Development and Exploration of Cultural Archetypes Using JMP

National behavioral patterns and trends are often explained in reference to poorly-defined cultural differences. These attributions lack broad contextual foundations and are thus inherently compromised. Yet culturally-based assessments, whether explicit or implicit, are often used in decision making at all levels, from individual actions to international relations. This paper outlines a straightforward semi-quantitative method developed in JMP for characterizing these important differences using freely available data. This method exploits the cross-platform flexibility of JMP, integrating the Cluster, Principal Components, Discriminant, PLS, and Screening platforms with the Contingency personality of the Fit Y by X platform. The method is generally applicable to circumstances where cultural issues may impact important decisions, but is demonstrated here in the realm of international economic development.

Introduction

Cultural archetypes are often utilized as convenient shorthand for alluding to common features among national or societal groupings without clear reference to a system or context beyond the group (*e.g.*, European). Cultural distinctions should be used with care, lest they be tainted by exaggeration borne of prejudice or ignorance. Yet as one international marketing expert aptly stated, although "teamwork training is big business in the United States, there is hardly a market for it in Japan" [1]. There is reasonable debate over how much to rely on such distinctions in decisional analysis. Resolution of that debate is far beyond the scope of effort entertained here as it hinges on deeply held philosophical beliefs. Instead, the following analysis is offered to illustrate a practical approach toward making contextually rich cultural comparisons where desired.

The topic of foreign direct investment (FDI) is explored here as a case study – specifically, the inward-flowing new "greenfield" investments (IGFDI) that tend to have the greatest impact on developing economies. First, a diverse set of potentially relevant socioeconomic data are used to aggregate societies into self-similar clusters. These clusters are considered to be cultural archetypes within the context of the data selected to bound the analysis. Next, a framework is outlined for determining an appropriate number of these cultural archetypes to be used in subsequent modeling. The diverse socioeconomic data are aggregated into a substantially smaller number of thematic variables and these candidate key factors are screened to develop a simple model that captures in a semi-quantitative fashion the relationships between the key thematic factors and trends in IGFDI. The case study concludes with a discussion of the differentiating themes that most clearly distinguish the identified cultural archetypes.

FDI is important to most economies as it constitutes a source of development funding for business enterprise. FDI categorizes a minimum 10% ownership stake in a firm, coupled with active management

responsibilities [2]. Such investments are illiquid and demand ongoing attention. IGFDI is especially important to developing economies because it signifies foreign investors' willingness to place a substantial long-term bet on the nation's economic soundness, in a manner that foreign portfolio investment does not. The study of IGFDI in relation to cultural archetypes indicates which behaviors, attitudes, traits and circumstances inspire investor confidence in the aggregate, and which do not.

The importance of FDI is underscored by the efforts of the United Nations Conference on Trade and Development over the past 30 years to measure it and to offer a forward-looking indicator of investment potential at the national level [3]. The UNCTAD econometric model for projecting national FDI potential uses 12 measurable input variables spanning economic output, energy use, exports, imports, telecommunications, and FDI stock – integrated FDI from prior years – as well as a more subjective country risk indicator. These inputs are relevant and unlikely to spark substantial controversy, but the model also demonstrates limited predictive capacity. The correlation coefficient between annual country potential score and any subsequent-year country performance score [4] ranged from -0.032 to 0.196 in the period from 2005 to 2010. Admittedly, predictions of this sort are notoriously difficult, as local conditions vary due to unforeseen events, and investor behavior is not always purely rational. But these results do lead to the question of whether a more predictive model might be achieved by incorporating a broader range of socioeconomic data, or aggregating societies by cultural similarity, or both.

A functional system of archetypes for addressing far-reaching international challenges must be drawn from factors that are truly world-spanning and broadly measurable, enabling meaningful cross-cultural comparisons. The goal is to arrive at a sufficiently nuanced view of cultural differentiators to prime human analysts with a balanced and coherent comprehension (*i.e.*, *gestalt*) that provides heuristic benefit for practical decision-making – for example, deciding what kind of information should be considered in developing a quantitative model for predicting FDI flows.

Several organizations provide open-source data and assessments on many nations, addressing topics of socioeconomic import such as gross national product, population density, business freedoms, etc. Various efforts have also been made to characterize national cultures on the basis of fundamental behaviors and attitudes, such as the Global Leadership and Organizational Behavior Effectiveness (GLOBE) study of 62 societies [5]. The GLOBE indices and a wide variety of socio-economic data are used here in concert to characterize 57 of these societies as self-similar clusters, forming the basis of clearly differentiated and globally relevant archetypes.

Cultural Archetype Development

Categorization should begin by mutual consideration of many potentially relevant variables – a wide-funnel approach acknowledging the very limited initial understanding that characterizes exploratory analysis. Lacking *a priori* rationale for emphasizing particular information, all available data were given equal weight in determining the clusters of societies that constitute candidate cultural archetypes. A total of 90 columns of broad-based socioeconomic data were collected, spanning business infrastructure and governmental policies, energy generation and consumption, economic and political freedoms, crime and punishment, corruption, health, education, demographic trends, and societal leadership preferences and characteristics (GLOBE). These data were applied to 57 societies representing all major populated regions of the world.

On a very basic level, the task of this analysis is to relate a set of societal groupings to IGFDI performance among each group. Cumulative per capita IGFDI [6] into the aforementioned 57 societies for the eight years from 2003 to 2010 is shown in Figure 1. The distribution of IGFDI among societies spans a broad range and is clearly uneven. The per capita IGFDI of Qatar, for example, exceeds $110,000 and is more than five times the next ranking society of Singapore, yet more than half of these societies have received per capita investment of less than $2,000 over the same time span.

