Expanding the Landscape of Cross-Border Flow Restrictions: Modern Tools and Historical Perspectives*

Katharina Bergant
International Monetary Fund & CEPR

Andrés Fernández
International Monetary Fund

Ken Teoh
International Monetary Fund

Martín Uribe
Columbia University & NBER

December 29, 2025

Abstract

Employing large language models to analyze official documents, we construct a comprehensive record of daily changes in *de jure* restrictions on cross-border flows worldwide since the 1950s. Our analysis uncovers the wide array of instruments used to regulate cross-border financial flows and documents their evolving prevalence over the past seven decades. The fine granularity of the new measures allows us to characterize cross-country and time-series variation across eight categories of restrictions, further distinguishing by flow, direction, instrument type, and overall policy stance. We exploit the high frequency nature of the new data to document novel patterns in the use of these restrictions, as well as their relationship to crises, and political economy determinants. We validate our measures against established indicators of capital account regulation and show that our LLM-based classifications both replicate and substantially extend these benchmarks along multiple dimensions. Finally, we examine policymakers' stated motivations for adopting these restrictions and account for the intensive margin of these policy actions.

Keywords: Cross-border flows; Controls; Large language models

JEL codes: F32, F38, F41.

*We thank comments by Tobias Adrian, Chris Erceg, Domenico Giannone, Tarek Hassan, Anton Korinek, Matteo Maggiori, Sole Martinez-Peria, Tjoervi Olafsson, Dennis Quinn, Jesse Schreger, and feedback from participants in seminars held at the IMF and the NBER-SI. We especially thank Lucas Canteroz-Paz, Ariadne Maridena Checo de los Santos, Laura Costa de Carvalho, Eastina Deng, Michael Gottschalk, Balint Kidd, Ese Onokpasa, Manuel Perez, Sprihaa Singh, Chloe Tse, and Ashvik Viswanathan for excellent research assistance. The views in this paper do not represent the views of the IMF, its Executive Board, or IMF management.

1 Introduction

The onset of the Bretton Woods system after World War II ushered in an international monetary system (IMS) characterized by a panoply of restrictions to cross-border flows. Since then, the IMF has undergone several transformations and, along with them, the tools and the strength with which policymakers have put sand in the wheels of international finance have also evolved. An illustrative example of this was President Nixon's historic decision on August 15, 1971, to suspend the convertibility of the US dollar into gold, which was preceded by a wide array of measures aimed at curbing capital outflows, and limiting foreign credit and imports.[1] The US was not alone in imposing restrictions, with several countries implementing a range of financial restrictions aimed at controlling cross-border flows. The tide turned, however, in the 1980s and 1990s, when several countries embraced globalization and implemented market reforms and capital account liberalization, signalling a shift towards greater openness in trade and finance. The current debate around deglobalization and the reconfiguration of global supply chains has brought the topic of broader restrictions to cross-border flows back to the forefront, highlighting the potential for renewed restrictions on cross-border flows.[2]

This paper seeks to inform this debate by providing a granular, systematic account of cross-border flow restrictions spanning the last seven decades.[3] By harnessing recent advances in artificial intelligence (AI) and machine learning (ML), our analysis covers over 40,000 policy changes since the early 1950s across a broad set of countries, enabling us to trace the dynamics of cross-border restrictions with unprecedented granularity. This high-resolution account of policy measures allows us not only to examine the evolution of these measures but also to delve into the main determinants in the use of these restrictions and the underlying motivations for their imposition. The new techniques also allow us to refine the measures, taking into account the intensity with which they have been deployed, among other characteristics.

[1] Some of the measures implemented by President Nixon included the Interest Equalization Tax aimed at curbing capital outflows, and the Voluntary Foreign Credit Restraint Program which limited foreign credit. These measures were introduced in the previous Kennedy and Johnson administrations. Further import restrictions through the imposition of temporary surcharges were also added during the Nixon administration (Bordo, 2016).

[2] See for instance the Mar. 14 op-ed in the Financial Times by G. Tett arguing that "tariffs on goods may be a prelude to tariffs on money".

[3] These datasets will be made publicly available via the authors' website and may also be accessed here: BTFU Data. Further details of the datasets are provided in Appendix E.

The primary source for our analysis is the International Monetary Fund (IMF)'s Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER). The IMF has published the AREAER annually since 1950, in accordance with the provisions in its Articles of Agreement. It provides a detailed and comprehensive *textual* account of countries' restrictions on international transactions of all IMF member countries. As a result, it offers a valuable historical perspective on the evolution of cross-border flow restrictions across both advanced and developing economies. However, as the global financial system has grown more complex, so too have the measures reported in the AREAER. The restrictions covered span a broad range of areas — foreign exchange markets, capital flows, current account transactions, among others—and the increasing granularity of these measures has made the text both voluminous and complex, posing a formidable challenge to analyze it systematically with traditional methods. This is where modern machine learning techniques, LLMs in particular, prove especially useful. As they are particularly adept at processing and analyzing large volumes of unstructured text, this enables us to classify and organize the vast and diverse set of restrictions in the AREAER in a consistent way. These techniques allow us to tackle the complexities of the dataset and efficiently extracting meaningful insights. This, in turn, enables a more comprehensive, consistent, and replicable analysis of the evolution of cross-border flow restrictions.

To measure cross-border flow restrictions in a comprehensive and structured manner, we develop a methodology that organizes the restrictions that have been present in the AREAERs since the first edition into eight distinct categories. We focus on restrictions where a commitment to consistent reporting exists, as per the IMF's Articles of Agreement.[4] These categories include: (1) foreign exchange market, (2) arrangements for payments and receipts, (3) residents and non-resident accounts, (4) import payments, (5) export proceeds, (6) invisible transactions and current transfers, (7) capital account transactions,[5] and (8) provisions specific to the financial sector. We then aggregate all categories into an integrated Balance of Payments Flows

[4]A notable example of a category that is not mandated to be reported is trade restrictions, which our analysis excludes.

[5]Throughout the paper, we follow the terminology used in the AREAER, where "capital account" refers to transactions that correspond to the modern financial account in the balance of payments (e.g., credit operations, securities flows, direct investment, and real estate transactions). This differs from the narrower "capital account" definition in the BPM6 framework, which records non-produced, non-financial assets and capital transfers. For consistency with the historical AREAER classification, we retain the original labeling but interpret these items as financial-account transactions.

restrictions index (iBoP). The "*i*ntegrated" characteristic of our measure come from the fact that it encompasses the full spectrum of restrictions, from *direct* impediments to payments in current account transactions and flows in the financial account, to regulations that might *indirectly* impact these flows via restrictions in FX markets, in resident and non-resident accounts, and broader financial sector controls.[6]

Our approach allows us to further break down the complexities of the restrictions. In addition to categorizing them into the eight dimensions mentioned above, our methodology captures several key characteristics of the changes in regulation. These include the direction of change (whether the restriction was tightened or loosened), the type of restriction implemented (price-based, quantity-based, or administrative), whether restrictions are being applied to inflows or outflows, and the specific date of implementation, including the day/month/year. This detailed classification allows us to track not only the evolution of specific policy tools but also the dynamics of the regulatory stance over time, which, in turn, provides a comprehensive picture of the changing landscape of cross-border restrictions across different countries and periods.

Two separate versions of iBoP are built: one capturing *changes* (iBoP-C) and another one measuring the *stance* of overall restrictions (iBoP-S). iBoP-C quantifies changes in restrictions over time by aggregating tightening and loosening actions in the AREAER's Changes section, allowing for high-frequency tracking of policy adjustments from 1950 to 2022. It captures net changes at the country-category level at daily frequencies. In contrast, iBoP-S measures the overall level of restrictions by identifying the existence of restrictions in various subcategories from 1995 onwards. It averages binary indicators across subcategories to capture the existence of restrictions at the extensive margin. These two indices complement each other, with iBoP-C focusing on policy adjustments and iBoP-S offering a snapshot of the restrictiveness of the stance of a country. [7]

[6]The comprehensive and integrated nature of the various categories in iBoP imply that some restrictions may not be impacting cross-border flows directly. For instance, a tax on the purchase or sale of US dollars in FX markets may deter transactions between residents and hence not directly impact cross-border flows. However, it may indirectly deter residents from transacting with non-residents and vice versa. We believe that a comprehensive account of the various tools that have been deployed by policymakers must include these different types of restrictions. Since the measures allow for alternative aggregation, researchers will be able to build alternative measures of restrictions tailored to the research question at hand.

[7]As later explained, in one of the extensions of our work, we additionally offer an intensity-weighted measure of policy adjustments over time iBoP-C-intensity that assigns a score to each

Our machine-learning classification methodology follows a five-step procedure. First, a sample of AREAER narratives, drawn from a range of countries and time periods, is manually labeled according to pre-specified classification dimensions – such as the category of restrictions, direction of change, and flow direction – to generate training data. Second, these labels are reviewed and refined through iterative discussions, with assistance from LLM-generated predictions to resolve ambiguities and improve inter-annotator reliability. Third, the training data is augmented by incorporating post-1995 AREAER classifications for the category dimension where such labels are systematically available. Fourth, a domain-adapted language model is fine-tuned on this combined dataset to classify the full corpus of AREAER text. Finally, the model's output is evaluated against held-out samples to ensure performance and consistency. This approach reduces the reliance on manual classification, enhances scalability, and allows for consistent identification of restrictions across a large and evolving text corpus. Our baseline classification model, adapted to the text of the AREAERs, outperforms a number of other machine learning models across all dimensions of classifications. Furthermore, it also achieves a performance comparable to human annotators along several dimensions.[8] [9]

Our measures allow us to document several stylized facts about the long-run evolution of cross-border flow measures. First, liberalization trends in cross-border flow restrictions have not evolved linearly over time. While countries have overall liberalized their financial flows over the past seven decades, periods of significant change in the global monetary system, notably the breakdown of Bretton Woods, saw a significant tightening of restrictions for several categories, notably FX and financial market restrictions. Second, we document that financial liberalization has occurred at an uneven pace, with higher income countries liberalizing faster and to a greater extent than lower income countries. This pattern appears robust

narrative based on their subjectively-assessed impact on cross-border flows.

[8]For instance, in classifying direction of change, human annotators achieve an accuracy of 90.7 percent, which is comparable to our baseline model's (BERT-DAPT) accuracy of 87.0 percent. Our baseline model also outperforms other models across most classification tasks, including word-count and autoregressive generative models (OpenAI's GPT and Meta's LLama). See Section 4.2.4 for further details.

[9]An important methodological step is to exclude categories of restrictions that fall outside of the core reporting requirements of the AREAERs. These include restrictions related to national security, trade, and domestic macro-prudential requirements. While an analysis of these categories of restrictions are of interest in their own right, we find the reporting of these restrictions tend to be inconsistent across countries and over time. Hence, we apply our ML methodology to identify these restrictions in our sample and exclude them systematically from our database.

across all categories of restrictions that we examine. Interestingly, disaggregating by income, we find a more extensive use of current account restrictions (i.e., import payment and export proceed restrictions) to manage cross-border flows around the end of the Bretton-Woods regime among low-income countries. Third, we find that liberalization of cross-border restrictions have primarily concerned quantity-based restrictions. Administrative-based and price-based measures restrictions saw some loosening over time, but at a relatively more modest pace. Fourth, we document that the liberalization trend in BoP flows has been largely driven by the more rapid pace in loosening of outflow restrictions, relative to that of inflow restrictions.

A more granular decomposition of measures reveals notable patterns in various categories of restrictions. In particular, we further disaggregate the number of categories from 8 to 24 categories.[10] Measures related to the trade of gold and banknotes spiked around the end of the Bretton-Woods regime, alongside repatriation and surrender requirements. We also see a large number of tightenings in capital account measures related to direct investments, real estate transactions, and credit operations around that period. Finally, in the more recent period following the Great Financial Crisis, we find a tightening of restrictions on commercial banks but a loosening of restrictions on non-bank institutions.

The analysis of the iBoP-S, which quantifies the existence of restrictions along the extensive margin since 1995, shows broadly similar patterns in the evolution of restrictions in the more recent sample period. This set of measures significantly expands on the set of measures previously analyzed in Fernandez et al. (2016), from 32 categories in the capital account category, to more than 200 categories of restrictions across all eight categories of restriction.[11] In particular, we find significant heterogeneity in the stance of restrictions across categories, with current account (exports, imports, and invisible transactions) being the most open in general, while financial sector and resident/non-resident account restrictions being the least liberalized. We also observe a continued trend towards greater openness in the past two decades. Similar to patterns in the Changes index (iBoP-C), we find that higher income countries tend to be more liberalized.

[10] In doing so, we trade-off some of the classification performance for more information about the categories of restrictions in our dataset.

[11] This level of disaggregation is not available for the Changes Index, iBoP-C, as the AREAER does not report Changes measures at a similar level of granularity until 2016. The ability of the machine-learning approach to extend categorical classifications to a similar level of disaggregation is limited by the information presented in the narrative text. See Section 5.2 for more details.

We compare our newly constructed measures with those from previous studies, specifically focusing on the works of Quinn (1997), Chinn and Ito (2006), Fernandez et al. (2016) and Ilzetzki et al. (2021), which we also describe in the literature review. As part of these exercises, we also replicate the widely used dataset of Fernandez et al. (2016) using our LLM-based approach. Since that dataset was fully hand-coded and underwent quality control and peer-review, the fact that our measures correlate strongly with it provides an important external benchmark and increases confidence in the reliability of our methodology. Furthermore, our measures co-move with those in the other studies mentioned. However, our index captures additional dynamics, reflecting its more comprehensive coverage across eight distinct dimensions of restrictions. Most importantly, the new measures account for many more dimensions in the spectrum of cross-border flow restrictions, going beyond what previous studies have considered. Previous studies generally summarize restrictions with one number per country per year. We demonstrate how the trade-off between granularity and comprehensiveness can be effectively managed, offering a richer and more detailed view of BoP restrictions without losing broad coverage. This enhanced granularity allows for a detailed decomposition of the evolving trends in cross-border restrictions.

We also examine how countries use cross-border flow restrictions in practice. We show that measures are rarely used in isolation: authorities often "bunch" multiple tools on the same day and also "stagger" interventions over short windows, with over half of measures having a neighbouring action within ±10 days and over three-quarters within ±30 days—patterns. These events include all combinations of types of restrictions and therefore, complicate single-policy event studies. In a simple regression exercise, we find that the number of measures rise markedly during crises, roughly doubling relative to median annual use. The results are particularly driven by tightening measures and policies targeted towards outflows. Furthermore, the effects are concentrated in currency and sovereign debt crises. Finally, the use of these measures is negatively correlated with institutional quality, consistent with the political-economy forces shaping policy responses.

In a first extension of our work, we systematically study a subset of policy changes and hand collect official statements to classify the *motivations* put forth by authorities. We analyzed a number of measures deemed "macro critical" in the IMF's Taxonomy [12],

[12] We describe this smaller subset of measures in more detail in Section 3.2

which revealed six main (non-mutually exclusive) motivations behind the imposition of these controls. The most common reasons include a fear of disruptive outflows, where measures aim to prevent capital flight during crises, and the fear of floating, which involves managing exchange rate volatility or maintaining monetary policy autonomy. Fear of overborrowing, concerning excessive risk-taking due to large foreign inflows, is another motivation documented, while long-term goals focused on fostering financial market stability and increasing capital openness also feature in the mind of policymakers. Geopolitical considerations and miscellaneous reasons, such as tax revenue generation also appear, but are less frequent. Still, it is noteworthy that about one third of the motivations is not linked to the state of the business cycle. Low income countries have been more likely to impose controls due to concerns about disruptive outflows or floating exchange rates. High-income countries, on the other hand, have been more concerned with overborrowing, particularly through the use of inflow controls.

While iBoP-C improves upon existing measures by capturing the timing and direction of each regulatory action, it still treats all changes equally regardless of their economic significance. This limitation also applies to previous works such as Chinn and Ito (2006) and Fernandez et al. (2016). A notable exception is the work by Quinn (1997), which classifies measures according to a interpretive measure of intensity. Motivated by this, a second extension we explored is to incorporate *intensity* weights in the spirit of Quinn (1997) but in a systematic way using LLM-based classification. This allows us to classify the intensity of each measure—ranging from minor procedural adjustments to sweeping prohibitions—and incorporate these weights into a refined iBoP-C-intensity index. Qualitatively, we find the results from the intensity-weighted indices to be broadly aligned with those of the unweighted index. Quantitatively, however, we find the liberalization trends starting from the 90s to be less pronounced, particularly when compared to the tightening episodes in the 70s and 80s. This finding indicates that many of the measures used during the end of the Bretton Woods were more economically intensive than the subsequent loosening. Nonetheless, the finding of a broad liberalization over time continues to hold.

Our work is divided into nine sections, including this introduction. Section 2 provides a literature review highlighting our contribution relative to previous works. Sections 3 and 4 describe our data sources and the LLM-based methodology used to construct the new indices. Section 5 presents the main stylized facts on the evolution of cross-border flow restrictions since 1950. Section 6 analyzes how countries deploy these tools in practice, through documenting bunching, staggering, and crisis-related usage patterns. Section 7 validates our methodology by comparing the new indices with existing benchmarks. Section 8 presents the two extensions of our work — documenting policymakers' stated motivations for capital controls and incorporating intensity weights. Section 9 concludes.

2 Literature review.

Our work speaks to four separate strands of the literature. First and foremost, our work contributes to previous efforts devoted to constructing *de jure* measures of capital account restrictions. Three well-known studies that have covered a large cross-section of advanced and developing countries on a historical dimension are Quinn (1997), Chinn and Ito (2006), and Fernandez et al. (2016).[13] While the three studies all use the same primary source, the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER), they display important differences that are worth summarizing briefly in order to better highlight our contribution. [14]

The work by Quinn (1997) constructs indicators on capital account and financial current account regulations based upon coding of the narrative contained in the AREAER, with higher levels representing least regulated and most open regimes. It also aims at quantifying intensity in measures of capital controls through a subjective scale. The set of measures covers six categories: payment for imports; receipts from exports; payment for invisibles; receipts from invisibles; capital flows by residents; and by nonresidents. Measures of restrictions in Chinn and Ito (2006) are built with the first principle component of four binary variables from the AREAERs that capture the existence of multiple exchange rates, restrictions on the current account

[13]Other studies that have zoomed in on a subset of countries when building alternative measures are Pasricha (2012), Ahmed and Zlate (2014), Guisinger and Brune (2017), Pasricha et al. (2018), Binici and Das (2021), Acosta-Henao et al. (2025), among others. Two separate on-going studies by Baba et al. (2025) and Li (2025) also build measures on capital account restrictions based on the AREAERs. A separate work that captures *de facto* capital controls through the existence of parallel exchange markets is Ilzetzki et al. (2021). The work by Quinn (1997) has been subsequently expanded and refined in Quinn and Toyoda (2008) and Alesina et al. (2024).

[14]Two separate unpublished studies by Baba et al. (2025) and Li (2025) also build measures on capital account restrictions based on the AREAERs. While the former use the Yes/No status of restrictions to quantify the stance of restrictions in capital account, the latter extends the method in Quinn (1997) to account for intensity in these restrictions.

and capital account transactions, and the requirement of the surrender of export proceeds. Aggregated measures by country in these two studies are available on a yearly frequency from the 1950s and 1970, respectively. Importantly, neither of the two approaches distinguish across the direction or type of flows. The work by Fernandez et al. (2016) exploits a change in the AREAERs after 1995 where more information was added in an effort to reflect the increased complexity of capital control policies. This allows the authors to build more granular yearly measures that focus on capital account restrictions across different assets, while also disaggregating controls by the direction of flows and the residency of the economic agents.[15] This granularity, however, comes at the cost of not extending the coverage of restrictions to beyond those in the capital account and to years earlier than 1995.

