Applied Economics and Policy Working Paper Series

AEP WP NO. 2021-18

A COMPARATIVE ASSESSMENT OF RESILIENCE MEASUREMENT APPROACHES

Joanna Upton, Elizabeth Tennant, Kathryn J. Florella and Christopher B. Barrett

Abstract

As development and humanitarian agencies increasingly advance the objective of 'building resilience', three resilience measurement methods have come into especially widespread use: the Resilience Indicators for Measurement and Analysis approach developed by FAO, the multi-dimensional index approach developed by TANGO International, and the probabilistic approach of Cissé and Barrett. We compare performance across those three methods using nationally representative panel data from Ethiopia and Niger. We find that the three measures exhibit significantly different distributions and orderings among households, and they vary significantly in the households they identify as resilient or least resilient. All three measures exhibit only modest out-of-sample predictive accuracy, generating many false negatives and false positives relative to the food security outcome measure whose resilience they are meant to reflect. It remains unclear what these measures capture and what value they add beyond more established wellbeing measures such as the food consumption score or real expenditures. Significant room exists for improvement in resilience measurement to better guide and evaluate development resilience interventions.

Charles H. Dyson School of Applied Economics and Management

School of Hotel Administration

Samuel Curtis Johnson Graduate School of Management

Cornell SC Johnson College of Business

Caveat utilitor:
A comparative assessment of resilience measurement approaches

September 2021 revised version

Abstract
As development and humanitarian agencies increasingly advance the objective of 'building resilience', three resilience measurement methods have come into especially widespread use: the Resilience Indicators for Measurement and Analysis approach developed by FAO, the multi-dimensional index approach developed by TANGO International, and the probabilistic approach of Cissé and Barrett. We compare performance across those three methods using nationally representative panel data from Ethiopia and Niger. We find that the three measures exhibit significantly different distributions and orderings among households, and they vary significantly in the households they identify as resilient or least resilient. All three measures exhibit only modest out-of-sample predictive accuracy, generating many false negatives and false positives relative to the food security outcome measure whose resilience they are meant to reflect. It remains unclear what these measures capture and what value they add beyond more established wellbeing measures such as the food consumption score or real expenditures. Significant room exists for improvement in resilience measurement to better guide and evaluate development resilience interventions.

Key words: Ethiopia, factor analysis, food security, Niger, poverty, predictive performance, risk, shocks, targeting

Caveat utilitor: A comparative assessment of resilience measurement approaches

I. Introduction

Poverty and food insecurity remain pressing concerns globally not just because of their distressing prevalence but because they are dynamic and stochastic. Many of the world's poorest communities are also those most threatened by climate, conflict, disease, and other shocks. Increasing attention to the roles that risk and shocks[1] play in the persistence of poverty and food insecurity underpins the rapid growth in popularity of the concept of development resilience. The resilience concept shifts focus beyond the need to raise (time-varying and stochastic) levels of wellbeing, towards shock-and-stress-proofing wellbeing over time, especially among poor populations vulnerable to a wide array of risks and shocks. Scholarship and related policy documents have therefore defined resilience for development as the ability to achieve and maintain an acceptable standard of wellbeing even in the face of shocks and stressors (Barrett and Constas 2014), or as "the capacity that ensures adverse stressors and shocks do not have long-lasting adverse development consequences" (Constas et al. 2014).

Development and humanitarian agencies increasingly emphasize 'resilience' in an effort to integrate traditionally siloed domains of humanitarian assistance, governance, food security, and economic growth, attempting to bridge from short-run emergency response to longer-term development strategy. As Constas et al. (2014, p.4) put it, "In a world where conventional approaches to dealing with humanitarian aid and development assistance have been questioned, resilience has captured the attention of many audiences because it provides a new perspective on how to effectively plan for and analyze the effects of shocks and stressors that threaten the well-being of vulnerable populations."

Employment of the concept requires measurement, however. Measurement is always a challenge for latent variables, like resilience, or poverty, or food security. Because latent variables are not directly observable, they must be inferred through the collection and analysis of directly observable variables believed to capture a specific conceptualization of the latent variable. Careful evaluation of the comparative performance of different candidate measures of latent variables is an important step in methods development to ensure the responsible design, evaluation and targeting of interventions. That has been the experience with the measurement of other important, latent variables such as food security or poverty (Lipton and Ravallion 1995; Webb et al. 2006; Barrett 2010; Maxwell et al. 2014, Vaitla et al. 2017, Ravallion 2019). Although multiple resilience measures have emerged and become widely employed in recent years, detailed comparative evaluation has been lacking. This paper helps to fill that important void.

Indeed, "both the definition of resilience and the methodologies used to measure it are heavily contested" (Jones and Tanner, 2017, p. 229), a point echoed by Watts (2016). A recent scoping review identifies three distinct conceptualizations of resilience – and a variety of specific definitions within each conceptualization – each underpinning different measurement approaches (Barrett et al. 2021). Of these, two conceptualizations currently dominate, resulting in distinct measurement methods that have diffused into the broader literature and been adopted as standard practice by specific donors,

[1] We use the term 'risk' to reflect exposure ex ante to possible adverse events, as distinct from a 'shock', which reflects an ex post realization of an adverse event.

governments, or communities of practice. Each method is quantitative and implementable using longitudinal (i.e., panel) household survey data. But since they are rooted in different conceptualizations of resilience, one might reasonably conjecture that they do not exhibit similar patterns when applied to the same population. Given how much current development and humanitarian programming follows from the widespread aspiration to build resilience to shocks and stressors, it seems time to assess how well these different measures perform.

Perhaps the dominant conceptualization among operational agencies treats resilience as the "capacity that ensures stressors and shocks do not have long-lasting adverse development consequences" (Constas et al. 2014, p.4). Two rival methods have emerged for operationalizing the resilience-as-capacity concept. The first, and to date most-widely used, quantitative measure of resilience is Resilience Indicators for Measurement and Analysis (RIMA), developed by the Food and Agriculture Organization of the United Nations (FAO). The method was substantially updated and re-released as RIMA-II (FAO 2016; hereafter simply "RIMA"). As we describe in greater detail below, RIMA uses factor analysis to estimate four latent variables, labelled "pillars" - Access to Basic Services (ABS), Assets (AS), Social Safety Nets (SSN), and Adaptive Capacity (AC) – from standard household and community survey data. In a second stage of estimation, the estimated pillars are combined into an overall resilience capacity index (RCI), scaled from a minimum of 0 to a maximum of 100, which is then associated statistically with food security indicators, such as food consumption score (FCS) or dietary diversity, using structural equation modeling. This approach has been implemented in some form by FAO researchers for a wide range of analyses (e.g., Alinovi et al. 2008, 2010; d'Errico et al. 2017, 2018a, 2018b), and is now the recommended tool within United Nations affiliate organizations (e.g., FAO, WFP, UNICEF, and IFAD), and under the Comprehensive Africa Agricultural Development Programme (CAADP), among others.

RCI is thus an explanatory variable, conceptualized as intermediating between risk and shocks, on the one hand, and development outcomes like food security or poverty, on the other. RIMA is typically not used to classify households as 'resilient' or 'not resilient', i.e., for binary classification of households as poor/non-poor or food secure/insecure. Of course, that poses a substantial challenge for evaluating performance in 'building resilience', a widely-advanced objective that necessarily renders resilience a dependent variable. On the rare occasions when they do use a resilient/not resilient outcome classification, FAO recommends classifying households based on their observed "realized resilience", as reflected in the simple retrospective period-on-period change in a wellbeing measure in longitudinal data – i.e., realized resilience implies a non-negative change in the outcome indicator of interest, like FCS (FAO 2016).

