Anomalies and Links to Market: Supranational Evidence Abroad from New Market Efficiency Measures

XI DONG, YAN LI, LIANA LI, DAVID E. RAPACH, and GUOFU ZHOU*

ABSTRACT

Cross-sectional anomalies and time-series market returns are jointly determined in equilibrium, suggesting a necessity to unify the two central literatures for more general understanding of asset pricing. Examining 44 non-U.S. countries, we find that representative cross-sectional anomalies are mostly insignificant at the country level, but become significant once aggregated to the supranational level. After aggregation, supposedly "market-neutral" long-minus-short anomaly returns predict developed-market returns, while "market-exposed" long-or-short anomaly returns predict emerging-market returns. Furthermore, characteristics—foundational to cross-sectional predictability—become useful in time-series predictability to some extent after supranational aggregation as well. We rationalize international anomaly-market links by decomposing them into novel measures of foundational market efficiency concepts, including inter-temporal, systematic importance of mispricing, relative importance of price randomness, and asymmetric mispricing correction speed. The first and the latter two factors shape the links differently across markets of different maturity. We address data-mining in testing international anomaly-market linkages by assuming anomalies and their market-return predictability are both data-mined domestically.

*XI DONG is at Baruch College, the City University of New York. YAN LI is at the Southwestern University of Finance and Economics. LIANA (YANRAN) LI is at Baruch College, the City University of New York. DAVID E. RAPACH is at the Federal Reserve Bank of Atlanta. GUOFU ZHOU is at Washington University in St. Louis. We would like to thank Jun (Jonathan) Wang, Lin Peng, Dashan Huang, Roger Loh, Ying Wu, and all the participants at the "AI Era in Finance" Symposium 2024 and the seminars at Baruch College, the City University of New York, Singapore Management University, Peking University, and Southwest University of Finance and Economics for their helpful comments and suggestions.

JEL Classification: G11, G14
Keywords: Return predictability, Market excess return, Cross-sectional anomaly return, International market efficiency, Machine learning, Limits of arbitrage

I. Introduction

In financial research, the quest for decoding the predictability of stock returns stands at the frontier. Within the vast expanse of this literature, two predominant strands emerge: The first is primarily concerned with the prediction of cross-sectional stock returns, or in simpler terms, the identification of anomalies (e.g., McLean and Pontiff (2016), Hou, Xue, and Zhang (2020)). The second strand utilizes a range of economic and financial variables to predict time-series market excess returns (e.g., Nelson (1976), Campbell (1987), Fama and French (1988, 1989), Pástor and Stambaugh (2009)). Following the debate between Engelberg et al. (2023) and Dong et al. (2022) on U.S.-based domestic evidence, there has been a surge in interest in studying whether and how these two strands of literature can be linked together.[1] The answer to this question is pivotal as it sits at the potential turning point of unifying two central stands of asset pricing literature. Given asset prices are jointly determined in equilibrium in the cross section at the stock level and in the time series at the macro level, the unification of these two fields is a necessary step for a more general understanding of asset pricing and market efficiency.[2]

In this paper, we examine the links between anomalies and market returns by investigating a large set of representative anomalies and their relation with market excess returns, using data from 44 countries, and across major alternative databases. Our investigation is characterized by several novel insights and methodologies. First, we aggregate data across countries using market capitalization weighting to form "supranational" level anomalies and market excess returns. Our findings reveal that these supranational anomalies, which capture the mispricing forces operating over and beyond the borders of countries, exhibit predictive patterns for market returns that significantly diverge from those observed using conventional country-level analyses. This finding suggests the importance of macro-, supranational-level, as opposed to micro-, country-level, inefficiency in generating both anomalies and their links with the time-series macro-level market returns, implying the superiority of considering supranational-level factors in developing international asset pricing models. Second, we reveal that anomaly-market links are drastically different across developed markets vs. emerging markets–markets of different maturity. Long-short anomaly returns, which are supposed to be "market-neutral", predict the market returns of the former while long *or* short anomaly returns predict the market returns of the latter markets. Third, to understand international anomaly-market links, we extend the theoretical framework of Dong et al. (2022) and, in particular,

[1] One year after the publication of both papers, the collective Google cites have reached 160. Engelberg et al. (2023) also won JFQA's William Sharpe Best Paper Award.

[2] The necessity of this unification is analogous to the attempt in physics to unify quantum mechanics governing the micro world and general relativity governing the macro world, as there should be a general rule governing both micro and macro worlds.

propose a novel decomposition of the anomaly-market linkage to extract cross-country differences in market efficiency along three foundational dimensions: the inter-temporal systematic importance of mispricing, the relative importance of price randomness, and asymmetric mispricing correction speed. Fourth, following Engelberg et al. (2023), we use market-wide anomaly characteristics to predict market returns and find that anomaly characteristics can work at some extent after supranational aggregation. We explain the difference between the predictive power of anomaly characteristics and anomaly returns in the context of our framework. Fifth, we propose a novel approach to account for the data mining concerns stemming from U.S. domestic data and show that our findings remain even if we assume anomalies and their predictive power of market excess return are both data-mined domestically. Finally, we provide the first international trading evidence, as well as other results, to support the channels underlying our findings.

Our tests focus on 100 anomalies, primarily derived from Dong et al. (2022), which discovers that these anomaly returns can predict the U.S. market return.[3] We build anomalies in two ways. The first way is a conventional, country-level approach, predominantly used by prior literature, where anomalies are constructed within each country. The second way is a supranational level approach, where we aggregate the returns of each anomaly across all countries in a selected country set using each country's lagged market capitalization as the weight. Our study constitutes four sets of results.

In the first set of results, we start by verifying the replicability of anomalies abroad. Our results show that only a small portion (17.49%) of anomalies can be successfully replicated in each foreign country on average based on the 10% significance level. To the extent that anomalies may capture mispricing, our finding suggests that anomalies are very noisy proxies abroad for mispricing at the country level. However, after we aggregate the anomalies at the supranational level by favoring the systematically important countries through value weighting, replication rates substantially increase. In some cases, most anomalies can be replicated.

We then examine the predictive ability of these anomaly returns on market excess returns at both the country and supranational levels, connecting cross-sectional predictability with time-series market return predictability internationally. To alleviate the issue of short time series in international data, we consider both out-of-sample (OOS) and in-sample (IS) tests, the latter of which can be more powerful in small samples (Inoue and Kilian (2005)). We primarily employ the simple predictor average shrinkage method introduced in Dong et al. (2022), as it is the only method applicable for both out-of-sample (OOS) and in-sample (IS) tests in a high-dimensional setting for aggregating information across anomalies. The method is theoretically and empirically demonstrated

[3]In Internet Appendix Table A4, we check robustness using the 153 anomalies from Jensen et al. (2023).

to be the most robust method, under many reasonable conditions, for diversifying away idiosyncratic noise and avoiding overfitting (e.g., Yuan and Zhou (2022); DeMiguel et al. (2009)). Notwithstanding, we also employ a range of shrinkage techniques, including forecast-combination and machine learning, and obtain robust results.

We find strikingly different patterns for supranational vs. country-level tests. In out-of-sample (OOS) tests, the minimal significance of anomaly returns observed at the country level may imply that anomaly returns at this level may also be ineffective in forecasting country-specific market returns. In fact, the OOS statistics R^2_{OS} indicate that there is only moderate economic significance for developed countries like G6 (G7 countries excluding the US) under one shrinkage method, with the maximum R^2_{OS} reaching 1.40%. In sharp contrast, at the supranational level, the OOS R^2_{OS}, ranging from 0.91% to 6.92%, are all sizable and statistically significant under all six shrinkage methods, exceeding the 0.5% threshold for economic significance (Campbell and Thompson (2008)). For IS tests, although country-level predictability becomes stronger, it remains much weaker than supranational-level predictability.

Overall, our first set of results advocates a supranational perspective on anomalies and their linkage to market returns, departing from prior literature's conventional country-level perspective. The strong supranational performance is the result of several reasons. First, the mispricing that is not only relevant for the cross section but also for the aggregate time series, henceforth referred to as the intertemporal, systematic important mispricing, is more likely to originate from economically more important countries. As Samuelson's Dictum dictates, the informational inefficiency is stronger at the more macro level (see, e.g., Xiao et al. (2022), Xiao et al. (2023)). Second, value-weighted aggregation diversifies the noise at the country level and is more likely to capture systematically important intertemporal mispricing that affects all countries.

In the second set of results, we extend the framework of Dong et al. (2022) to provide novel insights into the relation between anomalies and market returns in the international setting.[4] Their framework outlines a data-generating process (DGP) to illustrate the predictive capability of a *representative* long-short anomaly portfolio return. We first identify a richer set of mechanisms embedded in their framework that can generate a linkage between anomaly and market returns in the international setting than in the U.S. setting. The first mechanism is mispricing correction persistence (MCP), where long and short legs (extreme deciles) of anomalies capture salient mispricing that is persistently corrected over extended periods of time. The second mechanism is the inter-temporally systematic importance of the mispricing (ISIoM), in the sense that the correction of misrpicing begins from the most salient, mispriced segment of the market–represented by

[4]This approach is more consistent with an interpretation that anomalies reflect mispricing. Whether our findings can be explained in a risk-based framework is an interesting topic for future research.

anomaly long or short legs–and gradually extends to the broad markets, consistent with the view of prominent asset managers.[5] These MCP and ISIoM mechanisms allow both anomaly long and short legs to predict the market return. However, Dong et al. (2022) show that overpricing appears to dominate underpricing in MCP and ISIoM in the U.S. market such that only short-leg anomaly returns can predict the market return. It is unclear whether overpricing dominance applies to the international markets. The third mechanism explains why, counter-intuitively, the "supposedly market-neutral" long-short anomaly return can be more powerful than the long or short anomaly return in predicting the market return. This paradox arises because the U.S. market return encompasses a significant martingale component, consistent with the conventional belief that the U.S. market is informationally very efficient. Thus, exactly because the market is hard to predict, the "market-neutral" long-short anomaly return helps filter out the noise introduced by the unpredictable component. However, noise reduction comes with a cost: long-short cancels out the MCP and ISIoM in both the short and long legs to some degree. Consequently, whether long-short works better than long and short depends on the interplay between the asymmetry of MCP and ISIoM vs. the relative importance of noise in different country markets.

We then extend the data-generating framework to allow for heterogeneity around the world. Central to our analysis is the decomposition of the "beta", the predictive coefficient of a representative long-short anomaly return. This decomposition yields three critical components that reflect market efficiency along distinctive foundational notions. The first component is $\tilde{\beta}_S$, the coefficient of using the short-leg anomaly return to predict market excess return. It reflects the inter-temporally systematic importance of the mispricing (ISIoM) captured in the anomaly's short leg. Under the assumption persistent mispricing dominates transient mispricing, a more positive $\tilde{\beta}_S$ signifies greater inter-temporal systematic importance in the sense that the overpricing within the short-leg segment of the market is informative about the overpricing in the aggregate market. In the international context, the systematic importance of mispricing captured by anomaly long- or short-leg returns differs across supranations.[6] The second component is the "autocovariance ratio", measured as the autocovariance difference between the long- and the short-leg return, divided by the autocovariance of the short-leg return. This ratio measures the asymmetry of MCP, which reflects the asymmetry in the speed of mispricing correction in a country. A negative value indicates overpricing is more slowly corrected than underpricing, which we

[5]For example, in the same year when Dong et al. (2022) was published, the legendary asset manager Jeremy Grantham of GMO published a note stating that "at the end of the great bubbles [in history] it seems as if the confidence termites attack the most speculative and vulnerable first and work their way up, sometimes quite slowly, to the blue chips."

[6]The systematic importance also differs across anomalies. See Dong et al. (2024) for a detailed discussion.

refer to as overpricing dominance. We find the ratio is negative for almost all countries, suggesting that persistent bubbles are a more important mispricing concept in almost all countries than persistent underpricing. This persistence-based notion of overpricing dominance, which is informative about the horizon of bubbles, complements prior studies focusing on the magnitude of overpricing and is critical for understanding market inefficiencies and bubbles (e.g., Hong et al. (2000), Stambaugh, Yu, and Yuan (2012, 2015), Avramov et al. (2013), Baba Yara et al. (2020)). The third component is the "variance ratio", measured as the variance of the short-leg return divided by the variance of the long-short portfolio return. We demonstrate analytically that a higher variance ratio indicates a higher importance of the unpredictable component of a country's stock market prices relative to the country's persistent, and thus predictable, overpricing component. To the extent that price following a random walk is a foundational notion of market efficiency, the variance ratio can be interpreted as a market-level measure of the relative importance of price randomness of a country. Under the null random walk hypothesis, in a market where all information is efficiently priced, we shall expect all variability in the price to be unpredictable and none of it comes from mispricing correction.

Finally, we apply our decomposition to understanding how heterogeneity in market efficiency across different supranations shapes the linkage between anomalies and market returns. We triple-sort countries into supranations based on the above three components. We then use the supranational-aggregated anomaly returns to predict the corresponding market returns. We find that the supranation with high (low) values of all three components, denoted as HHH (LLL), achieves the highest (lowest) OOS R^2_{OS} as high (low) as 6.47 (-0.44), while the supranations with high values in two or one of the three components obtain OOS R^2_{OS} values that lie in between the HHH and LLL supranations.

In our third set of results, we delve deeper into the concept of supranational anomalies more thoroughly by adopting a widely recognized division of the global market into two distinct, non-overlapping groups: developed and emerging markets. This division is particularly relevant and economically meaningful in our context, as the two supranations exhibit very different market maturity, which implies distinct values of the three components of market efficiency based on our decomposition. Through this decomposition, we uncover novel insights on how market maturity impacts the anomaly-market linkage.

We first examine the three-factor alpha of DLRZ100 anomalies in these two markets. Existing works such as Jacobs (2016) use developed and emerging markets to proxy for markets with different maturity and show that mature (i.e., developed) markets surprisingly have equal or more significant mispricing than immature (i.e., emerging) markets. On the surface, our results at the country level appear to be consistent with his conclusion. Indeed, the three-factor alpha of the Emerging countries is smaller than that of the Developed countries. However, when analyzed at the supranations level, both the three-factor alpha in the Emerging supranation substantially rises, becoming 32% higher than that observed in the Developed supranation. This finding suggests that prior research that uses a country-level analysis may have underestimated the macro, supranational-level inefficiencies captured by emerging-market anomaly returns. The stark contrast between the supranational and country-level analyses underscores the value of adopting a supranational perspective to better understand anomalies and their linkage with the market return.

We then examine anomaly returns' predictability on their corresponding supranational market returns. Counter-intuitively, the Emerging long-short anomaly returns, despite being more significant than the Developed ones at the supranational level, lag far behind the Developed supranation in forecasting market returns. However, a different pattern emerges if we examine the long and short anomaly returns separately. Both long- *and* short Emerging supranational anomaly returns strongly predict emerging-market returns, generating OOS R^2_{OS} of 3.60 and 3.69, respectively.

Next, we turn to the decomposition to understand the difference between developed and emerging markets. We find that $\tilde{\beta}_S$ is significantly smaller whereas the value of autocovariance ratio and the variance ratio are significantly higher in the Developed supranation than those in the Emerging supranation. This suggests that although the persistent overpricing captured by anomalies is systematically more important for the Emerging supranation than in the Developed supranation, the speed of mispricing correction is much less asymmetric in the Emerging supranation. In other words, both underpricing and overpricing are corrected slowly in emerging markets. Consequently, the noise-reduction mechanism of the long-short return in the emerging supranation is not very effective. Furthermore, the relative importance of price randomness is low in the emerging supranation, further reducing the effectiveness of the noise-reduction mechanism of the long-short return. However, unlike the Developed supranation, both underpricing and overpricing captured by anomaly legs are inter-temporally, systematically important in the Emerging supranation. Taken together, while the inter-temporal systematic importance of both overpricing and underpricing is stronger in less mature markets, strengthening the predictive ability of both long and short-leg anomaly returns on market returns, the less asymmetric speed in mispricing correction and the lower the relative importance of price randomness weakens the linkage between long-minus-short anomaly returns and market returns for less matured markets.

