Beyond Capital: How Financial Inclusion, Governance, and Gender Gaps Shape Global Startups

Swaraj Rai

Quantitative Analysis Center, Wesleyan University

Certificate in Applied Data Science

Professor Maryam Gooyabadi

April 9, 2025

2. Abstract

This study investigates how institutional quality, financial inclusion, and gender equity influence entrepreneurial activity across countries. Using a panel dataset covering 125 countries over four time points (2011–2021), the analysis employs multivariate OLS, fixed and random effects panel models, mixed-effects modeling, and both static and temporal clustering techniques. Financial access and governance quality consistently emerged as robust predictors of formal business formation across countries. The Women's Entrepreneurship Feasibility Score (WEFS), a novel composite index developed in this study, demonstrated strong within-country explanatory power—highlighting the critical role of gender-inclusive legal and economic systems. Clustering analyses revealed distinct ecosystem profiles and startup trajectories, offering deeper insight into structural heterogeneity. Policy implications include strengthening governance, expanding digital financial infrastructure, and implementing targeted reforms to reduce gender-based barriers in entrepreneurship ecosystems.

3. Introduction

In an era of global economic uncertainty, entrepreneurship has emerged as a critical driver of job creation, innovation, and resilience. Particularly in low- and middle-income countries (LMICs), new business formation serves as a mechanism for absorbing youth labor, fostering inclusive growth, and catalyzing structural transformation. Yet, despite widespread recognition of its importance, the predictors of entrepreneurial activity remain contested and under-theorized. Why do some countries consistently generate high startup activity while others lag behind, even after accounting for macroeconomic fundamentals?

This study investigates the institutional, final, and demographic drivers of formal entrepreneurship across countries using a cross-national panel of 125 countries between 2011 and 2021 and employing an interdisciplinary, data-driven approach. Specifically, it asks: *What are the most robust predictors of new business density at the national level, and how can they be modeled using time series techniques and structural regressions?* Initially employing Granger causality and ARIMAX methods, limitations due to the short temporal depth led to alternative methods that will be later explored. Attempts to implement Arellano-Bond and Correlated Random Effects models were also constrained by singular matrix errors due to limited within-country variation. As a result, the study pivots to robust fixed effects, random effects, and mixed effects models—alongside clustering techniques—to identify both predictive variables and country-level patterns in entrepreneurial ecosystems. While a dynamic pooled OLS model was explored to capture persistence in startup activity, its simplifying assumptions and structural limitations led to its de-emphasis in the final analysis.

This paper aims to contribute to a growing body of literature that bridges the gap between economic theory and empirical modeling of entrepreneurship ecosystems.

The focus on *new business density*—defined as the number of newly registered limited liability companies per 1,000 working-age adults—provides a measurable, comparable indicator of entrepreneurial activity across time and countries. While prior studies often rely on macroeconomic or firm-level data, this research adopts a blended framework that incorporates both institutional and demographic predictors.

Of particular interest are variables related to governance quality, gender-based labor disparities, financial inclusion, and government investment in education—each of which represents a potential lever for policymakers.

This study centers three key hypotheses:

1. **Financial access enables entrepreneurship**: Broader access to formal financial services increases the likelihood of business formation by reducing capital constraints and risk exposure.

2. **Governance quality facilitates market entry**: Effective institutions—measured by regulatory quality, rule of law, and political stability—reduce barriers to entry and improve the ease of doing business.

3. **Gender-inclusive ecosystems drive business density**: Societies that reduce structural gaps in labor force participation and invest in women's entrepreneurial capacity will experience more robust startup activity.

The original plan, to test these hypotheses, was to utilize a hybrid methodological approach. By which, first a Granger causality test was attempted to assess short-term predictive relationships between exogenous variables and new business density. Followed by a structural ARIMAX model, which would have hoped to help evaluate the magnitude and direction of these effects within a time series framework. Finally, multiple OLS regressions with exogenous controls were planned to validate and refine these relationships across different model specifications. The last specification (multiple OLS regressions) was the only successful methodological approach and the core focus of this paper - given the limitations imposed by the data (limited temporal depth). Clustering techniques—both static (feature-based) and temporal (via Dynamic Time Warping)—complement the regression results by revealing behavioral groupings across countries that transcend geography or income group. The results consistently highlight the predictive strength of WEFS, governance, and financial inclusion, offering evidence that entrepreneurship is shaped not only by capital, but by the systems of access, equity, and policy that determine who gets to participate.

The dataset compiled for this study spans 125 countries and includes variables from the World Bank, Global Findex, and proprietary entrepreneurship indicators like the Women's Entrepreneurship Feasibility Score (WEFS). The panel includes both high-income and developing countries, allowing for cross-comparison. All variables were tested for stationarity, imputed where necessary, and normalized to ensure comparability.

The value of this research lies not only in its technical rigor but also in its policy relevance. As governments and international institutions seek scalable solutions to unemployment, especially among youth and women, understanding the levers of entrepreneurial activity becomes paramount. By highlighting the predictive power of financial inclusion, gender equity, and institutional quality, this study offers actionable insights for policymakers, development agencies, and ecosystem builders.

In sum, this paper aims to contribute both to academic literature on entrepreneurship and to the practical understanding of how institutional and demographic drivers can be leveraged to foster inclusive economic growth. The findings underscore the importance of thinking beyond capital and credit—toward the systemic enablers that allow new businesses to emerge, survive, and scale.

4. Literature Review

Entrepreneurship has increasingly been recognized as a vital engine for economic development, particularly in emerging and low-income economies. It serves as a mechanism for job creation, innovation, and resilience in contexts where traditional sectors may fail to absorb labor market entrants. Scholars such as Naudé (2010) emphasize the role of entrepreneurship in facilitating inclusive growth and institutional flexibility, while Acs, Desai, and Hessels (2008) argue that differences in institutional quality explain much of the variance in entrepreneurial activity across countries. Despite broad consensus on its macroeconomic significance, the drivers of entrepreneurship—especially at the national level—remain contested and under-theorized.

A growing literature has focused on the role of financial inclusion as a catalyst for entrepreneurship. Access to savings, credit, and digital payments reduces liquidity constraints and facilitates risk-taking, particularly among marginalized groups. The World Bank's Global Findex Database 2021 reveals that increases in formal financial access are strongly associated with productive investment behaviors, especially in low-income and rural populations (Demirgüç-Kunt et al., 2022). Beck, Demirgüç-Kunt, and Levine (2007) further demonstrate that financial system depth not only fosters firm creation but also reduces income inequality by enabling broader participation in the formal economy. However, the impact of financial inclusion may be conditional. As some scholars argue, capital access alone is insufficient when not accompanied by institutional reforms and targeted capacity-building (Banerjee & Duflo, 2011).

Institutional quality also plays a central role in enabling entrepreneurship. North (1990) famously defined institutions as the "rules of the game" in a society, arguing that predictable and efficient institutions lower transaction costs and promote innovation. Empirical work supports this view. Aidis, Estrin, and

Mickiewicz (2008), using comparative panel data, find that governance indicators—such as regulatory quality, rule of law, and political stability—are positively associated with small and medium enterprise (SME) formation. These findings are echoed by World Bank cross-country analyses showing that weak institutions correlate with higher informality and lower business density (World Bank, 2023). Nonetheless, institutions are slow to change and politically difficult to reform. This makes it all the more urgent to understand how institutional variables interact with other factors such as gender equity, education, and financial infrastructure.

The gendered dimensions of entrepreneurship are increasingly attracting scholarly and policy attention. Women face compounded barriers in starting and growing businesses, ranging from limited access to finance and education to discriminatory social norms and legal constraints. Minniti and Naudé (2010) emphasize that female entrepreneurs tend to operate in smaller, lower-growth sectors and face disproportionate administrative burdens. Elam et al. (2021), drawing on the Global Entrepreneurship Monitor, find that women's entrepreneurial activity is highly sensitive to ecosystem-level factors such as mentorship, childcare infrastructure, and societal support. The Women's Entrepreneurship Feasibility Score (WEFS), a composite indicator designed to capture the gendered context of business formation, has emerged as a promising variable in this space. Preliminary studies indicate that higher WEFS scores predict a significant increase in new women-led business registrations across LMICs (GEM, 2021).

Education and human capital investment represent another important pillar of entrepreneurship research. Building on Becker's (1964) theory of human capital, researchers have consistently found a positive correlation between education and entrepreneurial outcomes. Van der Sluis, Van Praag, and Vijverberg (2008), in a review of over 100 empirical studies, conclude that educational attainment increases both the likelihood of starting a business and the survival rate of startups. However, the relationship is non-linear and context-dependent. In some settings, higher education leads to stronger preferences for wage employment, especially in public sector–dominated economies (Amorós & Bosma, 2014). Moreover, the returns to education are mediated by gender and region, suggesting the need to analyze these interactions more carefully.

While these dimensions—financial inclusion, institutional quality, gender equity, and education—have been widely studied, most existing research examines them in isolation. A large portion of the literature relies on cross-sectional or firm-level data, often neglecting the dynamic, macro-level relationships that evolve over time. Crucially, few studies explore how these drivers interact or change in predictive significance across countries and years. The use of time series methods such as Granger causality testing and ARIMAX modeling remains rare in entrepreneurship research, despite their ability to distinguish between short-term forecasting power and long-run structural influence. This study addresses that gap by integrating temporal and structural analysis in a unified empirical framework, providing new insights into how institutional and demographic factors shape national entrepreneurship ecosystems over time.

5. Data & Methodology

1. Data Sources

This study draws on a cross-country panel dataset spanning the years 2011, 2014, 2017, and 2021. The unit of analysis is the country-year. The primary dependent variable is **New Business Density Rate (NBDR)**, defined as the number of new limited liability company (LLC) registrations per 1,000 adults aged 15–64, sourced from the World Bank's Entrepreneurship Database.

Independent variables were selected based on theoretical relevance, empirical precedent, and data availability. These include:

1. **Financial Access**: Share of adults (15+) with an account at a financial institution or mobile money service, sourced from the Global Findex Database. Serves as a proxy for credit and financial service access.