Our work complements and extends these three studies along several important dimensions. First, our measures of cross-border flow restrictions are more comprehensive by covering the *entire set* of regulations in the AREAERs.[16] Concretely, our work adds the voluminous set of restrictions on flows that go beyond those related to the capital account (as in Fernandez et al. (2016)). Moreover, while Quinn (1997) and Chinn and Ito (2006) do consider some restrictions on current account flows, they do so only on a few selected items. Other dimensions that our work includes which are absent in these studies are FX market restrictions, regulation on payments and receipts, limits on residents and non-residents accounts, and measures specific to the financial sector. As our analysis documents, these are important levers that policymakers have used in practice, which ought to be included in a comprehensive historical account of cross-border restrictions.[17] Second, our work allows us to have much more granularity in the measures we build while covering the entire history in the AREAERs since 1950. This increased granularity materializes in the expanded set of eight categories of restrictions that we consider, as well as in the

[15] This work consolidates and extends previous work from Schindler (2009), Klein (2012), and Fernández et al. (2015).

[16] The only information from the AREAERs that we consistently leave out is that pertaining to exchange rate regimes. For comprehensive accounts of exchange rate regimes see Levy-Yeyati and Sturzenegger (2003), Reinhart and Rogoff (2004), and Ilzetzki et al. (2021). We will map our measures to some of these *de facto* exchange rate regimes that the literature has studied.

[17] As mentioned before, in a strict sense, some of the dimensions considered in our work do not *directly* impact cross-border flows, e.g. restrictions on the FX market or on the financial market, as they may entail transactions of assets across residents. Arguably, they may impact them in an *indirect* way. As data is provided on a granular basis, users can tailor the measures to reflect more direct measures impacting flows.

ability to classify the changes in restrictions by the type and direction of flow, among other characteristics that our methodology also accounts for. Moreover, the higher granularity provided by the daily frequency is a key improvement relative to the annual frequency used in all previous studies which often represents a limitation in empirical work.

Third, by netting out the number of tightening and loosening measures, we get a proxy for the intensity with which countries have enacted regulation through the use of these policy tools, which has been challenging to capture in most of the previous work.[18] Furthermore, by providing a disaggregation between price-based, quantity-based and administrative measures, we document how much *de jure* policies have relied (or not) on instruments that can be more prone to adjustment along an intensive margin, as price based instruments are. Fourth, the use of a systematic approach to classify measures through an LLM distinguishes from previous work by making our measures more consistent, efficient, and replicable by others. They also make our measures much more adaptable to future updates and refinements.

The second strand of literature to which our work contributes is the one studying the use, motivation and effects of capital account restrictions. Fernández et al. (2015), Eichengreen and Rose (2014) and Acosta-Henao et al. (2025) describe a sticky use of capital controls, while Pasricha (2022) characterizes their deployment though a policy reaction function. The literature is much more scarce in documenting empirically the motivation for their use. Magud et al. (2018) is an exception, providing anecdotal evidence on the "fears" that have driven policymakers to resort to these measures. Our work complements and extends their work through a systematic analysis of country authorities' official documents, expanding the set of motivations that official records provide. While a large part of our results point to short-term rationales for authorities to implement controls, we also show that policymakers can have a longer-term view of capital controls. In particular, we find a substantial share of controls are adjusted to achieve long-term liberalization objectives. This aligns with our finding on a longer historical sample of controls that show cross-border restrictions have liberalized over time.

[18] Netting out may be an imprecise measure for intensity because there may be measures that are more restrictive than others. Acosta-Henao et al. (2025) studies direct measures of intensity in capital controls for selected EMs, but warns that doing so may bias against only a few price-based measures. This critique that does not apply to our measures as they encompass both price based and non price based regulations. In addition, we do offer an intensity-based measure of the changes in an extension of our work in Section 8.2.

De jure measures of capital controls have often been used to examine the impact of controls on financial stability and real exchange rates, among others, resulting in a voluminous empirical literature recently surveyed in Rebucci and Ma (2020), Erten et al. (2021) and Bianchi and Lorenzoni (2022).[19] Overall, the evidence surveyed on the real effects of controls has been mixed, with some finding positive effects on financial stability and reductions in real exchange rate pressures, while others finding no effects. The lack of consensus is largely a consequence of the considerable challenges in obtaining a clear identification of such effects. The challenges arise for various reasons. On the one hand, the timing of capital controls may not be properly captured by annual indexes as those mentioned above. Whilst such slow- moving indexes may capture whether restrictions are present or not (extensive margin), they can miss variations in the intensity of restrictions over the business cycle. On the other hand, there may be reverse causality and endogeneity problems, as well as omitted variables that simultaneously have an impact on both the use of capital controls and the fluctuations in the variables of interest.

In order to address some of these challenges, studies have resorted to several strategies. In a study that uses meta-analysis, Magud et al. (2018) find that capital inflow controls tend to be more effective relative to outflow controls in altering the composition of capital flows toward longer-term flows and reducing real exchange rate appreciation. A number of other studies have resorted to using higher-frequency data that better capture time variation in capital controls, at the cost of focusing on a relatively smaller sample of countries and/or time. Using quarterly data of 50 emerging market economies from 2005 to 2013, Ghosh et al. (2017) analyze the set of policy instruments that countries employ to respond to capital inflow surges. Using quarterly information from local press releases and news bulletins for 19 emerging market economies from 2002 to 2013 and adopting a fixed-effects model, Ahmed and Zlate (2014) find that capital controls have a significant negative impact on both total and portfolio net inflows. Using monthly data from 2000 to 2008 for five EMs, Baba and Kokenyne (2011) find that capital controls are associated with a temporary decline in capital inflows and a lengthening of maturities. In an effort to directly address endogeneity problems, Erten and Ocampo (2017) use binary variables on

[19] The literature has also aimed at quantifying the impact of capital controls on additional variables such as long term growth, output and consumption volatility, real exchange rate effects, monetary policy independence, and international spillovers. Another useful survey covering broader macro prudential policies is Bianchi and Mendoza (2020).

whether countries have bilateral investment treaties with the United States or whether they signed the EU membership agreement as instruments for whether countries use various capital control measures. They find that capital controls result in a reduction in real exchange rate appreciation. In what is perhaps to highest frequency analysis until now, using weekly frequency data for 60 countries from 2009 through 2011 and employing a propensity-score matching method to address selection, Forbes et al. (2015) find that capital flow measures have a robust effect on reducing financial vulnerability but find no evidence of a significant impact on net capital inflows or real exchange rates.[20] We contribute to this strand of the literature by making use of the novel data that we construct and, in particular, exploiting the daily frequency that allows for a cleaner identification. To the best of our knowledge, we are the first to use this higher frequency when documenting the use of controls on BoP flows. We also improve upon previous work by considering the wide array of restrictions that our granular data account for.

A third strand of the literature that our work speaks to is the recent new and rich setups that consider the use of capital controls in conjunction with several other policies. In particular, recent analytical frameworks developed in Basu et al. (2020) and Adrian et al. (2021) have explored within an integrated policy framework, how capital control policies can be combined with an array of other tools such as macroprudential instruments as well as interventions in FX markets. Our work complements these efforts by providing empirical evidence of how policies across many dimensions of the BoP flows have intertwined. Future studies can build upon our work to continue to further expand the set of policies within an integrated framework.[21]

A fourth strand of literature to which our work contributes is the rapidly growing use of large-scale NLP and AI methods to extract structured economic information

[20] Following the seminal work by Forbes (2007), an increasing body of work is resorting to microdata to assess the effects of capital controls - see the works of Alfaro et al. (2017), Andreasen et al. (2024), Fabiani et al. (2022), and Fabiani et al. (2025).

[21] A recent complementary contribution is Capelle and Pellegrino (2025), who develop a dynamic spatial general-equilibrium model to back out country- and time-varying "Revealed Capital Account Openness" wedges from observed cross-border portfolios. Their wedges summarize all impediments to international investment—including capital controls but also political risk, financial development, and taxation—into an implicit shadow tax consistent with actual investment patterns. While this provides a *de facto* measure of financial openness, our LLM-based indicators instead capture the *de jure* policy stance at a high level of granularity and frequency. Taken together, these approaches help illuminate the gap between formal capital account regulation and the realized degree of international financial integration.

from unstructured policy texts or to measure policy actions. Clayton et al. (2025) make use of LLMs to classify geoeconomic pressure episodes across countries. Similar applications are Ottonello et al. (2024) on firms' political speech and Fang et al. (2025) on decoding China's industrial policy. Aruoba and Drechsel (2024) show that applying machine-learning techniques to internal FOMC documents can recover monetary policy shocks from the Fed's information set. At a broader methodological level, Korinek (2025) emphasizes how generative AI and AI-agent systems can transform the research workflow—from data extraction to classification and validation—and Dell (2024) provides a comprehensive survey of deep-learning methods for economists, highlighting the gains from moving beyond traditional text-processing tools. Our contribution builds directly on these advances by developing a domain-adapted LLM pipeline tailored to the semi-structured regulatory text in the AREAERs, enabling high-frequency, fine-grained measurement of de jure cross-border flow restrictions at global scale.

3 Sources

3.1 The Annual Report on Exchange Arrangements and Exchange Restrictions

Our primary data source on financial and exchange restrictions is the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER). First published in 1950, the AREAER is among the IMF's oldest and most comprehensive textual datasets, compiled annually in accordance with the IMF's Articles of Agreement. The report systematically cover a broad range of restrictions related financial flows associated to current account transactions, international payments and transfers, foreign exchange market operations, capital and financial account transactions, and regulation on the financial sector related to cross-border flows.

Data sources. Data from the AREAER can be accessed through two main websites. Reports from 1999 onward are publicly available for download via structured queries on the IMF's AREAER website.[22] Reports covering earlier years (1950–1999) are available as archived PDF files from the IMF e-library.[23]

Figure 1 depicts the front cover and first page of the inaugural 1950 AREAER

[22]Restrictions can be downloaded by year, country, and type of restriction from the IMF's AREAER

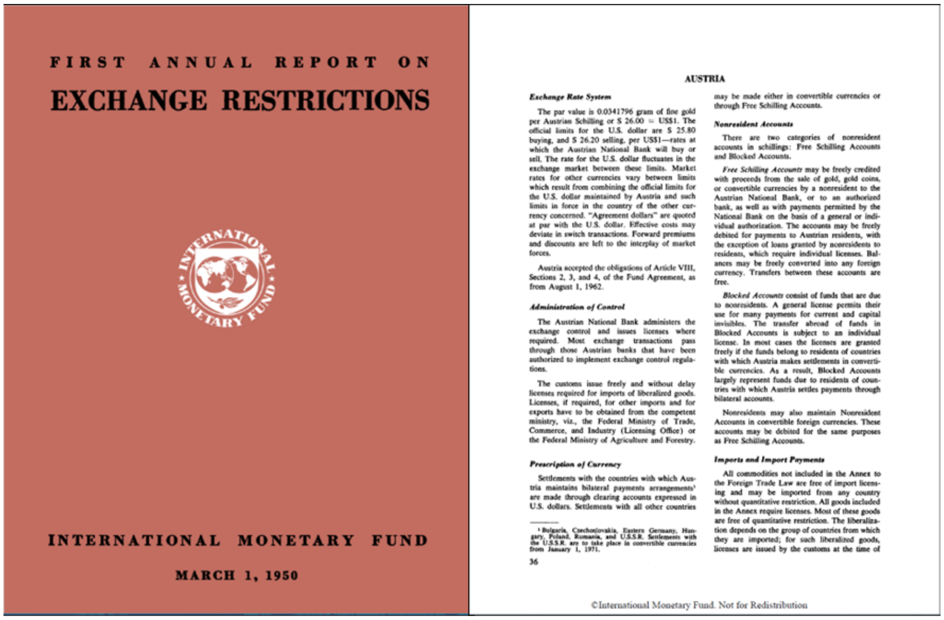


Figure 1: Front cover and first page of the inaugural 1950 AREAER report.

report, illustrating the format in the early editions. Such format differs significantly from those downloadable on the IMF's AREAER website for the post-1999 period. In particular, the former format described restrictions in an extended narrative, without much classification into subcategories.[24] The changing formatting of the AREAERs in the earlier years thus requires careful effort to collect and organize the data into a machine readable form.

To construct our dataset, we downloaded all historical PDF reports from 1950 to 1999 and extracted textual information on restrictions using Optical Character Recognition (OCR) software (Google Tesseract). In cases where sections were difficult to delineate clearly from OCR outputs, we applied a combination of large language model (LLM) prompts and manual verification to parse and clean the data. This processed historical dataset was then combined with structured data from 1999–2022, directly downloaded from the AREAER website.

Coverage. An examination of the early editions of the AREAER underscores the IMF's responsibility to monitor and report cross-border related restrictions under its Articles of Agreement. For example, the preface of the inaugural 1950 report

website: https://www.elibrary-areaer.imf.org/Pages/ChapterQuery.aspx

[23]These annual reports can be downloaded at https://www.elibrary.imf.org.

[24]An important exception is the Changes section of the AREAER, which has maintained a similar consistent format over time.

states: "*We recognize the many difficulties which have caused the maintenance and spread of restrictions in international economic relations during the period of postwar transition, but if trends of improvement in the underlying world conditions continue, members should find the task of removing exchange restrictions less difficult in the future*" (IMF, 1950). The same report also explicitly mentions the IMF's commitment to annual reporting: "*Not later than three years after the date on which the Fund begins operations and in each year thereafter, the Fund shall report on the restrictions still in force*" (IMF, 1950).

The restrictions covered in the AREAER relate to capital and financial account restrictions as well as the obligation of member countries in Article VIII, which includes restrictions on making payments and transfers for current international transfers, discriminatory currency arrangements or multiple currency practices, and restrictions on the convertibility of foreign held balances. They also cover controls on capital transfers (Article VI).

AREAER reports rely on inputs provided annually by member-country authorities through standardized questionnaires. The number of IMF member countries, and consequently the coverage of AREAER reports, has significantly expanded over time, reaching 197 countries in recent reports.

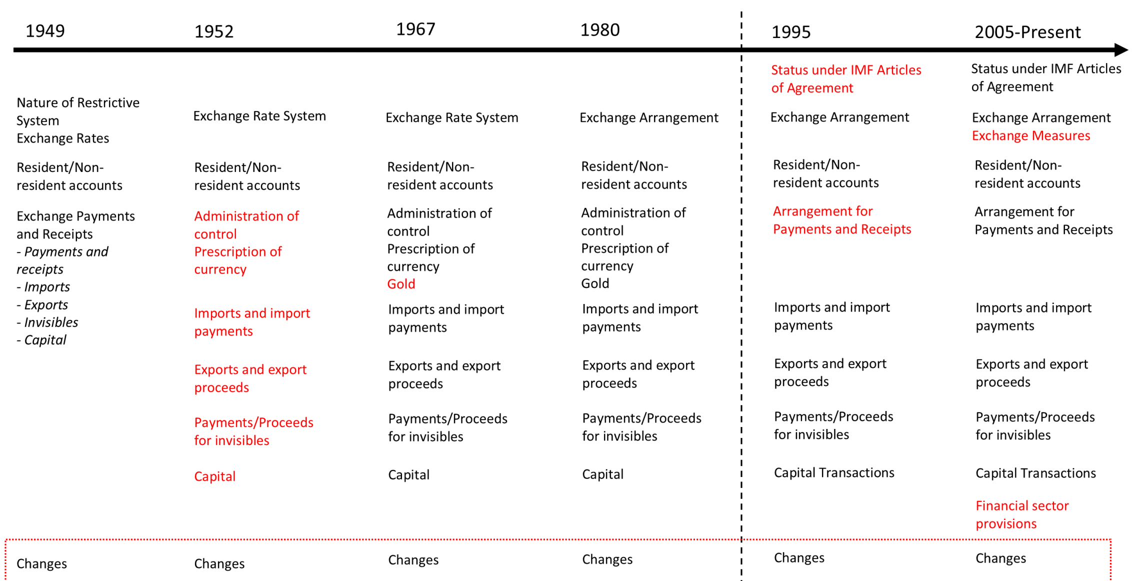
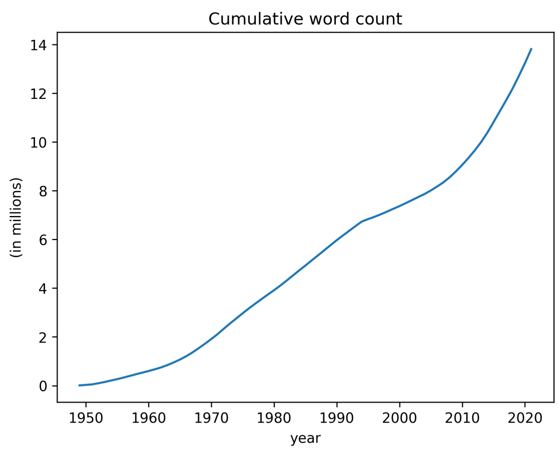


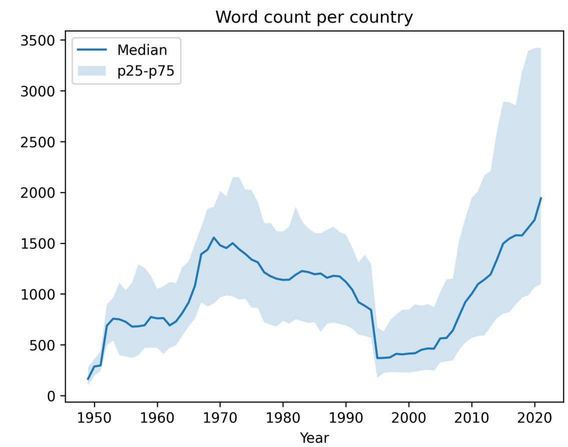
Figure 2: Sections in AREAER reports over time.

Structure of AREAER reports. AREAER reports feature a rich and evolving structure, with the current editions comprising more than 14 distinct sections. Figure

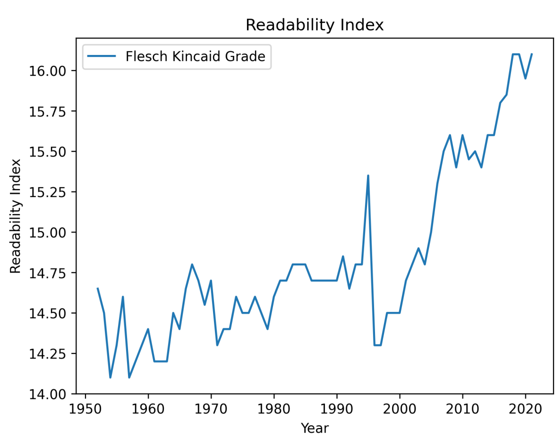
2 provides a historical overview of the evolving structure of these reports. Many core sections—such as those that refer to restrictions on payments and receipts, import payments, export proceeds, invisibles, and capital/financial transactions—have existed consistently since the early editions, albeit with changes in their classification and organization. Notably, the section explicitly covering policy changes has been present since the first report in 1950 (highlighted in the bottom of Figure 2), forming the basis for our iBoP-C construction.



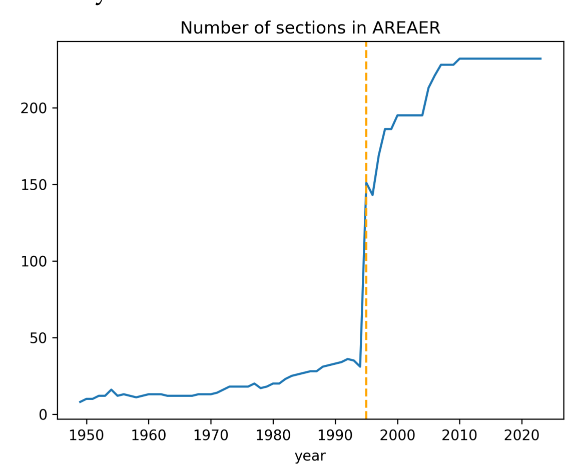
(a) Cumulative number of words in all AREAER reports.



(b) Number of words in AREAER reports per country.



(c) Readability index of AREAER reports over time.



(d) Number of sections in AREAER reports over time.

Figure 3: Overview of structure and content of AREAER reports over time.

Figure 3 summarizes trends in the structure and content of AREAER reports. Over the decades, reports have grown significantly in both volume and complexity. By

2022, the total cumulative text reached approximately 14 million words (Panel A). At the country level, the length of each report's section has also expanded considerably in recent decades (Panel B). Additionally, text complexity has increased, with recent reports requiring college-level comprehension or above (Panel C). A back-of-the-envelope estimate suggests that reading the entire historical content would require roughly 14 months of continuous reading.