In the fourth set of results, we take the angle of Engelberg et al. (2023) and examine whether the market-wide anomaly characteristics predict market returns internationally. Given that characteristics are the central cross-sectional return predictors, Engelberg et al. (2023) is well justified to examine the relation between characteristics and the market return. To understand the predictive power of characteristics, we further extend the data-generating framework. We show that characteristics introduce noise into time-series prediction primarily due to a time-varying component unrelated to mispricing.

The impact of this noise is exacerbated by the persistence nature of characteristics. The persistence of characteristics may also capture a persistent difference in cross-sectional returns, allowing them to be strong cross-sectional return predictors while being not useful for the time-series prediction. In addition, we show anomaly characteristics become useful in time-series predictability to some extent after supranational aggregation. Our findings suggest that although anomaly characteristics could introduce more noise compared to anomaly returns, supranational aggregation may still help mitigate them to some extent.

In the fifth set of results, we address data mining concerns, which are prevalent in the cross-sectional literature (e.g., Harvey et al. (2016), Hou et al. (2020)). In the new literature bridging cross-sectional and time-series return predictability, there is a lack of any approach to address date mining. We develop a novel approach to address such concerns in this setting. Our approach allows to address two central data-mining concerns simultaneously or separately. First, anomalies are discovered in the US market. Second the linkage between anomalies and the market return is also discovered in the US. market. In the first version of our approach, we remove the components of supranational predictors that co-move with their US counterparts, effectively treating domestic anomalies as completely data-mined ones. In the second, stronger version, we also remove the components of supranational market excess returns that co-move with the US market excess return. Given that our first set of results suggest that systematically important supranations are likely to be the important contributors to the anomaly-market linkage, removing the systematic important U.S. market return introduces a very strong adjustment. It essentially assumes that the existence of U.S. anomalies and their linkage with the U.S. market return are both data-mined. Even after both adjustments, we still observe that anomaly returns can predict market returns abroad.

In the last set of results, we provide international evidence to support other testable restrictions on the unique limits of arbitrage that generate the anomaly-market return linkage. First, employing proprietary data, we provide the first comprehensive analysis on the relation between anomaly returns and short selling at the market level in international markets. Short selling is an archetypal form of arbitrage and particularly relevant given overpricing dominance is one central mechanism underlying our findings. We find that increase in supranational long-short anomaly returns predicts increase in market-wide short selling in that supranation. The results support that high cross-sectional anomaly returns foreshadow market-wide arbitrageurs' correction of overpricing. Second, we classify anomalies into subgroups using three proxies for limits of arbitrage: bid-ask spread, idiosyncratic volatility, and market capitalization. We find that the long-short returns of anomalies with relatively stronger (weak) limits of arbitrage in their short legs than in their long legs display more (less) out-of-sample predictive ability for the market return. Third, we segment the sample periods based on different friction measures. We find significant increases in out-of-sample market return predictability during periods of heightened friction.

Literature and Contribution Our study contributes to a nascent but fast-growing literature seeking to bridge or unify the cross-sectional anomaly literature and the time-series market return predictability literature. Originating from a debate between Engelberg et al. (2023) and Dong et al. (2022), our study introduces fresh perspectives and insights on this endeavour, resolving various puzzles on several aspects.

First, by contrasting results aggregated at the supranational level with those obtained at the country level, we provide a novel perspective for future research to explore anomalies and their relationship with market returns abroad. First, our supranational aggregation highlights the critical role of systematically important mispricing across individual countries and the potential for noise diversification across nations. Second, the supranational aggregation approach also yields new insights that were previously considered puzzling in the cross-sectional anomaly literature. For instance, the difference between our supranational-level results and Jacobs (2016)'s result on the developed- vs. emerging-market anomalies challenges the prior literature's view on how market maturity is related to anomalies. The high replication rate of anomalies internationally reported in Jensen et al. (2023) is also largely attributable to supranational aggregation. Third, the anomaly literature has predominantly focused on equal-weighted anomaly returns. This focus overlooks the economic significance of mispricing and conflicts with the reality that all asset pricing models use value-weighted factors and prioritize explaining away the alphas of larger firms (e.g., Fama and French (1996)). Extending this equal-weighting logic to the international literature, a great number of studies draw their inferences by averaging anomaly results equally across countries, obscuring systematically important international economic forces and counter-intuitively assuming each country is equally important. We posit that the supranational aggregation approach offers a fresh perspective on numerous questions previously explored (and dismissed for lack of significance) in the international literature.

Second, we extend the data-generating framework of Dong et al. (2022), and provide both theoretical explanations and empirical evidence on why anomaly returns can be connected to market returns in different fashions in international markets, enriching the comprehension of anomaly-market linkages. We propose a decomposition that introduces a novel and straightforward way to use anomalies to evaluate international differences in market-level price efficiency across three foundational dimensions: inter-temporal, systematic importance of mispricing, relative importance of price randomness, and asymmetric mispricing correction speed. These anomaly-based efficiency measures are straightforward to compute for international markets, where high-quality earnings data have only become available in the recent decade. Our three market efficiency proxies also complement the widely-used measures from seminal papers, such as the price informativeness measure of Bai et al. (2016), the variance ratio of Lo and MacKinlay (1988), and the governance measure of Gompers et al. (2003).

Third, by examining evidence abroad, we contribute to the ongoing debate between using anomaly characteristics (Engelberg et al. (2023)), versus anomaly returns (Dong et al. (2022)). In the cross-sectional return predictability literature, the rivalry between characteristics and betas has been a central topic (see Daniel and Titman (1997) and the subsequent extensive body of research). In contrast, within the nascent field of anomaly-market linkage, the current debate centers around characteristics vs. returns. On the one hand, we offer insights into why characteristics might not perform as effectively as returns in forecasting time-series market returns, yet still hold considerable value in predicting cross-sectional returns. On the other hand, we also show that characteristics display some market return predictability through supranational aggregation.

Fourth, by investigating the anomaly-market relationship in international contexts, we offer out-of-sample evidence that addresses data mining concerns associated with findings in the U.S. market. Methodologically, we introduce a novel approach to alleviate data mining concerns in testing anomaly-market links.

II. Intuition on Predictability

In this section, we employ a stylized data-generating framework to provide intuition on how long-short anomaly portfolio returns and aggregate anomaly characteristics predict the market return.

A. Data-Generating Process

Assume that the prices for the long and short legs of a representative anomaly portfolio are exposed to a common martingale component with period t increment f_t, while the long-leg (short-leg) price contains a stationary component $u_{L,t} \leq 0$ ($u_{S,t} \geq 0$) reflecting the level of underpricing (overpricing), independent from the common martingale component. The loading for the long leg on f_t is k_L and for the short leg on f_t is k_S. $\frac{k_L}{k_S} = k$ which varies across countries. The notion of mispricing inherently implies the existence of a stationary component within the price structure, premised on the idea that any deviation due to mispricing is temporary and eventually corrected, even though the adjustment towards equilibrium may span multiple periods. The log return (in terms of price changes) in each leg is then given by

$$r_{l,t} = k_l f_t + \Delta u_{l,t} \quad \text{for } l = L, S, \tag{1}$$

where $r_{L,t}$ ($r_{S,t}$) is the long-leg (short-leg) return, $\Delta u_{l,t}$ is the change in mispricing, and $\text{cov}(f_t, \Delta u_{L,t}) = \text{cov}(f_t, \Delta u_{S,t}) = 0$. Using equation (1), the long-short anomaly portfolio return $r_{LS,t}$ can be expressed as

$$r_{LS,t} = (k_L - k_S) f_t + \Delta u_{L,t} - \Delta u_{S,t}. \tag{2}$$

For expositional ease, assume that the long and short legs together comprise the market.

$$r_{M,t} = \frac{1}{2}[(k_L + k_S) f_t + \Delta u_{L,t} + \Delta u_{S,t}], \tag{3}$$

where $r_{M,t}$ is the market return.[7]

According to the Wold representation theorem, the stationary component in each leg related to mispricing (i.e., the pricing error) can be generally expressed as

$$u_{l,t} = \sum_{j=0}^{\infty} \psi_{l,j} v_{l,t-j} \quad \text{for } l = L, S, \tag{4}$$

where $\psi_{l,0} = 1$, $v_{L,t} \leq 0$ ($v_{S,t} \geq 0$) is a serially uncorrelated underpricing (overpricing) shock, $\text{var}(v_{l,t}) \geq 0$, $\sum_{j=1}^{\infty} \psi_{l,j}^2 < \infty$ (square summability), and $\psi_{l,j} \geq 0$ for $j \geq 1$ (to ensure that $u_{L,t} \leq 0$ and $u_{S,t} \geq 0$). For simplicity, we assume that $v_{L,t}$ and $v_{S,t}$ are uncorrelated. When $\text{var}(v_{L,t}) = \text{var}(v_{S,t}) = 0$, there is no mispricing and hence the market return in equation (3) reduces to $r_{M,t} = f_t$.

Equation (4) provides a comprehensive representation for the mispricing component in each leg, as any stationary autoregressive moving-average (ARMA) process can be expressed via the infinite-order moving average(MA) process. The equation expresses the current-period level of mispricing upon both current and past mispricing shocks. The MA process can be interpreted as an impulse-response function: $\psi_{l,j}$ is the response (ceteris paribus) of $u_{l,t+j}$ for $j \geq 0$ to a period t unit mispricing shock.

Taking the first difference of equation (4), we derive a formula for the change in mispricing as follows,

$$\Delta u_{l,t} = \sum_{j=0}^{\infty} \tilde{\psi}_{l,j} v_{l,t-j} \quad \text{for } l = L, S, \tag{5}$$

[7] Following much of the cross-sectional literature, the anomaly portfolios in Section IV are based on stocks sorted into decile portfolios, and each long-short anomaly portfolio goes long (short) the tenth (first) decile portfolio.

where $\tilde{\psi}_{l,0} = \psi_{l,0} = 1$ and $\tilde{\psi}_{l,j} = \psi_{l,j} - \psi_{l,j-1}$ for $j \geq 1$. To simplify the exposition, we consider the mispricing correction assumption

$$\tilde{\psi}_{l,j} = \psi_{l,j} - \psi_{l,j-1} \leq 0 \quad \text{for } j \geq 1, \tag{6}$$

which suggests that due to sufficiently active arbitrage, the mispricing stemming from a mispricing shock at period t will not increase in any subsequent period.

B. *Mispricing Correction Persistence*

Consider a predictive regression accessing the relationship between the current long- or short-leg return of the anomaly portfolio and the next period's market return:

$$r_{M,t+1} = \alpha_l + \beta_l r_{l,t} + \varepsilon_{l,t+1} \quad \text{for } l = S, L, \tag{7}$$

where $\varepsilon_{l,t+1}$ is a zero-mean, serially uncorrelated disturbance term. Using equations (1) and (3), the standardized slope coefficient in equation (7) is given by

$$\tilde{\beta}_l = \frac{0.5\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t})}{[k_l^2 \text{var}(f_t) + \text{var}(\Delta u_{l,t})]^{0.5}} \quad \text{for } l = L, S. \tag{8}$$

Equation (8) indicates that the predictive ability of the long- or short-leg return depends on $\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t})$. Based on equation (5), the latter is given by

$$\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t}) = \left[(\psi_{l,1} - 1) + \sum_{j=1}^{\infty} (\psi_{l,j} - \psi_{l,j-1})(\psi_{l,j+1} - \psi_{l,j})\right] \text{var}(v_{l,t}) \;\; \text{for } l = L, S. \tag{9}$$

Our international empirical results generally indicate that $\tilde{\beta}_l > 0$, so that $\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t}) > 0$, especially for the short leg. This means that mispricing in a country usually take more than one month to correct, resulting in violation of both semi-strong and weak forms of market efficiency.

To understand the conditions that produce $\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t}) > 0$, we can use equation (5) to write the changes in the level of mispricing for the current and next period as

$$\Delta u_{l,t} = v_{l,t} + \sum_{j=1}^{\infty} (\psi_{l,j} - \psi_{l,j-1}) v_{l,t-j} \quad \text{for } l = L, S, \tag{10}$$

$$\Delta u_{l,t+1} = v_{l,t+1} + (\psi_{l,1} - 1) v_{l,t} + \sum_{j=2}^{\infty} (\psi_{l,j} - \psi_{l,j-1}) v_{l,t+1-j} \quad \text{for } l = L, S, \tag{11}$$

respectively. Equations (10) and (11) show how a new (i.e., period t) pricing shock affects the current and future changes in mispricing in opposite directions. Take an overpricing shock ($\upsilon_{S,t} > 0$) as an example. According to equation (10), such a shock at $\upsilon_{S,t}$ amplifies the current overpricing, resulting in a positive change in overpricing for the period t. Subsequently, the overpricing initiated by the period t shock is corrected in proportion to $\psi_{S,1} - 1 \leq 0$ in period $t+1$, corresponding to a nonpositive change in overpricing, as indicated by (11). Thus, the consecutive changes in mispricing due to a new overpricing shock can generate negative serial dependence in the short-leg return. In contrast, old pricing shocks ($\upsilon_{S,t-j}$ for $j \geq 1$) can produce positive serial dependence in the short-leg return. According to equations (10) and (11), the changes in overpricing corresponding to these old shocks are nonpositive in consecutive periods; for example, the overpricing associated with $\upsilon_{S,t-1}$ is corrected in proportion to $\psi_{S,1} - 1 \leq 0$ in t and $\psi_{S,2} - \psi_{S,1} \leq 0$ in $t+1$.

To sum up, the autocovariance between consecutive changes in mispricing, $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})$, is determined by two opposing effects: (i) the extent to which the overpricing associated with a new shock is corrected in the next period, and (ii) the extent to which the overpricing associated with old overpricing shocks is corrected in the current and next periods. These two effects are evident in the expression in brackets in equation (9), which accounts for all of the consecutive pairs of return responses to new and old overpricing shocks. The first term, $\psi_{S,1} - 1$, is the product of the period t and period $t+1$ return responses to a new unit overpricing shock, which captures the potential reversal due to the immediate correction of the overpricing induced by the new shock. The second term, $\sum_{j=1}^{\infty}(\psi_{S,j} - \psi_{S,j-1})(\psi_{S,j+1} - \psi_{S,j}) \geq 0$, reflects the potential momentum due to the persistent correction of the overpricing induced by old shocks. For example, $(\psi_{S,1} - 1)(\psi_{S,2} - \psi_{S,1})$ is the product of the period t and period $t+1$ overpricing corrections corresponding to a period $t-1$ unit overpricing shock, $(\psi_{S,2} - \psi_{S,1})(\psi_{S,3} - \psi_{S,2})$ is the product of the period t and period $t+1$ overpricing corrections corresponding to a period $t-2$ unit overpricing shock, and so forth.[8]

In order to have $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) > 0$, the return momentum generated by the correction of the overpricing induced by old shocks needs to outweigh the magnitude of the return reversal generated by the immediate correction of the overpricing induced by a new shock. When the momentum effect dominates the return reversal effect, MCP is sufficiently strong that $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) > 0$ in equation (9) and $\tilde{\beta}_S > 0$ in equation

[8]As $\psi_{S,1}$ decreases, the degree of overpricing correction in the period immediately after the shock increases, so that the magnitude of the reversal effect increases. In the extreme, $\psi_{S,1} = 0$, which implies that $\psi_{S,j} = 0$ for $j \geq 2$, so that the overpricing shock fully corrects in one period and $\psi_{S,1} - 1 = -1$. In this case, the second term in brackets is zero and $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) = -\text{var}(\upsilon_{S,t})$ in equation (9).