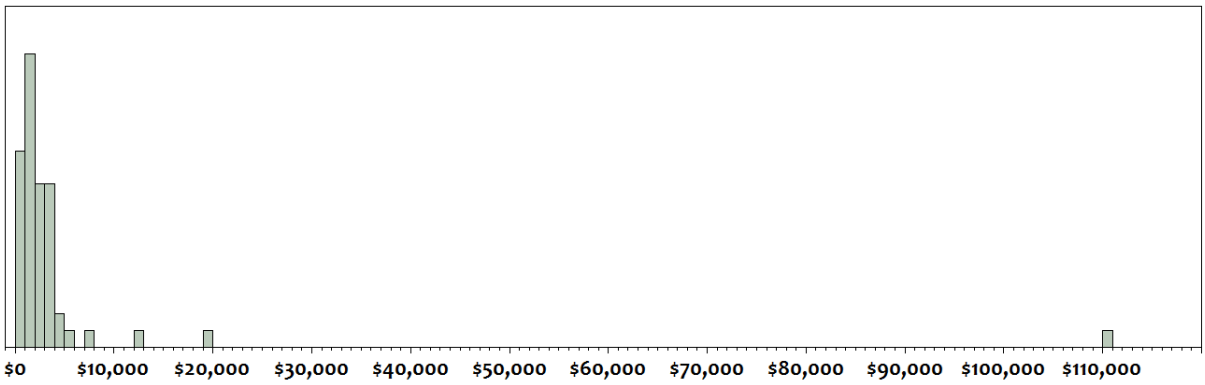

Figure 1: Distribution of cumulative IGFDI by nation, 2003 – 2010.

Hierarchical clustering (Ward's method) was used to characterize self-similar groupings of societies. An example showing six clusters is provided in Figure 2. The first principal component was used as the ordering column because the results of this method depend somewhat upon the row order of the data. The number of clusters was varied from two to nine, and the cluster designations were saved in separate columns in the data table for later use.

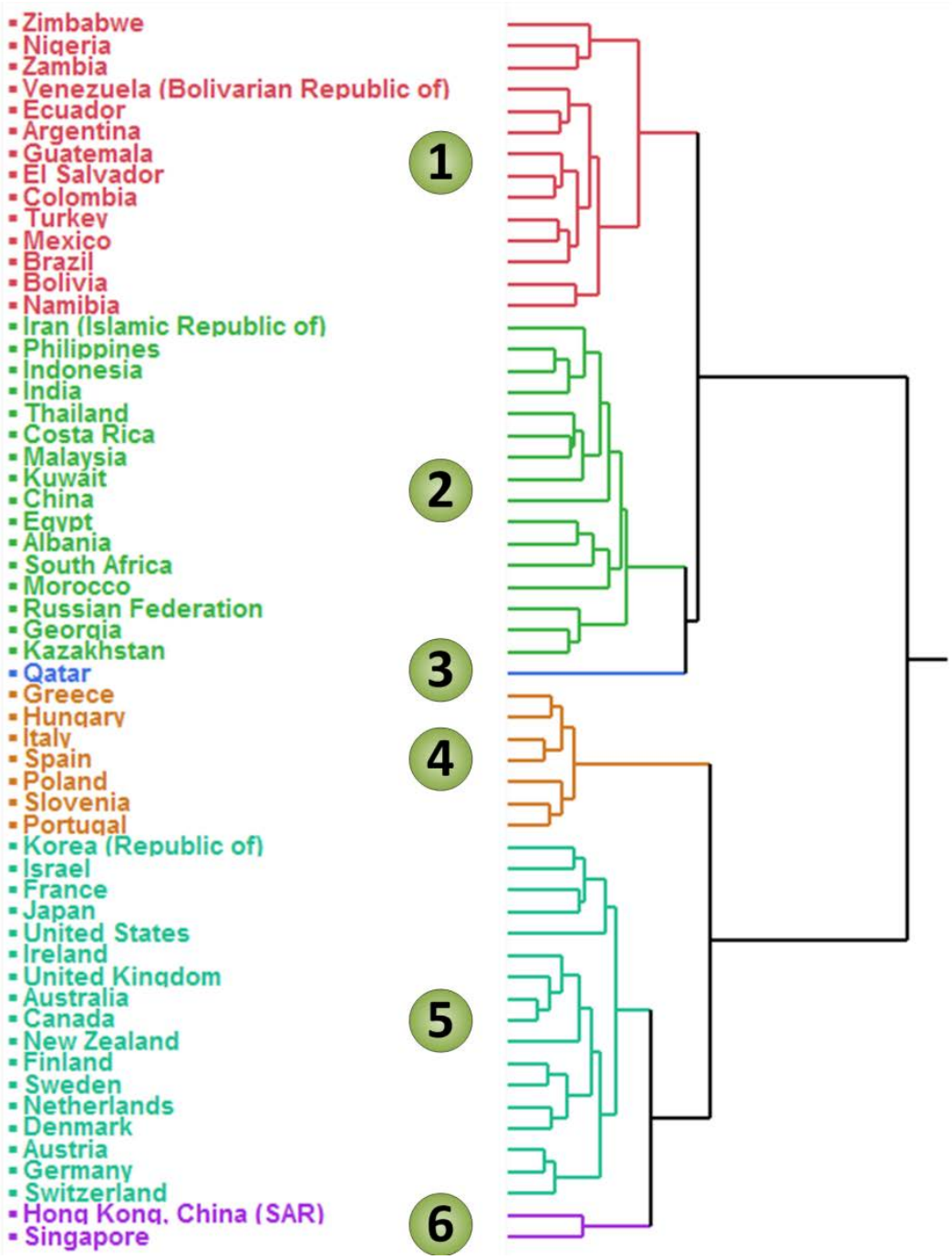

Figure 2: Societal clusters based on diverse socioeconomic data. Dendrogram spacing is by Euclidean distance.

Before moving on to archetype development, there is value in considering the alternative of using partial least squares (PLS) to form a model linking the input data to the output – in this instance, integrated per capita IGFDI from 2003 to 2010, inclusive. PLS can be an excellent modeling option when the number

of input variables exceeds the number of output data rows, as it does in this instance (90 inputs *vs.* 57 societies). Applied to this data set, PLS (JMP 10, NIPALS method) explains 99.3% of the variation in per capita IGFDI with 12 latent (composite) variables. But 46 of the 90 input variables are deemed important by Wold's criterion, and each is a composite of the original 90 inputs. The strength of PLS is in developing a reasonable model when the data volume is substantial, and intuitive comprehension of the linkages between inputs and outputs is relatively unimportant. The approach shown here is substantially less capable in regard to the statistical explanation of observed variation, but the results are more intuitive for explanation to human decision makers.

