Regulation & Governance

Regulation & Governance

ORIGINAL ARTICLE

On the Performance-Based Legitimacy of Financial Action Task Force: A Quantitative Exploration

Guo Cheng[1] | Bo Wen[2]

[1]Department of Economics, University of Chicago, Chicago, Illinois, USA | [2]Department of Government and Public Administration, University of Macau, Macau SAR, China

Correspondence: Bo Wen (bowen@um.edu.mo)

Received: 1 October 2024 | **Revised:** 28 August 2025 | **Accepted:** 22 September 2025

Funding: This work was generously supported by the National Natural Science Foundation of China under Grant Number [72004189], by the University of Macau's Start-up Research Grant (SRG) under Grant Number [SRG2024-00005-FSS], and by the University of Macau's Multi-Year Research Grant—Level II (Grant No. MYRG-GRG2024-00033-FSS).

Keywords: compliance and effectiveness | Financial Action Task Force (FATF) | informal governance | legitimacy | soft laws

ABSTRACT

How legitimate is Financial Action Task Force (FATF)? We address this question with cross-country evidence spanning performance-wise aspects—*standards, evaluation, strategy*. FATF's legitimacy needs strengthening. *Standards-wise,* compliance to FATF's Recommendations generally contributes to both FATF's and external effectiveness proxies, yet the latter has much weaker evidence. Granularly, FATF's Recommendations pose challenges for good-faith developing states, since capacity constraints hinder compliance and effectiveness conversion. *Evaluation-wise*, the effectiveness appraisal may be biased due to heavy reliance on compliance ratings, evidenced by weak correlations with external corruption control and financial transparency indicators. *Strategy-wise*, though enhanced follow-up boosts persistent compliance, the dominantly punitive strategy system is detrimental to good-faith yet low-capacity states. Theoretically, we build an extended framework on Mitchell's theory by deepening the analysis of effectiveness and justifying its applicability to non-binding institutions. We also enrich the dialogue on the legitimacy of international organizations by presenting a quantitative, objective approach using official data.

1 | Introduction

Financial security, transparency, and traceability are conducive to national prosperity (Vishwanath and Kaufmann 2001; Alcaide Muñoz et al. 2017; Janský et al. 2023), and essential for a thriving global society (Farazmand 1999; Mudacumura 2013). In the contemporary landscape of digital economy, countering money laundering is increasingly vital for tracing, identifying, and preventing illicit proceeds (Gelemerova 2009; Mugarura 2011). For instance, Sharman and Chaikin (2009) demonstrate that corruption, as a predicate crime, could be effectively reduced through anti-money laundering (AML) systems. The advent of digital finance introduces new complexities to AML tasks, such as ensuring privacy and data protection while leveraging technologies for enhanced risk assessments (FATF 2021). Additionally, the anonymity and opacity inherent in some digital financial transactions pose significant challenges to detecting and preventing money laundering (IMF 2023).

The Financial Action Task Force (FATF), as the global financial security watchdog, has coordinated and advanced international AML efforts for over three decades. Although FATF's governance receives widespread recognition, skepticism on its legitimacy exists. Yet, the existing literature does not address the concerns in a systematic, evidence-based and theory-driven way. Empirical research on FATF emphasizes

Guo Cheng and Bo Wen contributed equally to this study. The names are listed in alphabetical order to reflect this equal contribution.

regional characteristics in AML tasks compliance but lacks theoretical grounds, and often overlooks the relationship with effectiveness and FATF's influence. Conversely, critiques on the legitimacy of FATF predominantly rely on case-specific and speculative arguments, lacking cross-country evidence. Concurrently, the theoretical realm of global financial governance suffers from a lack of research (Chong and Lopez-De-Silanes 2015; Woo et al. 2016; Brooks et al. 2019; Jones and Knaack 2019; Abedin et al. 2024). Although the general literature on international organizations (IO) does provide useful perspectives to examine legitimacy, the mainstream legitimacy evaluations rely on elements not easily quantifiable (normative dimension), or surveyed perceptions (subjective dimension), which are unsuitable for an objective, quantitative inquiry we aim for.

In this article, we demonstrate the applicability of Mitchell's compliance theory to examine the governance of non-binding institutions such as FATF, though it was originally developed for binding international environmental laws and agreements (IEL/IEA). We enrich this framework by an extensive discussion on effectiveness, covering the econometric definition, measurement methodology, effective standards design, the relationship with compliance and the mediation from capacity factors. We also contribute to the ongoing discourse on the legitimacy of IO, particularly informal institutions like FATF, by showcasing the potential of objective, quantitative approaches in assessing their legitimacy. The analysis focuses on three critical criteria of performance—(a) effective standards design, (b) impartial and precise evaluation, (c) diversified strategies to boost compliance and effectiveness, consistent with our extended framework built upon Mitchell's theory. Data-wise, we use official, publicly available statistics from FATF, World Bank (WB), United Nations Development Program (UNDP), International Monetary Fund (IMF), Basel Institute on Governance and other credible sources, mitigating concerns about internal and external validity common in survey or experimental data.

The legitimacy of FATF is contested. First with respect to *standards*, though the compliance to FATF's Recommendations is strongly correlated with FATF's effectiveness proxy from a cross-sectional analysis, the panel analysis evidence suggests a much weaker link with external effectiveness indicators such as Basel AML Index (non-FATF component), casting doubt on the effectiveness of standards at the global level. In a nuanced dimension, evidence indicates that FATF's Recommendations pose challenges for good-faith nations with capacity constraints, hindering compliance and, as uniquely argued in our study, impeding the conversion from compliance to effectiveness.

Second in terms of *evaluation*, we recognize the potential biases from the dominant reliance on compliance results when evaluating the effectiveness. The correlations between FATF's effectiveness index and Basel indicators for corruption control and financial transparency are low or even negative, further revealing the existence of measurement biases. Yet, the possibility of political biases against developing nations is largely ruled out because of the peer-reviewed evaluation mechanism among the regional body members. On the other hand, compliance ratings may be inflated by superficial compliance in developing nations, providing an alternative explanation for their low compliance-to-effectiveness conversion ratio, beyond the capacity mediation mechanism.

Third regarding *strategy*, a causal analysis using an event study design demonstrates that the enhanced follow-up policy fosters sustained compliance, affirming its efficacy. However, the dominantly punitive strategy system is problematic for low-performing yet good-faith states with capacity constraints, arousing suspicions on political manipulation and compromising legitimacy beliefs. Yet, since the presidency of Elisa de Anda Madrazo of Mexico, FATF has increasingly adopted a supportive strategy system, emphasizing cognitive and generative approaches. This shift is sufficiently justified by our theory and empirical evidence.

In the subsequent sections, we first examine FATF's regime structure and establish theoretical foundations for compliance, effectiveness, and legitimacy. Next, we present the data, formulate research questions and hypotheses, and detail the research design. Empirical findings and discussion follow, culminating in a synthesis of evidence with policy recommendations in the institutional reflection section.

2 | FATF: History and Structure

Established in 1989 by the G7, FATF is an intergovernmental organization dedicated to AML and combating the financing of terrorism (CFT). For over three decades, FATF has developed comprehensive standards to prevent illicit financial activities and mitigate associated social harms, including terrorism, corruption, drug trafficking, and human trafficking. By fostering global cooperation, FATF seeks to enhance international financial security and unite jurisdictions in countering these offenses. As of July 2025, FATF comprises 38 member jurisdictions and 2 regional organizations. Additionally, its governance framework includes nine associate members, known as FATF-Style Regional Bodies (FSRBs). Together, FATF and FSRBs conduct evaluations across different regions worldwide. Although not a legal regulator, FATF has garnered wide support from G20 countries, with over 200 jurisdictions joining its network. Its regularly updated standards, peer-reviewed evaluation system, and formidable sanction mechanisms are key factors underpinning its authority.

However, FATF faces criticism for perceived discrimination, political manipulation, and lack of credibility in its evaluations. Littrell (2022) finds that FATF's effectiveness ratings respond most strongly to whether the assessed jurisdiction is a black-majority country. Politically, Chohan (2019) states that manipulation pervades the operations of FATF, pointing out that India is seeking co-chairmanship in the FATF's Asia-Pacific network to work against Pakistan. Case-Ruchala and Nance (2020) contend that skepticism about the validity of FATF's sanction lists stems from perceptions that political preferences eclipse technical expertise in their formulation. Regarding evaluation reliability, Levi et al. (2018) criticize that FATF's evaluations rely little on comparable, consistent and quantitative data across countries, casting doubt on its credibility. This aligns with Pontes

et al. (2022), who argue that strong performance in FATF's metrics falls far short of achieving true effectiveness in practice. The evidence suggests that FATF may have strayed from its principled objective, necessitating a rigorous examination of its legitimacy.

Before evaluating the legitimacy of FATF, a comprehensive understanding of its structure is essential. In FATF's regime, the *mutual evaluation* process constitutes the central component. This process assesses states' compliance with FATF's Recommendations and, from the fourth round, the effectiveness of their AML systems. Peer-reviewed evaluation reports summarize performance, with jurisdictions within the same FSRB mutually evaluating their AML progress. A standard mutual evaluation cycle typically spans 8 to 10 years. Between April 1998 and July 2025, three full evaluation rounds were completed, while the fourth round is nearing completion.

The methodology for mutual evaluations has undergone continuous refinement. Initially, FATF's 40 Recommendations, outlining AML measures, was the only component of the assessment. Jurisdictions are assessed on each recommendation, receiving a grade reflecting their technical compliance level. Besides *Not Applicable*, there are four possible grades—*Compliance (C), Large Compliance (LC), Partial Compliance (PC), No Compliance (NC)*. This reflects a typical new public management (NPM) approach, emphasizing performance metrics focused on activities and processes rather than goals achieved (Garland 1996).

Subsequently, nine additional recommendations on combating terrorist financing were incorporated into the original set. The fourth round of mutual evaluations consolidates these 49 recommendations into 40, organized across seven thematic sections. Notably, this round introduces a pioneering set of *Immediate Outcomes* for *effectiveness* assessment—marking a strategic shift from NPM principles towards an outcome-oriented governance framework. Possible grades on effectiveness include four levels—*High Effectiveness (HE), Substantial Effectiveness (SE), Moderate Effectiveness (ME), Low Effectiveness (LE)*. Similar to the sections in recommendations, the immediate outcomes are grouped into three *intermediate outcomes* that contribute to the high-level objective of AML. Technical compliance and effectiveness ratings are crucial, as they directly determine whether a jurisdiction undergoes enhanced follow-up or even gray-list designation.

This study focuses on the *fourth-round* mutual evaluation for the following reasons. First, the evaluation methodology comprehensively includes both reformative standards and effectiveness indicators, which (ideally) avoids measuring activities solely under the NPM mode. Meanwhile, FATF applies a risk-based approach (previously rule-based), taking the national risk profiles into account when measuring technical compliance. Second, as of July 2025, the fourth round is the most current cycle of mutual evaluations, capturing up-to-date global developments in AML systems and their evolving implementation challenges. Third, the fourth-round dataset, publicly accessible and systematically organized with minimal missing values, ensures high quality. Finally, the existing literature primarily examines prior rounds, with scant attention to the fourth round. We pioneer the analysis on the fourth-round data, delivering original research outputs to facilitate future studies.

The fourth-round mutual evaluation consists of three major procedures—initial mutual evaluation, interim follow-up process, final assessment. In the latest version of FATF's Procedures Manual (2023), the final assessment is replaced with the fifth-round initial evaluation to improve efficiency.

Among the above procedures, the initial mutual evaluation is the foremost assessment. A publication named *Mutual Evaluation Report* will be published after initial evaluation. The results on this report directly determine whether a jurisdiction will enter a *regular follow-up* (the default case) or an *enhanced follow-up*. Furthermore, for jurisdictions that fall considerably below the standards, they might be listed as the *Jurisdictions under Increased Monitoring*—the well-known *Gray List*. Black or gray-listing has been widely studied and proven to carry significant consequences, while little research examines the follow-up process, where we aim to contribute. The criteria for a state entering enhanced follow-up are outlined as follows (FATF 2022): in the initial mutual evaluation report,

1. In *Recommendations*, the state has eight or more NC/PC ratings OR
2. In *Recommendations*, any of R.3, 5, 10, 11, 20 is/are rated NC/PC OR
3. In *Immediate Outcomes*, the state has seven or more LE/ME ratings OR
4. In *Immediate Outcomes*, the state has four or more LE ratings.

Supervision intensity differs between regular and enhanced follow-up. Regular follow-up only requires one report before the final assessment, but enhanced follow-up typically requests three reports. In practice, we observe a much higher number of enhanced follow-up documents. During interim follow-up, only technical compliance is re-assessed, while effectiveness remains not updated. Both technical compliance and effectiveness are re-rated during the final assessment or the fifth-round initial evaluation, which marks the completion of the fourth round and the beginning of the fifth round (Figure 1).

3 | Legitimacy of FATF: Theory and Existing Evidence

3.1 | IO Literature

The study on FATF's governance is closely related to the literature of international organizations (IO), which has been a cornerstone of international relations (IR) research and recently intrigues scholarship in public administration (PA) (Bauer et al. 2016; Moloney and Rosenbloom 2020). The IR scholars explore why IOs exist and their role in global governance with approaches of realist, liberal, rational-institutionalist, constructivist, and so forth. The worldwide recognition and network expansion of FATF is consistent with the liberalist and constructivist view that IOs are powerful with autonomy to build international ties,

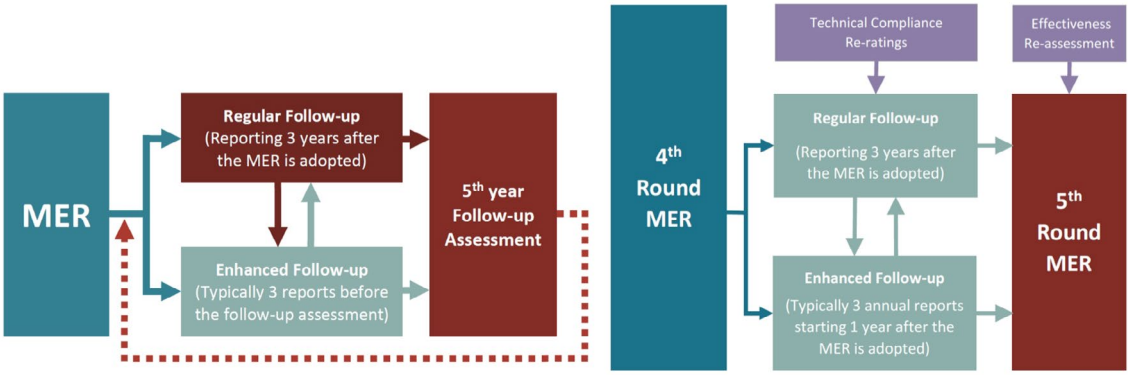

FIGURE 1 | FATF fourth-round process. *Note:* the above process charts come from the FATF Fourth-round Procedures Manual January 2021 (panel a) and February 2023 (panel b), respectively. "MER" stands for the initial mutual evaluation report. After the initial evaluation, member states either enter the regular follow-up or enhanced follow-up depending on the technical compliance and effectiveness ratings. Technical compliance is re-rated during the follow-up process while effectiveness is only re-rated during the fifth-year final assessment (before reform) or the initial evaluation of the fifth round (after reform). The reform streamlines the process by skipping the final assessment which is redundant with the next round initial evaluation.

create social knowledge and norms (Moravcsik 1997; Barnett and Finnemore 1999), while the criticism of political manipulation is in line with the realist view that IOs are tools employed by powerful states to advance their national interests (Jervis 1998; Sterling-Folker 2006; Mearsheimer 2017).

Legitimacy, a pivotal theme, is typically analyzed from *normative* and *subjective* dimensions (Karlsson-Vinkhuyzen and Vihma 2009). *Normative* legitimacy (level) is the extent to which IOs have the rightful authority to govern, based on their adherence level to moral principles like justice and democracy (Steffek 2003; Buchanan and Keohane 2006; Beetham 2013; Dellmuth et al. 2019). Specific aspects cover the validity of decisions (Zürn 2004), the justification for interventions, the sources of authority, the principles to limit power (Kyllönen 2006; Beetham 2013), and the level of transparency (Karlsson-Vinkhuyzen and Vihma 2009). Though relatively objective and qualitatively comparable, many of these elements remain vague and challenging to quantify.