A second, but closely related method developed by TANGO International likewise constructs a RCI based on factor analysis of a wide range of indicators, as detailed below, to estimate three latent variables: absorptive, adaptive, and transformative capacities (Smith and Frankenberger 2018). Like FAO, the TANGO method yields an RCI that is an explanatory variable, not an outcome itself. When necessary to classify households as resilient or not, TANGO also uses the same "realized resilience" approach as FAO – i.e., a household exhibits realized resilience by a non-negative change in the wellbeing indicator of interest between successive rounds in longitudinal data. TANGO analyses commonly use regression analysis to associate estimated resilience capacities with observed change in wellbeing over time. Since 2018, the TANGO approach serves as the basis for recommended resilience analysis under projects funded by the United States Agency for International Development, many of which require data collection expressly to construct these measures (USAID, Henly-Shapard and Sagara, 2018).

The third measure now in widespread use derives from a different conceptualization of "resilience as a normative condition" (Barrett et al. 2021), anchoring the concept and measure in normative well-being standards, just as poverty assessments employ a poverty line as a reference point and food security measures likewise use some normative threshold. This follows from the principle that resilience "should be indexed to a given development outcome (e.g., food security, poverty, health) with a normative threshold" (Constas et al. 2014, p. 7). Cissé and Barrett (2018, hereafter C&B) introduced a measure that derives directly from this alternative conceptualization, which follows from the theoretical framework of Barrett and Constas (2014). Thus the C&B measure is an outcome measure anchored to some normative standard of wellbeing, a lefthand side variable like the realized resilience measures, in contrast to the RIMA and TANGO RCIs that are explanatory variables not tied to any specific normative standard.

The C&B measure therefore employs different statistical methods than the RIMA and TANGO measures. In particular, C&B uses ordinary least squares regression to estimate the mean and variance of a household wellbeing variable conditional on its characteristics and experiences – potentially including shocks or risk exposure – and from that derives the probability of a household reaching or exceeding a pre-specified, normative standard – e.g., a poverty line or FCS level. That conditional probability is the household's resilience score (RS), which necessarily falls in the [0,1] interval. When one multiplies the RS by 100, the C&B measure therefore uses the same cardinal scale as the RIMA and TANGO RCIs, although the RS reflects a probability while the RCIs are unitless index measures. The C&B method has now been adopted in a range of academic studies in several contexts (Cissé and Ikegami 2016, Upton et al. 2016, Alloush 2019, Knippenberg et al. 2019, Phadera et al. 2019, Premand and Stoeffler 2020, Vaitla et al. 2020).

To date resilience measures have been used largely to build a narrative to motivate development or humanitarian interventions, much less for targeting or impact evaluation (Barrett et al. 2021). But targeting and impact evaluation applications have become increasingly widespread as these measures have diffused into broader use (e.g., Smith et al. 2015, Cissé and Ikegami 2016, Knippenberg and Hoddinott 2017, Smith and Frankeberger 2018, Phadera et al. 2019, Premand and Stoeffler 2020). With program targeting and evaluation increasingly resting on these measures, it becomes especially important to explore whether the measure one chooses matters to how one assesses a (sub-) population's resilience or the impacts of an intervention on their resilience, or to who one might target for intervention.

At a more fundamental level, one might also ask what value resilience measures add beyond that provided by conventional poverty or food security measures and how closely the two correspond. The resilience community argues that two major improvements come from the new conceptualization(s) and associated measures.

The first is that by offering probabilistic assessments, allowing for time paths of recovery, and by expressly identifying capacities (i.e., targetable characteristics) that significantly dampen shocks' impacts, resilience measures address both ex ante and ex post exposure to stressors and shocks, thereby integrating the tools of conventional ex post poverty or food security measurement with those of the (much smaller) vulnerability measurement literature (e.g., Ligon and Schechter 2003; Kamanou and Morduch 2004; Grimm et al. 2016, Dang and Lanjouw 2017).

The second claim is that although conventional poverty and food security measures often work reasonably well to target the chronically poor or food insecure, they perform far worse in identifying transitory deprivation, and thus are least useful in responding to major shocks, i.e., when the humanitarian community most needs them. This puts a premium on developing measures expressly aimed at the sensitivity of wellbeing to shocks and those interventions or capacities that insulate distinct subpopulations from especially adverse effects of disasters.

If these resilience measures add value, however, there must be some measure of correspondence with the normative poverty or food security measures to which resilience is ostensibly anchored. It remains an open question how well these measures correlate with the underlying wellbeing measures that are meant to be their focus, as well as how closely they correspond with one another. There has been little empirical validation of these methods, especially comparatively or with respect to their performance predicting out-of-sample the wellbeing outcomes that anchor the resilience concept.[2] Do these measures process raw data in fundamentally similar ways such that they are essentially substitutes for one another? If so, then methods differences are more a matter of taste than of consequence. Given that the concept is rooted in the stochastic dynamics of more familiar individual, household, and community wellbeing representing poverty and food insecurity, do they correspond reasonably well with more widely used poverty and food security measures? Perhaps most importantly, if resilience reflects "the capacity that ensures adverse stressors and shocks do not have long-lasting adverse development consequences" (Constas et al. 2014), as the prevailing conceptualization among donor and operational agencies holds, then we should care whether these measures exhibit skill in predicting development outcomes out-of-sample.

We fill this gap with statistical comparison of the three prevailing resilience measurement approaches – TANGO, RIMA, and C&B – using nationally representative panel data from Niger and Ethiopia, two countries where resilience is a major policy topic. We compare only these three methods both because they have the greatest traction to date in the literature and in policy circles, and because they can be implemented using exactly the same data, ensuring that any differences in findings arise solely due to differences in the methods used. Other resilience measurement methods currently in use require qualitative data and/or quantitative data at higher frequency (Knippenberg et al. 2019) or using countries, rather than households, as units of observation (Smerlak and Vaitla 2017). We implement each of the three methods precisely following its developers' directions as to how to build household-and-period-specific RCI or RS measures and in classifying households' resilience by each method.[3] We then compare the three resulting resilience measures using several methods, as we describe below.

[2] Knippenberg et al. (2019) compares inferences under C&B and the new MIRA method they develop (discussed further below), but not with either RIMA or TANGO's RCIs, nor do they use any of the comparative performance measures reported here. Jones and d'Errico (2019) compare RIMA with a new Subjective self-Evaluated Resilience Score using variants of the correlation and distribution tests we report here. As we finished this manuscript, we discovered that Alloush (2019) uses one of the tests we report, comparing the performance of C&B and TANGO in predicting future wellbeing indicators – such as poverty, household income per capita, food expenditure, and wealth – in panel data from South Africa, finding results qualitatively very similar to ours. After our manuscript first circulated, d'Errico and Smith (2019) produced a comparison of the RIMA and TANGO approaches to one another in terms of their distributions and relative ranking of different groups' resilience capacity.

[3] To ensure fidelity to the methods developers' best practices, we had extensive discussions and email exchanges with the RIMA and TANGO teams during the drafting of this paper, shared the initial draft of this paper with them and broader user communities for comments, and revised our analyses and draft to account for their helpful corrections and cricitisms. We are grateful for their assistance in clarifying finer points of their approaches.

Our results reveal that each of the three currently-dominant resilience measures disappoints. They are inconsistent with one another. The distribution of measures within sample varies significantly across measures. Population-level descriptions of resilience differ by method even when using the same data. The rank correlation of households' estimated RCI or RS among approaches is modest, and they do not classify the same households as falling near the bottom of the distribution. So any household-level targeting that one might do would differ significantly based purely on the method employed. Because the TANGO and RIMA methods divorce their realized resilience classification of households from their RCI estimation procedure, correspondence between these measures is not assured. Indeed, each RCI measure fails a basic test of internal consistency; households classified as "realized resilient" do not have significantly higher RCI than those classified as "not resilient" by either the RIMA or TANGO method, even though RCI is defined as capacities that confer resilience, thus should correspond with realizations. None of the measures offers obvious value addition in predicting wellbeing out of sample, relative to simply using more familiar and easier-to-compute wellbeing measures.