2. **Governance Quality**: World Bank's "Control of Corruption" estimate, used as a proxy for institutional quality. The Worldwide Governance Indicators (WGI) project assesses governance across six key dimensions:
    a. Voice and Accountability
    b. Political Stability and Absence of Violence/Terrorism
    c. Government Effectiveness
    d. Regulatory Quality
    e. Rule of Law
    f. Control of Corruption

3. **Education Investment**: Government spending on education as a percentage of GDP, reflecting long-term investment in human capital.

4. **Female Population Share** and the **Labor Force Participation Gender Gap**[1] (male minus female LFP rate), from World Bank indicators: Calculated as the male LFP rate minus the female LFP rate. A proxy for structural gender inequality in access to economic opportunity.

[1] Utilized to craft WEFS metric

5. **GDP (constant 2015 USD) and Population Growth**: Included as Macroeconomic controls.

6. **Dominant Employment Sector**: The dominant sector of employment in a given country-year entry. Classified into either agriculture, industry, and services.

7. **Informal Economy Share**[2]: Percentage of GDP derived from informal activities, using ILO estimates. High informality can hinder formal entrepreneurship.

8. **Entrepreneurship Score**[3]: A composite measure from the World Bank's *Women, Business and the Law* dataset, capturing gender parity in business-related legal frameworks.

While other variables (e.g., Ease of Doing Business, Gini Index, Literacy Rates) were initially considered, they were excluded due to excessive missingness. After cleaning and filtering for complete data, the final panel includes 391 observations across 125 countries, with 82 countries having three or more years of data, enabling limited panel analysis.

2. Construction of the Women's Entrepreneurship Feasibility Score (WEFS)

To operationalize gendered feasibility in entrepreneurship, this study constructs a novel index called the **Women's Entrepreneurship Feasibility Score (WEFS)**. Unlike additive composites, WEFS captures structural constraints via a penalty-adjusted formula:

[2] Utilized to craft WEFS metric
[3] Utilized to craft WEFS metric

$$WEFS = \frac{Entrepreneurship\ Score \times (1 - \frac{LFP_{gap}}{LFP_{male}})}{Informal\ Economy\ \%}$$

The WEFS metric synthesizes three core components::

1) **Entrepreneurship Score** is sourced from the *World Bank's Women, Business and the Law* dataset and represents a composite score (0–100) based on legal frameworks that affect women's ability to start and run a business. Specifically, it aggregates responses to questions such as:

1. Can a woman legally register a business in the same way as a man?
2. Are there gender-based restrictions on obtaining a business license?
3. Are there laws prohibiting gender-based discrimination in access to credit or property ownership?
4. Are there equal rights for women to sign contracts, register assets, or appear in court?

The higher the score, the more legally "feasible" it is for women to pursue entrepreneurship in that country, with 100 representing full legal parity.

2) $|\mathbf{LFP_{female}} - \mathbf{LFP_{male}}| / \mathbf{LFP_{male}}$

- Where $\mathbf{LFP_{female}} - \mathbf{LFP_{male}}$ represents $\mathbf{LFP_{Gender\ Gap}}$ - this term represents the **normalized gender gap in labor force participation**, capturing structural inequities in economic participation. A larger gap (i.e., lower female LFP relative to male LFP) reduces the

WEFS score, penalizing environments where women are systematically excluded from the workforce.

3) Informal Economy Share (% GDP): This inclusion came from the recognition that female-led businesses have empirically been shown as underrepresented in official census estimates globally (Elam et al. (2021).

This formulation ensures that legal parity is penalized in environments where women are economically excluded or where informality dilutes the ability to operate formally. The result is a scaled, comparative index that better reflects real-world feasibility, especially in LMICs where gendered entrepreneurship data is sparse. The WEFS was normalized and included as a core predictor in all major regression and clustering models.

3. Methodological Approach

Initially, time-series models such as Granger causality and ARIMAX were attempted to analyze short-term predictive relationships and structural dynamics. However, due to the dataset's limited temporal depth - offering only 4 observations per country-year combination at most - these approaches proved to be infeasible. Time-series diagnostics revealed insufficient data for stationarity testing and model stability, prompting a pivot away from stochastic time series techniques.

The methodological framework was subsequently revised to focus on regression and clustering techniques better suited to short-panel data. Several econometric models were implemented to identify robust predictors of new business density, each offering a different lens on the data:

1. Multivariate Ordinary Least Squares (OLS):

Served as a foundational model to estimate the direction and statistical significance of core predictors, including financial access, governance quality, and gender equity (WEFS).

2. Fixed Effects (FE) and Random Effects (RE):

These panel models were used to account for unobserved heterogeneity across countries. The FE model captured within-country change, while the RE model allowed for both within- and between-country variation. A Hausman test indicated that the RE model was preferred, but results were triangulated across both specifications.

3. Mixed-Effects (Hierarchical) Model:

Implemented as a random-intercepts model with country-level variation, this approach confirmed the predictive significance of key variables while accounting for nesting across space.

4. Region-Level Fixed Effects (Region FE):

An alternative model specification controlling for sub-regional heterogeneity provided a broader spatial lens, particularly in the absence of sufficient within-country variation.

5. Stepwise AIC OLS:

Applied to optimize variable selection through the Akaike Information Criterion (AIC), thus aiding in model parsimony and addressing potential multicollinearity.

6. Dynamic Pooled OLS (Log-Linear with Lag):

Initially explored to capture entrepreneurial momentum through a lagged dependent variable, this model offered insight into persistence effects. However, due to its simplifying assumptions—such as homogeneity in startup momentum across countries—it is interpreted cautiously and not emphasized in the final analysis.

7. Clustering Techniques (Static and Temporal):

To complement regression-based inference, two forms of clustering were applied. Feature-based static clustering grouped countries based on summary statistics of their NBDR behavior, while Dynamic Time Warping (DTW) clustering grouped them by the trajectory of startup activity over time.

Finally, exploratory efforts to implement Correlated Random Effects (CRE/Mundlak) and Arellano-Bond GMM estimators were conducted. However, both models proved computationally infeasible due to the short time dimension (T = 3–4) and limited within-country variation, resulting in matrix singularity errors. These are noted in the limitations section.

Together, this multi-model framework enables a robust examination of entrepreneurial activity drivers, with convergence across RE, FE, and mixed-effects models pointing to the centrality of financial inclusion, governance quality, and gender equity (via the WEFS metric) in explaining new business formation globally.

4. Limitations of the Methodology

While the methodology employed in this study offers valuable insights into the structural factors influencing business formation, several limitations merit acknowledgment:

1. **Limited Temporal Depth**

   The dataset covers only a small n number of years for most countries (typically between 2-4), which constrains the reliability of longitudinal modeling and weakens the power of time-series techniques such as Granger causality. A longer panel would enhance the ability to detect dynamic effects and structural breaks over time, allowing time series analysis. Until then, the models outlined in the previous section and further expounded in the results section will suffice.

2. **Data Availability and Selection Bias**

   In building upon the limited data at hand, mentioned previously, it also needs to be acknowledged that not all countries report consistently. As a result, the sample is unbalanced and restricted to countries with at least four years of complete data, which may introduce selection bias. The resulting analysis may overrepresent countries with better data infrastructure, typically higher-income or institutionally stronger states, skewing generalizability.

3. **Endogeneity and Omitted Variable Bias**

   Despite efforts to include institutional, financial, and demographic predictors, the complex ecosystem in which businesses are formed likely includes unobserved factors—such as political shocks, informal sector dynamics, or local cultural norms—that were not captured. These omissions may introduce endogeneity, where causality flows both ways or is mediated by latent variables.

Efforts to implement Correlated Random Effects (CRE) using the Mundlak correction, and Arellano-Bond (AB) dynamic panel estimation were undertaken to address potential endogeneity and lag dependence. However, both models failed due to computational singularity and matrix instability—attributable to the dataset's short panel length (T = 3–4 years) and insufficient within-country variation. These limitations constrained the ability to model dynamic or endogenous relationships robustly.

4. **Measurement and Construct Validity Issues**

   Several key variables—such as governance scores, financial access metrics, and the WEFS index—are derived from survey data, expert opinions, or composite formulations. These measures are prone to subjectivity. Additionally, while indices like WEFS are innovative and attempt to capture multidimensional barriers, their aggregation methods may mask important nuances or reinforce dominant narratives. Country-level metrics also obscure within-country disparities and legal or cultural variation, which can significantly influence business formation but remain unobserved in this analysis.

The limited sample size and panel depth also prevented the use of cross-validation or holdout sample testing, which would have strengthened model reliability. As such, internal model diagnostics are the primary basis for evaluating fit and robustness.

5. **Granger Causality and Causal Inference**

   Granger causality was initially attempted to test short-term predictive relationships, but the method was abandoned early in the project due to insufficient temporal observations and non-stationarity. More generally, while Granger causality can establish temporal ordering, it does not confirm causation and is sensitive to omitted variable bias. Without instrumental variables or natural experiments, causal inference remains outside the scope of this study.

6. **Sectoral and Legal Variation**

   The study uses national-level business density as the dependent variable, but the definition, legal form, and sectoral distribution of businesses differ widely across countries. An LLC being the proxy used for new business formation insinuates that what is in one country may differ in definition to another. This heterogeneity in definitions limits comparability.

Despite these constraints, the methodological framework remains a meaningful starting point for identifying patterns in global entrepreneurship formation. By combining econometric modeling with composite measures like WEFS, this study contributes to a growing body of research emphasizing the importance of inclusive institutional environments in shaping entrepreneurial ecosystems. Future research should aim to build on this work by incorporating longer time series horizons, implementing causal identification strategies such as instrumental variables or natural/quasi-natural experiments, and exploring sub-national heterogeneity to refine the understanding of entrepreneurial ecosystems.

