Romania	Debt/Equity	Eurostat	2–3	EUR		1.00
South Africa	Debt/Equity	Johannesburg Stock Exchange	2–3	ZAF	Y	0.64
Sri Lanka	Debt/Equity	Colombo Stock Exchange	2–3	USD	Y	0.30
Thailand	Debt/Equity	Bank of Thailand	3–4	USD	Y	0.92
Turkey	Debt/Equity	Central Bank of Turkey	2–3	USD		1.00
Ukraine	Debt/Equity	National Bank of Ukraine	2–3	USD		1.00

<sup>a</sup>The Czech Republic and Korea are not part of the IMF's classification of emerging market economies, but are included in private sector classifications of EMs such as leading investment benchmark indices.

<sup>b</sup>India's portfolio flow data are recorded on a reporting day basis rather than on a trading day basis, which may contribute to a lower correlation between data used in the KP dataset and Balance of Payments data.

Source: Authorities' data, BOPS, and IMF.

More broadly speaking, there is very limited literature discussing methods to track capital flows in real time. One important study is Miao and Pant (2012), which highlights the serious limitations of traditional capital flow data for surveillance and policymaking purposes. The authors propose two methods to address this data gap. One focuses on net capital flows (i.e., the financial account balance), which is estimated using timely data on trade in goods and changes in reserves, augmented with several global and EM financial variables (such as the VIX, US bond prices, and EM stock prices).<sup>20</sup> In the other method, an error correction model is applied to EPFR data and several financial variables to obtain an estimate of gross portfolio flows to EM regions. However, both methods are still subject to considerable time lags,

and the study does not assess predictive performance relative to other proxies.

#### 4.1 | Nowcasting horse race

We use several alternative portfolio flow proxies in a nowcasting setting with quarterly portfolio flows from the BoP as the dependent variable, where the accuracy of nowcasts is measured as the root mean squared forecast error (RMSFE). The portfolio flow proxies include:

1. Weekly EPFR data, published with a lag of about 7 days.

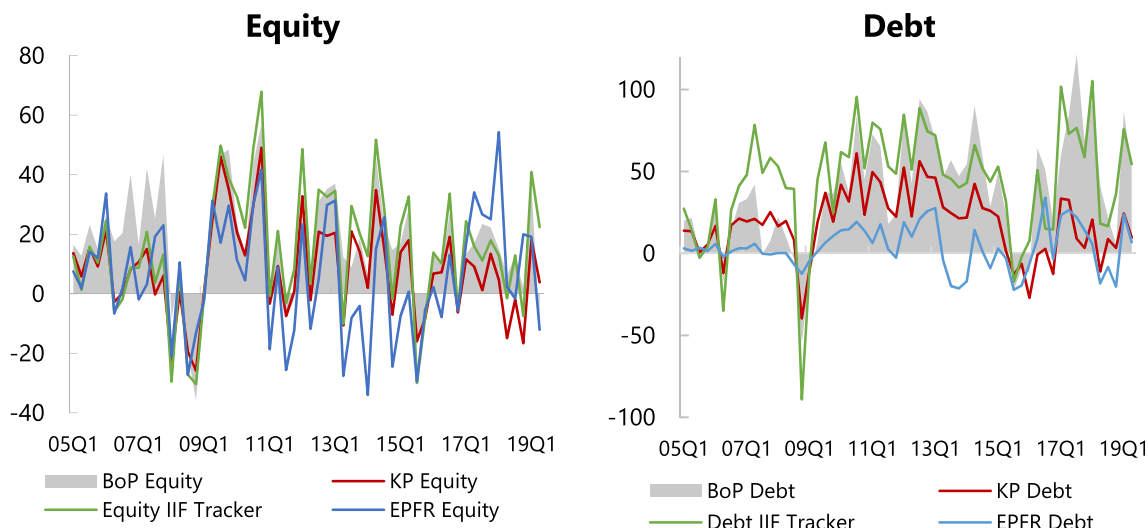


FIGURE 12 Comparison balance of payments and EM portfolio flow proxies, USD billion. Source: IMF BOPS, EPFR, IIF [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

2. Monthly EPFR data, published with a lag of about 15 days.
3. Daily IIF data, published with a lag of 1–3 days.
4. Monthly IIF data, published with a lag of about 1–5 days.

We also include the monthly data from the KP dataset, described in the prior section. The underlying data are typically available after about 2–3 months.