The key takeaway is that there remains considerable room for improvement on all fronts, perhaps especially in the two index-based RCI methods (RIMA and TANGO) on which the donor and practitioner communities depend most heavily right now, as they are systematically less related to observed wellbeing outcomes. Much like the challenges of food insecurity measurement (Vaitla et al. 2017), even though these three main resilience measures all attempt to reflect the same latent variable, they clearly reflect different underlying concepts and correspond differently with established wellbeing indicators, but in ways that analysts rarely acknowledge. So *caveat utilitor*: user beware!

II. Resilience Measurement Approaches

We do not replicate and compare all published resilience measurement methods, in large part because diverse data requirements preclude many direct comparisons. Some methods rely on qualitative (or mixed methods) and a highly context- or community-specific process, including BRACED (DFID 2014), CoBRA (UNDP 2013), the Characteristics of a Disaster-Resilient Community approach (Twigg 2009), or the USAID Measurement Framework for Community Resilience (USAID 2013), among others. Béné et al. (2017), Ansah et al. (2019), and Barrett et al. (2021) summarize a range of approaches, each making the important point that there is often a tension between the nuanced reality of resilience on the ground and the desire for rigorous, quantitative measures. A few other quantitative methods use quite different sorts of household or individual data, and/or take a different approach that is not directly comparable. For example, Smerlak and Vaitla (2017) develop a cross-country time series approach focusing on the recovery path of food security indicators; but the method is infeasible without long time series thus far unavailable at household or individual level. Knippenberg and Hoddinott (2017) use similar data, but estimate a recovery path, building on cross-sectional work by Vollenweider (2015). Knippenberg et al. (2019) use high frequency (monthly) data on subjective shock assessment in Malawi to investigate shock persistence and its correlates using the Measurement Indicators for Resilience Analysis protocol. These methods have appealing features but do not lend themselves to comparison with the three core methods in a data set we could find. So we omit them from the exercise we describe below.

a. The Resilience Indicators for Measurement and Analysis – II (RIMA) method

FAO developed the RIMA approach as the concept of resilience gained traction in development circles in the late 2000s. Development and humanitarian agencies were increasingly pressed to design programs to improve resilience, to monitor and evaluate these investments, and to target interventions to the least resilient. RIMA was the first widely used method to provide a measurement approach meant to inform such programming. RIMA includes a theoretical and empirical framework with the main objectives of measuring the key determinants of a household's resilience capacity so as to guide policy and evaluate interventions. As the recommended approach for UN and other organizations, RIMA has been implemented for several impact assessments and other analyses. Alinovi et al. (2008, 2010) first examined the approach using data on households in Palestine and Kenya. An FAO study (2015) applies RIMA to resilience and targeting in Niger. Garbero and Chichaibelu (2019) use it to examine impacts of a small-scale irrigation project in Ethiopia. D'Errico et al. (2018a) describe resilience in Tanzania and Uganda. Brück et al. (2019) examine the impacts of conflict on resilience capacity in Gaza. The FAO has issued practitioner guidance documents on RIMA, the 2016 updates of which serve as the basis for our replication (FAO 2016). CAADP advises African states and partners to use RIMA in resilience analysis.

The four pillars that make up the RIMA resilience capacity index (RCI) are estimated from factor analysis using the principal factor or sufficient factors to explain ≥96 percent of sample covariance. Estimation of the overall RCI uses a structural equation model, specifically the Multiple Causes Multiple Indicators (MIMIC) model (Jöreskog and Goldberger 1974). This two-stage technique estimates resilience capacity as a latent variable, mediating an effect between the four pillars – Access to Basic Services (ABS), Adaptive Capacity (AC), Assets (AST) and Social Safety Nets (SSN) – and wellbeing outcomes, *W*, such as FCS. Mathematically, the model uses the following two equations,

$$RCI = [\beta_1, \beta_2, \dots, \beta_n] * [ABS, AST, SSN, AC] + [\varepsilon_1] \qquad (1)$$

where the pillars (*ABS, AST, SSN*, and *AC*) reflect the underlying latent variable, RCI, through which they have a joint effect on wellbeing outcomes, specified as,

$$[W_1, W_2, \dots, W_n] = [\alpha_1, \alpha_2, \dots, \alpha_n] * RCI + [\varepsilon_2, \varepsilon_3, \dots, \varepsilon_n] \qquad (2)$$

where ε_n are error terms. Informed by previous research on resilience, vulnerability and food security, RIMA also proposes the set of variables that comprise each pillar.

RCI reflects a household's resilience capacities. It is not used to identify households as "resilient" or not. For this resilience classification task, the FAO (and TANGO) approach is to use 'realized resilience', *R,* which is simply the difference between the wellbeing outcome after a shock, or at the end of a period, relative to the prior period:

$$R = W_t - W_{t-1} \qquad (3)$$

We construct $\tilde{R}$ on the same [0,100] scale as C&B's RS measure by normalization, $\tilde{R} = \left[R - R^{min}\right] * 100/\left[R^{max} - R^{min}\right]$ where R^{max} and R^{min} are the maximum and minimum values, respectively, in the *R* series. Alternatively, they use the binary indicator version, *I*, that takes value one if $R \geq 0$, and zero otherwise.

Based on the definition of resilient units as the ones that do not suffer a loss in wellbeing in the face of shocks, RIMA estimates a Probit model of the association between suffering a loss in wellbeing (I) and the variables previously used to generate the resilience pillars. As these analyses do not compare directly across approaches, we set this aside, and focus just on the RCI, and I and R/$\widetilde{\mathrm{R}}$ resilience indicators.

It is important to note that measures of shock exposure are not directly included in the RIMA model. RIMA uses the measures of shock exposure just to determine the households that were affected by a shock and suffered a loss in wellbeing when defining *I* and *R*. Consequently, the impact of shocks on wellbeing is not directly estimated, and the relationship between shocks, resilience capacity and wellbeing is not formalized in an integrated model. D'Errico et al. (2018a) nonetheless estimate a Probit model with suffering a loss in wellbeing as the dependent variable, and include both RCI and measures of shock exposure as explanatory variables. The buffering capacity of RCI in the face of shocks is therefore directly tested in this revised application of RIMA.

b. The TANGO method

TANGO International developed a significantly modified variation on RIMA, first with a program evaluation in Ethiopia (Smith et al., 2015), and subsequently in Bangladesh (Smith and Frankeberger, 2018). The TANGO method focuses on household and community level capacities – "absorptive," "adaptive," and "transformative" – that are hypothesized to promote resilience. They define these following Constas et al. (2014) as "the capacity that ensures adverse stressors and shocks do not have long-lasting adverse development consequences." Since 2018, the TANGO approach has served as the basis for recommended resilience analysis under USAID funded projects (Henly-Shapard and Sagara, 2018), and the associated TANGO resilience survey module is now being systematically collected to evaluate resilience investments by USAID.

The TANGO method bears important similarities to RIMA. Both use factor analysis to construct a resilience capacity index (RCI) of latent variables – four pillars in the RIMA case, three capacities in TANGO's – estimated using data series that prior studies and/or common practice in development suggest are associated with these concepts. Again like RIMA, TANGO does not use RCI to classify households as resilient or not. Instead, they use the same 'realized resilience' measure, *R* (re-scaled ($\widetilde{\mathrm{R}}$), or *I* in its binary transform) as in equation 3. Two key differences between RIMA and TANGO are (i) in how one incorporates variables into the four pillars versus three capacities, and the resulting modified factor analysis methods employed, and (ii) the link to a wellbeing outcome in RIMA's MIMIC modeling approach.