More popularly, scholars adopt the *subjective* form of legitimacy. Hurd (1999) defines legitimacy as the perception that an IO's rules ought to be obeyed. Subsequent scholars similarly define legitimacy as the belief that the IO's authority is appropriately exercised (Tallberg and Zürn 2019). Recent empirical work examines how institutional characteristics and qualities (Dellmuth et al. 2019; Jongen and Scholte 2024), trust and distrust (Verhoest et al. 2025) determine legitimacy beliefs through surveys or experiments. These mainstream legitimacy forms are subjective judgment which may vary considerably, since every state or individual evaluate legitimacy according to their own criteria. Yet, legitimacy beliefs, widely recognized as stemming from IO *purpose, procedure, and performance (3P)* (Scharpf 1999; Scott 2005; Hurd 2008; Bernauer and Gampfer 2013; Binder and Heupel 2015; Lenz and Viola 2017; Scholte and Tallberg 2018; Tallberg et al. 2018; Tallberg and Zürn 2019; Dellmuth et al. 2019), can be validly proxied by these elements. Quantifying the 3P component(s) with objective evidence enables a persuasive inference of the legitimacy level, as demonstrated by our FATF case study in what follows.

Not every IO is formal with legal enforcement power. Vabulas and Snidal (2013) propose a continuum of international institutions with different formality, from decentralized cooperation ("low end"), informal arrangements with non-binding agreements ("middle point") to highly formalized organizations ("high end"). Conceptualizing institutional formality as a spectrum rather than a dichotomy is pioneering, which blurs the sharp distinction between formal IOs (FIOs) and informal IOs (IIOs). Moreover, they emphasize the increasing importance of IIOs in world politics like "G groups" and its derivatives. FATF, established by the G7, is a vivid example of IIOs that play an important role in global financial governance. When deciding the formality of an IO, a designer would find IIOs can maintain more flexibility, grant states greater autonomy, lower short-term transaction costs, minimize bureaucracy, and manage high certainty, compared with the formal counterparts. However, the FIOs have been dominantly researched, leaving limited attention to the influence of IIOs (Stone 2013; Vabulas and Snidal 2013), where our study of FATF contributes.

3.2 | FATF Literature

Beyond case studies of specific states discussed earlier in the *FATF: History and Structure* Section, the FATF-focused literature includes quantitative research, such as analyses of technical compliance in prior rounds. Researchers focus on the trends and region-level determinants of technical compliance. Johnson (2008) compares the self-assessment results pre 2003 and the mutual evaluation results post 2003 and finds that the technical compliance level declined. Verdugo Yepes (2011) examines third-round compliance using multiple regressions to identify correlated variables, finding that domestic regulatory quality and economic development positively influence compliance. Additionally, Verdugo Yepes notes that the strong

correlation with development highlights a potential weakness in the standards or their compliance assessment. Mekpor et al. (2018) conduct research similar to Verdugo Yepes (2011), but utilize data from both the third and fourth rounds. They similarly identify factors such as technology, regulatory quality, and banking sector structure as significant determinants of compliance. The current empirical evidence has two limitations: (a) regional factors are chosen without solid theoretical grounds, (b) the causal relation or association with FATF's institutional governance is largely ignored.

Another important aspect of AML tasks is effectiveness. Although introducing effectiveness ratings is acknowledged as a positive step by FATF, concerns regarding the credibility of these measurements persist. Pol (2018) questions FATF's effectiveness methodology by arguing that it miscounts outputs and activities as outcomes. Littrell (2022) finds that though there is no evidence of racial bias in compliance ratings, the effectiveness grades are troublingly low for black-majority nations. While for non-FATF effectiveness indicators, Bensassi and Raz (2025) evaluate the correlation between trade-related fraud activities (proxied by the trade gap) and compliance with FATF's standards. They find that 15.3% of the fraud reduction is associated with the adoption of FATF's Recommendations. Yet, limited literature systematically investigates the relationship between technical compliance and effectiveness measured by either FATF or non-FATF sources.

3.3 | Mitchell's Framework and Extensions Proposed

When examining FATF's governance, the literature rarely proceeds with a theoretical framework. There is no sufficient theory analyzing financial governance, especially in the context of money laundering (Chong and Lopez-De-Silanes 2015; Woo et al. 2016; Brooks et al. 2019; Jones and Knaack 2019; Abedin et al. 2024). The general IO literature does provide perspectives on the influence and legitimacy, but we find empirical inquiries based on objective evidence rare in the mainstream works. In this study, we build upon Mitchell's compliance theory, originally discussed in the context of international environmental laws and agreements (IEL/IEA). We demonstrate that Mitchell's theory is valuable in assessing the legitimacy of FATF, and enrich the framework by drawing inspirations from econometrics and the IO literature. The extended framework is applicable to other IIOs or FIOs from the perspective of compliance and effectiveness.

Originating from Mitchell's 1993 framework and refined over recent decades, the *compliance theory* defines compliance, discusses sources of non-compliance, explores the relationship with IEL/IEA legitimacy, and provides insights into institutional design. In our study, FATF's regional technical compliance and effectiveness ratings fit seamlessly in this framework. Although FATF's regime consists of standards that are not legally binding, this difference does not negate the extension of Mitchell's framework to FATF fundamentally.

In the spirit of "continuous formality" (Vabulas and Snidal 2013), FATF, as an informal IO with soft laws, has more flexibility in standards implementation and punishment mechanism. Violation with a treaty carries inevitable legal responsibilities, while unsatisfactory compliance with FATF's standards may also result in sanctions through black or gray-listing, leading to reputation damage, economic loss from reduced trade, and political isolation from other member states. In this sense, FATF's standards are also "binding", not from a legal channel, but from peer pressure. Meanwhile, the evaluated jurisdictions voluntarily join FATF's network, indicating their respect and willingness to accept FATF's governance. Through both voluntary participation and peer supervision among member states, FATF functions almost equivalently to an "international law regulator", even without legally binding power. In the following paragraphs, we discuss Mitchell's framework for general rules (either legally binding or non-binding), though the original arguments are made regarding binding IEL/IEA.

Mitchell (1993) defines compliance as the degree to which a state adheres to a regime's explicit rules. Mitchell identifies two types of compliance. The first is *rule-induced compliance*, which is the portion of compliance directly resulting from the rules. We generalize this concept to *IO-induced compliance*, which comes from the influence of the IO, either the rules, enforcement measures, or other channels. The second is *coincidental compliance*, which occurs independently of the IO's influence. Non-compliance has two forms. *Good-faith non-compliance* occurs when states exert genuine efforts but fail to meet the requirements. Mitchell (2008) argues that one important reason for good-faith non-compliance is incapacities, which can be administrative, financial, or technological. Similar arguments are made by other scholars (Karlsson-Vinkhuyzen and Vihma 2009), emphasizing on technical, human, and political capacity building. By contrast, *intentional non-compliance*, where states reject or minimally adhere to the standards, may stem from cost–benefit trade-offs or ideological differences.

Regarding effectiveness, we employ the definition of *problem-solving effectiveness*, referring to the degree that a problem is resolved by the means (Young 1999). There is caveat when discussing effectiveness. In Mitchell's discussion (Mitchell 1993, 2001, 2008, 2014, 2021, Chayes, Chayes and Mitchell 1998), effectiveness of the rules includes both its problem-solving effectiveness and the influence on states' compliance. For effectiveness in this paper, we refer solely to the *effectiveness of a state's system (or IO standards) in addressing a specific issue*. In the case of FATF, we may define the effectiveness as follows:

The extent that a state's system including institutions, measures, policies (or FATF's Recommendations) reduce(s) the amount of money laundering.

Effectiveness is hard to measure directly. Econometrically, it is the treatment effect on the outcome (variable) of concern, requiring counterfactual inquiries. Here, we conceptualize the outcome as a quantitative status (e.g., the amount of money laundering), distinct from effectiveness—the quantitative change (e.g., the reduction in money laundering achieved by a state's system). In the other contexts (such as FATF's Immediate Outcomes) outside our framework, an outcome typically denotes a specific goal (e.g., the amount of money laundering is reduced by 90% by a state's system). For clarity, outcome(s) in

what follows refers to the quantitative status, unless otherwise specified in the parentheses.

Additionally, the problem context may exacerbate difficulties, for instance, in measuring the outcome(s) of concern. In the money-laundering scenario, it has been challenging to quantify the amount of illicit money due to the shadowy nature. Thus, IOs need *indirect* proxies (we use "proxies" or "indicators" interchangeably) of effectiveness. The compliance degree to the standards is an important form of indirect indicators, as the higher compliance is, the greater effectiveness is expected. Yet, a key assumption underpinning the suitability of compliance as an effectiveness proxy is the standards are effective in theory. Following our previous definition of effectiveness, effective standards in theory are as follows:

A series of proposed actions that can indirectly or directly resolve the problem in a logical sense.

Our definition is consistent with Mitchel's (1993) argument that the standards are not valuable unless sufficient compliance to them can lead to desired effectiveness. Adding to Mitchell's discussion, we argue that even if the standards are highly effective in theory so that compliance generates substantial effectiveness *on average*, there still could be *regional heterogeneity* in how much compliance is translated into effectiveness. Regional incapacities (administrative, financial, and technological) that are dominant sources of non-compliance, are also sources for low conversion efficiency from compliance to effectiveness.

How to design effective standards? First, *direct* measures are the most effective. In the case of money laundering, policies that enhance financial transparency raise the difficulty in money laundering, thus curbing it directly. Identifying the *chain of events* is conducive to finding *indirect* measures. For instance, corruption is a typical predicate crime (upstream event) to money laundering. States with sound policies that control corruption and ensure public transparency are expected to spot rare corruption, reducing the money laundered from corruption indirectly. Similar measures may include policies that prevent human trafficking and illegal drug trade. Regarding downstream crimes, we may consider terrorist financing. Terrorists often rely on cross-border transfers of funds, where money laundering processes such as layering and integration are inevitable (FATF 2008; Al-Suwaidi and Nobanee 2021; Shah 2025). Measures such as strengthening the national security system can combat terrorist activities, impeding related money laundering crimes indirectly.

Compliance to well-designed standards makes effectiveness (of the whole state's system) promising, yet inferring effectiveness solely from compliance can bias the evaluation. These two different concepts will be tautological in practice if compliance is the only proxy. IOs need other indirect proxies for effectiveness, for instance, a series of *stage outcomes* relevant to the measures. Public perception about the corruption is an outcome related to the anti-corruption policies (upstream measures), while the number of terrorist attacks is associated with the anti-terrorist financing tasks (downstream measures). Monitoring the stage outcomes informs how successfully effectiveness is transmitted via the chain of events. Yet, the accuracy of effectiveness proxies varies depending on their "logic distance" to the event of concern, with closer proximity yielding greater accuracy. We summarize these properties in Table 1 below.

Constructively, IOs should improve member states' compliance and effectiveness through a *strategy system*. Mitchell (2008) suggests a selected combination of punitive, remunerative, preclusive, generative, cognitive, or normative strategies. Before understanding each strategy, we revisit four behavioral types that operate from two different behavioral logics. An *interests-maximizer* follows a logic of *consequences*, weighing costs against benefits of compliance. A *positive* interests-maximizer believes benefits outweigh costs, while a *negative* interests-maximizer believes the opposite. On the other hand, a state may behave by a logic of *appropriateness*. A *complier* believes it right to comply, while a *defier* believes it wrong. While positive interests-maximizers and compliers are expected to have good

TABLE 1 | Effectiveness proxy types.

Effectiveness proxy type	Description	Logic distance to the event of concern	Example: AML context
Outcome(s) of Concern	The amount of the event of concern	0	The amount of money laundering activities
Stage Outcome(s)	The outcomes relevant to upstream or downstream measures that indirectly solve the problem	1	Public perception of corruption, the amount of terrorist financing
Compliance to Direct Measures	The compliance degree to the measures that directly solve the problem	1	Compliance to financial transparency practices
Compliance to Indirect Measures	The compliance degree to the measures that indirectly solve the problem	2	Compliance to anti-corruption/ anti-terrorism policies

Note: the *outcome(s) of concern* that quantifies the event of concern directly is the first-best effectiveness proxy. This is because econometrically, effectiveness is the treatment effect on the outcome(s) of concern from the influence of the state's system. Other effectiveness proxies differ in their "logic distance" to the event of concern. The *stage outcome(s)* and *compliance to direct measures* affect the outcome(s) of concern directly—distance 1. The *compliance to indirect measures* affects stage outcome(s), which then affects the outcome(s) of concern—distance 2. Proxies with closer logic proximity to the event of concern are more accurate and deserve higher weights when inferring the ultimate effectiveness.

performance because of self-motivation, their counterparts are likely to underperform.

Behavior-driven non-compliance should be addressed contingent to the behavioral logic. For negative interests-maximizers, strategies that manipulate costs and benefits of behaviors are in effect. Ex post strategies, either *punitive* strategies (impose punishment after non-compliance), or *remunerative* strategies (offer rewards after compliance), change the cost–benefit structure after the action, while ex ante strategies, either *preclusive* (increase the difficulty in non-compliant behaviors) or *generative* (make compliance easier with resources in advance) make the change before the action. For defiers, both *cognitive* (instill more knowledge of choices) or *normative* (change values and norms) strategies that operate through behavioral appropriateness are suitable to establish positive attitudes towards compliance.

In expectation, positive interests-maximizers and compliers have satisfactory performance, yet they might still struggle with low compliance or effectiveness. Their poor performance is normally *incapacity-driven*. Insufficient administrative, financial, or technological resources hinder good-faith states from achieving compliance or translating it into effectiveness. The best strategies to help are *generative* strategies that equip states with necessary resources in advance. *Remunerative* strategies to reward performance progress are also conducive to building a virtuous cycle of sustained improvement. Table 2 summarizes the applicability of different strategies contingent on behavioral types.

3.4 | Legitimacy: A Quantitative, Objective Inference Scheme

Following a comprehensive discussion on compliance, effectiveness, and the strategy system, we conclude this section with a quantitative, objective inference scheme for legitimacy, centering the three critical components of IO *performance*—standards, evaluation, strategy. A *legitimate* IO should possess the following performance-related characteristics,

1. Design achievable, cost-efficient, and universally applicable standards by complying to which the member states can achieve a reasonable level of effectiveness in expectation, regardless of states' capacity levels.
2. Deploy an impartial and precise evaluation system where both compliance and effectiveness are accurately measured and timely updated to the greatest extent. Self-identified effectiveness proxies should have sufficient correlations with similar proxies from external credible sources to ensure robustness.
3. Adopt a diversified strategy system that boosts long-term compliance and effectiveness at an affordable cost, within an efficient time range. Different strategies are adequately applied to member states with different behavioral logics or incapacities.

The legitimacy of IOs can by *partly* inferred from the extent to which the criteria are satisfied. First regarding the *standards*, the proposed standards should enable good-faith states to achieve both high compliance and effectiveness, which are quantifiable elements. Ideally, this result has be robust to incapacities so that states with any development levels could benefit from adopting the standards. Different from illy-made choices, incapacities, especially in the financial and technological forms, are vulnerabilities what developing states should not be blamed for.

Second regarding the *evaluation system*, impartiality and precision is the primary baseline to follow. Biases that come from political manipulation, persistent measurement errors, and incomplete effectiveness proxies critically undermine accurate assessment of the performance. Cross-validation with external credible indicators is conductive to verifying the robustness of evaluation results. Quantitatively, high correlations between IO-identified and external proxies justify the credibility of both.

Third regarding the *strategy system*, IOs should employ a mix of strategies contingent on the causes behind poor performance. Behavior-driven and incapacity-driven underperformance necessitates distinct strategies. Punitive strategies should be exercised with extra caution because it easily discourages good-faith states with incapacities. The efficacy of strategies to promote compliance can be quantitatively assessed using econometric tools such as differences-in-difference and event study models.

A flow chart is presented to summarize our theoretical framework built on Mitchell's compliance theory and the IO literature (Figure 2). In the following sections, we evaluate the performance-based legitimacy of FATF against these criteria

TABLE 2 | Behavioral types, performance, and strategies.