The sample period for the nowcasting exercise runs from 2005:Q1 until 2019:Q2. In our baseline model, we use aggregate flows for those emerging market countries where all the portfolio flow proxies are available (separately for equity and debt flows).<sup>21</sup> The selection is guided by data availability. Only those countries with data from all high-frequency proxies are included. Given the limited number of countries with daily flows data, the subset of emerging markets might not be representative for the entire emerging market universe. Therefore, we only predict quarterly BoP portfolio flows of the countries that are part of the sample and not those to all emerging markets. We do not run panel regressions but instead sum the capital flow proxies as well as BoP flows in each quarter. Aggregate flows are scaled by the combined quarterly GDP of countries included in the sample. We estimate a pseudo-out-of-sample forecast error, using only data that would have been available as of the relevant quarter. Therefore, the number of observations used in the regression is quite small at the beginning of the sample period and increases over time.

In the first step, observations for each portfolio flow proxy are summed up incrementally for each quarter. For example, daily capital flows are summed for days 1–2, days 1–3, days 1–4, and so forth, up until the last

observation of the quarter is included. Similarly, weekly/monthly data are summed over an increasing number of weeks/months in each quarter.

In the second step, for each of these sums the following model is estimated:

$$\text{BoP}_{\text{quart}} = c + \beta_{ij} \cdot \text{capflow\_sum}_{ij},$$

where  $\text{BoP}_{\text{quart}}$  is quarterly portfolio flow data from the balance of payments,  $c$  is a constant term, and  $\text{capflow\_sum}_i$  is the sum of the high-frequency proxies for portfolio flows from the first observation to the  $i$ -th observation of the quarter, and  $j$  represents the portfolio flow proxy used.

This model is used to predict quarterly BoP portfolio flows using available data from 2005:Q1 onwards. For example, to predict the quarterly BoP flows in 2007:Q1, the model is estimated using the observations for 2005:Q1–2006:Q4, while for the prediction for 2012:Q2, the model is estimated using the observations for 2005:Q1–2007:Q1. For each quarter,  $i$  nowcasts are produced by multiplying the estimated coefficient from the regression and  $\text{capflow\_sum}_i$  for that quarter (and adding the constant term).

We evaluate the predictive content of each portfolio flow proxy by computing the root-mean squared forecast error for each  $i$ , for the period from 2007:Q1 until 2019:Q2:

$$\text{RMSFE}_{ij} = \sqrt{\frac{1}{T} \sum (y_{tij} - \hat{y}_{tij})^2},$$

where RMSFE is the root-mean squared forecast error for sum  $i$  and proxy  $j$ ,  $T$  is the number of quarters for which

**TABLE 4** Correlation with balance of payments portfolio flows (2010:Q1–2019:Q2)

	<b>Equity</b>		<b>Debt</b>	
All EMs	IIF Monthly	0.86	IIF Monthly	0.83
	EPFR Monthly	0.67	EPFR Monthly	0.68
	KP	0.85	KP	0.64
Brazil	IIF Monthly	1.00	IIF Monthly	0.99
	EPFR Monthly	0.56	EPFR Monthly	0.04
	KP	1.00	KP	1.00
Bulgaria	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	0.11	EPFR Monthly	0.03
	KP	1.00	KP	1.00
Chile	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	−0.09	EPFR Monthly	0.18
	KP	1.00	KP	1.00
Czech Republic	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	0.13	EPFR Monthly	0.07
	KP	0.96	KP	1.00
Hungary	IIF Monthly	0.99	IIF Monthly	0.73
	EPFR Monthly	0.14	EPFR Monthly	0.20
	KP	1.00	KP	1.00
India	IIF Monthly	0.95	IIF Monthly	0.94
	EPFR Monthly	0.18	EPFR Monthly	0.16
	KP	0.94	KP	0.93
Korea	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	0.26	EPFR Monthly	0.33
	KP	1.00	KP	1.00
Lebanon	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	0.16	EPFR Monthly	0.34
	KP	0.96	KP	0.97
Mexico	IIF Monthly	0.97	IIF Monthly	0.82
	EPFR Monthly	0.22	EPFR Monthly	0.21
	KP	0.97	KP	0.82
Pakistan	IIF Monthly	0.77	IIF Monthly	1.00
	EPFR Monthly	0.10	EPFR Monthly	−0.04
	KP	1.00	KP	1.00
Philippines	IIF Monthly	0.90	IIF Monthly	0.98
	EPFR Monthly	0.40	EPFR Monthly	0.31
	KP	1.00	KP	1.00
Poland	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	0.35	EPFR Monthly	0.47
	KP	1.00	KP	1.00
Romania	IIF Monthly	N/A	IIF Monthly	N/A
	EPFR Monthly	0.01	EPFR Monthly	0.11
	KP	1.00	KP	1.00

(Continues)



TABLE 4 (Continued)