(8). This condition can be expressed as follows,

$$\sum_{j=1}^{\infty}(\psi_{S,j}-\psi_{S,j-1})(\psi_{S,j+1}-\psi_{S,j}) > -(\psi_{S,1}-1). \tag{12}$$

An analogous condition to equation (12) applies to $\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t})$ and $\tilde{\beta}_L$. Equation (12) is consistent with various theoretical models that can explain strong MCP.[9]

Next, consider a predictive regression based on the long-short anomaly portfolio return:

$$r_{M,t+1} = \alpha_{LS} + \beta_{LS} r_{LS,t} + \varepsilon_{LS,t+1}. \tag{13}$$

The standardized slope coefficient in equation (13) is given by

$$\tilde{\beta}_{LS} = \frac{0.5[\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) - \text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})]}{[(k_L - k_S)^2 \text{var}(f_t) + \text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t})]^{0.5}}. \tag{14}$$

Empirically, we find that for $\tilde{\beta}_{LS} < 0$, which holds when

$$\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) > \text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}), \tag{15}$$

in other words, when the MCP for overpricing exceeds underpricing. Our finding that long-short anomaly portfolio returns negatively predict the market return using the global data is consistent with the U.S. evidence in Dong et al. (2022). It is also in line with the relative importance of overpricing in the U.S. market where Hong, Lim, and Stein (2000), Stambaugh, Yu, and Yuan (2012, 2015), and Avramov et al. (2013) find that the short legs of anomaly portfolios are largely responsible for the profitability of long-short anomaly portfolio returns.

C. *Noise Reduction*

In our setup, the long-short anomaly portfolio return, which is designed to be "market neutral", can produce a better predictive signal for predicting the market return than the short-leg return To see this, for simplicity, we assume in this and the next subsection that $\tilde{\beta}_L = 0$, which is in line with the relatively weak predictive ability of long-leg returns. In

[9] Andrei and Cujean (2017) show that when information about mispricing spreads among traders at an accelerated rate, immediate correction of a current pricing error is dominated by the corrections of previous pricing errors, resulting in return momentum. From a behavioral perspective, Chan, Jegadeesh, and Lakonishok (1996), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999), and Da, Gurun, and Warachka (2014) explain return momentum as underreaction to news. Gârleanu and Pedersen (2013, 2016) and Dong, Kang, and Peress (2020) show that various considerations lead arbitrageurs to slowly allocate capital to correct mispricing.

this case, $\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) = 0$, so equation (14) becomes

$$\tilde{\beta}_{LS}\big|_{\tilde{\beta}_L=0} = \frac{-0.5\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{[(k_L - k_S)^2\text{var}(f_t) + \text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t})]^{0.5}}. \tag{16}$$

The magnitude of $\tilde{\beta}_{LS}$ in equation (16) is greater than that of $\tilde{\beta}_S$ in equation (8) when

$$\text{var}(r_{LS,t}) < \text{var}(r_{S,t}), \text{ or equivalently, } k_S^2\text{var}(f_t) > \text{var}(\Delta u_{L,t}) + (k_L - k_S)^2\text{var}(f_t). \tag{17}$$

Intuitively, when k is close to 1 ($k_L - k_S$ is close to 0), using $r_{L,t} - r_{S,t}$ in lieu of $r_{S,t}$ as the predictor in the predictive regression nearly removes the common unpredictable component f_t, thereby filtering noise from the predictor to provide a sharper predictive signal for the market return. Thus, the natural design of "market neutral" actually improves predictability by mitigating the unpredictable component f_t. Consistent with this intuition, we empirically find that $\text{var}(r_{LS,t})$ is considerably smaller than $\text{var}(r_{S,t})$.

D. Predictability Decomposition

Global data enables us to identify the variations in predictability across different markets, with empirical evidence indicating superior predictive performance of long-short anomaly returns in developed markets compared to emerging ones. To better understand the driving force behind the diversity of international market excess return predictability, we decompose the predictive coefficients derived above. We start with the negative ratio of the standardized slope coefficient of the representative long-short anomaly return to the standardized slope coefficient of its short-leg return based on (8) and (14). To expedite exposition, we set $k_S = 1$:

$$-\frac{\tilde{\beta}_{LS}}{\tilde{\beta}_S} = -\frac{\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) - \text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}\left[\frac{\text{var}(f_t) + \text{var}(\Delta u_{S,t})}{(k-1)^2\text{var}(f_t) + \text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t})}\right]^{0.5} \tag{18}$$

We define

$$\textit{autocovariance ratio} = \frac{\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) - \text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})} \tag{19}$$

and

$$\textit{variance ratio} = \frac{\text{var}(f_t) + \text{var}(\Delta u_{S,t})}{(k-1)^2\text{var}(f_t) + \text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t})} \tag{20}$$

Empirically, the "autocovariance ratio" can be easily calculated as the difference between the auto-covariance of the long leg return and that of the short leg return, divided by the auto-covariance of the short leg return. Accordingly, the "variance ratio" can also be easily calculated as the ratio of the variance of the short leg to that of the long-short leg. In other words, we can measure them as follows using information from anomaly portfolio return,

$$Autocovariance\ ratio = \frac{ACOV(L) - ACOV(S)}{ACOV(S)} \tag{21}$$

$$Variance\ ratio = \frac{VAR(S)}{VAR(LS)} \tag{22}$$

where $ACOV(L)$ ($ACOV(S)$) stands for the auto-covariance of the long (short) leg anomaly return series and the $VAR(S)$ ($ACOV(LS)$) stands for the variance of the short-leg (long-short) anomaly return series. [10]

Taking one step further, the autocovariance ratio can indicate asymmetric mispricing correction persistence between the long and the short leg. For example, when there are reasons like short-selling constraints that impede the pace of correction on the short-leg overpricing, we shall expect more persistent correction in the short leg (reflected in a high $ACOV(S)$) when compared to the long leg. Thus, a more negative autocovariance ratio suggests more asymmetry in the correction persistence between the long-leg underpricing and the short-leg overpricing.

Moreover, we can rewrite equation (20) as

$$variance\ ratio = \frac{\frac{\mathrm{var}(f_t)}{\mathrm{var}(\Delta u_{S,t})} + 1}{(k-1)^2 \frac{\mathrm{var}(f_t)}{\mathrm{var}(\Delta u_{S,t})} + \frac{\mathrm{var}(\Delta u_{L,t})}{\mathrm{var}(\Delta u_{S,t})} + 1}, \quad \text{where} \quad k = \frac{k_L}{k_S} \tag{23}$$

With $\mathrm{var}(\Delta u_{L,t})$ and $\mathrm{var}(\Delta u_{S,t})$ being comparable and k approximates to 1, the variance ratio can be a good proxy of the ratio $\frac{\mathrm{var}(f_t)}{\mathrm{var}(\Delta u_{S,t})}$. The ratio $\frac{\mathrm{var}(f_t)}{\mathrm{var}(\Delta u_{S,t})}$ is an indicator of the relative importance of price randomness, as higher $\mathrm{var}(f_t)$ and lower $\mathrm{var}(\Delta u_{S,t})$

[10] The anomaly we refer to in this data-generating process is a representative anomaly that captures the systematic mispricing correction of a particular market. We propose that one should first aggregate the information from a group of anomaly returns (e.g., taking the average of the group of anomaly returns) to reduce the noise in individual anomalies. Then use the aggregated return to calculate the autocovariance and variance ratio in the equation above.

means pricing variability comes more from the unpredictable component f_t instead of the mispricing-related $\Delta u_{S,t}$.

To sum up, we can rewrite equation (18) as

$$-\frac{\tilde{\beta}_{LS}}{\tilde{\beta}_S} = -Autocovariance\ ratio * (Variance\ ratio)^{0.5} \tag{24}$$

The autocovariance ratio is indicative of asymmetric mispricing correction persistence between long and short-leg returns. The variance ratio serves as an indicator of price randomness, where there's higher variability in the unpredictable martingale component, as opposed to the mispricing-related component.

After moving the $\tilde{\beta}_S$ to the right-hand side and taking the logarithm version of (24), we have[11]:

$$log(-\tilde{\beta}_{LS}) = log(\tilde{\beta}_S) + log(-autocovariance\ ratio) + log(sqrt\_variance\ ratio) \tag{25}$$

Therefore, we posit that the efficacy of the long-short anomaly is maximized under conditions when (1) there's a high value of $\tilde{\beta}_S$, representing a high degree of systematic importance of overpricing, (2) there's significant asymmetry in the MCP between the long and the short leg, fostering a "return momentum effect" for prediction, and (3) "long-minus-short" effectively mitigates the influence of the large unpredictable martingale component.

So far, we successfully decoded the relationship between long-short anomaly portfolio returns and the market excess return, as well as explained the heterogeneity of the predictive power across different markets. More importantly, through the comparison in the predictive power of market excess return, we provide a novel and straightforward way to extract market efficiency information along three dimensions.

E. Aggregate Anomaly Characteristics

Now let's switch to the anomaly characteristics and explore the way they forecast the market excess return. Suppose that the characteristic $c_{l,t}$ is related to the mispricing component in each leg as follows:

$$c_{l,t} = \phi_l + u_{l,t} + \eta_{l,t} \quad \text{for } l = L, S. \tag{26}$$

[11]Based on our data generating process, the signs of $-\tilde{\beta}_{LS}$ and *autocovariance ratio* are both negative. Thus, we put a negative sign in front of them before taking the logarithm transition.

Here, ϕ_l is the unconditional property of the long- and short-leg stocks. For example, the long-leg stocks have a higher beta than the short-leg stocks. $u_{l,t}$ is the time-varying overpricing or underpricing (For simplicity, we set the coefficient on $u_{l,t}$ in Equation (26) to unity.). ϕ_l and $u_{l,t}$ both help identify the stocks with high and low returns in the cross-section, supporting the usefulness of sorting based on characteristics for predicting cross-sectional returns. $\eta_{l,t}$ is the component that is unrelated to mispricing. It can be related to fundamentals that are immediately priced in, thus connecting to the martingale component f_t in the return process outlined in equation (1). More broadly, $\eta_{l,t}$ can also relate to some characteristic-specific attributes and be serially correlated. For example, given that many characteristics are observed at frequencies lower than a month, $\eta_{l,t}$ can reflect the persistent stale information that is already priced in. Aggregating the characteristic based on Equation (26), we have

$$c_t = 0.5(c_{L,t} + c_{S,t}) = 0.5(\phi_L + u_{L,t} + \eta_{L,t} + \phi_S + u_{S,t} + \eta_{S,t}). \tag{27}$$

Because many characteristics are fairly persistent or nearly nonstationary, they are unfit for time-series predictive regression (Engelberg et al. (2023)). We de-trend the series by taking the first difference of Equation (27):

$$\Delta c_t = 0.5(\Delta u_{L,t} + \Delta u_{S,t} + \Delta\eta_{L,t} + \Delta\eta_{S,t}). \tag{28}$$

For the new DGP, the standardized slope coefficient for a predictive regression relating the difference of the aggregate characteristic in Equation (28) to next period's market return can be rewritten as

$$\tilde{\beta}_{\Delta c} = \frac{0.5[\mathrm{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) + \mathrm{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})]}{[\mathrm{var}(\Delta u_{L,t}) + \mathrm{var}(\Delta u_{S,t}) + \mathrm{var}(\Delta\eta_{L,t}) + \mathrm{var}(\Delta\eta_{S,t})]^{0.5}}. \tag{29}$$

In line with the relatively weak predictive ability of long-leg returns, as in Section II.C, we assume that $\mathrm{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) = \tilde{\beta}_L = 0$. Equation (29) then becomes

$$\tilde{\beta}_{\Delta c}\big|_{\tilde{\beta}_L=0} = \frac{0.5\mathrm{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{[\mathrm{var}(\Delta u_{L,t}) + \mathrm{var}(\Delta u_{S,t}) + \mathrm{var}(\Delta\eta_{L,t}) + \mathrm{var}(\Delta\eta_{S,t})]^{0.5}}. \tag{30}$$

Comparing the standardized slope coefficient in Equation (30) to that in Equation (16) that uses the long-short return as the predictor, the presence of $\mathrm{var}(\Delta\eta_{L,t})$ and $\mathrm{var}(\Delta\eta_{S,t})$ in the denominator of Equation (30) adds noise to the predictive signal, which dilutes the predictive signal in Equation (30) compared to that in Equation (16). As a result, the long-short return is a less noisy predictor of the next period's market return than the aggregate characteristic.

Furthermore, the persistence of the uninformative component of the characteristic can be attributed to the difference in the nature of data between characteristics and returns. For example, many characteristics update less frequently than returns or rely on overlapping data, such as in the case of volatility- and beta-based anomalies. Consequently, a significant portion of the monthly change in the characteristic could stem from some redundant overlapping information that is already priced in. Indeed, we verify that most of the characteristics are still persistent even after first differencing, suggesting that $\Delta\eta_{L,t}$ and $\Delta\eta_{S,t}$ could be positively autocorrelated. For simplicity, let's assume that they follow an AR1 process. Under this assumption, $\text{var}(\Delta\eta_{L,t})$ and $\text{var}(\Delta\eta_{S,t})$ increases with the degree of autocorrelation in $\Delta\eta_{L,t}$ and $\Delta\eta_{S,t}$. Intuitively, the more persistent the uninformative component of a characteristic is, the greater the noise it introduces. This is because more persistent noise revert less, thereby resulting in a bigger variance.

Finally, our data-generating process also provides one reason why characteristics can still be effective in predicting the cross-sectional returns, even with the presence of timing-varying noise from $\eta_{l,t}$. Indeed, the stable time-invariant information ϕ_l and the time-varying information $u_{l,t}$ both provide information about the cross-section of returns. However, the former is not useful for predicting the time-varying return. As a result, akin to the firm fixed effect, the time-invariant (or, loosely speaking, the highly persistent) informative component about returns can work very well in predicting the cross-sectional difference in returns but not the month-to-month time variation in returns.

III. Methodology

This section describes the construction of forecasts and their evaluation using both statistical and economic criteria. We start with introducing the out-of-sample forecast, following Dong et al. (2022). Then we cover in-sample forecast as well to increase the sample size and power. All the following descriptions are based on the case of predicting the market excess return of a particular market (either a supranation or a country) using predictors from the same market.

A. *Forecast Construction*

Assume that we have multiple predictors (in our context, 100 long-short anomaly portfolio returns) to forecast the market excess return ($r_{M,t}$). To generate $\hat{r}_{M,t+1|t}$, a forecast of the month $t+1$ market excess return, we use all information available through month t. All of our market excess return forecasts are out of sample, as we only use data available through month t to forecast $r_{M,t+1}$.

When it comes to the forecast of market excess return, the most popular benchmark in the literature so far is the prevailing mean forecast, which implicitly assumes that the market excess return is unpredictable (apart from its mean value). The prevailing mean forecast is simply the average of the market excess return observations available at the time of forecast formation. The prevailing mean forecast is difficult to beat in practice (e.g., Goyal and Welch (2008)) due to the inherently small predictable component in the monthly market excess return (i.e., return data are quite noisy). Thus, successful out-of-sample strategies effectively shrink the market excess return forecast toward the prevailing mean benchmark to reduce the likelihood of overreacting to the noise in return data.

We compare the prevailing mean benchmark to the six forecasts summarized below, each of which incorporates the information in the group of predictors.

Predictor Average An alternative strategy for guarding against overfitting is to first combine the predictors themselves into a small number of variables and then use the reduced set of variables as predictors in a low-dimensional predictive regression. Intuitively, as discussed in Section II, we consolidate the predictors to filter the noise in the individual predictors. The predictor average forecast is based on OLS estimation of a univariate predictive regression in which the lagged cross-sectional average of the predictors serves as the explanatory variable.

Principal Component We can also combine the predictors by extracting the first principal component from the set of predictors. The principal component forecast uses the lagged principal component as the explanatory variable in a univariate predictive regression estimated via OLS.