An important structural change occurred in 1995, with the text going from being largely unstructured to being semi-structured, with the number of sections expanded significantly (Panel D). These changes in the structure of the reports make machine learning methods much more suitable to the analysis and ensure efficiency and consistency.

3.2 The Capital Flow Management Taxonomy

One of the extensions of our work will document the motivations underlying capital flow management (CFM) measures. For this particular analysis, we draw on a complementary source: the IMF's Taxonomy of Capital Flow Management Measures. This publicly available source is a much smaller subset of the measures outlined in the IMF's AREAER.[25] There are a few important characteristics of these measures. First, they have been assessed by IMF staff as being macro-critical, i.e. they limit capital flows such that they significantly influence a country's present or prospective domestic or balance of payments stability. Second, and importantly, the IMF taxonomy encompasses measures discussed in published IMF staff reports since the adoption of the Institutional View on the Liberalization and Management of Capital Flows in 2012. Thus, it is important to note that, while the IMF Taxonomy includes changes to capital flow policies, it does not reflect the full stock of measures determining a country's capital account openness. Instead, it is focused on identifying new measures introduced after the adoption of the Institutional View. For measures that predate the Institutional View, the taxonomy includes only assessments of their recalibration—whether through tightening, easing, or adjustments to enforcement—and their subsequent removal when deemed necessary. If a measure was implemented before 2012 and has not been recalibrated it will not feature in the taxonomy. Given our interest in documenting the motivations behind the use of

[25] The dataset can be accessed online through the IMF Data portal (https://www.imf.org/en/Data).

CFMs, it seems natural to start our analysis on the introductions of CFMs that have been assessed as macro-critical by others.[26]

Appendix Figure A1 presents the distribution of CFMs in the IMF's taxonomy. There are a total of 537 changes recorded, distributed as follows: 153 introductions, 111 tightenings, 222 eases, 49 removals, and 2 extensions. Our analysis on motivations will focus on the 153 introductions of new measures.

4 New Measures of Cross-Border Restrictions

Available measures of cross-border restrictions in the literature have often focused exclusively on policies that directly affect the capital account (e.g. Fernandez et al., 2016) or have only partly included restrictions in current transfers that are not disaggregated from those in the capital account (Chinn and Ito, 2006). A broader multi-dimensional approach that comprehensively accounts for the various types of measures in the AREAER that countries have employed is thus warranted.

There are at least three challenges that arise when trying to accomplish this goal. First, what categories of restrictions should one study? Second, given the large volume of text in the AREAER reports, how can this large amount of text be processed in a systematic way? Third, how can one satisfactorily account for changes in the structure of the AREAER, in a way that allows for a consistent measure of current and capital account restrictions over time and across countries? The next subsections will describe how we address these challenges.

4.1 Defining categories of restrictions

We begin with a discussion of the categories of restrictions we study. To guide our selection of categories, we focus on restrictions that affect current and capital accounts in the Balance of Payments. Starting with the 13 sections in the AREAER, we group similar topics in the AREAER into 8 relevant categories of restrictions: FX markets, arrangements for payment and receipts, resident and non-resident accounts, import payments, export proceeds, invisible transactions and current transfers, capital account transactions, and provisions specific to the financial sector.[27]

[26]See Binici and Das (2021) for an in-depth analysis of the measures in the IMF's CFM taxonomy.

[27]The mapping between the 2023 AREAER and the categories is as follows. Measures on FX markets come from Sections II and III.F-H; arrangements for payment and receipts come from Section IV;

The categories of restrictions are defined as follows:

1. **FX markets**
   Includes (1) exchange restrictions and multiple currency practices (MCPs) maintained by a member country, (2) foreign exchange transactions subject to a special tax, fees, or other mandatory cost, (3) foreign exchange transactions subsidized using separate, non-market exchange rates, (4) restrictions imposed on foreign exchange markets. We exclude measures related to exchange rate arrangements or monetary policy frameworks.[28]

2. **Arrangement for payments and receipts**
   Includes (1) official requirements affecting the selection of currency and the method of settlement for transactions with other countries, (2) agreements that prescribe specific rules for payments to each other, including cases in which private parties are also obligated to use specific currencies, (3) separate rules for trading in gold with foreign countries, (4) regulations governing the physical movement of means of payment between countries. Because they do not relate to cross-border flows, we exclude from our analysis text referring to arrangements related to the use of foreign exchange among residents, the administration of control, and controls on domestic ownership of currency and gold.

3. **Residents and non-resident accounts**
   Indicates (1) whether resident accounts that are maintained in the national currency or in foreign currency abroad are allowed and describes how they are treated and the facilities and limitations attached to such accounts, (2) whether local nonresident accounts maintained in the national currency or in foreign currency are allowed and describes how they are treated and the facilities and

resident and non-resident accounts come from Sections V and VI; imports and import payments come from Section VII, exports and export proceeds come from Section VII; invisible transactions and current transfers come from Sections IX and X; capital account transactions come from Section XI, and provisions specific to the financial sector come from Section XII. Since 1950, however, Sections have varied in number and scope (see 2). Within these categories, we retain all relevant measures except for restrictions that are not directly related to cross-border flows or fall outside the IMF's reporting mandate. See Section 4.3 for further details.

[28] For comprehensive treatments of exchange rate arrangements and monetary policy regimes, see Levy-Yeyati and Sturzenegger (2003), and Ilzetzki et al. (2021). These works provide de facto classifications of exchange-rate and monetary frameworks that complement the de jure restrictions analyzed in this paper.

limitations attached to such accounts. Because they do not relate to cross-border flows, we exclude restrictions on resident accounts held domestically.

4. **Import payments**
   Describes the nature and extent of exchange and trade restrictions on import payments. Includes: (1) foreign exchange budgets, (2) financing requirements of imports, and (3) documentation requirements for the release of foreign exchange for imports. We exclude restrictions that directly impact the trade of goods and services (e.g., import licenses, taxes, and tariffs).[29]

5. **Export proceeds**
   Describes restrictions on the use of export proceeds. Includes: (1) repatriation and surrender requirements for exporters, (2) export financing requirements, and (3) export documentation requirements. As before, we exclude restrictions that directly impact the trade of goods and services (e.g., export licenses and taxes).

6. **Invisible transactions and current transfers**
   Describes (1) procedures for effecting payments abroad in connection with current transactions in invisibles, with reference to prior approval requirements, the existence of quantitative and indicative limits, and/or bona fide tests, and (2) regulations governing exchange receipts derived from transactions in invisibles, including descriptions of any limitations on their conversion into domestic currency and the use of those receipts.

7. **Capital account transactions**
   Describes regulations influencing both inward and outward capital flows. Includes (1) repatriation and surrender requirements, (2) controls on capital and money market securities, (3) controls on credit operations, (4) controls on direct investment, (5) controls on real estate transactions, (6) controls on personal capital transactions.

8. **Provisions specific to the financial sector**

[29] While the AREAER does report a number of trade-related measures, we exclude them from our baseline categories. A detailed review of the historical reports revealed that these trade measures are likely recorded inconsistently across countries and over time. As noted in the Introduction, this reflects the fact that the IMF's Articles of Agreement mandate consisted of reporting only for balance-of-payments–relevant financial flows, not for trade policy measures.

> Describes (1) regulations specific to commercial banks and other credit institutions, such as monetary, prudential, and FX market restrictions, (2) controls specific to institutions, such as insurance companies, pension funds, investment firms (including brokers, dealers, or advisory firms), and other securities firms (including collective investment funds). When classifying restrictions in this category we include those that pertain to cross-border flows, and exclude primarily *domestic* macro-prudential policy measures, including restrictions on local FX lending, purchases of locally issued securities, treatment of FX deposit accounts, reserve requirements, liquid asset requirements, interest rate controls, and credit controls.[30]

In choosing these eight categories of restrictions, we exclude certain parts of the AREAER: detail information about a country's status with the IMF (Section I in the 2023 AREAER vintage), hence is not relevant to cross-border restrictions; description of a country's monetary policy framework and exchange rate regimes (Sections III.A to III.E), which is not a focus of our measures; and restrictions on digital currencies and crypto assets (Section XIII), added only in 2022.

The comprehensive and integrated nature of the various categories aims to capture not only restrictions that impact cross-border flows directly, but also those that have more indirect implications. For instance, a tax on the purchase or sale of US dollars in FX markets, even if not directly targeted at minimizing cross-border flows, may end up deterring deterring residents from transacting with non-residents and vice versa. Likewise, these restrictions include taxes to imports, or on the financial flows linked to these transactions that indirectly impact cross-border flows through their effect on trade. As the granularity of our measure allows for alternative aggregation, researchers can built alternative measures of restrictions tailored to the research question at hand.

4.2 A machine learning approach to classifying measures

Our methodological approach to classifying AREAER restrictions leverages recent tools in natural language processing, particularly Large Language Models (LLMs).

[30] A section explicitly gathering provisions to the financial sector appears in the AREAER only from 1995. This topic, however, is pervasive since the start of the AREAERs. The LLM is able to focus on substance of the category classified despite it not appearing explicitly in the AREAERs before that year.

The decision to adopt a machine learning-based classification methodology stems from two primary challenges inherent in constructing a consistent and extensive dataset from textual narratives.

The first challenge is maintaining consistency and efficiency in manual annotation. With traditional manual annotation, differences inevitably arise due to subjective interpretations by annotators, potentially leading to inconsistencies in classification. Additionally, there are substantial changes in the AREAER structure and coverage over the seven-decade span from 1950 to 2022. This may require adjusting rules for classification to accommodate differences in the formatting of the narratives. Manually annotated labels would need to be revisited every time an adjustment is made, which can be challenging to do consistently in a large textual database. The second challenge is classifying narratives at scale. As previously noted, the amount of text describing restrictions is significant and manual annotation is more prone to errors when analyzing large quantities.

To address these challenges, we leverage large language models (LLMs) to automate and enhance the annotation process. Our methodology proceeds in several clearly defined steps. First, we generate a carefully annotated training dataset through manual coding of a sample of AREAER narratives drawn from a broad range of countries and years. Second, we validate and refine the training dataset through iterative discussions among annotators, supplemented by LLM-generated predictions, to resolve ambiguities and reduce inter-coder discrepancies. Third, we further augment this manually labeled dataset with publicly available classifications from recent AREAER reports. Incorporating existing AREAER labels significantly expands the size of training data, improving the accuracy and robustness of our classification models. Lastly, we employ LLMs trained on this training dataset to systematically classify narratives from all AREAER reports.

4.2.1 Step 1: Manual Annotation of Labels

In the initial phase, human annotators label a sample of observations from the AREAER Changes section across five dimensions:

1. **Direction of Change (tighten / loosen / neutral)**
   Indicates whether the measure tightens or loosens capital flow restrictions.

Measures that do not affect existing restrictions, or where the direction is unclear, are labeled as "neutral."

2. **Direction of Flow (inflow / outflow / neutral)**
   Specifies whether a policy measure affects capital inflows, outflows, or both. If the measure impacts both inflows and outflows equally, or its focus cannot be determined, the label is "neutral."

3. **Category of Restriction**
   Identifies which of the eight categories is being restricted, such as FX markets, payments and receipts, resident/non-resident accounts, import payments, export proceeds, invisible transactions and current transfers, capital account transactions, or financial sector restrictions. A single measure may affect multiple categories. [31]

4. **Type of Restriction (price / quantity / administrative)**
   Classifies whether the restriction uses a price-based instrument (e.g., taxes, duties, fees), quantity-based instrument (e.g., limits, quotas, surrender or repatriation requirements), or an administrative instrument (e.g., licensing, approval requirements). Measures can in principle combine all three types of restrictions.

5. **Numerical Information (numerical / non-numerical)**
   Indicates whether the textual description of the measure contains explicit numerical information, such as a limit on price or duration.

4.2.2 Step 2: Review of Manual Annotations

We investigate the reliability of our manual annotations by comparing the labels assigned by two trained humans from the pool of annotators.[32] This analysis focuses on observations from the pre-1994 sample period, where the AREAER reports are lengthier and potentially more difficult to label consistently.

[31] Note that category labels are systematically available only from 1995 onward, though partial data exists for earlier years. Because detailed category labels exist in the AREAER database only from 1995 onward, "category of restriction" is classified using a machine-learning approach only in the pre-1995 sample. For post-1995 observations, we use official category labels provided in the AREAER reports.

[32] The pool of annotators comprised the set of co-authors and a pool of research assistants from Columbia University with prior experience labeling restrictions in the post-1995 sample period. For several months, the group would meet weekly to discuss manual classifications, ensuring consistency in classification.

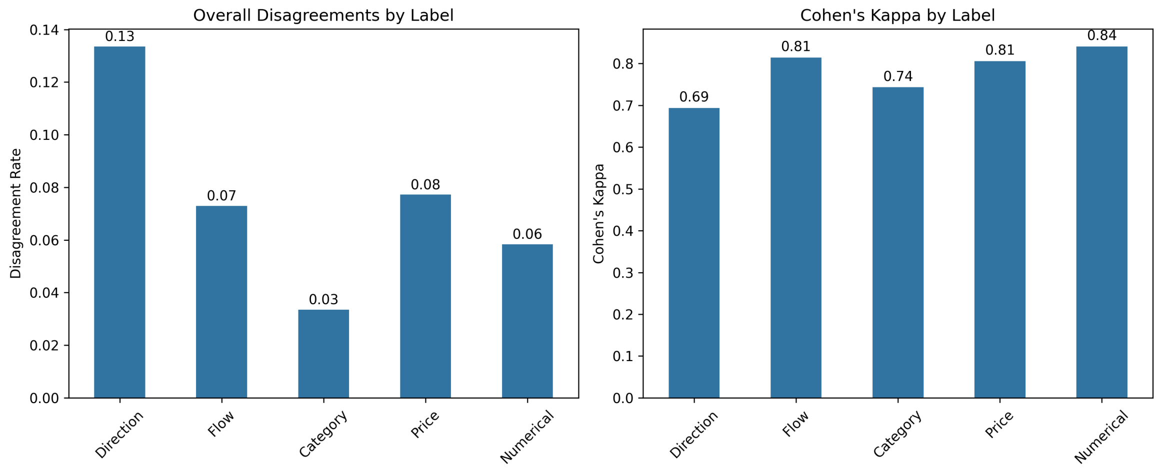


Figure 4: Disagreement rate between annotators (left) and Cohen's kappa (right) by dimension. The sample consists of observations from the pre-1994 AREAER sample.

Figure 4 displays two measures of agreement between the two annotators. In the left panel, the disagreement is calculated as the number of observations on which annotators assign different labels divided by the total number of observations.[33] These results show that labeling direction of change (tighten, loosen, or neutral) appears the most challenging, with a disagreement rate of 13%. The second most difficult dimension is type of restriction (price vs. non-price), which shows a 7% disagreement rate. By contrast, category of restriction has the smallest disagreement rate (3%).

In the right panel of Figure 4, we report Cohen's kappa for each dimension. Unlike simple disagreement rates, Cohen's kappa adjusts for the likelihood of chance agreement, which becomes particularly relevant if labels are imbalanced in the sample. Formally, for each dimension we calculate

$$\kappa = \frac{P_o - P_e}{1 - P_e},$$

where P_o is the observed agreement rate (i.e., the proportion of instances on which annotators agree), and P_e is the expected agreement given each annotator's individual label frequencies. Averaging kappa over all labels within a dimension yields values ranging from 0.69 to 0.84. Values greater than 0.60 reflect substantial agreement, and values exceeding 0.80 represent close to full agreement (Landis and Koch, 1977).

[33]Equivalently, we can define disagreement rate as $1 - \frac{\text{no. observations where annotators agree}}{\text{total observations}}$.

Consequently, these results suggest that the manual annotations demonstrate a generally high degree of reliability, even in the more ambiguous pre-1994 setting.

Refining Annotator Discrepancies. Despite these relatively strong reliability metrics, some differences inevitably remain, particularly for dimensions such as direction of change. To address these discrepancies, we follow a two-part approach. First, any observation on which the two annotators disagree is jointly re-examined by at least one co-author with domain expertise. This step resolves discrepancies that may be due to a lack of clarity of the policy context of the AREAER entries. Second, we use a large language model (LLM) to predict labels and generate an explanation for its choices.[34] We compare these model predictions to the refined manual labels and resolve any remaining discrepancies via discussion.

Appendix Figure A3 describes the extent to which this reconciliation process reduces disagreements between annotators and the LLM. It shows that discussions between annotators often lead to a lower overall disagreement rate between them. Furthermore, revisiting their predictions after comparing them with the LLM's predictions also leads to further reduction in the disagreement rate. This suggest that combining a manual labeling approach with an assistive LLM can improve overall accuracy of manually annotated data.

Across labels, disagreement rates decline most markedly for the direction, with disagreements often resolved in favor of the model's predictions once annotators and domain experts revisit the text. By contrast, the category of restriction dimension shows only minor improvements occur after introducing model predictions for this dimension. This indicates that the LLM's predictions for this dimension, absent further finetuning, yield less accurate labels than the expert-reviewed annotations. These are further documented in Appendix Figure C1.

4.2.3 Step 3: Augmented Training Data

To develop a comprehensive training set, we combine the manually annotated labels with pre-existing labels from the post-1995 AREAER:

- Manual Annotations: The primary set of hand-labeled observations, covering all five dimensions described above. This consists of between 2000 observations

[34] We do not provide the model with any examples of manually coded labels at this stage (zero-shot prompting). The prompts used for each label dimension are available upon request.

(price / numerical) to 3000 observations (direction / flow) per label.[35]

- Post-1995 Category Labels: For observations after 1995, labels for categories of restrictions are already present in the AREAER.[36] We incorporate these as ground truth for that single dimension, thus expanding our labeled dataset.
- Supplementary Sections (III.A–III.E): We also include observations from post-1995 AREAER sections that do not directly relate to capital flows (e.g., exchange arrangements). These observations are labeled "neutral" for the direction and flow dimensions.

We merge these sources into a single dataset and split it into training and testing subsets. In particular, 80% of the labeled examples go to training, while the remaining 20% form a hold-out set for unbiased performance evaluation.

Appendix Table A3 shows the distribution of labeled observations across the five classification dimensions—Direction, Flow, Category, Price, and Numerical. Except for the category label, the training sample for all other labels consist of between 2000 to 3000 training examples, which consists of between 3 to 4 percent of the full sample of observations. For category labels, we additionally incorporate the full set of post-1995 narratives as examples, given that category labels are already systematically recorded in the Changes section. This brings the total number of training examples for the category label to 35 percent of the full sample. Appendix Figures C3–C7 show the distribution of labels across various characteristics of observations. The share of labeled observations tend to be higher in the post-1995 AREAER sample period, in part because manually annotated observations are supplemented with additional observations of labels from the AREAER. Notwithstanding, the share of labeled observations broadly includes observations across different income groups and region.

4.2.4 Step 4: Training of large language models

We focus on a Domain-Adapted BERT model as our primary approach for the final prediction tasks. However, to contextualize its performance, we also consider a range of other classifiers for comparison.

[35] Manually annotated labels for category of restriction are replaced with labels from the AREAER in the post-1995 sample.

[36] Labels for other dimensions (i.e., direction of change, direction of flow, type of restrictions, numerical information) are not present in the AREAER.

Domain-Adapted BERT Although large language models (LLMs) can exhibit strong generalization, recent studies suggest that pre-training on domain-specific corpora can yield significant performance gains (Gururangan et al., 2020). Given the specialized nature of the AREAER corpus, we continue pre-training a BERT model (110 million parameters) on narratives of restrictions extracted from the entire collection of AREAER reports (1950–2022), including both stance and changes sections. Through domain-adaptive pre-training, the embeddings are specialized to AREAER-specific terminology and context, which can potentially improve classification accuracy.[37]

Comparative Models In addition to Domain-Adapted BERT, we evaluate the following alternative methods to gauge relative performance:

- Word-Count Models: TF-IDF weighted representations with classifiers, including logistic regression, random forests, and support vector machines.
- Open-Source Generative Models: We also include unmodified (i.e., off-the-shelf) BERT (438 million parameters) and Llama 3.2 (3 billion parameters) to illustrate how they perform without further adaptation to the AREAER domain.
- Closed-Source Generative Models: Frontier systems like OpenAI's GPT are tested via in-context (few-shot) learning.