	Equity		Debt	
South Africa	IIF Monthly	0.75	IIF Monthly	0.64
	EPFR Monthly	0.25	EPFR Monthly	0.51
	KP	0.74	KP	0.64
Sri Lanka	IIF Monthly	0.89	IIF Monthly	N/A
	EPFR Monthly	0.12	EPFR Monthly	0.26
	KP	0.91	KP	0.25
Thailand	IIF Monthly	0.80	IIF Monthly	0.75
	EPFR Monthly	0.11	EPFR Monthly	0.06
	KP	1.00	KP	0.65
Turkey	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	0.54	EPFR Monthly	0.56
	KP	1.00	KP	1.00
Ukraine	IIF Monthly	1.00	IIF Monthly	1.00
	EPFR Monthly	-0.03	EPFR Monthly	0.23
	KP	1.00	KP	1.00

Source: Authors' calculations, data as of July 2020.

a forecast (or “nowcast”) is produced,  $t$  is the time in quarters,  $y$  is actual portfolio flows for that quarter, and  $\hat{y}$  is the forecast for that quarter.

The results of this exercise are shown in Figure 13 for equity flows and in Figure 14 for debt flows. As expected, forecast errors for all proxies generally decline over the course of the quarter as more data become available. Daily IIF data and weekly EPFR data are available as early as 2 and 7 days into the quarter, much earlier than either of the monthly proxies. RMSFEs based on the first few observations of daily and weekly proxies are relatively high, but the forecast performance of the daily IIF data improves notably once 20–30 days' worth of data are included.

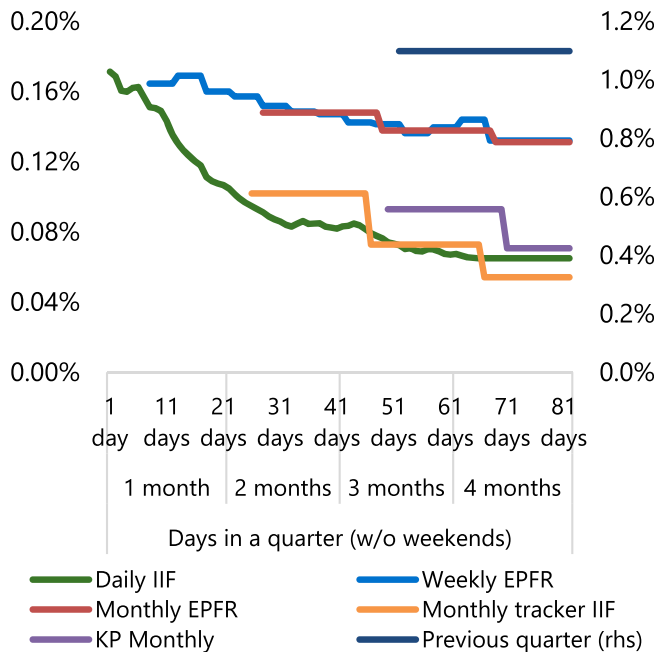
The monthly proxies first become available about 33 and 45 days into the quarter for IIF and EPFR data, respectively. The IIF's monthly tracker consistently performs better than either of the EPFR datasets (weekly or monthly), but only outperforms the IIF daily flows data after the second data release in the quarter (on average). Similarly, EPFR's monthly data mostly underperforms the weekly data, with the exception of debt flows towards the end of the quarter. We also compare the RMSFEs of the various proxies to that of a benchmark autoregressive regression with one lag and a constant term. All proxies outperform this benchmark by a wide margin, with all RMSFEs at least 80% lower.

The improvements in forecast performance of quarterly capital flows using IIF compared to EPFR data are economically important. Expressed as a share of absolute

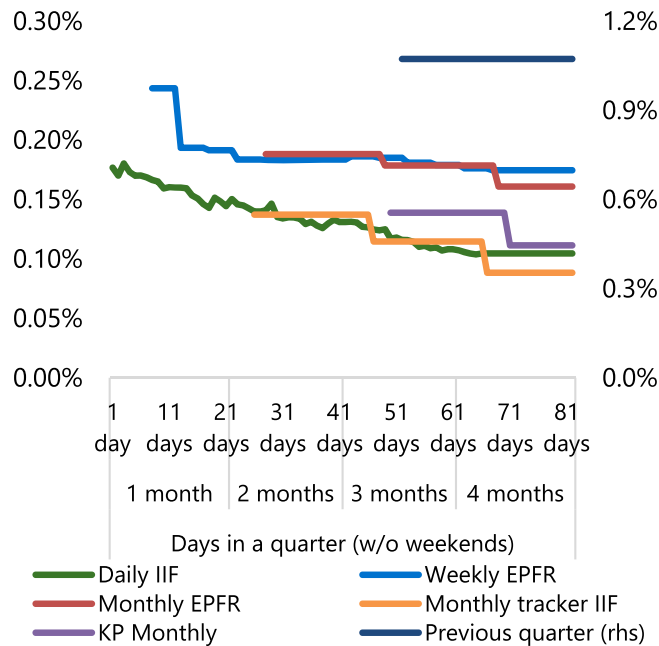
average quarterly flows, RMSFEs of debt and equity flows based on daily and monthly IIF data outperform weekly and monthly EPFR data, respectively, by around 20% early in the quarter and 50% at the end of the quarter. Compared to the benchmark regression, RMSFEs of IIF flows as a share of quarterly flows are consistently lower while EPFR data only achieve marginally lower RMSFEs at the end of the quarter.