Each of the primary three capacities – absorptive, adaptive, and transformative – consist of several sub-components, some of which are themselves a composite of several indicators. For example, absorptive capacity is built from indicators that reflect bonding social capital, asset ownership, savings, informal safety nets, and disaster preparedness and mitigation.[4] With the exception of a few singular indicators and/or additive indices from household or community survey data, these sub-components are likewise constructed using factor analysis as a data reduction technique on a vector of household- or community-level indicators. These are in turn combined using factor analysis into first one of the three capacities and then in yet another (nested) factor analysis into the overall RCI.

Using panel (baseline and end line) data on households, with baseline and end line observations, Smith and Frankenberger (2018) specify the following procedure for each of the factor analyses:[5]

- Factor loadings are calculated using baseline values of the component variables;

[4] Table 2 describes the RCI sub-components. All the underlying indicators are described in the Appendix.
[5] We thank Lisa Smith for graciously providing helpful methodological guidance on implementation of the method.

- Variables with estimated factor loadings that have the "wrong" (i.e., unexpected) sign are dropped, as are those with a Kaiser-Meyer-Olkin (KMO) statistic of less than 0.5, and the factor analysis repeated until no such variables remain;
- End line variables are standardized using the baseline mean and standard deviation (rather than the full panel data);
- Factor loadings are then used in linear combination with the standardized end line variables, to sum up to a single index.

A more standard approach to factor analysis[6] would simplify these steps in using pooled data (i.e., all the included survey rounds) to calculate the factor, then creating the singular index using variables within sample, rather than those standardized using out-of-period means and standard errors, and not excluding variables whose factor loading have unexpected signs. It is likewise unusual practice to include the same variables in multiple capacity indices. We nonetheless endeavor to follow the TANGO procedure as closely as possible in generating TANGO RCI measures this analysis.[7] The component indices are then included in the same fashion in higher-level indices, so as to produce the overall RCI.

Finally, similar to RIMA, TANGO estimates the relationship between RCI and wellbeing through two different regression models, one cross-sectional and another using panel data. Smith and Frankenberger (2018) explore the association between RCI and the intertemporal change in the wellbeing indicator of interest for household *i*, both directly and in mediating the partial correlation of observed change in wellbeing with a shock that hits between period t-1 and period t, S_t:

$$Y_{it} - Y_{it-1} = \alpha + \beta_1 S_{it} + \beta_2 RCI_{it-1} + \gamma S_{it} * RCI_{it-1} + \beta_3 X_{it} + \mu_i + \varepsilon_{it} \quad (4)$$

where X is a vector of household characteristics for household *i* at time *t*, μ represents a household fixed effect that controls for time invariant household unobservables, and ε is the regression error term. The dependent variable is merely the continuous change in wellbeing measure that gets discretized into the resilience classification variable, *R*, per equation 3. We set this analytical extension aside and focus on the RCI, and the I and $\mathrm{R}/\widetilde{\mathrm{R}}$ resilience classification indicators, as there is not a reliable way to benchmark different measures' results in estimating equation 4.

c. The Cissé and Barrett method

The Cissé and Barrett (2018) method was developed to directly implement the Barrett and Constas (2014) theory of development resilience, which defines resilience in terms of having an acceptably high likelihood of remaining above the poverty line (or other meaningful wellbeing threshold) even in the face of shocks and stressors. The approach has now been implemented by academic researchers in a variety of contexts (Upton et al. 2016, Cissé and Barrett 2018, Alloush 2019, Knippenberg et al. 2019, Vaitla et al. 2020) and is increasingly used for impact evaluation (Cissé and Ikegami 2016,

---

[6] As reflected by standard factor analysis guidelines, and what one obtains for example when relying uniquely on pre-packaged commands as available in Stata, R, or other data analysis programs.

[7] In some cases, we lack observations of certain variables in both periods, and hence cannot apply out-of-period means and standard errors or factor loadings (in which case we implement the 'pooled' or single-period approach). As a robustness check, we construct the full index using the simplified factor analysis approach as described in the text. We find that while some of the sub-indices differ in nature, and may lead to different results in other analyses, the results of our comparisons across indices are not substantively affected. We proceed as such with the TANGO method; but include the 'simplified' factor loadings in the Appendix and can provide full results of that approach on request.

Phadera et al. 2019, Premand and Stoeffler 2020), with some unpublished efforts we are aware of to employ the method for targeting project beneficiaries.

The C&B method uses standard ordinary least squares (OLS) regression approach to estimate the household-specific conditional mean of wellbeing, the residuals from which can then be used to estimate the household-specific conditional variance. Combining those two estimated conditional moment functions, and assuming a two-parameter distribution (such as beta, exponential, gamma, normal, student-t, etc.) one can thereby estimate the conditional probability of satisfying some normative wellbeing standard. The first step is an OLS regression to estimate the conditional mean of the wellbeing indicator:

$$W_{i,t} = \sum_k \beta_k W_{i,t-1}^k + \gamma_1 X_{i,t} + \Omega S_{i,t} + \varepsilon_{i,t} \tag{5}$$

In equation 5, the superscript *k* indicates the polynomial order (e.g., 2 is quadratic, 3 is cubic, etc.) to allow for possible nonlinear dynamics under a first-order Markov process assumption, following the empirical literature on the estimation of poverty dynamics (Barrett et al. 2016). *X* contains a series of time-varying household and community characteristics. *S* represents shock or stressor indicators, including climate and price, and could be interacted with the *X* and/or *W* variables, if so desired (although those interactions are omitted from equation 5 so as to reduce clutter). Noting that under the mean zero residual assumption, $E[\varepsilon_{i,t}^2] = V(W_t | X_t, S_t, W_{t-1})$, the analyst then uses the squared equation 5 regression residuals as the dependent variable in a second, conditional variance of wellbeing equation as a function of the same (or potentially other) explanatory variables:

$$\varepsilon_{i,t}^2 = \sum_k \alpha_k W_{i,t-1}^k + \delta_1 X_{i,t} + \vartheta S_{i,t} + \omega_{i,t} \tag{6}$$

Using the two estimated conditional moment functions (equations 5 and 6), and an assumed two-parameter probability distribution (e.g., normal, lognormal, gamma, exponential, beta), one then computes the household-and-period-specific conditional probability density function of wellbeing. The inverse cumulative probability above the poverty line, FCS threshold (or other normative wellbeing standard), given the values of other covariates, yields the resilience score. Given the dynamics incorporated into equations 5 and 6, for each time period *s* up to *T* periods into the future the analyst constructs the resilience score, RS $(\rho_{i,s})_{s=0}^{T}$ as:

$$\rho_{i,s} \equiv \Pr\left(W_{i,t+s} \geq \underline{W} | W_{i,t}, X_{i,t}, S_{i,t}\right) = F\left(\underline{W}, W_{i,t}, X_{i,t}, S_{i,t}\right) \tag{7}$$

Where $F(\cdot)$ is an assumed two-parameter inverse cumulative density function. In this paper we use the gamma distribution, following Cissé and Barrett (2018).

The resilience score, RS, represents the estimated conditional probability of having an acceptable level of wellbeing in a given period. RS necessarily falls in the [0,1] interval. Multiply by 100 and it provides a comparably-scaled comparison to the RIMA and TANGO RCI, but with the important difference that RS is in percentage (i.e., probability) units, while the factor analysis-based TANGO and RIMA RCIs are unitless indices by construction. The inclusion of (a polynomial in) the lagged value of the wellbeing measure and the possibility of time-varying explanatory variables permits estimation of a complete time series of household-specific RS estimates. That is presently infeasible in the RIMA or TANGO methods. As with RIMA and TANGO, the core data requirements are multiple observations of an outcome variable over time for the same households (or individuals), as well as sufficient other data on explanatory variables so as to plausibly examine the controlled association between resilience and shocks or covariates of interest.

