

What Do Fragility Indices Measure?

Assessing Measurement Procedures and Statistical Proximity

Sebastian Ziaja

Forthcoming in *Zeitschrift für Vergleichende Politikwissenschaft (ZfVP)*

Abstract

This article examines nine fragility indices. Their content validity is assessed by reviewing conceptualization, measurement and aggregation methods. Their convergent/divergent validity is assessed via principal component analysis and multidimensional scaling. These techniques are capable of determining the dimensionality of and the statistical proximity within the examined sample of indices. Both the conceptual and the statistical analysis support the hypothesis that there is a group of “holistic” fragility indices which are of little use for investigating the causes and consequences of fragility. The remaining indices address more specific aspects of fragility and produce empirically distinguishable results.

Keywords: fragility, measurement, multivariate methods

Zusammenfassung

Dieser Artikel untersucht neun Fragilitätsindizes. Die Inhaltsvalidität der Indizes wird durch die Überprüfung von Konzeptualisierung, Messung und Aggregationsmethoden bewertet. Die konvergente/diskriminante Validität wird mittels Hauptkomponentenanalyse und multidimensionaler Skalierung bewertet. Diese Verfahren erlauben die Untersuchung von Dimensionalität und statistischer Ähnlichkeit innerhalb der Gruppe von Indizes. Sowohl die konzeptionelle als auch die statistische Analyse stützen die Hypothese, dass es eine Gruppe „holistischer“ Fragilitätsindizes gibt, die für die Erforschung der Ursachen und Folgen von Fragilität von geringem Nutzen sind. Die restlichen Indizes befassen sich mit spezifischen Aspekten von Fragilität und produzieren empirisch unterscheidbare Ergebnisse.

Schlüsselwörter: Fragilität, Messung, multivariate Verfahren

Sebastian Ziaja

German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE), Tulpenfeld
6, 53113 Bonn, Germany

and University of Essex, Department of Government, United Kingdom

email: sebastian.ziaja@die-gdi.de

1 Introduction¹

“Fragility” has become a catchword in international development and security debates. It refers to the issue of dysfunctional states and societies, which Western donors have rediscovered as one of the major global challenges. Fragile states are said to threaten industrialized nations by providing safe-havens for terrorists. They are said to pose a central obstacle to fighting poverty and securing peace across the globe, thus potentially provoking migration (OECD 2008, p. 11). The increased importance of fragility in the perception of the development community gave rise to a number of indices attempting to quantify the concept. These indices have not yet been systematically analyzed regarding their utility for quantitative research. This article intends to close this gap by presenting a comparative analysis of nine fragility indices.

Fragility indices are widely considered to be a welcome addition to the family of social science datasets (Marshall 2008, p. 2). But using them without previous scrutiny is a dangerous practice and may corrupt both qualitative and quantitative results. Even low degrees of random error – which are often accepted by researchers as long as better data is not available – can cause serious problems in regression analysis (Herrera and Kapur 2007, p. 7; Jackman 2008, 126-128). Our knowledge about the quality of current fragility indices is still fragmentary and mostly limited to conceptual issues.² Only one recent publication offers a statistical critique, focussing on the aggregation and ranking practice of some fragility indices (Gutiérrez Sanín 2009). Some individual indices have been scrutinized more profoundly. The best-researched one is, probably due to its comparatively old age and for being part of the widely-used Worldwide Governance Indicators (WGI), the WGI Political Stability and Absence of Violence index (Kaufmann et al. 2008).³ Other studies are concerned with stock-taking of potentially useful indicators (USAID 2006) or they focus on “early warning approaches” (Nyheim 2009), which are often of qualitative nature and do not lend themselves to immediate application in quantitative research. For the full set of quantitative fragility

¹ For comments and advice, I would like to thank Jörn Grävingholt, Javier Fabra Mata, participants of the ZfVP Workshop “Indizes in der Vergleichenden Politikwissenschaft” in May 2009 in Leipzig and two anonymous referees.

² Conceptual critiques of fragility indices can be found in Marshall 2008; Wulf and Debiel 2009; Call 2010; and Wennmann 2010.

³ Some of the most prominent critiques are Arndt and Oman 2006; Kurtz and Schrank 2007; Thomas 2007; Langbein and Knack 2010.

indices that is available today, there is no systematic analysis of their measurement procedures or of their statistical properties.⁴

The central question of this article is: What do fragility indices measure? To answer this question, I first examine the measurement procedure of each index with an analytical framework developed by Gerardo Munck and Jay Verkuilen (2002; 2009). It is assessed whether the conceptual logic of the indices is maintained in the procedure that leads to the generation of the index scores. In the methodological literature, this criterion has been labelled “content validity” (Adcock and Collier 2001, p. 538-540) or “concept-measure consistency” (Goertz 2006, p. 95-127). I then examine the statistical proximities between indices to determine whether similarities and differences in index design are reflected accordingly in the scores of the indices. It is assessed whether indices that were most similar to each other in the conceptual analysis are also similar to each other in the statistical analysis, and whether indices that should measure different things according to their conceptual design also differ in the scores they produces. This criterion has been labelled “convergent/discriminant validity” (Adcock and Collier 2001, p. 540-542).

I conduct the statistical analysis only within the set of fragility indices, although convergent and discriminant validation is usually conducted across a range of more diverse variables. I do not examine the indices’ proximity to democracy, economic well-being or other concepts because the set of fragility indices is in itself sufficiently diverse for this kind of analysis. Given the large number of indices examined, I do not statistically assess the internal validity and reliability of the indices and I do not discuss all relevant measurement decisions in detail. As most indices do not provide original data, I do not statistically assess the impact of method factors on indices either.⁵

I choose the approach described above to provide a first joint validation of the whole spectrum of fragility indices. The article gives researchers some guidance on the conceptual and statistical properties of these indices and identifies the most relevant patterns among them. It does not provide an ultimate assessment of the validity of fragility indices because any assessment of measurement quality needs to be made with reference to the intended application (Adcock and Collier 2001, p. 533). There is no appropriate reference application in the young field of fragility research yet. Beyond its substantial interest, this article also has a methodological focus. Network analysis and multidimensional scaling are applied. These

⁴ Some earlier parts of this article and more descriptive information on the indices can be found in the *Users’ Guide on Measuring Fragility* (Fabra Mata and Ziaja 2009).

⁵ See Bollen 1993 for such an approach applied to democracy indices.

techniques are useful for conveniently visualizing various properties of datasets, but they have seldom been used for the analysis of political science indicators.

The article proceeds as follows: In the next section, indices which potentially measure fragility are identified. In section three, the measurement procedures of these indices are assessed. In section four, multivariate techniques are used to assess the statistical proximity of the index scores. In section five, the implications of the analysis for the validity of individual indices are discussed. The conclusion summarizes the results and proposes ways to improve the measurement of fragility.

2 Defining fragility

“Fragility” is not a well-defined term. It emerged in the development and security discourse to describe a situation in which the state is “not capable or willing” to promote development and to provide security (e.g., DFID 2005, p. 7). It is often unclear whether fragility is attributed to a society as a whole or only to the state and its institutions (cp. Fabra Mata and Ziaja 2009, p. 5-6). Affected states are labelled “fragile states”, a term which summarizes various previously used denominations referring to countries with a dysfunctional, deteriorating or absent central authority, including “weak”, “failed”, “failing” and “collapsed states”.⁶ The renewed focus on the state as a guarantor of stability had begun in the 1990s, when state authority in Haiti and Somalia disintegrated (e.g., Helman and Ratner 1992). But only after the terrorist attacks on 11 September 2001, the term “failed states” entered the mainstream discourse. Soon after these events, the National Security Strategy of the United States declared: “America is now threatened less by conquering states than we are by failing ones” (The White House 2002, p. 1). The perception that state fragility threatened not only the populations of the affected countries (causing poverty, hunger and violence) but also those of the developed countries (through terrorism, uncontrolled migration and the degradation of globally relevant natural resources) put fragility high on the agenda of the international community (e.g., OECD 2008; ERD 2009). As the debate had started in the field of policy making, many definitions are comprehensive – in order to accommodate various political interests – rather than parsimonious and operational. The debate often conflates the bandwidth of states with the strength of states (cp. Fukuyama 2004, p. 6-14). It also conflates willingness and capacity (cp. Ziaja and Fabra Mata 2010, p. 2).

A rather new term in the field of political science, “(state) fragility” has not been through an academic debate as intense and rigorous as “democracy” or, more recently, “governance”. In

⁶ See OECD 2008. For discussions of various terminologies, see Gros 1996 and Rotberg 2004.

the 1980s, the state had been “brought back in”, but the relevant strand of literature focused more on developed than on poor nations (e.g., Evans et al. 1985). Some promising approaches to explaining the nature of “weak states” can be found (Jackson and Rosberg 1982; Migdal 1988; Desch 1996; Holsti 1996). They have been largely ignored by fragility indices.⁷ Explicitly academic publications on the issue of fragility are mostly recent and always highly policy-oriented (Brinkerhoff 2008; Carment, Prest and Samy 2009). Finally, the debate has no quasi-consensual minimum definition as democracy research has in polyarchy (Dahl 1971).