6. Empirical Results

6.1 Summary Statistics and Correlation Diagnostics

Exploratory analysis revealed wide cross-country variability in new business density rates (NBDR), with clear differences by region and income group. In analyzing NBDR over time, as seen on *Figure 3*, findings suggest a recovery after the 2008–2010 financial crisis, a dip in 2014–2017, and a post-COVID spike by 2021.

Figure 2: Log-transformation was applied to normalize the distribution for regression analysis. Outliers present illustrate the need for Cook's Distance to isolate outliers in Deterministic modeling.

Figure 1: A correlation matrix confirmed moderate associations between financial access, WEFS, and governance. Variance Inflation Factors (VIFs) for all predictors were below 5, indicating no significant multicollinearity.

6.2 Single, Multivariate, and Stepwise OLS

Table 4: Baseline OLS regressions confirmed that:

- **Financial access** ($p < 0.001$),
- **Governance quality** ($p < 0.001$), and
- **WEFS** ($p \approx 0.088$)

were all positively associated with business density.

After excluding influential outliers (*Figure 7*) (e.g., Australia, South Africa, Kosovo), WEFS gained strength ($p = 0.006$), as seen on *Table 7*, reinforcing its role as a key structural variable.

Table 5: Stepwise AIC regression retained WEFS, governance, and financial access, while GDP, education spending, and demographic controls were excluded.

6.3 Panel Models: Fixed, Random, and Mixed Effects

Table 5: To account for panel structure, both Random Effects (RE) and Fixed Effects (FE) models were estimated. A Hausman test ($p > 0.5$) suggested RE was consistent and preferred, but results from both models were reported.

- **RE model:**
  - Financial Access ($\beta = 2.82$, $p = 0.001$)
  - Governance ($\beta = 0.91$, $p = 0.026$)
  - WEFS ($\beta = 0.66$, $p = 0.005$)

- **FE model:**
  - WEFS ($\beta = 1.73$, $p = 0.004$) became stronger
  - Governance and financial access lost significance, indicating they explain more **between-country** than **within-country** variation.

- **Mixed-effects model** (random intercepts by country):
  - Confirmed results from RE model
  - Intra-class correlation (ICC ≈ 0.80) indicated strong country-level clustering in entrepreneurial outcomes

6.4 Region-Level Fixed Effects

Table 5: To control for broad geographic patterns, a model with sub-region fixed effects was estimated. South Asia and Sub-Saharan Africa showed significantly lower baseline startup activity ($p < 0.001$), even

after controlling for structural predictors. Governance and WEFS remained significant in this model as well.

6.5 Dynamic Panel OLS (Lagged Dependent Variable)

Table 5: A Dynamic Pooled OLS model was explored using lagged log-transformed NBDR. The lag term was significant ($p < 0.001$), confirming “entrepreneurial momentum.” However, due to its simplifying assumptions (e.g., uniform momentum across countries) and lack of country-level error terms, this model is reported for completeness but not emphasized in final interpretation.

6.6 Clustering Analysis

Two complementary clustering techniques were employed:

(a) Static Clustering (K-Means on NBDR Features):

Figures 10, 11, 12, 13: Countries were grouped into five clusters based on NBDR characteristics: average level, slope, variance, range, autocorrelation.

- Cluster 1 (e.g., Australia): High NBDR, strong institutions
- Clusters 3 & 5: Underperformers with low WEFS and governance
- Radar charts visualized standardized cluster profiles

(b) Temporal Clustering (DTW with K-Medoids):

Figures 14, 15, 16, 17: This method grouped countries based on their NBDR trajectories over time.

- Cluster 4 (UAE, Australia): High and consistent growth
- Cluster 1 (many SSA countries): Flat or declining trends
- Cluster assignment correlated with structural features (e.g., WEFS, governance)

Together, these clustering techniques provided deeper insight into how countries with similar structural traits or growth patterns behave over time.

6.7 Forecasting (Directional)

Figure 18: Using the Dynamic Pooled OLS model, one-year-ahead forecasts were generated and visualized via choropleth maps. These predictions were treated directionally rather than precisely due to model limitations. Countries forecasted to perform well included Western Europe and East Asia. Low predicted NBDR levels persisted across South Asia and Sub-Saharan Africa.

Outliers were filtered using ±2 standard deviations to improve interpretability.

6.8 Key Takeaways

1) *Table 5:* Across all valid models—OLS, RE, FE, and Mixed Effects—three variables consistently emerged as robust predictors of entrepreneurial activity:

- **Financial Access**: A strong enabling factor, particularly in RE and Mixed models
- **Governance Quality**: Significant in all specifications, though more important between countries

- **WEFS**: Gained strength post outlier exclusion, remained robust across all model types

2) *Figures 10-17*: Clustering confirmed that structural similarity, not geography or income group, best explains entrepreneurial ecosystem behavior.

3) *Table 5 and Figure 18*: While Dynamic Pooled OLS showed strong fit, its simplifying assumptions make FE and Mixed models more defensible.

7. Discussion

The empirical results of this study reinforce the central role of institutional quality, financial inclusion, and gender equity in shaping entrepreneurial ecosystems. Across OLS, panel regressions, mixed-effects models, and clustering techniques, these three factors consistently emerged as the strongest predictors of formal business formation. Their stability across model types and subsamples strengthens confidence in their importance, particularly in policy and development contexts.

Notably, the Women's Entrepreneurship Feasibility Score (WEFS) demonstrated strong and consistent effects—especially in fixed-effects and mixed models. Its enhanced significance after outlier exclusion, and its ability to predict within-country variation over time, suggests that gender-inclusive legal and structural environments matter not only across countries, but within them as they evolve. This highlights the value of multidimensional proxies that go beyond traditional macroeconomic indicators.

In contrast, governance quality and financial access exhibited stronger explanatory power in random effects models, which capture cross-country differences. This supports the idea that institutional strength and financial infrastructure are foundational drivers of startup ecosystems, but are often slower to shift within countries over short time horizons. These findings collectively point to the importance of tailoring interventions: countries with low WEFS may benefit from targeted legal reform, while those with institutional weaknesses may require more fundamental systemic rebuilding.

Attempts to implement more sophisticated dynamic panel models, such as Arellano-Bond (AB) and Correlated Random Effects (CRE), were ultimately unsuccessful due to computational singularity and insufficient panel length ($T = 3–4$). While this limits causal inference and prohibits deeper investigation into endogeneity, the convergence across models that were feasible provides strong descriptive and predictive evidence.

These findings contribute meaningfully to the theoretical literature on entrepreneurship development by emphasizing several key insights:

1. **Entrepreneurial ecosystems are structurally embedded.** Entrepreneurship emerges from a systemic interaction of financial access, legal environment, and institutional trust—not from capital access alone. Countries must build comprehensive support structures rather than siloed reforms.

2. **Within-country dynamics matter.** The stronger performance of WEFS in fixed-effects models suggests that internal reforms—particularly around gender equity—can shift entrepreneurship rates over time, independent of baseline economic conditions

3. **Cross-country variance is structurally driven.** Differences in governance and financial inclusion explain a large share of the cross-national variation, emphasizing the need for context-sensitive models and avoiding "one-size-fits-all" policy solutions.

4. **Entrepreneurial momentum is real, but caution is warranted.** Whilst dynamic pooled OLS did highlight temporal persistence in startup rates, the model's simplifying assumptions limit its reliability. Nonetheless, the notion of path dependence remains theoretically plausible and merits deeper study in future research with longer time horizons available.

5. **Gender-informed modeling improves predictive accuracy.** The WEFS index meaningfully enhanced model performance across multiple specifications. This suggests that gender-focused metrics can and should be included in entrepreneurship research—not as supplements, but as central structural variables.

Together, these findings support a shift away from capital-centric models of entrepreneurship and toward multi-dimensional frameworks that incorporate demographic, institutional, and equity-centered lenses. As the next section will argue, such frameworks are not only more accurate—they are also more actionable for policymakers.

8. Policy Recommendations

Based on the findings of this study, several actionable strategies emerge for governments, development agencies, and ecosystem builders aiming to foster entrepreneurial ecosystems:

1. Invest in Gender-Inclusive Legal and Economic Reform

The Women’s Entrepreneurship Feasibility Score (WEFS) consistently emerged as a strong predictor of startup activity, especially within countries and over time. Policymakers should prioritize reforms that reduce gender-based legal restrictions, improve women’s labor force participation for a desired parity, and reduce informality in female-led businesses. Targeted reforms in property rights, contract enforcement, and anti-discrimination protections can improve real-world feasibility for women entrepreneurs, particularly within informal economies.

2. Build Institutional Capacity and Accountability

While governance quality was most predictive between countries, it also remains a foundational pillar for long-term ecosystem strength. Efforts to reduce corruption, improve bureaucratic efficiency, and enhance regulatory clarity are critical for streamlining new entrepreneurship. In countries scoring low on governance, institutional development—including digital government services and civil service reform—can lay the groundwork for sustainable entrepreneurship growth.

3. Expand Financial Access — Especially Digitally

Financial inclusion showed strong effects across pooled and panel models. Expanding access to financial institutions, particularly through digital platforms like mobile banking and e-wallets, can reduce entry barriers and support small-scale entrepreneurship. Initiatives should be tailored to rural areas, youth, and informal entrepreneurs, with a strong emphasis on gender-sensitive design. A successful case study of such reforms comes from Kenya, with the positive reception in the mobile-banking app “M-PESA”.

4. Deploy Region-Specific Interventions Based on Cluster Diagnostics

Clustering results revealed that entrepreneurial ecosystems vary more by structural composition than geography alone. Instead of one-size-fits-all programs, development agencies should tailor cluster-sensitive support strategies. For example - findings from Dynamic Clusters suggest:

- Cluster 1 countries (low-NBDR, low governance, low WEFS) may need foundational institutional reform.
- Cluster 4 countries (high WEFS, high governance) may benefit from innovation financing and scale-stage support.

Dynamic cluster trajectories also suggest where momentum can be reinforced or where stagnation needs to be disrupted - providing all the more reasoning for strengthening data collection and allowing temporal analysis in future work.