PLS The first principal component explains as much variation as possible in the predictors themselves. However, from a forecasting standpoint, we are interested in explaining the target variable. Instead of extracting a factor that explains as much of the variation in the predictors as possible, Kelly and Pruitt (2013, 2015) develop a three-pass regression filter to construct a target-relevant factor from a set of predictors that is maximally correlated with the target variable. The lagged target-relevant factor then serves as the explanatory variable in a univariate predictive regression estimated via OLS. The three-pass regression filter is essentially a version of PLS.

Combination ENet When the number of predictors is large, the simple combination forecast can be too conservative in the sense that it "overshrinks" the forecast toward the prevailing mean, thereby neglecting too much of the relevant information in the predictor variables. Using insights from Diebold and Shin (2019), Rapach and Zhou (2020) and Han et al. (2021) employ the ENet to refine the simple combination forecast. Instead of averaging across all of the individual univariate predictive regression forecasts, the combination ENet (C-ENet) forecast takes the average of the individual forecasts selected by the ENet in a Granger and Ramanathan (1984) multiple regression relating the actual market excess return to the individual univariate forecasts.

ENet The ENet forecast is based on the multiple predictive regression fitted via the ENet instead of OLS. The ENet (Zou and Hastie (2005)) relies on penalized regression to guard against overfitting. The ENet penalty term includes both ℓ_1 (LASSO) and ℓ_2 (ridge; Hoerl and Kennard (1970)) components. The ℓ_1 component permits shrinkage to zero, so that the ENet performs variable selection. Based on Flynn, Hurvich, and Simonoff (2013), we use the Hurvich and Tsai (1989) corrected version of the Akaike (1973) information criterion to select the value of the regularization parameter governing the degree of shrinkage. The ENet directly addresses overfitting by shrinking the slope coefficients of the fitted model, which results in shrinking the forecast toward the prevailing mean benchmark.

Simple Combination Instead of OLS estimation of the multiple predictive regression, forecast combination begins by computing a set of forecasts based on OLS estimation of univariate predictive regressions that include each lagged predictor (in turn). The simple combination forecast is the arithmetic mean of the individual univariate forecasts. Rapach, Strauss, and Zhou (2010) show that the simple combination forecast exerts a strong shrinkage effect.

B. *Forecast Evaluation—Statistical Accuracy*

To measure the market excess return forecasts in terms of statistical accuracy, we calculate MSFE. Denote the errors for the prevailing mean benchmark and a competing forecast by

$$\hat{e}_{0,t|t-1} = r_{M,t} - \hat{r}^{\mathrm{PM}}_{M,t|t-1}, \tag{31}$$

$$\hat{e}_{1,t|t-1} = r_{M,t} - \hat{r}_{M,t|t-1}, \tag{32}$$

respectively, where $\hat{r}^{\mathrm{PM}}_{M,t|t-1}$ is the prevailing mean benchmark forecast and $\hat{r}_{M,t|t-1}$ generically denotes a competing forecast. The sample MSFE is given by

$$\widehat{\mathrm{MSFE}}_j = \frac{1}{T}\sum_{t=1}^{T} \hat{e}^2_{j,t|t-1} \quad \text{for } j = 0, 1, \tag{33}$$

where T is the number of out-of-sample observations. Following the Clark and West (2007) procedure, we test for a difference in the population MSFEs, which can be conveniently implemented in a simple regression framework:

$$\underbrace{d_{t|t-1} + \left(\hat{r}^{\mathrm{PM}}_{M,t|t-1} - \hat{r}_{M,t|t-1}\right)^2}_{f_{t|t-1}} = \mu + \varepsilon_t, \tag{34}$$

where $\hat{d}_{t|t-1} = \hat{e}^2_{0,t|t-1} - \hat{e}^2_{1,t|t-1}$ is the period t loss differential. The t-statistic corresponding to the OLS estimate of μ in equation (34) is used to test

$$H_0\colon \text{MSFE}_0 \leq \text{MSFE}_1 \ (\mu \leq 0) \text{ versus } H_A\colon \text{MSFE}_0 > \text{MSFE}_1 \ (\mu > 0), \tag{35}$$

where MSFE_j is the population MSFE for $j = 0, 1$.[12] The t-statistic is computed using a heteroskedasticity- and autocorrelation-consistent (HAC) standard error (Newey and West (1987)).

It is common to report the Campbell and Thompson (2008) R^2_{OS} statistic when comparing MSFEs for the prevailing mean benchmark and a competing market excess return forecast:

$$R^2_{\text{OS}} = 1 - \frac{\widehat{\text{MSFE}}_1}{\widehat{\text{MSFE}}_0}. \tag{36}$$

Equation (36) gives the proportional reduction in the sample MSFE for the competing forecast vis-à-vis the prevailing mean benchmark. Using the Clark and West (2007) statistic to test equatiion (35) is tantamount to testing $H_0\colon R^2_{\text{OS}} \leq 0$ against $H_A\colon R^2_{\text{OS}} > 0$ (in population). Because the predictable component in the monthly market excess return is necessarily limited, the R^2_{OS} statistic will be small. Nevertheless, based on the market Sharpe ratio, Campbell and Thompson (2008) suggest that a monthly R^2_{OS} statistic as small as 0.5% can signal economic significance. As described in Section III.B, we also assess the economic significance of market return forecasts more directly by measuring their economic value to an investor.

In addition, we examine whether out-of-sample return predictability (as measured by the R^2_{OS} statistic) is related to market frictions. To the extent that greater frictions exacerbate limits of arbitrage, we expect anomaly portfolio returns to generate stronger out-of-sample gains during high-friction periods. To test whether out-of-sample return predictability changes with the state of market frictions, we augment the Clark and West (2007) framework in equation (34) as follows:

$$f_{t|t-1} = \mu + \xi I_t + \varepsilon_t, \tag{37}$$

[12] The well-known Diebold and Mariano (1995) and West (1996) (DMW) procedure uses $d_{t|t-1}$ instead of $f_{t|t-1}$ as the dependent variable in equation (34). Clark and McCracken (2001) and McCracken (2007) show that the DMW test tends to be severely undersized when comparing forecasts from nested models (as in our application), in which case it has little power to detect improvements in forecast accuracy. Clark and West (2007) adjust the DMW test statistic so that its asymptotic distribution is well approximated by the standard normal.

where I_t is an indicator variable that equals one (zero) if market frictions are high (low). We use the t-statistic corresponding to the OLS estimate of ξ in equation (37) to test

$$H_0\colon\ R^2_{\text{OS,high}} \leq R^2_{\text{OS,low}}\ (\xi \leq 0) \text{ versus } H_A\colon\ R^2_{\text{OS,high}} > R^2_{\text{OS,low}}\ (\xi > 0), \tag{38}$$

where $R^2_{\text{OS,high}}$ ($R^2_{\text{OS,low}}$) is the value of the R^2_{OS} statistic during periods of high (low) market frictions. We again compute the t-statistic using a HAC standard error.

C. *Forecast Evaluation—In-sample Insights*

While out-of-sample testing is often regarded as more stringent, it might not be the most reliable method for assessing the predictive capability of anomaly returns on an international scale. There is a trade-off between the in-sample tests and out-of-sample tests of predictability, as out-of-sample analysis is based on sample-splitting that involves a loss of information and hence lower power in small samples (Inoue and Kilian (2005)). Given that the average sample period for country-level analysis in our international prediction tests spans only 354 months, it is necessary to consider in-sample prediction capabilities as well. This approach is crucial that it maximizes the use of available data to provide more comprehensive insights into the predictive strength of anomaly returns.

IV. Anomaly Construction

We construct 100 anomalies, hereafter referred to as DLRZ100, based on the works of Dong et al. (2022) and substitute anomalies that lack international data.[13] These anomalies are derived using data from CRSP, Compustat NA, and Compustat Global, and are cross-verified with major alternative databases including Datastream, Worldscope, and IBES Global. DLRZ100 covers a variety of categories, including but not limited to value versus growth, profitability, investment, issuance activity, momentum, and trading frictions. Anomalies that are constructed as interactions of two distinct signals (as these anomalies fundamentally rely on multiple anomalies) or indicator variables (such as IPO) are not incorporated into DLRZ100.

We first construct DLRZ100 at the country level. Our sample is restricted to common stocks traded on the main exchanges of their respective countries, with market equity higher than the 20^{th} percentile of NYSE. Additionally, We require these stocks to be identified as primary stocks of the underlying company according to Compustat. The assignment of stocks to different countries is based on the countries of the exchanges

[13] We substitute the 9 anomalies due to international data availability. All of our results hold if we only use the 91 anomalies originated from Dong et al. (2022). We also use a set of 153 anomalies from Jensen et al. (2023), hereafter JKP153, for robustness check in the appendix.

where they are listed. For each anomaly of each country, we sort the stocks into decile groups based on the related characteristics. Subsequently, we form the decile portfolio returns using a value-weighting approach. The decile portfolios are ranked by relative returns. In constructing the long-short anomaly returns, we long the tenth decile portfolio return and short the first decile portfolio.[14] To compute the market excess return for a specific country, we value-weight the monthly excess returns of all primary common stocks, whose underlying firms are incorporated in that country, from the country's main exchanges.[15]

Next, we aggregate each country-level anomaly return, as well as market excess return, into the supranational level, incorporating all available countries in a given supranation. The weight of each country is determined by its total market capitalization, thereby assigning greater weights to countries that hold systematic importance. We begin with the following supranations abroad: G6, G19, and World. The G6(G19) supranation comprises the countries of the G7(G20), excluding the United States. The World supranation covers all non-US countries with sufficient data. The sample period spans 1986:01 to 2021:12. Table I reports summary statistics at the supranational level. We average the supranational market capitalization and the supranational market excess return across 420 months, from January 1986 to December 2021. The supranational long-short anomaly returns, are averaged first across DLRZ100 anomalies, then over time. The world supranation incorporates all countries selected in the sample, and therefore the highest market capitalization. All supranations have positive long-short anomaly portfolio returns on average.

V. Supranational vs. Country-level Analysis

In this section, we delve into the performance of anomalies as well as their ability to predict the market excess return, examining the performance at both aggregated supranational level and conventional country level.

A. Anomaly Replication

We start with evaluating the performance of the DLRZ100 anomalies at both the country and supranational levels in Table II in terms of their raw returns, CAPM alphas, and Fama-French three-factor alphas. (We also calculate the replication rate for the JKP153 anomaly set in Table A4 for robustness.) The ratios of long-short anomaly

[14] We obtain the direction of long vs short of each anomaly from the paper where it is initially discovered to avoid look-ahead bias.

[15] The way we construct the country-level market excess return follows the same logic as Fama and French (1993).

Table I
Summary Statistics

This table reports the summary statistics for market capitalization, market excess return, and monthly portfolio returns of DLRZ100 at the supranational level, from 1986:01 to 2021:12. Both the supranational-level anomaly returns and market excess returns are constructed as market-cap-weighted returns across all available countries within the supranation. Row 2 shows the number of countries in each supranation. Row 3 shows the number of months in each supranational sample. Row 4 shows the supranational market capitalization (in billions) averaged across time. Row 5 shows the supranational market excess return averaged across time. Row 6 shows the supranational anomaly return, averaged first across anomalies, then over time.

supranation	G6	G19	World
Number of countries	6	18	44
N	432	432	432
Average market capitalization	9967	15316	20327
Average market excess return	0.45%	0.50%	0.52%
Average long-short anomaly return	0.21%	0.25%	0.24%

returns (and CAPM alpha) that satisfy the 10% significance level at the country level are computed and averaged across all available countries within each supranation. [16]. All replication rates experience a substantial increase when the country-level anomaly portfolio returns are aggregated to the supranational level. Notably, as we aggregate all available countries together to form the "World" supranation, the number of replicable anomalies almost doubles when aggregated to the supranational level(increasing from 17.49% to 40.00% for raw returns, 23.20% to 44.00% for CAPM alphas, and 24.57% to 46.00% for three-factor alphas). The substantial increase is not driven by the few big countries, as the replication rate of market-cap-weighted country level replication rate is still much less than the supranational level replication rate. Even for the G6 supranation, which already comprises the most systematically significant countries overseas, supranational aggregation enhances anomaly performance (rising from 21.36% to 33.00% for raw returns, 34.55% to 39.00% for CAPM alphas, and 24.57% to 46.00% for three-factor alphas 35.55% to 37.00%). Overall, our results advocate for the study of international anomalies at the supranational level, where noise from anomalies of each country is reduced and systematically important mispricing is captured at the more macro level. This perspective may challenge the prevailing view in prior literature, which typically presents evidence for anomalies at the level of individual countries.

[16]Since the international time-series data for anomaly returns are short, we propose that a 10% significance level is an adequate criterion for assessing predictability.

Table II

Replication Rate at the Country vs. Supranational Level

This table presents the percentage of anomalies that can be replicated at a 90% confidence interval at both the country level and supranation level based on the DLRZ100 anomaly pool from 1986:01 to 2021:12. The country level replication rate is averaged across all available countries within each supranation. The replication rates are computed using both raw long-short anomaly returns (Panel A), their CAPM alpha (Panel B), and their Fama-French three-factor alpha (Panel C). For each anomaly, the supranational-level anomaly return is constructed as the market-cap-weighted average of all available country-level anomaly returns within that supranation. The replication rates at the supranational level are then calculated using the supranational-level anomaly return (Columns 3).

	Country level	Supranational level
Panel A: Raw Return		
G6	21.36%	33.00%
G19	19.38%	38.00%
World	17.49%	40.00%
Panel B: CAPM alpha		
G6	34.55%	39.00%
G19	26.34%	42.00%
World	23.20%	44.00%
Panel C: FF3 alpha		
G6	35.55%	37.00%
G19	30.58%	39.00%
World	24.57%	46.00%

B. Market Excess Return Forecast

Next, we use long-short anomaly returns to forecast market excess return, connecting cross-sectional predictability with time-series market return predictability internationally. We use the first 10 years (1986:01 to 1995:12) of the full sample period as the initial in-sample estimation period. The following five years (1996:01 to 2000:12) serve as the initial holdout out-of-sample period for computing the C-ENet forecast so that the rest of the samples (2001:01 to 2021:12) constitute the out-of-sample period for forecast evaluation. Because the methods in Section III.A require non-missing predictor data (except for the predictor average), for an anomaly with a missing return in a given month, we fill in the missing return with the cross-sectional average for the available anomaly returns in that month. Due to the short sample period of the international data, we conducted the in-sample forecasts as well in the same setting and all the results hold.

Table III reports R^2_{OS} statistics for monthly market excess return forecasts based on the 100 long-short anomaly portfolio returns at both supranational and country levels using the set of countries in G6, G19, and World. For supranational prediction, both the anomaly returns and market excess returns are constructed as market-cap-weighted returns across all available countries, where the aggregated supranational anomaly portfolio returns are used to predict its supranational market excess return. For country-level prediction, each country's anomaly returns are used to predict its own market return first. Then the country level R^2_{OS} is averaged across all available countries within each supranation. In sharp contrast, the supranational forecasts outperform their country-level counterparts significantly. All the supranational forecasts using the six shrinkage methods generate sizable R^2_{OS} greater than the Campbell and Thompson (2008) threshold of 0.5%, indicating sufficient economic significance and improved accuracy than the prevailing mean benchmark; all of them are statistically significant as well according to the Clark and West (2007) statistics. Take the the predictor average method as an example, the R^2_{OS} are economically and statistically significant in all supranations, with considerable R^2_{OS} of 5.58%, 5.09%, and 5.13%, respectively. Conversely, without the supranational aggregation, the country-level forecasts on average only show significance in the G6, which consists of the top systematically important countries. Take the best-performing method "Avg" as an example, the equal-weighted country level R^2_{OS} is only economically and moderately statistically significant in countries of G6 (1.40%) and not even economically significant in countries of G19 (-0.75%) and World (-0.39%). [17] Beyond the individual country level, supranational aggregation captures the crucial cross-country, inter-temporal prediction power. This phenomenon occurs when the correction for certain types of mispricing happens in some countries before others. In addition to the one-month horizon forecast, we also perform out-of-sample predictions on multiperiod market excess return, $r^{(h)}_{M,t} = (1/h)\sum_{j=1}^{h} r_{M,t+(j-1)}$. The results can be found in Table A2 in Appendix. Moreover, we also checked our results by using JKP153 anomaly set (Table A4), and building Dong et al. (2022)'s anomalies in Datastream and Worldscope database (Table A3). The comprehensive set of JKP153 aims to encompass all anomalies documented in cross-sectional literature. Our findings confirm that this "All Anomalies" collection also forecasts international market returns, albeit with weaker statistical significance. The results suggest that this "All Anomalies" set is effective but less precise in capturing the

[17] We also calculate the market-capitalization-weighted average R^2_{OS} across all available countries within each supranation. For example, using the "Avg" method, the R^2_{OS} values for G6, G19, and the World are 2.31%, 0.47% and 0.67%, respectively, with only the R^2_{OS} for G6 being statistically significant. This indicates that the outperformance at the supranational level is not merely driven by "big-country" effects. Furthermore, the better performance of the market-cap-weighted average compared to the equal-weighted average suggests that larger countries tend to capture systematically important mispricing, which can also help to predict smaller countries after supranational aggregation. This cross-country effect is not as apparent when using a simple equal-weighted or even a market-cap-weighted approach.

systematically important mispricing concept in the theoretical framework of Dong et al. (2022). For further details, see Dong et al. (2024).