A minimum working definition of fragility is, however, necessary to serve as a baseline when analyzing the indices. I derive such a definition from Max Weber’s (1976) pivotal definition of the state as the bearer of the monopoly on the legitimate use of force. State fragility is thus the degree to which the state is lacking a monopoly on the legitimate use of force.⁸ The state requires a certain amount of repressive capacity to uphold this monopoly. The amount of repressive capacity that is required depends on the resistance that non-state actors apply to counter the state’s drive for predominance. This conceptualization allows the state’s optimal strength to differ: a state with low repressive capacity can be less fragile than one with high repressive capacity if the former is confronted with a population that is disproportionately less inclined to resist their state, compared to the population of the latter.⁹ It would thus be necessary to measure the monopoly of force in a way that considers both repressive capacity and repressive necessity.¹⁰

Before turning to the analysis of fragility indices, their population has to be defined. There is no generally acknowledged list of fragility indices which would designate the universe of cases to be analysed. Here, I define those indices as fragility indices which are intended to measure fragility (a teleological criterion) and those which are used to measure fragility (a practical criterion). Only indices which are publicly available and free of charge have been considered.¹¹

Table 1 lists the concepts which the selected indices intend to measure. Four indices comply very clearly with the teleological criterion: The Country Indicators for Foreign Policy (CIFP) Fragility Index, the Failed States Index, the Index of State Weakness in the Developing World

⁷ An exception is the CIFP Fragility Index (Carment, Prest and Samy 2009, p. 84-89).

⁸ This article will not deal in greater detail with the issue of legitimacy; for the relationship between legitimacy and fragility, see Bellina et al. 2009; Call 2010.

⁹ The idea of relative capacity is, of course, not new; see, for example, Kugler and Domke (1986) who measure “relative political capacity” as actual over expected extraction of resources.

¹⁰ One could also start with Charles Tilly’s (1985) concept of the state as a protection racket and arrive at a similar operationalization.

¹¹ Fabra Mata and Ziaja (2009, p. 109-111) provide a list of all indices reviewed for potentially being fragility indices.

and the State Fragility Index. They have been published with the explicit purpose of identifying fragile, failed or weak states. All of their definitions refer to the monopoly of violence or, closely related, the state's capacity to uphold security. The Bertelsmann Transformation Index (BTI) "Weak Stateness Index" is a side-product of a more comprehensive index measuring democracy and market economy.¹² It identifies states which do not have "functioning administration structures" and which can not "secure [their] monopoly on the use of force" (Bertelsmann Stiftung 2007, p. 6) and thus qualifies as a fragility index.¹³ The World Bank's approach to identifying fragile states with the International Development Association's (IDA) Resource Allocation Index is now officially acknowledged on the organization's website (World Bank 2010).¹⁴

[Table 1 about here]

The three remaining indices do not qualify as fragility indices without doubt. Both the Political Instability Index and the WGI Political Stability index refer to political (in-)stability, which is assumed to occur in fragile states with a higher probability than elsewhere. The same is true for violent conflicts. Their probability is measured by the Peace and Conflict Instability Ledger (PCIL) Risk Ratio. I include these indices because the WGI Political Stability index is used as a measure of fragility in the literature (e.g. Bratton and Chang 2006) and because the other two could be considered appropriate by the literature with the same reasoning.

The origin of the indices is one factor which is not covered by my analytical framework but which must be considered when interpreting the indices. Table 2 lists the index producers. The producers are non-profit, with the exception of the Economist Intelligence Unit, a British advisory company which forms part of the Economist media group. Of the remaining indices, three are university-made (in the United States and Canada), two are published by the World Bank, two by US-think tanks and one by a German private foundation which is closely related to a commercial media group.¹⁵ Table 2 also cites the index publications and informs about

¹² For an assessment of the overall BTI index, see Müller and Pickel 2007.

¹³ The index is used for a map in a brochure and mentioned in the documentation of the methodology (Bertelsmann Stiftung 2007; 2008). The producers say, however, that they would construct a "proper index of state fragility" differently (personal communication).

¹⁴ It is, for example, used in the upcoming World Development Report "Conflict, Security and Development" (<http://wdr2011.worldbank.org/>) and by the International Monetary Fund (IMF 2010).

¹⁵ See Arndt 2008 on the organizational motivations for publishing governance indices.

data availability and coverage. Most fragility indices have a very short temporal coverage, starting only in the past five to fifteen years.¹⁶

[Table 2 about here]

3 The measurement procedure

The analysis of the measurement procedure has a twofold purpose: It is an analysis carried out in its own right to assess the content validity of the indices; and it prepares the ground for the multivariate analyses, as their results can only be interpreted in a meaningful way after clarifying how the scores of the individual indices were generated. I will interpret the results from both stages of the analysis jointly at the end of the paper.

The literature on measuring democracy is one of the most advanced literatures dealing with the cross-country measurement of social science concepts. I will apply a well-received framework from this literature, described by Gerardo Munck and Jay Verkuilen (2002; 2009) and further specified by Thomas Müller and Susanne Pickel (2007), to analyze the measurement procedure of the indices under investigation. This procedure comprises three challenges: conceptualization, measurement and aggregation (Munck and Verkuilen 2009, p. 15). Conceptualization refers to the task of concept specification, i.e. identifying as many components as necessary and as few as possible to describe a concept, and of conceptual logic, i.e. avoiding redundancy or conflation within the “concept tree” (the hierarchical representation of the relationship between the components). The components are referred to as concept “attributes”. Attributes are the constitutive elements of a concept and have to be identified correctly to achieve valid measurements. Measurement refers to the translation of concept attributes into numerical data.¹⁷ The deviation of the data from the unobserved “true” values should be minimized for both systematic and random deviation, i.e. the measurement should be valid and reliable. Aggregation refers to the concentration of the data of the attributes into a joint score that retains validity and reliability with respect to the underlying concept. All steps of the measurement procedure have to be documented in a way that allows other scholars to repeat them, i.e. the measurement procedure should be replicable.

¹⁶ This is the major reason why the World Bank uses the IDA Resource Allocation Index to measure fragility: for internal purposes, World Bank staff can draw on a time-series starting in the 1970s.

¹⁷ “Measurement” thus refers to measurement in the narrower sense, limited to the quantification of single components, while “measurement procedure” refers to measurement in the broader sense, including previous conceptualization and subsequent aggregation.

I will focus only on the most salient issues when presenting the results of this analysis. Given the large conceptual differences between indices, I do not provide an “index of fragility index quality” as Müller and Pickel (2007, p. 517-518) provide for democracy indices; it would be more misleading than helpful in this case.

3.1 Conceptualization

Which attributes do fragility indices consist of and how are they organized? Table 1 gives an overview of the concepts that fragility indices intend to measure. Some concepts are not sufficiently systematized to derive the index attributes, which disrupts the linkage between concept and measurement at the very start. Systematized or not, most indices refer to state functions. For example, the CIFP Fragility Index defines fragility as “the extent to which the actual institutions, functions, and processes of a state fail to accord with the strong image of a sovereign state” (Carment, Prest and Samy 2009, p. 84). But which functions is this “sovereign state” expected to satisfy? One function which is mentioned frequently is the monopoly on the use of force. Indices also refer to “public services” (Fund for Peace 2009b), “legitimate, transparent and accountable political institutions” and “fostering [...] economic growth” (Rice and Patrick 2008, p. 3). This is much broader than the working definition sketched out above. I will label the indices with the most comprehensive definitions “holistic” indices. They include the Failed States Index, the Index of State Weakness, the State Fragility Index and the CIFP Fragility Index. Looking at the operationalizations presented in Table 3 below supports this categorization: these indices cover a broad range of attributes related to security, governance, social welfare and economic development.

The supposedly holistic indices are all motivated by policy concerns. As their primary objective is identifying states that pose obstacles to promoting development and to consolidating peace, they are very inclusive and aspire to cover a broad range of deficiencies which can affect a state and its society. This may be a welcome trait when looking for an answer to the question of which states the international community must worry about most. But it does also involve a decision which is difficult to defend: The indices have to trade political, social, economic and security concerns off against each other. While these factors do covary in practice, they do not always occur in the same constellation. Given two comparable countries, one affected by a small but violent conflict, one by a large-scale economic crisis, it would require careful theoretical consideration or substantial empirical evidence to

convincingly argue how these countries perform with respect to each other. Such considerations are not provided by the indices.¹⁸

There are also concerns regarding the nature of individual attributes of the holistic indices. “[R]easonable public services” (Fund for Peace 2009b) need to be defined against a benchmark. Following the argumentation leading to the working definition of fragility presented above, it would make sense to define these social welfare attributes relative to the expectations of a population. States can be stable for long periods of time without providing, for example, universal health care, if the population does not expect such a service.