Behavioral type	Behavioral logic	Compliance likelihood	Likely poor performance driver	Recommended strategies
Positive Interests-Maximizer	Logic of consequences (benefits > costs)	High	Incapacities	Generative, Remunerative
Negative Interests-Maximizer	Logic of consequences (costs > benefits)	Low	Behavioral (cost–benefit calculation)	Punitive, Remunerative, Preclusive, Generative
Complier	Logic of appropriateness (right to comply)	High	Incapacities	Generative, Remunerative
Defier	Logic of appropriateness (wrong to comply)	Low	Behavioral (appropriateness)	Cognitive, Normative

Note: the table summarizes different behavioral types, logics, and recommended strategies for poor performance.

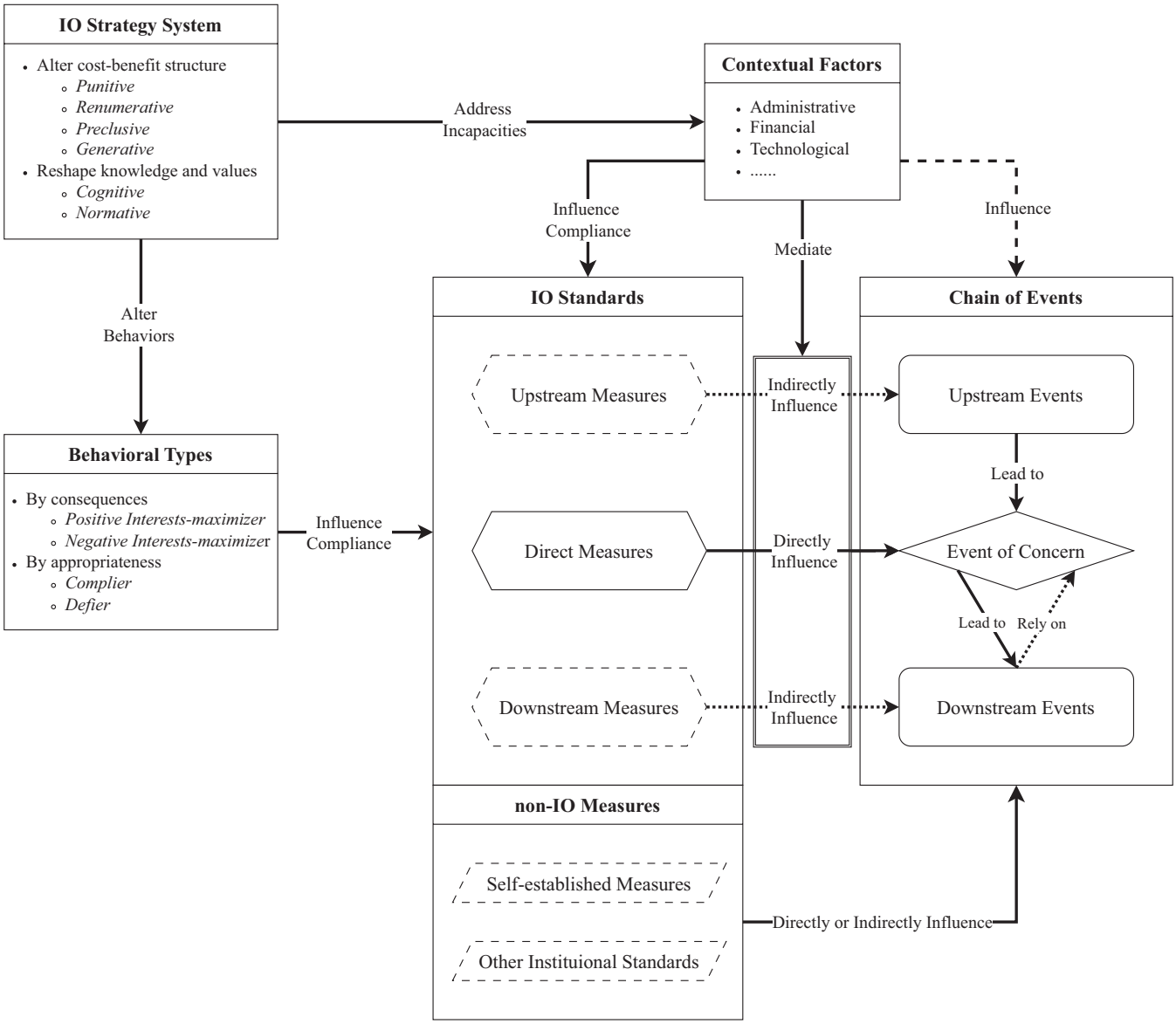

FIGURE 2 | Theoretical flow chart. *Note:* the flow chart synthesizes how an IO can properly exert its power. "Event of concern" is the event that the IO aims to influence (e.g., (reduce) money laundering, (encourage) waste recycling), centering in the "chain of events" with other "upstream events" leading to it or, and "downstream events" following or relying on it. "IO Standards" should consist of direct measures that influence the event of concern directly, and (or) indirect measures that influence it indirectly via the channel of upstream and downstream events. There might be other legitimate measures not covered by the IO, either state self-established or outlined by other institutions. "Contextual factors" play an important role in determining the compliance to the IO standards and mediate its relationship with standards effectiveness. "Behavioral types" also influence compliance. IO should employ its "strategy system" to enhance states' performance (if unsatisfactory), either driven by behaviors or incapacities. The IO evaluation system, which is not shown in the chart, should ensure impartial and timely evaluation on compliance to the IO standards and effectiveness of *the whole state's system* (constituted of both the IO standards and other measures).

with empirical evidence. Policy recommendations are given in the end.

4 | Data

The primary dataset is the ratings from FATF's fourth-round mutual evaluations. The ratings consistent of two components—40 Technical Recommendations and 11 Immediate Outcomes (Goals). Starting from January 2014 until July 2025, 194 regions have completed their initial mutual evaluation, among which 125 jurisdictions have subsequent reports after the initial assessment. Regarding the follow-up process, only 28 regions (14.43%) entered the default regular follow-up after the initial evaluation, which reflects the stringent standard of FATF. Among the 166 jurisdictions that entered the enhanced follow-up, 137 (82.53%) regions violated both technical compliance and effectiveness criteria, while 10 (6.02%) only violated the technical compliance requirements and 19 (11.45%) only violated the effectiveness requirements. Table 3 summarizes key follow-up statistics by continent.

In addition to the fourth-round ratings, we incorporate the third-round data for necessary robustness checks. Although

data from different rounds are not directly comparable due to evolving standards, they can be aligned using the mapping from the FATF Recommendations Manual (2012 Edition). Most new standards have exact correspondence with the previous ones, which affirms the approach's validity. For very few recommendations (R.1 *Assessing risks & applying a risk-based approach*, and R.7 *Targeted financial sanctions related to proliferation*), we approximate them by the average ratings of all recommendations and average ratings of special recommendations on terrorist financing in the third-round data, respectively.

To quantitatively analyze these categorical ratings, we follow a similar approach as the previous researchers (Johnson 2008; Verdugo Yepes 2011; Mekpor et al. 2018). For compliance ratings, we encode "No Compliance" as 0, "Partial Compliance" as $\frac{1}{3}$, "Large Compliance" as $\frac{2}{3}$, and "Compliance" as 1 for each individual recommendation. For recommendations with "Not Applicable" ratings, we fill them with the average rating of other recommendations. With these numeric ratings, we develop a set of compliance indices. For the overall technical compliance index (TCI) that we present in the main text, it is an Individual-Average Technical Compliance Index (I-TCI) that averages all individual recommendations ratings ranging from 0 (no compliance) to 1 (full compliance). Secondly, we independently consider seven sectional compliance indices by averaging ratings in each recommendation section. The sectional indices deliver a multi-dimensional picture of the performance. Notation-wise, we use "TCI – sectional theme" as the succinct and informative format. For robustness purpose, we also develop a Section-Average Technical Compliance Index (S-TCI) that averages the compliance ratings by sectional component. The results of S-TCI are placed in the appendix (Figures A1–A3; Tables A1–A5) as robustness checks. Table 4 reports different types of TCIs by continent.

For the effectiveness ratings, we apply the same logic to each individual immediate outcome (goal). We encode "Low Effectiveness" as 0, "Moderate Effectiveness" as $\frac{1}{3}$, "Substantial Effectiveness" as $\frac{2}{3}$, and "High Effectiveness" as 1, and take the average of the 11 individual ratings to get the Individual-Average Effectiveness Index (I-EI), or average the three sectional means to get the Section-Average Effectiveness Index (S-EI). For the sectional effectiveness indices (in terms of intermediate outcomes (goals)), we denote them as "EI—sectional theme". Notably, the effectiveness ratings and indices here should be interpreted as FATF-identified proxies of effectiveness, not the actual effectiveness of the state's system, which is the unobserved treatment effects on the outcome of concern. Strictly speaking, the effectiveness ratings are limited inferences since the effectiveness of a specific state's system cannot be observed directly, and is almost infeasible to identify under causal frameworks. Table 5 reports different types of EIs by continent.

TABLE 3 | Follow-up statistics by continent.

Counts	Global	Europe	Asia	North America	Oceania	Africa	South America
States with MER available	194	49	43	30	12	48	12
States in Regular Follow-up	28	15	8	4	1	0	0
States in Enhanced Follow-up	166	34	35	26	11	48	12
Pass Rate	14.43%	30.61%	18.60%	13.33%	8.33%	0.00%	0.00%

Note: the table summarizes the FATF fourth-round follow-up statistics by continent. The first row shows the number of states that completed the initial evaluation (with mutual evaluation reports (MER) available). The second row shows the number of states that entered the default regular follow-up, and the third row shows that of enhanced follow-up. The *pass rate* is the proportion of states that passed the regular follow-up criteria.

TABLE 4 | Technical Compliance Index by continent.

Technical compliance index (TCI)	Global	Europe	Asia	North America	Oceania	Africa	South America
A: Policy and Coordination	0.56	0.64	0.60	0.61	0.49	0.39	0.71
B: ML and Confiscation	0.71	0.78	0.69	0.77	0.65	0.61	0.75
C: Terrorist Financing	0.43	0.53	0.47	0.50	0.43	0.25	0.43
D: Preventive Measures	0.62	0.67	0.67	0.68	0.55	0.50	0.65
E: Transparency and Beneficiary	0.39	0.52	0.40	0.42	0.33	0.25	0.43
F: Authorities and Others	0.62	0.67	0.65	0.65	0.56	0.52	0.71
G: International Cooperation	0.65	0.73	0.64	0.67	0.57	0.58	0.69
Individual Average (I-TCI)	0.60	0.66	0.63	0.64	0.53	0.48	0.65
Sectional Average (S-TCI)	0.57	0.65	0.59	0.62	0.51	0.44	0.62

Note: the table summarizes the different variations of technical compliance indices (sectional, individual-average, section-average) from the FATF fourth-round initial evaluations by continent.

TABLE 5 | Effectiveness Index by continent.

Effectiveness index (EI)	Global	Europe	Asia	North America	Oceania	Africa	South America
A: Risk Mitigation	0.40	0.57	0.47	0.41	0.43	0.11	0.46
B: Crime Proceeds Prevention	0.22	0.34	0.26	0.24	0.21	0.04	0.25
C: Threat Detection	0.26	0.41	0.32	0.28	0.20	0.07	0.28
Individual Average (I-EI)	0.28	0.42	0.33	0.29	0.25	0.07	0.30
Sectional Average (S-EI)	0.29	0.44	0.35	0.31	0.28	0.07	0.33

Note: the table summarizes the different variations of effectiveness indices (sectional, individual-average, section-average) from the FATF fourth-round initial evaluations by continent.

For other covariates, we obtain publicly available data from WB, UNDP, IMF, Basel Institute on Governance, and other credible sources. Specific data cover governance, education, technology, human development, financial development, foreign direct investment, Basel AML Index, and so forth.

5 | Research Questions, Hypotheses and Theoretical Foundations

5.1 | Determinants of Compliance

By examining compliance determinants, we seek to address gaps in existing research through a theory-guided approach. This analysis evaluates whether regional incapacities hinder compliance (in other words, whether FATF's standards disproportionately challenge developing states), and whether FATF's interventions enhance compliance. We refer to the empirical literature, FATF's methodology, and Mitchell's arguments (1993) to form a set of potential determinants. The FATF methodology manual (2023) designates relevant regional features as *context* or *contextual factors*. One particular category FATF emphasizes is *structural* factors, termed *administrative* factors in our theoretical framework, including institutional accountability, transparency, political stability, and other governance aspects. Administrative incapacities do not justify low compliance with FATF's standards (FATF 2023), as governance quality is primarily determined by governmental policy *choices*. Yet, we would like to draw particular attention to the other factors that Mitchell argues (1993)—*financial* and *technological* incapacities, which are *inevitable dilemmas* during early stages of development. Financial factors are examined in depth due to their critical relevance to the monetary nature of AML. We also include other socioeconomic factors such as education and foreign direct investment, following the existing literature.

Beyond identifying regional factors influencing compliance, we examine the correlation between compliance and evaluation stages. Specifically, we investigate whether the fourth-round follow-up compliance ratings significantly differ from prior ratings, *conditional on regional factors*. Ratings from the final third-round assessment and initial fourth-round evaluation serve as the baseline. Following the fourth-round initial evaluation, jurisdictions enter either regular or enhanced follow-up. Consequently, follow-up report ratings reflect jurisdictions' compliance achievement under FATF's supervision during the fourth round. Two hypotheses are formed:

Hypothesis 1. (H1): *There are administrative, technological, financial and socioeconomic factors that are significantly correlated with the compliance degree.*

Hypothesis 2. (H2): *Holding the regional covariates fixed, fourth-round follow-up compliance is still significantly higher than that before the follow-up.*

The above hypotheses only argue the correlation between technical compliance and independent variables. To rigorously investigate the causal effects from FATF's governance, we need to examine the evaluation process, testing for the presence of *FATF-induced compliance*. If there is, one further inquiry centers whether the policy effects fade, persist, or strengthen over time. Thus, we form a causal hypothesis:

Hypothesis 3. (H3): *There is considerable and persistent compliance induced by certain FATF's interventions.*

5.2 | Linking Compliance and Effectiveness

Only effective standards justify the compliance to them. Standards should be designed to ensure compliance correlates strongly with effectiveness, making greater compliance preferable to lesser (Mitchell 1993). Assessing effectiveness reveals both individual state's performance and the validity of FATF's Recommendations. As compliance seldom fully translates into effectiveness, monitoring effectiveness performance is essential. First, due to the mediation from regional factors, the effectiveness might not be perfect even if fully compliant. Secondly, an excessive focus on compliance may understate international soft law's effectiveness (Meyer 2014). Hence, we form the following hypotheses:

Hypothesis 4. (H4): *FATF's effectiveness proxies are sufficiently correlated with the compliance level, supporting the validity of FATF's Recommendations.*

Hypothesis 5. (H5): *Compliance does not fully translate into effectiveness, as proxied by FATF's indicators, with technical compliance typically yielding lower effectiveness on average.*

Effectiveness may also vary across states even with the same compliance. In our extended framework of Mitchell's theory, we argue that administrative, financial, and technological factors play an important role in governing the relationship

between compliance and effectiveness. A natural hypothesis is as follows:

Hypothesis 6. (H6): *There are administrative, technological, and financial factors that mediate the relationship between FATF's effectiveness proxy and compliance, which indicates jurisdiction-level heterogeneity in this relationship.*

Ideally, the FATF's effectiveness proxies serve as reliable inferences for the actual effectiveness, given the fundamentally different assessment approach from technical compliance. Yet, concerns of validity exist. Measurements from the same organization have inevitable correlations with each other (Kaufmann and Kraay 2024). When designing the standards, FATF has to make sure the proposed recommendations are effective, which implies a *natural and harmless* correlation between standards compliance and actual effectiveness. Yet, within the same institution, information that is used for evaluation share a lot in common, which *harmfully inflates* the correlation between measurements. Furthermore, the primary basis for assessing effectiveness is the compliance with FATF's Recommendations, questioning the validity of FATF's effectiveness proxies. For instance, *Immediate Outcome (Goal) 1* benefits primarily from the implementation of *Recommendation 1, 2, 33, 34*, and some elements of *Recommendation 15* (FATF 2023). Consequently, FATF's effectiveness proxies may be biased upwards (downwards) for states with high (low) compliance, and a strong correlation between compliance and these proxies does not necessarily validate FATF's Recommendations.