One distinction between cross-sectional and time-series predictability lies in the nature of the data utilized: the former derives from a return-predictor relationship extracted from thousands of stock observations across the cross-section, whereas the latter relies on a few hundred monthly observations over time. This distinction could result in reduced testing power for time-series analyses, especially since international time-series data often start much later than U.S. data, typically around 1990, resulting in much shorter samples. To address concerns about the short sample size, we conduct in-sample prediction as shown in Table A1, to enhance testing power. By applying the Predictor Average method to aggregate the DLRZ100 anomalies, we observe strong predictability of supranational long-short anomaly returns in the in-sample tests. Unlike out-of-sample (OOS) testing, which estimates parameters using a subset of the full data in each recursive regression, in-sample (IS) testing utilizes all available data in the full sample, offering increased power in smaller samples (Inoue and Kilian (2005)).

Table III

Supranational vs Country-Level Market Excess Return Forecast

This table reports Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percentage for market excess return forecast using DLRZ100 long-short anomaly portfolio returns. The out-of-sample period covers 2001:01 to 2021:12. For supranational prediction, both the anomaly returns and market excess returns are constructed as market-cap-weighted returns across all available countries selected to form the supranation. For country-level prediction, the average R^2_{OS} is calculated for market excess return forecasts using the same set of countries. Avg is a univariate predictive regression forecast based on the cross-sectional average of the 100 long-short anomaly portfolio returns. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the 100 long-short anomaly portfolio returns. C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. The ENet forecast is based on elastic net estimation of a multiple predictive regression that includes all 100 of the long-short anomaly portfolio returns. Combine is the arithmetic mean of univariate predictive regression forecasts based on the 100 individual long-short anomaly portfolio returns (in turn). Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics.

	Supranational						Country Level					
	Avg	PC	PLS	C-Enet	Enet	Combine	Avg	PC	PLS	C-Enet	Enet	Combine
G6	5.58***	2.72**	6.92***	4.23**	2.15**	0.91**	1.4*	1.03	-1.16	-0.77	-1.67	0.46
G19	5.09***	2.45**	7.46***	3.46**	3.8**	0.79**	-0.75	-1.54	-5.12	-0.71	-2.65	0.11
World	5.13***	2.49*	7.22***	1.89**	4.16**	0.79**	-0.39	-0.65	-5.31	-0.53	-2.58	0.07

Overall, we successfully extend the linkage between the cross-sectional return predictability and the time-series predictability abroad at the supranational level, instead of the traditional country level. supranational aggregation helps to capture the intertemporal systematically important mispricing at the macro level and diversify noise from anomalies of individual countries. The results also echo Samuelson's Dictum that informational inefficiency is stronger at the broader macro level (Xiao et al. (2022), Xiao et al. (2023)), providing insights to future studies on global market efficiency.

VI. Decoding Heterogeneity in Market Predictability

In this section, we delve deeper into the driving force behind the heterogeneity in market excess return predictability. We start by estimating the components' variance in anomaly legs and verifying the decomposition framework described in the data-generating process in Section II.D, and enhance our examination with a distinct and economically meaningful division of all countries into developed and emerging ones.

A. Analyzing Variability Components using System of Equations

Following Section II.D, we begin by evaluating the extent to which the "variance ratio" serves as a proxy for the relative importance of price randomness ($\text{var}(f_t)/\text{var}(\Delta u_{S,t})$) as shown in Equation (23). It appears that this linear relationship is unverifiable since k, $\text{var}(f_t)$, $\Delta u_{L,t}$, and $\Delta u_{S,t}$ are all theoretical concepts that seem empirically unmeasurable. However, we show that we can estimate the variability of the embedded components: k, $\text{var}(f_t)$, $\text{var}(\Delta u_{L,t})$ and $\text{var}(\Delta u_{S,t})$ using a system of equations in different ways. Assuming that the time series of the long-leg, short-leg, and long-short return of a representative anomaly are stationary, the variance of them can be written as:

$$\text{var}(L) = k^2 \text{var}(f_t) + \text{var}(\Delta u_{L,t}) \tag{39}$$

$$\text{var}(S) = \text{var}(f_t) + \text{var}(\Delta u_{S,t}) \tag{40}$$

$$\text{var}(L - S) = (k-1)^2 \text{var}(f_t) + \text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t}) \tag{41}$$

First, we approximate $\text{var}(\Delta u_{L,t})/\text{var}(\Delta u_{S,t})$ for all countries. $\Delta u_{L,t}$ ($\Delta u_{S,t}$) denotes the change in underpricing (overpricing) in the long (short) leg return of the representative anomaly. Hence, we use the ratio between the variance of the FF3 alpha of each leg, which captures the portion of the returns that cannot be explained by the exposure to the risk factors, to proxy $\text{var}(\Delta u_{L,t})/\text{var}(\Delta u_{S,t})$:

$$\frac{\text{var}(\Delta u_{L,t})}{\text{var}(\Delta u_{S,t})} = \frac{\text{var}(\alpha_L)}{\text{var}(\alpha_S)} \tag{42}$$

where

$$r_{l,t} = \alpha_l + \beta_M r_{M,t} + \beta_{SMB} r_{SMB,t} + \beta_{HML} r_{HML,t} + \epsilon_{l,t} \quad \text{for } \ l = L, S \tag{43}$$

Using equations (39), (40), (41), and (42), along with the averages of the long-short, long-leg, and short-leg returns of the DLRZ100 anomaly as proxies for the representative anomaly's corresponding components, we can solve for k, $\text{var}(f_t)$, $\text{var}(\Delta u_{L,t})$, and $\text{var}(\Delta u_{S,t})$. Additionally, we calculate the correlation between the "variance ratio" and $\text{var}(f_t)/\text{var}(\Delta u_{S,t})$. In Panel A of Table IV, we show the correlation estimation between the two measures is nearly 1 (0.99), suggesting one could be a strong proxy of the other. The summary statistics of the estimated parameters k, $\text{var}(f_t)$, $\text{var}(\Delta u_{L,t})$ align with our prior expectations: (1) k is a parameter close to 1. (2) Pricing variability is significantly more attributable to the martingale component f_t than to the serially correlated component $\Delta u_{S,t}$, with $\text{var}(f_t)/\text{var}(\Delta u_{S,t})$ averaging 19.01 across all countries. Otherwise, taking arbitrage would be very easy.

Next, we approximate k in the system of equations to estimate k, $\text{var}(f_t)$, $\text{var}(\Delta u_{L,t})$, and $\text{var}(\Delta u_{S,t})$. The analytical solutions of $\text{var}(f_t)$, $\text{var}(\Delta u_{L,t})$ and $\text{var}(\Delta u_{S,t})$ are:

$$\text{var}(f_t) = \frac{\text{var}(L) + \text{var}(S) - \text{var}(L - S)}{2k} \tag{44}$$

$$\text{var}(\Delta u_{L,t}) = \text{var}(L) - \frac{k}{2}[\text{var}(L) + \text{var}(S) - \text{var}(L - S)] \tag{45}$$

$$\text{var}(\Delta u_{S,t}) = \text{var}(S) - \frac{1}{2k}[\text{var}(L) + \text{var}(S) - \text{var}(L - S)] \tag{46}$$

We need a reasonable estimate for k to reach the analytical solutions above. We start with an estimation process where we allow all possible values of k to be considered. Specifically, to make sure $\text{var}(\Delta u_{L,t}), \text{var}(\Delta u_{S,t}) > 0$, k satisfies the following:

$$\frac{\text{var}(L) + \text{var}(S) - \text{var}(L - S)}{2\text{var}(S)} < k < \frac{2\text{var}(L)}{\text{var}(L) + \text{var}(S) - \text{var}(L - S)} \tag{47}$$

We discretize k into a series of values with a small increment within the upper and lower bounds for all countries. For each value of k, we can estimate the analytical solutions of $\text{var}(f_t)$, $\text{var}(\Delta u_{L,t})$, and $\text{var}(\Delta u_{S,t})$, and thus estimate $\frac{\text{var}(f_t)}{\text{var}(\Delta u_{S,t})}$. We also restrict that pricing variability regarding the change in underpricing cannot exceed that of overpricing ($\text{var}(\Delta u_{L,t}) < \text{var}(\Delta u_{S,t})$). [18] After gathering all possible values for $\frac{\text{var}(f_t)}{\text{var}(\Delta u_{S,t})}$ in all countries, we calculate the correlation between it and the "variance ratio".

In addition, we also use two alternative ways to approximate k. The first method estimates k using the ratio of the coefficient estimates from two regressions: the regression of $r_{L,t}$ on $r_{L+S,t}$ and the regression $r_{S,t}$ on $r_{L+S,t}$:

$$\tilde{\beta}_{L\_L+S} = \frac{k(k+1)\text{var}(f_t) + \text{var}(\Delta u_{L,t})}{\text{var}(r_{L+S,t})} \tag{48}$$

$$\tilde{\beta}_{S\_L+S} = \frac{(k+1)\text{var}(f_t) + \text{var}(\Delta u_{S,t})}{\text{var}(r_{L+S,t})} \tag{49}$$

where $r_{L+S,t} = r_{L,t} + r_{S,t}$. The ratio between (48) and (49) is:

$$\frac{\tilde{\beta}_{L\_L+S}}{\tilde{\beta}_{S\_L+S}} = \frac{k(k+1)\text{var}(f_t) + \text{var}(\Delta u_{L,t})}{(k+1)\text{var}(f_t) + \text{var}(\Delta u_{S,t})} \tag{50}$$

Since most pricing variability should come from non-autocorrelated component f_t, $\text{var}(f_t) \gg \text{var}(\Delta u_{L,t})$ and $\text{var}(f_t) \gg \text{var}(\Delta u_{S,t})$. Hence,

$$\frac{\tilde{\beta}_{L\_L+S}}{\tilde{\beta}_{S\_L+S}} \approx k \tag{51}$$

In the second way, we estimate k as the square root of the ratio between $\text{var}(L)$ and $\text{var}(S)$. Similarly, when $\text{var}(f_t) \gg \text{var}(\Delta u_{L,t})$ and $\text{var}(f_t) \gg \text{var}(\Delta u_{S,t})$, the ratio between equation (39) and equation (40) becomes:

$$\frac{\text{var}(L)}{\text{var}(S)} \approx k^2 \tag{52}$$

[18] Due to factors such as short-sale constraints Miller (1977), we should expect a higher magnitude of overpricing than underpricing in general. In addition, the correction of overpricing shall be more persistent than that of underpricing.

Panel B and C show the estimation when we discretize k within its upper and lower bound and when we approximate k using the two alternative methods, respectively. Our findings also align with expectations. The correlation between the "variance ratio" and $\text{var}(f_t)/\text{var}(\Delta u_{S,t})$ are high (0.86 in Panel B, and 0.98 and 0.99 in Panel C), suggesting that the "variance ratio" is a good proxy for the relative importance of price randomness across countries.

Overall, the consistency presented in Panels B and C of Table IV confirms that the simple variance ratio effectively measures price randomness. Given its ease of estimation and interpretation, we opt to employ this variance ratio in subsequent analyses to measure price randomness. This decision is validated by the similarity of results when substituting the variance ratio with $\text{var}(f_t)/\text{var}(\Delta u_{S,t})$, ensuring our methodology remains robust.

Table IV

Analysis on Estimation of Variability Components in Anomaly Returns

This table displays the main findings of our estimation on the components of anomaly return variability. Panel A approximates each country's $var(\Delta u_{L,t})/var(\Delta u_{S,t})$ using its $var(\alpha_L)/var(\alpha_S)$, where $var(\alpha_L)$ ($var(\alpha_S)$) is the variance of the intercept from regressing the average of the long leg of the DLRZ100 anomaly returns on the country level Fama-French 3 factors. Panel B provides the estimation when k is discretized within its upper and lower bounds. Panel C presents the estimation when k is approximated using the ratio of regression coefficients (Method 1) and the square root of $\text{var}(L)/\text{var}(S)$ (Method 2). Specifically, in Method 1, we approximate k as the ratio of the coefficient estimates from two regressions: The first regression is the regression of the average of DZLR100 anomalies' long legs ($r_{L,t}$) on the sum of the average of DZLR100 anomalies' long legs and the average of DZLR100 anomalies' short legs ($r_{L+S,t}$). The second regression is the regression of the average of DZLR100 anomalies' short legs ($r_{S,t}$) on the sum of the average of DZLR100 anomalies' long legs and the average of DZLR100 anomalies' short legs ($r_{L+S,t}$).

Panel A: Approximate $var(\Delta u_{L,t})/var(\Delta u_{S,t})$				
Correlation between $var(f_t)/var(\Delta u_{S,t})$ and the variance ratio				**0.99**
	Mean	SD	Min	Max
k	0.86635	0.05016	0.7393	0.96442
$var(f_t)$	0.00689	0.00348	0.00318	0.02077
$var(\Delta u_{L,t})$	0.00033	0.00023	0.00008	0.00111
$var(\Delta u_{S,t})$	0.00046	0.00032	0.00010	0.00142
$var(f_t)/var(\Delta u_{S,t})$	19.01245	10.03443	4.33219	53.70379
Panel B: Discretize k within the upper and lower bounds				

(Continued)

Correlation between $var(f_t)/var(\Delta u_{S,t})$ and the variance ratio				**0.86**
	Mean	SD	Min	Max
k	0.88754	0.05373	0.69766	1.0567
$var(f_t)$	0.00669	0.00334	0.00303	0.02093
$var(\Delta u_{L,t})$	0.00020	0.00019	0.00000	0.00123
$var(\Delta u_{S,t})$	0.00065	0.00047	0.00009	0.00273
$var(f_t)/var(\Delta u_{S,t})$	13.93493	8.17188	1.89536	57.86897
Panel C: Approximate k				
Method 1				
Correlation between $var(f_t)/var(\Delta u_{S,t})$ and the variance ratio				**0.98**
	Mean	SD	Min	Max
k	0.86148	0.05278	0.72010	0.96406
$var(f_t)$	0.00693	0.00349	0.00319	0.02085
$var(\Delta u_{L,t})$	0.00036	0.00024	0.00008	0.00117
$var(\Delta u_{S,t})$	0.00042	0.00029	0.00009	0.00134
$var(f_t)/var(\Delta u_{S,t})$	20.77069	10.61264	4.99982	56.26608
Method 2				
Correlation between $var(f_t)/var(\Delta u_{S,t})$ and the variance ratio				**0.99**
	Mean	SD	Min	Max
k	0.86577	0.04976	0.73955	0.96450
$var(f_t)$	0.00689	0.00348	0.00318	0.02077
$var(\Delta u_{L,t})$	0.00034	0.00022	0.00008	0.00111
$var(\Delta u_{S,t})$	0.00046	0.00032	0.00010	0.00142
$var(f_t)/var(\Delta u_{S,t})$	19.33217	10.19884	4.19923	54.27266

B. *Decomposition Analysis on Market Predictability*

As discussed in Section II.D, the predictability of market excess returns by long-short anomaly returns is driven by three primary components. The first one is the $\tilde{\beta}_S$, the predictive power of the short-leg return of a representative anomaly via overpricing correction persistence, measuring inter-temporal systematic importance of mispricing. The second one is the *autocovariance ratio*, which proxies the level of asymmetric mispricing correction speed between the short leg and the long leg. A higher level of asymmetry, proxied by a higher value of (more negative) autocovariance ratio, indicates overpricing is more slowly corrected than underpricing. The third component is the *variance ratio*, measuring the market-level relative importance of price randomness. A higher variance ratio reflects more price variation coming from the unpredictable martingale component rather than the mispricing-related component so that the noise cancellation works better via “long-minus-short”. In this section, we show that the supranations formed by countries with higher values of the three components outperforms the supranations consisting of countries with lower values of them.