The results also show that the KP dataset has a similar performance to the IIF's monthly tracker, although the monthly data from countries contained in the KP dataset is subject to a greater release lag within the current quarter. The better forecast performance of IIF and KP data is consistent with our prior, given that EPFR data is conceptually different from the other data sources (as discussed in Section 2). One caveat regarding the IIF's monthly EM tracker and the KP dataset is that observations for the latest 1–2 months of data are revised with each monthly data release for the EM aggregate estimate and for more than half of the countries included in the country level flows. This may bias our results towards a smaller forecast error for these datasets as the other data sources generally do not get revised.

The main findings from the nowcasting horse race are that (1) all portfolio flow proxies considered have significant predictive content for BoP portfolio flows, (2) IIF proxies generally outperform EPFR proxies, and (3) higher frequency proxies (e.g., daily IIF data) generally outperform lower-frequency proxies, especially for equity flows in the first half of the “current” quarter.



**FIGURE 13** Equity portfolio flows to EMs (root mean squared forecasting error (RMSFE) in % of GDP). Using all countries for which equity and debt flows are available from all data sources. For equity flows, the sample contains Brazil, India, Pakistan, Philippines, South Africa, Korea, Sri Lanka, Thailand, and Turkey. For debt flows, the sample contains Hungary, India, Poland, South Africa, Thailand and Turkey. *Source:* BOPS, EPFR, IIF, IMF [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 14** Debt portfolio flows to EMs (Root mean squared forecasting error (RMSFE) in % of GDP). Using all countries for which equity and debt flows are available from all data sources. For equity flows, the sample contains Brazil, India, Pakistan, Philippines, South Africa, Korea, Sri Lanka, Thailand, and Turkey. For debt flows, the sample contains Hungary, India, Poland, South Africa, Thailand and Turkey. *Source:* BOPS, EPFR, IIF, IMF [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

We test the robustness of these findings through a range of alternative specifications. For example, we conduct the same exercise using a shorter sample period that begins in 2010:Q1, which did not substantially change average forecast errors. In an extension inspired by the coincident indicators proposed by Miao and Pant (2012), we include averages from the first to the *i*-th quarterly observation of global financial variables used in the literature on capital flow drivers (VIX, US Treasury yields) as additional predictors. The inclusion of these variables does not consistently reduce the RMSFEs and the relative magnitude of forecast errors does not change substantively.

We also run the nowcasting exercise using aggregate EM flows from each source (for equity flows, debt flows, and the sum of equity and debt flows), as well as for individual EM countries where data are available from all three data sources. The results are presented in Annex IV (Data S1). For the EM aggregate exercise, the monthly proxy data often perform better than in our baseline results. Notably, the IIF's monthly tracker generally outperforms other indicators for both total flows and equity flows, while EPFR tends to outperform for debt flows.

## 5 | CONCLUSION

In this paper, we provide an overview of the main capital flow data sources used in the empirical literature and among policymakers. IMF balance of payments statistics remain a key data source and serve as an important benchmark for assessing the properties of other datasets. Data availability has improved significantly in recent years, especially with respect to high frequency proxies for portfolio flows at the daily, weekly, and monthly frequencies. We find that fund flow data have been used by a sizeable share of empirical studies (around 14%) over the past decade, allowing researchers to gain deeper insight into the drivers of rapid shifts in investor behaviour.

This paper also sheds light on the conceptual and empirical differences between fund flow data and portfolio flows as measured in the IMF's balance of payments statistics. These differences highlight that findings in the empirical literature are likely shaped by unique properties of fund flow data. A greater focus on high frequency data that are consistent with balance of payments accounting principles seems warranted but may have been hampered by data subscription requirements and limitations of existing datasets.

We help address this gap by providing a monthly portfolio flow dataset designed specifically for academic use. The dataset covers 18 of the largest emerging market economies and tracks the quarterly balance of payments data of these countries closely. The dataset may be particularly useful for analyzing rapid shifts in portfolio flows, such as those witnessed during the COVID-19 pandemic. The dataset may also help facilitate future research investigating whether the role of external drivers (relative to domestic factors) is as important as suggested by papers using fund flow data (which are by construction subject to common factors).