Supervised Fine-Tuning After further pre-training, Domain-Adapted BERT is fine-tuned on our labeled training set. We train separate linear classification heads for each dimension (direction, flow, category, price, numerical). We adopt similar fine-tuning strategies for the non-domain BERT and Llama. Word count models are fine-tuned with non-linear classification heads. In contrast, the closed-source LLMs rely on few-shot prompting rather than parameter updates.[38]

Label Model	Direction	Flow	Category	Price	Numerical
BERT (DAPT)	**0.870**	**0.806**	0.816	**0.767**	0.942
BERT (base)	0.854	0.797	**0.822**	0.696	**0.949**
GPT 4o mini	0.857	0.679	0.638	0.567	0.917
Llama 32 3B	0.854	0.742	0.761	0.658	0.899
TFIDF + Logistic Reg.	0.752	0.758	0.687	0.638	0.899
TFIDF + RF	0.738	0.734	0.546	0.600	0.919
TFIDF + SVC	0.772	0.770	0.755	0.651	0.904

Table 1: Comparison of model performance. Performance metric reported is the model accuracy. The sample consists of a 20% randomly selected held-out sample of labeled data.

4.2.5 Step 5: Evaluation of model performance

Table 1 reports the performance of the models along all five dimensions of labels. Across all dimensions, the performance of BERT models, both base and domain adapted, surpasses those of larger models, including GPT-4o-mini and Llama 3.2 3B. These models also tend to exceed those of word-count base model, indicating that the non-linear patterns of the text matters for its interpretation of the labels. The domain-adapted pre-trained (DAPT) model tends to perform better than the base model, although the non-adapted model marginally outperforms the adapted model when classifying the categories and numerical labels. Appendix Tables C1 to C5 considers other indicators of performance for each dimension of classification: precision, recall, F-1 scores. These alternative metrics provide a similar picture of the relative performance across models.

4.3 Refining the baseline iBoP-C measures

While our approach aims for broad coverage that leverages all available information on restrictions in the AREAER, some entries may inadvertently fall outside the IMF's

[37]We used the WordPiece tokenizer from bert-base-uncased to process the text, splitting on subword units at a maximum sequence length of 128. For each 128-token chunk of the text, 20 percent of tokens are randomly masked as part of the training objective. The model is then trained on 20 epochs of the entire corpus. Perplexity, a common measure of how well a language model predicts a sample (lower is better), evaluated on a held-out sample showed a reduction from 17.28 to 2.19, indicating a substantial improvement in the model's predictive accuracy.

[38]Prompts for generative LLM applications are provided in Appendix D.1.

reporting mandate. Under the Articles of Agreement, member countries are required to report restrictions on current international payments and transfers (Art. VIII, Sec. 2(a)), discriminatory currency arrangements or multiple-currency practices (Art. VIII, Sec. 3), and capital controls (Art. VI, Sec. 3). While restrictions beyond this scope (e.g., national-security or purely trade measures) could be of significant economic interests, we find that they are more likely to be inconsistently reported in the AREAERs. If left unfiltered, could distort the time profile of restrictions across countries. We therefore refine the baseline set of restrictions by excluding four categories of restrictions:

- *Exchange-regime descriptions.*[39] Restrictions that describe exchange-rate regimes or monetary-policy frameworks (AREAER III.A–E) rather than specific cross-border measures. Although such frameworks may influence the use of capital controls, they represent policy regimes rather than restrictions per se. This refinement primarily affects Category 1 (foreign-exchange measures).

- *National security related restrictions.* Measures imposed for national-security or geopolitical reasons (AREAER II.B), which may appear in any category. These measures do not constitute de jure capital or current-account restrictions within the IMF's mandate and are systematically excluded.

- *Trade-related restrictions.* Measures that directly affect the flow of goods or services—such as import duties, export taxes, embargoes, or export-licensing requirements—rather than the financial transactions associated with those flows. These correspond to AREAER VII.D–F and VIII.D–E and mainly affect Categories 4 (Import Payments) and 5 (Export Proceeds).

- *Domestic macroprudential measures (MPM).*[40] Domestic regulations unrelated to cross-border flows (AREAER XII.A.4–6, XII.A.9.a, XII.B.1.c–d, XII.B.2.c–d, and XII.B.3.c–d). These affect Category 8 (Financial-sector restrictions) and include measures such as reserve requirements on foreign-exchange deposits and interest-rate controls.

[39] The IMF's Article of Agreement includes member countries' obligations to report changes in exchange arrangements. However, we opt to exclude these from our dataset as they are comprehensively covered elsewhere, notably in Reinhart and Rogoff (2004).

[40] The IMF reports these measures in a separately administered survey to member country. Link: https://www.elibrary-areaer.imf.org/Macroprudential/Pages/iMaPPDatabase.aspx

Classification. The training sample for each refinement label consists of positive and negative examples drawn from the relevant AREAER categories. The examples are restricted to post-2016 narratives, where AREAER categorical labels are available. Positive examples consist of examples from the AREAER categories defined under each refinement label. Negative examples consist of examples from the same category of restriction (Category 1-8) that do not fall under the AREAER categories classified as the positive examples. In the case of national security-related restrictions, we use examples from across all eight categories, since we find that that these restrictions may also show up in other categories. To ensure a balanced panel, we randomly down-sample negative examples to match the number of positive examples.

To construct the trade-related labels, we impose an additional two step procedure to further enhance the accuracy of the classification. In the first step, we construct a training sample with positive examples from trade-related categories (AREAER VII.D–F, VIII.D–E) and negative examples from other entries in Categories 4 and 5. These examples are used to train a model that classifies trade-related measures. In the second step, we filter for restrictions explicitly referring to payments by matching trade and payment keywords. Restrictions that are trade-related in the first step or not payment-related in the second are excluded from the final sample.

Because detailed AREAER category labels are available only from 2016 onward, applying these refinements consistently to earlier years requires prediction. We therefore train Large Language Models (LLMs) to extend these refinements to the full historical sample. All models are based on our domain-adapted baseline model (BERT-DAPT; see Section 4.2). Appendix Table A2 provides the performance of the model for each of the refinement label.

Appendix Table A1 summarizes the distribution of restrictions by year and refinement category. The refined sample—which excludes trade-related, domestic macroprudential (MPM), national security-related restrictions, and exchange regime measures—accounts for roughly 60 percent of all restrictions reported in the *AREAER* Changes section between 1950 and 2022. Trade-related measures represent about 28 percent of all entries, domestic MPM measures account for 2.4 percent, national security related measures account for 4.10 percent, while exchange regime based measures account for 6.30 percent. These measures are excluded in the construction of the baseline sample used in subsequent analyses.

Duplicated entries. The raw dataset of measures extracted from AREAER reports contain a non-trivial share of duplicated narratives. In particular, we identify approximately 8 percent of narratives with similar country-date identifiers. Of these, approximately 4 percent also have similar category identifiers. We treat these duplicated entries as follows. For narratives that share similar country-date-category metadata, we keep a unique entry. These narratives are otherwise undistinguishable using information contained in the AREAERs. For narratives that share similar country-date metadata but have different category metadata, we keep all entries in the construction of the numerical measures discussed below. The reason is because countries may often record multiple entries for a given measure if they assess these measures to apply to multiple categories. Nonetheless, in Section 6, we retain only unique narratives given that the application focuses on the bunching and staggered use of these measures.

4.4 Constructing Numerical Measures

Having classified narratives of restrictions, the next step involves constructing our measure of cross-border flows' restrictions. We consider two separate measures of restrictions: a measure capturing changes in restrictions (iBoP-C) and a measure capturing the stance of restrictions (iBoP-S).

The change index, iBoP-C. iBoP-C measures changes in cross-border restrictions over time, starting from 1950 until 2022. The index relies on information from the Changes section of the AREAER. As previously discussed, this section reports changes in restrictions from all sections in the AREAER at the daily level starting from 1950. Moreover, the reporting structure has remained consistent until today. After classifying these changes, our baseline measure aggregates these changes by summing all tightenings reported at the country-year-category level and subtracting all loosenings reported at the same country-year-category level. We then cumulate these net tightenings from the first year until the most recent year.[41] Formally, for

[41] As discussed below, this aggregation methodology does not account for differences in the economic significance of each measure. We address this limitation to some extent by considering an intensity-weighted index that assigns a score to each measure based on their subjectively-assessed impact on cross-border flows. See Section 8.2 for additional details.

country i in year t and category c, iBoP-C is given by

$$iBoP\_C_{itc} = \sum_{\tau \leq t} \left(\sum_{k \in \mathcal{T}_{i\tau c}} Tightening_{k,i\tau c} - \sum_{l \in \mathcal{L}_{i\tau c}} Loosening_{l,i\tau c} \right) \tag{1}$$

where $Tightening_{k,itc}$ is the k change that tightens restrictions in country i, category c and year t, and similarly $Loosening_{l,itc}$ is the l change that tightens restrictions in country i, category c and year t. The set of measures are captured in $\mathcal{T}$ and $\mathcal{L}$, which can vary across countries, time and category. While we report results at the annual frequency, analogue measures are available at a quarterly, monthly, and even daily level given that changes are reported at the daily frequency.

iBoP-C allows us to document adjustments to cross-border restrictions at a high frequency and therefore show whether countries have been actively loosening or tightening different measures. However, it does not lend naturally to measuring the overall stance of restrictions across countries. Doing so requires knowing the initial level of restrictiveness in controls by country. As a step towards addressing these concerns, we also develop a stance index, iBoP-S, that aims to capture the extensive margin of cross-border restrictions over time.

The stance index, iBoP-S. While iBoP-C is constructed using the section on changes in the AREAERs, iBoP-S leverages on the remaining sections that provide information on the stance of restrictions at the country-year-category level. In contrast to iBoP-C, iBoP-S identifies the *existence or not* of restrictions for each of the subcategories that we consider. We employ information on granular classifications of the stance of restrictions in the AREAER starting in 1995.[42]

A key feature of the iBoP-S construction is that it respects the hierarchical structure of the AREAER. The AREAER organizes restrictions in a nested taxonomy: broad categories (e.g. capital and money market instruments) contain subcategories (e.g. equity inflows and outflows), which themselves contain more granular subcategories (e.g., equity purchases by non-residents and equity sales by residents abroad). The stance index mirrors this structure and is constructed using a bottom up approach.

At the lowest level, each subcategory s is a binary indicator $\mathcal{I}_{s,it} \in 0, 1$, with 1 indicating that a restriction exists and 0 indicating that no or minimal restrictions

[42] In the 2022 version of the AREAER, there are a total of 232 subsections of restrictions relevant to the categories we cover. Given that the structure of the AREAERs differ significantly in the pre-1995 period, this measure is available only starting from 1995. In future work, we aim to release a standardize measure that extends iBoP-S from 1950.

exist. Intermediate parent catgories aggregate their children categories by taking simple average of the stance indicators of all nodes directly below them. Missing values or non-numerical values are ignored in computing the averages. Higher-level categories aggregate the stance values of their immediate children categories in the same way.

Formally, let the AREAER define a hierarchy for each category c, where every node n has a set of immediate child nodes $\mathcal{C}(n)$. The stance value for any node is defined recursively as

$$Stance_{it}(n) = \begin{cases} \mathcal{I}_{n,it}, & \text{if } \mathcal{C}(n) = \varnothing, \\ \dfrac{1}{|\mathcal{C}(n)|} \displaystyle\sum_{m \in \mathcal{C}(n)} Stance_{it}(m), & \text{if } \mathcal{C}(n) \neq \varnothing. \end{cases} \tag{2}$$

The stance index for a top-level AREAER category c (i.e., the same eight categories used in the iBoP-C index) for country i in year t is then defined as

$$iBoP\_S_{itc} = Stance_{it}(c). \tag{3}$$

Thus, the value of each category is an equally weighted average over all subcomponents, where wieghts arise naturally from the nested structure of the taxonomy. Unlike the iBoP-C, iBoP-S index is available only at the annual frequency, because the stance of restrictions is reported annually in the AREEAER.

4.4.1 Caveats and Limitations

While our measurement improves upon most of the literature in capturing the direction of policy adjustments by counting the number of reported tightening and loosening changes within each category per year, a limitation of this approach is that it weighs all reported changes equally. In practice, relatively minor administrative measures —e.g., changes in documentation requirements— are given the same weight as substantial legislative reforms —e.g., the imposition of capital requirements on the banking sector. To address this, in an extension presented in Section 8.2, we extend our approach by incorporating intensity-based classifications, in the spirit of the ordinal scales used by Quinn (1997) but implemented systematically with LLMs. This refinement allows us to distinguish between minor procedural adjustments and sweeping prohibitions, and to show how accounting for intensity refines both the measurement of policy trends and the interpretation of their economic relevance.

An important characteristic of iBoP-C is that it identifies the frequency and net direction (tightening versus loosening) of policy changes rather than absolute restrictiveness. Thus, while our index accurately documents whether restrictions have increased or decreased in a given year, it cannot indicate the absolute openness or restrictiveness across countries. Consequently, this measure should not be used to rank countries by their overall restrictiveness. We address absolute openness levels with iBoP-S that we have from 1995 onward.

Another limitation to any work that builds upon the AREAER as main source is related to the completeness and consistency of the reporting in these documents, especially given evolving reporting standards and coverage over time. Early AREAER reports, particularly before the mid-1990s, often have less detailed or potentially incomplete reporting. Consequently, earlier years may understate the frequency and scope of policy changes. Our methodology addresses this limitation by deploying large language models trained on the full set of AREAER text to classify changes. Given that the same information is used to train the model, the classification should be consistent over time. Moreover, we also carefully examine country-specific narratives in developing our training sample, including resolving ambiguous cases and ensuring that these cases are consistently labeled within our training sample.[43]

Despite these caveats, our methodology offers significant advantages. By providing comprehensive coverage of restrictions in the AREAER over seven decades, it substantially extends the historical scope beyond previous datasets while simultaneously allowing for a level of granularity in the policy actions that is far beyond of what other studies have done in terms of frequency, type, and direction of the policy actions. Moreover, our systematic classification approach greatly improves consistency, comparability, and reproducibility compared to traditional narrative analyses. This should allow researchers to draw robust insights into long-term trends and patterns of cross-border flow restrictions.

[43]As mentioned in the literature review, some studies have relied on additional sources other than the AREAERs. This, however, comes at the cost of covering a smaller sample of countries and/or time in order to remain consistent.

5 Stylized Facts

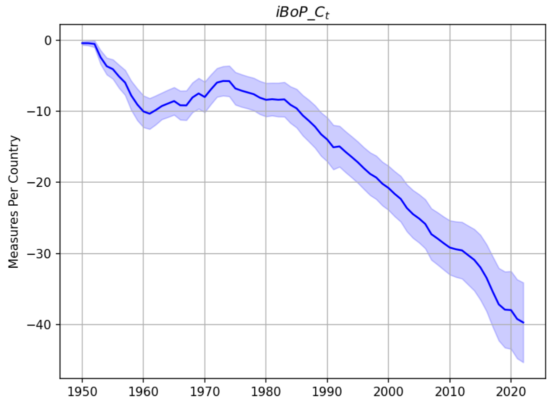
This section presents several stylized facts about the long-run evolution of cross-border flow restrictions using the iBoP-C index. First, we find that liberalization trends in cross-border flow restrictions have not evolved linearly over time. While countries have overall liberalized their financial flows over the past seven decades, periods of significant change in the global monetary system, notably the breakdown of Bretton Woods, saw a significant tightening of restrictions for several categories. Second, we document that financial liberalization has occurred at an uneven pace, with higher income countries liberalizing faster and to a greater extent than lower income countries. This pattern is robust across all categories of restrictions that we examine. Interestingly, disaggregating by income, we find a more extensive use of current account restrictions (i.e., import payment and export proceed restrictions) to manage cross-border flows around the end of the Bretton-Woods regime among low-income countries. Third, we find that liberalization of cross-border restrictions have primarily concerned quantity-based restrictions. Administrative-based and price-based measures restrictions saw some loosening over time, but at a relatively more modest pace. Finally, we document that the liberalization trend in BoP flows has been largely driven by the more rapid pace in loosening of outflow restrictions, relative to that of inflow restrictions.

5.1 A Long-Run Perspective: 1950-present

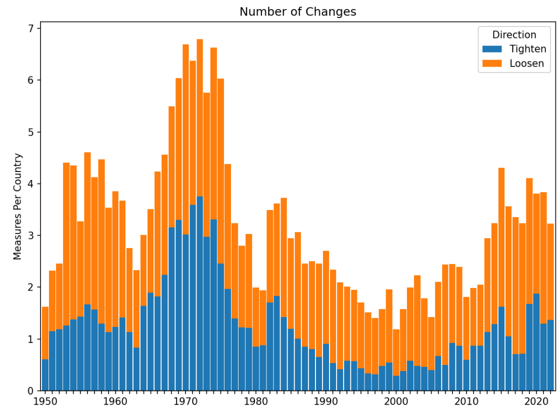
5.1.1 Stylized Fact 1: Liberalization trends in cross-border flow restrictions have not evolved linearly.

Cross-border restrictions display a clear stop-go pattern rather than a monotonic decline. After an initial period of liberalization in the 1950s, countries saw sharply tightened restrictions from the early 1960s through the early 1980s. This tightening coincided with rising pressures on the Bretton Woods system and then the transition to generalized floating, when many more countries relied more heavilyh on FX and financial sector regulations to manage balance-of-payment pressures and exchange rate volatility.

The annual number of policy changes show pronounced clustering, with countries adjusting restrictions in a coordinated manner rather than randomly through time.



(a) Cumulative net tightening



(b) Number of changes

Figure 5: Left panel: Average cumulative net tightening index (iBoP-C) (negative values indicate net loosening). Right panel: Average number of tightening and loosening measures per country. Shaded area denotes the standard error of the mean $(SD_t/\sqrt{(N_t)}$ where SD_t is the standard deviation of Changes index. The overall series agregates all categories by country-year.

This bunching behavior reflects episodes of systemic stresses in the international monetary system, a feature that we will explore further in Section 6.

Category-level evidence shows that the 1970s tightening was concentrated in FX and financial sector restrictions, with some increases in export-proceeds requirements (Figure 6). Despite these episodic reversals, most categories exhibit clear, sustained liberalization from the the mid-1980s onwards, consistent with global financial deregulation and deeper integration of countries into international markets.

