Testing the Ruchir Sharma Framework on MSCI Emerging Markets

Belgutei Undur-Orgil (ubelgutei@proton.me)

ABSTRACT

This project applies Ruchir Sharma's *Breakout Nations* framework to the MSCI Emerging Markets index to identify which EMs exhibit the strongest structural conditions for sustained growth today, and how these conditions differ across countries. The objective is to move beyond the aggregate "Emerging Market" label by decomposing economic performance into seven pillars. We posit that sustainable outperformance is not random, but the result of specific structural conditions.

My interest in this project lies in a specific observation made during my exchange semester at the University of Seoul. While analyzing global equity benchmarks in *Special Issues in Business and Financial Accounting* course, I came across MSCI Emerging Market Index. Despite South Korea possessing infrastructure and technological advancement that often exceeds Western European standards, it didn't sit well with me that this country is classified alongside developing nations with vastly inferior structural metrics.

That contradiction made me question: If Korea is "Emerging", then what are the requirements? What makes one country structurally 'better' than another? To resolve this, I turned to Ruchir Sharma's framework laid in his book *Breakout Nations: In Pursuit of the Next Economic Miracles.* Unlike rigid index inclusion criteria, Sharma's approach assesses the structure of the country's economy such as demographics, debt, and productivity. This approach provides more nuanced lens to evaluate which "Emerging" markets are breaking out, and which (like Korea) may have already graduated in everything but name.

I. Data & Methodology

MSCI Inc. (2025) defines the market classification framework underlying the 24 countries used in this project. Countries are first mapped to ISO3 codes and then filtered to retain only those that can be consistently identified across both the World Bank and IMF datasets.

The analysis and ranking window is 2010-2024. In practice, the raw data extraction step has included earlier years when necessary to derive variables (for example, multi-year growth rates). However, all reported scoring, rankings and headline exhibits are computed within the 2010-2024 to maintain comparability across countries and avoid earlier-period data coverage irregularities.

Data construction follows a two-stage workflow because of personal preferences. The data preparation stage is implemented in Python, where raw series are pulled from source systems, cleaned, harmonized by ISO3 identifiers, reshaped into a country-year panel, and exported into a baseline dataset (e.g., a Parquet panel file). The scoring and reporting stage is done in R, where features are applied as needed, indicators are normalized, pillar scores are computed, composite breakout scores are formed under NA-values rules, and all tables and figures are then produced.

The model draws on two primary data sources. World Bank *World Development Indicators* (WDI) is assessed through API pulls. IMF *World Economic Outlook* (WEO) series are sourced through a bulk-download extract that is parsed into tidy country-year format and then merged with the main panel prior to scoring. In addition to these baseline sources, governance and financial inclusion datasets were cataloged but excluded for later improvements on the model.

The indicator framework is organized into seven pillars that represent distinct structural channels for growth and breakout potential inspired by Sharma's breakout-nations framework (Sharma, 2012). *Demographic* tailwinds and labor-force dynamics shape the long-run ceiling for growth and middle-class formation. *Fiscal* strength act as the "floor." Weak fiscal/macro fundamentals constrain sustained growth and amplify crisis risk. *Investment* depth proxy the "engine" of productivity growth and tradables capacity. *Foreign direct investment* signals capital inflows, technology transfer, and integration into global production networks. *Trade* as external openness and balanced demand influence scalability and resilience. *Commodity* dependence increases exposure to global cycle volatility and can weaken institutional/industrial diversification incentives. *Income* level anchors "where a country is" in its' development stage.

Indicators are measured in different units and ranges (e.g., inflation in %, reserves in USD, manufacturing share in % of GDP). Without normalization, indicators with larger numeric ranges would dominate mechanically. To compare "UAE vs. Brazil" on a consistent footing, we transform each indicator to a standardized score. Before standardization, each indicator is winsorized at the $1^{st}/99^{th}$ percentiles. This caps extreme observations that reflected crisis spikes and reporting anomalies.