The C&B method compares each household's resilience score, ρ_i, against a minimally acceptable likelihood of achieving some normative wellbeing standard, like a poverty line.[8, 9] As with the wellbeing threshold itself, this probability threshold may vary based on the assessment goals. One may, for example, consider those "resilient" – i.e., $R = 1$, comparable to that in equation 3 – who have a relatively low probability of falling above a relatively high poverty threshold. Or one might classify as "resilient" those who have a very high probability of avoiding falling below a very low threshold. Adjusting these thresholds also allows for purposeful adjustment of the likelihood of errors of exclusion or inclusion in targeting based on RS (Upton et al. 2016). For the purposes of direct comparison with RIMA and TANGO in the analyses that follow, we designate as "resilient" the exact same percentage of households as identified by the RIMA/TANGO binary resilience indicator, *I*. This ensures that we compare among a similarly-sized subpopulation.

III. Comparative Methods

Construction of the RIMA RCI, C&B score, and the resilience classifiers I and R/$\tilde{R}$ require selection of a wellbeing outcome indicator. We use two key candidates available across data sets: real consumption expenditures, the primary poverty indicator used globally, and dietary diversity as measured by the Food Consumption Score (FCS, see Weisman et al. 2009), perhaps the most widely used food security indicator globally. As the results are very similar – and also due to greater data issues with consumption expenditures, noted below – we discuss results for the FCS, relegating consumption expenditures results to the Appendix.

In this section, we explain the statistical comparisons we use among the resilience (RCI and RS) measures and resilience classifiers (I, R / $\tilde{R}$). We use the same exact panel data, from two separate countries, to construct these measures for each household under each of the three methods. This assures that any differences that emerge arise entirely due to variation in how each measure uses the data, not to some idiosyncratic feature of a single data set.

We will conduct all analyses at the household level. We then supplement these assessments by comparing each measure's performance for distinct subpopulations defined by indicators commonly used for household-level targeting, such as female-headedness, livestock or land ownership, or seasonal migration. We do this because the household level analyses might raise (at least) two types of concerns.[10]

First, household-level correspondence with FCS or expenditures measures will necessarily be imperfect because of stochastic realizations and measurement error. So perfect correlation with the

[8] We compute only the current period RS, and set aside the possibility of a time sequence, because that is not estimable under the other methods.

[9] The C&B resilience score can then be aggregated across a population and then decomposed into subgroups following the FGT class of poverty measures (Foster et al. 1984), to look at the prevalence (and depth) of resilience at more aggregate population and sub-population levels, an appealing feature not available in the RIMA or TANGO methods.

[10] We thank editor Andrew Foster and an anonymous reviewer for pushing us to undertake group-level analyses.

focal wellbeing measure is unrealistic. But at a bare minimum, one should find at least statistically significant positive correlation.

Second, because poverty and food insecurity typically exhibit spatial concentration, even within low-income countries, development and humanitarian agencies often use geographic targeting to target communities. In some cases, they leave household-level targeting to communities. But in a great many other cases, within targeted communities, agencies use indicator targeting to focus resources on intended beneficiaries thought most likely to be poor, food insecure, or not resilient. If stochasticity and measurement error offset among individual households within a subpopulation, then a measure that does not correspond well with household-level realizations should fare much better at the more aggregate, subpopulation level. As we demonstrate, the core messages from the household-level messages carry through to the more aggregate assessments.

a. Distributions

The first comparison of interest is whether the various measures generate similar distributions when computed from the same data. Each measure differs in its construction. So one naturally wonders, are the measures' distributions similar but centered or spread differently, so that the difference is essentially one of scaling? Or do they depict fundamentally different distributions? To examine this question we simply compare the kernel density estimates of the distribution of household-specific RCI, RS, and $\tilde{R}$ measures (all scaled to between 0 and 100), for each measure in each country. This comparison matters to understanding how one's choice of methods may influence simple descriptions of the headcount prevalence of resilience among households and of the magnitudes of changes needed to 'build resilience'.

b. Rankings of households by RCI, RS, and $\tilde{R}$

Whether or not the distributions of the measures are similar, do they order households similarly? Put differently, do different measures rank households differently from least to most resilient? This has implications for sub-group descriptions and for targeting interventions toward the least (or most) resilient households. If any agency were, for example, to have a budget sufficient to assist some fixed number of households, how would the measurement method used affect resource distribution?

To answer that question, we use two comparisons. First, for each pair of measures we estimate the Spearman rank correlation coefficient, a nonparametric measure of the monotonic ordinal relationship between two variables –i.e., their ranking, without regard to the distance between or concentration of observations (Hauke and Kossowski, 2011). A high rank correlation coefficient signals that a relatively higher resilience capacity or resilience score in one measure corresponds to a higher ranking in the other. This coefficient estimate is especially useful when comparing unitless measures like the RIMA and TANGO RCI or when comparing between measures that use different units (e.g., C&B as distinct from RIMA or TANGO), and/or are very differently distributed. We also compute the rank correlation coefficients between each RCI/RS measure, $\tilde{\mathrm{R}}$, and the underlying wellbeing measures they are meant to reflect, per the wellbeing-based definitions that conceptually underpin each measure. Targeting has typically been based on poverty or food security indicators (or proxies for those indicators), so how does use of resilience measures differ? While one would not expect perfect correlation, else the resilience measures would be redundant to established wellbeing measures, intuitively one prefers a reasonably high magnitude, positive correlation in rankings.

Second, because most resilience programming focuses on the poorest and most vulnerable, we are especially concerned about the correspondence among measures in ranking the *worst off*. If, for example, different resilience measures do not coincide for the wealthiest, but correspond closely in identifying the least resilient households, imperfect correspondence may suffice for targeting purposes. We hence also identify which share of those ranked in the bottom 20% by each measure are likewise found in the bottom 20% by the other measures, as well as by the wellbeing indicator itself.

c. Internal consistency between RCI and I

Conceptually, an increase in RCI or RS signals a higher likelihood of being resilient. Therefore, households classified as resilient by the R or I classifier (equation 3 for RIMA and TANGO) should exhibit statistically significantly higher RCI than households classified as not resilient. We test for this internal consistency by performing one-sided t-tests of the null of no difference in the continuous RCI measures between the two subpopulations classified as I=0 and I=1 by that same method. We replicate this analysis for the CB method. Note, however, that because RS is a direct estimate of resilience outcomes (as distinct from capacities), it mechanically satisfies the internal consistency check so long as the underlying equations (5 and 6) exhibit any significant explanatory power. We repeat this exercise at the subpopulation level as a check whether any apparent underperformance at household-level is perhaps due just to idiosyncratic realizations of stochastic shocks and to measurement error.

d. Out-of-sample prediction of wellbeing

A key performance test of any measure, especially one estimated in a population-representative sample, is how well it estimates out of sample, i.e., when applied to observations not from the original estimation (sub-)sample. We implement several different out-of-sample predictive performance comparisons.

First, we test out-of-sample performance in the time series, looking at how accurately RCI/RS/$\tilde{R}$/I measures from one period predict wellbeing measures in a future period. Conceptually, the objective is to establish whether measures generated in data today provide tolerably accurate predictions of wellbeing of those same measured units in the future. We consider this the most important out-of-sample predictive accuracy test because the concept of resilience is rooted in stochastic wellbeing dynamics, so a resilience measure should positively covary with future wellbeing, i.e., can predict how likely the household is to be well off in the future. This test reflects the targeting challenge humanitarian and development agencies face all the time: can we use today's status to predict tomorrow's wellbeing in the absence of an intervention?