Specifically, we triple-sort all countries based on their $\tilde{\beta}_S$, autocovariance ratio, and variance ratio. First, we aggregate the country-level DLRZ100 anomaly returns using the Predictor Average method and then calculate each country's $\tilde{\beta}_S$, autocovariance ratio, and variance ratio. Next, we equally divide the countries into top and bottom groups based on their respective values of $\tilde{\beta}_S$, autocovariance ratio, and variance ratio, each considered separately. Finally, we form eight non-overlapping supranations based on the countries' positions (either 'top' or 'bottom') in these divisions. We construct supranational-level anomaly returns and market excess returns for the eight supranations.

Table VI shows the out-of-sample prediction results of the eight supranational market excess returns using their supranational-level anomaly returns. Consistent with our expectations, the best-performing supranation is constructed using countries fall in the “top” groups of $\tilde{\beta}_S$, autocovariance ratio, and variance ratio (the supranation marked as "HHH" in Panel A). The second-tier supranations are those that satisfy two of the above conditions(the supranations marked as "HHL", "HLH", and "LHH" in Panel B) and the third-tier supranations satisfy one of them (the supranations marked as "HLL", "LHL", and "LLH" in Panel C). The last tier has all of their components fall into the “bottom” group. (the supranations m"LLL" Panel D). The average of the first tier achieves the remarkable highest out-of-sample prediction performance with statistical significance (6.47%), followed by the second tier (3.86%). In contrast, the third and the last tier do not show any economic or statistical significance(-0.26% and -0.44%), even aggregated into the supranational level. Overall, the prediction outcomes presented in Table VI align with what is outlined in our framework in equation (25). Specifically, supranations composed of countries with high values in the three components generally demonstrate superior forecast performance.

C. *Developed vs Emerging Markets Analysis*

Now we perform a distinct and economically meaningful cut based on the level of market maturity — by grouping countries into Developed and Emerging [19] to examine

[19]The countries classified as 'Developed' are those whose equity markets are designated as 'Developed Markets' by MSCI. The remaining countries in our sample are classified as 'Emerging.

Table VI

Supranational OOS prediction: Triple Sort by $\tilde{\beta}_S$, the autocovariance ratio, and the variance ratio

This table reports the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percentage at the supranational level using DLRZ100 long-short anomaly portfolio returns to predict market excess returns. Countries are divided into eight groups based on their $\tilde{\beta}_S$, autocovariance ratio, and variance ratio via independent triple sorting. Long-short anomaly returns are aggregated using the Predictor Average method in each country. We construct supranational-level anomaly returns and market excess returns for the eight supranations in a value-weighting way, where countries with larger market capitalization receive a higher weight. Using supranational anomaly returns. supranation "H(L)H(L)H(L)" consists of countries that fall in the high(low) category of $\tilde{\beta}_S$, high(low) category of autocovariance ratio, and high(low) category of variance ratio. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics observed in supranational-level prediction.

Panel A : All three components in the top group			
HHH			1st tier average
6.47***			6.47***
Panel B: Two of three components in the top group			
HHL	HLH	LHH	2nd tier average
4.11**	1.96***	5.52***	3.86**
Panel C: One of three components in the top group			
HLL	LHL	LLH	3rd tier average
0.06	1.75**	-2.6	-0.26
Panel D: All three components in the bottom group			
LLL			4th tier average
-0.44			-0.44

whether and why the 100 anomalies as well as their market excess return predictability behave differently. First, we examine the mispricing level of DLRZ100 in the Developed and the Emerging markets at both the country level and the supranational level following Jacobs (2016), by regressing the averaged DLRZ100 anomaly returnS on global Fama-French three factors. To make a fair comparison, we do the test for Developed markets using two sample periods: 1986:01-2021:12 (to use all available data) and 1991:01-2021:12 (same sample period as the Emerging markets). In our country-level analysis, presented in the second column of Panel A1 of Table VII, we see higher mispricing level from developed markets in the overall level of mispricing between Developed and Emerging countries as indicated by anomalies (0.313 vs. 0.235). However, the conclusion reverses as we aggregate the anomalies to the supranational level to capture systematically important mispricing. We discover a notably higher level of mispricing, denoted by a higher alpha value, in the Emerging market (0.370) than in the Developed market (0.281). The supranational aggregation effect is significantly smaller for Developed markets (-0.177) than for Emerging markets. Our results suggest that the conventional way of studying anomalies in individual countries may have underestimated the macro-level mispricing information, therefore providing an incomplete picture of market inefficiencies of the Emerging supranation.

Although the Emerging supranation has the most salient mispricing level, the prediction power of their long-short anomaly portfolio returns is much lower than the Developed supranation (1.26% vs. 3.24%). However, we model it in Section II.B that this result does not undermine our economic rationale but rather enriches it, as the Emerging supranation is embedded with almost symmetrical mispricing correction speed (MCS) in the long and short leg of anomaly portfolio returns. The empirical evidence is provided in Panel B of Table VII. We compare the predictive performance of long-short returns, long-leg excess returns, and short-leg excess returns in the Developed and Emerging supranations. All forecasts are performed using the Predictor Average method to guard against overfitting. For the Developed supranation, the R^2_{OS} by supranational short legs (1.4%; 1.57%) are notably higher than the long-leg excess returns (0.32%; 0.93%), which is consistent with stronger MCP for overpricing vis-à-vis underpricing explained in Section II.B. Comparing the long-short and short-leg excess returns, the former (5.02%; 3.24%) outperforms the latter (1.4%; 1.57%) sizably for forecasting the monthly market excess return, as the long-short return filters the common component that is unrelated to the future market excess return. Conversely, the Emerging supranation exhibits the much-better-performing long-leg excess return (3.60%) and short-leg excess return (3.69%). There's not a big difference between the R^2_{OS} statistics achieved by long-leg excess return and the one achieved by short-leg excess return. This negligible discrepancy stems from the relatively symmetrical speed of correction in the overpricing and underpricing of the Emerging supranation. It is therefore reasonable that the long-short excess return (1.26%) does not perform as effectively as the separate long- and short-leg excess return separately in this specific supranation. While filtering out the common component that bears no relation to the future market excess return, the long-short return also cancels out the mispricing correction persistence that is embedded equivalently in both the long and the short leg.

Next, to obtain further insights, we analyze the values attributed to Developed and Emerging markets regarding the three dimensions of market efficiency outlined in our earlier decomposition framework in equation (25). We calculate $\log(\tilde{\beta}_S)$, $\log(-\text{autocovariance\_ratio})$, and $\log(\text{sqrt\_variance\_ratio})$ for all countries using the time series of the aggregated DLRZ100 anomaly returns via the *Predictor Average* method. Intuitively, the three log transformations are positively related to the predictive power of the short-leg return, the level of asymmetry, and the price informativeness. To better understand the heterogeneity in predictability, we conduct a market-capitalization-weighted T-test

Table VII
Developed markets vs. Emerging markets

This table displays the anomaly replication rates and Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percentage using DLRZ100 anomaly portfolio returns. Panel Apresent the global-three-factor alpha of DLRZ100 anomalies at both the country level and supranational level. The sample period is 1986:01 to 2021:12 and 1991:01 to 2021:12 for the Developed markets, and 1991:01 to 2021:12 for the Emerging markets. Panel B shows the out-of-sample R^2 (R^2_{OS}) statistics in percentage for the Developed supranation and Emerging supranation at the supranational level. The out-of-sample period covers 1996:01 to 2021:12 and 2001:01 to 2021:12 for the Developed supranation, and 2001:01 to 2021:12 for the Emerging supranation. Columns 1, 2, and 3 show the prediction performance of long minus short leg anomaly return, long leg anomaly return, and short leg anomaly return based on DLRZ100 anomaly portfolio returns, respectively. All columns perform the Predictor Average forecast, which is a univariate predictive regression forecast based on the cross-sectional average of the 100 long-short anomaly portfolio returns. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics.

Panel A: Global FF3 alpha				
	Sample Period	Equal Weighted Country level	Supranational level	
Developed	1986:01 to 2021:12	0.313***	0.271***	
Developed	1991:01 to 2021:12	0.308***	0.281***	
Emerging	1991:01 to 2021:12	0.235***	0.370***	
(Developed - Emerging) Supranational Aggregation Effect			-0.177***	
Panel B: OOS prediction				
	Sample Period	LongShort	Long	Short
Developed	1986:01 to 2021:12	5.02***	0.32	1.4*
Developed	1991:01 to 2021:12	3.24**	0.93	1.57
Emerging	1991:01 to 2021:12	1.26*	3.60**	3.69**

comparing Developed and Emerging countries' $\log(\tilde{\beta}_S)$, log(−autocovariance_ratio), and log(sqrt_variance_ratio).

The findings, presented in Table VIII, reveal that the Developed countries in general exhibit lower predictive power of the short-leg return (−0.41***), a higher degree of asymmetric MCS (0.70***), and a higher degree of market-level price randomness (0.18***) than the Emerging markets. On the one hand, the asymmetry between the speed of correction between overpricing and underpricing is low in Emerging markets. Thus, the "long-minus-short" construction makes the MCP embedded in each leg cancel out each other. In addition, emerging markets also experience a lower level of the relative importance of price randomness (denoted by a lower value of variance ratio), so not much unpredictable component to cancel out via the "long-minus-short" construction when compared to developed markets. On the other hand, a higher level of MCP, as denoted by a higher $\tilde{\beta}_S$ and a lower value of autocovariance ratio, allows the emerging markets to use the long- and the short-leg separately to predict the market excess return. Therefore, long-short anomaly portfolio returns perform better in the market excess return predictability in developed markets and long- *and* short-leg returns perform better in the Emerging markets.

Table VIII

Developed vs. Emerging Difference in Decomposition

This table reports market-capitalization-weighted T-test comparing Developed and Emerging countries' $\log(\tilde{\beta}_S)$, $\log(-\text{autocovariance\_ratio})$, and $\log(\text{sqrt\_variance\_ratio})$. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	$\log(\tilde{\beta}_S)$	$\log(-\text{autocov\_ratio})$	$\log(\text{sqrt\_var\_ratio})$
Developed vs. Emerging	-0.41***	0.70***	0.18***

Collectively, our analyses above explore and decode the heterogeneity in the market excess predictability across diverse markets. Aligning with the decomposition framework detailed in Section II.D, we find that supranations characterized by higher inter-temporal, systematic importance of mispricing (ISIoM), more asymmetric mispricing correction speed, and greater relative importance of price randomness exhibit superior performance in predicting market excess returns. In addition, our results indicate that long-short anomaly returns significantly forecast market excess returns in the developed markets. In contrast, both the long- and short-leg anomaly returns are significant predictors of market excess returns in Emerging markets. This distinction is supported by the findings that emerging markets generally exhibit higher ISIoM in both the long-leg underpricing and the short-leg overpricing, less asymmetry in the speed of mispricing correction, and lower relative importance of price randomness.

VII. Anomaly Characteristic Analysis

In this section, we check the market return predictability of the aggregate anomaly characteristics. Following Engelberg et al. (2023), we construct market-wide anomaly characteristics based on DLRZ100 anomalies and use them to forecast the next-month market excess return at both the country level and the supranational level. For supranational prediction, both the anomaly characteristics and market excess returns are constructed in a market-capitalization-weighted way across all available countries selected to form the supranation. We do the Augmented Dickey-Fuller (1979) test for every anomaly characteristic and adjust the nonstationarity following Engelberg et al. (2023). We then use the Predictor Average method to aggregate the DLRZ100 adjusted anomaly characteristics at both the country level and the supranational level. For supranational prediction, both the anomaly characteristics and market excess returns are constructed in a market-capitalization-weighted way across all available countries selected to form the supranation. For country-level prediction, the average R^2_{OS} is calculated for market excess return forecasts using the same set of countries. The out-of-sample prediction performance is shown in Table IX. Consistent with the finding of Engelberg et al. (2023), all country-level anomaly characteristics fail to predict market excess return. With all country-level R^2_{OS} being negative, the country-level market-wide anomaly characteristics even underperform the prevailing mean benchmark. However, after we aggregate the anomaly characteristics into supranational level, they work much better. We observe economically and statistically significant out-of-sample predictability in G6 (3.55%) and Developed (3.28%), and economically significant out-of-sample predictability in Emerging (0.69%). Our findings suggest that although anomaly characteristics may introduce more noise compared to anomaly returns, as indicated in Section II.E, supranational aggregation may still help to mitigate them to some extent.

Table IX

OOS Prediction using Anomaly Characteristics

The table reports Campbell and Thompson (2008) R^2 (R^2_{OS}) statistics in percent for market excess return forecasts based on DLRZ100 anomaly characteristics at both country level and supranational level. We do the Augmented Dickey-Fuller (1979) test for every anomaly characteristic and adjust the nonstationarity following Engelberg et al. (2023). We then use the Predictor Average method to aggregate the DLRZ100 adjusted anomaly characteristics at both the country level and the supranational level. For supranational prediction, both the anomaly characteristics and market excess returns are constructed in a market-capitalization-weighted way across all available countries selected to form the supranation. For country-level prediction, the average R^2_{OS} is calculated for market excess return forecasts using the same set of countries. The out-of-sample period is 1996:01 to 2021:12 for G6, G19, and World, and 2001:01 to 2021:12 for Emerging. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively, for the positive R^2_{OS} statistics.

Region	Country Level	Supranational Level
G6	-0.44	3.55***
G19	-1.02	-1.41
World	-0.63	-1.03
Developed	-0.36	2.28**
Emerging	-0.77	0.69

VIII. Allowing for Domestic Data Mining

Given the ongoing debates and attention surrounding the potential spuriousness of anomalies (e.g., Harvey et al. (2016) and Hou et al. (2020)), we take measures to ensure that our supranational out-of-sample time-series predictability is not driven by data mining issue. To investigate the possibility of p-hacking in anomalies, a natural approach is to examine their replicability beyond the initial sample where they were first discovered, for example, in markets other than the U.S. (Lu et al. (2017)). Similarly, we also accounted for the potential data mining on U.S. market return predictability. Thus, we construct new predictors and target variables by removing the common component between the U.S. and the G6, G19, and World supranation in the anomaly returns and market excess returns respectively, via recursive regressions.