5. Promote Ecosystem-Level Thinking Over One-Off Reforms

Fragmented policies that target only capital access or only education are unlikely to succeed. Sustainable entrepreneurship requires interaction-based strategies that align legal reforms, financial infrastructure, education, and institutional capacity. Cross-sectoral approaches—such as public-private incubators, regional SME hubs, and multi-stakeholder coalitions—can help act on the relevant findings of this study.

9. Limitations

While this study offers valuable insights into the structural drivers of entrepreneurial activity, several methodological and data-related limitations warrant acknowledgment. These caveats are essential for contextualizing the findings and setting a foundation for future research - further elaborated upon in the following section.

1. Limited Temporal Depth

The dataset includes only four discrete time points (2011, 2014, 2017, 2021) for most countries, with many countries having only 3 observations. This restricts the use of longitudinal methods like Granger causality, ARIMA, or GARCH, and severely limits the within-country variability necessary to robustly estimate dynamic panel models.

This short time dimension also affects the interpretability of Fixed Effects models, which rely on within-unit variation. While FE results were reported, they must be interpreted with caution.

2. Infeasibility of Advanced Panel Methods

Attempts to implement Arellano-Bond (AB) and Correlated Random Effects (CRE) models were abandoned due to computational singularity and insufficient within-country variation. These methods, though well-suited to panel data in theory, require more time points and deeper variation than this dataset permits.

This inability to model lagged effects and unobserved heterogeneity more precisely limits causal interpretation and underscores the importance of cautious inference.

3. Dynamic Pooled OLS — Predictive, Not Diagnostic

While the Dynamic Pooled OLS model showed strong performance on R^2, it assumes uniform momentum across countries and does not control for country-specific shocks or feedback loops. It lacks panel-specific error terms, which are critical in modeling differentiated temporal trajectories.

Therefore, while useful for forecasting and confirming the presence of temporal inertia, its results should be treated as suggestive, not definitive.

4. Measurement and Construct Validity

Several key indicators, including:

- Governance (Control of Corruption),
- Financial Access
- WEFS

are based on expert assessments, composite indices, or survey-reported data. Meaning that they are prone to Subjectivity, Variation in definition across countries, and Hidden biases in data availability.

In the case of WEFS, while the index provides clear predictive value, it applies equal weight to components and does not capture within-country disparities or intersectional barriers.

5. Limited Generalizability Due to Data Availability

The final dataset includes 391 observations from 125 countries, but the sample is unbalanced. Countries with better statistical infrastructure—typically middle- or high-income—are overrepresented, while fragile, conflict-affected, or data-poor states are underrepresented.

As a result, policy recommendations may not fully apply to countries/environments omitted from the analysis.

6. Omitted Variable Bias and Endogeneity

Despite a multidimensional modeling approach, entrepreneurship ecosystems are complex. Several relevant factors could not be included, such as:

- Sectoral business registration trends
- Local political shocks
- Sub-national disparities

This opens the possibility of endogeneity in model estimates, where omitted variables influence both the independent and dependent variables.

7. Clustering Sensitivity to Preprocessing Choices

Both static (K-means) and dynamic (DTW-based) clustering depend on:

- Initial random seeds
- Scaling/normalization methods
- Outlier sensitivity

Different preprocessing pipelines may yield different cluster boundaries. While the present clusters are meaningful and interpretable, their boundaries should not be treated as fixed or universally replicable.

8. Absence of Causal Identification Strategy

This study is predictive, not causal. None of the models implement instrumental variables, regression discontinuity designs, or natural experiments. As a result, findings should not be interpreted as definitive proof of causality—only as evidence of consistent, statistically significant relationships.

Despite these limitations, this study contributes to a replicable (all data available publicly), data-driven framework for analyzing entrepreneurship formation across countries. It introduces a novel gender feasibility index, applies multiple validated econometric approaches, and leverages clustering to uncover deeper structural groupings. Future work can build on this foundation by expanding temporal coverage, incorporating subnational data, and applying causal inference techniques to further validate and refine these insights.

10. Conclusion & Future Work

This study set out to understand the structural drivers of entrepreneurial activity across countries using a uniquely constructed panel dataset with variables spanning the domains of governmental quality, macroeconomy, and demographic composition. Through the use of panel regressions, mixed-effects models, and clustering techniques, the findings highlight the critical roles of financial access, institutional quality, and gender-inclusive environments—captured via the Women's Entrepreneurship Feasibility Score (WEFS)—in shaping national patterns of new business formation.

At the empirical core, financial access and governance emerged as foundational between-country predictors, while WEFS demonstrated unique explanatory power within countries over time, especially in fixed-effects and hierarchical models. Clustering analyses revealed that countries behave more similarly when grouped by structural traits and growth trajectories than by income or geographic region alone. These insights affirm that fostering entrepreneurship requires not only capital but enabling systems of inclusion, trust, and infrastructure.

While dynamic and causal modeling approaches such as Arellano-Bond, CRE, and Granger causality were explored, they proved infeasible given the panel's short time dimension and limited within-country variation. Their exclusion reinforces the importance of acknowledging the constraints of global development data and using appropriate models that respect those constraints.

However, in order to strengthen and extend this work in the future, several paths are recommended:

1. Expand Temporal and Subnational Coverage

Longer time series would enable the use of true dynamic panel estimators (e.g., Arellano-Bond, GMM) and better capture structural breaks, lags, and nonlinear growth. Subnational data (e.g., state-level NBDR or regional WEFS) could help uncover intra-country variation and local differences.

2. Integrate Causal Identification Strategies

Future studies should implement instrumental variable techniques, regression discontinuity designs, or natural/quasi-natural experiments to establish causality and isolate exogenous variation in governance reforms, financial access rollouts, or gender equity programs.

3. Incorporate Informality and Sectoral Granularity

Formal entrepreneurship proxies often underrepresented real-world activity, especially in LMICs. Incorporating proxies for informal sector entrepreneurship (e.g., mobile money usage, street vendor density) or analyzing sector-specific trends (e.g., services vs. agriculture) would provide more nuanced insights.

4. Co-Design WEFS Extensions with Experts

While WEFS showed consistent statistical power, its formulation remains deductive. Future iterations should be co-designed with women entrepreneurs, policymakers, and legal scholars to better reflect lived experience and avoid overly top-down measurement strategies.

In sum, this study provides a wholly replicable, scalable, and policy-relevant framework for understanding the conditions under which entrepreneurship flourishes or falters across countries. It affirms that sustainable entrepreneurship is not driven solely by capital—but by the structures of access that allow people to act on their ideas. As the global economy navigates a wave of digital, environmental, and demographic change, entrepreneurship policy must move beyond financial levers to embrace systemic inclusion, adaptive institutions, and grounded realities.

11. Citation List

1. Acs, Z. J., Desai, S., & Hessels, J. (2008). Entrepreneurship, economic development and institutions. *Small Business Economics*, 31(3), 219–234. https://doi.org/10.1007/s11187-008-9135-9
2. Aidis, R., Estrin, S., & Mickiewicz, T. (2008). Institutions and entrepreneurship development in Russia: A comparative perspective. *Journal of Business Venturing*, 23(6), 656–672. https://doi.org/10.1016/j.jbusvent.2008.01.005
3. Amorós, J. E., & Bosma, N. (2014). Global Entrepreneurship Monitor: 2013 Global Report. *Global Entrepreneurship Research Association (GERA)*. https://www.gemconsortium.org/report/48580

4. Banerjee, A. V., & Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. PublicAffairs.
5. Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.
6. Beck, T., Demirgüç-Kunt, A., & Levine, R. (2007). Finance, inequality and the poor. *Journal of Economic Growth*, 12(1), 27–49. https://doi.org/10.1007/s10887-007-9010-6
7. Demirgüç-Kunt, A., Klapper, L., Singer, D., & Ansar, S. (2022). *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. The World Bank. https://documents1.worldbank.org/curated/en/099818107072234182/pdf/IDU06a834fe908933040670a6560f44e3f4d35b7.pdf
8. Elam, A. B., Brush, C. G., Greene, P. G., Baumer, B., Dean, M., Heavlow, R., & Greene, F. J. (2023, November 28). *Global Entrepreneurship Monitor research underscores resilience of women entrepreneurs and opportunities for policymakers*. Global Entrepreneurship Monitor. https://www.gemconsortium.org/news/global-entrepreneurship-monitor-research-underscores-resilience-of-women-entrepreneurs-and-opportunities-for-policymakers
9. Minniti, M., & Naudé, W. (2010). What do we know about the patterns and determinants of female entrepreneurship across countries? *The European Journal of Development Research*, 22(3), 277–293. https://doi.org/10.1057/ejdr.2010.17
10. Naudé, W. (2010). Entrepreneurship, developing countries, and development economics: New approaches and insights. *Small Business Economics*, 34(1), 1–12. https://doi.org/10.1007/s11187-009-9198-2
11. North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.
12. United Nations Statistics Division. (2023). *Standard Country or Area Codes for Statistical Use (M49)*. https://unstats.un.org/unsd/methodology/m49/

13. Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2008). Education and entrepreneurship selection and performance: A review of the empirical literature. *Journal of Economic Surveys*, 22(5), 795–841. https://doi.org/10.1111/j.1467-6419.2008.00550.x
14. World Bank. (2023). *World Development Indicators*. https://databank.worldbank.org/source/world-development-indicators
15. World Bank. (2022). *Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. https://globalfindex.worldbank.org/
16. World Bank. (2022). *Women, Business and the Law 2022*. https://wbl.worldbank.org/
17. World Bank. (2023). *Worldwide Governance Indicators*. https://info.worldbank.org/governance/wgi/
18. World Bank. (2023). *Entrepreneurship Database* [Data set]. The World Bank. https://www.worldbank.org/en/programs/entrepreneurship
19. World Bank. (2024). *Informal Economy Database*. https://www.worldbank.org/en/research/brief/informal-economy-database