We undertake two distinct out-of-sample predictive performance tests in the time domain. First, we graphically examine the correspondence between the binary realized resilience classification in the second-to-last period in the data, and those who fall above the FCS threshold of 35, following Weisman et al. (2009), in the last period; i.e., how each realized resilience indicator directly associates with next-period wellbeing. We report the conditional frequencies for those who fall above or below the FCS threshold in both, either or neither periods. That is, conditional on resilience classification in the prior period, what proportion winds up food secure or not? These probabilities reflect the measure-specific rates of correct classification (as resilient/food secure or not), false positives (i.e., type I errors in which the household would subsequently prove food insecure but is mistakenly classified as resilient), and false negatives (i.e., type II errors in which the household would subsequently prove food secure but is mistakenly classified as not resilient). We superimpose on those scatter plots the locally weighted polynomial (LOWESS) nonparametric regression of final period wellbeing outcome on the prior period resilience measure so as to summarize the average relationship between them to check for monotonicity, as one should expect a higher resilience measure to be positively associated with subsequent wellbeing.

Second, we run the bivariate regression of observed wellbeing outcomes in the last survey wave on the predicted, continuous, RCI/RS and R for the final survey wave as estimated in data from all-but-final survey waves:

$$W_{real,i,t} = \alpha + \gamma \, \widehat{RCI}/\widehat{RS}_{i,t-1} + u_{i,t} \tag{8}$$

The greater the explanatory power of the RCI/RS, the lower the Root Mean Square Error (RMSE) from estimating equation 8. We test the null hypothesis of equal RMSE across models based on different measures of the explanatory variable by bootstrapping each regression, using the same number of clusters as in the original sample, and testing the resulting distributions of each RMSE to generate a statistical test of relative, forward-looking, out-of-sample predictive performance in the time domain.

Then we repeat the statistical predictive performance evaluation with cross-sectional out-of-sample prediction.[11] Conceptually, this provides an indicator as to how reliably one could use data from one program area to target programming in a nearby area where data were not collected. To perform this test we estimate the three resilience measures with a random draw of 75% of the sample from two survey waves, predict the measures in the other 25% sub-sample off the estimated factor loadings and/or regression coefficient estimates from that estimation sub-sample, and then compare the predicted RCI/RS with actual wellbeing indicators in the smaller, testing sub-sample, again using equation 8 and a test for differences in RMSE among different measures. We cross-validate the procedure by performing this random draw and prediction tasks a total of ten times, and report the mean RMSE across iterations.

IV. Data

a. Living Standards Measurement Surveys for Niger and Ethiopia

For our analyses we use the Living Standards Measurement Survey – Integrated Surveys on Agriculture (LSMS-ISA) data from Ethiopia and Niger. LSMS-ISA data are recurring, nationally representative population surveys collected by the World Bank in collaboration with national governments and other partners. These countries represent two of the regions of focus for resilience measurement efforts: the Horn of East Africa, and the Sahel of West Africa. The data are comparable but also vary somewhat in structure (e.g., number and timing of rounds, sampling protocols), providing slightly different opportunities to explore these methods and provide robust comparisons. If our findings are consistent between these two data sets in these different contexts, then that builds confidence that any differences we identify among methods arise due to differences in how one constructs the measures, rather than something idiosyncratic to a specific data set.

The LSMS-ISA project started in Ethiopia in 2011-2012, building on previous government efforts with the Agricultural Sample Survey (AgSS), with additional waves collected in 2013-2014 and 2015-2016. The first wave, which excluded urban areas, stratified within representative regions based

[11] There is no analogous, graphical way to do the out-of-sample performance evaluation in cross-section, so we only use the statistical approach in cross-section.

on rural versus small town status.[12] Sample households were then selected within Enumeration Areas (EAs), with EA quotas for the most populous regions to assure a representative sample for those regions, and a representative sample across the cluster of remaining smaller regions. In total, 290 EAs (3466 households) were chosen in rural areas in the first wave, and 43 (503 households) in small towns. Urban areas were added in 2013-2014. Sample sizes were calculated so as to be representative for the largest regions and Addis Ababa, a total of 1500 urban households. A third wave was implemented across all households in 2015-2016.[13] Household visits occurred in January – April in each survey round. This period follows the *meher* harvest in the agricultural areas, and falls during the *belg* rains and planting in *belg*-receiving areas. In pastoral areas, it is a lean season, also near the start of the *diraac/sugum* or *gu/genna* rains in the Northern and Southern parts of the country, respectively.

The first Niger survey was launched in 2011-2012, as the Niger *Enqu*ête *National sur les Conditions de Vie des Ménages et Agriculture*, with one follow-up wave in 2014-2015. The sampling was designed to be representative across eight major regions, and the capital region of Niamey, with further stratification across four livelihood zones: agricultural, agro-pastoral, pastoral, and urban. The number of enumeration areas, called *zones de dénombrement* (ZDs), was assessed as proportional within each region based on the 2001 census.[14] Each household was visited twice, once during the post-planting or lean period (*soudure*), and once during the post-harvest period. For wave 1, the first visit was between July and September 2011, and the second visit November 2011 and January 2012; for wave 2, the first visit was between September and November 2014, and the second visit was between January and March 2015. The rainy season generally falls between June and September, with planting between May and July and harvesting starting in October to as late as January. While survey timing shifted between waves, both pairs of visits fell so as to have a high likelihood of reaching households in the post-planting and post-harvest periods.[15]

b. Outcome Variables

Both RIMA and C&B explicitly integrate a wellbeing outcome into the construction of the RCI or RS, and all three methods do so for categorical resilience (R or I). For an additional cross-check on our comparisons, we performed our analyses with two different indicators for wellbeing: the Food Consumption Score (FCS, a measure of dietary diversity) and real per adult equivalent[16] consumption expenditures. The results are quite similar across different outcome measures. We present FCS primarily because it is collected in the same way across countries, so is most directly comparable. Cross-country differences in the precise content, recall period, and aggregation levels of survey expenditure modules can generate differences in expenditure measures. So to emphasize the most comparable metric, and conserve on space here, we relegate the expenditures based assessment to the Appendix.

---

[12] Small towns were designated to be those with fewer than 10,000 residents based on the 2007 census, with "urban centers" later designated as those with over 10,000 residents.

[13] In the interest of using the three wave panel, our analysis excludes the urban households. The full sampling design, by region and livelihood, and maps of EAs, are provided in the Appendix.

[14] The Appendix includes the sample design by region and livelihood zone for the first wave, and a map of the ZDs.

[15] One might naturally worry about seasonality. Unfortunately, there is no seasonal variation in the Ethiopia data. Although the Niger data include both post-planting and post-harvest survey waves, there are only two of each. We therefore cannot do any comparative out-of-sample prediction at seasonal frequency. While the precise levels of estimated food security and resilience may vary seasonally, the comparisons that are this paper's focus should be unaffected by seasonality. First, all metrics compared should suffer similarly. Second, our predictive performance is assessed either in a cross-section (for the RCI measures) or in draws on the full panel (C&B), so the same seasonal conditions are present in both testing and training samples.

[16] We calculated adult equivalent units using the World Health Organization conversion codes, as discussed by Dercon and Krishnan (1998) and previously applied in Ethiopia. The full equivalency table is provided in the Appendix.