Finally, we assess the predictive content of various portfolio flow proxies in a nowcasting “horse race.” The results are encouraging in that all portfolio flow proxies have significant predictive content for balance of payments portfolio flows. Among the various predictors, IIF portfolio flow trackers generally outperform EPFR fund flow data, and higher frequency proxies generally outperform lower-frequency data, especially in the first half of the current quarter.

Overall, this paper highlights the importance for researchers and policymakers to understand in detail the capital flow data and proxies they rely upon. This will help economists utilize the best-suited dataset depending on the question that is asked, and will help inform how the answers are framed. On this basis, the growing number of capital flow datasets provide valuable opportunities for advancing economic research and informing policy decisions in real time.

## ACKNOWLEDGEMENTS

We are grateful for helpful suggestions by Gustavo Adler, Swarnali Ahmed, Serkan Arslanalp, Roland Beck, Mahir Binici, Robin Brooks, Eugenio Cerutti, Nathan Converse, Dimitris Drakopoulos, Thomas Elkjaer, Eric Fischer, Jonathan Fortun, Joji Ishikawa, Sebnem Kalemli-Özcan, Yevgeniya Korniyenko, Swapnan-Kumar Pradhan, Gian Maria Milesi-Ferretti, Gurnain Pasricha, Daniel Rodriguez, Russell Rollow, Ratna Sahay, Todd Schneider, Beatrice Scheubel, Ling Zhu and an anonymous referee. We would also like to acknowledge the valuable comments from participants in several IMF seminars on monitoring capital flows, including during presentations given to the IMF’s Capital Flows Group, the Asia & Pacific Department, the Western Hemisphere Department, and two ICD lectures on “Tracking, Interpreting and Forecasting Capital Flows Developments.” Open Access funding enabled and organized by Projekt DEAL.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Robin Koepke  <https://orcid.org/0000-0002-9308-9619>

Simon Paetzold  <https://orcid.org/0000-0002-1887-7156>

## ENDNOTES

- <sup>1</sup> The numbers referenced for IIF and IMF data relate to the concept of “gross” portfolio inflows, defined as non-resident purchases of EM stocks and bonds. Section 2 provides a detailed discussion of differences in scope and data coverage of BoP, EPFR and IIF data.
- <sup>2</sup> An important benefit of using capital flow data on a BoP basis is that they are consistent with other macroeconomic accounts, such as current account data, government finance statistics, and the international investment position. Capital flow data are often analyzed through the prism of external financing (of a current account deficit) and financing of a government’s budget deficit. Indeed, a key motivation of the voluminous literature on the role of external (“push”) and domestic (“pull”) drivers of capital flows has been to advance our understanding of the degree to which foreign capital inflows are likely to be a *temporary* rather than *permanent* source of external finance (e.g., Calvo et al., 1993; Chuhan et al., 1998; Fernández-Arias, 1996; surveyed in Koepke, 2019). This motivation argues for the use of BoP-consistent data that can be analyzed and interpreted in conjunction with other standard macroeconomic data.
- <sup>3</sup> For the academic dataset, we build on the work of Koepke (2019), which surveys 34 empirical studies to analyze the empirical drivers of capital flows to emerging markets. We extend this sample of studies by including an additional 54 studies in this same strand of literature, focusing on papers with a large number of citations and/or published in leading economic journals. Overall, we evaluate a total of 88 studies published over the course of the last 27 years and listed in Annex III (Data S1). For the policy dataset, we review 220 reports published since 2010, including financial stability reports by G-20 emerging market central banks and reports of policy-oriented international financial institutions, notably the IMF’s Global Financial Stability Reports, the BIS’s Annual Reports, and the World Bank’s Global Development Prospects report.
- <sup>4</sup> Information on data frequency or data source was unavailable or ambiguous in some papers such that the percentages do not sum to 100%.
- <sup>5</sup> The BPM6 (IMF 2009) is available at <https://www.imf.org/external/pubs/ft/bop/2007/pdf/bpm6.pdf>. The BPM6 compilation guide (IMF 2014) is available at <https://www.imf.org/external/pubs/ft/bop/2014/pdf/guide.pdf>.
- <sup>6</sup> The BPM6 notes that “residence of each institutional unit is the economic territory with which it has the strongest connection, expressed as its center of predominant economic interest” (§4.113).
- <sup>7</sup> It is worth noting that the term “flow” has a different meaning in the capital flow literature than in the terminology of balance of payments statistics. In the capital flow literature, a “flow” relates to transactions. In macroeconomic statistics, a “flow” refers to all changes in the stock of a variable from one period to the next, including transactions, valuation changes, and other flows (see BPM6 §2.2).