To tackle this critical issue, we have to develop or adopt effectiveness proxies from credible, independent external sources unaffiliated with FATF. A key limitation of alternative effectiveness indicators, similar to FATF's, is their inability to precisely capture actual effectiveness. Different indicators inform us about effectiveness from different perspectives, yet none provides an accurate measurement, particularly in AML, where much data remains covert. Yet, a high correlation between FATF's and external effectiveness proxies justifies the robustness of these measurements. We propose the following hypothesis:

Hypothesis 7. (H7): *Alternative effectiveness proxies from external credible sources correlate sufficiently with FATF's proxies, and Hypothesis 4–6 still hold.*

6 | Research Design

6.1 | Hypothesis 1 and 2

In this section, we articulate the empirical strategies with rigorous mathematical expressions. To study the first question about compliance determinants, we construct a set of multi-regressions with administrative, financial, technological, and socioeconomic factors. We carry out the regressions by adding the four categories one by one. For *administrative* factors, we proxy them by the worldwide governance indicators compiled by World Bank from diverse credible data sources. Although these indicators measure governance quality from six different perspectives, correlations among indicators are all above 0.65. Including all of them in the regression might be troublesome to reveal the relations. Hence, we choose two representative governance indicators, voice and accountability (measuring institutional transparency and accountability) and regulatory quality (measuring governmental regulatory quality in the private sector).

The *technological* factor is proxied by the population proportion that have access to the internet. The *financial* factors are captured by the multi-layered financial development indices from IMF. For *socioeconomic* factors, we consider the foreign direct investment as a proportion out of GDP, and proxy education level by the government education expenditure proportion out of GDP. A dummy variable, FU, is defined to indicate whether a jurisdiction is in the follow-up process (FU = 1) or prior to it (FU = 0). The general model specification to study Hypothesis 1 and 2 is as follows:

$$\mathrm{TCI}_{it} = \alpha + \tau \mathrm{FU}_{it} + \beta^T X_{it} + \epsilon_{it}$$

where TCI_{it} is the technical compliance index, α is the intercept, FU_{it} is the follow-up indicator, X_{it} is a set of regional covariates. i represents jurisdiction, t represents year. τ captures the average difference between the compliance during the fourth-round follow-up and before.

6.2 | Hypothesis 3

To address Hypothesis 3 on FATF-induced compliance, technically we employ an event study model with two-way fixed effects. Policy-wise, we focus on the treatment of enhanced follow-up versus the regular follow-up, with the former featuring higher supervision intensity. We define EFU as a dummy variable to indicate whether a state receives the treatment of enhanced follow-up (EFU = 1) or regular follow-up (EFU = 0).

Regarding the dynamics of treatment effects, we consider the timeline as the stages of evaluations, defined by the ordinal sequence of follow-up reports. Specifically, the report takes the form "jth regular (or enhanced) follow-up report" where j indicates the time order of publication. We define a variable STAGE to indicate the stage of the follow-up process. STAGE = 0 for ratings on the fourth-round initial evaluation (baseline stage). STAGE = −1 for ratings on the third-round final assessment. STAGE = j for ratings on the jth follow-up report where $j \leq 4$. For the fifth or subsequent follow-ups, fourth-round final assessment, fifth-round initial evaluation, we assign 5 to STAGE. The event study specification is as follows:

$$\mathrm{TCI}_{it} = \alpha_i + \gamma_t + \sum_{j=-1,1,2,3,4,5} \tau_j \mathrm{EFU}_i \times \mathbf{1}\left[\mathrm{STAGE}_{it} = j\right] + \beta^T X_{it} + \epsilon_{it}$$

where TCI_{it} is the technical compliance index, α_i is the jurisdiction fixed effects, γ_t is the year fixed effects, EFU_i is the enhanced follow-up indicator, $\mathbf{1}\left[\mathrm{STAGE}_{it} = j\right]$ are stage indicators, X_{it} is a vector of relevant regional covariates (optional). i

TABLE 6 | Determinants of compliance (I-TCI).

	OLS				
	(1)	(2)	(3)	(4)	(5)
	Admin	**Admin, tech, socioecon**	**Admin, tech, socioecon and fin (1)**	**Admin, tech, socioecon and fin (2)**	**Admin, tech, socioecon and fin (3)**
In 4th-round Follow-up	0.183***	0.135***	0.137***	0.139***	0.137***
	(0.0135)	(0.0138)	(0.0140)	(0.0141)	(0.0143)
Voice and Accountability	−0.0530	−0.0380	−0.0267	−0.0571	−0.0483
	(0.0645)	(0.0623)	(0.0636)	(0.0693)	(0.0699)
Regulatory Quality	0.523***	0.119*	0.0434	0.0466	0.0865
	(0.0604)	(0.0690)	(0.0831)	(0.0820)	(0.0864)
Access to Internet (% of Population)		0.335***	0.322***	0.304***	0.295***
		(0.0375)	(0.0378)	(0.0398)	(0.0409)
Foreign Direct Investment (% of GDP)		0.0204*	0.0212*	0.0216*	0.0248**
		(0.0115)	(0.0118)	(0.0126)	(0.0125)
Government Education Expenditure (% of GDP)		−0.677	−0.670	−0.636	−0.266
		(0.517)	(0.514)	(0.520)	(0.550)
Financial Development			0.0778*		
			(0.0437)		
Financial Institutions Development				0.127**	
				(0.0635)	
Financial Markets Development				−0.00148	
				(0.0378)	
Financial Institutions Access					0.0735**
					(0.0318)
Financial Institutions Depth					−0.0500
					(0.0529)
Financial Institutions Efficiency					0.145*
					(0.0741)
Financial Markets Access					0.00702
					(0.0359)
Financial Markets Depth					0.0511
					(0.0531)
Financial Markets Efficiency					−0.0163
					(0.0281)
Constant	0.228***	0.302***	0.315***	0.307***	0.219***
	(0.0284)	(0.0267)	(0.0276)	(0.0290)	(0.0509)
No. of Observations	364	364	364	364	364
R^2	0.445	0.549	0.552	0.555	0.562

Note: the dependent variable is Individual-Average Technical Compliance Index. Admin, Tech, Socioecon and Fin represents administrative, technological, socioeconomic, and financial factors. The financial factors are examined from the overall level to more nuanced levels by a top-down approach. Robust standard errors are used. *0.1, **0.05, ***0.01.

represents jurisdiction, j represents stage, t represents year. τ_j captures the dynamic treatment effects of enhanced follow-up over stages.

6.3 | Hypothesis 4 and 5

Regarding Hypothesis 4 and 5 about the conversion between compliance and effectiveness, it is suitable to use a zero-intercept regression across regions:

$$\mathrm{EI}_{it} = \rho \mathrm{TCI}_{it} + \epsilon_{it}$$

where EI_{it} is the effectiveness index, TCI_{it} is the technical compliance index. i represents jurisdiction, t represents year. Here ρ is the parameter of interest, interpreted as the average proportion of technical compliance converted into effectiveness. We name ρ as the conversion coefficient. We assume zero-intercept in the specification—if completely violating all FATF's standards, on average, the state's AML system achieves null effectiveness proxied by FATF's indicators. The analysis is essentially cross-sectional if considering the evaluation stages as the timeline, as FATF's effectiveness ratings are provided only during the initial evaluation while not updated in the follow-ups.

6.4 | Hypothesis 6

To study the mediation mechanism in the compliance-effectiveness nexus, we may consider candidate variables similar to those in the compliance determinants analysis. Yet, to concisely illustrate, we seek a single comprehensive mediator, Human Development Index (HDI), to elucidate the overarching mechanism first. HDI, though developed by the UNDP to assess *living conditions*, it strongly correlates (see in Figure A4) with administrative regulatory quality (0.81), technological advancement (0.87), financial development (0.77). Hence, HDI is a good proxy for overall development that synthesizes relevant capacities. After examining the overall mediation, we dive into more nuanced dimensions of the administrative, technological, and financial factors. Methodologically, we still use zero-intercept regressions but incorporate interaction terms to address the Hypothesis 6:

$$\mathrm{EI}_{it} = \rho^T X_{it} \times \mathrm{TCI}_{it} + \epsilon_{it}$$

where EI_{it} is the effectiveness index, TCI_{it} is the technical compliance index, X_{it} is a vector of mediator(s). Now the conversion rate is determined by both ρ and the vector X_{it} where X_{it} mediates the conversion from compliance to effectiveness.

TABLE 7 | Estimated treatment effects from enhanced follow-up on compliance (I-TCI).

	DID			Event study
	(1)	(2)	(3)	(4)
$EFU \times POST$	0.147***	0.178***	0.0413***	
	(0.0115)	(0.00958)	(0.0104)	
$EFU \times 1[STAGE = -1]$				0.00252
				(0.0192)
$EFU \times 1[STAGE = 1]$				0.0141
				(0.00906)
$EFU \times 1[STAGE = 2]$				0.0552***
				(0.0130)
$EFU \times 1[STAGE = 3]$				0.0611***
				(0.0173)
$EFU \times 1[STAGE = 4]$				0.0581***
				(0.0174)
$EFU \times 1[STAGE = 5]$				0.0777***
				(0.0214)
Constant	0.520***	0.507***	0.563***	0.558***
	(0.0114)	(0.00389)	(0.00421)	(0.00649)
Jurisdiction FE	N	Y	Y	Y
Year FE	N	N	Y	Y
No. of observations	664	664	664	664

Note: the dependent variable is Individual-Average Technical Compliance Index. *EFU* is the treatment indicator for enhanced follow-up ($EFU = 1$) versus regular follow-up ($EFU = 0$). *POST* is the simple stage indicator where 1 means in the fourth-round follow-up and 0 means otherwise. $1[STAGE = j], j = -1,1,2,3,4,5$ are nuanced stage indicators where 1 means in the stage j and 0 means otherwise. The coefficients of $EFU \times POST$ are the simple DID estimates, while the coefficients of $EFU \times 1[STAGE = j], j = -1,1,2,3,4,5$ are the event study estimates. Standard errors are clustered at the jurisdictions level. ***0.01.

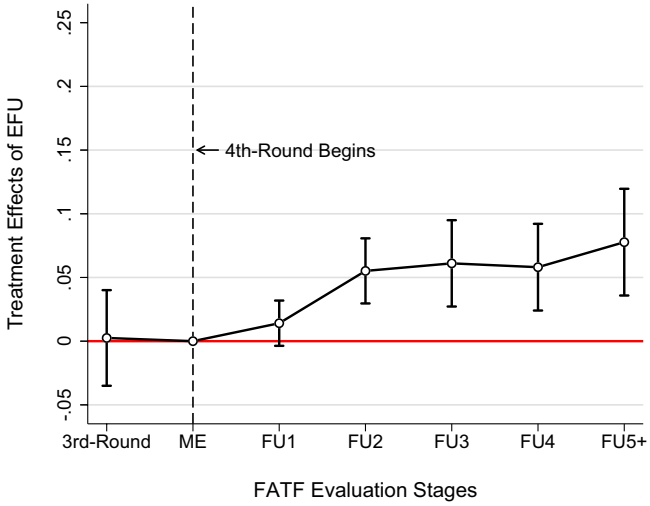

FIGURE 3 | Time trend of treatment effects from enhanced follow-up on compliance (I-TCI). *Note:* the figure shows the time trend of treatment effects from the enhanced follow-up on the Individual-Average Technical Compliance Index, compared with the regular follow-up. The *y*-axis shows the treatment effects during different stages. The *x*-axis shows the timeline of evaluation as stages: "3rd-Round" means the third-round final assessment, "ME" means the fourth-round initial mutual evaluation (the baseline period), "FU1" means the first follow-up (the same logic applies for "FU2," "FU3," "FU4"), "FU5+" means the fifth or later follow-ups.

6.5 | Hypothesis 7

To test Hypothesis 7 about the validity of FATF's effectiveness proxies, we build upon an external indicator called Basel AML Index measuring the ML risks. The index uses a composite methodology, aggregating data from several publicly available sources, such as FATF, Transparency International, and Global Initiative against Transnational Organized Crime. It evaluates risks across five domains with different weights: *Quality of AML/CFT Systems* (50%), *Corruption and Fraud Risks* (17.5%), *Financial Transparency* (17.5%), *Public Transparency* (5%), *Legal and Political Risks* (10%). Since the composite Basel AML Index measures the risk not effectiveness, and incorporates a significant proportion of FATF's reports (35% in the domain of *Quality of AML/CFT framework*), we invert the index, exclude the FATF component and renormalize to obtain a Basel non-FATF Effectiveness Index, ranging from 0 (null effectiveness) to 1 (full effectiveness). We obtain the historical data from 2012 to 2024 covering over 150 countries. Owing to good yearly coverage, we employ panel methods with two-way fixed effects, enhancing the rigor in analysis. The mathematical expressions are similar to those for Hypothesis 4–6 with additional jurisdiction and year fixed effects.

7 | Empirical Findings and Discussion

7.1 | Incapacities: Key Obstacles to Compliance

Capacities matter for compliance. Overall, we identify regulatory quality (administrative), access to Internet (technological), financial institutions development (financial), and foreign direct investment (socioeconomic) as key factors positively correlated with the compliance at the 5% significance level. Notably, the significance level of regulatory quality drops when the technological and financial factors appear together, suggesting the latter two types are more important in determining the compliance. For the financial factors, the development of financial institutions plays a more dominant role than financial markets, especially in the aspects of access and efficiency. Foreign direct investment also shows positive association with compliance, since highly compliant states probably have better financial security, attracting more foreign investment. Detailed regression results are reported in Table 6.

These findings prove Hypothesis 1 and provide evidence for Mitchell's arguments (1993) about incapacities and non-compliance. In the case of FATF, all three types of factors (administrative, technological, financial) are critical to determining the compliance. Strong capacities in these areas appear essential for compliance with FATF's Recommendations, suggesting that the standards are not universally feasible to all states. For developing regions, though administrative deficiencies cannot be a valid excuse for non-compliance, they can hardly embrace high technological or financial benefits. Low-compliant states may act in good faith, yet their efforts towards high compliance could be impeded by relevant capacity constraints.

7.2 | FATF-Induced Compliance

Identifying FATF-induced compliance is crucial for assessing the efficacy of its interventions. The prior section's results show that fourth-round follow-up ratings reflect at least 13.5% higher compliance than earlier ratings, controlling for covariates, thus confirming Hypothesis 2. This suggests that FATF acts to enhance compliance, yet formal causal analysis is required to identify treatment effects.

The simple DID estimate with two-way fixed effects suggests that enhanced follow-up boosts 4.13% more compliance than regular follow-up. Event study estimates reveal nuanced dynamics, with the effects on compliance generally strengthening throughout the follow-up process. In particular, the magnitude increases from 1.41% in the first follow-up to 7.77% in the fifth or further follow-ups. Similar patterns are observed among most sectional compliance indices, except for Section F and G. The evidence confirms Hypothesis 3, demonstrating that FATF's enhanced follow-up significantly increases overall compliance with persistent and escalating effects, despite volatile patterns in some sections. Detailed estimates are reported in Table 7.

One caveat of the DID or event study models is the reliance on the parallel trend assumption. The time trend of the overall compliance shows that the estimate one period before the baseline is not significantly different from zero at the 5% significance level, indicating no evidence of a pre-trend (Figure 3). However, not all sectional compliance indices meet the parallel trend assumption, necessitating cautious interpretation of the results (Figure 4).