12. Appendix

Figure 1: Correlation Matrix of Independent Variables

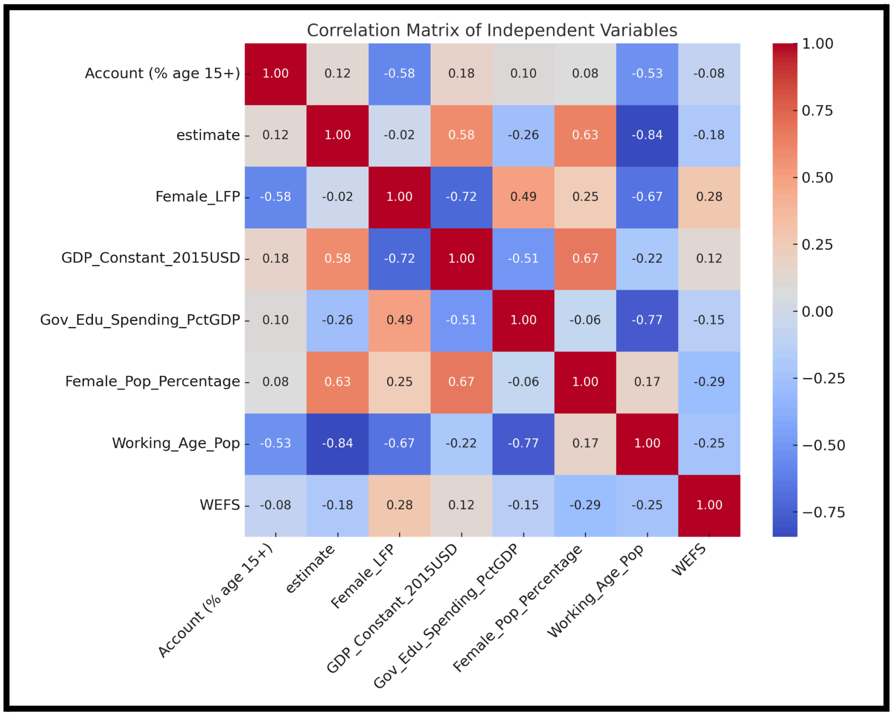

Figure 2: Log-Transformed Distribution of New Business Density Rate

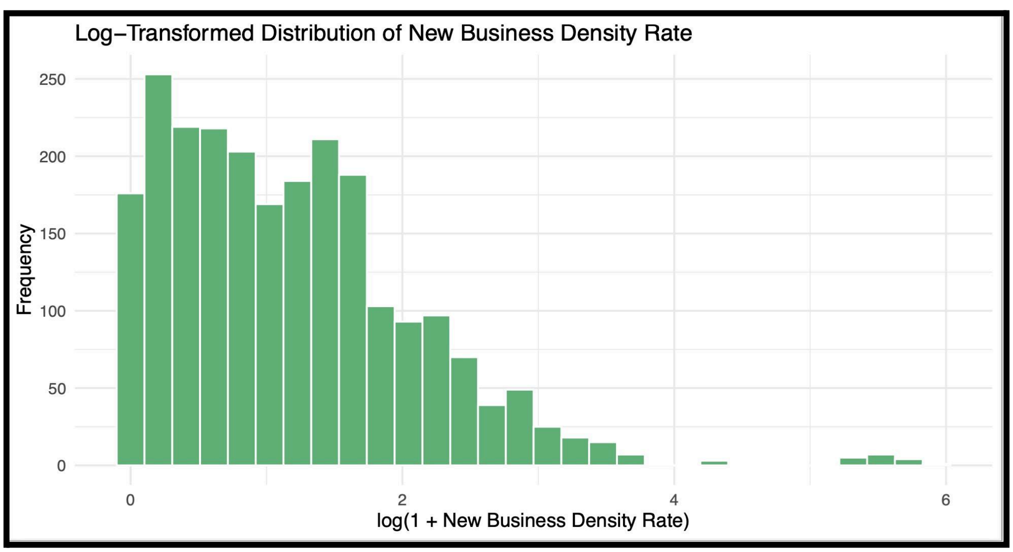

Figure 3: Average New Business Density Rate (2008-2022)

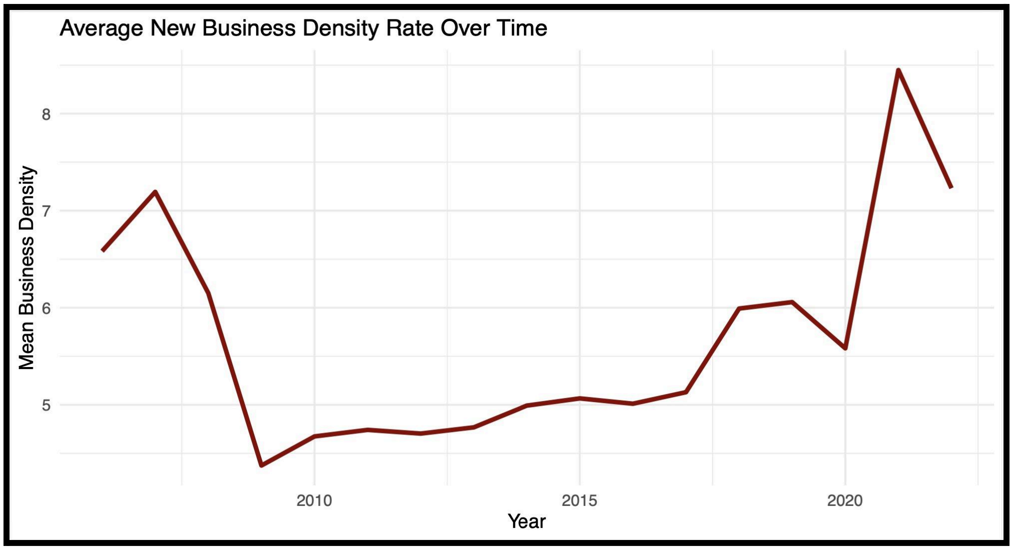

Figure 4: WEFS Formulation

$$\text{WEFS} = \frac{\text{Entrepreneurship Score} \times (1 - \text{LFP Gap Ratio})}{\text{Informal Economy Share (\%)}}$$

Table 1: Independent Variables and Mechanism by which they operate

Domain	**Variable Name**	**Description**
Finance	Account (% age 15+)	Share of adults with a financial account (Global Findex)
Governance	Estimate (WGI)	Control of corruption proxy (World Bank Governance Indicator)
Gender	WEFS	Women's Entrepreneurship Feasibility Score
Macroeconomic	GDP_Constant_2015USD	Constant-price GDP (control)

Demographic	Female_Pop_Percentage	% of population that is female – proxy for Gender Distribution
Demographic	Working_Age_Pop	% of population age 15–64 – proxy for Age Distribution
Education	Gov_Edu_Spending_PctGDP	Government expenditure on education as % of GDP

Table 2: Deterministic Models, Use Case, and Limitations

Model	**What It Does**	**Why It's Used**	**Limitations**
Multivariate OLS	Estimates relationships without accounting for panel data structure.	Interpretable, fast; good for initial diagnostics.	No panel structure; assumes independent observations.
Stepwise Regression (AIC)	Selects predictors by optimizing model fit using Akaike Information Criterion.	Reduces noise, avoids overfitting, keeps model parsimonious.	Can exclude theory-driven predictors; risk of overfitting to training data.
Fixed Effects (FE)	Controls for time-invariant traits by focusing on within-country variation.	Ideal when isolating the effect of policy or access changes within countries.	Drops time-invariant variables; requires within-unit variability.
Random Effects (RE)	Assumes country-specific effects are random and uncorrelated with predictors.	Useful for comparing across countries and estimating effects of slower-changing variables.	Biased if random effects are correlated with predictors.

Mixed Effects	Combines fixed and random effects in a hierarchical structure.	Captures both country-specific variation and general effects.	More complex interpretation; sensitive to specification.
Pooled OLS	Treats all observations as one large dataset, ignoring country or time structure.	Baseline model to estimate overall effects across countries and years.	Ignores unobserved heterogeneity and time/country effects.
Dynamic Pooled OLS	Adds lagged dependent variable to capture momentum, pooled across panel.	Models temporal persistence; **best-performing** model for forecasting and explanation.	Assumes uniform dynamics across countries; no country-specific error structure.

Table 3: Cluster Types, Use Case, and Limitations

Clustering Method	**What It Does**	**Why It's Used**	**Limitations**
K-Means Clustering	Groups countries based on similar values of selected indicators (e.g., WEFS, finance, governance).	Segments countries into **static** “entrepreneurship ecosystems” based on structural conditions.	Assumes spherical clusters; sensitive to outliers and variable scaling.
K-Medoid (with DTW)	Clusters countries based on **trajectory** of startup activity using Dynamic Time Warping (DTW).	Captures **temporal patterns** in NBDR; useful when comparing how countries evolve over time.	Requires ≥4 years of data; medoids are less interpretable than centroids.