- <sup>8</sup> The BPM6 specifies that “a single transaction between two counterparties thus gives rise to four entries. [...] As a liability of one unit is mirrored in a financial asset of another unit, for instance, they should be identically valued, allocated in time, and classified to avoid inconsistencies in aggregating balance sheets of units into regional or global totals” (§3.29).
- <sup>9</sup> The BPM6 notes that “market prices are the basis for valuation in the international accounts” (§3.67). However, some financial instruments (e.g., loans) are recorded at nominal values.
- <sup>10</sup> Pradhan and Silva (2019) show that the data contained in the LBS align relatively closely with the IMF’s International Investment Position Statistics but there are several sources of discrepancies, including differences in reporting populations, mix of data sources, treatment of instruments, and country-specific factors.
- <sup>11</sup> The LBS provide bilateral data for both claims and liabilities whereas the CBS provide data mainly for claims and very limited data on liabilities (overall total and local currency liabilities in host countries).
- <sup>12</sup> More specifically, the data serve as a proxy for non-residents’ net acquisition of EM assets, that is, inward portfolio flows to EMs. For some countries, the sum of daily or monthly observations is identical to the official quarterly BoP data on inward portfolio flows for the same period.
- <sup>13</sup> For example, the paper entitled “Capital Flows are Fickle: Anytime, Anywhere” (Bluedorn et al., 2013) mainly focuses on volatile components of capital flows. Similarly, a recent paper on “Capital Flows at Risk” (Gelos et al., 2019) focuses on portfolio equity and debt flows. The use of the term “capital flows” is thus similar to the use of the word “America” for the United States of America (a stylistic device referred to as *totum pro parte*, the whole for a part).
- <sup>14</sup> The study includes the following countries with high-quality fund holding data: Australia, Canada, Denmark, European Monetary Union, Norway, Sweden, Switzerland, the United States, and the United Kingdom.
- <sup>15</sup> In particular, the data are based on transactions between residents and non-residents of emerging market countries. The data are not subject to valuation effects and reflect transactions at market prices.
- <sup>16</sup> The countries included are the Brazil, Bulgaria, Chile, Czech Republic, Hungary, India, Lebanon, Mexico, Pakistan, Philippines, Poland, Romania, South Africa, Korea, Sri Lanka, Thailand, Turkey, Ukraine. Of these countries, the Czech Republic and Korea are not part of the IMF’s classification of emerging markets, but are included in private sector classifications of EMs such as leading investment benchmark indices. Data for Bulgaria, Hungary, India, Romania, and South Africa were converted to USD using monthly average exchange rates.
- <sup>17</sup> Among the proxy data, we use government debt flows for Mexico, non-government debt flows for Thailand, transactions on national stock exchanges by non-residents for South Africa and Sri Lanka and portfolio inflows by non-residents for India.
- <sup>18</sup> A potential reason for these discrepancies could be the use of proxy data and incomplete data for capital flows (i.e., stock exchange data in the case of equity flows for Sri Lanka and South Africa, the exclusion of government debt flows for

Thailand, and the exclusion of private sector debt flows for Mexico and Sri Lanka).

- <sup>19</sup> In this section, the terms “nowcast,” “forecast,” and “prediction” are used interchangeably.
- <sup>20</sup> As noted in Miao and Pant (2012), variants of this approach have been used by several studies to obtain timely proxies of capital flow movements, including Calvo et al. (2004, 2008). One limitation of this approach is that trade in services, primary income, and secondary income play major roles in some EMs, causing errors in the estimation. The IIF provides monthly estimates of net capital flows for 23 EMs, using the same basic approach.
- <sup>21</sup> For equity flows, the sample contains Brazil, India, Pakistan, Philippines, South Africa, Korea, Sri Lanka, Thailand, and Turkey. Pakistan and Sri Lanka were excluded prior to 2010 and Brazil before 2008 as daily IIF data was unavailable for these countries beforehand. For debt flows, the sample contains Hungary, India, Poland, South Africa, Thailand, and Turkey. Thailand was excluded prior to 2010 as daily IIF data was unavailable.