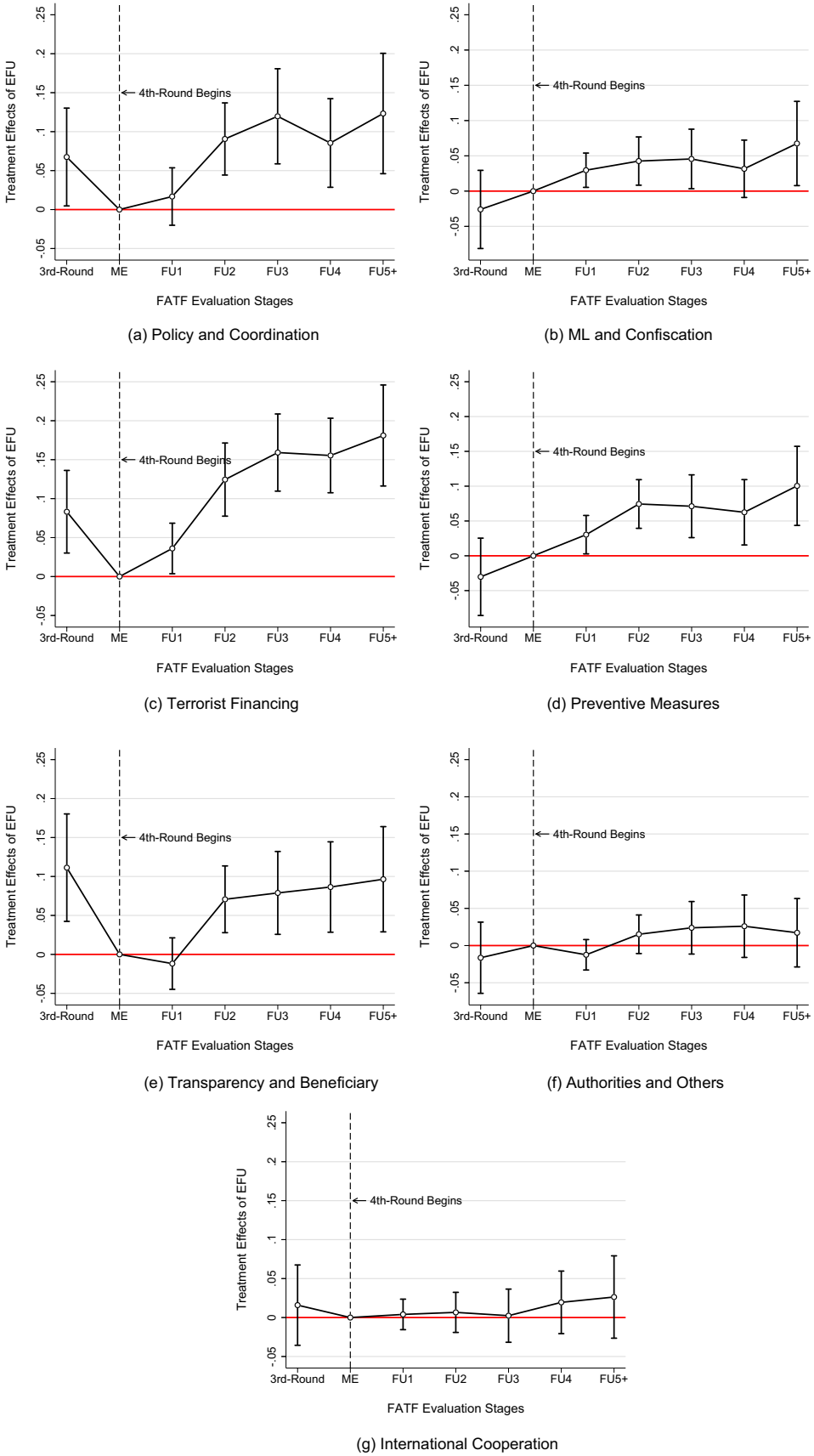

FIGURE 4 | Time trend of treatment effects from enhanced follow-up on compliance (sectional TCIs). *Note:* the figure shows the results of similar analysis (as in Figure 3) for sectional technical compliance indices. The seven sections include (a) Policy and Coordination, (b) ML and Confiscation, (c) Terrorist Financing, (d) Preventive Measures, (e) Transparency and Beneficiary, (f) Authorities and Others, (g) International Cooperation.

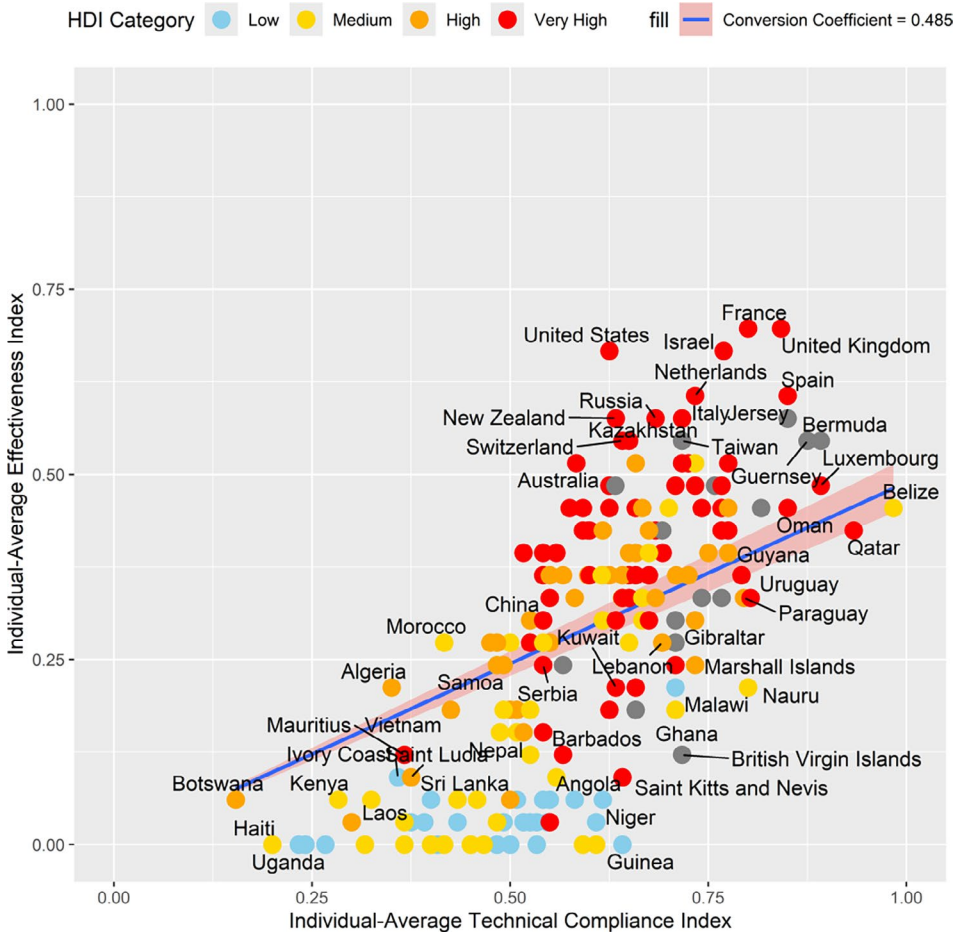

FIGURE 5 | Scatter Plot of Effectiveness Index (I-EI) and Technical Compliance Index (I-TCI). *Note:* the above scatter plot shows the relationship between Individual-Average Effectiveness Index and Technical Compliance Index. The fitted zero-intercept linear coefficient is 0.485. The regions are mapped with Human Development Index (HDI) categories—Low, Medium, High, and Very High. Gray dots are regions with missing HDI values.

7.3 | Compliance to Effectiveness: Conversion Efficiency Matters

Compliance is critical, as is its conversion efficiency into effectiveness. Fourth-round ratings reveal thatregions with high compliance typically achieve reasonable effectiveness (as shown in Figure 5). In a zero-intercept regression, the conversion coefficient ρ is 48.5% significant at the 1% level, indicating that jurisdictions on average convert 48.5% of compliance into effectiveness. This evidence first proves our Hypothesis 4 that FATF's effectiveness proxy is sufficiently correlated with compliance overall, justifying the validity in FATF's Recommendations in terms of FATF's effectiveness indicators. Secondly, Hypothesis 5 about incomplete conversion is proved since the conversion coefficient ρ is lower than 1 (perfect conversion benchmark).

Furthermore, a closer examination of the scatter plot informs that certain regions fall significantly below the regression line, while some exceed the average. This indicates considerable heterogeneity in the conversion efficiency ρ among regions. In the analysis of interaction between TCI and HDI, the coefficient is significant at the 1% level with a magnitude of 65.3%. The *R*-squared rises from 0.802 (without interaction) to 0.882 (with interaction), indicating a substantially improved model fit with the interaction setup. This evidence confirms Hypothesis 6 at the top level, demonstrating that overall development, proxied by HDI (strongly correlated with administrative, financial, technological capacities), significantly mediates the conversion from compliance to effectiveness.

To make the results more intuitive, we examine the mediation using the HDI category—Low, Medium, High, Very High. Findings indicate that the conversion coefficient improves significantly from the Low-HDI to the Very-High-HDI group. The Low-HDI group's estimate is 7.91%, while the Very-High-HDI group's estimate is 61.3%, approximately 7.75 times higher (Table 8).

After analyzing the overall mediation mechanism, we explore how specific regional factors influence the conversion ratio (Table 9). The administrative, technological, and financial factors identified as compliance determinants are examined as mediators again. It is evident that the interaction with the technological factor remains significant at the 5% level across all specifications, with the coefficient over 24%. The interaction with overall financial development is also significant, with a larger coefficient of 40.6%. Specifically, financial markets efficiency exhibits a 26.8% interaction magnitude, significant at the 1% level. In contrast, the interaction with regulatory quality is marginally significant with a smaller magnitude. These

TABLE 8 | From compliance (I-TCI) to effectiveness (I-EI)—HDI as a mediator.

	OLS		
	(1)	(2)	(3)
	No mediation	Mediation: HDI value	Mediation: HDI 4 categories
TCI	0.485***		
	(0.0181)		
$HDI \times TCI$		0.653***	
		(0.0178)	
$HDI_{LOW} \times TCI$			0.0791***
			(0.0195)
$HDI_{MEDIUM} \times TCI$			0.345***
			(0.0387)
$HDI_{HIGH} \times TCI$			0.512***
			(0.0197)
$HDI_{VERYHIGH} \times TCI$			0.613***
			(0.0227)
No. of observations	178	178	178
R^2	0.802	0.882	0.898

Note: the dependent variable is Individual-Average Effectiveness Index, and independent variable is Individual-Average Technical Compliance Index. The relationship is mediated by Human Development Index (HDI), ranging from 0 to 1. We consider continuous HDI and HDI categories (Low, Medium, High, Very High) defined by United Nations. Robust standard errors are used. Zero-intercept is assumed. ***0.01.

findings prompt a deeper examination of regional incapacities. Beyond leading to good-faith non-compliance, technological and financial deficiencies significantly explain why developing regions, even if highly compliant through tremendous efforts, often achieve suboptimal effectiveness.

7.4 | Accounting for Biases

The discussion in the preceding sections assumes that FATF's ratings accurately reflect states' performance. Yet, compliance and effectiveness ratings may be biased due to measurement limitations, political manipulation, or speculative behavior.

First, to address concerns regarding biases in effectiveness evaluation (due to the heavy reliance on compliance results), we analyze its correlations with the Basel Effectiveness Index and the component indices. In the first column of the correlation heatmap (Figure 6), we observe FATF's effectiveness index has 0.65 correlation with the overall Basel index, yet most of which is driven by the Basel FATF component (correlation is 0.89). The correlation with Basel non-FATF component is only 0.30. Among the five sectional components, the *Quality of AML/CFT Framework* section exhibits the highest correlation at 0.85, largely due to the inclusion of the FATF component within this section. However, none of the remaining four domains has correlations higher than 0.60. Specifically, the *Corruption Risks* and *Financial Transparency* domains show correlations of 0.20 and −0.11, indicating low or negative relationships with FATF's effectiveness index. These findings partially reject Hypothesis 7, showing that FATF's effectiveness proxy correlates *weakly* with credible external metrics such as the Basel indices, especially in corruption control and financial transparency.

Despite insufficient correlations with Basel indices, Hypothesis 4–6 may still hold, requiring careful scrutiny before deeming them invalid. Panel analysis results (Table 10) show a positive conversion ratio of 1.74%, but none of the coefficients, even when mediated by HDI, are significant at the 10% level. This indicates a limited impact of compliance on Basel non-FATF effectiveness indicator. However, the Very-High-HDI group exhibits a significantly higher conversion ratio of 7.49%, while other groups show marginally negative coefficients. Hence, overall development still mediates the relationship between compliance and Basel proxy independent of FATF. This analysis does not explore more nuanced regional factors due to the limited sample size relative to the number of variables and fixed effects, which constrains the model to yield meaningful results.

Second, concerns regarding political manipulation of ratings exist. As FATF is led by developed countries, evaluations may exhibit biases against developing nations, particularly in assessing effectiveness. Consequently, even if highly compliant, developing regions may receive lower effectiveness ratings, offering an alternative explanation for their low conversion ratio. However, FATF's peer-reviewed evaluation mechanism within each FATF-Style Regional Body (FSRB) largely mitigates this concern. Jurisdictions within the same FSRB typically share similar

TABLE 9 | From compliance (I-TCI) to effectiveness (I-EI)—Administrative, technological, financial factors as mediators.

	OLS		
	(1)	(2)	(3)
	Admin, tech, and fin (1)	Admin, tech, and fin (2)	Admin, tech, and fin (3)
Regulatory Quality × TCI	0.239*	0.186	0.165
	(0.142)	(0.178)	(0.189)
Access to Internet (% of Population) × TCI	0.285**	0.283**	0.240**
	(0.121)	(0.119)	(0.115)
Financial Development × TCI	0.406***		
	(0.0841)		
Financial Institutions Development × TCI		0.292	
		(0.178)	
Financial Markets Development × TCI		0.172*	
		(0.0893)	
Financial Institutions Access × TCI			0.151*
			(0.0793)
Financial Institutions Depth × TCI			0.0500
			(0.126)
Financial Institutions Efficiency × TCI			0.161
			(0.102)
Financial Markets Access × TCI			0.129
			(0.0778)
Financial Markets Depth × TCI			−0.149
			(0.130)
Financial Markets Efficiency × TCI			0.268***
			(0.0649)
No. of observations	108	108	108
R^2	0.918	0.919	0.928

Note: the dependent variable is Individual-Average Effectiveness Index, and independent variable is Individual-Average Technical Compliance Index. The relationship is mediated by regional factors spanning administrative, technological, and different layers of financial factors. Robust standard errors are used. Zero-intercept is assumed. *0.1, **0.05, ***0.01.

development levels, minimizing the potential for discrimination when conducting mutual evaluations. For instance, *the Action Group against Money Laundering in Central Africa (GABAC)*, comprising Democratic Republic of Congo, Cameroon, Congo, Chad, Central African Republic, Equatorial Guinea, Gabon—most with low to medium HDI (marginally high for Gabon)—deploys a team of experts from within the group to evaluate member states, excluding participation from developed nations. Additionally, the mediation mechanism from overall development remains robust for Basel non-FATF effectiveness proxy, supporting the capacity mediation hypothesis and countering claims of systematic biases in effectiveness ratings against developing nations.

Third, developing states may pursue superficial compliance (Teichmann 2020; Nizzero 2023; Pandey 2024), alternatively explaining why these states exhibit low conversion efficiency into effectiveness. On one hand, possibly due to limited awareness of AML/CFT tasks' importance yet threatened by soft sanctions, developing nations (defiers) have to demonstrate superficially high compliance to satisfy evaluators without substantial efforts. On the other hand, this speculative behavior may stem from cost–benefit trade-offs or capacity constraints. Complying with FATF's Recommendations demands substantial resources, leading developing nations (compliers or interests-maximizers) to adopt a perfunctory approach to meet criteria, as the costs of genuine compliance often exceed the benefits or capacities.