Figure 5: IV Univariate

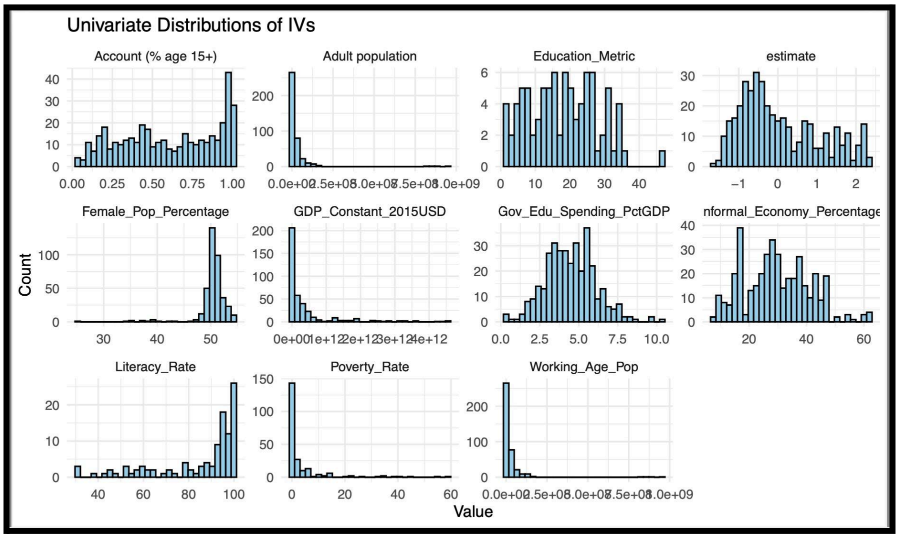

Figure 6: NBDR and IV Bivariate

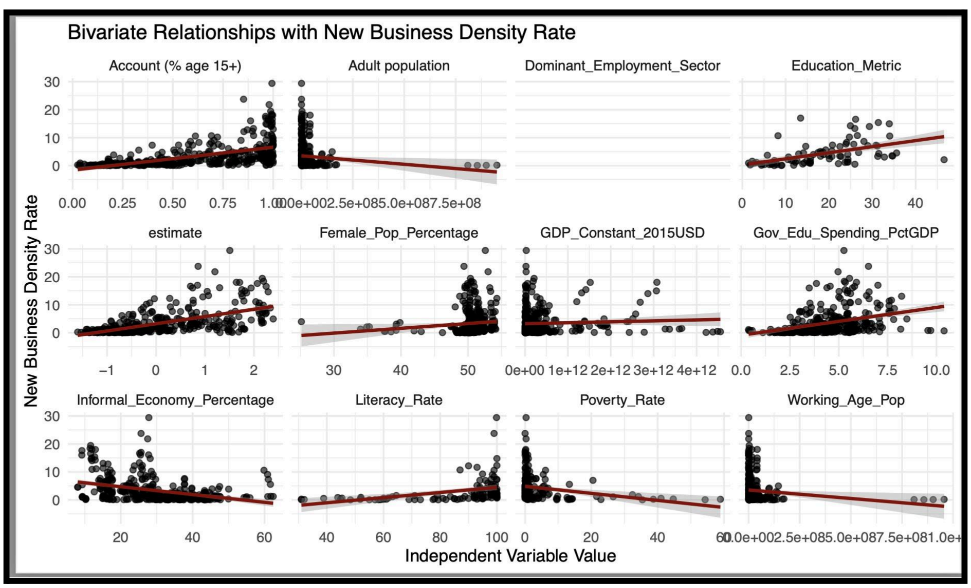

Table 4: OLS Single Univariate

	WEFS				WGI Estimate				GDP				GOV SPEND EDU			
Predictors	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	-0.06	0.39	-0.82 – 0.71	0.884	3.34	0.20	2.94 – 3.73	**<0.001**	3.38	0.28	2.83 – 3.93	**<0.001**	-0.96	0.70	-2.35 – 0.42	0.171
wefs	1.29	0.12	1.06 – 1.52	**<0.001**												
estimate					2.63	0.19	2.25 – 3.02	**<0.001**								
gdp constant 2015usd									0.00	0.00	-0.00 – 0.00	0.285				
gov edu spending pct gdp													1.00	0.15	0.71 – 1.29	**<0.001**
account age 15																
working age pop																
population growth																
Observations	340				340				340				340			
R^2 / R^2 adjusted	0.265 / 0.263				0.351 / 0.349				0.003 / 0.000				0.119 / 0.116			
AIC	1905.062				1862.837				2008.590				1966.785			

	FIN ACCOUNT AGE 15+				WORKING AGE POP				POP GROWTH			
Predictors	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	-1.74	0.47	-2.65 – -0.82	**<0.001**	3.69	0.26	3.18 – 4.20	**<0.001**	4.35	0.32	3.72 – 4.98	**<0.001**
wefs												
estimate												
gdp constant 2015usd												
gov edu spending pct gdp												
account age 15	8.52	0.68	7.19 – 9.86	**<0.001**								
working age pop					-0.00	0.00	-0.00 – -0.00	**0.017**				
population growth									-0.81	0.20	-1.20 – -0.42	**<0.001**
Observations	340				340				340			
R^2 / R^2 adjusted	0.318 / 0.316				0.017 / 0.014				0.047 / 0.044			
AIC	1879.794				2003.969				1993.317			

Table 5: OLS Multivariate

	Multivar OLS				Pooled OLS				Random Effects				Fixed Effects			
Predictors	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	-0.02	0.98	-1.95 – 1.91	0.987	-0.06	0.95	-1.92 – 1.81	0.951	-0.27	1.02	-2.28 – 1.73	0.788				
account age 15	4.57	1.11	2.39 – 6.74	**<0.001**	3.67	1.03	1.63 – 5.70	**<0.001**	2.82	0.86	1.13 – 4.50	**0.001**	1.92	1.01	-0.07 – 3.91	0.058
wefs	0.30	0.18	-0.04 – 0.65	0.088	0.22	0.18	-0.12 – 0.57	0.201	0.66	0.24	0.20 – 1.12	**0.005**	1.73	0.59	0.56 – 2.89	**0.004**
estimate	1.54	0.36	0.84 – 2.25	**<0.001**	1.40	0.37	0.68 – 2.12	**<0.001**	0.91	0.41	0.11 – 1.71	**0.026**	0.38	0.69	-0.99 – 1.75	0.583
gdp constant 2015usd	-0.00	0.00	-0.00 – -0.00	**<0.001**												
gov edu spending pct gdp	0.04	0.14	-0.24 – 0.32	0.792	0.13	0.15	-0.15 – 0.42	0.360	0.01	0.15	-0.29 – 0.31	0.940	-0.16	0.21	-0.57 – 0.25	0.454
working age pop	0.00	0.00	-0.00 – 0.00	0.900												
population growth	0.09	0.18	-0.26 – 0.45	0.604												
sub region [Australia and New Zealand]																
sub region [Central Asia]																
sub region [Eastern Asia]																
sub region [Eastern Europe]																
sub region [Latin America and the Caribbean]																
sub region [Northern Africa]																
sub region [Northern Europe]																
sub region [South-eastern Asia]																
sub region [Southern Asia]																
sub region [Southern Europe]																
sub region [Sub-Saharan Africa]																
sub region [Western Asia]																
sub region [Western Europe]																
lag density																
Random Effects																
σ^2																
τ_{00}																
ICC																
N																
Observations	340				340				340				340			
R^2 / R^2 adjusted	0.423 / 0.411				0.385 / 0.378				0.188 / 0.179				0.077 / -0.410			
AIC	1834.565				1850.334				1265.883				1135.952			

	Region Fixed Effects				Mixed Effects				Dynamic Pooled OLS				Stepwise AIC			
Predictors	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	-1.01	1.21	-3.40 – 1.38	0.404	-0.25	1.01	-2.23 – 1.73	0.801	-1.09	0.30	-1.69 – -0.49	**<0.001**	0.35	0.69	-1.00 – 1.70	0.610
account age 15	4.33	1.02	2.32 – 6.35	**<0.001**	2.87	0.86	1.18 – 4.56	**0.001**					4.38	1.00	2.41 – 6.36	**<0.001**
wefs	0.25	0.21	-0.16 – 0.66	0.230	0.62	0.23	0.17 – 1.07	**0.007**	0.00	0.06	-0.10 – 0.11	0.942	0.31	0.17	-0.03 – 0.64	0.070
estimate	0.38	0.37	-0.35 – 1.11	0.309	0.95	0.40	0.16 – 1.74	**0.018**	0.27	0.10	0.08 – 0.46	**0.005**	1.56	0.34	0.90 – 2.22	**<0.001**
gdp constant 2015usd									-0.00	0.00	-0.00 – 0.00	0.188	-0.00	0.00	-0.00 – -0.00	**<0.001**
gov edu spending pct gdp	0.08	0.14	-0.19 – 0.35	0.558	0.02	0.15	-0.28 – 0.32	0.883	0.10	0.04	0.03 – 0.17	**0.008**				
working age pop																
population growth																
sub region [Australia and New Zealand]	9.72	1.64	6.49 – 12.95	**<0.001**					-0.37	0.46	-1.26 – 0.53	0.419				
sub region [Central Asia]	0.14	1.30	-2.42 – 2.70	0.916					0.40	0.34	-0.27 – 1.08	0.243				
sub region [Eastern Asia]	-1.49	1.44	-4.31 – 1.34	0.302					0.37	0.39	-0.39 – 1.14	0.338				
sub region [Eastern Europe]	0.80	0.99	-1.15 – 2.74	0.421					0.90	0.26	0.39 – 1.42	**0.001**				
sub region [Latin America and the Caribbean]	0.28	0.89	-1.47 – 2.02	0.756					0.52	0.23	0.05 – 0.98	**0.029**				
sub region [Northern Africa]	0.10	1.55	-2.94 – 3.15	0.946					0.51	0.41	-0.30 – 1.31	0.215				
sub region [Northern Europe]	3.99	1.10	1.82 – 6.16	**<0.001**					0.35	0.30	-0.24 – 0.93	0.246				
sub region [South-eastern Asia]	-0.73	1.10	-2.89 – 1.43	0.506					0.42	0.29	-0.15 – 0.99	0.145				
sub region [Southern Asia]	-0.67	1.04	-2.71 – 1.37	0.519					-0.58	0.27	-1.12 – -0.04	**0.036**				
sub region [Southern Europe]	0.31	0.94	-1.54 – 2.17	0.741					0.87	0.24	0.39 – 1.35	**<0.001**				
sub region [Sub-Saharan Africa]	0.06	0.82	-1.55 – 1.66	0.945					-0.03	0.21	-0.46 – 0.39	0.877				
sub region [Western Asia]	3.20	1.02	1.18 – 5.21	**0.002**					0.81	0.28	0.27 – 1.35	**0.004**				
sub region [Western Europe]	-1.61	1.31	-4.17 – 0.96	0.219					0.47	0.35	-0.21 – 1.16	0.173				
lag density									0.20	0.01	0.17 – 0.23	**<0.001**				
Random Effects																
σ^2					2.48											
τ_{00}					9.64 $_{country}$											
ICC					0.80											
N					114 $_{country}$											
Observations	340				340				340				340			
R^2 / R^2 adjusted	0.547 / 0.523				0.385 / 0.874				0.718 / 0.702				0.423 / 0.416			
AIC	1772.795				1565.929				869.505				1828.964			