Some indicators are "risk indicators" where higher values are structurally worse (e.g., inflation, debt burden, commodity dependence). For those indicators we applied:

$$x' = -x$$

This ensures that, after transformation, larger values consistently imply stronger structural conditions across all series.

For each indicator x, after winsorization and any sign flip, we computed a pooled z-score across all countries and all years:

$$z = \frac{x - \mu}{\sigma}$$

, where μ and σ are the sample mean and standard deviation computed over the full panel (all country-year observations with non-missing values). Each pillar score is computed as the simple average of its constituent z-scores:

$$P_{i,t}^{(k)} = \frac{1}{m_k} \sum_{j=1}^{m_k} z_{i,t,j}$$

, where $P_{i,t}^{(k)}$ is pillar k for country i in year t and m_k is the number of indicators assigned to pillar k. If some indicators are missing, the pillar score is averaged over available indicators.

The main composite index is the mean across available pillars:

$$S_{i,t} = \frac{1}{K} \sum_{k=1}^{K} P_{i,t}^{(k)}$$

We also compute a coverage-aware "penalized" version to reduce inflation from partial pillar coverage:

$$S_{i,t}^{pen} = S_{i,t} \cdot \left(\frac{pillars\ available}{K} \right)$$

In the headline comparisons we focus on fully covered country-years (7/7 pillars).

II. Findings & Discussion

Before analyzing individual national performance, we assessed the aggregate statistical properties of the Breakout Score model across the full panel (2010-2022). Table 1 summarizes the distribution of scores for our two primary model specifications: the "Naive" unweighted average and the "Penalized" structural score.

Table 1: Descriptive Statistics of Breakout Scores (2010-2022)

Specification	Mean (μ)	Std Dev (σ)	Interpretation
Naive	0.026	0.321	Raw average of available pillars.
Penalized	-0.004	0.271	Adjust for missing data and structural fragility.

The "Penalized" score, which serves as the basis for our final ranking, shows a mean near zero (-0.004) and tighter standard deviation (0.271) compared to the "Naive" model. This contradiction in volatility confirms that our penalization mechanism filters out noise, "shrinking" the outliers who score highly merely due to missing data.

Applying this penalized framework to the 2022 headline year, we classified the 23 MSCI Emerging Markets (excl. Taiwan) into four distinct tiers. This hierarchy shows a significant divergence between

"market consensus" and structural reality, specifically the shift in leadership from the traditional BRICS giants toward more industrial countries.

First, regarding the dependency on commodities and specifically "Curse of Oil", Sharma typically warns that resource-rich nations risk complacency, allowing currency overvaluation to hollow out industrial diversity. The United Arab Emirates defies this stereotype; while it benefits from Income score ($z=1.83$), its leadership is structurally supported by a massive Fiscal buffer ($z=1.03$) and, crucially, positive Demographic momentum ($z=0.61$). Unlike the stagnant rentier states often critiqued in development economics, the UAE's data reflects a successful pivot toward a diversified service economy.

Furthermore, a central search in Breakout Nations is identifying the successor to China's manufacturing dominance. Our result identifies Malaysia as the empirically correct answer, presenting the most balanced structural profile in the index with positive contributions across ($z=0.87$), FDI ($z=0.43$), and Investment ($z=0.17$). Malaysia is actively absorbing the "China Plus One" supply chain reconfiguration. Unlike other contenders that lack the necessary infrastructure or political stability, Malaysia's scores indicate it possesses the structural hardware required to capitalize on global trade shifts.