## REFERENCES

- Ananchotikul, N., & Zhang, L. (2014). Portfolio Flows, Global Risk Aversion and Asset Prices in Emerging Markets. IMF Working Paper no. 14-156. International Monetary Fund.
- Avdjiev, S., Chui, M. K. F., & Shin, H. S. (2014). Non-financial corporations from emerging market economies and capital flows. *BIS Quarterly Review*, (December), 67–77.
- Avdjiev, S., Hardy, B., Kalemli-Özcan, S., & Servén, L. (2020). *Gross capital flows by banks, corporates, and sovereigns*. The World Bank.
- Bank for International Settlements. (2009). Capital Flows and Emerging Market Economies. CGFS Papers No. 33. Bank for International Settlements.
- Bank for International Settlements. (2019). Reporting Guidelines for the BIS International Banking Statistics. <https://www.bis.org/statistics/bankstatsguide.htm>
- Bergant, K., & Schmitz, M. (2018). *Valuation effects and capital flows-security level evidence from euro area investors*. Trinity College Dublin. School of Social Sciences & Philosophy.
- Bertaut, C., Bressler, B., & Curcuru, S. E. (2019). Globalization and the Geography of Capital Flows. <https://www.federalreserve.gov/econres/notes/feds-notes/globalization-and-the-geography-of-capital-flows-20190906.htm>
- Bluedorn, J. C., Duttagupta, R., Guajardo, J., & Topalova, P. (2013). Capital Flows are Fickle: Anytime, Anywhere. IMF Working Paper no. 13-183. International Monetary Fund.
- Bonizzi, B. (2017). Institutional investors’ allocation to emerging markets: A panel approach to asset demand. *Journal of International Financial Markets, Institutions and Money*, 47, 47–64.
- Broner, F., Didier, T., Erce, A., & Schmukler, S. L. (2013). Gross capital flows: Dynamics and crises. *Journal of Monetary Economics*, 60(1), 113–133.
- Calvo, G., Izquierdo, A., & Mejía, L.-F. (2004). On the Empirics of Sudden Stops: The Relevance of Balance-Sheet Effects. NBER Working Paper no. 10520. National Bureau of Economic Research.
- Calvo, G., Izquierdo, A., & Mejía, L.-F. (2008). Systemic Sudden Stops: The Relevance of Balance-Sheet Effects and Financial