Category	Count	Share (%)
FX markets	4239	10.3
Payments and receipts	6040	14.7
Resident and non-resident accounts	3123	7.6
Import payments	4313	10.5
Export proceeds	2939	7.2
Invisible transactions	5114	12.5
Capital account transactions	8023	19.6
Financial sector	3102	7.6
Overall	41030	100.0

Table 2: Number of changes by category

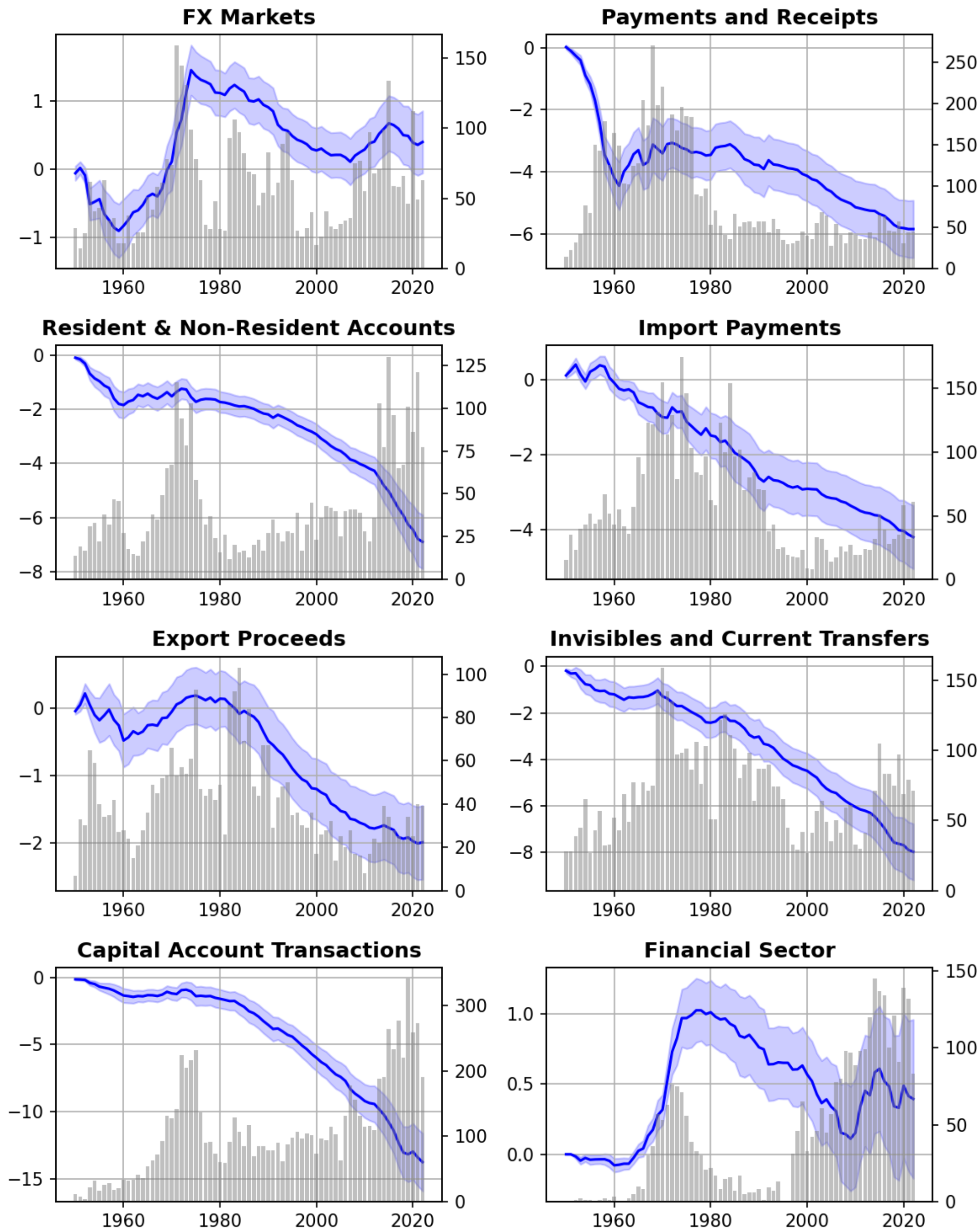


Figure 6: Blue lines show cumulative net tightening for each category (left axis); grey vertical bars show annual counts of policy measures (right axis).

Table 2 showss that policymakers rely on a wide array of instruments to manage cross-border flows, beyond the set of capital account restrictions typically focused in the literature (see for example Fernandez et al. (2016)). In our sample, capital account restrictions account for only 20 percent of all measures. On the other hand, a third of all measures are related to current-account payments (import, exports, and invisibles), while the remainder relates to FX market and financial sector restrictions. Focusing solely on capital account measures would therefore overlook a substantial share of tools used historically to regulate cross-border flows.

5.1.2 Stylized Fact 2: Financial liberalization occurred unevenly, with high-income countries liberalizing faster and to a greater extent than lower-income countries.

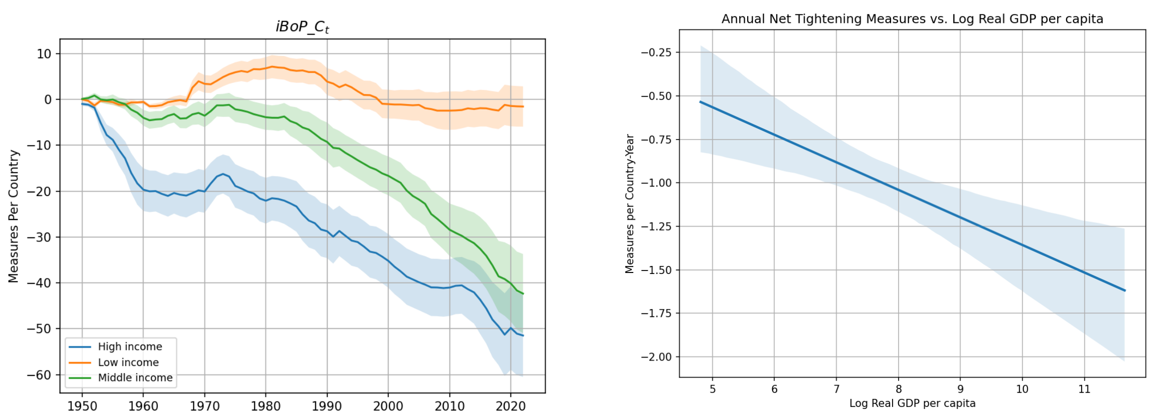


Figure 7: Left panel: average cumulative net tightening (iBoP-C) by income group (2024 World Bank Income Group Classification). Right panel: Annual net tightening measures per country-year vs log real GDP per capita (2015 USD). The line shows a linear regression fit, the error band shows the 95 percent confidence interval of the regression estimate.

There is substantial cross-country heterogeneity in the timing and extent of liberalization across country groups. High income countries began liberalizing in the 1960s and continued steadily thereafter, aside from a brief reversal during the Bretton Woods collapse. Middle income countries followed a similar trajectory but with a lag of nearly two decades. By contrast, low-income countries tightened restrictions in the 1970s and have remained comparatively restrictive even in recent decades[44]. In Section 6.3, we document how these patterns can be explained by differences in institutional quality as well as political, financial, and macroeconomic risk in the

[44] These patterns align closely with de facto financial globalization measures documented in Capelle and Pellegrino (2025).

economy. Cross-sectional relationship between income per capital and net tightening actions reinforces these patterns. Countries with higher income levels loosen restrictions more frequently, consistent with the intuition that these countries have stronger capacity to manage cross-border flow volatility.

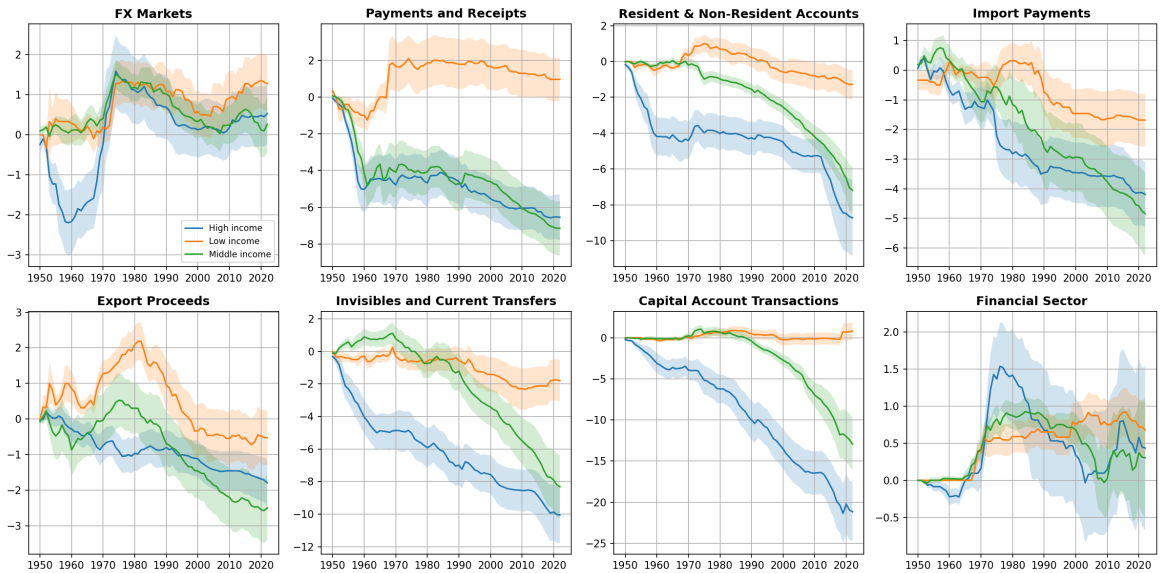


Figure 8: Average cumulative net tightening (iBoP-C) for each category, computed over countries within each income group (2024 World Bank classification).

Category-level results (Figure 8) show that high-income countries led liberalization across nearly all dimensions, especially on capital account measures and regulations on payments and receipts. Even so, these countries experienced targeted episodes of tightening in the 1970s, especially in FX market and financial sector restrictions. Middle-income countries followed similar sequencing with a lag, particular for capital account restrictions, whereas low-income countries saw limited liberalization and relied more heavily on current account restrictions (import and export-related measures). However, because cumulative changes do not capture initial restrictiveness, these results may not reflect differences in the absolute restrictiveness of cross-border flow restrictions across income groups.

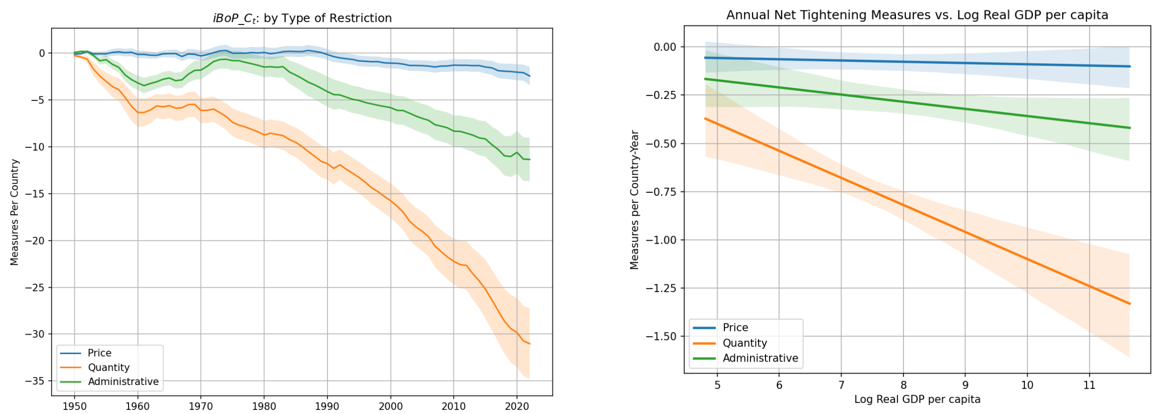


Figure 9: Left panel: Average cumulative net tightening (iBoP-C) by type of restriction (price-based, quantity-based, administrative). Right panel: Annual net tightening measures per country-year vs log real GDP per capita (2015 USD). The line shows a linear regression fit, the error band shows the 95 percent confidence interval of the regression estimate. See Appendix Figure A8 for category-level indices.

5.1.3 Stylized Fact 3: Quantity-based and administrative-based restrictions have seen progressive loosening, while price-based restrictions tend to persist.

We further decompose restrictions into price-based restrictions, quantity-based restrictions, and administrative-based restrictions. Price-based restrictions directly alter the costs of transactions. These include taxes, fees, subsidies, interest premiums, surrender and repatriation requirements, and reserve requirements. Quantity-based restrictions limit the volume or amount of transactions and encompass quotas, caps, bans and suspensions. Administrative restrictions impose procedural or institutional requirements that do not directly target cost or volume, such as licenses, authorization requirements, reporting requirements, maturity requirements, holding period requirements, or bilateral payments agreements.

Quantitative measures show the strongest long-run liberalization. These measures were widely used during the Bretton Woods period but were steadily liberalized starting in the 1980s as countries shifted toward more market-based allocation mechanisms. Administrative measures show mild tightening in the 1970s but broadly liberalized afterwards. By contrast, price based measures display little long-run variation.

There are also notable differences in the extent to which different types of measures are liberalized by income group. High income countries are more likely to report loosening all three types of measures, with a larger number of loosenings for quantity-based and administrative based measures. On the other hand, lower income countries exhibit fewer easings along each of these three types of measures.

5.1.4 Stylized Fact 4: Outflow restrictions have liberalized more rapidly than inflow restrictions

Our dataset allows us to classify restrictions by flow direction. Broadly, we distinguish between outflow restrictions (i.e., measures limiting flows out of a country) and inflow restrictions (i.e., measures limiting flows into a country). Figure 10 (left panel) shows that while both inflow and outflow restrictions have liberalized considerably after 1980, outflow restrictions were loosened earlier and more extensively. This asymmetry is consistent with longstanding policy justifications in the literature. In particular, outflow controls have historically been deployed as tools to contain capital flight and stabilize exchange rates during periods of financial stress (Chang et al. (2024)). Once macroeconomic conditions normalize, these restrictions tend to be removed relatively quickly. By contrast, inflow restrictions are often used to manage financial stability risk, through mitigating excessive borrowing, which could lead to credit booms and busts. As a result, inflow measures tend to be more persistent and liberalized more gradually.

Nonetheless, there are notable differences across income groups. High income countries systematically loosen both inflow and outflow restrictions, with a greater likelihood of removing outflow restrictions. On the other hand, low income countries are less likely to loosen outflow restrictions, suggesting that these countries continue to rely heavily on these types of restrictions to manage external risks. We document the use of these types of measures during crises episodes in Section 6.2.

5.2 A More Granular View of the Past Seven Decades

We next examine a more granular breakdown of restriction subcategories in the AREAER to further understand specific policy tools countries have employed. While the AREAER has consistently reported broad categories, detailed subcategory information is systematically available only from 1995 onward, and even then, comprehensive classification was not consistently reported until after 2016.

To extend the subcategory classification to the full sample of measures, we em-

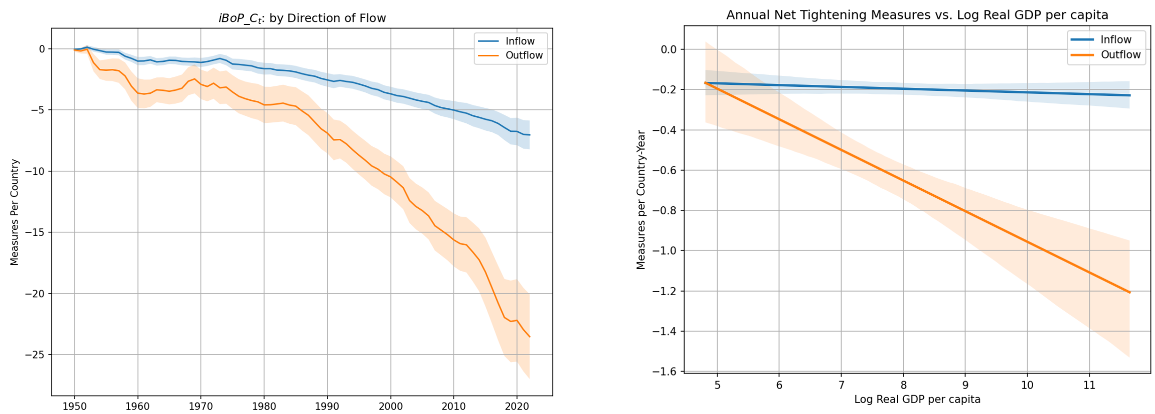


Figure 10: Left panel: Average cumulative net tightening (iBoP-C) for inflow and outflow restrictions. Right panel: Annual net tightening measures per country-year vs log real GDP per capita (2015 USD). Charts omit restrictions that affect both inflows and outflows or do not have clear flow implications. The line shows a linear regression fit, the error band shows the 95 percent confidence interval of the regression estimate. See Appendix Figure A9 for category-level indices.

ploy our baseline LLM fine-tuned on the AREAER text (BERT-DAPT). The training and validation data is constructed using the subcategory classifications from the AREAERs in the post-1995 sample. To ensure there are sufficient examples for each subcategory, we manually group these subcategories based on their economic similarity. A separate multi-class classification head is then attached to the LLM and fine-tuned for each of the eight broad categories to distinguish between subcategories within a given broad category. We compare the performance of the model, trading off between achieving greater granularity in the subcategory classifications and the accuracy of the model. This iterative procedure leaves us with a final grouping of 24 distinct subcategories shown in Table 3.[45]

Figure 11 presents cumulative net tightening trends as well as a count of measures in each of the 24 subcategories over time. We observe several notable patterns. First, restrictions on the trade of gold and control of banknotes were significantly tightened around the end of Bretton Woods, indicating that it was a significant policy lever. We also see a pick up in repatriation and surrender requirements, both on export

[45]Appendix Table A4 presents the performance metrics by category. The average model accuracy across categories is 83 percent. Note that since this classification task relies on the aggregate categories being available. Users should also account for classification errors arising from the classification of the eight broad categories.

Subcategory	Count	Share (%)
Overall	41030	100.0
Exchange taxes and subsidies	271	0.7
Restrictions in the foreign exchange market	3968	9.7
Prescription of currency requirements	2575	6.3
Trade of gold	867	2.1
Control of banknotes	2598	6.3
Restrictions on resident accounts	1597	3.9
Restrictions on nonresident accounts	1526	3.7
Import financing requirements	3128	7.6
Import documentation requirements	1185	2.9
Repatriation and surrender requirement of export proceeds	2625	6.4
Export financing and documentation requirements	314	0.8
Invisible payments: trade and investment	771	1.9
Invisible payments: travel and personal	2962	7.2
Invisible proceeds	1381	3.4
Repatriation and surrender requirement on capital transactions	156	0.4
Capital account: capital and money market	2850	6.9
Capital account: credit operations	1777	4.3
Capital account: direct investment	2064	5.0
Capital account: real estate	507	1.2
Capital account: personal capital	669	1.6
Restrictions on commercial banks	2249	5.5
Restrictions on insurance companies	356	0.9
Restrictions on pension funds	284	0.7
Restrictions on investment and collective investment funds	213	0.5

Table 3: Number of changes by subcategory

proceeds as well as on capital transactions during that period. Finally, we observe that financial sector restrictions tend to be more concentrated on commercial banking activity. While these restrictions have tightened in the recent years, particularly following the GFC, restrictions on the non-banking sector (institutional investors, pension funds, and collective investment funds) saw broad loosening over time.

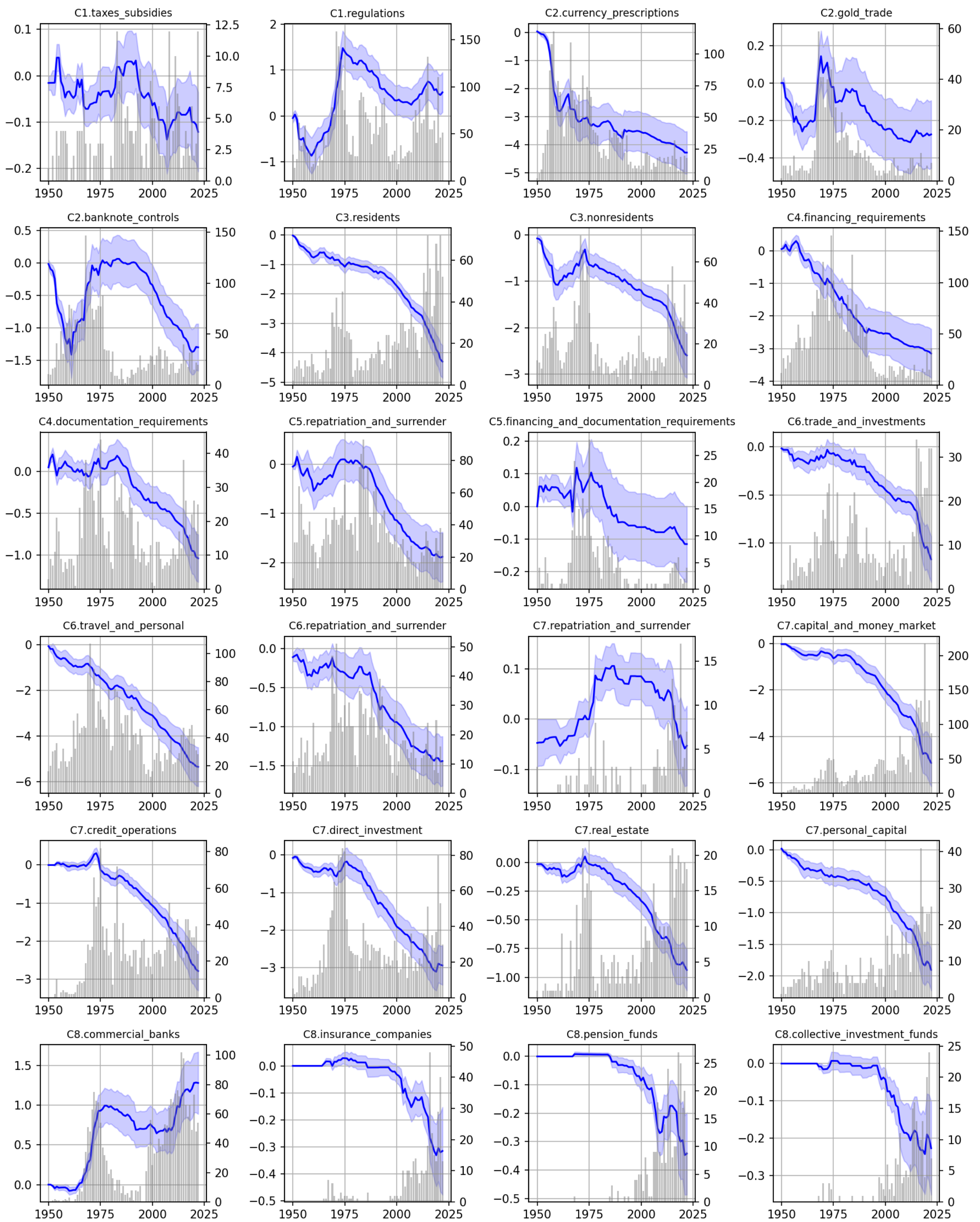


Figure 11: Average cumulative net tightening and number of measures by subcategory.