How do the other indices perform? The BTI Weak Stateness Index provides the most parsimonious definition, as it is limited to the attributes “functioning administration structures” and “monopoly on the use of force” (Bertelsmann Stiftung 2007: 6). From an academic standpoint, this is a welcome trait, since a certain degree of parsimony is a prerequisite for quantitatively distinguishing concepts and investigating the causes and consequences of state fragility. The four holistic indices are not suitable for this task because they include assumed causes and consequences of state fragility in their definitions (cp. Gutiérrez Sanín 2009, p. 5). Due to their model-based design, the PCIL Risk Ratio and the WGI Political Stability index do not define attributes in a conceptual sense. They just use indicators to feed their models. The authors of the PCIL Risk Ratio use the same indicators as the “global model for forecasting political instability” published by the Political Instability Task Force (Goldstone et al. 2010). The model is based on one of the largest data mining projects in the field of political science and aims at predicting various types of civil war or political violence. Two other indices refer to the Political Instability Task Force as a source of inspiration: the Political Instability Index and the State Fragility Index (EIU 2009c; Marshall and Goldstone 2007: 3). This is in a way surprising, as these indices aim at measuring “social unrest” and “fragility” respectively, which is not necessarily equivalent to civil war or political violence. The State Fragility Index does not provide a descriptive definition of its subject of interest, only an operational one. The attributes of the IDA Resource Allocation Index all refer to policies and institutions, which could the index more sensitive to sudden changes compared to the structural indices. Unfortunately, most attributes of the index go far beyond the core issues of fragility and include elements such as economic policy. The full list of attributes for all indices is presented in Table 3 below (which also informs about the aggregation methods).

¹⁸ The literature on measuring democracy provides examples on how multi-dimensional concepts can be convincingly operationalized by building on a mature corpus of theory (e.g. Bollen 1993; Munck 2009).

An example of violating the conceptual logic, i.e. the combination of attributes, is provided by the Index of State Weakness. The producers, in an attempt to build their index in “a straightforward and transparent fashion” (Rice and Patrick 2008, p. 7), measure each of their attributes with five indicators (see Table 3). It is, however, improbable that each of these dimensions is best measured with exactly five indicators. This increases the danger of conflation (e.g., within the “political basket”, the WGI Government Effectiveness index and the WGI Voice and Accountability index) and of redundancy (e.g., the WGI Government Effectiveness index could also refer to the “social welfare basket” as it includes judgements about social infrastructure).

3.2 Measurement

How do the indices translate their attributes into numerical data? The first issue to resolve is whether the indices select the right indicators to measure their attributes. Many do not discuss the selection of indicators, which makes them very vulnerable to criticism. Even otherwise highly transparent indices do not explain their choice of indicators in detail. In the case of WGI Political Stability index, for example, only a list of (apparently reasonably selected) indicators is provided, but no justification (Kaufmann, Kraay and Mastruzzi 2008, p. 73). Confronted with repeated criticism, the authors of the Worldwide Governance Indicators have published several responses which discuss general issues of indicator validity in detail (Kaufmann, Kraay and Mastruzzi 2007; Kaufmann and Kraay 2008). One of these is the allegation that their indices are subject to a cultural bias inherent in their indicators which are based on expert survey data. Across all indices, however, there is little or no documented deliberation on whether indicators actually measure the right thing, regardless of whether they draw on 83 indicators (CIFP Fragility Index) or on two (BTI Weak Stateness Index).

Another issue of selection regards the measurement of “political” attributes. Legitimacy is here measured with democracy indices. Regardless of the fact that legitimacy may come in alternative institutional shapes, one question arises: must repressive but stable regimes be considered fragile, just because it is assumed that, in the long run, they will not be able to accommodate social demands as good as democracies can? It could be more useful to use the term “fragile” only with regard to countries with incapable governments that are likely to break down soon. If legitimacy is to be included, its role and measurement must be better justified.

The selection of indicators is, however, not only determined by conceptual reasoning. The choice is severely restricted by data availability. Many indicators which would be valuable for measuring fragility are not available for the most affected countries. In the field of security,

for example, crime rates and police coverage are not available (Rotberg and Gisselquist 2008: p. 44-46). When approximately comparable public statistics are available, they are often the only source of information and can thus not be validated with alternative sources. One example is the Inter-Agency Child Mortality Estimation Group (IACMEG), constituted jointly by the World Health Organization, UNICEF, the World Bank, the UN Population Division and others (You et al. 2010). While it is laudable that institutions join forces to produce the best-possible data, such a concentration makes cross-validation impossible. Child mortality is used by all holistic fragility indices, the PCIL Risk Ratio and the Political Instability Index.

Looking at the larger picture, which data sources do fragility indices draw on? Figure 1 represents the relationships between sources and indices. Indices are represented by dark grey circles, institutions providing data (“sources”) by light grey circles. The radius of the circles is a logarithmic function of the number of indicators used or provided.¹⁹ The grey lines indicate that an index uses indicators from the respective source. The width of each line represents the impact which a source exerts on the overall score of a particular index. The black arrows that connect the fragility indices point at those indices which use indicators from another index. Since there are several World Bank sources and indices, they are displayed individually. They are grouped together and encircled with a black outline.

[Figure 1 about here]

The graph is based on a matrix which includes all indices, indicators and data sources. A spring-embedder algorithm reduces the multidimensional information contained in this matrix to fit it on a plane and to avoid overlap of elements in order to create a well-arranged visualization which is at the same time a good representation of the underlying data. More connected nodes are positioned at the centre of the graph, less connected ones at the border.²⁰ Remaining overlaps have been removed by hand.

The network unveils that fragility indices are heavily dependent upon data provided by international organizations, mainly the World Bank and United Nations agencies. Commercial providers of data are important for the Political Instability Index and the WGI Political Stability index. The former is produced by the Economist Intelligence Unit, which at the same time provides the bulk of expert data used in the index. The WGI Political Stability index

¹⁹ For a detailed treatment on visualizing data in network graphs, see Krempel 2005.

²⁰ With the exception of the IDA Resource Allocation Index and the Country Policy and Institutional Assessments, which have been moved closer to the other World Bank sources to group them together.

draws on several commercial and non-commercial expert polls. Academic sources play a comparatively small role, with the exception of a cluster based at the University of Maryland and the George Mason University. This cluster emerges from a number of institutional and personal overlaps and consists of, *inter alia*, the Center for Systemic Peace including the Polity IV project, the Political Instability Task Force, the Minorities at Risk Project and the State Fragility Index. Three indices stick out for providing original data: BTI Weak Stateness Index, the IDA Resource Allocation Index and the Failed States Index. The former two conduct expert assessments. The latter is fed by a “Conflict Assessment System Tool” which performs a content analysis on a large corpus of electronically available news. Unfortunately, the producers do not uncover the entire data generation procedure. They state that the index scores are also based on public statistics and expert calibration rounds, without specifying the impact. The IDA Resource Allocation index relies on the Country Policy and Institutional Assessment (CPIA), which is produced by World Bank experts with the explicit purpose to transparently allocate aid funds from the Bank’s International Development Association (IDA). The BTI Weak Stateness index uses data from one source only: the BTI Country Assessments. These assessments are carried out by one expert and one reviewer, followed by calibration rounds. The overall purpose of the BTI is measuring market-oriented democracy. It is possible that this source suffers from “halo effects” which make the coders unconsciously contaminate specific questions with the overall goal of the questionnaire.

Overall, fragility indices draw data from a common pool rather than creating or uncovering competing sources. While each index has individual sources, all except the BTI Weak Stateness Index draw on the central nodes to a large degree (World Bank, United Nations agencies, Economist Intelligence Unit, Center for Systemic Peace), making it improbable that resulting scores will differ significantly. Most indicators are highly aggregate socio-economic measures which may be proxies for theoretically relevant attributes of state fragility, but which lack precision in distinguishing these from other phenomena like human development.

[Table 3 about here]

3.3 Aggregation

How are the scores of the individual attributes combined into an overall score? Most indices apply similar aggregation methods based on the simple addition of equally weighted attributes (see Table 3). Additive aggregation rules imply that low values on one indicator can be partly compensated by high values on another indicator (Munck 2009, p. 50). Whether

compensation is valid depends on the number of dimensions specified in the index concept. Aggregation can be justified either by claiming that all selected indicators describe the same unidimensional phenomenon, or by claiming that the indicators, even though they do not covary, are parts “combined to form a whole” (Munck and Verkuilen 2009, p. 30). The WGI Political Stability index clearly pursues the unidimensional strategy. The four holistic fragility indices pursue the parts-of-a-whole strategy. In a unidimensional setting, it is easy to argue that one indicator may compensate for another: a skirmish with secessionist rebels and a political assassination are similar to some extent. In a multidimensional setting, the justification of allowing indicators to compensate for each other is more problematic: which degree of child mortality compensates for which level of corruption? Such considerations have to be made for any multidimensional index. The justification is difficult when no common unit of measurement is available, no “numeraire” (cp. Gutiérrez Sanín 2009). The holistic fragility indices push the envelope quite far without providing sufficient justification. For example, the CIFP Fragility Index consists of six equally weighted attributes which are aggregated by taking the arithmetic average (i.e. addition and subsequent rescaling). A country reaches 83% of the maximum (inverted) score if it fails completely in the security & crime attribute while receiving maximum scores in all other attributes. This is empirically improbable, but it depicts the dangers of using additive aggregation procedures when faced with attributes which can hardly be compensated for by other attributes. A second issue in this context is heterogeneity of countries with equal scores. If countries with the same score differ significantly across the attributes, it is difficult to argue that the aggregation of the attributes is justified: What do we learn if Saudi-Arabia and Somalia, for substantially different reasons, are put in the same category? In general, fragility index methodologies do not discuss the robustness of scores with regard to alternative aggregation choices.²¹ Measures of uncertainty that result from the aggregation process are only provided by the PCIL Risk Ratio and the WGI Political Stability Index. This information is essential for deciding whether one country can be considered less fragile than another, or whether the difference could be due to chance. While unviable weighting methods may distort results strongly, as Treier and Jackman (2008) show for the Polity IV dataset, weighting is not discussed sufficiently by most indices. The authors of the Index of State Weakness declare that there is no “formula” on how to weight components of fragility indices. They choose to go with equal weighting (Rice and Patrick 2008, p. 26, endnote 33). Most other indices follow the same approach, forgetting that equal weighting is a decision that needs substantive justification nonetheless. Moreover, there is a

²¹ An exception is the WGI Political Stability Index.

weighting effect of differently sized categories inherent in several indices: indicators in less populated categories carry more weight than indicators in categories which are fed by many indicators (see Table 2). Extensive considerations on weighting are made by the authors of the PCIL Risk Ratio and the WGI Political Stability index. These indices are based on models with data-driven weighting schemes.