The case of South Korea represents an interesting validation of the Breakout Nations thesis, illustrating how structural excellence can overwhelm demographic adversity. The author of the framework cites Korea as "gold medalist" that established "a class by itself" due to its rare ability to stay at the cutting edge of fast-changing industries (Sharma, 2012). South Korea's high ranking is primarily driven by its Investment pillar ($z = +1.41$), a score nearly three standard deviations above the emerging market mean, similarly how Sharma sees Korea has become the "Germany of Asia" as the global standard for high-end manufacturing powerhouse. This massive positive score, driven by high R&D and capital formation represents the accumulated effects of policies that prioritize export-oriented manufacturing sector. Conversely, the model also shows the long-term headwind described by Sharma: the nation's rapidly aging population and low fertility rates. However, the overall positive score demonstrates that the high productivity of Investment pillar is powerful enough to offset this structural ceiling. Structural health of the Korean economy reflects a "Graduated Nation", one that has succeeded in moving beyond the reliance on labor-force growth (the core EM model) to one sustained by capital-deepening and technological specialization. The model suggests that the external classification of Korea within the MSCI EM index reflects market criteria rather than its internal, world-class structural economy.

Special cases must be made with countries including India as this country is explicitly framed as a 50/50 breakout call, because the demographic story is conditional and can be derailed by policy and political economy. He argues that India's elite narrative leans too heavily on "demographic dividend" while underweighting the hard constraints; educating the young, creating enough productive jobs, and

moving labor into higher-productivity urban employment rather than leaving it stuck in low-productivity agriculture. That conditionality is why, for the demographic pillar we chose to emphasize momentum and absorption capacity rather than a static snapshot of the age pyramid. This model used growth-type measure derived from the working-age level series and paired with female labor force participation and urbanization dynamics. In other words, we tried to measure not just whether India has a lot of working-age people, but whether it is still getting a meaningful incremental demographic push and whether that push is likely to translate into output through participation and structural transformation. India can have a large working-age cohort and still not score strongly if the incremental working-age growth is not exceptional in percentage terms and if participation/absorption proxies are weak relative to peers. Female labor force participation is the most common driver here: across EM peers, countries that score well on the demographic pillar often do so because they combine demographic momentum with high participation and/or fast structural transformation. If India is weak on participation relative to the sample, that can pull the combined demographics pillar slightly below zero even if the working-age population itself is rising. In a composite built from z-scores, "slightly negative" is often the signature of "one strong element, one weak element, net near the middle."

He also flags specific failure modes: welfare-state drift that pushes deficit and debt higher, and crony capitalism that worsens "rules vs. connections" environment for investors. The two biggest "drags" of India are Fiscal ($z=-0.58$) and Income ($z=-0.15$), with secondary drags from FDI ($z=-0.29$) and Trade ($z=-0.15$). Sharma warns that deficits were expanding quickly and that debt had reached levels that are high for a major developing country. On FDI and investment quality, Sharma's argument is also strikingly aligned with our direction: he describes foreign investment as "well below the required totals" and worries that weak governance and graft push capital away from productive domestic deployment.

In my current setup, higher the GDP per capita mechanically improves the score, so a low level becomes a large negative z-score (India's -0.9). But Sharma treats low income as runway and India's per-capita income being far below China's is one reason he sees more "room to grow". In other words: our model is currently using income level partly as a proxy for institutional/structural maturity, while Sharma uses it partly as a proxy for catch-up potential. Linking these perspectives clarifies why India is one of the special cases. If income is treated as a monotonic "higher is better", the model will systematically penalize lower-income economies even when their core appeal is precisely the size of their convergence gap; if income is treated as a non-linear variable or moved outside the composite as a conditioning descriptor, the framework becomes closer to Sharma's intended logic. Therefore, we can already see potential improvements that can be made to this model.

Another interesting observation is regarding Czechia. This country has ranked among the stronger structural cases because its non-demographic pillars such as solid investment and industrial depth,

meaningful trade integration and a macro profile that while not uniformly strong on every subcomponent does not trigger the instability flags that dominate the lowest-ranked countries. This pattern is quite consistent with the qualitative characterization in Sharma's European discussion, where Czechia is grouped with Poland as part of the region's "sweet spot" in which institutional discipline and external integration make convergence plausible.