5.3 Overview of the Stance Index, iBoP-S

Having discussed the stylized facts from the Changes index (iBoP-C), we now turn to the Stance index (iBoP-S). Whereas iBoP-C captures adjustments in restrictions, iBoP-S measures the extensive margin of cross-border flow restrictions. We report iBoP-S beginning in 1995, when the AREAER first introduced systematically disaggregated measures across all categories.[46]

To classify the stance of restrictions based on descriptions of restrictions, we employ a baseline LLM fine-tuned on AREAER texts (BERT-DAPT). The training and validation dataset is constructed from the binary "yes/no" indicators that reflect the presence of restrictions reported in the AREAERs. For capital account transactions, we replace these indicators with those from FKRSU to maintain consistency in the coding with the latter database. The per-category training sample is constructed using full de-duplicated narrative-indicator pairs from the 1995-2017 sample, while the per-category validation sample is constructed from 500 randomly selected de-duplicated narrative-indicator pairs from the 2018-2022 sample. To allow for the relationship between narrative-indicators to differ across categories, we train separate models for each category.[47]

We rely on predicted values instead of the raw yes/no indicators from AREAER reports for two reasons. First, textual descriptions often contain richer and more accurate information than the status columns, as noted by Fernandez et al. (2016). Second, the status columns are self-reported indicators, which may be inconsistent across countries and over time. Training a model on examples from multiple countries and periods ensures greater consistency. Where no narrative description is available, we fall back on the reported yes/no indicators, consistent with Fernandez et al. (2016).

Once predictions are generated, iBoP-S is constructed bottom-up by aggregating stance values across subcategories. At the most granular level, each subcategory takes a binary value (0 = no restriction, 1 = restriction in place). Missing information is treated as not available. We then average across subcategories within each group, moving upward until we obtain stance values for the eight broad categories. This

[46]While the LLM can be used to extend the Stance index further back in time, the structural break in the reporting of the AREAERs in 1995 makes it difficult to generate a continuous Stance index over time. We leave the task for future extensions.

[47]Appendix Table A5 reports the model performance metrics across categories based on a held-out test sample. The average accuracy across categories is 0.92.

approach parallels that of Fernandez et al. (2016) for capital account restrictions but extends it to the remaining seven categories reported in the AREAER.

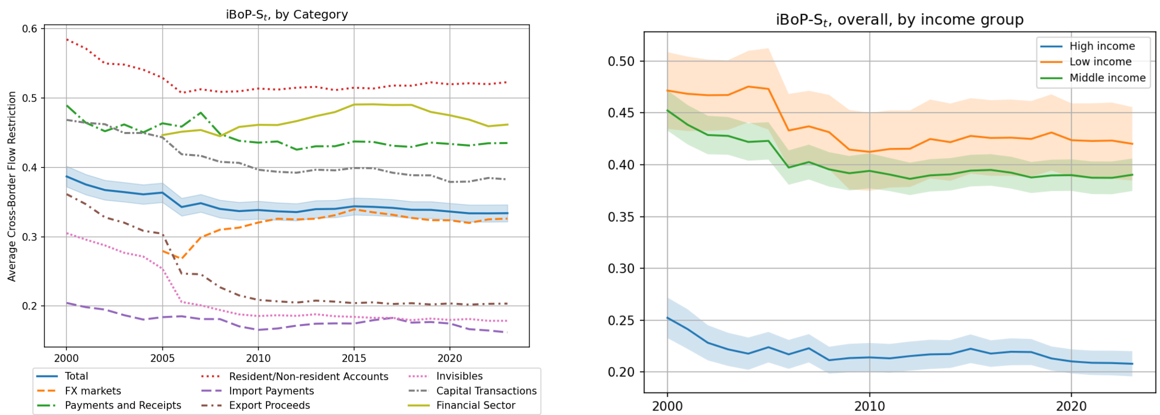


Figure 12: Left panel: Average overall stance (iBoP-S) by category. Right panel: Overall stance by income group.

Figure 12 illustrates the overall stance of restrictions (between 0 and 1), with lower values indicating a greater extent of liberalization. The left panel shows significant heterogeneity in the stance of restrictions across categories, with current account (exports, imports, and invisibles) being the most open in general, whereas financial sector and resident/non-resident account restrictions being the least liberalized. Given that the stance index measures the extensive margin of restrictions, there tends to be more limited variation. Nonetheless, we continue to observe a trend towards greater openness over the past two decades.

The right panel shows that higher income countries tend to be the most liberalized, consistent with the results from the cumulative changes index (iBoP-C). Middle and low income countries appear to be twice as restrictive as high income countries. We find these stylized facts to be qualitatively similar at the category level (Appendix Figure A10). In some cases, such as for capital account restrictions, middle and low income countries appear to be three times as restrictive as low income countries. This suggest that there continues to be a gap in the degree to which countries deploy cross-border flow measures, with middle and low income countries relying on these measures to manage cross-border flows as compared to high income countries.

6 Use of cross-border flow restrictions

With comprehensive data on cross-border flow restrictions now available, it is important to reassess how these instruments are deployed. This has been done for capital flow measures (most prominently in Fernández et al. (2015), Eichengreen and Rose (2014) and Pasricha (2022)) and should be extended to the other seven categories of cross-border flow restrictions. As mentioned earlier, capital flow restrictions only constitute less than 20% of all cross-border restrictions which is why focusing on them misses a significant share of policy tools used by countries to steer international flows. In addition to this, if capital flow measures are combined with tools from other categories, existing analyses could have attributed the total effect solely to the capital flow measure. While the preceding section documented aggregate patterns across countries and time, this section investigates the country-level use of these tools to manage the (political) economy.

This section presents novel evidence on the timing, crisis-related deployment, and cross-country heterogeneity in the use of cross-border restrictions. First, we document that these measures are rarely implemented in isolation; instead, they tend to cluster within narrow time windows. Over 70% of measures are accompanied by at least one additional measure within 30 days, indicating that policymakers employ a broad toolkit in rapid succession. Second, we find that the likelihood of imposing cross-border restrictions rises significantly during crises, particularly those involving sovereign debt or currency distress. Finally, cross-country heterogeneity analysis reveals that jurisdictions with weaker institutional quality and elevated macro-financial risks exhibit a markedly higher propensity to adopt cross-border restrictions.

6.1 Bunching and Staggering

We begin by analyzing the temporal clustering of measures at the country level. Several novel patterns emerge.

First, measures are rarely implemented in isolation. The first row of Table 4 reports the incidence of *bunching*, defined as the adoption of multiple measures on the same day: 35% of all measures coincide with at least one other measure. This indicates deliberate policy coordination, with some episodes involving extensive simultaneous interventions. Figure 13 complements this by showing that 62% of bunching events involve two measures, 29% involve three to five, and 6.3% involve six to ten. 2.7% of bunching events even record more than 10 measures used simultaneously.

Window	Single Measure	More than One Measure
± 0 days	65%	35%
± 1 days	59%	41%
± 3 days	54%	46%
± 5 days	49%	51%
± 10 days	40%	60%
± 20 days	30%	70%
± 30 days	23%	77%

Table 4: Share of measures in bunching and staggering for different horizons. This table shows the share of policy measures that are stand-alone, bunched with other measures, or staggered within ± 1, 3, 5, 10, 20 and 30 day windows.

Moving to the lower lines of Table 4, we look at wider time horizons and analyze whether countries do staggered interventions, i.e. they use more than one measure within a certain time frame. Unlike bunching, staggering may reflect sequential adjustments or recalibration following earlier interventions. Line 5 shows that almost 60% of measures are accompanied by another measure within 10 days, and 77% within 30 days. This pervasive clustering complicates causal inference: when case studies analyze the effects of a cross-border measure on macroeconomic variables (such as capital flows or economic output), effects of closely spaced interventions might be falsely attributed to the event in question. While this complicates the evaluation of the policies, at least the new dataset can now provide full information needed to address these confounding factors.

As a second novel finding, we show in Table 5 that countries combine a very broad spectrum of tools. Each line focuses the share of other measures given that the measure was used in combination with at least one other measure within 30 days. While some tools - such as measures on capital accounts - are more common than others during staggering episodes, it shows that policy makers use literally all combinations possible. Very often, measures are also combined with another measures in the same category (diagonal in blue in the table). This most likely reflects reflect sequential adjustments or recalibration. Finally, the data also shows that both, bunching and staggering happen over time and across countries with different stages of development and exchange rate regimes. In other words, they are not limited to

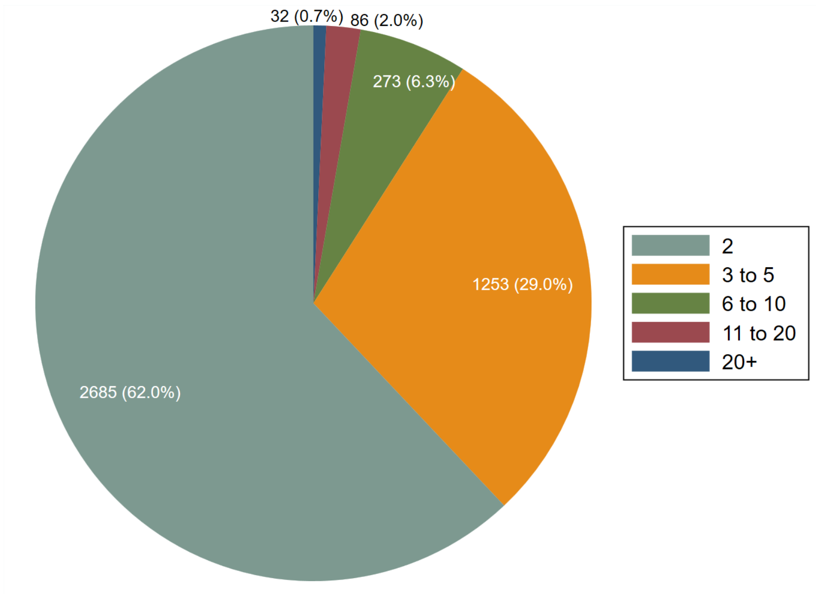


Figure 13: Number of measures on the same date.This figure shows the different types of bunching that occur within the dataset. Each category (2, 3 to 5, 6 to 10, 11 to 20, 20+) represents the number of measures adopted on the same day. The relative size of each category is conditional on there being at least two measures implemented on the same day.

a special episode or a specific countries, underscoring their global and persistent nature (Figure A13).

	FX Markets	Pyts & Rcpts	NR Act.	Import Pyts	Export Pcds	Invis Trans	Capital Act.	Fin. Sector
FX Markets	20	12	8	12	9	15	17	7
Pyts & Rcpts	9	27	11	11	7	13	17	5
NR. Acc	9	14	14	10	6	16	23	6
Import Pyts	10	12	8	22	12	17	16	4
Export Pcds	10	12	8	16	13	15	20	4
Inv. Tra	11	12	10	15	9	19	19	4
CapM	10	11	10	11	8	15	27	8
Fin. Sector	10	9	13	7	6	11	29	15
AVG	**11**	**14**	**10**	**13**	**9**	**15**	**21**	**7**

Table 5: Conditional shares of other measures used in a staggering episode (-30 to +30 days).This table shows the share of measures that are implemented within a ±30 day window, conditional on a measure from a certain category (listed on the LHS) being implemented.

6.2 Use of cross-border flows and crises

When do countries use these measures and how does this correlate with the occurrence of crises? To address this, we use the following specification:

$$\text{No of measures}_{i,t} = \alpha + \beta * Crisis_{i,t} + \gamma * \Delta realGDP_{i,t} + \mu_i + \mu_t + \epsilon_{i,t} \tag{4}$$

where the left-hand side variable is the number of measures taken, either the total, inflows, outflows, tightenings, loosenings, or just the subcategories of measures for country i in year t. Since we want to focus on the active use of these policies, we use the sum of actions within t, rather than the stance that would be measured with the iBoP-C or iBoP-S. The coefficient of interest on the right-hand-side is β to show whether countries use the measures significantly more during crises (as defined by Laeven and Valencia, 2018). In addition to real GDP growth, we also control for country as well as time fixed effects to account for unobservable determinants at those levels.

Table 6 shows in column (1) that countries adopt significantly more measures during crises. Usage is more than double the median annual use, which corresponds to one measure. This increase is driven by measures targeting FX markets, non-resident accounts, import payments, export proceeds, and invisible transactions. In contrast, financial sector and capital account measures do not exhibit a systematic crisis response. If a paper - like much of the existing literature - would focus only on the latter, an insignificant finding would present an incomplete picture. Table 7 disaggregates by direction and shows that crises trigger both tightening and loosening measures, albeit tightening measures are more prominent. When distinguishing by flow type, countries predominantly adopt outflow measures, consistent with existing literature (Chang et al., 2024).[48] Finally, Table 8 differentiates crises by origin — banking, currency, or sovereign debt — and finds that the effect is concentrated in currency and sovereign debt crises, where tightenings and outflow measures dominate. The correlation with banking crises is not statistically significant. This contrast is intuitive as currency and sovereign debt crises often directly affect cross-border flows while banking crisis can often be contained domestically.

[48]We do not collapse by flow type and direction at the same time as this leaves us with a high share or zero values and therefore non-normality of the left-hand-side variable.

	(1) All	(2) FX	(3) Pay & Rec	(4) NR Acc	(5) Imp	(6) Exp	(7) Inv. Tra	(8) CapM	(9) FS
Crisis Dummy	**1.263*****	**0.249*****	0.077	0.179***	0.309***	0.121***	0.168**	0.144	0.014
	(0.31)	(0.06)	(0.06)	(0.06)	(0.07)	(0.04)	(0.07)	(0.11)	(0.05)
Pseudo R2	0.311	0.173	0.220	0.127	0.237	0.194	0.229	0.194	0.185
Observations	7989	7989	7989	7989	7989	7989	7989	7989	7989
△ Real GDP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Use of cross-border measures during crisis - split by categories. This table shows estimation results from specification (4). The dependent variable is the number of policy measures adopted by country i in year t. Columns (2)–(7) focus on the number of policy measures within each category. $Crisis_{i,t}$ is a dummy variable indicating the occurrence of a systemic banking crisis as defined by Laeven and Valencia (2018). All regressions include controls for real GDP growth, country and time fixed effects. Standard errors are in parentheses.

	(1) Tightening	(2) Loosening	(3) Net tightening	(4) Inflow	(5) Outflow	(6) Net inflow
Crisis Dummy	0.728***	0.452**	0.277	0.240***	0.580***	-0.340**
	(0.15)	(0.21)	(0.20)	(0.08)	(0.19)	(0.17)
Pseudo R2	0.269	0.256	0.148	0.239	0.237	0.183
Observations	7989	7989	7989	7989	7989	7989
△ Real GDP	Yes	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Use of cross-border measures during crisis - split by direction.This table shows estimation results from specification (4). The dependent variable is the number of policy measures adopted by country i in year t. Each column corresponds to the type of policy measure adopted: tightening, loosening, net tightening (number of tightening - number of loosening measures), inflow, outflow & net inflow (number of inflow - number of outflow measures). All regressions include controls for real GDP growth, country and time fixed effects. Standard errors are in parentheses.

6.3 Use of cross-border restrictions and the political economy

Recent evidence suggests that non-advanced economies rely more heavily on cross-border restrictions than advanced economies (e.g., during the COVID-19 period, Bergant and Forbes (2023)). One potential mechanism for this could be higher levels of risk and lower levels of institutional quality. In addition to this, consistent with

	(1) All	(2) Tightening	(3) Loosening	(4) Inflow	(5) Outflow	(6) Net inflow
Banking Crisis	0.028	0.352	-0.391	0.016	-0.053	0.069
	(0.51)	(0.24)	(0.34)	(0.12)	(0.31)	(0.28)
Currency Crisis	1.507***	0.789***	0.629**	0.243**	0.700***	-0.457**
	(0.41)	(0.19)	(0.27)	(0.10)	(0.25)	(0.22)
Sovereign Debt Crisis	2.532***	1.586***	0.885*	0.335**	1.538***	-1.203***
	(0.70)	(0.33)	(0.47)	(0.17)	(0.42)	(0.38)
Pseudo R2	0.313	0.271	0.256	0.239	0.238	0.184
Observations	7989	7989	7989	7989	7989	7989
△ Real GDP	Yes	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Use of cross-border measures during crisis - split by crisis type. This table shows estimation results from specification (4). The $Crisis_{i,t}$ regressor is categorized into 3 types: banking, currency and sovereign debt crises as defined by Laeven and Valencia (2018). The dependent variable is the number of policy measures adopted by country i in year t. Columns (2) - (6) focus on specific types of measures: tightening, loosening, inflow, outflow & net inflow. All regressions include controls for real GDP growth, country and time fixed effects. Standard errors are in parentheses.

opportunistic political cycles, Müller (2023) shows that macroprudential policies are more often adjusted around elections. We test these hypotheses for cross-border flow restrictions using the following specification:

$$\text{No of measures}_{i,t} = \alpha + \beta * institutional_{i,t} + \gamma * \Delta realGDP_{i,t} + \mu_i + \mu_t + \epsilon_{i,t} \qquad (5)$$

where we use the data in institutional quality provided by The PRS Group (2023) and data on elections provided by Scartascini et al. (2021). Table 9 first reports results without country fixed effects to capture cross-country heterogeneity while time fixed effects control for global shocks. Most institutional quality indicators are significantly negatively correlated with the use of cross-border measures: lower institutional quality (i.e. more conflicts, higher tensions, or higher risks) is associated with greater reliance on restrictions. For example, countries that are not able to offer stable governments (measured by a government's ability to carry out declared programs) and investment policies (measured by contract viability, profits repatriation, and payment delays) often resort to more actions regarding cross-border restrictions.

While the first 12 characteristics speak mainly to political risk, we also analyze the ICRG composite risk index, which combines political, financial, and economic risk.[49] We find that overall risk is strongly correlated with measure adoption—even after controlling for country fixed effects. Thus, within-country variation also matters: governments resort to more restrictions during periods of heightened risk. Finally, incorporating election data from Scartascini et al. (2021) reveals that countries adopt or adjust more cross-border restrictions in years of executive or legislative elections. This effect is significant across all categories and aligns with Müller (2023), who documents opportunistic use of macroprudential policies around elections.