All index providers could improve transparency of documentation and the provision of data to enable replication. Previously, the PCIL Risk Ratio performed excellent by providing the full dataset plus replication scripts for statistical software online. Since the publication of the 2010 edition, the links have unfortunately disappeared. The BTI Weak Stateness Index, the World Bank indices and, more recently, the Failed States Index provide disaggregate data in spreadsheet format for download (see Table 2). The Index of State Weakness and the State Fragility Index perform worst by not even providing data in HTML-format on their websites, but only within the respective PDF publications, which are not easily transformed into data files.

4 A multivariate analysis of the index scores

Having retraced the measurement procedure of fragility indices, I will now check whether those indices which have shown similar conceptual traits also have similar statistical traits. Unexpected deviations could be a sign for deficient validity or reliability. While a stand-alone convergence validation of different measures is always relative to unknown systematic biases jointly underlying these measures (Müller and Pickel 2007, p. 515), the preceding analysis of conceptualization, measurement and aggregation provides a frame of reference against which the results can be interpreted.

The violin plots in Figure 2 summarize the statistical properties of the indices. The grey area is a density plot representing an index's distribution. The black box reaches from the 25th to the 75th percentile. The white circle marks the mean. The scores were standardized to a zero to one “best to worst” scale to align them with each other. While the multivariate methods used here do not require any specific distributional properties, it can be argued that an index should represent the whole spectrum of possible values and – assuming an approximately normal distribution of the “true” underlying values – not end abruptly with a large number of cases clustered at either end of the scale (Goertz 2008, p. 115). Most indices have reasonable distributions. The PCIL Risk Ratio is not capable of distinguishing at the lower end (non-fragile states); the estimated risk is very low for a large number of states. The BTI Weak Stateness Index and the State Fragility Index end abruptly at zero, which indicates that they struggle with differentiating less fragile states.

[Figure 2 about here]

A traditional way of examining statistical proximity between variables are bivariate correlation tables. They are informative, but they are not very reader-friendly. Instead of correlation tables, I will use two multivariate methods which are appropriate for graphically representing statistical proximities, and which offer additional benefits such as identifying dimensionality: principal component analysis (PCA) and multidimensional scaling (MDS) (Borg and Groenen 2005). PCA is used to check the joint dimensionality of the fragility indices: How much of the joint variance of the nine indices analyzed can be projected onto one single component? Only if indices do not deviate strongly in this analysis, we can assume that they might be measuring the same (always provided that their concepts comply). PCA results for the two first components can be graphed on a two-dimensional plot. MDS is then used to take a closer look at the variance between the indices: How similar are the scores produced by each index? MDS projects data onto a number of factors that is fixed *ex ante*. The goal is to graph multiple dimensions on a plane or in a three-dimensional space, representing their variance as accurately as possible in the form of geometrical distances. Thus, MDS can display, depending on the model fit, almost the entire variance between all indices at once, not just the first two components. The disadvantage of MDS is that it does not allow an assessment of dimensionality as PCA does.

The PCA results propose that the indices can be represented on one or two dimensions (see Figure 3). The second eigenvalue is slightly less than 1, which is the criterion for counting a component as a separate dimension. This solution is plotted on the right hand graph, with index positions represented as linear combinations of the first two components. The closer an index is to the circle (of radius one), the better it is represented by these two components. It shows that all indices load strongly on component 1, which could thus be termed the “meta fragility” component. The IDA Resource Allocation Index and the PCIL Risk Ratio stick out. They span the second dimension into contrary directions.

[Figure 3 about here]

To better represent the relationship between indices and display as much variance as possible, I turn to MDS. The algorithm used in MDS starts in an initial configuration and tries to minimize the stress of that configuration by changing it iteratively. The configuration with the

smallest amount of stress best represents the similarities across indices. The first graph in Figure 4 shows a solution that interprets the index scores as ordinal variables, i.e. it examines similarities of the ranks between indices. Distances can be directly interpreted as dissimilarities. The axes and potential rotations of the graph are not relevant for the interpretation. The solution shows nicely that the core fragility indices rank countries in a similar manner, while the WGI Political Stability index, the PCIL Risk Ratio and the Political Instability Index rank countries differently and are thus more distant from each other. Given that the latter two are inspired by the PITF global model on predicting state failure, it may seem a bit surprising that one finds them that far apart. As noted above, however, they have different purposes: the PCIL Risk Ratio intends to measure the risk of being affected by violent conflict, the Political Instability Index intends to measure the risk of “social and political unrest or upheaval” (EIU 2009a, p. 15). This divergence leads to different operationalizations which explain the observed dissimilarity.²²

[Figure 4 about here]

The MDS results have to be interpreted with care, however. The the solving algorithm may get stuck in a local minimum under adverse circumstances, falling short of providing the optimal solution. This can be tested by running repeated stress optimizations with random starting configurations. The second graph shows such a solution, which has almost zero stress and all but three indices cluster on one spot. Such traits are a sign for a “degenerate solution” where the applied scale level does not provide enough distinctions for the algorithm to properly display dissimilarities (Borg and Groenen 2005, p. 219).

To produce useful results, one may increase the measurement level to an interval scale. This solution is shown in the third graph. The solution is again rotated (which has no meaning), but the same indices end up near the borders of the plane which supports the proposition that they are the indices with most dissimilar results (also in scores, not only in ranks). Again, the holistic fragility indices cluster, especially those based on similar data. The Failed States Index, partly informed by content analysis, is a bit removed from that cluster. The BTI Weak Stateness Index, even though it does not belong to the holistic fragility indices as defined on the outset, produces more similar results than the rather broadly designed Political Instability Index. This result may be due to actual similarities in the occurrence of a defective monopoly

²² A rather low bivariate correlation of 0.53 supports the MDS result. The advantage of MDS, however, is that all the different relationships can be grasped at once via the graphical display, dispensing of the need to comb through the correlation table row by row.

of violence and fragility in a broad sense, but it may as well be due to a bias affecting BTI-coders who are socialized with existing conflict and governance measures that are supposed to correlate highly with deficient stateness.

While results using interval scaling are robust to random initial configurations, differences in surrounding conditions limit the explanatory power of this analysis. Results may be distorted by the reduction of the sample size to 113, which is due to the limited coverage of some indices. Another possible source of distortion are differences in chronological coverage. The CIFP Fragility Index is, on the internet, available only for 2007. The Political Instability Index value had to be interpolated from the 2007 and 2009/10 scores, because no 2008 edition was published. Differences in the actual time-spans that the indices refer to could be another source of distortion. The WGI Political Stability index 2008 is mostly based on indicators from 2008, the Failed States Index 2008 is built from texts produced in the year 2007 and the indicators contributing to the 2008 Index of State Weakness have a time-lag of about two to four years. Despite these potential disturbances, the conceptual and statistical results are in general reconcilable.

5 What do fragility indices measure?

What conclusion can be made from the previous analyses regarding the validity of each of the nine fragility indices? The multivariate analysis could be interpreted as evidence for high rates of validity among the holistic fragility indices. Their similarity is not surprising, however, since they use almost the same data for very similar conceptual attributes. The fact that the Failed States Index is close in spite of relying partly on alternative data should not be overestimated, as the degree by which the score is determined through content analysis is not released by the producers. While measuring the same, it is not clear, due to conceptual and methodological deficiencies, what these indices actually measure.

Comparing the concepts with the working definition sketched above, most concepts are much broader. In the measurement process, assumptions about causes and consequences of fragility are implicitly introduced. Repressive capacity is mostly measured indirectly by GDP per capita (which leads to higher tax revenue which leads to manoeuvring space for governments) and similar measures which approximate the state's institutional capacities (e.g. WGI Government Effectiveness). Repressive necessity is measured indirectly by outcome indicators of social welfare, like health and education indicators. Citizens who receive services from the state feel less inclined to challenge its rule. But expectations regarding the provision of services differ, and no index is capable of providing relative measures of

repressive necessity. Only survey data or expert coded assessments of state-bandwidth expectations could provide this.