- Integration. NBER Working Paper no. 14026. National Bureau of Economic Research.
- Calvo, G. A., Leiderman, L., & Reinhart, C. M. (1993). Capital Inflows and Real Exchange Rate Appreciation in Latin America: the Role of External Factors. *IMF Staff Papers*: 108-151. International Monetary Fund.
- Calvo, G. A., & Reinhart, C. M. (2002). Fear of floating. *The Quarterly Journal of Economics*, 117(2), 379-408.
- Cerdeiro, D. A., & Komaromi A. (2019). Financial Openness and Capital Inflows to Emerging Markets: In Search of Robust Evidence. IMF Working Paper no. 19-194. International Monetary Fund.
- Cetorelli, N., & Goldberg, L. S. (2011). Global banks and international shock transmission: evidence from the crisis. *IMF Economic Review*, 59(1), 41-76.
- Chuhan, P., Claessens, S., & Mamingi, N. (1998). Equity and bond flows to Latin America and Asia: The role of global and country factors. *Journal of Development Economics*, 55(2), 439-463.
- Converse, N., Levy-Yeyati, E., & Williams, T. (2020). How ETFs Amplify the Global Financial Cycle in Emerging Markets. Federal Reserve International Finance Discussion Papers no. 1268. Board of Governors of the Federal Reserve System.
- Coppola, A., Maggiori, M., Neiman, B., & Schreger, J. (2020). Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens. NBER Working Paper no. 26855. National Bureau of Economic Research.
- Damgaard, J., Elkjaer, T., & Johannesen, N. (2019). What Is Real and What Is Not in the Global FDI Network? IMF Working Paper no. 19-274. International Monetary Fund.
- Dooley, M. (1996). A survey of literature on controls of international capital transactions. *IMF Staff Papers*, 43(4), 639-687.
- Farnham, S., & Koepke, R. (2015). Daily portfolio flows user guide. IIF Research Note.
- Farnham, S., & Tiftik, E. (2018). Introducing EM Portfolio Flows Tracker 4.0. IIF Methodology Note. [https://www.iif.com/Portals/0/Files/portfolio\\_flows\\_tracker\\_4.0\\_methodology\\_note.pdf](https://www.iif.com/Portals/0/Files/portfolio_flows_tracker_4.0_methodology_note.pdf)
- Fernández-Arias, E. (1996). The new wave of private capital inflows: Push or pull? *Journal of Development Economics*, 48(2), 389-418.
- Financial Stability Board (FSB) and International Monetary Fund (IMF). (2009). The Financial Crisis and Information Gaps: Report to the G-20 Finance Ministers and Central Bank Governors. IMF Staff and the FSB Secretariat (October 29, 2009). <https://www.imf.org/external/np/g20/pdf/102909.pdf>
- Forbes, K. J., & Warnock, F. E. (2012). Capital flow waves: Surges, stops, flight, and retrenchment. *Journal of International Economics*, 88(2), 235-251.
- Fortun, J. (2020). IIF Capital Flows Tracker - August 2020. IIF Research Note.
- Fratzscher, M. (2012). Capital flows, push versus pull factors and the global financial crisis. *Journal of International Economics*, 88(2), 341-356.
- Gelos, R. G., Gornicka, L., Koepke, R., Sahay, R., & Sgherri, S. (2019). Capital Flows at Risk: Taming the Ebbs and Flows. IMF Working Paper no. 19/279. International Monetary Fund.
- Hu, G., David McLean, R., Pontiff, J., & Wang, Q. (2014). The year-end trading activities of institutional investors: Evidence from daily trades. *The Review of Financial Studies*, 27(5), 1593-1614.
- Informa Financial Intelligence. (2020a). EPFR Data Offering. Retrieved May 01, 2020, from [https://financialintelligence.informa.com/~/\\_media/informa-shop-window/financial/files/pdfs/epfr\\_data\\_offering.pdf](https://financialintelligence.informa.com/~/_media/informa-shop-window/financial/files/pdfs/epfr_data_offering.pdf)
- Informa Financial Intelligence. (2020b). EPFR Global Fund Data Collection. Retrieved May 10, 2020.
- Institute of International Finance. (2007). *Institute of International Finance: The first 25 years*. The Institute of International Finance.
- Institute of International Finance. (2020). *Capital flows tracker - April 2020*. The Institute of International Finance.
- International Monetary Fund. (2007). FAQs on Conversion from BPM5 to BPM6 (including FAQs on BPM6 Sign Convention). <https://www.imf.org/external/pubs/ft/bop/2007/bpm6faq.pdf>
- International Monetary Fund. (2009). *Balance of payments and international investment manual sixth edition*. International Monetary Fund.
- International Monetary Fund. (2014). *Balance of payments manual, sixth edition compilation guide*. International Monetary Fund.
- International Monetary Fund. (2015). *Coordinated direct investment survey guide*. International Monetary Fund.
- International Monetary Fund. (2017). *Coordinated portfolio investment survey guide*. International Monetary Fund. <http://data.imf.org/?sk=B981B4E3-4E58-467E-9B90-9DE0C3367363&slid=1481574691948>
- International Monetary Fund. (2020). Global Financial Stability Report: Markets in the Time of COVID-19. <https://www.imf.org/en/Publications/GFSR/Issues/2020/04/14/global-financial-stability-report-april-2020#Chapter1>
- Koepke, R. (2019). What drives capital flows to emerging markets? A survey of the empirical literature. *Journal of Economic Surveys*, 33(2), 516-540.
- Koepke, R., & Kunii, K. (2015). IIF Flows Alerts: An Early Warning System for EM Capital Flows. IIF Research Note. Institute of International Finance.
- Koepke, R., & Mohammed, S. (2014a). Introducing the IIF Portfolio Flows Tracker. IIF Research Note. Institute of International Finance.
- Koepke, R., & Mohammed, S. (2014b). Portfolio Flows Tracker FAQ. IIF Research Note. Institute of International Finance.
- Koepke, R. (2018). Fed policy expectations and portfolio flows to emerging markets. *Journal of International Financial Markets, Institutions and Money*, 55, 170-194.
- Lane, P. R., & Milesi-Ferretti, G. M. (2018). The external wealth of nations revisited: international financial integration in the aftermath of the global financial crisis. *IMF Economic Review*, 66(1), 189-222.
- Miao, Y., & Pant, M. (2012). Coincident Indicators of Capital Flows. IMF Working Paper no. 12-55. International Monetary Fund.
- Milesi-Ferretti, G.-M., & Tille, C. (2011). The great retrenchment: International capital flows during the global financial crisis. *Economic Policy*, 26(66), 289-346.
- Milne, A. (2014). *The control and management of international capital flows: A review of the literature*. Mimeo.

- Morris, S., Shim, I., & Shin, H. S. (2017). Redemption risk and cash hoarding by asset managers. *Journal of Monetary Economics*, 89, 71–87.
- Pradhan, S.-K., & Silva, J.-F. (2019). Using Mirror Data to Track International Banking. IFC Working Papers no. 19.
- Puy, D. (2016). Mutual funds flows and the geography of contagion. *Journal of International Money and Finance*, 60, 73–93.
- Shek, J., Shim, I., & Shin, H. S. (2018). Investor redemptions and fund manager sales of emerging market bonds: How are they related? *Review of Finance*, 22(1), 207–241.
- Vegh, C. A., & Vuletin, G. (2012). Overcoming the Fear of Free Falling: Monetary Policy Graduation in Emerging Markets, NBER Working Paper no. 18175. National Bureau of Economic Research.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Koepke, R., & Paetzold, S. (2024). Capital flow data—A guide for empirical analysis and real-time tracking. *International Journal of Finance & Economics*, 29(1), 311–331. <https://doi.org/10.1002/ijfe.2687>