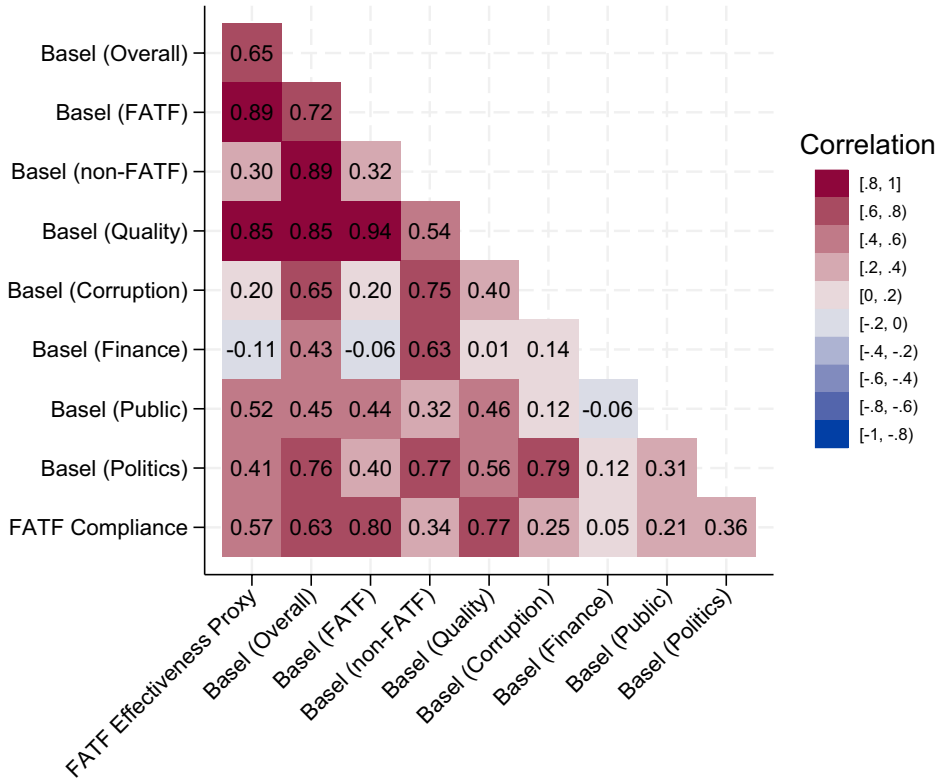

FIGURE 6 | Correlation heatmap between FATF Indices (I-TCI, I-EI) and Basel Indices. *Note:* the heatmap shows the correlations between FATF indices and Basel Effectiveness indices (the inverse of the original Basel AML indices), ranging from 0 (worst) to 1 (optimal). "FATF Effectiveness Proxy" is the Individual-Average Effectiveness Index, "FATF Compliance" is the Individual-Average Technical Compliance Index. "Basel (Overall)" is the Basel Effectiveness Index (overall). "Basel (FATF)" is the FATF component of Basel Index, while "Basel (non-FATF)" is the non-FATF component of Basel Index. "Basel (Quality/Corruption/Finance/Public/Politics)" are the five domain components of Basel Index, referring to Quality of AML/CFT framework, Corruption and fraud risks, Financial transparency and standards, Public transparency and accountability, Political and legal risks, respectively.

8 | Institutional Reflection

In this section, we evaluate FATF's performance-based legitimacy using prior empirical evidence. We examine the three performance-related aspects sequentially, offering policy recommendations to address identified deficiencies if any.

8.1 | Standards

The evidence for the effectiveness in FATF's standards is mixed. First, while a cross-sectional analysis indicates a sufficient correlation between compliance and FATF's effectiveness index, more rigorous panel analysis shows compliance does not significantly contribute to Basel non-FATF effectiveness index. At the global level, FATF should refine its standards design to enhance states' average effectiveness, prioritizing alignment with credible external effectiveness indicators.

Second, incapacities, especially technological and financial deficiencies, are shown to hinder compliance and impede conversion to effectiveness, resulting in considerable regional heterogeneity. Hence, FATF's Recommendations may create additional difficulties for developing nations to comply to. Implementing sophisticated AML/CFT activities outlined by FATF is resource-intensive and time-consuming (Gombiro et al. 2015; Sittlington and Harvey 2018; Podder 2022). Developing states may have the goodwill to adhere to FATF's standards and pursue effectiveness, but often lack necessary capacities to achieve these objectives. Unsatisfactory performance due to innocent developmental constraints can easily place good-faith nations on the gray- or black-list, exposing them to international sanctions. FATF should address these concerns by recognizing capacity gaps and formulating standards that are feasible across varying capacities to the greatest extent possible.

8.2 | Evaluation

The precision of FATF's evaluation is contested. First, the weak or negative correlations between FATF's effectiveness proxies and Basel non-FATF indicators, particularly in corruption control and financial transparency, suggest limited reliability in FATF's effectiveness evaluation. Methodologically, while immediate effectiveness definitions are largely valid and relevant to ultimate AML effectiveness, the evaluation does not directly measure effectiveness—the unobserved treatment effect on money laundering levels, which is generally infeasible to infer for specific states. Instead, the ratings serve as incomplete and biased proxies of effectiveness, heavily reliant on compliance

TABLE 10 | From compliance (I-TCI) to Basel non-FATF effectiveness—HDI as a mediator.

	OLS		
	(1)	(2)	(3)
	No mediation	**Mediation: HDI value**	**Mediation: HDI 4 categories**
TCI	0.0174		
	(0.0634)		
$HDI \times TCI$		0.0638	
		(0.0888)	
$HDI_{LOW} \times TCI$			−0.0354
			(0.0713)
$HDI_{MEDIUM} \times TCI$			−0.0253
			(0.0631)
$HDI_{HIGH} \times TCI$			−0.00145
			(0.0673)
$HDI_{VERYHIGH} \times TCI$			0.0749
			(0.0681)
Jurisdiction FE	Y	Y	Y
Year FE	Y	Y	Y
No. of observations	405	405	405
R^2	0.864	0.864	0.867

Note: the dependent variable is Basel non-FATF Effectiveness Index, and independent variable is Individual-Average Technical Compliance Index. The relationship is mediated by Human Development Index (HDI), ranging from 0 to 1. We consider continuous HDI and HDI categories (Low, Medium, High, Very High) defined by United Nations. Standard errors are clustered at the jurisdiction level.

evaluations with misleadingly strong correlations. To address legitimacy concerns stemming from measurement issues, developing an expanded set of theme-based effectiveness proxies beyond compliance is beneficial, with Basel AML Index composition as a valuable reference for the topic partition. Weighing components based on their logic proximity to the event of concern also enhances the measurement accuracy, for instance, the compliance to *direct* measures should have a higher weight than that to *indirect* measures. Follow-up reassessment of effectiveness is also highly recommended, since timely updates are crucial to monitor the latest progress.

Regarding impartiality, the peer-reviewed evaluation mechanism largely mitigates systematic political manipulation on ratings, particularly effectiveness, against developing nations. However, the potential for individual instances of political manipulation persists. FATF should publicly respond to these contestations and provide further justifications for the impartial evaluations. Even if being impartial and precise in measurement methodology, the ratings may still be biased, for instance, in the case of superficial compliance. Possible measures to address this speculative behavior are mentioned in the following *Strategy* Section.

8.3 | Strategy

The efficacy of FATF's strategy system requires strengthening. Though our causal analysis affirms that the enhanced follow-up effectively fosters long-term compliance, the strategy system has long been dominantly, problematically punitive. Harsh measures, such as gray- and black-listing, when applied to good-faith regions with temporarily poor performance, may raise suspicions of injustice and political manipulation within FATF. This misuse of punitive strategies significantly erodes legitimacy beliefs towards FATF, even if there is no political manipulation. For instance, Pakistani scholars contend that despite substantial compliance efforts, Pakistan faced gray-listing due to political reasons (Mukhtar 2018; Chohan 2020; Ibrahim 2021; Amin et al. 2022).

Remarkably, since the presidency of Elisa de Anda Madrazo of Mexico in July 2024, FATF has prioritized advancing financial inclusion to support low-capacity states. The proposed strategies are primarily *cognitive* and *generative*, particularly benefiting good-faith nations with capacity constraints. In the cognitive strategical dimension (effective for defiers), FATF is developing guidance on AML effectiveness in the low-capacity context to equip the developing nations with necessary awareness and knowledge. We recommend tailored training sessions, designed for developing regions, to assist regional officials in implementing this guidance. In the generative strategical dimension (effective for compliers and interests-maximizers), FATF is uniting an alliance of technical assistance providers, including Multilateral Development Banks (MDBs) to foster effective framework development. This scheme is strongly supported by our empirical evidence that the technological and financial constraints significantly impede compliance and effectiveness conversion.

To implement measures targeting states in need, FATF should first establish criteria to identify low-capacity regions. We recommend prioritizing the identification of technological and financial deficiencies in institutional access and market efficiency based on our evidence. Eligible states may apply for financial aid, technical support, and tailored training from FATF. The usage of these resources should be monitored, with states required to report during follow-up evaluations, ensuring proper allocation to AML/CFT activities. FATF should also regularly update the list of low-capacity regions based on the progress and capacity improvement to avoid overreliance from recipient states. To specifically address potential superficial compliance, we recommend FATF conduct unannounced spot checks on regions with high compliance but low effectiveness to ensure genuine progress.

Additionally, we discuss remunerative and punitive strategies that restructure the costs and benefits. In the remunerative strategical dimension, states that have made remarkable progress or maintained at a high level deserve financial rewards, for instance, grants or assess to low-interest loans from FATF, World Bank or IMF. Enhanced ex post benefits may transform a negative interest-maximizer into a positive one. The remunerative strategies also facilitate sustained progress for compliers and positive interest-maximizers with capacity-constrained performance, provided rewards are reinvested in future AML/CFT activities. Punitive strategies such as gray-listing, which are only applicable to negative interests-maximizers, should be exercised with particular caution to avoid harming good-faith states that struggle with performance.

9 | Conclusion

Synthesizing the evidence presented, we may arrive at the conclusion that the performance-based legitimacy of FATF is contested. With respect to *standards effectiveness*, at the global level, though the cross-sectional evidence suggests the compliance is strongly correlated with FATF's effectiveness proxy, more rigorous panel evidence does not support its significant contribution to effectiveness measured by Basel Institute on Governance. Region-wise, the FATF's Recommendations pose challenges to developing nations with low capacities, since these constraints, particularly technological and financial deficiencies, limit the compliance and conversion into effectiveness.

In terms of *evaluation impartiality and precision*, the measurement methodology of FATF's effectiveness indicators is problematic for overly relying on the compliance component. The weak correlations between FATF's effectiveness index and external corruption control, financial transparency indicators also suggest the measurement biases. Yet, on the good side, the peer-reviewed evaluation system minimizes the possibility of structural political biases against developing nations, ensuring impartiality. Behavioral factors, such as superficial compliance, may also bias the ratings, necessitating surprise inspection occasionally.

Regarding *strategy efficacy*, the enhanced follow-up is demonstrated to induce persistent compliance compared with the regular follow-up, justifying the positive impact from FATF's interventions. However, punitive strategies with sanctions, such as gray-listing, do not facilitate better performance for good-faith yet low-capacity states. Conversely, these states tend to lose legitimacy beliefs towards FATF and speculate political manipulation on weaponizing FATF against them.

Despite imperfect legitimacy, we recognize FATF's remarkable shift to a financially inclusive regime, particularly supporting developing nations in recent 2 years. The rationality of the cognitive and generative strategies proposed is strongly corroborated by our theory and cross-country evidence. We put forwards additional guidance on the implementation of these strategies, and discuss the applicability of remunerative and punitive strategies with caveats in practice. It is understandable that a standards-wise reform to enhance feasibility across diverse developmental levels is challenging, yet this limitation can be compensated with a supportive strategy system that FATF is currently exploring. Evaluation precision is also essential for an accurate reflection of performance. We recommend developing theme-based effectiveness proxies weighted by their logic proximity to money laundering activities, adopting a methodology similar to the Basel AML Index.

Theoretically, we extend the Mitchell's framework to non-binding institutions such as FATF, demonstrating its universality. We build on this framework by extensively enriching the discussion on effectiveness. First, we interpret effectiveness from a rigorous econometric perspective, laying the quantitative foundation. Yet, we point out that effectiveness, as an unobserved treatment effect, is hard to measure directly, necessitating the proxies to capture it. The effectiveness proxies differ in their logic distance to the outcome(s) of concern, with more proximate proxies being more accurate. We also discuss how to design effective standards by examining the chain of events. In terms of its relationship with compliance, we propose a mediation mechanism from capacities (administrative, technological, financial), where higher-capacity states are expected to achieve higher effectiveness given the same compliance. We finally discuss the appropriateness of different strategies contingent on whether the poor performance is driven by behaviors or incapacities.

More broadly, we enhance the existing methodology for assessing the legitimacy of IO by proposing an objective, quantitative, and performance-based inference framework. Three crucial criteria are emphasized—(a) effective standards design, (b) impartial and precise evaluation, (c) diversified strategies to boost compliance and effectiveness. While our work showcases the potential of objective, data-driven inquiries into the legitimacy of IO (either binding or non-binding), we acknowledge that the analysis remains incomplete. The compliance, effectiveness data facilitate a thorough investigation into the performance such as FATF, yet it is not suited for quantifying the purpose and procedure dimensions. Performance-based (il)legitimacy does not necessarily reflect these other two aspects, which require separate investigation. Future research can contribute, for instance, by measuring the level of transparency and democracy in the decision process (procedural and normative dimension). Metrics such as the composition of meeting participants (nationality, gender, race), the meeting frequency, the number of

related public documents, the quality and quantity of evidence supporting conclusions provide objective indicators for inferring legitimacy. This series of objective, quantitative approaches complement, rather than challenge or surpass, mainstream methods such as qualitative normative comparisons and subjective belief surveys.

Empirically, though we strive for the optimal robustness of our results, there are a few limitations. First, we cannot establish a causal link between capacity constraints and compliance due to challenges in identifying instrumental variables, or other exogenous shocks. Second, regarding the relationship between compliance and FATF's effectiveness index, we perform a cross-sectional analysis, as effectiveness ratings are only available in initial evaluations. Ideally, re-rating the effectiveness during follow-up assessments would enable rigorous panel analysis with two-way fixed effects, similar to our approach with the Basel non-FATF effectiveness index. Third, when examining the treatment effects of enhanced follow-up on compliance, the parallel trends assumption holds broadly, yet pre-trends in some of the sectional compliance indices require cautious interpretation.

Money laundering is far from a victimless crime—it siphons resources, corrodes governance, and erodes trust in global financial systems. As the de-facto guardian of anti-money laundering standards, FATF plays an indispensable role. Yet as our study reveals, even an institution as central as FATF must evolve to sustain its legitimacy in an increasingly complex and unequal world. By offering a performance-based, quantitative approach to evaluate legitimacy, we not only deepen theoretical debates on compliance and institutional authority but also provide a practical framework through which global governance can be both assessed and improved. Ultimately, this paper is not just about FATF—it is about how we think about legitimacy, fairness, and effectiveness in international cooperation. The standards we inherit are not beyond scrutiny. They must be revisited, recalibrated, and, when necessary, reimagined. As we confront hidden harms like money laundering, the way forward lies not only in stronger rules, but in more reflexive institutions—ones willing to examine themselves as closely as they examine others.

Acknowledgments

The authors thank the Basel Institute on Governance for providing access to the Expert Edition Plus dataset on anti-money laundering (AML). This work was generously supported by the National Natural Science Foundation of China under Grant Number [72004189], by the University of Macau's Start-up Research Grant (SRG) under Grant Number [SRG2024-00005-FSS], and by the University of Macau's Multi-Year Research Grant—Level II (Grant No. MYRG-GRG2024-00033-FSS).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors declare that the data underlying this article titled "On the Performance-based Legitimacy of Financial Action Task Force: A Quantitative Exploration" will be shared on reasonable request to the corresponding author. Tremendous efforts were made by the authors to collect, track, and process the official data from the website of the Financial Action Task Force (FATF) alongside complementary data from the World Bank, International Monetary Fund, United Nations Development Program, Basel Institute on Governance, and other public sources.

References

Abedin, W. A. F. A., S. B. Coulibaly, H. A. F. E. Z. Ghanem, E. S. W. A. R. Prasad, and M. A. R. I. L. O. U. Uy. 2024. "Reforms for a 21st Century Global Financial Architecture." Special report.

Alcaide Muñoz, L., M. P. Rodríguez Bolívar, and A. M. López Hernández. 2017. "Transparency in Governments: A Meta-Analytic Review of Incentives for Digital Versus Hard-Copy Public Financial Disclosures." *American Review of Public Administration* 47, no. 5: 550–573.

Al-Suwaidi, N. A., and H. Nobanee. 2021. "Anti-Money Laundering and Anti-Terrorism Financing: A Survey of the Existing Literature and a Future Research Agenda." *Journal of Money Laundering Control* 24, no. 2: 396–426.