In a nutshell, how good are the individual indices? The *BTI Weak Stateness Index* stands out for trying to measure the monopoly of violence itself, which is desirable when looking for a narrow conceptualization of state fragility. A limited number of coders and a potential expert-bias may limit its reliability and validity. Nonetheless, the index is a useful alternative to violence indicators based on structural data. The *CIFP Fragility Index* has a good theoretical foundation. Especially the authority-legitimacy-capacity framework is promising. The way it is operationalized, however, leads to the most holistic measure of all indices. Its applicability in regression analysis is questionable. The *Failed States Index* is partly based on content analysis. It could be of great use for validating indices based on structural data. Unfortunately, its methodology has not been sufficiently documented and fully disaggregate data is not available. The content analysis results can not be separated from the other information that is used and the impact of the content analysis on the final scores is unknown. The *IDA Resource Allocation Index* measures policy, which could be an advantage with regard to indices purely based on outcome indicators. It does, however, include components such as economic liberalization which should not appear in measures of state fragility. Beyond validity concerns, the small geographical coverage of the openly available data reduces the index's applicability in quantitative research. The *Index of State Weakness* intends to be transparent and accessible, placing the symmetry of index design over validity. It does not introduce any relevant theoretical or methodological innovations. Overall, it is not recommendable for academic purposes. The *PCIL Risk Ratio* is based on an established model for predicting conflict and translates these results into an index. It does not comply with the minimum definition of fragility sketched above. For the purpose of measuring the probability of conflict, however, it should be more useful than other fragility indices. The *Political Instability Index* tries to measure social unrest based on a model predicting civil war, which casts doubt on its validity. Statistically, it measures something different from both the PCIL Risk Ratio and the holistic indices, but its conceptualization is not sufficiently documented to assess what it measures exactly. The *State Fragility Index* distinguishes effectiveness and legitimacy of various state functions, which is a welcome theoretical contribution. The overall index struggles with the same issues as the other holistic indices, although it is much more economical with indicators. The *WGI Political Stability Index* has a broad database and high construct validity due to their data-driven aggregation process; the lack of theoretical justification for the inclusion of certain indicators casts doubt on its validity to measure what

it claims to measure, i.e. the probability of unconstitutional change of government. What could be of use for researchers is that the index scores actually differ to some degree from the holistic indices' scores.

6 Conclusion

This article assessed the measurement procedures and the statistical proximity of nine fragility indices. The measurement procedures were assessed with an analytical framework distinguishing conceptualization, measurement and aggregation (cp. Munck and Verkuilen 2009). The statistical proximity was assessed using multivariate scaling techniques. Three indices proved very similar in conceptual and statistical matters: The CIFP Fragility Index, the Index of State Weakness and the State Fragility Index produce very similar results across a common sample of 113 countries. They use a broad spectrum of indicators, ranging from political to economic and social issues, to measure a rather holistic concept of fragility. The Failed States Index, building on a similarly holistic concept, draws on content analysis as an additional source of data and produces only slightly different scores. Regarding their maximalist concept specifications, however, it is difficult to argue that these measurements are valid representations of a useful definition of fragility. Various dimensions are integrated without sufficiently justifying why and to what amount economic issues, for instance, can compensate for security issues. Regarding the choice of indicators, some indices go far beyond immediate political influence and include even geographic factors, e.g. “arable/fertile land availability” (CIFP Fragility Index). It is interesting to see that the holistic fragility indices, in the similarity plot generated with multidimensional scaling, are centred between three more specific indices. This supports the claim that fragility, as defined by the holistic indices, adds up to a rather unspecific mix of development-adverse circumstances. Overall, the holistic indices include potential conditions for and consequences of fragility affecting both state and society, frustrating the possibility to investigate these mechanisms. All fragility indices are capable of distinguishing countries at the extremes of their scores, but this is only a limited value added compared to qualitative ad-hoc considerations. Using the holistic fragility indices in regression analysis is, in general, not recommendable. Some of their subcategories, however, might be of use for particular research questions (e.g. “political efficiency” from the State Fragility Index).

Although the ideas motivating research on state fragility are not new, much remains to be done to better quantify fundamental functional deficiencies of developing and developed states. Some issues could be resolved in the short run, using the available data. More adequate categorization and standardization methods would create more consistent typologies and

potentially time-invariant scores.²³ Increasing the validation efforts would make fragility indices more trustworthy, for example by nomological validation: How well do established hypotheses hold when tested with different indices?²⁴ Such assessments presuppose clearly stated, systematized concepts.

The central problem of fragility indices is their aspiration to measure the broad picture of many kinds of problems encountered in developing countries. While the broad picture might be convenient at times, measuring better specified “partial regimes”²⁵ of fragility is crucial for investigating its causes and consequences.²⁶ The proposal of Carment, Prest and Samy (2009) to distinguish “authority”, “capacity” and “legitimacy” is a good start, but their operationalization is too broad.²⁷

Some other conceptual clarifications are desirable, such as explaining the relationship between fragility and political instability or state capacity. Making the frame of reference explicit is another issue to be addressed (state vs. regime vs. government). It is also necessary to better specify the time interval in which fragility should and can be measured. The implications of subnational variation due to secessionist provinces or due to decreasing projection of state power beyond the capital also need to be considered, although measuring these on a global scale is beyond the reach of current data. Finally, it has to be clarified whether fragility is to be measured on an absolute scale or on a scale relative to a society's expectations. The latter option may be more sensible, but it is confronted with severe measurement problems. New expert assessments or global public surveys are two (imperfect) ways to proceed. In the long run, new original data is necessary to better measure fragility.

²³ See Fabra Mata and Ziaja (2010, p. 32) for an overview of categorization methods applied by fragility indices. An excellent discussion on methodological choices, including a convincing standardization method, can be found in the Index of African Governance (Rotberg and Gisselquist 2008).

²⁴ On nomological validation, see Adcock and Collier (2001, p. 542).

²⁵ Collier and Levitsky (2009) use this term with respect to democracy concepts.

²⁶ Coppedge (2002, p. 39) arrives at a similar conclusion with regard to democracy: “The highest priority for improving the measurement of democracy is therefore improving the measurement of disaggregated attributes of democracy.”

²⁷ Call (2010) tries to apply this approach to less broadly defined “capacity”, “security” and “legitimacy gaps”.

Literature

- Adcock, R. and D. Collier. 2001. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American Political Science Review* 95(3): 529-546.
- Arndt, C. 2008. The Politics of Governance Ratings. *International Public Management Journal* 11(3): 275-297.
- Arndt, C. and C. Oman. 2006. *Uses and Abuses of Governance Indicators*. Paris: OECD Development Centre.
- Baliamoune-Lutz, M. and M. McGillivray. 2008. State fragility: Concept and measurement. *UNU-WIDER Research Paper* 2008/44.
- Bellina, S., D. Darbon, S. S. Eriksen and O. J. Sending. 2009. *The Legitimacy of the State in Fragile Situations*. Oslo: Norwegian Agency for Development Cooperation (Norad Report 20/2009).
- Bertelsmann Stiftung (ed.). 2007. *Bertelsmann Transformation Index 2008: Political Management in International Comparison*. Brochure. http://bti2008.bertelsmann-transformation-index.de/fileadmin/pdf/Anlagen_BTI_2008/BTI_2008_Brochure_EN.pdf
- Bertelsmann Stiftung (ed.). 2008. *Bertelsmann Transformation Index 2008: Political Management in International Comparison*. Gütersloh: Verlag Bertelsmann Stiftung.
- Bertelsmann Stiftung (ed.). 2009. *About BTI, Ranking and Country Reports*. <http://www.bertelsmann-transformation-index.de/11.0.html?&L=1>.
- Bollen, K. 1993. Liberal Democracy. *American Journal of Political Science* 37(4): 1207-1230.
- Borg, I. and P. J. F. Groenen. 2005. *Modern Multidimensional Scaling: Theory and Applications*. 2nd Edition. New York: Springer.
- Bratton, M., and E. C. C. Chang. 2006. State building and democratization in sub-Saharan Africa: Forwards, backwards, or together? *Comparative Political Studies* 39(9): 1059-1083.
- Brinkerhoff, D. W. 2007. Governance in post-conflict societies: Rebuilding fragile states. New York: Routledge.
- Call, C. T. 2010. Beyond the "Failed State": Toward Conceptual Alternatives. *European Journal of International Relations*. Online first, doi:10.1177/1354066109353137.
- Carment, D., S. Prest and Y. Samy 2009. *Security, Development and the Fragile State: Bridging the Gap between Theory and Policy*. Abingdon: Routledge Studies in Intervention and Statebuilding.
- Carment, D., Y. Samy and S. Prest. 2008. State Fragility and Implications for Aid Allocation: An Empirical Analysis. *Conflict Management and Peace Science* 25(4): 349-373.
- Collier, D. and S. Levitsky. 2009. Democracy: Conceptual Hierarchies in Comparative Research. In: D. Collier and J. Gerring (eds.). *Concepts and Method in the Social Science: The Tradition of Giovanni Sartori*, London: Routledge. 269-288.
- Coppedge, M. 2002. Democracy and Dimensions. *Comparative Political Studies* 35(1): 35-39.
- Country Indicators for Foreign Policy (CIFP). 2009. Country Ranking Table 2007. http://www.carleton.ca/cifp/app/ffs_ranking.php.
- Dahl, R. 1971. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.
- Department for International Development (DFID). 2005. Why we need to work more effectively in fragile states. London: Department for International Development.
- Desch, M. C. 1996. War and strong states, Peace and Weak states? *International Organization* 50(2): 237-268.
- Economist Intelligence Unit (EIU). 2009a. *Manning the Barricades: Who's at risk as deepening economic distress foments social unrest*. London: EIU.
- Economist Intelligence Unit (EIU). 2009b. *Social Unrest*. http://viewswire.eiu.com/site_info.asp?info_name=social_unrest_table.
- Economist Intelligence Unit (EIU). 2009c. *Political Instability Index: Vulnerability to social and political unrest*. http://viewswire.eiu.com/index.asp?layout=VWArticleVW3&article_id=874361472.ent&task=view&id=102&Itemid=891.
- European Report on Development (ERD). 2009. *Overcoming Fragility in Africa*. San Domenico di Fiesole: European University Institute.
- Evans, P. B., D. Rueschemeyer and T. Skocpol. 1985. *Bringing the state back in*. Cambridge: Cambridge University Press.
- Fabra Mata, J. and S. Ziaja. 2009. *Users' Guide on Measuring Fragility*. Oslo: United Nations Development Programme.
- Foreign Policy and Fund for Peace. 2008. The Failed States Index 2008. *Foreign Policy* 167: 64-68.
- Fukuyama, F. 2004. *State-Building*. Ithaca, New York: Cornell University Press.
- Fund for Peace. 2009a. *Failed States Index Scores 2008*. http://www.fundforpeace.org/web/index.php?option=com_content&task=view&id=292&Itemid=452.
- Fund for Peace. 2009b. *Failed States Index FAQ*. http://www.fundforpeace.org/web/index.php?option=com_cont
- Goertz, G. and H. Starr, eds. 2006. *Necessary Conditions: Theory, Methodology, and Applications*. Lanham: Rowan & Littlefield Publishers, Inc.
- Goertz, G. 2006. *Social Science Concepts. A User's Guide*. Princeton: Princeton University Press.