Figure 1: Pillar Heatmap by Countries
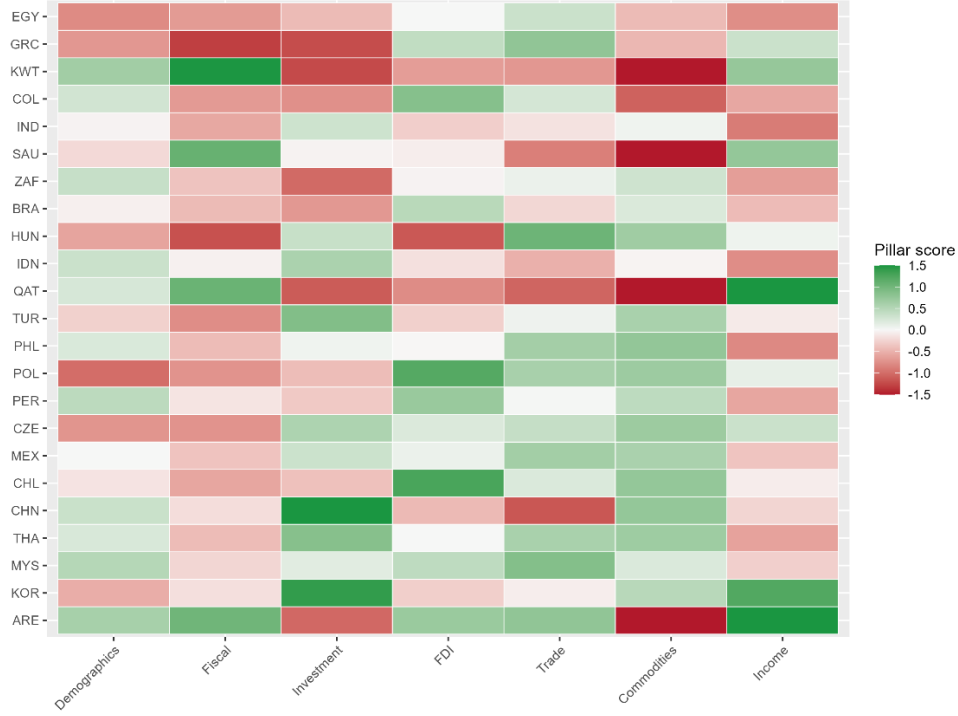
When we reconfigured the demographic component, Czechia's composite score fell primarily because the Demographics pillar itself deteriorated sharply, while other pillars remained essentially unchanged. In particular, the Demographics pillar in the first iteration placed weight on the working-age population level series (*SP.POP.1564.TO*). Because this is a level variable, cross-sectional standardization (z-scoring) makes it behave as a proxy for country scale: large-population EMs mechanically score high and smaller EMs mechanically score low, irrespective of whether their demographic dynamics are favorable in the sense relevant to breakout theory. Czechia's working-age population level is structurally small relative to the MSCI EM peer group, so once that level enters the pillar as a scored component, it exerts a persistent negative pull on Czechia's Demographics score and on the composite breakout score.

This finding matters for interpretation because it highlights a broader identification issue in empirical "breakout" scoring: demographic *momentum* is conceptually distinct from demographic *scale*. Sharma's narrative does not treat Czechia as attractive because it has a large workforce rather it treats the country's attractiveness as stemming from institutional quality and EU-linked integration that enable steady growth despite limited demographic scale. When the Demographics pillar inadvertently rewards

scale, the model becomes biased against small but structurally strong economies. In this sense, Czechia's decline in ranking is informative: it signals that the demographic specification is sensitive to whether the pillar is capturing the intended mechanism (tailwind and absorption) or an unintended correlation (population size).

The key implication is that Czechia's weaker result should not be read as evidence against its structural convergence thesis and instead, it motivates a refinement in the demographic measurement strategy. Payoff is not that the first version of our project "gets the ranking right," but that it makes disagreements interpretable: when the book and the model diverge (India being the clearest case), the divergence can be decomposed into measurable drivers (fiscal drag, external position, investment/FDI integration, and how "income level" is treated), rather than remaining a purely narrative debate.