Amin, M., A. Gul, S. Akhtar, and A. Naz. 2022. "Financial Action Task Force (FATF) and the Pakistan Stance: An Overview of the Mandate, Charter, Effectiveness in the Political Scenario of Pakistan." *Journal of Policy Research (JPR)* 8, no. 3: 473–480.

Barnett, M. N., and M. Finnemore. 1999. "The Politics, Power, and Pathologies of International Organizations." *International Organization* 53, no. 4: 699–732.

Bauer, M. W., S. Eckhard, J. Ege, and C. Knill. 2016. "A Public Administration Perspective on International Organizations." In *International Bureaucracy: Challenges and Lessons for Public Administration Research*, 1–12. Palgrave Macmillan UK.

Beetham, D. 2013. *The Legitimation of Power*. Bloomsbury Publishing.

Bensassi, S., and A. F. Raz. 2025. "Combating Trade-Related Fraud: Do the Financial Action Task Force Recommendations Bite?" *Economica* 92, no. 365: 322–347.

Bernauer, T., and R. Gampfer. 2013. "Effects of Civil Society Involvement on Popular Legitimacy of Global Environmental Governance." *Global Environmental Change* 23, no. 2: 439–449.

Binder, M., and M. Heupel. 2015. "The Legitimacy of the UN Security Council: Evidence From Recent General Assembly Debates." *International Studies Quarterly* 59, no. 2: 238–250.

Brooks, C., E. Fenton, L. Schopohl, and J. Walker. 2019. "Why Does Research in Finance Have So Little Impact?" *Critical Perspectives on Accounting* 58: 24–52.

Buchanan, A., and R. O. Keohane. 2006. "The Legitimacy of Global Governance Institutions." *Ethics & International Affairs* 20, no. 4: 405–437.

Case-Ruchala, D., and M. Nance. 2020. "FATF Blacklists Don't Work the Way You Think They Do." In *Second International Research Conference on Empirical Approaches to AML and Financial Crime Suppression*, Nassau, Bahamas, 27–29.

Chayes, A., A. H. Chayes, and R. B. Mitchell. 1998. "Managing Compliance: A Comparative Perspective." *Engaging Countries: Strengthening Compliance With International Environmental Accords* 39: 44–45.

Chohan, U. W. 2019. "The FATF in the Global Financial Architecture: Challenges and Implications."

Chohan, U. W. 2020. "The Political Economy of the FATF & IMF in Pakistan During 2019 (Part 1)."

Chong, A., and F. Lopez-De-Silanes. 2015. "Money Laundering and Its Regulation." *Economics and Politics* 27, no. 1: 78–123.

Dellmuth, L. M., J. A. Scholte, and J. Tallberg. 2019. "Institutional Sources of Legitimacy for International Organisations: Beyond Procedure Versus Performance." *Review of International Studies* 45, no. 4: 627–646.

Farazmand, A. 1999. "Globalization and Public Administration." *Public Administration Review* 59: 509–522.

FATF. 2008. *Money Laundering and Terrorist Financing: Vulnerabilities of Commercial Websites and Internet Payment Systems*, 23. FATF/GAFI. Recuperado.

FATF. 2021. "Opportunities and Challenges of New Technologies for AML/CFT." FATF, Paris, France.

FATF. 2022. *Consolidated Processes and Procedures for Mutual Evaluations and Follow-Up: "Universal Procedures", September 2022*. FATF. www.fatf-gafi.org/publications/mutualevaluations/documents/universal-procedures.html.

FATF. 2023. *Procedures for the FATF Fourth Round of AML/CFT Mutual Evaluations, Updated February 2023*. FATF. www.fatf-gafi.org/publications/mutualevaluations/documents/4th-round-procedures.html.

Garland, D. 1996. "The Limits of the Sovereign State: Strategies of Crime Control in Contemporary Society." *British Journal of Criminology* 36, no. 4: 445–471.

Gelemerova, L. 2009. "On the Frontline Against Money-Laundering: The Regulatory Minefield." *Crime, Law and Social Change* 52, no. 1: 33–55.

Gombiro, C., M. Jantjies, and N. Mavetera. 2015. "A Conceptual Framework for Detecting Financial Crime in Mobile Money Transactions." *Journal of Governance and Regulation* 4: 727–734.

Hurd, I. 1999. "Legitimacy and Authority in International Politics." *International Organization* 53, no. 2: 379–408.

Hurd, I. 2008. *After Anarchy: legitimacy and Power in the United Nations Security Council*. Princeton University Press.

Ibrahim, S. A. 2021. "Pakistan's FATF Conundrum." *Pakistan Horizon* 74, no. 2–3: 163–183.

IMF. 2023. "Anti-Money Laundering and Combating the Financing of Terrorism." *IMF Discussion on Financial Integrity*.

Janský, P., M. Palanský, and D. Wójcik. 2023. "Shallow and Uneven Progress Towards Global Financial Transparency: Evidence From the Financial Secrecy Index." *Geoforum* 141: 103728.

Jervis, R. 1998. "Realism in the Study of World Politics." *International Organization* 52, no. 4: 971–991.

Johnson, J. 2008. "Third Round FATF Mutual Evaluations Indicate Declining Compliance." *Journal of Money Laundering Control* 11, no. 1: 47–66.

Jones, E., and P. Knaack. 2019. "Global Financial Regulation: Shortcomings and Reform Options." *Global Policy* 10, no. 2: 193–206.

Jongen, H., and J. A. Scholte. 2024. "Institutional Sources of Legitimacy in Multistakeholder Global Governance at ICANN." *Regulation & Governance* 18, no. 3: 1018–1039.

Karlsson-Vinkhuyzen, S. I., and A. Vihma. 2009. "Comparing the Legitimacy and Effectiveness of Global Hard and Soft Law: An Analytical Framework." *Regulation & Governance* 3, no. 4: 400–420.

Kaufmann, D., and A. Kraay. 2024. "The Worldwide Governance Indicators: Methodology and 2024 Update." SSRN 5154675.

Kyllönen, S. 2006. "Ympäristöongelmat ja demokraattinen legimiteetti." *Oikeus* 35, no. 4: 595–610.

Lenz, T., and L. A. Viola. 2017. "Legitimacy and Institutional Change in International Organizations: A Cognitive Approach." *Review of International Studies* 43, no. 5: 939–961.

Levi, M., P. Reuter, and T. Halliday. 2018. "Can the AML System Be Evaluated Without Better Data?" *Crime, Law and Social Change* 69, no. 2: 307–328.

Littrell, C. 2022. "Biases in National Anti-Money Laundering Risk Assessments." SSRN 4137532.

Mearsheimer, J. J. 2017. "The False Promise of International Institutions." In *International Organization*, 237–282. Routledge.

Mekpor, E. S., A. Aboagye, and J. Welbeck. 2018. "The Determinants of Anti-Money Laundering Compliance Among the Financial Action Task Force (FATF) Member States." *Journal of Financial Regulation and Compliance* 26, no. 3: 442–459.

Meyer, T. 2014. "How Compliance Understates Effectiveness." *AJIL Unbound* 108: 93–98.

Mitchell, R. B. 1993. "Compliance Theory: A Synthesis." *Review of European Community & International Environmental Law* 2: 327–334.

Mitchell, R. B. 2001. "Institutional Aspects of Implementation, Compliance, and Effectiveness." In *International Relations and Global Climate Change*, 221–244. MIT Press.

Mitchell, R. B. 2008. "Compliance Theory: Compliance, Effectiveness, and Behavior Change in International Environmental Law."

Mitchell, R. B. 2014. "Compliance Theory: An Overview 1. Improving Compliance With International Environmental Law." 3–28.

Mitchell, R. B. 2021. "887 C51 Compliance Theory." In *The Oxford Handbook of International Environmental Law*, 887-C51. Oxford University Press.

Moloney, K., and D. H. Rosenbloom. 2020. "Creating Space for Public Administration in International Organization Studies." *American Review of Public Administration* 50, no. 3: 227–243.

Moravcsik, A. 1997. "Taking Preferences Seriously: A Liberal Theory of International Politics." *International Organization* 51, no. 4: 513–553.

Mudacumura, G. M. 2013. "Accountability and Transparency: Cornerstones of Development and Democratic Governance." In *Challenges to Democratic Governance in Developing Countries*, 37–55. Springer International Publishing.

Mugarura, N. 2011. "The Institutional Framework Against Money Laundering and Its Underlying Predicate Crimes." *Journal of Financial Regulation and Compliance* 19, no. 2: 174–194.

Mukhtar, A. 2018. "Money Laundering, Terror Financing and FATF: Implications for Pakistan." *Journal of Current Affairs* 3, no. 1: 27–56.

Nizzero, M. 2023. "Anti-Money Laundering and Countering Terrorist Financing From an IR, Criminology, and Compliance Perspective: A Theoretical Marriage of Convenience." In *Countering Terrorist and Criminal Financing*, 19–30. CRC Press.

Pandey, D. 2024. "A Comparative Analysis of the FIUs and FATF Compliance of Canada, Australia, The Netherlands and India." *Journal of Money Laundering Control* 27, no. 5: 886–900.

Podder, S. 2022. "Leveraging the Provisions of Open Banking to Fight Financial Crimes." In *Financial Technology and the Law: Combating Financial Crime*, 19–46. Springer International Publishing.

Pol, R. F. 2018. "Anti-Money Laundering Effectiveness: Assessing Outcomes or Ticking Boxes?" *Journal of Money Laundering Control* 21, no. 2: 215–230.

Pontes, R., N. Lewis, P. McFarlane, and P. Craig. 2022. "Anti-Money Laundering in the United Kingdom: New Directions for a More Effective Regime." *Journal of Money Laundering Control* 25, no. 2: 401–413.

Scharpf, F. 1999. *Governing in Europe: Effective and Democratic?* Oxford University Press.

Scholte, J. A., and J. Tallberg. 2018. "Theorizing The Institutional Sources of Global Governance Legitimacy. Legitimacy in Global Governance: Sources, Processes, and Consequences." 56–74.

Scott, W. R. 2005. "Institutional Theory: Contributing to a Theoretical Research Program." In *Great Minds in Management: The Process of Theory Development*, vol. 37, 460–484. Oxford University Press.

Shah, P. 2025. The Inter-Relationship of Money Laundering and Terrorist Financing. AML UAE Discussion.

Sharman, J. C., and D. Chaikin. 2009. "Corruption and Anti-Money-Laundering Systems: Putting a Luxury Good to Work." *Governance* 22, no. 1: 27–45.

Sittlington, S., and J. Harvey. 2018. "Prevention of Money Laundering and the Role of Asset Recovery." *Crime, Law and Social Change* 70, no. 4: 421–441.

Steffek, J. 2003. "The Legitimation of International Governance: A Discourse Approach." *European Journal of International Relations* 9, no. 2: 249–275.

Sterling-Folker, J. 2006. *Realist Approaches. Making Sense of International Relations Theory*, 13. *Boulder, Lynne Rienner Publishers, Inc*.

Stone, R. W. 2013. "Informal Governance in International Organizations: Introduction to the Special Issue." *Review of International Organizations* 8: 121–136.

Tallberg, J., K. Bäckstrand, and J. A. Scholte. 2018. *Legitimacy in Global Governance: Sources, Processes, and Consequences*. Oxford University Press.

Tallberg, J., and M. Zürn. 2019. "The Legitimacy and Legitimation of International Organizations: Introduction and Framework." *Review of International Organizations* 14: 581–606.

Teichmann, F. M. J. 2020. "Money-Laundering and Terrorism-Financing Compliance–Unsolved Issues." *Journal of Money Laundering Control* 23, no. 1: 90–95.

Vabulas, F., and D. Snidal. 2013. "Organization Without Delegation: Informal Intergovernmental Organizations (IIGOs) and the Spectrum of Intergovernmental Arrangements." *Review of International Organizations* 8: 193–220.

Verdugo Yepes, C. 2011. Compliance with the AML/CFT international standard: Lessons from a cross-country analysis.

Verhoest, K., M. Maggetti, E. Guaschino, and J. Wynen. 2025. "How Trust Matters for the Performance and Legitimacy of Regulatory Regimes: The Differential Impact of Watchful Trust and Good-Faith Trust." *Regulation & Governance* 19, no. 1: 3–20.

Vishwanath, T., and D. Kaufmann. 2001. "Toward Transparency: New Approaches and Their Application to Financial Markets." *World Bank Research Observer* 16, no. 1: 41–57.

Woo, J. J., M. Ramesh, M. Howlett, and M. K. Coban. 2016. "Dynamics of Global Financial Governance: Constraints, Opportunities, and Capacities in Asia." *Policy and Society* 35, no. 3: 269–282.

Young, O. R. 1999. *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms*. MIT Press.

Zürn, M. 2004. "Global Governance and Legitimacy Problems." *Government and Opposition* 39, no. 2: 260–287.

Appendix A

TABLE A1 | Determinants of compliance (S-TCI).

	OLS				
	(1)	(2)	(3)	(4)	(5)
	Admin	Admin, tech, socioecon	Admin, tech, socioecon and fin (1)	Admin, tech, socioecon and fin (2)	Admin, tech, socioecon and fin (3)
In 4th-round Follow-up	0.155***	0.109***	0.112***	0.113***	0.112***
	(0.0132)	(0.0136)	(0.0138)	(0.0139)	(0.0142)
Voice and Accountability	−0.0406	−0.0393	−0.0268	−0.0580	−0.0547
	(0.0592)	(0.0572)	(0.0585)	(0.0630)	(0.0638)
Regulatory Quality	0.521***	0.142**	0.0591	0.0624	0.108
	(0.0561)	(0.0651)	(0.0769)	(0.0759)	(0.0781)
Access to Internet (% of Population)		0.313***	0.299***	0.280***	0.269***
		(0.0377)	(0.0380)	(0.0397)	(0.0412)
Foreign Direct Investment (% of GDP)		0.0281***	0.0290***	0.0294***	0.0320***
		(0.00854)	(0.00885)	(0.00950)	(0.00992)
Government Education Expenditure (% of GDP)		−0.370	−0.363	−0.327	0.0497
		(0.505)	(0.503)	(0.507)	(0.538)
Financial Development			0.0858**		
			(0.0402)		
Financial Institutions Development				0.133**	
				(0.0589)	
Financial Markets Development				0.00149	
				(0.0341)	
Financial Institutions Access					0.0835***
					(0.0304)
Financial Institutions Depth					−0.0488
					(0.0476)
Financial Institutions Efficiency					0.127*
					(0.0723)
Financial Markets Access					0.0139
					(0.0347)
Financial Markets Depth					0.0383
					(0.0507)
Financial Markets Efficiency					−0.00595
					(0.0267)
Constant	0.220***	0.285***	0.298***	0.290***	0.211***
	(0.0272)	(0.0257)	(0.0266)	(0.0278)	(0.0491)
No. of Observations	364	364	364	364	364
R^2	0.430	0.535	0.539	0.542	0.550

Note: the dependent variable is Section-Average Technical Compliance Index. Admin, Tech, Socioecon and Fin represents administrative, technological, socioeconomic, and financial factors. The financial factors are examined from the overall level to more nuanced levels by a top-down approach. Robust standard errors are used. *0.1, **0.05, ***0.01.

TABLE A2 | Estimated treatment effects from enhanced follow-up on compliance (S-TCI).

	DID			Event study
	(1)	(2)	(3)	(4)
$EFU \times POST$	0.127***	0.158***	0.0451***	
	(0.0114)	(0.00909)	(0.00956)	
$EFU \times 1[STAGE = -1]$				0.0293*
				(0.0174)
$EFU \times 1[STAGE = 1]$				0.0132
				(0.00904)
$EFU \times 1[STAGE = 2]$				0.0606***
				(0.0131)
$EFU \times 1[STAGE = 3]$				0.0715***
				(0.0172)
$EFU \times 1[STAGE = 4]$				0.0667***
				(0.0159)
$EFU \times 1[STAGE = 5]$				0.0875***
				(0.0208)
Constant	0.513***	0.501***	0.547***	0.535***
	(0.0114)	(0.00370)	(0.00389)	(0.00611)
Jurisdiction FE	N	Y	Y	Y
Year FE	N	N	Y	Y
No. of observations	664	664	664	664

Note: the dependent variable is Section-Average Technical Compliance Index. *EFU* is the treatment indicator for enhanced follow-up ($EFU = 1$) versus regular follow-up ($EFU = 0$). *POST* is the simple stage indicator where 1 means in the fourth-round follow-up and 0 means otherwise. $1[STAGE = j], j = -1,1,2,3,4,5$ are nuanced stage indicators where 1 means in the stage j and 0 otherwise. The coefficients of $EFU \times POST$ are the simple DID estimates, while the coefficients of $EFU \times 1[STAGE = j], j = -1,1,2,3,4,5$ are the event study estimates. Standard errors are clustered at the jurisdictions level. *0.1, ***0.01.