- Goertz, G. 2008. Concepts, Theories and Numbers. In: Robert E. Goodin, Janet M. Box-Steffensmeier, Henry E. Brady and D. Collier (eds.). *The Oxford Handbook of Political Methodology*, Oxford and New York: Oxford University Press. 97-119.
- Goldstone, J. A., R. H. Bates, D. L. Epstein, T. R. Gurr, M. B. Lustik, M. G. Marshall, J. Ulfelder and M. Woodward. 2010. A Global Model for Forecasting Political Instability. *American Journal of Political Science* 54(1): 190-208.
- Gros, J. G. 1996. Towards a taxonomy of failed states in the New World Order: decaying Somalia, Liberia, Rwanda and Haiti. *Third World Quarterly* 17(3): 455-472.
- Gutiérrez Sanín, F. 2009. The quandaries of coding and ranking: Evaluating poor state performance indexes. London: London School of Economics (Crisis States Research Centre Working Paper 58).
- Hartgen, K. and S. Klasen. 2010. *Fragility and MDG Progress: How Useful is the Fragility Concept?* San Domenico di Fiesole: European University Institute (Robert Schuman Centre for Advanced Studies Working Paper 2010/20).
- Helman, G. B. and S. R. Ratner. 1992. Saving Failed States. *Foreign Policy* Winter 1992-1993 (89): 3-20.
- Herrera, Y. M. and D. Kapur. 2007. Improving Data Quality. *Political Analysis Advance Access* 15(4): 365-386.
- Hewitt, J. J., J. Wilkenfeld and T. R. Gurr, eds. 2009. *Peace and Conflict 2010*. College Park: Center for International Development and Conflict Management, University of Maryland.
- Hewitt, J. J. 2007a. *Peace and Conflict Instability Ledger: Description of Data and Model Estimation*. College Park: Center for International Development and Conflict Management, University of Maryland.
- Hewitt, J. J. 2007b. The Peace and Conflict Instability Ledger: Ranking States on Future Risks. In: J. J. Hewitt, J. Wilkenfeld and T. R. Gurr (eds.). *Peace and Conflict 2008*. College Park: Center for International Development and Conflict Management, University of Maryland. 5-20.
- Holsti, K. J. 1996. *The State, War, and the State of War*. Cambridge: Cambridge University Press.
- Independent Evaluation Group (IEG). 2006. *Engaging with Fragile States: An IEG Review of World Bank Support to Low-Income Countries Under Stress*. Washington D.C.: The World Bank.
- International Monetary Fund (IMF). 2010. Regional Economic Outlook: Sub-Saharan Africa: Back to High Growth? Washington, D.C.: International Monetary Fund.
- Jackman, S. 2008. Measurement. In: R. E. Goodin, J. M. Box-Steffensmeier, H. E. Brady and D. Collier (eds.). *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press. 119-152.
- Jackson, R. H. and C. G. Rosberg. 1982. Why Africa's Weak States Persist: The Empirical and the Juridical in Statehood. *World Politics* 35(1): 1-24.
- Kaufmann, D. and A. Kraay. 2008. Governance Indicators: Where Are We, Where Should We Be Going? *The World Bank Research Observer* 23(1): 1-30.
- Kaufmann, D., A. Kraay and M. Mastruzzi. 2007. *Worldwide Governance Indicators Project: Answering the Critics*. World Bank Policy Research Working Paper 4149. Washington, D.C.: World Bank.
- Kaufmann, D., A. Kraay and M. Mastruzzi. 2008. *Governance Matters VII: Aggregate and Individual Governance Indicators 1996-2007*. World Bank Policy Research Working Paper No. 4654. Washington D.C.: The World Bank.
- Kaufmann, D., A. Kraay, and M. Mastruzzi. 2009. *Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996-2008*. World Bank Policy Research Working Paper No. 4978. Washington D.C.: The World Bank.
- Krempel, L. 2005. *Visualisierung komplexer Strukturen: Grundlagen der Darstellung mehrdimensionaler Netzwerke*. Frankfurt a. M.: Campus.
- Kugler, J. and W. Domke. 1986. Comparing the Strength of Nations. *Comparative Political Studies* 19(1): 39-69.
- Kurtz, M. J. and A. Schrank. 2007. Growth and Governance: Models, Measures, and Mechanisms. *The Journal of Politics* 69(2): 538-554.
- Langbein, L. and S. Knack. 2010. The Worldwide Governance Indicators: Six, one, or none? *Journal of Development Studies* 46(2): 350-370.
- Marshall, M. G. 2008. *Fragility, Instability, and the Failure of States*. New York: Council on Foreign Relations.
- Marshall, M. G. and B. R. Cole. 2008. Global Report on Conflict, Governance and State Fragility 2008. *Foreign Policy Bulletin* 18: 3-21.
- Marshall, M. G. and J. A. Goldstone. 2007. Global Report on Conflict, Governance and State Fragility 2007. *Foreign Policy Bulletin* 17: 3-21.
- Migdal, J. S. 1988. *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton: Princeton University Press.
- Müller, T. and S. Pickel. 2007. Wie lässt sich Demokratie am besten messen? Zur Konzeptualität von Demokratie-Indizes. *Politische Vierteljahresschrift* 48(3): 511-539.
- Munck, G. L. 2009. *Measuring Democracy*. Baltimore: The John Hopkins University Press.
- Munck, G. L. and J. Verkuilen. 2002. Conceptualizing and Measuring Democracy: Evaluating Alternative Indices. *Comparative Political Studies* 35(5): 5-34.