Right-Hand-Side	Any Measure	Outflow Measures	Inflow Measures	Loosening	Tightening	Time FE	Country FE
Quality of Bureaucracy	-0.085	0.054	-0.048	0.031	-0.127***	Y	
Corruption	-0.447***	-0.218***	-0.226***	-0.059**	-0.202***	Y	
Democracy accountability	0.015	0.057	0.006	0.016	-0.037	Y	
Ethnic Tensions	-0.219**	-0.153**	-0.051	-0.103***	-0.053	Y	
External Conflicts	-0.296***	-0.186***	-0.164***	-0.021	-0.105***	Y	
Government Stability	-0.218**	-0.085	-0.087	-0.060**	-0.124***	Y	
International Conflict	-0.276***	-0.177***	-0.137***	-0.061***	-0.094***	Y	
Investment Profile	-0.408***	-0.211***	-0.218***	-0.053***	-0.183***	Y	
Law and Order	-0.233**	-0.084	-0.111*	-0.042*	-0.133***	Y	
Military in Politics	-0.091	-0.015	-0.009	-0.013	-0.064**	Y	
Religious Tensions	-0.255**	-0.206***	-0.109*	-0.093***	-0.053	Y	
Socioeconomic	-0.088	-0.032	-0.054	-0.004	-0.059**	Y	
Composite Risk Index	-0.048***	-0.021**	-0.027***	-0.007**	-0.026***	Y	
Composite Risk Index	-0.085***	-0.016***	-0.035*	-0.049***	-0.043***	Y	Y
Election Taking Place	0.529**	0.250*	0.298**	0.074	0.278***	Y	

Table 9: Correlation use of cross-border measures and institutional quality/election periods. This table shows estimation results from specification (5). The dependent variable is the number of policy measures adopted by country i in year t. The first 12 regressors measure factors of political risk: quality of bureaucracy, corruption, democratic accountability, ethnic tensions, external conflicts, government stability, international conflict, investment profile, law & order, military's involvement in politics, religious tensions and socioeconomic conditions. The ICRG Composite Index considers political, economic and financial risk factors. Higher scores indicate a higher level of institutional quality. *Election Taking Place*, constructed using election data from Scartascini et al. (2021), is a dummy variable which indicates the occurrence of executive or legislative elections in year t. Fixed effects are applied as noted in the last two columns.

[49] While economic risk contains the main macroeconomic indicators (output, inflation, fiscal budget), the financial risk focuses on variables such as exchange rate stability, import liquidity, and foreign debt.

7 Validation the LLM methodology

We validate the LLM-based methodology by comparing the constructed measures with existing manually coded measures of capital controls. The validation proceeds in two steps. First, we benchmark iBoP-S against the capital account indicators of Fernandez et al. (2016) (FKRSU). Because a subset of the indicators covered in iBoP-S encompasses all indicators covered by FKRSU, we are able to replicate their dataset directly. Second, we broaden the comparison by contrasting iBoP-C to other well-known indices, including Quinn and Toyoda (2008), Chinn and Ito (2006), and Ilzetzki et al. (2021), to assess how our more comprehensive coverage aligns with earlier approaches.

7.1 Comparison with FKRSU

We begin with a comparison of the iBoP-S capital account indicators with the indicators in FKRSU. Specifically, we focus on the overall stance of capital account restrictions (ka) and their inflow (kai) and outflow (kao) components, which covers 32 subcategories and six types of instruments: capital and money market securities, collective investment securities, derivatives, credit operations, direct investment, and real estate. The sample is restricted to the 100 countries from 1995 to 2019 available from FKRSU.

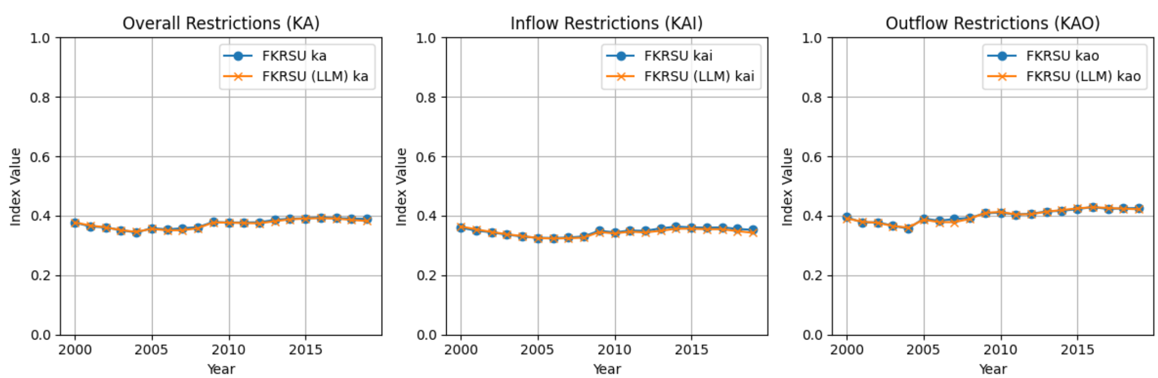


Figure 14: Average ka, kai, and kao indicators from the FKRSU (original) and FKRSU (LLM Extension). Sample restricted to 100 countries and 32 instruments from FRKSU (original) between 1995-2019.

Figure 14 shows that the average aggregated series from iBoP-S closely tracks the original dataset. Figure 15 plots the overall capital account restriction stance

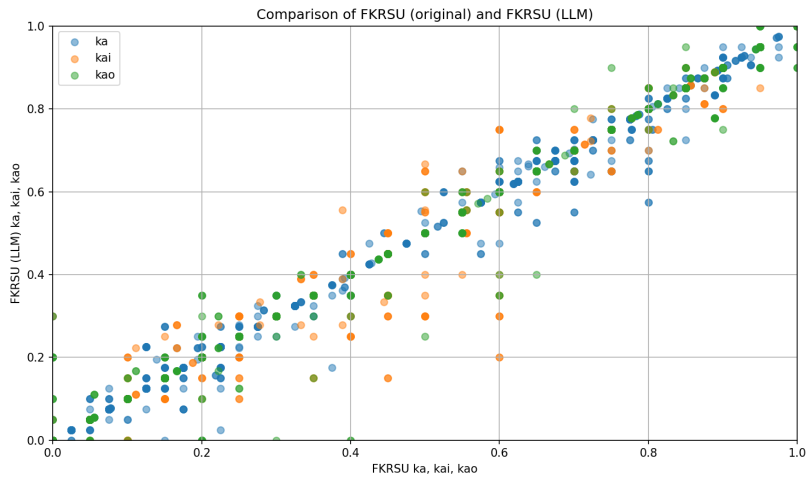


Figure 15: Comparison of capital restrictions (ka, kai, kao) values from FKRSU (original) and FKRSU (LLM Extension). Sample restricted to 100 countries and 32 instruments from FRKSU (original) between 1995-2019. Aggregated indices (ka, kai, kao) are reweighted using a similar methodology from FKRSU.

index (ka) by country-year. It shows most data points lie along the 45-degree line, indicating that the values of iBoP-S and FKRSU are highly consistent within the sample. Overall, this represents a compelling validation of the LLM methodology used in our work to construct cross-border flow restrictions.[50].

7.2 Comparison with other measures

We now compare our iBoP-C with measures coming from previous peer-reviewed studies that include a large panel of countries. Concretely, we focus on four well-known studies discussed in the Literature Review. The first three are the direct measures in Quinn and Toyoda (2008), Chinn and Ito (2006), and Fernandez et al. (2016))[51]. The fourth study is the indirect measures in Ilzetzki et al. (2021) who document the fraction of countries with dual, multiple or parallel exchange rates.[52]

[50] As part of the data release, we also provide users the updated and extended FKRSU indicators

[51] When computing statistics with the measures put together by Quinn (1997) we use simple averages between the current and capital account measures to make it more comparable to ours. Likewise, when using the measures by Fernandez et al. (2016), we use the most aggregate measures that averages inflow and outflow restrictions.

[52] Their index also draws on the IMF's AREAER complimented with the publications by Franz Pick over 1946-1998 and Pick and Sedillot (1971). It assigns a value of 1 to a country in each year when any one of three criteria is met. First, the country has an official (*de jure*) dual market for foreign exchange. Second, the country has a *de jure* system of multiple exchange rates. Third, there is an informal parallel

Figure 16 presents the time series of the five measures and Table 10 presents the corresponding correlations. When computing these descriptive statistics, we pool the data across countries. The most salient feature that comes out of these statistics is that our new measure commoves with the ones in previous studies, but the correlation is not perfect. In fact, all correlations are above or close to 0.9. It is also noteworthy that our index captures well the increase in restrictions surrounding the collapse of Bretton Woods and a stronger liberalization after the mid 1980s. This partly reflects the fact that, as mentioned before, the new measure enables us to capture more comprehensively the variety of tools used and captured in the eight categories that we cover.

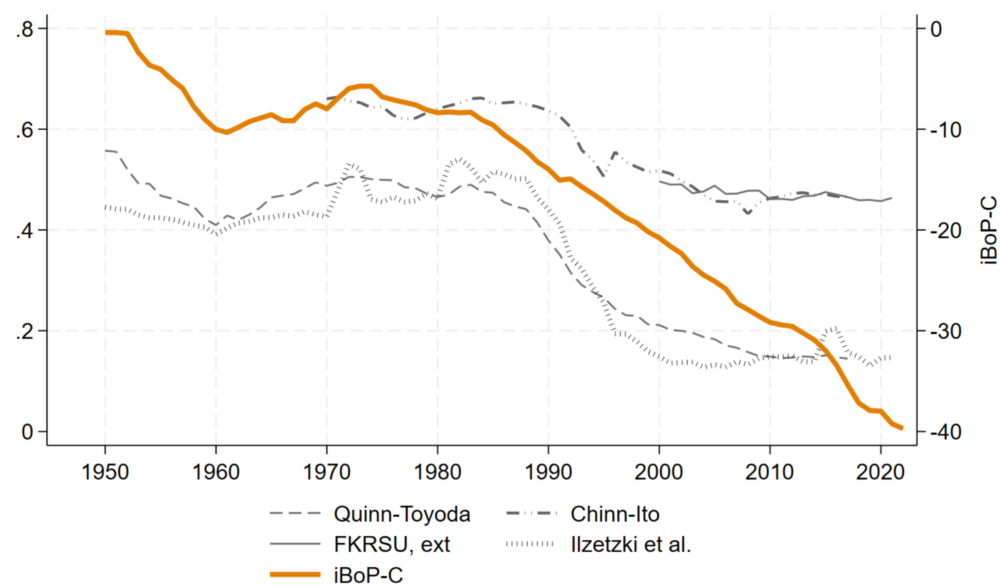


Figure 16: Comparison with other datasets mentioned in the text (Quinn and Toyoda 2008, Chinn and Ito 2006, Ilzetzki et al. 2021). "FKRSU, ext" refers to the measure by Fernandez et al. (2016) extended by our LLM as described in Section 7.1.

market (whether tolerated or illegal) and the parallel market has a premium of 10 percent or more within a 12-month moving window. The index takes a value of zero in any year when none of these three criteria is met.

	Quinn-Toyoda	Chinn-Ito	FKRSU, ext	Ilzetzki et al.	iBoP-C
Quinn-Toyoda	1				
Chinn-Ito	0.9683	1			
FKRSU, ext.	0.8230	0.5152	1		
Ilzetzki et al.	0.9543	0.9641	-0.1212	1	
iBoP-C	0.9672	0.9413	0.8801	0.8803	1

Table 10: Comparison with other datasets mentioned in the text (Quinn and Toyoda 2008, Chinn and Ito 2006, Ilzetzki et al. 2021). "FKRSU, ext" refers to the measure by Fernandez et al. (2016) extended by our LLM as described in Section 7.1.

8 Extensions and Refinements

8.1 Motivations for Capital Controls

The imposition of capital controls has been a topic of significant academic interest, but there remains a notable gap in the literature regarding a systematic, empirical account of *why* controls are introduced. While existing models of optimal capital controls emphasize motivations such as overborrowing for the use of inflow controls (see Bianchi and Mendoza (2018)), or to coordinate foreign investors not to run in case of a crisis with outflow controls (see Chang et al. (2024)), there is limited evidence on the actual reasons behind capital control measures as justified by government officials.

We systematically account for the official motivations in a subset of the measures considered. Our starting point is the work by Magud et al. (2018) who provide anecdotal evidence on the "*fears*" that drive policymakers to impose controls on cross-border flows.[53]

To explore the motivations for capital controls, we rely on a narrative approach utilizing official documents. Specifically, we analyze capital flow management (CFM) measures as categorized in the IMF taxonomy described before. This implies that we focus on the subset of measures labeled as “macro-critical” by IMF staff either introduced or changed post-2012. This includes 153 individual measures across 41 countries, spanning advanced economies, emerging markets, and developing economies.

[53]Their work highlights four fears: fear of exchange rate appreciation; fear of hot money; fear of large inflows; and fear of loss of monetary autonomy.

Our approach involves a manual search for official statements from governments regarding the motivations behind these measures. These statements were sourced from speeches, financial and monetary stability reports, IMF Article IV consultations, social media (e.g., Twitter), and other relevant channels. In total, we found official statements for 138 of the measures (90%), from which we identified motivations for 117 (76%). For the remaining measures, we supplemented our findings using ChatGPT to extract motivations.

While some measures are motivated by a single reason, others may reflect overlapping motivations, as capital controls often address multiple concerns simultaneously. Below is a detailed explanation of each of the six categories:[54]

- **Fear of Overborrowing**. This category refers to the concern that large foreign inflows may encourage excessive risk-taking, fuel asset-price bubbles, and create vulnerabilities to sudden withdrawals. These measures are often introduced either proactively (ex-ante) or reactively during a surge in inflows to prevent financial instability.

- **Fear of Disruptive Outflows**. Policymakers may introduce capital controls to prevent large-scale outflows, particularly during times of crisis or heightened financial uncertainty. These measures serve as precautionary measures to safeguard a country's foreign reserves or to stabilize the financial system. In some cases, they are used to mitigate the effects of capital flight during periods of economic turmoil.

- **Fear of Floating**. The fear of exchange rate volatility is a significant motivation for capital controls. This category encompasses multiple concerns, including a fear of exchange rate volatility leading to uncertainty in the economy; fear of appreciation, which can harm export competitiveness; fear of depreciation, which may reduce purchasing power or increase the burden of foreign-denominated debt. Lastly, any measure aimed at supporting a pegged exchange rate or maintaining monetary policy autonomy in the face of capital mobility would enter into this category.

[54] Our classification expands the four fears in Magud et al. (2018) framework to capture a broader range of economic, geopolitical, and long-term considerations that surfaced throughout our analysis. While Magud et al. (2018)'s fears of appreciation and loss of monetary autonomy overlap with our fear of floating, their fears of large inflows and of hot money bear resemblance to our fears of overborrowing and of disruptive outflows.

- **Geopolitics**. Capital controls may also be motivated by political or geopolitical considerations. These measures are often driven by national security concerns, including military conflicts or international sanctions. Governments may impose controls to restrict the flow of capital for strategic reasons or to comply with international mandates.

- **Long-term Goals**. Some capital controls are not reactive to short-term economic pressures but are instead aimed at long-term goals, such as fostering the stability and depth of domestic financial markets, or increasing capital account openness. These measures are designed to strengthen investor confidence and increase the resilience of the financial system, often leading to greater international connectedness over time. Despite regulating capital flows, these controls may de facto increase openness by fostering a more stable environment for investment.

- **Miscellaneous**. This category includes measures that do not fit neatly into the above classifications. Motivations for these measures include increasing tax revenues, or addressing other unique economic conditions not covered by the other categories.

Figure 17 reveals several key trends in the motivations behind the imposition of capital controls from our systematic analysis. The most common reasons for imposing restrictions are driven by concerns about disruptive outflows or volatile currency movements with 37% of the cases found to have this characteristic. This is followed by a fear of floating (31%) and of overborrowing (22%). Long-term goals not associated to business cycles are next with 20% of cases, and geopolitics and miscellaneous are the less frequent motivations with 13% and 3%, respectively.[55] It is notable, therefore, that about 1/3 of the motivations are not linked to the state of the business cycle.

Figure 18 explores how the motivations are related to income levels across countries, types of tools, direction of the flow, and the *de facto* exchange rate regime from Ilzetzki et al. (2021). Fears of floating and of disruptive outflows are more pervasive in low and middle income countries. Fear of overborrowing, in contrast, is more pervasive in high income countries (panel a). Bans and limits are the two types of

[55] As mentioned, the 6 types of motivations are not mutually exclusive, hence the percentages do not add to 100.

controls mostly used across all the motivations, with the exception of taxes, which is the most used in the case of fears of overborrowing. Surrender and repatriation requirements are the second most relevant type in the case of fears of disruptive outflows (panel b).

Fears of disruptive outflows, fear of floating, and geopolitical drivers are more characteristic of the motivations behind outflow measures compared to inflow measures. As expected, fear of overborrowing disproportionally impacts more inflow controls. Long-term goals are equally distributed among types of flows (panel c). Lastly, fears of overborrowing are more frequent in freely floating countries. "Freely falling" cases, as defined by Ilzetzki et al. (2021), are more pervasive when fears of floating and of disruptive outflows are identified (panel d).

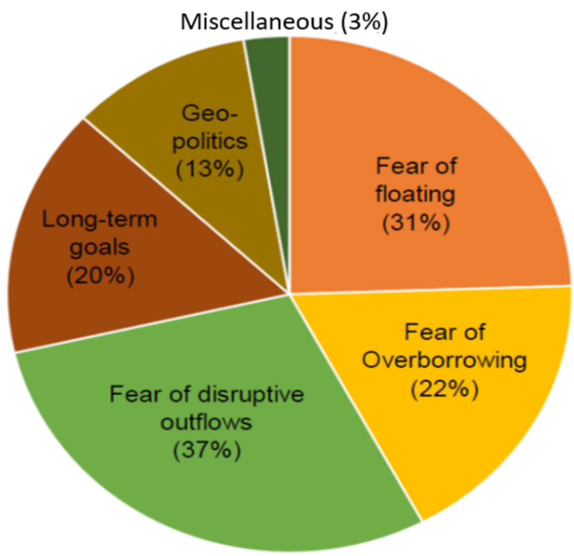
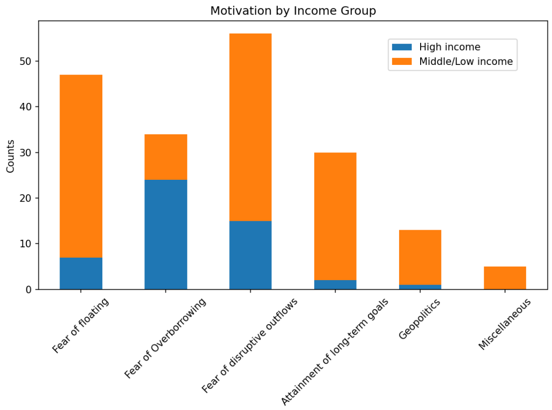


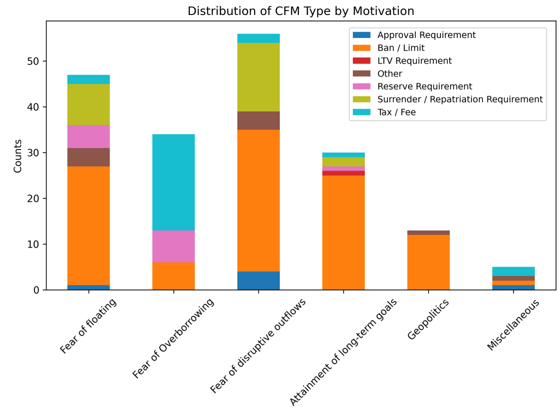
Figure 17: Motivations for use of Capital Controls.