This is also where it becomes productive to pivot from "country scoring" to "equity opportunity": developed markets tend to be information-saturated and heavily covered, so mispricing is harder to sustain, whereas emerging markets can combine adequate structural fundamentals with thinner information, weaker transmission, and segmented investor bases - conditions under which valuation gaps can persist. The logic is that to use breakout-score framework as a screen for "where the macro floor and ceiling are not obviously broken," then shift the research burden toward equity-market mechanisms based on coverage, governance, investor protection, liquidity, ownership structures, and information environment that can help us spot undervalued opportunities. Korea is the clean motivating example: it can look structurally strong on several macro dimensions while still trading at a discount that is better understood through geopolitical risks.

References

Sharma, R. (2013). "Breakout nations: In pursuit of the next economic miracles" (with a new epilogue). W. W. Norton & Company.

World Bank. (n.d.). World Development Indicators: [Data set]. World Bank DataBank. Retrieved December 03, 2025

International Monetary Fund. (n.d.). World Economic Outlook (WEO) [Data set]. IMF Data. Retrieved December 03, 2025

Table 2: The 2022 Breakout Hierarchy

Rank	Country	Score	Tier	# Pillars > 0	Demographics	Fiscal	Investment	FDI	Trades	Commodities	Income
1	United Arab Emirates	0.313	Tier 1	5	0.61	1.03	-1.04	0.71	0.78	-1.73	1.83
2	South Korea	0.291	Tier 2	3	-0.54	-0.18	1.41	-0.29	-0.07	0.49	1.22
3	Malaysia	0.24	Tier 1	5	0.5	-0.24	0.17	0.43	0.87	0.22	-0.29
4	Thailand	0.19	Tier 1	5	0.24	-0.43	0.86	0	0.6	0.68	-0.62
5	China	0.174	Tier 2	3	0.35	-0.19	2.15	-0.45	-1.16	0.75	-0.25
6	Chile	0.15	Tier 2	3	-0.14	-0.59	-0.39	1.28	0.22	0.76	-0.08
7	Mexico	0.133	Tier 2	5	0.01	-0.37	0.34	0.09	0.64	0.59	-0.37
8	Czechia	0.108	Tier 2	5	-0.72	-0.73	0.56	0.22	0.39	0.69	0.35
9	Peru	0.082	Tier 2	4	0.45	-0.14	-0.33	0.72	0.02	0.44	-0.59
10	Poland	0.066	Tier 2	4	-1	-0.73	-0.42	1.21	0.61	0.69	0.12
11	Philippines	0.064	Tier 2	4	0.23	-0.43	0.06	-0.01	0.64	0.76	-0.8
12	Turkey	-0.050	Tier 2	3	-0.28	-1.25	0.9	-0.29	0.07	0.6	-0.1
13	Qatar	-0.084	Tier 2	3	0.25	1.07	-1.13	-0.78	-1.06	-2.23	3.28
14	Indonesia	-0.087	Tier 2	2	0.35	-0.05	0.58	-0.17	-0.51	-0.03	-0.77
15	Hungary	-0.116	Tier 2	4	-0.61	-1.2	0.37	-1.15	1.04	0.67	0.07
16	Brazil	-0.164	Tier 2	2	-0.06	-0.44	-0.69	0.48	-0.23	0.22	-0.43
17	South Africa	-0.187	Tier 2	3	0.37	-0.38	-1.02	-0.04	0.1	0.32	-0.65
18	Saudi Arabia	-0.202	Tier 2	2	-0.22	1.09	-0.04	-0.07	-0.88	-2.05	0.75
19	India	-0.226	Tier 3	2	-0.04	-0.58	0.32	-0.29	-0.15	0.06	-0.9
20	Colombia	-0.239	Tier 4	3	0.29	-0.68	-0.75	0.86	0.27	-1.08	-0.58
21	Kuwait	-0.288	Tier 3	3	0.65	1.67	-1.24	-0.65	-0.7	-2.49	0.75
22	Greece	-0.332	Tier 3	3	-0.7	-1.48	-1.23	0.42	0.78	-0.47	0.34
23	Egypt	-0.395	Tier 4	2	-0.79	-0.68	-0.44	0	0.34	-0.44	-0.77