- Munck, G. L. and J. Verkuilen. 2009. Conceptualizing and Measuring Democracy. An Evaluation of Alternative Indices. In: Gerardo L. Munck (ed.). *Measuring Democracy*. Baltimore: The John Hopkins University Press. 13-37.
- Nyheim, D. 2009. *Preventing Violence, War, and State Collapse: The Future of Conflict Early Warning and Response*. Paris: Organisation for Economic Co-operation and Development.
- Organisation for Economic Co-operation and Development (OECD). 2008. *Concepts and Dilemmas of State Building in Fragile Situations*. Paris.
- Rice, S. E. and S. Patrick. 2008. *Index of State Weakness in the Developing World*. Washington D.C.: Brookings Institution.
- Rotberg, R. I. 2004. The Failure and Collapse of Nation-States. In: Robert I. Rotberg (ed.). *When States Fail: Causes and Consequences*. Princeton: Princeton University Press. 1-50.
- Rotberg, R. I. and R. M. Gisselquist. 2008. *Strengthening African Governance: Index of African Governance, Results and Rankings*. Cambridge: Kennedy School of Government's Program on Intrastate Conflict and Conflict Resolution.
- Thomas, M. A. 2007. What do the Worldwide Governance Indicators Measure? s.l.: The Paul H. Nitze School of Advanced International Studies; The Johns Hopkins University.
- Tilly, C. 1985. War making and state making as organized crime. In: P. B. Evans, D. Rueschemeyer, T. Skocpol (eds.). *Bringing the state back in*. Cambridge: Cambridge University Press. 169–187.
- Treier, S. and S. Jackman. 2008. Democracy as a Latent Variable. *American Journal of Political Science* 52(1): 201-217.
- United States Agency for International Development (USAID). 2006. *Measuring Fragility Indicators and Methods for Rating State Performance*. Silver Spring: USAID.
- Wennmann, A. 2010. *Grasping the Strengths of Fragile States: Aid Effectiveness Between "Top-down" and "Bottom-up" Statebuilding*. CCDP Working Paper 6. Geneva: The Centre on Conflict, Development and Peacebuilding.
- The White House. 2002. *The National Security Strategy of the United States of America*. Washington D.C.
- World Bank. 2009a. *IDA Resource Allocation Index (IRAI) - 2008*. <http://go.worldbank.org/S2THWI1X60>.
- World Bank. 2009b. *Governance Matters 2009: Data and Research Papers*. <http://info.worldbank.org/governance/wgi/resources.htm>.
- World Bank. 2009c. *Governance Matters 2009: Frequently Asked Questions*. <http://info.worldbank.org/governance/wgi/faq.htm>.
- World Bank. 2010. *Fragile and Conflict-Affected Countries: Definitions of Fragility and Conflict*. <http://go.worldbank.org/6B4932MAV0>.
- Wulf, H. and T. Debiel. 2009. *Conflict Early Warning and Response Mechanisms: Tools for Enhancing the Effectiveness of Regional Organisations? A comparative Study of the AU, ECOWAS, IGAD, ASEAN/ARF and PIF*. Crisis States Working Papers Series 49 (2).
- You, D., T. Wardlaw, P. Salama and G. Jones. 2010. Levels and trends in under-5 mortality, 1990–2008. *The Lancet* 375(9709): 100-103.
- Ziaja, S. and J. Fabra Mata. 2010. *State Fragility Indices: Potentials, Messages and Limitations*. Briefing Paper 10/2010. Bonn: Deutsches Institut für Entwicklungspolitik.

Table 1: Index concepts

Abbreviation	Index name	Concept measured and its definition
BTI-WS	“BTI Weak Stateness Index” (Bertelsmann Transformation Index) ²⁸	<i>Weak stateness</i> : “[...] successful transformation requires that a state ha[s] functioning administration structures and that it secure[s] its monopoly on the use of force.” (Bertelsmann Stiftung 2007, p. 6)
CIFP-FI	CIFP Fragility Index (Country Indicators for Foreign Policy)	<i>Fragility</i> : “Fragility is a measure of the extent to which the actual institutions, functions, and processes of a state fail to accord with the strong image of a sovereign state, the one reified in both state theory and international law.” (Carment, Prest and Samy 2009, p. 84)
FSI	Failed States Index	<i>State failure</i> : “A state that is failing has several attributes. One of the most common is the loss of physical control of its territory or a monopoly on the legitimate use of force. Other attributes of state failure include the erosion of legitimate authority to make collective decisions, an inability to provide reasonable public services, and the inability to interact with other states as a full member of the international community.” (Fund for Peace 2009b)
ISW	Index of State Weakness in the Developing World	<i>State weakness</i> : “We define weak states as countries that lack the essential capacity and/or will to fulfill four sets of critical government responsibilities: fostering an environment conducive to sustainable and equitable economic growth; establishing and maintaining legitimate, transparent, and accountable political institutions; securing their populations from violent conflict and controlling their territory; and meeting the basic human needs of their population.” (Rice and Patrick 2008, p. 3)
IRAI	IDA Resource Allocation Index / Country Policy and Institutional Assessment ²⁹	<i>Fragile states</i> : “‘Fragile states’ is the term used for countries facing particularly severe development challenges: weak institutional capacity, poor governance, and political instability. Often these countries experience ongoing violence as the residue of past severe conflict.” (World Bank 2010) ³⁰
PCIL-RR	PCIL Risk Ratio (Peace and Conflict Instability Ledger)	<i>State instability</i> : “[E]vents that create significant challenges to the stability of states. These include revolutionary wars, ethnic wars, adverse regime changes, and genocides or politicides.” (Hewitt et al. 2008b, p. 5)
PII	Political Instability Index	<i>Social and political unrest</i> : “We define social and political unrest or upheaval as those events or developments that pose a serious extra-parliamentary or extra-institutional threat to governments or the existing political order.” (EIU 2009, p. 15)
SFI	State Fragility Index	<i>State fragility</i> : “The State Fragility Index and Matrix [...] rates each country according to its level of fragility in both effectiveness and legitimacy across four dimensions: security, governance, economic development, and social development.” (Goldstone and Cole 2008, p. 9)
WGI-PV	WGI Political Stability and Absence of Violence (Worldwide Governance Indicators)	<i>Probability of unconstitutional change of government</i> : “Political Stability and Absence of Violence/Terrorism measures the perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means.” (World Bank 2009c)

²⁸ The index is used in Bertelsmann publications to measure “weak stateness” (Bertelsmann Stiftung 2007, p. 6), but it is not officially promoted as an index.

²⁹ The World Bank defines a list of fragile countries using the IDA Resource Allocation Index. Some publications use the index itself to measure fragility (e.g., Harttgen and Klasen 2010).

³⁰ In the World Bank, fragile states were formerly labelled “low-income countries under stress” (IEG 2006).

Table 2: Fragility indices, producers and coverage

Index name	Producer (type)	Publication / Data source	Data availability / file format	Years covered ³¹	Countries covered (2008)
BTI Weak Stateness Index	Bertelsmann Stiftung (foundation / media group)	Bertelsmann Stiftung 2008 / Bertelsmann Stiftung 2009	Fully disaggregate / xls	2006, 2008, 2010	125
CIFP Fragility Index	Carleton University (university)	Carment, Prest and Samy 2009 / CIFP 2009	Aggregate / html	2007 ³²	192
Failed States Index	Fund for Peace (think tank)	Foreign Policy and Fund for Peace 2008 / Fund for Peace 2009a	Partially disaggregate / html	2005-2010	177
Index of State Weakness	Brookings Institution (think tank)	Rice and Patrick 2008	Partially disaggregate / pdf	2008	141
IDA Resource Allocation Index	World Bank (international organization)	World Bank 2010 / World Bank 2009a	Fully disaggregate / xls	2005-2009 ³³	75
PCIL Risk Ratio	University of Maryland (university)	Hewitt 2007a / Hewitt 2007b	Aggregate / pdf	2008, 2010	160
Political Instability Index	Economist Intelligence Unit (consulting / media group)	EIU 2009a / EIU 2009b	Partially disaggregate / html	2007, 2009/10	165
State Fragility Index	George Mason University (university)	Marshall and Cole 2008	Partially disaggregate / pdf	1995, 2001, 2007-2009	162
WGI Political Stability	World Bank (international organization)	Kaufmann et al. 2009 / World Bank 2009b	Partially disaggregate / xls, csv ³⁴	1996, 1998, 2000-2009	209

³¹ The following analyses refers to the indices and index scores of 2008 editions, except for the CIFP Fragility Index (2007) and the Political Instability Index (arithmetic average of 2007 and 2009/10 scores).

³² Carment, Prest and Samy (2009) use a dataset from 1980-2006, but this data is not available online.

³³ The World Bank has data starting in the mid 1970s on many more countries, but this data is not disclosed.

³⁴ A file in csv-format is available via the new World Bank Data Catalogue (<http://data.worldbank.org/data-catalog>).

Table 3: Attributes, aggregation methods and indicator weights

Index	Attributes and aggregation method	Number of Indicators	Weight per indicator ³⁵
BTI Weak States Index	<i>(Monopoly of Violence + Basic Administration) / 2</i>	2	0.500
CIFP Fragility Index	<i>(Governance [12] + Economics [24] + Security & Crime [10] + Human Development [17] + Demography [10] + Environment [10]) / 6</i>	83	0.007-0.019
Failed States Index	<i>(Mounting Demographic Pressures + Massive Movement of Refugees or Internally Displaced Persons creating Complex Humanitarian Emergencies + Legacy of Vengeance-Seeking Group Grievance or Group Paranoia + Chronic and Sustained Human Flight + Uneven Economic Development along Group Lines + Sharp and/or Severe Economic Decline + Criminalization and/or Delegitimization of the State + Progressive Deterioration of Public Services + Suspension or Arbitrary Application of the Rule of Law and Widespread Violation of Human Rights + Security Apparatus Operates as a "State Within a State" + Rise of Factionalized Elites + Intervention of Other States or External Political Actors) / 12</i>	12 ³⁶	0.083
IDA Resource Allocation Index	<i>(Economic Management [3] + Structural Policies [3] + Policies for Social Inclusion/Equity [5] + Public Sector Management and Institutions [5]) / 4</i>	16	0.050-0.083
Index of State Weakness	<i>(Economic Basket [5] + Political Basket [5] + Security Basket [5] + Social Basket [5]) / 4</i>	20	0.050
PCIL Risk Ratio	No attributes specified. The index uses a logistic regression model to produce country scores. As variables, it uses Regime Consistency, Infant Mortality, Economic Openness, Militarization, Neighborhood War and dummy variables for Autocracy and Partial Democracy.	7	Not determined ³⁷
Political Instability Index	<i>(Underlying Vulnerability [12] + Economic Distress [3]) / 2</i> Three indicators have a weight of two.	15	0.038-0.200
State Fragility Index	<i>Effectiveness score + Legitimacy score</i> <i>= (Security Effectiveness [1] + Political Effectiveness [3] + Economic Effectiveness [2] + Social Effectiveness [1]) + (Security Legitimacy [1] + Political Legitimacy [4] + Economic Legitimacy [1] + Social Legitimacy [1])</i>	14	0.031-0.125
WGI Political Stability Indicator	No attributes specified. The index uses an unobserved components model which weights each indicator according to its correlation with the other indicators.	35 ³⁸	0.010-0.094 ³⁹

Data sources: see Table 1. Refers to the 2008 editions of indices; exceptions: CIFP Fragility Index (2007) and Political Instability Index (2009/10). The figures in brackets show the number of indicators used to measure the each attribute. If the number of indicators used is one for all attributes, the bracket is omitted.