8.2 Incorporating Intensity

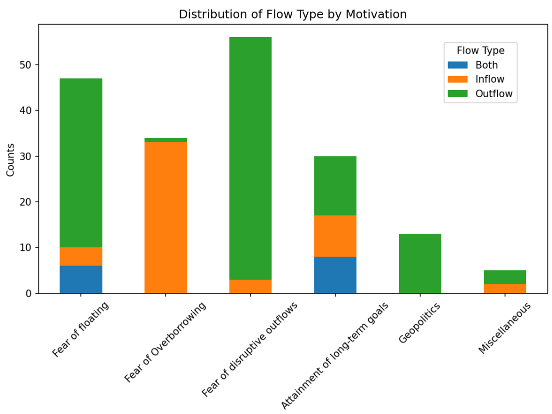
A common limitation of existing capital control measures is that they often capture only an extensive margin, e.g. through binary variables capturing the presence or not of controls, without measuring their economic *intensity*. By capturing changes to already active restrictions, the iBoP-C partly addresses this limitation insofar as the changes can capture the recalibration of already active restrictions that extensive margin measures would miss. However, because the measures are ultimately also



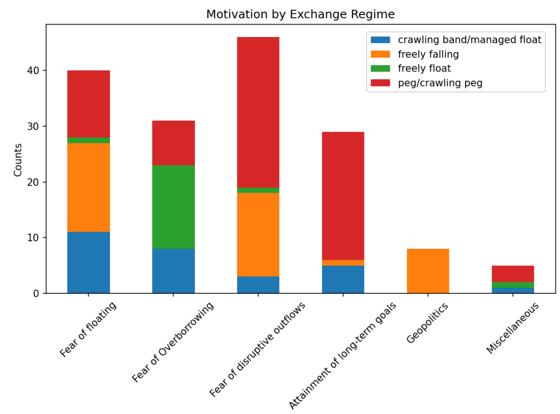
(a) By income group



(b) By type of CFM



(c) By direction of flow



(d) By exchange regime

Figure 18: Motivations of CFM policies in the Taxonomy, by income, type, direction of flow, and exchange regime.

binary (i.e., tightening/loosening) without accounting for the magnitude or severity of the measure, it continues to treat all reported changes as equal, regardless of their intensity. As a result, a minor documentation requirement and a sweeping prohibition receive the same weight, potentially biasing assessments of policy trends and their effects.

Previous work, most notably Quinn (1997), has emphasized that incorporating a measure of intensity can yield a more accurate depiction of policy restrictiveness. Quinn's methodology assigns an ordinal scoring of restrictions based on a subjective assessment of their scope, coverage, and severity. Such approach, however, has not been systematically extended to the breadth of instruments and historical coverage we analyze.

To address this limitation, we recompute our iBoP-C measure by weighting each reported change according to its intensity. Our methodology draws inspiration from Quinn's ordinal scoring approach but adapts it to the multi-category, multi-decade AREAER dataset. Each change in restriction is assigned an intensity score, $\alpha \in (0.1, 0.25, 0.5, 0.75, 1)$, based on the scope and nature of the measure: notification/-documentation, tax, quantitative limit, approval requirement, ban, respectively. The scoring thresholds were calibrated using a combination of expert review of AREAER entries and a generative LLM (GPT-4o-mini), supplemented with in-context examples, ensuring consistency across categories. The classification rules are customized to the eight categories of restrictions we track and were developed using hand-coded examples and Large Language Model (LLM) prompted on category-specific criteria (see Appendix D.2 for details). For instance, a complete prohibition on a broad class of capital transactions receives $\alpha = 1$, whereas a notification/documentation requirement receives $\alpha = 0.1$. Loosening measures are scored based on the intensity of the restriction being removed; tightening measures are scored based on the intensity of the new restriction. The resulting intensity-weighted iBoP-C aggregates net tightenings and loosenings as before, but with each measure scaled by its α-value.

Figure 19 compares the baseline (unweighted) and intensity-weighted iBoP-C. Many of the patterns documented at the category levels using the unweighted measures continue to hold when examining the intensity-weighted index. However, there are some interesting differences. Relative to the unweighted index, we find a slightly more pronounced tightening in the 70s and 80s and a more gradual liberalization trend in the subsequent periods. This indicates that many of tightenings put in place during the 1970s and 80s were more restrictive than the subsequent loosenings in the decades that followed. Complementary evidence presented in the Appendix (Fig.A11) decomposes the trend into the various levels of intensity ($\alpha's$) and shows how the stronger tightening in the 70s is associated with relatively more reliance on bans and approvals. The intensity weighted index indicates that the average country effectively loosened restrictions by around 10 measures as compared to around 40 measures by 2023 as implied by the unweighted measure.[56] Nonetheless, an important finding of this section is the findings established using our baseline unweighted measure largely hold even after accounting for different intensities across measures.

At the category level, we find that the intensity-weighted measures present a

[56]This feature of the data can also be explained to some extent by the assignment of weights less than one on average across measures, which modulates variation in the measure over time.

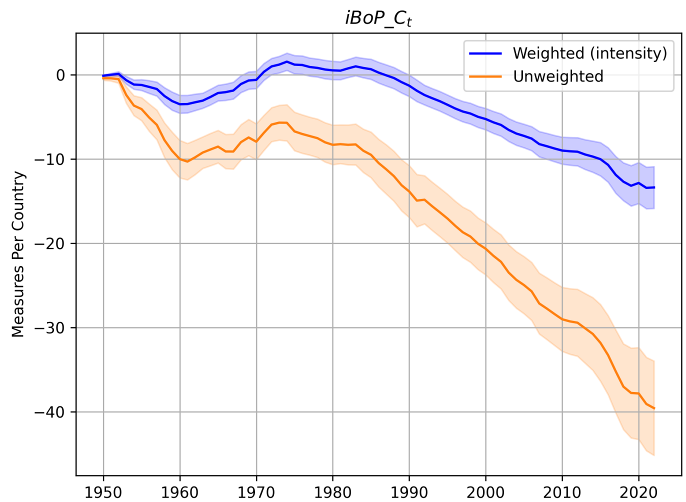


Figure 19: Cumulative net tightening, average, weighted by intensity

picture that is qualitatively consistent with those of the unweighted measures. This indicates that assigning each change an equal value as in our iBoP-C is a good approximation as there are no periods or countries that systematically use more/less intense controls. As with the aggregate iBoP-C index, we similarly find a more gradual liberalization over the past seven decades across all categories, except for FX markets and financial sector restrictions. Nonetheless, some differences emerge. We find a sharper tightening of FX market and financial sector restrictions among high income countries during the 1970s and 80s. We also find a more pronounced tightening in the intensity-weighted iBoP-C index for financial sector restrictions in the post-2010 sample for middle income countries. Finally, the intensity weighted measure suggest that outflow restrictions began to be effectively loosened starting from the 1990s, while the unweighted measure indicates a gradual loosening starting from the 1970s. We present an intensity-weighted version of the long-run stylized facts in Appendix B.2.

9 Conclusion

In this paper, we provide a detailed, systematic account of cross-border flow restrictions over the past seven decades, using modern machine learning techniques to analyze semi-structured official documents. By developing a new, high-resolution measure of restrictions, the iBoP, we are able to document the evolution of policies

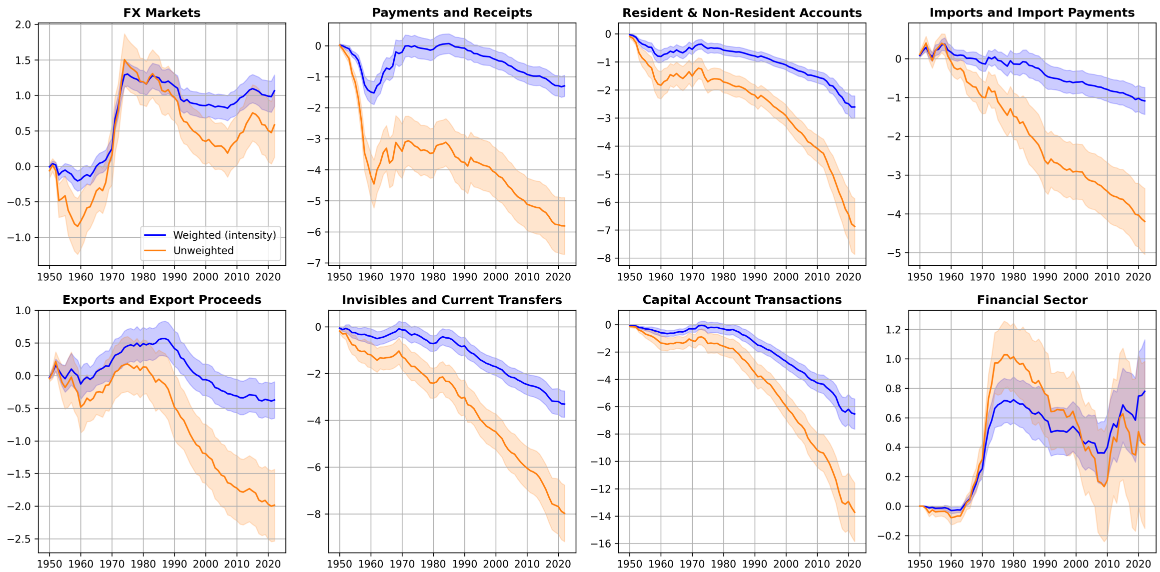


Figure 20: iBOP-C, weighted by intensity of restrictions by category

that have shaped the international monetary system, tracing changes across different countries and time periods with unprecedented granularity.

We document several stylized facts about the long-run evolution of cross-border flows and their use over time. First, we document that policymakers have relied on a broad set of tools to regulate cross-border flow restrictions, beyond those that directly impact the capital account. By systematically mapping this wider universe of policy tools, our work expands the conventional borders of what counts as a restriction on cross-border flows, offering a more comprehensive and empirically grounded view of countries' regulatory choices. These include restrictions related to current account transactions (import payments, export proceeds, and invisible transactions), but also measures targeted at the FX and financial markets, and methods of payments (use of gold and banknotes). We note a significant tightening of these restrictions during the end of the Bretton Woods era, a pattern that shows up only after considering these broader set of restrictions. Finally, we find significant heterogeneity in the use of cross-border flow restrictions across countries and over time. Higher income countries were more aggressive in pursuing the liberalization of these measures compared to lower income countries. Outflow restrictions were also the focus of most liberalization over time, but lower income countries continue to rely more heavily on outflow restrictions than higher income countries.

We also document how countries use the tools included in our new dataset. Its high frequency nature reveals that countries rarely use any of these tools in isolation, with over half of measures having a neighbouring action within ±10 days and over three-quarters within ±30 days—patterns. We also find that countries double their use during crisis, particularly outflow measures, where the effects are concentrated during sovereign debt and currency crises. Our simple regression framework also indicates that countries with lower institutional quality resort to BoP restrictions relatively more frequently, potentially compensating for other frictions in their policy tool kit.

We complement the construction of the new indices with a set of validation exercises and extensions that broaden their analytical value. First, we conduct a direct validation test by restricting iBoP-S to the subset of categories used in Fernandez et al. (2016). In this controlled setting, our LLM-based classifications reproduce the FKRSU series almost perfectly. Second, we compare the full iBoP-C series with the leading indicators of capital account regulation and show that while the indices co-move strongly, iBoP-C delivers substantially richer cross-country and time-series variation owing to its broader scope and higher frequency. Third, we extend the analysis by systematically extracting policymakers' stated motivations for adopting restrictions and, separately, by incorporating intensity weights that differentiate minor procedural changes from sweeping prohibitions. These refinements reveal a smoother long-run liberalization and sharper tightening episodes—especially during the Bretton Woods collapse—than suggested by unweighted measures. Together, these exercises confirm both the accuracy of our methodology and the additional insight gained from exploiting the full richness of the AREAER.

Looking ahead, several promising avenues remain. A natural next step is to further leverage LLMs to construct a comprehensive stance index that extends back to the 1950s, consolidating information on both the frequency and intensity of restrictions into a unified measure. Our dataset also opens the door to empirical applications assessing the effectiveness of restrictions in shaping capital flows, exchange rate pressures, and crisis dynamics, and to studying their interactions with monetary and macroprudential policies. Finally, the approach can be adapted to capture not only the *de jure* but also the *de facto* dimension of cross-border restrictions, thus offering a more complete picture of international financial integration. These extensions will deepen our understanding of how countries manage capital mobility and the implications for the stability of the international monetary system.

References

Acosta-Henao, M., Alfaro, L., and Fernández, A. (2025). Sticky capital controls. *Journal of International Economics*.

Adrian, T., Erceg, C. J., Kolasa, M., Linde, J., and Zabczyk, P. (2021). A quantitative microfounded model for the integrated policy framework. *IMF Working Paper*, (2021/292).

Ahmed, S. and Zlate, A. (2014). Capital flows to emerging market economies: A brave new world? *Journal of International Money and Finance*, 48:221–248.

Alesina, A., Furceri, D., Ostry, J. D., Papageorgiou, C., and Quinn, D. P. (2024). Structural Reforms and Elections: Evidence from a World-Wide New Dataset. *Journal of the European Economic Association*, 22(4):1936–1980.

Alfaro, L., Chari, A., and Kanczuk, F. (2017). The real effects of capital controls: Firm-level evidence from a policy experiment. *Journal of International Economics*, 108(C):191–210.

Andreasen, E., Bauducco, S., and Dardati, E. (2024). Capital controls and firm performance. *Journal of International Economics*, page 103897.

Aruoba, S. B. and Drechsel, T. (2024). Identifying monetary policy shocks: A natural language approach. Working paper, University of Maryland.

Baba, C., Cervantes, R., Darbar, S. M., Kokenyne, A., and Zotova, V. V. (2025). New Measures of Capital Flow Restrictions — AREAER Indices. Mimeo.

Baba, C. and Kokenyne, A. (2011). Effectiveness of Capital Controls in Selected Emerging Markets in the 2000's. IMF Working Papers 2011/281, International Monetary Fund.

Basu, S., Boz, E., Gopinath, G., Roch, F., and Unsal, F. (2020). A conceptual model for the integrated policy framework. *IMF Working Paper*, (2020/121).

Bergant, K. and Forbes, K. (2023). Policy packages and policy space: Lessons from covid-19. *European Economic Review*, 158:104499.

Bianchi, J. and Lorenzoni, G. (2022). The prudential use of capital controls and foreign currency reserves. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics: International Macroeconomics*, chapter 6.

Bianchi, J. and Mendoza, E. (2020). A Fisherian Approach to Financial Crises: Lessons from the Sudden Stops Literature. *Review of Economic Dynamics*, 37:254–283.

Bianchi, J. and Mendoza, E. G. (2018). Optimal time-consistent macroprudential policy. *Journal of Political Economy*, 126(2):588–634.

Binici, M. and Das, M. (2021). Recalibration of capital controls: Evidence from the IMF taxonomy. *Journal of International Money and Finance*, 110:102252.

Bordo, M. D. (2016). The Operation and Demise of the Bretton Woods System; 1958 to 1971. Economics Working Papers 16116, Hoover Institution, Stanford University.

Capelle, D. and Pellegrino, B. (2025). Unbalanced financial globalization. Working paper, International Monetary Fund and Columbia Business School. First version: September 2021; This version: June 2025.

Chang, R., Fernández, A., and Martinez, H. (2024). Capital Controls on Outflows: New Evidence and a Theoretical Framework. NBER Working Papers 32877, National Bureau of Economic Research, Inc.

Chinn, M. D. and Ito, H. (2006). What Matters for Financial Development? Capital Controls, Institutions, and Interactions. *Journal of Development Economics*, Volume 81, Issue 1:163–192.

Clayton, C., Coppola, A., Maggiori, M., and Schreger, J. (2025). Geoeconomic pressure. Working paper, National Bureau of Economic Research.

Dell, M. (2024). Deep learning for economists. Technical report, National Bureau of Economic Research.

Eichengreen, B. and Rose, A. (2014). Capital controls in the 21st century. *Journal of International Money and Finance*, 48:1–16.

Erten, B., Korinek, A., and Ocampo, J. A. (2021). Capital controls: Theory and evidence. *Journal of Economic Literature*, 59(1):45–89.

Erten, B. and Ocampo, J. A. (2017). Macroeconomic Effects of Capital Account Regulations. *IMF Economic Review*, 65(2):193–240.

Fabiani, A., Fernández, A., and Gupta, N. (2025). Capital Controls, Misallocation and Risk: Cross-country Evidence from Firms' Bond Issuanc. Mimeo.

Fabiani, A., Piñeros, M. L., Peydró, J.-L., and Soto, P. E. (2022). Capital controls, domestic macroprudential policy and the bank lending channel of monetary policy. *Journal of International Economics*, 139(C).

Fang, H., Li, M., and Lu, G. (2025). Decoding china's industrial policies. Technical report, National Bureau of Economic Research.

Fernandez, A., Klein, M., Rebucci, A., Schindler, M., and Uribe, M. (2016). Capital Control Measures: A New Dataset. *IMF Economic Review*, Volume 64:548–574.

Fernández, A., Rebucci, A., and Uribe, M. (2015). Are capital controls countercyclical? *Journal of Monetary Economics*, 76:1–14.

Forbes, K., Fratzscher, M., and Straub, R. (2015). Capital-flow management measures: What are they good for? *Journal of International Economics*, 96(S1):76–97.

Forbes, K. J. (2007). One cost of the chilean capital controls: Increased financial constraints for smaller traded firms. *Journal of International Economics*, 71(2):294–323.

Ghosh, M. A. R., Ostry, M. J. D., and Qureshi, M. S. (2017). Managing the Tide: How Do Emerging Markets Respond to Capital Flows? IMF Working Papers 2017/069, International Monetary Fund.

Guisinger, A. and Brune, N. E. (2017). The diffusion of financial liberalization in developing countries. *Foreign Policy Analysis*, 13(1):1–27.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., and Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks.

Ilzetzki, E., Reinhart, C., and Rogoff, K. (2021). Rethinking exchange rate regimes. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*.

Klein, M. W. (2012). Capital controls: Gates versus walls. Technical Report w18526, National Bureau of Economic Research.

Korinek, A. (2025). Generative AI for economic research: Use cases and implications. *Journal of Economic Literature*. Forthcoming.

Laeven, L. and Valencia, F. (2018). Systemic Banking Crises Revisited. *IMF Working Paper*, 18/206.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Levy-Yeyati, E. and Sturzenegger, F. (2003). To Float or to Fix: Evidence on the Impact of Exchange Rate Regimes on Growth. *American Economic Review*, 93(4):1173–1193.

Li, W. (2025). A new measure of capital flows management in emerging and developing. Mimeo.

Magud, N. E., Reinhart, C. M., and Rogoff, K. S. (2018). Capital controls: Myth and reality. *Annals of Economics and Finance*, 19(1):1–47.

Müller, K. (2023). Electoral cycles in macroprudential regulation. *American Economic Journal: Economic Policy*, 15(4):295–322.

Ottonello, P., Song, W., and Sotelo, S. (2024). An anatomy of firms' political speech. Working paper, National Bureau of Economic Research.

Pasricha, G. K. (2012). Recent trends in measures to manage capital flows in emerging economies. *North American Journal of Economics and Finance*, 23(3):286–309.

Pasricha, G. K. (2022). Estimated policy rules for capital controls. *Journal of International Money and Finance*, 122:102593.

Pasricha, G. K., Falagiarda, M., Bijsterbosch, M., and Aizenman, J. (2018). Domestic and multilateral effects of capital controls in emerging markets. *Journal of International Economics*, 115:48–58.

Quinn, D. (1997). The Correlates of Change in International Financial Regulation. *American Political Science Review*, Volumne 91, Issue 3:531–551.

Quinn, D. P. and Toyoda, A. M. (2008). Does capital account liberalization lead to growth? *The Review of Financial Studies*, 21(3):1403–1449.

Rebucci, A. and Ma, C. (2020). Capital controls: A survey of the new literature. *Oxford Research Encyclopedia of Economics and Finance*.

Reinhart, C. M. and Rogoff, K. S. (2004). The Modern History of Exchange Rate Arrangements: A Reinterpretation. *The Quarterly Journal of Economics*, 119(1):1–48.

Scartascini, C., Cruz, C., and Keefer, P. (2021). The database of political institutions (dpi): 2020. Version covering 1975–2020.

Schindler, M. (2009). Measuring financial integration: A new data set. *IMF Staff Papers*, 56(1):222–238.

The PRS Group (2023). International Country Risk Guide.