The highly uneven distribution of IGFDI in Figure 1 indicates that a scale-compression transformation might be appropriate. The penchant of human decision makers for relatively simple rankings spanning a modest number of levels, however, favors an ordinal binning scheme. Ordinal ranking also enables the use of the probability estimate from Pearson's χ^2 test statistic as a pattern-matching indicator for selecting an appropriate number of clusters and levels. When applied in hypothesis testing, this test indicates the probability that the differences within a given combination of clusters and levels could have occurred by chance. That meaning is lost in applying the probability across many combinations, but it provides a qualitative indication of relative suitability, with smaller probabilities implying a superior match.

The Contingency personality of the Fit Y by X platform was applied to a range of cluster counts and numbers of IGFDI ordinal bins. Pearson's χ^2 probability was extracted from the results report for each combination of clusters and bins. A matrix of these probabilities was generated, and then depicted using the Bubble Plot platform as shown in Figure 3. Choosing the number of clusters and ordinal bins (divisions) from among the available (blue) options is a matter of judgment. The remainder of the analysis shown here uses six clusters and eight ordinal bins.

Once the number of societal clusters to use for the remainder of the analysis has been chosen, there is value in providing a label for each cluster that captures a common theme. The labels that were selected for the six-cluster framework depicted in Figure 2 are listed in Table I.

Consolidation of Cultural Factors

The above clustering analysis of 57 societies considers a total of 90 socioeconomic and behavioral factors drawn from various sources. This diversity of source information was included in order to improve the opportunity of finding essential factors that may affect societal or broad cultural attractiveness for IGFDI. But much of the data are highly correlated, calling into question how many truly distinguishing cultural factors are actually available.

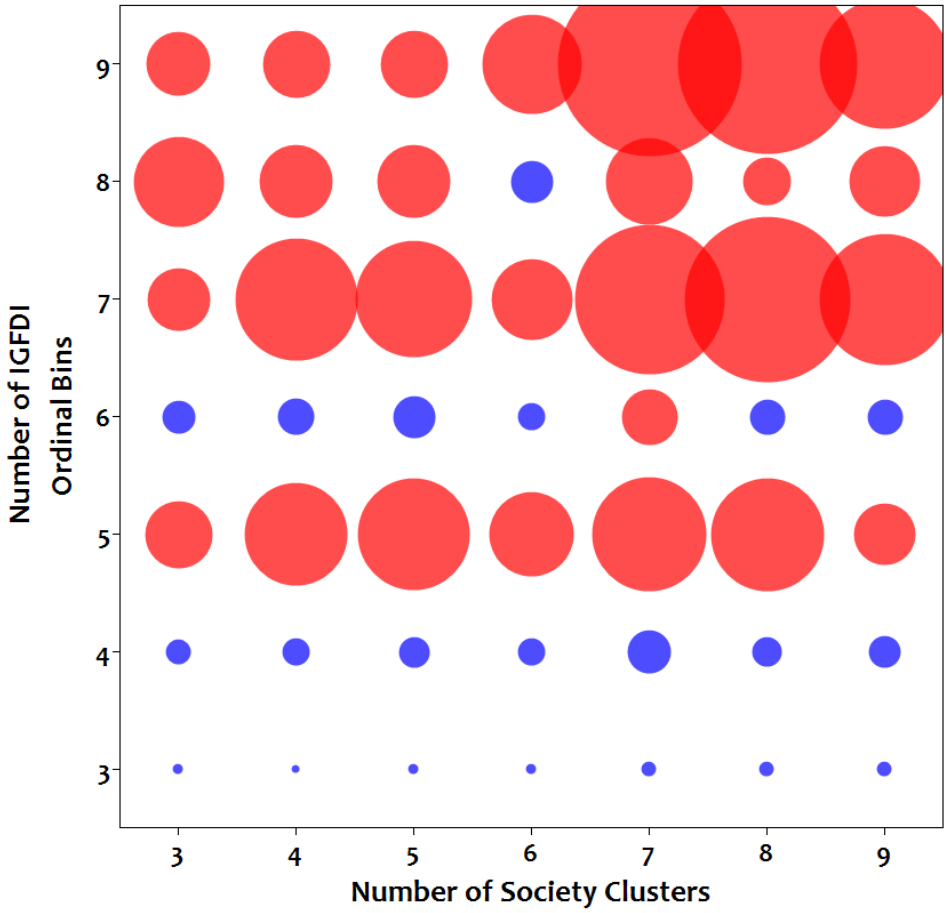

Figure 3: Pearson χ^2 probability array for selecting number of clusters and ordinal divisions. Bubbles are sized by probability estimate. The bubbles are colored blue for Pearson χ^2 probabilities of 1% or less, and are colored red if over 1% or the test result was flagged as suspect. Small bubbles (low probabilities) indicate better registration between the cluster and binning structures.

Table I. Descriptive Labels for the Six-Cluster Analytical Framework.

Cluster	Descriptive Label	Membership
1	Developing Economies	14
2	Transitional Economies	16
3	Qatar	1
4	Mediterranean & Eastern Europe	7
5	Developed Economies	17
6	Strategic Port Cities	2

Methods such as principal components analysis or partial least squares can be used to produce composite input variables emphasizing the most relevant data, but the complexity of the results can obfuscate subjective meaning. Moreover, where factors are closely correlated mathematically, yet have no rational basis for similarity, the essential themes identified through factor analysis may be unstable and heuristically counterproductive. Consequently, an effort was undertaken to more simplistically and overtly consolidate the 90 input factors into a substantially smaller set of meta-level groupings prior to developing a key factors model.

A correlation matrix was generated for the 90 input factors. The matrix was clustered as a square color-map showing clear demarcations between groups of highly similar data columns (Figure 4).