The FCS is constructed by dividing foods into eight different groups, weighted based on dietary quality, with the weights multiplied by the number of days each was consumed of the previous seven, and then summed across groups (Weisman et al. 2009).[17] A household with FCS ≥ 35 is typically considered food secure. A 7-day recall was used in both countries, collected for both seasons in Niger, only once per survey wave in Ethiopia.

c. Household and community characteristics

The three methods we replicate lay out the exact household and community controls they recommend with varying degrees of specificity. Where the original authors of each method are specific, we attempt to match them as precisely as possible given the data. Smith and Frankenberger (2018) specifically enumerate the variables and indicators to be used in the TANGO method. So as to maximize comparability across models, we therefore apply the indicators they prescribe, in the manner most in keeping with the intent expressed by each method's authors. The components of the TANGO RCI need to be adapted slightly in the data, as not all of the specific indicators used by Smith and Frankenberger (2018) – drawn from a special purpose survey – are available from the (quite extensive) LSMS-ISA surveys. Hence, we include the intersection of what is prescribed by the authors and the data collected in the LSMS surveys, and other available indicators that best reflect the concept, and had the authors check the details of our application of the method. For RIMA, we can replicate more exactly FAO (2016), as all of the prescribed indicators are included in the LSMS data. The RIMA indicators are a subset of those designated by TANGO. C&B offers the most flexibility in choice of explanatory variables, as the authors do not designate specific controls. In order to maximize the comparability, we use the larger set of controls prescribed by the TANGO for the X_{it} vector in each of the three methods. Several of these are included after the initial factor analysis on the sub-component, per the procedures specified by each method and reported above.

All models integrate household characteristics and assets that may affect household-level shock response and recovery. Following TANGO we include: household head age, sex, education, and participation in different livelihood activities; along with household size and breakdown by different age and sex groupings. RIMA integrates a subset of these as part of the human capital component of the Adaptive Capacity pillar. Household assets form part of the RCIs, and include durable and productive asset indices (constructed using principal components analysis), livestock holdings combined into Tropical Livestock Units,[18] and access to / value of agricultural land. Many other household and community-level indicators, such as access to information and services, presence of certain infrastructures, etc. are also included within the RCIs.

d. Climate Shocks

The concept of shocks is integral to all three approaches. The C&B method prescribes that shocks and stressors measures be included as explanatory variables, whenever possible. In both Niger and Ethiopia drought is widely believed the dominant shock impacting food security and other wellbeing measures. Hence, we include the Normalized Difference Vegetation Index (NDVI), a measure of vegetation density widely used to reflect water availability and other (e.g., pest, plant disease) natural shocks. The index is spatially calibrated to match the household data at the EA/ZD level. In Ethiopia,

[17] The food groups (and weights) are as follows: staples (2), pulses (3), vegetables (1), fruit (1), meat and fish (4), milk (4), sugar (0.5), oil (0.5).

[18] Livestock holdings are measuring using Tropical Livestock Units, with the following conversion factors: camels = 1; cattle = 0.7; horses = 0.8; donkeys/mules = 0.5; sheep/goats = 0.1; pigs = 0.2; and chickens = 0.01.

we use a 15 km radius, while in Niger we vary the radius of the NDVI indicator, increasing it from 15km in settled areas to 50km in pastoral locations to reflect extensive grazing patterns not relevant in the Ethiopia sample. NDVI corresponds most closely with crop and forage yield over the growing season. We therefore use the average NDVI of the growing months preceding the post-harvest period, during which consumption data were collected. We then calculate the z-score of NDVI for each survey location, relative to the entire period of data availability, and censor NDVI observations ≥ -1.5 standard deviations from the mean so as to focus on extreme adverse events, the roughly bottom 5% of period average realizations.

V. Results

a. Wellbeing Outcomes and Underlying Resilience Indicators

We begin by describing the outcome indicators and other characteristics used in the analyses. Table 1 shows the mean FCS across waves for both countries, along with the percentage classified as food insecure (FCS≤35). Between 21% (Wave 1, post-planting) and 11% (Wave 2, post-harvest) of households in Niger are considered food insecure. In Ethiopia, based on the same FCS threshold as Niger, roughly one-third of households are classified as food insecure in each survey wave.

Table 2 then describes the components of the RCIs, estimated following each method's developer's recommended practices (e.g., using just baseline values for the TANGO RCI but estimating the RIMA RCI jointly using all survey rounds). Many of these are composed of several different indicators (the full set is described in the appendix) Note that components that are drawn from factor analyses are unitless, hence comparing across countries is not meaningful. As noted previously, for the C&B model we include the sub-components of the TANGO index – as in, the components underneath each of the broader "resilience capacities" (Absorptive, Adaptive, and Transformative) as controls, given that the method does not specify exactly what to include other than that it include characteristics most likely to affect changes in wellbeing (with the majority intrinsically controlled for via the lagged wellbeing indicator).

b. Resilience Capacity Indexes, Resilience Scores, and their distributions

Our comparison starts by describing the resilience indicators, first through basic summary statistics of the key indicators (Table 3) and then in kernel density estimates of their distributions (Figure 1). Although each of the four distributions differs statistically significantly from each of the others (the Kolmogorov-Smirnov test p-values are all well below 0.01), the RIMA and TANGO RCI are noticeably more similar to one another than to the C&B RS or the realized resilience measure, $\widetilde{R}$ (intertemporal FCS difference). The RIMA RCI distribution appears as a more dispersed, less peaked variation of the TANGO RCI distribution. The similarity makes sense since these two measures are each generated by factor analysis of largely similar variables. The distribution of the C&B RS arises partly from the FCS threshold used in its construction. The mean would be higher (lower) and distribution shifted to the right (left) if we used a lower (higher) threshold for resilience classification. As a reminder, since the conditional probability threshold that identifies a unit as resilient has to be specified by the researcher, we selected the one that generates precisely the same sample share classified resilient as in RIMA and TANGO – which both use the method reflected in equation 3 – so as to facilitate further comparisons. Unlike the other three distributions, realized resilience ($\widetilde{R}$) exhibits a relatively Gaussian distribution centered near the midpoint between the observed sample maximum and minimum FCS changes, broadly consistent with what one would expect from iid shocks – whether real or measurement error – to the time series. Lastly, the marked differences in these distributions implies that each method would generate different estimates of the prevalence of high or low resilience or resilience capacities within the population, leading to different descriptive conclusions.

c. Ranking Households

Do measures order households similarly in terms of their resilience or resilience capacities? Figure 2 exhibits the rank order correlation coefficients across all resilience measures and FCS outcomes. The strength of the correlations varies significantly across combinations of measures, and to some degree across countries. Consistent with finding somewhat similar distributions, we find a relatively strong correlation (0.52-0.70) between the RIMA and TANGO RCIs. The RIMA RCI and C&B RS also correlate relatively strongly in Niger (less so in Ethiopia), with a correlation of 0.697 (0.441) between the RS and RIMA. The TANGO RCI and C&B RS correlate more weakly (0.39-0.47).

More striking differences emerge when we look at how well the measures rank correlate with the FCS outcome, with which the resilience measure are conceptually intended to correspond, albeit imperfectly. The TANGO RCI correlates weakly with FCS (0.22-0.25), RIMA RCI somewhat better (0.23-0.40), the C&B RS slightly better still (0.41-0.48), and $\tilde{R}$ best (0.52-0.56), although that is somewhat mechanical since FCS is the first term of the difference variable that is $\tilde{R}$. Indeed, realized resilience, $\tilde{R}$, does not correlate at all with the RIMA measures and only very weakly with the TANGO RCI and C&B RS, and negatively in one country for each. Intertemporal change in FCS ($\tilde{R}$) clearly orders households quite differently than do the other measures.

In most cases we are most likely to be concerned with identifying the least resilient households. So we concentrate on whether methods rank the same households in the bottom quintile. As shown in Figure 3, households ranked in the bottom 20% by one measure (the rows) are unlikely to appear in other measures' (the columns). Of the 20 comparisons, the maximum is that 60% match of membership in the bottom quintile of the resilience measure and the average is just 32%. So these measures do not identify the same households reliably, whether or not we focus on the bottom of the population distribution.