Figure 7: NBDR Outlier List

```
> print(influential_info)
               country year
20           Australia 2014
21           Australia 2017
22           Australia 2021
23             Austria 2011
24             Austria 2014
25             Austria 2017
78          Costa Rica 2014
94  Dominican Republic 2017
95    Egypt, Arab Rep. 2011
96    Egypt, Arab Rep. 2014
187             Kosovo 2011
188             Kosovo 2014
189             Kosovo 2017
202            Lesotho 2011
203            Lesotho 2017
222               Mali 2011
223               Mali 2014
229         Mauritania 2017
230             Mexico 2011
231             Mexico 2014
326       South Africa 2017
327       South Africa 2021
328              Spain 2011
```

Table 7: OLS Multivariate Table Outlier Comparison

	Full OLS Model				OLS w/o Influential Points			
Predictors	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	-0.02	0.98	-1.95 – 1.91	0.987	-0.65	0.67	-1.97 – 0.67	0.335
account age 15	4.57	1.11	2.39 – 6.74	**<0.001**	3.41	0.74	1.95 – 4.87	**<0.001**
wefs	0.30	0.18	-0.04 – 0.65	0.088	0.47	0.15	0.18 – 0.76	**0.001**
estimate	1.54	0.36	0.84 – 2.25	**<0.001**	0.72	0.24	0.25 – 1.20	**0.003**
gdp constant 2015usd	-0.00	0.00	-0.00 – -0.00	**<0.001**	-0.00	0.00	-0.00 – -0.00	**<0.001**
gov edu spending pct gdp	0.04	0.14	-0.24 – 0.32	0.792	0.22	0.10	0.02 – 0.41	**0.032**
working age pop	0.00	0.00	-0.00 – 0.00	0.900	0.00	0.00	-0.00 – 0.00	0.460
population growth	0.09	0.18	-0.26 – 0.45	0.604	-0.18	0.12	-0.42 – 0.06	0.139
Observations	340				317			
R^2 / R^2 adjusted	0.423 / 0.411				0.501 / 0.489			
AIC	1834.565				1438.809			

Figure 8: Elbow-Method to Derive Optimal K-Clusters (Static)

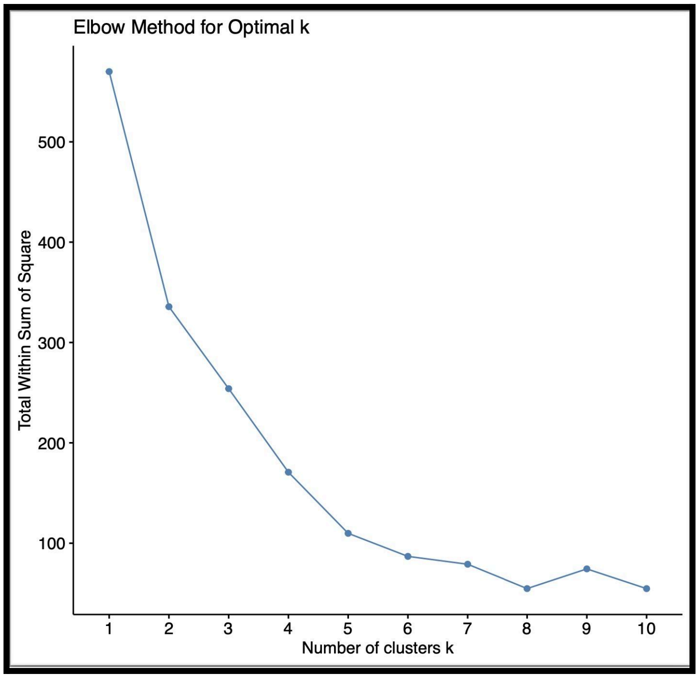

Figure 9: Gap Statistic to Derive Optimal K-Clusters (Static)

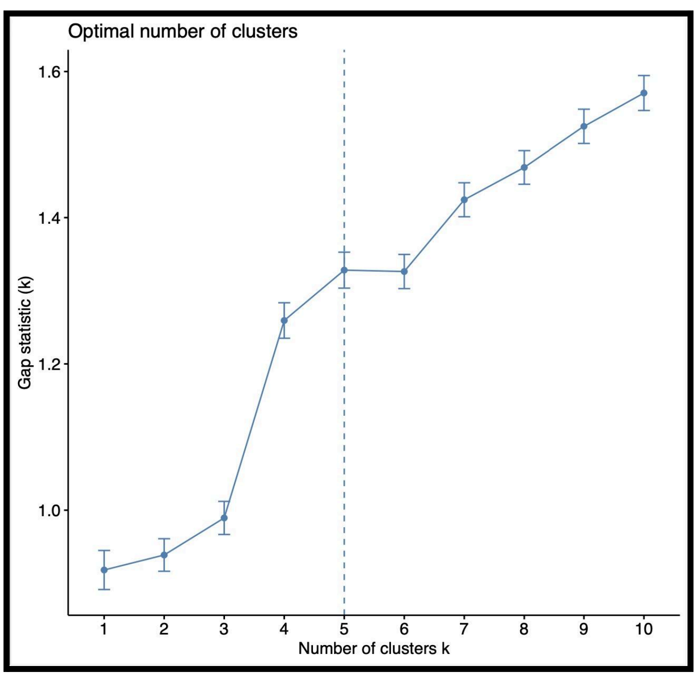

Figure 10: Cluster Geographic (Static)

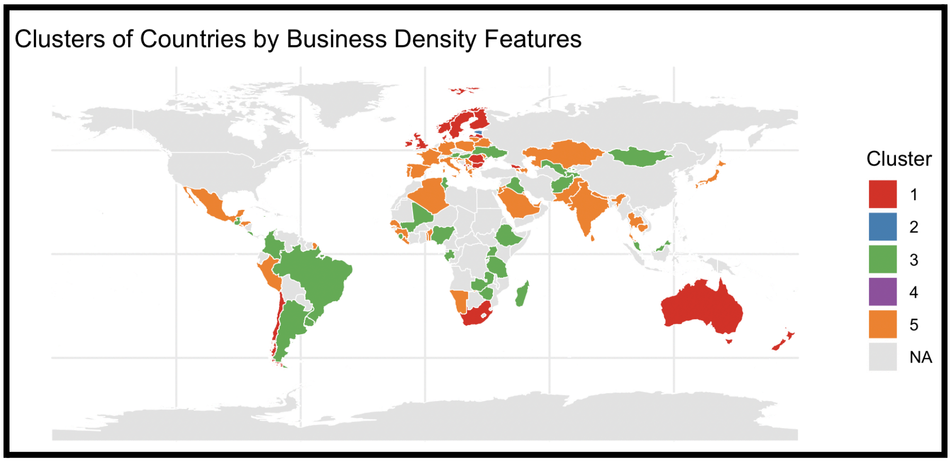

Figure 11: K-Means Cluster of Countries (Static)

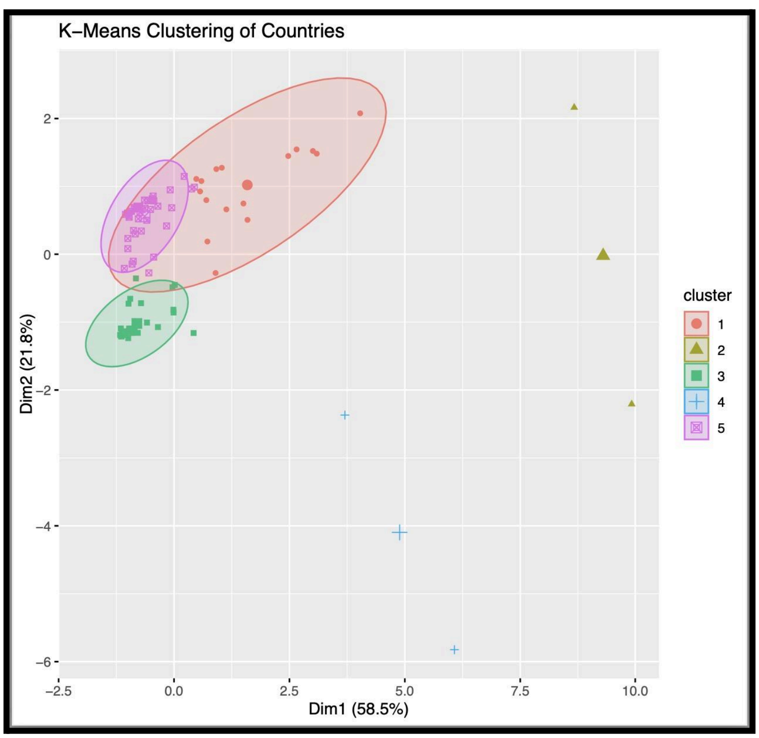

Figure 12: Average NBDR by Cluster (Static)