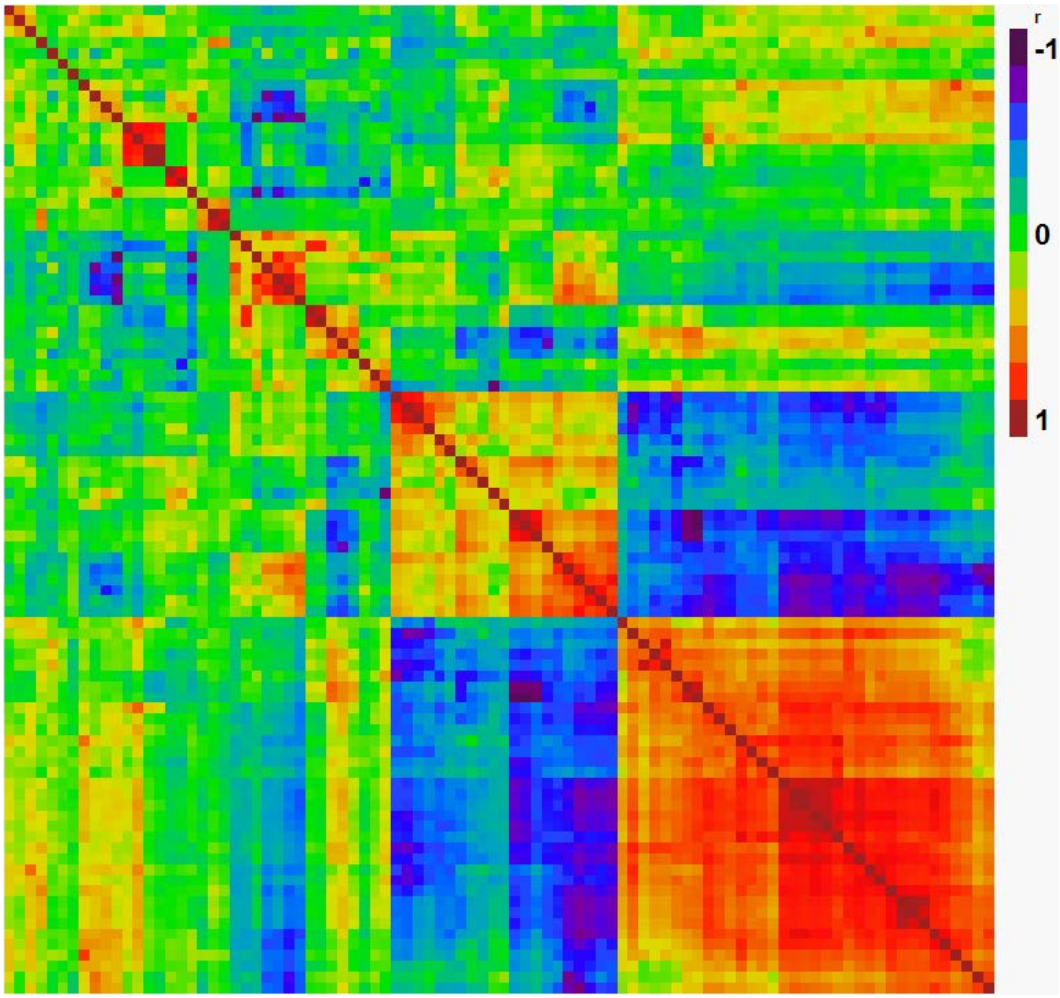

Figure 4: Correlation coefficient cluster map as a color-coded square matrix.

While the color-map confirms structure – indicative of the opportunity to employ a reduced-dimensionality data set – sorting the data columns into statistically balanced aggregates is a daunting

visual exercise. Consequently, the matrix was clustered using Ward's hierarchical method, with initial ordering on the first principal component of the data set (*i.e.*, the correlation matrix). As a compromise between manageability and granularity, 18 meta-groups were chosen (see Figure 5). Based on experience, a 5:1 consolidation has proven useful in the past, but this is another judgment decision.

Lateral motion along the dendrogram within Figure 5 represents Euclidean distance, showing that data groups (DG) 15 through 18 constitute a highly related super-group that differs dramatically from the remainder of the data set. Interestingly, many of the indicators produced by leading international organizations (*e.g.*, World Bank Global Indicators, World Economic Forum) are heavily correlated, such that the dimensionality of their analyses is far lower than might be implied from the number of different scores provided in their reports.

The 18 data groups identified in Figure 5 were then consolidated into representative meta-level data columns in order to serve as input variables for subsequent model development. Given the high level of correlation within each group, the average of the column-standardized data was used as the group surrogate.

Modeling the Consolidated Cultural Data Set

In order to discern which data groups might be most informative for IGFDI model development, the 18 data groups were pre-screened using the Discriminant platform. The Discriminant platform was used with stepwise variable selection to identify the data groups that enable differentiation of the six societal clusters established above. This was done to eliminate data groups that might correlate with FDI, but would not help to characterize the cultural differences pertaining to FDI. The fitting procedure that was followed was to first add all variables, and then remove variables step-wise until the next deletion had 0.1 or lesser p-value. This procedure is conservative from the standpoint of removing only those data groups that are clearly non-differentiating with respect to societal archetype membership.

The resulting discriminant model included 11 of the 18 covariates – groups 1, 2, 3, 5, 9, 10, 12, 13, 15, 17 and 18 in Figure 5 – and correctly categorized cluster membership for all but two societies (The Philippines and Turkey). The biplot depiction of that analysis is shown in Figure 6. Note that each societal cluster, or archetype, has now been given a descriptive label. Society color-coding is the same as for the dendrogram in Figure 2. The chosen data groups, or themes, readily differentiate the societal archetypes, but the biplot also underscores the unique nature of Cluster 3, with the nation of Qatar as its only member.