First, for each supranation j, we regress each of its anomaly i's return ($\mathrm{r}_{i,t-1}^{j}$) and market excess return ($\mathrm{r}_{m,t-1}^{j}$) on the corresponding domestic peers and get the recursive slope coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\mathrm{r}_{i,t-1}^{j} = \beta_0 + \beta_1{}^{*}\mathrm{r}_{i,t-1}^{US} + \epsilon_{t-1} \tag{53}$$

$$\mathrm{r}_{m,t-1}^{j} = \beta_0 + \beta_1{}^{*}\mathrm{r}_{m,t-1}^{US} + \epsilon_{t-1} \tag{54}$$

Next, we remove the comoving part between r_i^j (r_m^j) and r_i^{US} (r_m^{US}) that could suffer from domestic data mining issues from the original supranational anomaly return (market excess return):

$$\tilde{r}_{i,t}^{\mathrm{j}} = \mathrm{r}_{i,t}^{\mathrm{j}} - (\hat{\beta}_0 + \hat{\beta}_1{}^{*}\mathrm{r}_{i,t}^{\mathrm{US}}) \tag{55}$$

$$\tilde{r}_{m,t}^{\mathrm{j}} = \mathrm{r}_{i,t}^{\mathrm{j}} - (\hat{\beta}_0 + \hat{\beta}_1{}^{*}\mathrm{r}_{m,t}^{\mathrm{US}}) \tag{56}$$

Table X shows the out-of-sample prediction results when data mining is allowed. The left panel assumes anomalies are spurious, where we replace the r_i^j by $\tilde{r}_i^{\mathrm{j}}$ as new predictors. The right panel shows the results conditional on both the anomalies and their market return predictability are outcomes of data mining, where we replace the r_i^j by $\tilde{r}_i^{\mathrm{j}}$ as new predictors, and $\mathrm{r}_{m,t-1}^{j}$ by $\tilde{r}_{m,t}^{\mathrm{j}}$ as new target variables. We take the first five years to conduct the first recursive regression so our out-of-sample period covers 2006:01 to 2021:12. All supranations remain economically significant according to Campbell and Thompson (2008) in both panels; most of them are still statistically significant according to Clark and West (2007) statistics.

To conclude, we employ a direct method to tackle the data-mining issue frequently encountered in studies of financial anomalies. Our analysis demonstrates that our results remain robust even under the assumption that both the anomalies and their capability to predict market returns have been identified through data mining within domestic markets.

Table X

Supranational OOS prediction: Allowing for Domestic Data mining

This table reports the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percentage at the supranational level using DLRZ100 long-short anomaly portfolio returns, assuming all the anomalies are data-mined in U.S. (Panel A), and both the anomalies and their market return predictability are data-mined in U.S. (Panel B). The out-of-sample period covers 2006:01 to 2021:12. The supranational predictors in both Panel A and Panel B are constructed as the residual from the recursive regression where each supranational anomaly return is regressed on its contemptuous U.S. counterpart. The response variable in Panel A is the supranational market excess return. The response variable in Panel B is constructed as the residual from the recursive regression where each supranational market excess return is regressed on contemptuous U.S. market excess return. Avg is a univariate predictive regression forecast based on the cross-sectional average of the 100 long-short anomaly portfolio returns. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the 100 long-short anomaly portfolio returns. C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. The ENet forecast is based on elastic net estimation of a multiple predictive regression that includes all 100 of the long-short anomaly portfolio returns. Combine is the arithmetic mean of univariate predictive regression forecasts based on the 100 individual long-short anomaly portfolio returns (in turn). Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics observed in supranational-level prediction.

	Panel A: Assuming domestic anomaly returns are data-mined						Panel B: Assuming both domestic anomaly returns and their market return predictability are data-mined					
	Avg	PC	PLS	C-Enet	Enet	Combine	Avg	PC	PLS	C-Enet	Enet	Combine
G6	3.17**	2.82*	4.22**	5.41**	0.80	0.87*	3.72**	3.09*	2.85**	4.79*	-0.92	1.01*
G19	2.36**	2.53**	4.43**	1.01	2.07*	0.72**	2.23**	2.82**	2.58*	2.00	0.94	0.67*
World	2.48**	2.81*	4.27**	1.95	0.89	0.75*	2.7**	2.85*	3.04*	-1.28	-0.78	0.74*

IX. Systematical Asymmetric MCP across Supranations

To further substantiate the correlation between the systematical asymmetric mispricing correction persistence and market excess return predictability, we conduct a series of tests based on supranational-level anomaly portfolio returns from different aspects.

A. *Coefficient Analysis*

As explained in Section II, the asymmetry in MCP contributes to stronger MCP for overpricing vis-à-vis underpricing. Thus, a negative direction is expected on the predictive capacity of long-short anomaly returns of the G6, G19, and World supranations. To capture the information in all 100 of the supranational long-short anomaly portfolio returns, we examine recursive estimates of the slope coefficients in the predictive regressions underlying the predictor average forecasts.

The predictive regression underlying the predictor average forecast is given by

$$r_{M,t+1} = \alpha + \beta \bar{r}_{LS,t} + \varepsilon_{t+1}, \tag{57}$$

where $\bar{r}_{LS,t} = \frac{1}{n}\sum_{i=1}^{n} r^{i}_{LS,t}$ and $r^{i}_{LS,t}$ is the supranational long-short portfolio return for the ith anomaly for $i = 1, \ldots, n$ ($n = 100$).

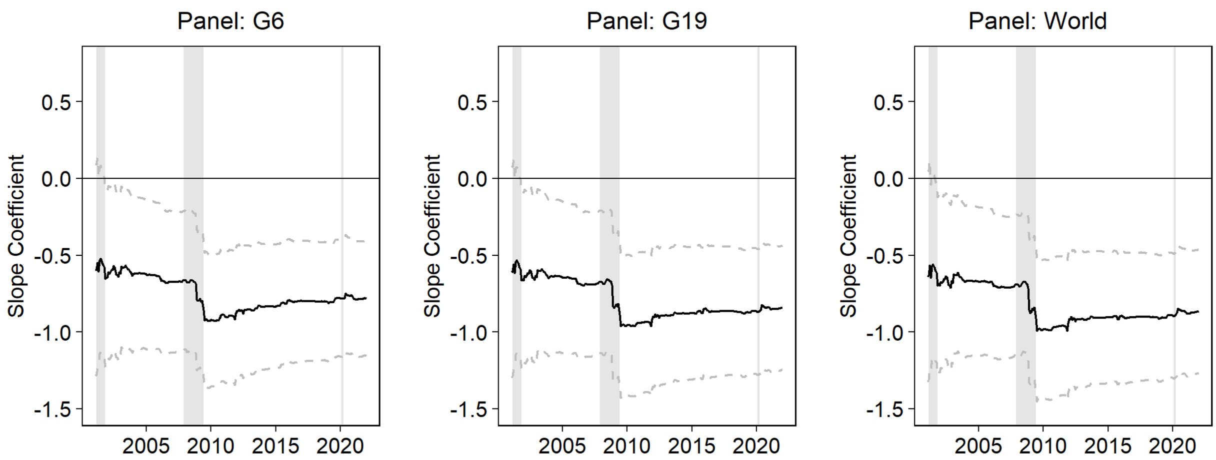

Figure 1. Recursive slope coefficient estimates. Solid lines depict standardized recursive slope coefficient estimates used to compute the predictor average forecasts of the supranational market excess return in G6, G19, and World supranations based on DLRZ100 long-short anomaly portfolio returns. The sample period covers 2001:01 to 2021:12. Dashed lines denote 90% confidence intervals. Vertical bars indicate business-cycle recessions as dated by the National Bureau of Economic Research.

The three panels of Figure 1 depict recursive estimates of the standardized slope coefficients in equation (57) for each supranation and their 90% confidence intervals when the explanatory variable is $\bar{r}_{LS,t}$. [20] The recursive estimates are negative for G6, G19, and World supranations, consistent with stronger MCP for overpricing vis-à-vis underpricing. As expected, the 90% confidence intervals tend to narrow as the estimation sample lengthens. The recursive estimates become significant around the mid-2000s and remain significant thereafter. The recursive estimates are quite stable in all panels of Figure 1 from the early to late 2000s. They then became larger in magnitude around the Global Financial Crisis and concomitant Great Recession and remained relatively stable thereafter, suggesting that systematical asymmetric mispricing correction persistence has not substantially diminished in importance since the mid-2000s.

B. *Arbitrage Trading*

Finally, we investigate whether long-short anomaly portfolio returns predict arbitrageurs' trading activities in the broad market. We aggregate the long-short anomaly portfolio returns at the supranational level and use the average of all DLRZ100 long-short anomaly returns to construct predictors for our predictive regressions. Given that asymmetric limits of arbitrage generate relatively strong overpricing correction persistence, we expect an increase in long-short anomaly returns to lead to an increase in arbitrageurs' short position; in other words, arbitrageurs' trading activities are more likely to confirm that shares were broadly overvalued. To measure arbitrageurs' activities, we construct a value-weighted short-selling indicator across all countries by aggregating the one-month percentage change in short interest, defined as the total number of uncovered shares sold short (sourced from Markit) divided by the total number of shares outstanding (from Compustat), at the country level. This country-level short-selling measure is then aggregated to the supranational level through market-cap weighting. The sample period covers 2002:01 to 2021:12. Both the dependent variables and the predictors are standardized before estimating the predictive regressions.

Table XI reports monthly predictive regression results for supranational arbitrageurs' short selling. The results support our conjecture. The coefficient estimates for predicting next-month short selling are significantly positive and economically substantial across all supranations: a one-standard-deviation increase in the supranational long-short anomaly return leads to a 0.21, 0.17, and 0.20 standard-deviation increase in the percentage change of arbitrageurs' short interest for the G6, G19, and World supranations, respectively. These findings align with the perspective that the arbitrageurs' trading activities are reflective of the ongoing mispricing correction process, particularly focusing on the continuation of overvaluation corrections.

[20]The horizontal axes in Figure 1 correspond to the forecast month, so that the month t estimate is based on data available through month $t-1$.

Table XI
Arbitrage Trading

This table reports ordinary least squares estimates of standardized slope coefficient estimates for univariate predictive regressions that use the predictor average strategy to combine the information in DLRZ100 long-short portfolio returns to predict the monthly percentage change in arbitrageurs' short interest. The sample period covers 2002:05 to 2021:12. DLRZ100 anomalies include 91 anomalies that have the same concept as those in Dong et al. (2022), and 9 anomalies replaced due to data availability. The predictor average uses the cross-sectional average of DLRZ100 long-short anomaly portfolio returns. Both the predictors and dependent variables are standardized. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Supranation	G6	G19	World
Aribitrage Short Market Position	0.203**	0.163**	0.188**

C. Anomaly Subgroups

As documented in the work in the study by Dong et al. (2022) focused on the U.S. market, asymmetric limits to arbitrage can generate asymmetric MCP. Greater arbitrage limits facilitate larger initial mispricing and a more enduring correction process. Stocks with larger bid-ask spreads, greater idiosyncratic volatility, and smaller market capitalization are typically viewed as having stronger limits of arbitrage. We, therefore, classify anomalies according to bid-ask spread (BA), idiosyncratic volatility (IDIO), and market capitalization (SIZE) , and compare the relative performance in the classified subgroups. For a given supranation, month, and anomaly characteristic, we first sort stocks into deciles as

Table XII

R^2_{OS} Statistics for Subgroups

This table reports Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percent for market excess return forecasts based on long-short anomaly portfolio returns for the subgroups of DLRZ100 long-short anomaly portfolio returns in the second column at the supranational level. The out-of-sample period covers 2001:01 to 2021:12. For supranational-level prediction, both the supranational-level anomaly returns and market excess returns are constructed as market-cap-weighted returns across all available countries within the supranation. We form the anomaly subgroups BA-POS, IDIO-POS, and SIZE-POS (the left panel) and BA-NEG, IDIO-NEG, and SIZE-NEG (the right panel). Avg is a univariate predictive regression forecast based on the cross-sectional average of the 100 long-short anomaly portfolio returns. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the 100 long-short anomaly portfolio returns. C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. The ENet forecast is based on elastic net estimation of a multiple predictive regression that includes all 100 of the long-short anomaly portfolio returns. Combine is the arithmetic mean of univariate predictive regression forecasts based on the 100 individual long-short anomaly portfolio returns (in turn). Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics observed in supranational-level prediction.

		Positive						Negative					
Region	Proxy	avg	pcr	plsr	cenet	enet	combine	avg	pcr	plsr	cenet	enet	combine
G6	BA	2.02**	-0.28	-3.38	-0.24	-1.32	0.06	4.89***	3.76**	8.02***	1.77**	3.79**	1.27**
	IDIO	1.41*	-2.18	-4.85	-0.96	-0.92	-0.18	5.36***	4.22**	7.32***	2.2***	3.42**	1.11**
	SIZE	4.33***	3.49**	6.05***	3.09***	3.29*	1.15**	3.28***	-0.11	-2.27	-1.52	-0.33	0.20
G19	BA	2.73**	-0.27	-0.77	-0.33	0.16	0.12	3.61**	3.31**	8.18***	2.18***	4.54**	1.08***
	IDIO	1.50**	-1.57	-2.62	-0.81	-0.30	-0.08	4.52***	3.96**	7.92***	2.14**	4.84**	0.95**
	SIZE	3.85***	3.33**	6.62***	2.44**	4.05**	0.99**	2.66***	-0.23	-0.10	-0.21	0.12	0.21
World	BA	3.23**	-0.26	-0.55	-0.38	0.20	0.14	3.50**	3.41**	7.78***	1.62**	4.59**	1.11**
	IDIO	1.21*	-1.80	-2.78	-0.40	-1.03	-0.10	4.85***	4.14**	7.96***	2.37**	5.00**	1.00**
	SIZE	4.27***	3.62**	7.14***	2.17**	3.77***	1.07**	2.20**	-0.36	-0.59	0.32	0.57	0.15

described in Section IV and compute the average values for a given proxy for the stocks in the long and short legs. We then compute the long-leg average value for the proxy minus the short-leg average value for the proxy for each month. Finally, we compute the time-series average of the differences for the in-sample period of each supranation (1986:01 to 2000:12). We denote the time-series averages for the differences for the bid-ask spread, idiosyncratic volatility, and size proxies by DTSA-BA, DTSA-IDIO, and DTSA-SIZE, respectively. We form the subgroups of anomalies for which DTSA-BA, DTSA-IDIO, and DTSA-SIZE, respectively, are negative (positive). In line with our out-of-sample focus, we exclude data from the forecast evaluation period when determining the subgroups. We expect greater market return predictability based on negative DTSA-BA, negative DTSA-IDIO, and positive DTSA-SIZE compared to positive DTSA-BA, positive DTSA-IDIO, and negative DTSA-SIZE, respectively, as the former three subgroups represent anomalies with asymmetrically stronger MCP in their short legs than their long legs. Table XII reports R^2_{OS} statistics for the different subgroups in each supranation using the strategies designed to guard against overfitting. Although the bid-ask spread, idiosyncratic volatility, and size are quite noisy proxies for limits of arbitrage, the results generally support the relevance of asymmetric limits of arbitrage and stronger MCP for overpricing vis-à-vis underpricing. For example, in G6, nearly all of the six forecasts for the subgroups of negative BA, negative IDIO, and positive SIZE are both economically and statistically significant. Almost every R^2_{OS} statistics of them are particularly large (e.g., 5.09%, 4.58%, and 4.09% under the predictor average method, respectively). In contrast, for each forecasting strategy in G6, the R^2_{OS} statistic for the positive BA, positive IDIO, and negative SIZE subgroup is always lower than the one of their peer subgroup. Only four are above the 0.5% threshold. The results for the subgroups of G19 and World deliver similar messages as G6, aligning with the intuition in Section II, with relatively stronger limits of arbitrage in the short leg producing greater market return predictability.

D. Market Frictions

In this section, we assess the influence of time-series market-level frictions on MCP. Frictions such as limited risk-bearing capacity and transaction costs lead arbitrageurs to adjust to mispricing gradually, resulting in MCP (Gârleanu and Pedersen (2013, 2016)). Therefore, the long-short anomaly portfolio returns should contain more relevant information for predicting the market excess return during times of high frictions, given our predictability finding is driven by arbitrageurs slowly correcting mispricing in the presence of asymmetric limits of arbitrage and stronger MCP for overpricing vis-à-vis underpricing. We investigate this issue by testing for an increase in the R^2_{OS} statistic during periods of high friction using equation (37).