³⁵ The weight per indicator can vary, even if no explicit weighting scheme is applied, when differently sized categories lead to different impacts of individual indicators on the overall index score. All weights calculated by the author unless otherwise stated.

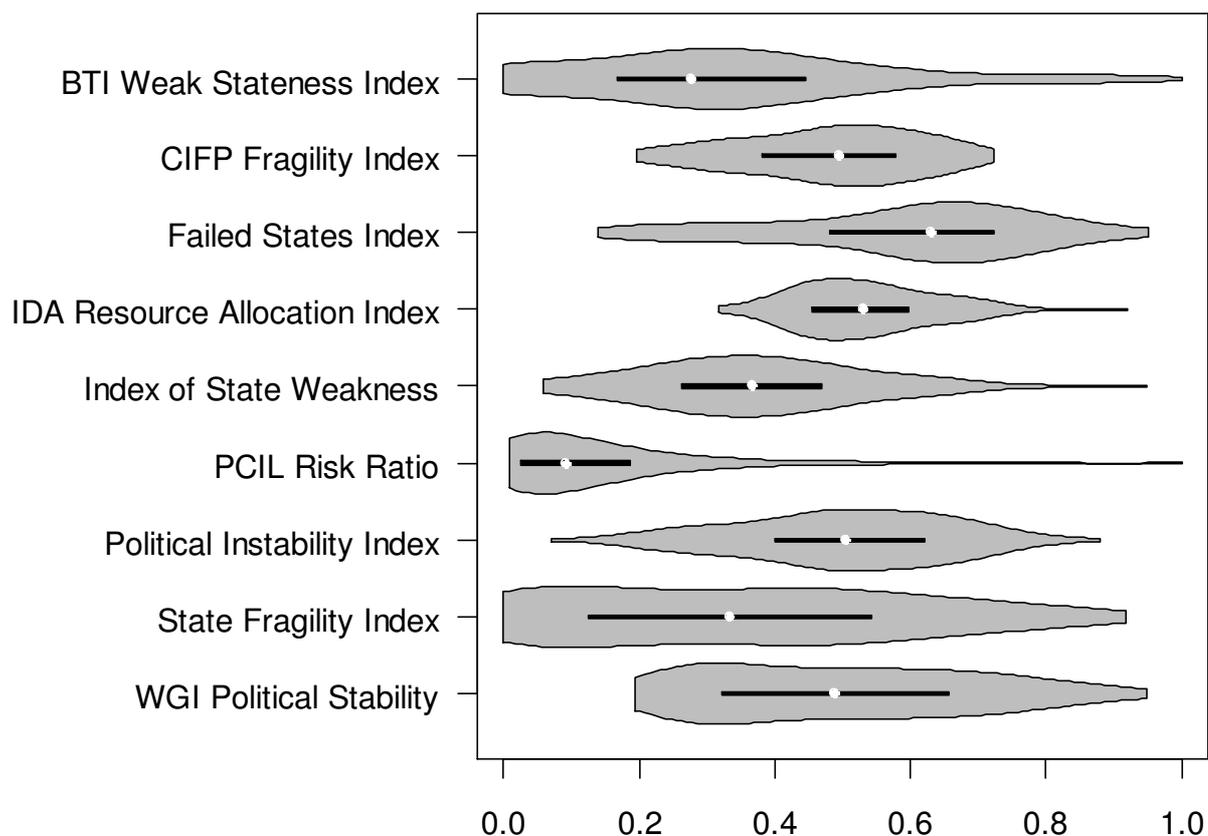
³⁶ The Failed States Index also draws on at least four structural indicators and calibrates scores by expert judgement. The exact design and impact of these components is not published.

³⁷ Since the index does not standardize the indicators, the modelled weights can not be calculated. According to the author, regime consistency has the strongest and militarization the lowest impact (personal communication).

³⁸ The WGI Political Stability index uses 35 indicators from 13 sources in the year 2008. The six Worldwide Governance Indicators jointly use many more indicators from 35 sources.

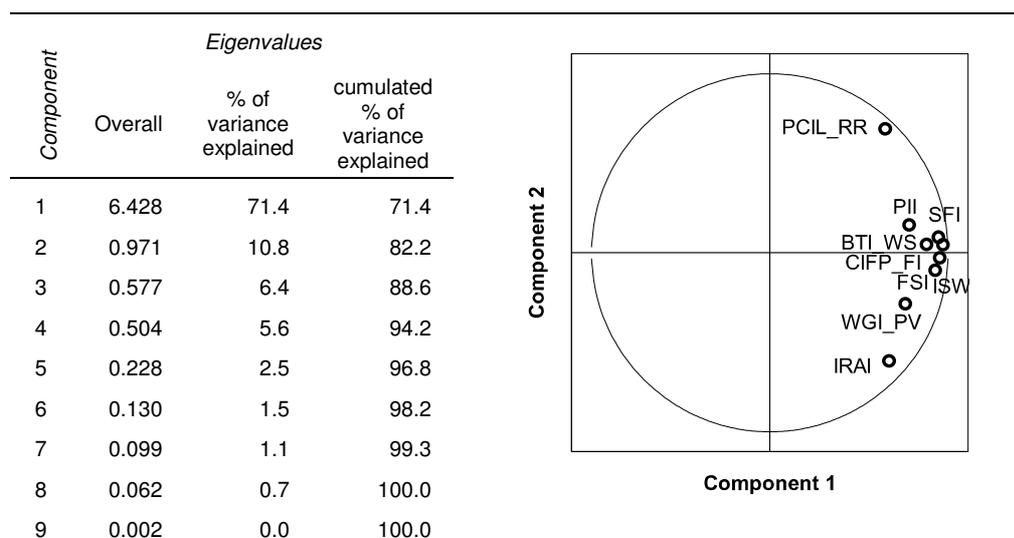
³⁹ These are weights provided by the index authors (World Bank 2009b).

Figure 2: Violin plots of fragility index scores 2008*



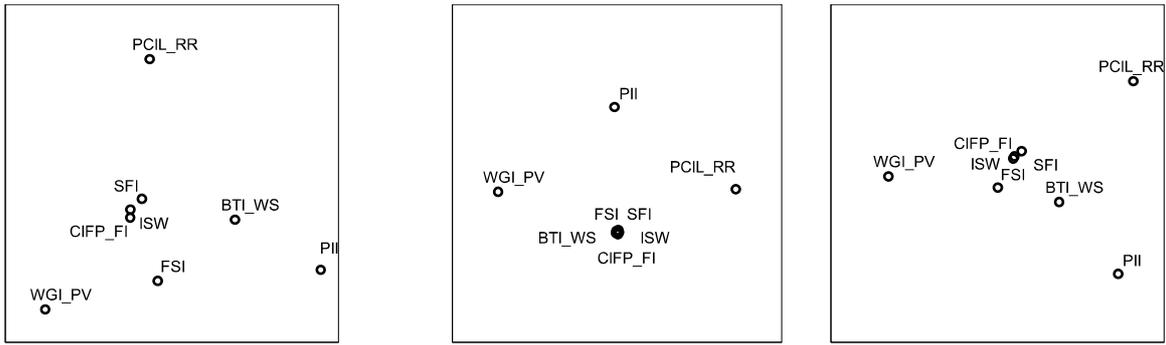
Data sources: see Table 2. *) CIFP Fragility Index: 2007 edition; Political Instability Index: arithmetic average of 2007 and 2009/10 editions. The BTI Weak Stateness Index, the IDA Resource Allocation Index, the Index of State Weakness and the WGI Political Stability index have been inverted to adhere to a „best to worst“ scale.

Figure 3: Principal component analysis of fragility index scores 2008*



Abbreviations: see Table 1. Data sources: see Table 2. Pairwise deletion. *) see remarks under Figure 2.

Figure 4: Multidimensional scaling of fragility index scores 2008*



Scale of dissimilarities:

Ordinal

Ordinal

Interval

Initial configuration:

Torgerson

Random (100 starts)

Torgerson

Normalized raw stress:

0.0016

0.0001

0.0042

Abbreviations: see Table 1. Data sources: see Table 2. $N = 113$, Listwise deletion. *) see remarks under Figure 2. The IDA Resource Allocation Index has been excluded to keep the number of observations sufficiently high.