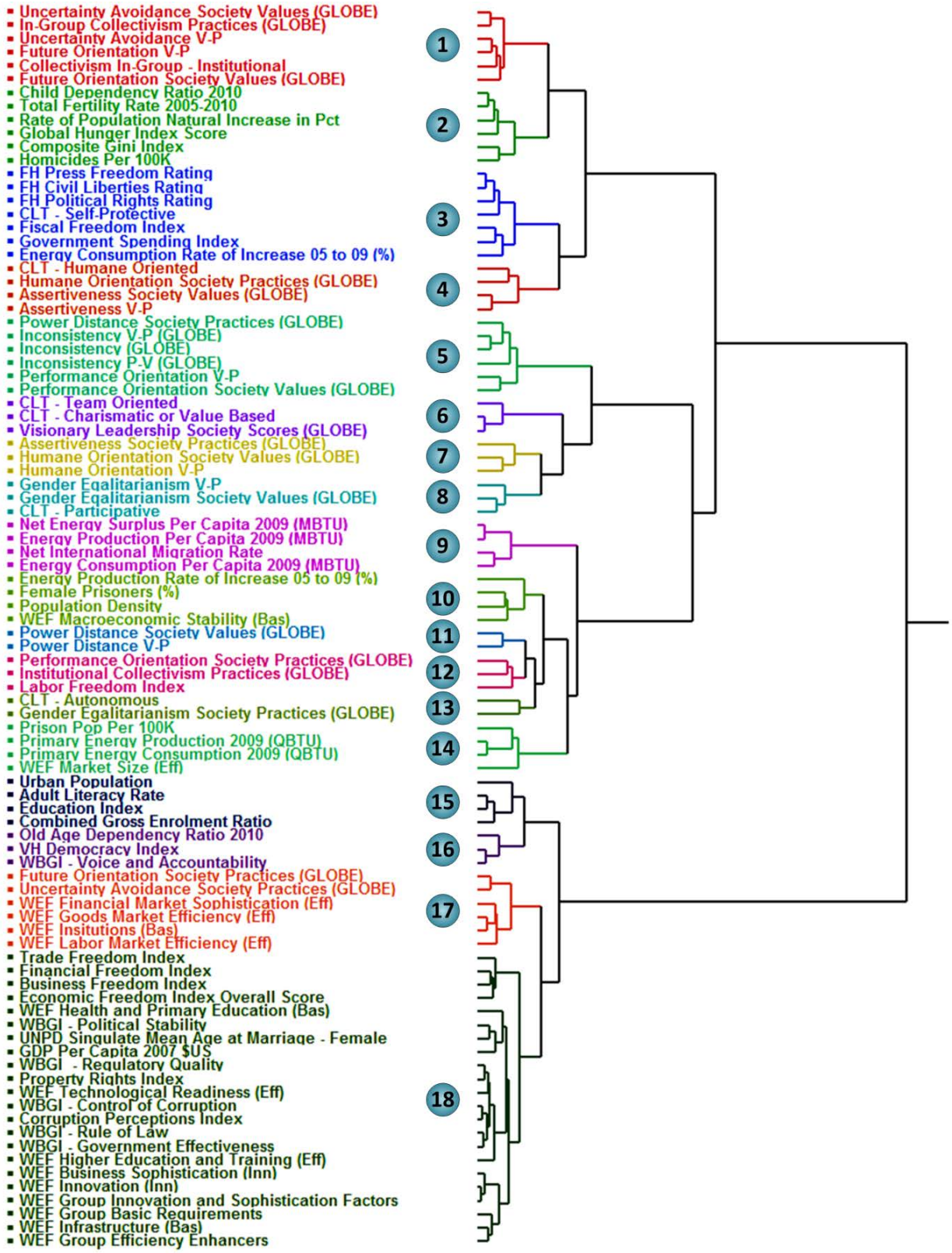

Figure 5: Correlation coefficient cluster map of 90 cultural variables.

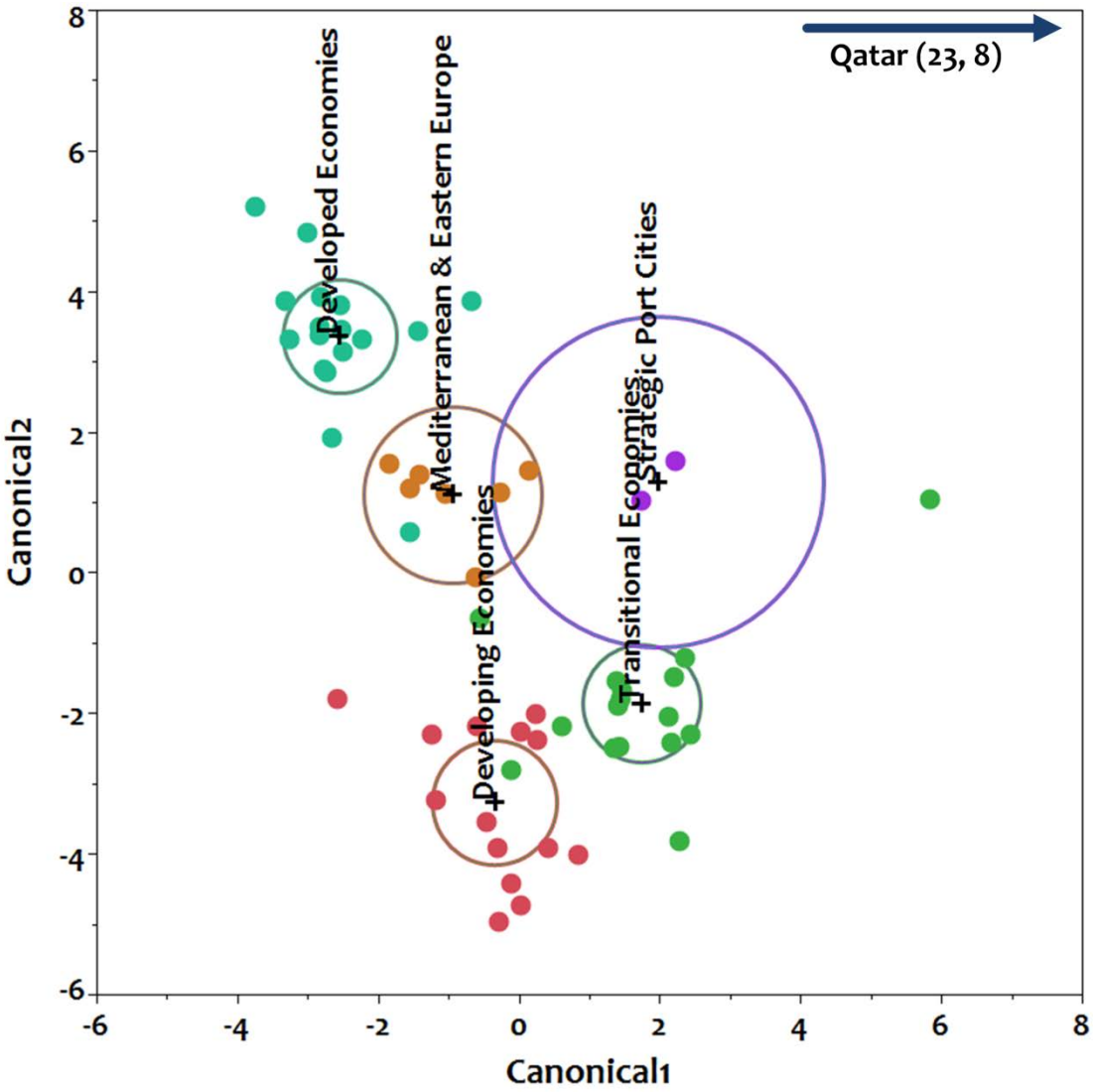

Figure 6: Separation of societal clusters by discriminant model.