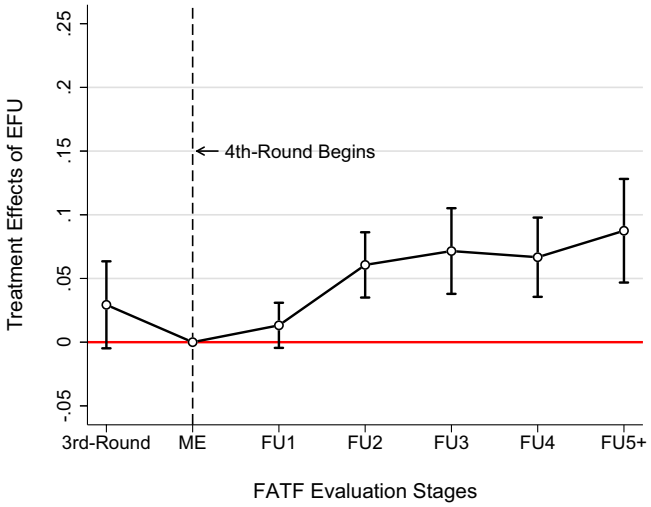

FIGURE A1 | Time trend of treatment effects from enhanced follow-up on compliance (S-TCI). *Note:* the figure shows the time trend of treatment effects from the enhanced follow-up on the Section-Average Technical Compliance Index, compared with the regular follow-up. The y-axis shows the treatment magnitude during different stages. The x-axis shows the timeline of evaluation as stages: "3rd-Round" means the third-round final assessment, "ME" means the fourth-round initial mutual evaluation (the baseline period), "FU1" means the first follow-up (the same logic applies for "FU2", "FU3", "FU4"), "FU5+" means the fifth or later follow-ups.

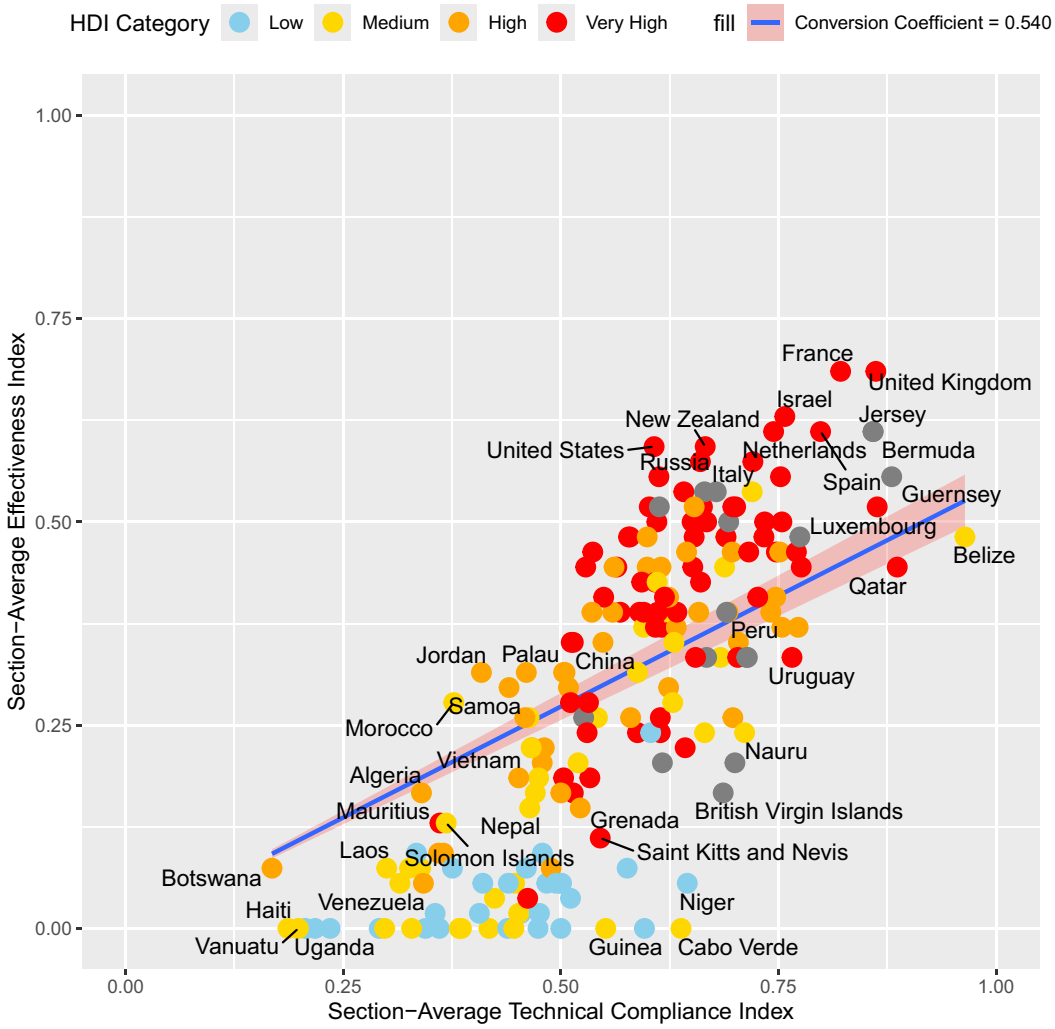

FIGURE A2 | Scatter Plot of Effectiveness Index (S-EI) and Technical Compliance Index (S-TCI). *Note:* the above scatter plot shows the relationship between Section-Average Effectiveness Index and Technical Compliance Index. The fitted zero-intercept linear coefficient is 0.540. The regions are mapped with Human Development Index (HDI) categories - Low, Medium, High, and Very High. Gray dots are regions with missing HDI values.

TABLE A3 | From compliance (S-TCI) to effectiveness (S-EI)—HDI as a mediator.

	OLS		
	(1)	(2)	(3)
	No mediation	**Mediation: HDI value**	**Mediation: HDI 4 categories**
TCI	0.540***		
	(0.0177)		
$HDI \times TCI$		0.718***	
		(0.0164)	
$HDI_{LOW} \times TCI$			0.101***
			(0.0234)
$HDI_{MEDIUM} \times TCI$			0.395***
			(0.0414)
$HDI_{HIGH} \times TCI$			0.574***
			(0.0212)
$HDI_{VERYHIGH} \times TCI$			0.661***
			(0.0194)
No. of observations	178	178	178
R^2	0.834	0.905	0.918

Note: the dependent variable is Section-Average Effectiveness Index, and independent variable is Section-Average Technical Compliance Index. The relationship is mediated by Human Development Index (HDI), ranging from 0 to 1. We consider continuous HDI and HDI categories (Low, Medium, High, Very High) defined by United Nations. Robust standard errors are used. Zero-intercept is assumed. ***0.01.

TABLE A4 | From compliance (S-TCI) to effectiveness (S-EI)—Administrative, technological, financial factors as mediators.

	OLS		
	(1)	(2)	(3)
	Admin, tech, and fin (1)	**Admin, tech, and fin (2)**	**Admin, tech, and fin (3)**
Regulatory Quality × TCI	0.325**	0.281	0.190
	(0.148)	(0.184)	(0.198)
Access to Internet (% of Population) × TCI	0.328***	0.327***	0.286***
	(0.121)	(0.120)	(0.107)
Financial Development × TCI	0.325***		
	(0.0863)		
Financial Institutions Development × TCI		0.236	
		(0.187)	
Financial Markets Development × TCI		0.136	
		(0.0952)	
Financial Institutions Access × TCI			0.117
			(0.0777)
Financial Institutions Depth × TCI			−0.0164
			(0.128)
Financial Institutions Efficiency × TCI			0.235**
			(0.113)
Financial Markets Access × TCI			0.0962
			(0.0756)
Financial Markets Depth × TCI			−0.109
			(0.127)
Financial Markets Efficiency × TCI			0.264***
			(0.0631)
No. of observations	108	108	108
R^2	0.925	0.925	0.936

Note: the dependent variable is Section-Average Effectiveness Index, and independent variable is Section-Average Technical Compliance Index. The relationship is mediated by regional factors spanning administrative, technological, and different layers of financial factors. Robust standard errors are used. Zero-intercept is assumed. **0.05, ***0.01.

TABLE A5 | From compliance (S-TCI) to Basel non-FATF effectiveness—HDI as a mediator.

	OLS		
	(1)	(2)	(3)
	No mediation	**Mediation: HDI value**	**Mediation: HDI 4 categories**
TCI	0.0431		
	(0.0680)		
$HDI \times TCI$		0.0990	
		(0.0956)	
$HDI_{LOW} \times TCI$			−0.0107
			(0.0673)
$HDI_{MEDIUM} \times TCI$			0.000199
			(0.0640)
$HDI_{HIGH} \times TCI$			0.0319
			(0.0743)
$HDI_{VERYHIGH} \times TCI$			0.108
			(0.0734)
Jurisdiction FE	Y	Y	Y
Year FE	Y	Y	Y
No. of observations	405	405	405
R^2	0.864	0.865	0.867

Note: the dependent variable is Basel non-FATF Effectiveness Index, and independent variable is Section-Average Technical Compliance Index. The relationship is mediated by Human Development Index (HDI), ranging from 0 to 1. We consider continuous HDI and HDI categories (Low, Medium, High, Very High) defined by United Nations. Standard errors are clustered at the jurisdiction level.

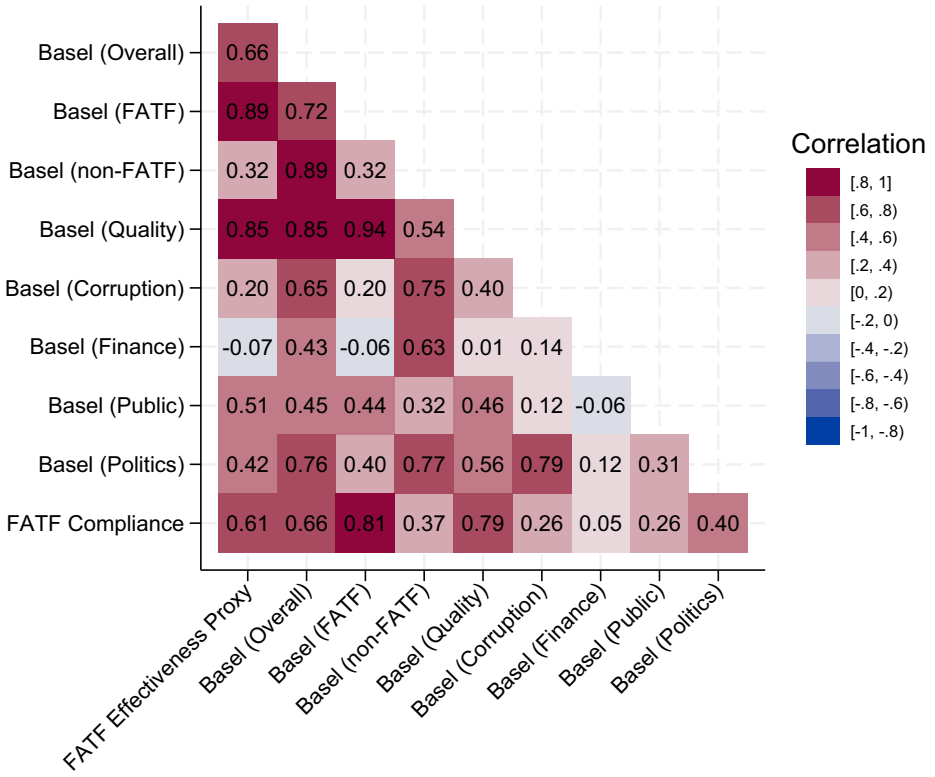

FIGURE A3 | Correlation heatmap between FATF Indices (S-TCI, S-EI) and Basel Indices. *Note:* the heatmap shows the correlations between FATF indices and Basel Effectiveness indices (the inverse of the original Basel AML indices), ranging from 0 (worst) to 1 (optimal). "FATF Effectiveness Proxy" is the Section-Average Effectiveness Index, "FATF Compliance" is the Section-Average Technical Compliance Index. "Basel (Overall)" is the Basel Effectiveness Index (overall). "Basel (FATF)" is the FATF component of Basel Index, while "Basel (non-FATF)" is the non-FATF component of Basel Index. "Basel (Quality/Corruption/Finance/Public/Politics)" are the five domain components of Basel Index, referring to Quality of AML/CFT framework, Corruption and fraud risks, Financial transparency and standards, Public transparency and accountability, Political and legal risks, respectively.

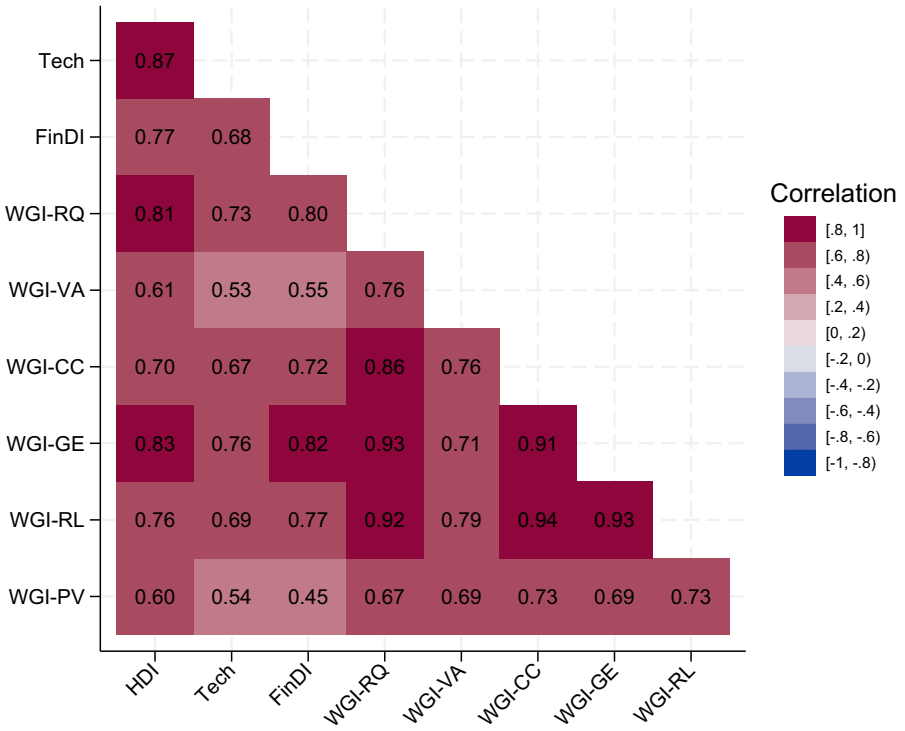

FIGURE A4 | Correlation heatmap between HDI and other regional covariates. *Note:* the heatmap shows the correlations between Human Development Index (HDI) and other regional covariates. "Tech" is the population proportion that has access to Internet. "FinDI" is the financial development index. "WGI-RQ/VA/CC/GE/RL/PV" is a series of worldwide governance indicators, denoting Regulatory Quality, Voice and Accountability, Control of Corruption, Government Effectiveness, Rule of Law, Political Stability and Absence of Violence/Terrorism, respectively.

Biographies

Guo Cheng is a recent graduate with a master's degree in computational social science (Economics Track) from the University of Chicago (UChicago). He now serves as a part-time graduate researcher at the UChicago. His research interests focus on the intersection of development economics, public policy, and sociology.

Bo Wen is currently an associate professor in the Department of Government and Public Administration (DGPA) at the University of Macau. His primary areas of study are public management, organization theory and behavior, institutional analysis, policy implementation, regulatory governance, and Chinese politics. His work appears in Public Administration Review, The China Quarterly, The American Review of Public Administration, International Public Management Journal, among others.