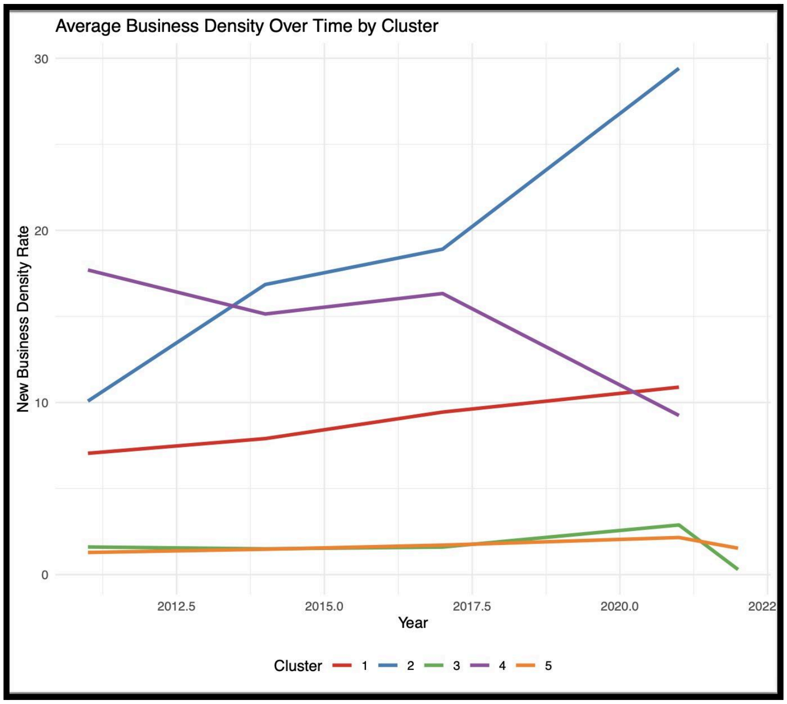

Figure 13: Cluster Profiles (Static NBDR Features)

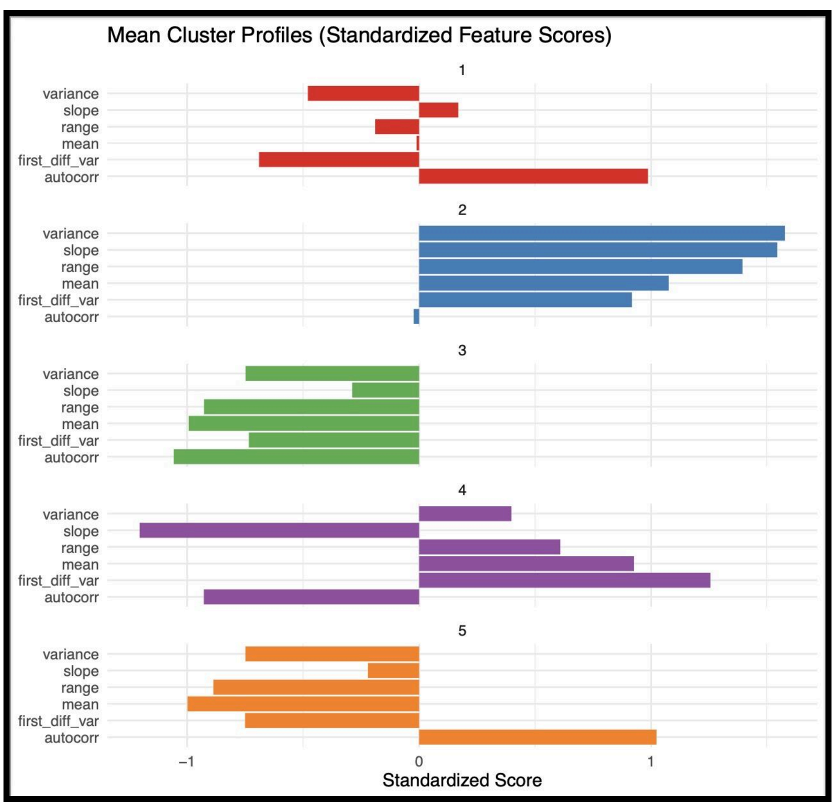

Figure 14: Cluster Geographic (DTW)

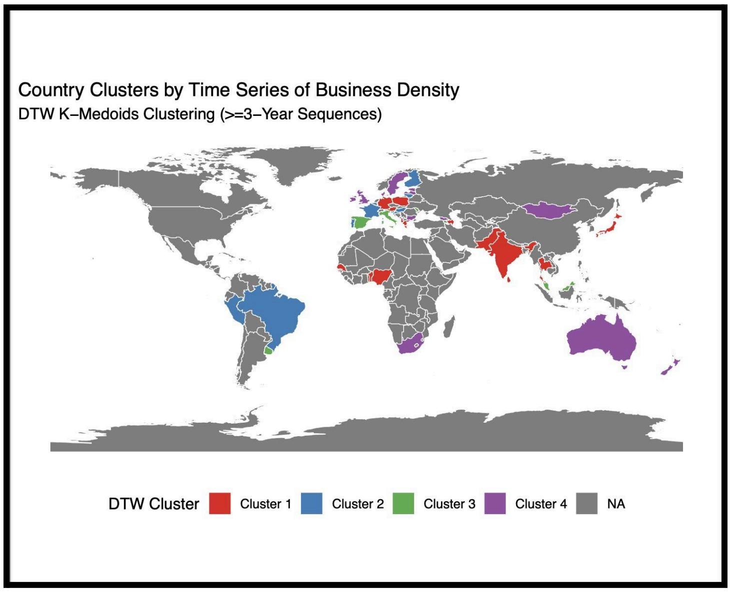

Figure 15: Cluster Profile (DTW)

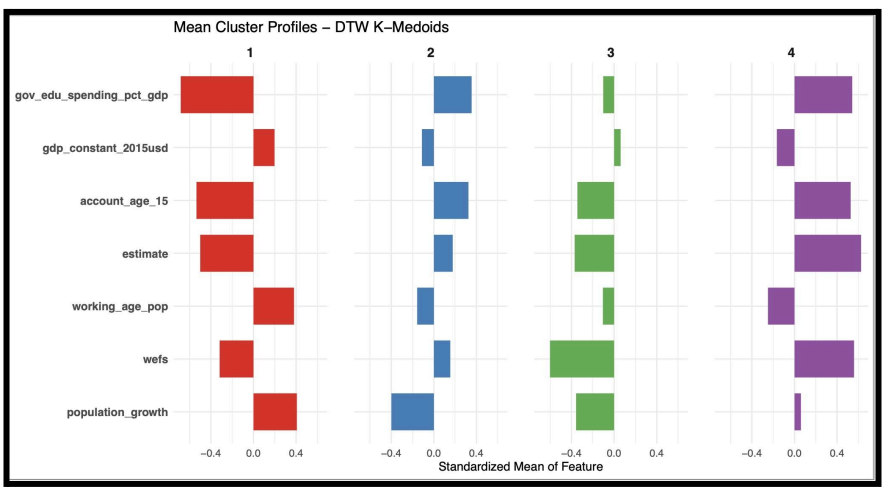

Figure 16: Average NBDR by Cluster (DTW)

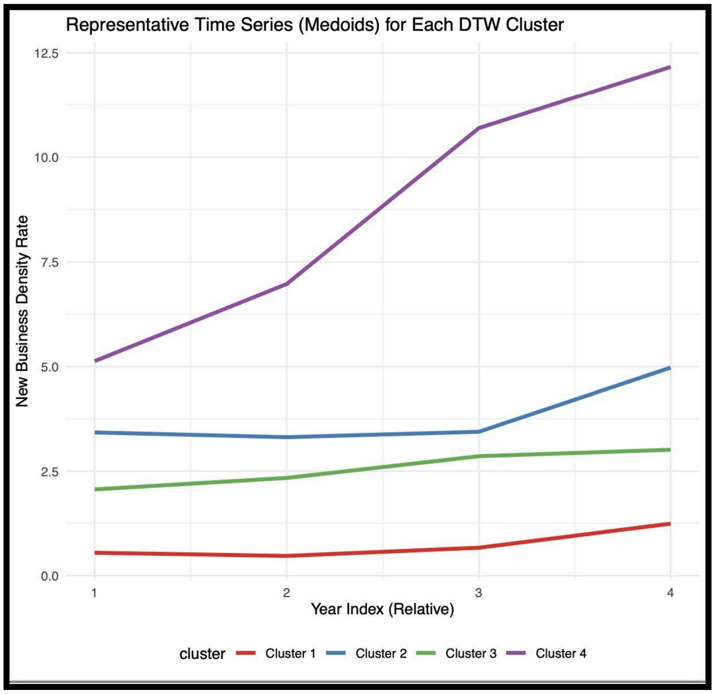

Figure 17: Sub-Regional Representation by Cluster (DTW):

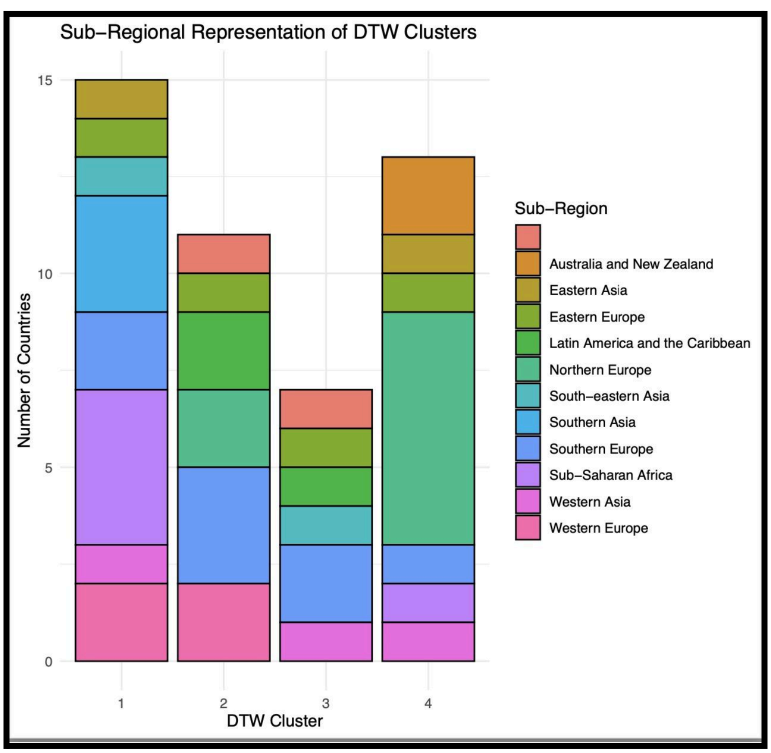

Figure 18: One-Year-Ahead Forecasts (2023) of NBDR (By Dynamic Pooled OLS):