After pre-screening, the Screening platform was applied to the selected data groups using the mean octile of per capita IGFDI as the output variable. Model refinement was carried out by removing terms and interactions with Student's t-test probabilities exceeding 0.1. The final model retains data groups DG01, DG02, DG03, DG09, and DG18, as well as the DG02*DG03, DG18*DG01 and DG18*DG18 interaction terms. The key model parameters are summarized in Table II. The model fit for archetype (cluster) mean octile IGFDI in terms of those key parameters is presented in Figure 7.

Table II. Key model parameters for IGFDI Mean Octile Fit.

Term	Thematic Label	Est.	Std. Err.	t Ratio	Prob>|t|
DG01	Uncertainty Avoidance Values	0.4881	0.2274	2.15	0.0368*
DG02	Ease of Child-Rearing	0.7933	0.1580	5.02	<.0001*
DG09	Energy Intensity	0.324	0.0952	3.41	0.0013*
DG18	Business Preparedness	1.397	0.2383	5.86	<.0001*
DG02*DG03	Ease of Child-Rearing * Civil Liberties	0.694	0.2411	2.88	0.0059*
DG18*DG01	Business Preparedness * Uncertainty Avoidance Values	1.115	0.2132	5.23	<.0001*
DG18*DG18	Business Preparedness * Business Preparedness	0.9528	0.1793	5.31	<.0001*
Intercept	Intercept	4.548	0.415	10.96	<.0001*

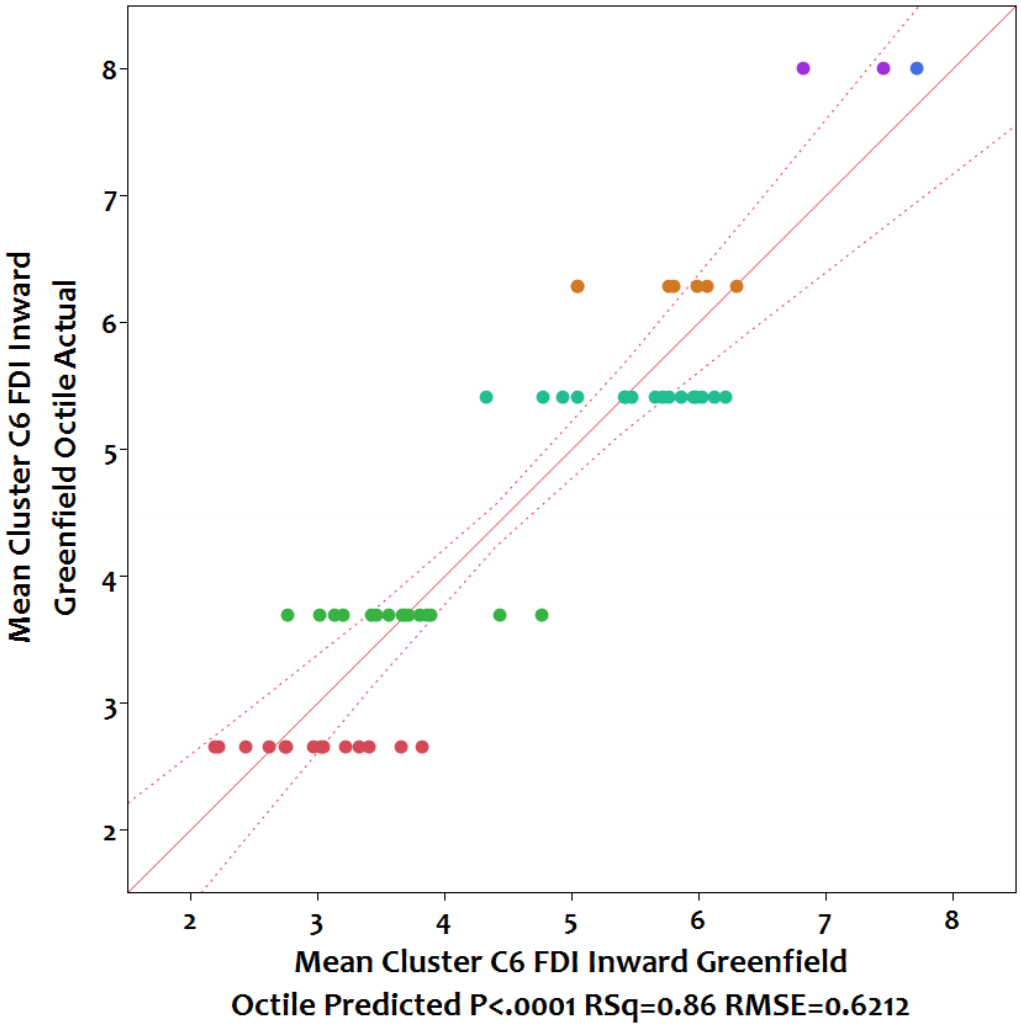

Figure 7: IGFDI model fit to five key culturally differentiating factors. Match between predicted and actual mean cluster IGFDI octiles.

Clearly, with R^2 of 0.86 (adjusted R^2 = 0.84) there is a reasonably good fit between the predicted and realized cluster mean octile of IGFDI. These model terms are summarized with subjective themes in Table III below, each with a representative input variable chosen for maximum correlation with the data group surrogate. The thematic descriptions are simplistic but the focus of this analysis is on making the results as easy as possible for decision makers to comprehend.

Table III. Meta-Group Themes and Representative Variables.

Group	Group Theme(s)	Representative Variable	Correlation
1	**Uncertainty Avoidance Values**. Concern for the future and avoidance of risks. Pride and reliance in local *vs.* more centralized collective action and distribution of resources.	Uncertainty Avoidance V-P (Values – Practices)	0.9376
2	**Ease of Child-Rearing**. Relative ability of general populace to care for basic needs of children. Characterized by low fertility rate, prevalence of hunger, violent crime, and concentration of wealth.	Child Dependency Ratio	0.9691
3	**Civil Liberties**. Political and press freedoms. Modest government spending. Reduced need for social posturing in leadership.	FH Civil Liberties Rating	0.8850
9	**Energy Intensity**. Intensity of development and use of energy resources. Inward migration of people from other nations.	Energy Production Per Capita	0.9818
18	**Business Preparedness**. Business freedoms, property rights, rule of law, political stability regulatory quality, infrastructure, and educated populace with specialization.	Government Effectiveness	0.9836

Note: The original values for data groups DG02 and DG03 have been inverted in order to aid comprehension, by presenting them in positive terms and with positive model parameters. In their original form, high values in these data groups correlated with difficulties in child-rearing and denial of civil liberties, respectively. All model terms have been corrected for consistency throughout.