The following proxies for market frictions from the literature are considered: (1) The level of innovations to aggregate liquidity. We construct it based on Amihud (2002). (2) Idiosyncratic volatility, which is widely believed to be a major implementation cost of short arbitrage (e.g., Pontiff (2006)). We measure aggregate idiosyncratic risk for a given month by first calculating the idiosyncratic volatility of individual stocks following Ang et al. (2006) and then computing the value-weighted average of the idiosyncratic volatilities for the individual stocks. Both proxies are initially constructed at the country level and subsequently aggregated to the supranational level using market-cap weighting. (3) Short fee (Asness et al. (2018)), which measures the cost of shorting stocks. We again aggregate to the market level by computing the value-weighted average of short fees for individual stocks.[21] For all of the proxies, we separate high- and low-friction regimes using the sample median in each supranation.

Table XIII reports differences in R^2_{OS} statistics (in percentage points) between high- and low-friction regimes for market excess return forecasts based on the 100 long-short anomaly portfolio returns in G6, G19, and World supranations.In support of the relevance of asymmetric limits of arbitrage and stronger MCP for overpricing relative to underpricing, most of the R^2_{OS} increased in the high-friction periods under the six methods for supranations. Some of these increases have considerable magnitudes, exceeding 10 percentage points at 1% significant level.

X. Conclusion

We find that anomalies and their prediction on market excess return can be extended abroad at the supranational level, but not at the country level. Long-short anomaly returns strongly predict market excess returns of the Developed supranation, while long *and* short anomaly returns strongly predict market excess returns of the Developed supranation. The results hold even after accounting for the possibility of data mining within domestic markets. In contrast, aggregated anomaly characteristics have limited power in predicting market excess returns, despite their perceived importance for cross-sectional return predictability and asset pricing in previous literature. We provide rationales for our findings, supporting that the predictive ability of long-short anomaly returns comes from three dimensions: The first is the inter-temporally systematic importance of the mispricing (ISIoM). The second is the asymmetry in the speed of overpricing correction to underpricing correction. The third is the market-level relative importance of price randomness , where higher randomness indicates there's higher variability in the unpredictable martingale component of stock returns as opposed to the mispricing-related

[21]The short fee is based on the cost-to-borrow score from Markit (https://ihsmarkit.com/index.HTML), beginning in 2002:01.

Table XIII

R^2_{OS} Statistic Differences Between High- and Low-Friction Regimes

This table reports percentage-point increases in Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics for market excess return forecasts using DLRZ100 long-short anomaly portfolio returns at the supranational level. The out-of-sample period covers 2001:01 to 2021:12. The increase is computed between high and low-friction periods, which are defined using the sample median of the variable in the second column. Avg is a univariate predictive regression forecast based on the cross-sectional average of the 100 long-short anomaly portfolio returns. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the 100 long-short anomaly portfolio returns. C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. The ENet forecast is based on elastic net estimation of a multiple predictive regression that includes all 100 of the long-short anomaly portfolio returns. Combine is the arithmetic mean of univariate predictive regression forecasts based on the 100 individual long-short anomaly portfolio returns (in turn). The Ens (ensemble) forecast takes the average forecast on the above 6 methods. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R2OS statistics observed in supranational level prediction.

Region	Friction measures	Avg	PC	PLS	C-Enet	Enet	Combine
G6	Aggregate Liquidity	12.82***	9.88***	19.75***	8.82***	10.29***	1.86
	Liquidity Innovations	9.66***	10.23***	16.77***	9.03***	7.27**	1.59
	IDIO	2.66***	2.34*	7.65***	1.16*	5.25**	0.61
	Short Fee	11.32***	10.79***	24.59***	7.34**	10.82***	1.66
G19	Aggregate Liquidity	8.33***	5.82**	14.07***	6.18**	5.53**	1.15
	Liquidity Innovations	2.01	4.54*	10.07***	4.68*	5.67**	0.94
	IDIO	2.47***	3.85*	11.42***	3.82**	3.73**	0.89
	Short Fee	12.72***	10.87***	24.96***	7.7***	9.01***	1.80
World	Aggregate Liquidity	10.56***	6.07*	12.85***	3.68	6.81**	1.19
	Liquidity Innovations	1.57	5.06*	6.55***	2.52	5.22*	0.79
	IDIO	1.63***	3.65*	8.43***	3.16	6.38***	0.79
	Short Fee	7.47***	5.53**	15.94***	2.05	5.5**	0.90

component. In addition, we show evidence of global trading and pricing to support the channel underlying our findings. To sum up, through exploring and providing a comprehensive analysis of the linkage between the cross-sectional and time-series return predictability aboard, we develop innovative methods and understandings towards the importance of systematic and macro-level market efficiency.

REFERENCES

Akaike, Hirotogu, 1973, Information theory and an extension of the maximum likelihood principle, in Boris N. Petrov, and Frigyes Csaki, eds., *Proceedings of the 2nd International Symposium on Information Theory* (Akadémiai Kiadó, Budapest).

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Economics* 5, 31–56.

Andrei, Daniel, and Julien Cujean, 2017, Information percolation, momentum and reversal, *Journal of Financial Economics* 123, 617–645.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Asness, Clifford, Andrea Frazzini, Ronen Israel, Tobias J. Moskowitz, and Lasse H. Pedersen, 2018, Size matters, if you control your junk, *Journal of Financial Economics* 129, 479–509.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2013, Anomalies and financial distress, *Journal of Financial Economics* 108, 139–159.

Baba Yara, Fahiz, Martijn Boons, and Andrea Tamoni, 2020, Persistent and transitory components of characteristics: Implications for asset pricing, *Martijn and Tamoni, Andrea, Persistent and Transitory Components of Characteristics: Implications for Asset Pricing (January 31, 2020)* .

Bai, Jennie, Thomas Philippon, and Alexi Savov, 2016, Have financial markets become more informative?, *Journal of Financial Economics* 122, 625–654.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 1509–1531.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *Journal of Finance* 51, 1681–1713.

Clark, Todd E., and Michael W. McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85–110.

Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

Da, Zhi, Umit G. Gurun, and Mitch Warachka, 2014, Frog in the pan: Continuous information and momentum, *Review of Financial Studies* 27, 2171–2218.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839–1885.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *the Journal of Finance* 52, 1–33.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The review of Financial studies* 22, 1915–1953.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

Diebold, Francis X., and Minchul Shin, 2019, Machine learning for regularized survey forecast combination: Partially-egalitarian Lasso and its derivatives, *International Journal of Forecasting* 35, 1679–1691.

Dong, Xi, Namho Kang, and Joel Peress, 2020, Slow arbitrage: Fund flows and mispricing in the frequency domain, Working paper, City University of New York, Baruch College, Bentley University, and INSEAD.

Dong, Xi, Yan Li, Yanran Li, David E. Rapach, and Guofu Zhou, 2024, Anomalies and the market return: Further evidence and discussion, Working paper, Baruch College, Southwestern University of Finance and Economics, Federal Reserve Bank of Atlanta, and Washington University in St. Louis.

Dong, Xi, Yan Li, David E. Rapach, and Guofu Zhou, 2022, Anomalies and the expected market return, *Journal of Finance* 77, 639–681.

Engelberg, Joseph, R David McLean, Jeffrey Pontiff, and Matthew C Ringgenberg, 2023, Do cross-sectional predictors contain systematic information?, *Journal of Financial and Quantitative Analysis* 58, 1172–1201.

Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 21, 3–25.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and the expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F, and Kenneth R French, 1996, Multifactor explanations of asset pricing anomalies, *The journal of finance* 51, 55–84.

Flynn, Cheryl J., Clifford M. Hurvich, and Jeffrey S. Simonoff, 2013, Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models, *Journal of the American Statistical Association* 108, 1031–1043.

Gârleanu, Nicolai, and Lasse Heje Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 66, 2309–2340.

Gârleanu, Nicolai, and Lasse Heje Pedersen, 2016, Dynamic portfolio choice with frictions, *Journal of Economic Theory* 165, 487–516.

Gompers, Paul, Joy Ishii, and Andrew Metrick, 2003, Corporate governance and equity prices, *The quarterly journal of economics* 118, 107–156.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Granger, Clive W. J., and Ramu Ramanathan, 1984, Improved methods of combining forecasts, *Journal of Forecasting* 3, 197–204.

Han, Yufeng, Ai He, David E. Rapach, and Guofu Zhou, 2021, Expected stock returns and firm characteristics: E-LASSO, assessment, and implications, Working paper, University of North Carolina, Charlotte, University of South Carolina, Saint Louis University, and Washington University in St. Louis.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ...And the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hoerl, Arthur E., and Robert W. Kennard, 1970, Ridge regression: Applications to nonorthogonal problems, *Technometrics* 12, 69–82.

Hong, Harrison, Terence Lim, and Jeremy C. Stein, 2000, Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies, *Journal of Finance* 55, 265–295.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Hurvich, Clifford M., and Chih-Ling Tsai, 1989, Regression and time series model selection in small samples, *Biometrika* 76, 297–307.

Inoue, Atsushi, and Lutz Kilian, 2005, In-sample or out-of-sample tests of predictability: Which one should we use?, *Econometric Reviews* 23, 371–402.

Jacobs, Heiko, 2016, Market maturity and mispricing, *Journal of financial economics* 122, 270–287.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *The Journal of Finance* 78, 2465–2518.

Kelly, Brian, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Kelly, Brian, and Seth Pruitt, 2015, The three-pass regression filter: A new approach to forecasting using many predictors, *Journal of Econometrics* 186, 294–316.

Lo, Andrew W, and A Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *The review of financial studies* 1, 41–66.

Lu, Xiaomeng, Robert F Stambaugh, and Yu Yuan, 2017, Anomalies abroad: Beyond data mining.

McCracken, Michael W., 2007, Asymptotics for out of sample tests of Granger causality, *Journal of Econometrics* 140, 719–752.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy return predictability? *Journal of Finance* 71, 5–32.

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151–1168.

Nelson, Charles R., 1976, Inflation and rates of return on common stocks, *Journal of Finance* 31, 471–483.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Pástor, Ľuboš, and Robert F. Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance* 64, 1583–1628.

Pontiff, Jeffrey, 2006, Costly arbitrage and the myth of idiosyncratic risk, *Journal of Accounting and Economics* 42, 35–52.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.

Rapach, David E., and Guofu Zhou, 2020, Time-series and cross-sectional stock return forecasting: New machine learning methods, in Emmanuel Jurczenko, ed., *Machine Learning for Asset Management: New Developments and Financial Applications* (Wiley, Hoboken, NJ).

Stambaugh, Robert F., Jiafeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 102, 288–302.

Stambaugh, Robert F., Jiafeng Yu, and Yu Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.

West, Kenneth D., 1996, Asymptotic inference about predictive ability, *Econometrica* 64, 1067–1084.

Xiao, Yaqing, Hongjun Yan, and Jinfan Zhang, 2022, A global version of samuelson's dictum, *American Economic Review: Insights* 4, 239–254.

Xiao, Yaqing, Hongjun Yan, and Jinfan Zhang, 2023, Global and local information efficiency: An examination of samuelsonâs dictum, Working paper, Capital University of Economics and Business, DePaul University, and Chinese University of Hong Kong (Shenzhen).

Yuan, Ming, and Guofu Zhou, 2022, Why naive 1/n diversification is not so naive, and how to beat it?, Working paper, Columbia University and Washington University in St. Louis - John M. Olin Business School.

Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67, 301–320.

Appendix

A. Tables

Table A1

In-Sample Supranational Forecast

This table reports ordinary least squares estimates of standardized slope coefficient estimates and t-stat for univariate predictive regressions that use the Predictor Average strategy to combine the information in supranational DLRZ100 long-short portfolio returns to predict the market excess return of the same supranation. The Predictor Average uses the cross-sectional average of DLRZ100 long-short anomaly portfolio returns. The sample period covers 1986:01 to 2021:12. The dependent variable is also standardized.

	G6	G19	World
Coefficient	-0.19***	-0.19***	-0.20***

Table A2

Supranational vs Country-Level $R^2_{\mathbf{OS}}$ Statistics

This table reports Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percentage, for multi-period market excess return forecasts, using DLRZ100 long-short anomaly portfolio returns. The forecasts are examined at the supranational level. The out-of-sample period covers 2001:01 to 2021:12. Both the anomaly returns and market excess returns are constructed as market-cap-weighted returns across all available countries within the supranation. Avg is a univariate predictive regression forecast based on the cross-sectional average of the 100 long-short anomaly portfolio returns. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the 100 long-short anomaly portfolio returns. C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. The ENet forecast is based on elastic net estimation of a multiple predictive regression that includes all 100 of the long-short anomaly portfolio returns. Combine is the arithmetic mean of univariate predictive regression forecasts based on the 100 individual long-short anomaly portfolio returns (in turn). Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics observed in supranational level prediction.

	Horizon	Avg	PC	PLS	C-Enet	Enet	Combine
G6	3-month	3.53**	3.21**	6.17***	3.06**	4.02**	0.93**
	6-month	4.94***	3.67**	5.83***	3.87**	1.29*	0.88**
	9-month	1.86**	1.15**	0.37**	0	0.08	0.25
	12-month	0.35	-0.01	-5.86	0.18	-2.36	-0.12
G19	3-month	2.66**	2.12*	5.29**	3.91**	3.52**	0.59*
	6-month	3.64**	2.17*	3.73**	0.79	0.41	0.44
	9-month	1.2*	0.39	-0.55	0	-0.42	0.01
	12-month	0.3	-0.09	-5.03	0.02	-2.22	-0.22
World	3-month	3.08**	2.7*	5.78**	1.29*	3.82**	0.7*
	6-month	4.02**	2.61**	4.25**	2.05*	2.69	0.55*
	9-month	1.56**	0.88	0.22*	-0.56	-0.43	0.09
	12-month	0.72	0.48	-3.67	0	-0.67	-0.14

Table A3
OOS Prediction using Datastream&Worldscope Database

This table reports Campbell and Thompson (2008) out-of-sample R2 (R^2_{OS}) statistics in percentage for market excess return forecasts using the long-short anomaly portfolio returns, based on Dong et al. (2022), at the supranational level. The out-of-sample period covers 2001:01 to 2021:12. For supranational-level prediction, both the supranational-level anomaly returns and market excess returns are constructed as market-cap-weighted returns across all available countries within the supranation, using data from Datastream and Worldscope. The six shrinkage methods in Section III.A are applied. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics observed in supranational level prediction.

	Avg	PC	PLS	C-Enet	Enet	Combine
G6	3.5**	2.86**	2.9**	1.69**	3.57*	0.85**
G19	2.73**	2.23**	1.61*	0.13	-0.1	0.61**
World	3.15**	2.42**	1.88**	2.69**	-1.76	0.75**

Table A4
Evaluating JKP153 Anomalies

This table displays the performance of the 153-anomaly portfolio returns based on Jensen et al. (2023). Panel A presents the percentage of anomalies that can be replicated at a 90% confidence interval at the supranation level for G6, G19, and World. Both the supranational-level anomaly returns and market excess returns are constructed as market-cap-weighted returns across all available countries within the supranation. The replication rates at the supranational level are shown in Columns 2 and 4. Panel B reports Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percentage for market excess return forecasts at the supranational level for the G6, G19, and the World supranations using the six shrinkage methods in Section III.A. The out-of-sample period covers 2001:01 to 2021:12. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for positive R^2_{OS} statistics.

Panel A: Replication Rate

	Raw	CAPM	FF3
G6	46.41%	51.63%	50.33%
G19	57.52%	62.09%	59.48%
World	58.17%	66.67%	64.05%

Panel B: OOS prediction

	Avg	PC	PLS	C-Enet	Enet	Combine
G6	3.77***	1.7**	4.14**	1.75*	-0.31	0.67**
G19	3.41**	1.59**	3.78**	2.05**	0.14	0.53**
World	3.48**	1.49**	3.67**	1.38*	-1.84	0.53**