Discussion of Key Cultural Factors Influencing IGFDI Flow

Clearly, each cultural archetype (society cluster) is linked with a different set of typical circumstances that yield differing IGFDI outcomes. But each archetype is now framed against the others by its relative standing in terms of the key thematic model variables, and there is a framework for assessing the impact of each factor in turn.

Uncertainty Avoidance Values

The GLOBE study captured the important difference between cultural values and practices. Uncertainty Avoidance Values and Uncertainty Avoidance V-P – meaning Values less Practices scores – are highly correlated. In other words, Values in this context point to a level of cultural *concern* regarding Uncertainty Avoidance, but those *Values* are actually negatively correlated with cultural *Practices* (correlation coefficient -0.6089). The GLOBE Future Orientation Values and Future Orientation V-P are also embodied within this thematic factor in the model shown above, and there is a similarly constructed difference between Values and Practices in Future Orientation as well (correlation coefficient -0.4007).

Consequently, and perhaps counter-intuitively, high scores in the Uncertainty Avoidance Values theme do not imply strong uncertainty avoidance in practice, or relatively strong planning for the future. Uncertainty Avoidance Values is a relatively weak driver of IGFDI among the identified key factors. In concert with Business Preparedness, however, its leverage is more than doubled. Notionally, Business Preparedness constitutes evidence of Uncertainty Avoidance and Future Orientation practices in action. Mathematically, Business Preparedness and DG17, which captures those practices, are highly correlated (corr. coeff. 0.823).

Ease of Child-Rearing

Average family size, the availability of food, and the concentration of income as described by the Gini index all affect the absolute or relative child-rearing challenges faced by typical parents within a society. The incidence of violent crime as indicated by the homicide rate is also somewhat correlated (0.6064) with the Gini index. The rate of population growth is intimately related to the fertility, family size, and availability of food as indicated by the Global Hunger Index. The most representative single variable encompassed by this theme is the Child Dependency Ratio. This ratio is intended to track the economically unproductive portion of the population ages zero to 14 relative to the assumed productive portion ages 15 to 60. The developing economies tend to be most impacted by this theme, which is strongly and negatively linked to IGFDI.

Energy Intensity

This theme refers to the per capita rates of energy production and consumption within societies. Historically, economic progress has been achieved in tandem with access to sources of power beyond direct human labor. Exogenous energy amplifies the impact of human capital. It also attracts human migration, presumably to partake in the economic benefits that accrue.

The model developed here is correlative, and in the absence of natural experiments to corroborate cause-and-effect relationships, potential implications of causation must be considered reservedly. For this

theme, the direction of causality in general is unclear. Certainly, IGFDI will tend to cause increased Energy Intensity as investments in plant and equipment infrastructure tend to directly increase energy usage, magnified by economic ripple effects. On the other hand, societies with increasing per capita energy use may also be expected to attract IGFDI from investors seeking to take advantage of the resulting increased human productivity.

Qatar is a special case cluster with a membership of one. It is an oil-rich state with a relatively modest population of less than one million people, and high Energy Intensity. A high level of IGFDI is involved in exploiting Qatar's oil and natural gas resources. In that instance the direction of causality is clear.

Business Preparedness

Nearly one-fourth of the data used in this analysis fall under the highly correlated Business Preparedness umbrella. Various indices characterizing business freedoms, property rights, rule of law, political stability, regulatory quality, and level of infrastructure all point to a supportive environment for voluntary economic activity. It seems quite reasonable for investors to consider the level of Business Preparedness within a society before making large, long-term investments. Business opportunities must be weighed against the freedom to act upon them.

Differentiation of Archetypes

A fair side-by-side comparison of the identified archetypes must consider the relative impacts of the model terms. The archetypes are summarized by average impact of the model terms in Figure 8. Also, each archetype has been given a summary label to replace the society cluster number, from Developing Economies (1) to Strategic Port Cities (6).

Strategic Port Cities

These cities attract IGFDI by virtue of their strong Business Preparedness and relatively high Ease of Child-Rearing, owing to well-ordered social conditions, low birth rates, and sufficient wealth to ensure a consistent food supply. The main feature differentiating this archetype from the Developed Economies is their somewhat higher level of Uncertainty Avoidance Values.

Developed Economies

This archetype features the lowest level of Uncertainty Avoidance Values, in conjunction with relatively high Business Preparedness. Within the model developed here, this combination detracts from IGFDI. It is worth considering that in these economies the Mergers and Acquisitions form of FDI may be more preferable to investors, as opposed to greenfield investments aimed at building new facilities.

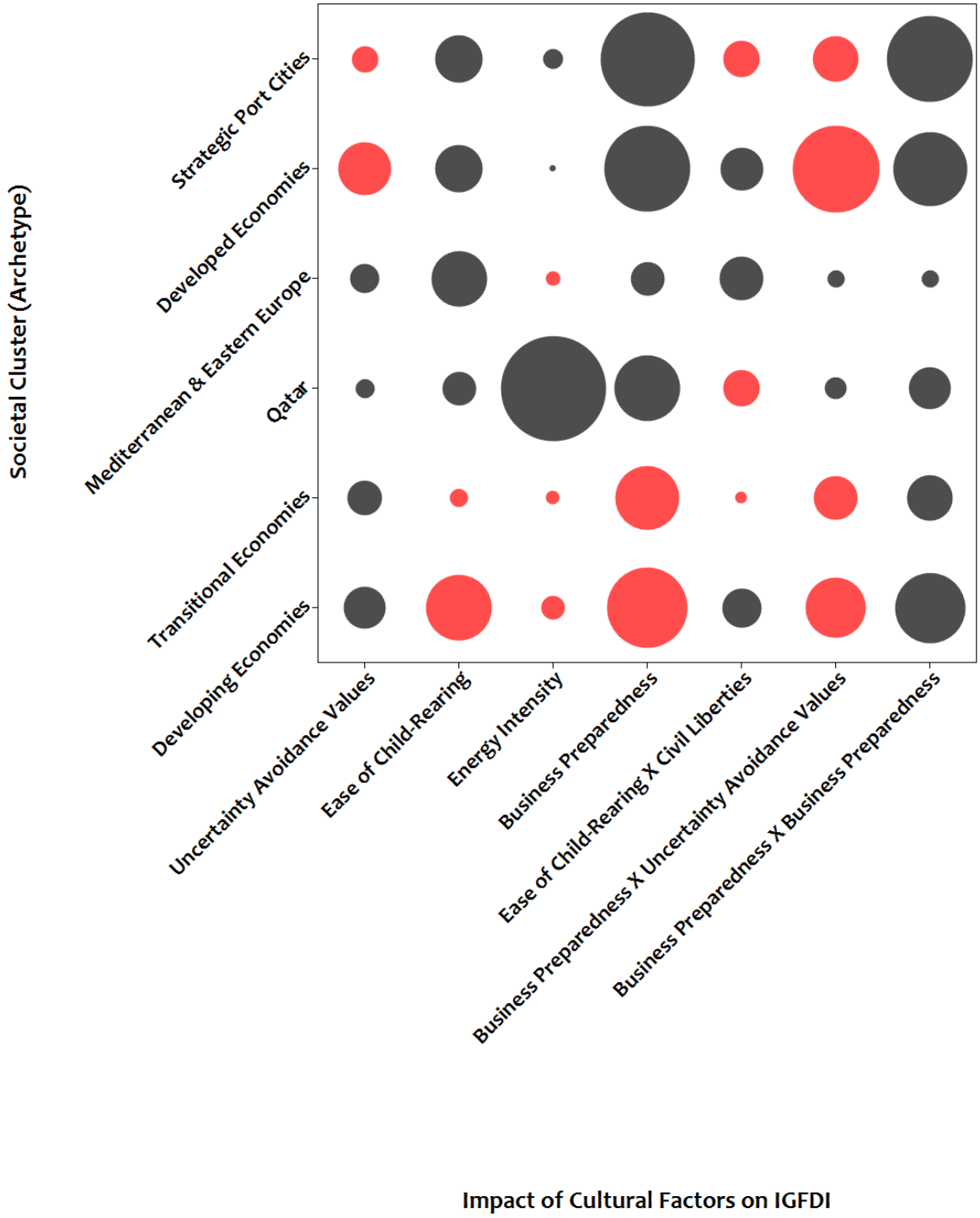

Figure 8: Impact of model terms. Bubble size represents absolute value of product of average thematic variable by archetype and corresponding standardized model effect. Bubble color represents sign: black is positive, red is negative. Sum of black area less red area across all key cultural factors yields relative IGFDI for each societal cluster.

Mediterranean & Eastern Europe

The main difference between this group and the Developed Economies is a relative lack of Business Preparedness. Ease of Child-Rearing is highest in this group, owing to low population growth and well-ordered societies, and that dominates the contribution to IGFDI in the model.

Qatar

This special case archetype is heavily dominated by the contribution of Energy Intensity. The contribution of that theme to the other archetypes is almost negligible. This finding merely reinforces that Qatar is anomalous, receiving extreme per capita IGFDI by virtue of the opportunities generated in exploiting its oil and natural gas reserves. If Energy Intensity were disregarded as a differentiating factor, Qatar would appear most similar to the Mediterranean & Eastern Europe archetype.

Transitioning Economies

Relative to the Mediterranean & Eastern Europe group, the Transitioning Economies suffer from moderately lower Ease of Child-Rearing and Business Preparedness.

Developing Economies

This archetype is characterized by the most challenging environment for Child-Rearing and the lowest Business Preparedness, leading to the lowest IGFDI. The second-order term for Business Preparedness acts to mitigate what would otherwise be a very low level of IGFDI. This could be viewed as a spurious artifact of model development; however, there is an alternative rational argument. Poor Business Preparedness is a logical deterrent to investors, but extremely poor Business Preparedness also indicates a significant frontier, where lack of sound structure for commerce engenders elevated risk, but some investors will judge the opportunities to outweigh the risks.

Concluding Remarks

The two most critical themes driving differences in the flow of IGFDI to six cultural archetypes have been determined to be Business Preparedness and Ease of Child-Rearing. These findings were obtained by consolidating a data set comprising 90 separate input variables. First, 18 highly correlated thematic data groups were generated. These were culled to 11 differentiating themes, and then finally, to five themes governing IGFDI. Among these, Energy Intensity strongly impacts only one archetype representing a single unusual nation. A strong systematic element is embedded in this thematic approach, but a measure of subjective judgment is also essential. Ultimately, the findings are plausible and readily understood from an ordinal or qualitative perspective, at the expense of modest losses in quantitative fidelity.

The potential benefit of a broader sociological perspective than currently utilized by UNCTAD in generating its predictions for the flow of FDI has also been established. Many slowly-changing factors may influence the flow of FDI in ways that a model based on readily measurable immediate markers of economic activity cannot capture.

In conclusion, an approach to developing contextually framed cultural archetypes has been demonstrated using inward-flowing greenfield foreign direct investment as an example. The approach is inherently semi-quantitative, and it demands a measure of subjective judgment in order to characterize relevant themes. But it also extracts intuitive clarity from voluminous data. This clarity can stand on its own as a useful tool for many purposes, or be an initial step toward the development of new hypotheses to be evaluated with more quantitative tools.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2012-6513C.

References

1. de Mooij, M.K., *Global Marketing and Advertising: Understanding Cultural Paradoxes*. Third ed2009: SAGE Publications, Inc. 344 pages.
2. United Nations Conference on Trade and Development: Division on Investment and Enterprise, *UNCTAD Training Manual on Statistics for FDI and the Operations of TNCs*, 2009, United Nations: Switzerland. p. 162.
3. United Nations Conference on Trade and Development, *Trade and Development Report, 1981-2011*, 2012, United Nations: Switzerland. p. 138.
4. UNCTAD, *Annex Table 28: Inward FDI Performance and Potential Index ranking, 1990-2010*, 2011.
5. House, R.J., et al., eds. *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. 2004, SAGE Publications, Inc.: Thousand Oaks, California. 848 pages.
6. UNCTAD, *Web table 19. Value of greenfield FDI projects, by destination, 2003-2011*, 2012.