Together with the previous findings about the distributions of each measure, it is clear that these measures generate significantly different descriptions of the same population using the same data, in terms of both the prevalence of high or low resilience (capacities) and the identity of those with high or low measures. One might reasonably contend that the RIMA and TANGO RCI measures, intended more as explanatory variables capturing resilience capacities – do not need to correspond well with contemporaneous wellbeing measures like FCS, unlike the $\tilde{R}$ or C&B RS measures, as they are meant to predict households' future state, not describe current ones. Such concerns lead naturally to our upcoming comparisons, especially of out-of-sample predictive performance.

d. Internal Consistency

A binary indicator of resilience – taking value 1 if resilient, 0 if not resilient –should be positively correlated with the corresponding continuous measure of resilience or resilience capacities. We commonly overlook such internal consistency issues because binary indicators are typically defined by discretizing an underlying continuous measure, thereby automatically generating that correlation. There exists precisely such a mechanical positive correlation between I and R/$\tilde{R}$ by virtue of the definition of I. Likewise, in the case of the C&B measure the correspondence follows automatically from the definition of the binary classification as a function of the continuous RS measure. That same mechanical correspondence does not exist, however, between the RIMA or TANGO RCI measures and the realized resilience employed for classification in studies that use those methods, because R (and its transforms, I and $\tilde{R}$) is computed entirely separately from the RCI indices.

Table 4 reports the difference in mean RCI or RS across the groups classified as not resilient versus resilient by each method. The resilient, as determined by the binary realized resilience measure I, comprise 60% of the Niger sample and 53% in Ethiopia. In the left-hand block of Table 4, we show the resilience classifications that follow from the C&B method and test the hypothesis that the RS (and FCS) differ between not resilient and resilient subsamples. We expect to and easily do reject the null hypothesis of equal means in favor of the one-sided alternate hypothesis that the mean RS score for households classified resilient exceeds the mean RS for those classified not resilient. In the righthand block of Table 4, the continuous $\tilde{R}$ measure is likewise significantly higher for the resilient group than the non-resilient group under the I measure, as it must be since both measures monotonic transforms of the same R measure**.** Those results are therefore somewhat tautological.

Mean FCS – the wellbeing measure of interest – differs statistically significantly between the resilient and not resilient groups as classified by either the C&B or the realized resilience method. That is an encouraging result, as it does not follow automatically from the construction of the resilience classifier but does signal that those classified as resilient are more food secure than those deemed not resilient, as one would expect.

By contrast, in the righthand block of Table 4 we see that the mean TANGO and RIMA RCIs for subsamples classified as realized resilient (I=1) never statistically significantly exceed those for the non-resilient. This signals an internal inconsistency between the binary resilience classifier – realized resilience – and the continuous RCI measure. The internal inconsistency in the RIMA and TANGO methods between the factor analysis-based RCI measure and the household binary resilience classification measure that does not use the RCI raises important concerns. Since higher resilience capacity should translate into a higher likelihood of being resilient, this suggests that one cannot simultaneously believe either method's measures in both the RCI and R/I domains. One must choose, and presumably favor the RCI measure that has been more the focus of the method's development, not the less thoughtfully developed R measure in each method. This naturally leads to the question of how these methods perform in out-of-sample prediction.

e. Subpopulation-level correspondence

As discussed early, idiosyncratic risk and measurement error may make it unrealistic to expect high correlation between resilience measures and the underlying wellbeing concept to which they are targeted (FCS in the present case). But at subpopulation scale the noise arising from stochasticity and measurement error should largely cancel out. So we replicate a version of Table 4 looking at the resilience classification associated with each measure and FCS-based food security classification, for distinct subpopulation defined by a commonly-used indicator targeting variable: female headship, high dependency ratio, an uneducated household head, low or no livestock holdings, landlessness, no receipt of remittances, and seasonal migration. We would expect an effective measure of food security resilience to generate similar pattern of differences as the underlying FCS measure at the group level. Indicators associated with statistically significantly differences in food security status should likewise manifest a statistically significantly association in the corresponding resilience measure.

As shown in Table 5, the basic patterns found in Table 4 persist at targetable subpopulation scale. Even allowing for aggregation to reduce the noise arising from idiosyncratic household-level shocks and measurement error, we find each of the resilience indicators has quite imperfect correspondence with the actual FCS classifications into food secure or food insecure. Of the 14 country-indicator pairs we study, the C&B binary classification matches the FCS intergroup differences in 11 cases, 7 of them with inter-group differences that are statistically significant by both measures. While this is more consistent than the alternative binary metric, only half of the country-indicator pairs matched by ordering and statistical significance. This does not seem an especially impressive performance for a subpopulation measure. The RIMA/TANGO binary indicator does even worse, only matching FCS groupwise difference orderings in 4 of 14 cases, and with only 2 of those statistically significant for both I and FCS.

f. Out-of-sample predictive performance

As described earlier, in studying out-of-sample predictive performance, we first look at graphical descriptions and conditional frequencies relating resilience scores for the second-to-last time period (horizontal axis), compared against a depiction of the outcome measures in the final period (vertical axis), with lines at the resilience cut-off in the second-to-last period and the FCS threshold in the last period (Figure 4). This allows visual comparison of how the binary realized resilience classification predicts the subsequent realization of the wellbeing measure. We then report four conditional frequencies: of having been correctly classified as resilient (upper right) or not (bottom left), and the likelihood of having been mis-classified as resilient (bottom right) or as not resilient (upper left). If the prediction of food security based on the last-period resilience score were perfect, there would be no red dots.

The C&B score succeeds in correctly classifying 58% of households in Niger, and 64% in Ethiopia. Although perhaps not impressive, these are statistically significantly better than a coin flip. Type I errors (mis-classified as resilient) are very low in Niger (2%), somewhat higher in Ethiopia (15%); whereas Type II errors (mis-classified as not resilient) are much higher, 40% and 20% in Niger and Ethiopia, respectively. This is relatively promising, in that Type I errors are more problematic from a humanitarian perspective, which prioritizes not missing those who prove food insecure, but still does not reflect great accuracy. The Lowess estimates show a monotonically positive association between prior period resilience and subsequent food security, as they should.

The lower panel results are a bit more grim, reflecting the weak power of R in predicting FCS. Realized resilience correctly classifies 52% of households in Niger, 51% in Ethiopia, not statistically significantly better than a coin flip. Type I errors are somewhat higher than C&B in both cases, at 5% and 20%, as are Type II errors at 43% and 29%, in Niger and Ethiopia, respectively. Notably, the Lowess estimates show a counter-intuitive, non-monotonic relationship; both those with systematically lower and higher R are more likely to be food secure in the next period. This signals a serious weakness in the use of 'realized resilience', R (or its transforms), as a predictor of future wellbeing.

Finally, we report cross-sectional and forward-looking out-of-sample statistical tests of the predictive accuracy of the continuous RCI and RS measures with respect to wellbeing. The upper and lower panels of Table 6 report the RMSE of the equation 8 regression models of the cross-sectional and time series out-of-sample predictions, respectively, of FCS based on RCI/RS for each method, for both countries. Recall that RMSE reflects the standard deviation of the unexplained variance in FCS, thus a lower RMSE indicates better prediction.

The C&B RS predicts wellbeing significantly more accurately than either RIMA or TANGO RCS in both the cross-sectional and intertemporal domains, and is more accurate intertemporally than the RIMA/TANGO classifier, R.[19] No such ordering is available among the RIMA RCI, TANGO RCI and R measures, each of which dominates the others in one or more test. But perhaps the most important out-of-sample forecast accuracy finding is that even the best (C&B RS) resilience measure does not consistently improve upon the far simpler approach of using the most recent FCS observation to predict the next period's food security status. The ratio of the RMSE to the sample mean (displayed in the bottom row) reflects the relative dispersion of the forecast errors, the normalized standard error. Although no strict criterion exists to determine an acceptable RMSE, a widespread rule of thumb is that a normalized standard error of 0.2 or less signals strong predictive skill while 0.5 or greater indicates poor predictive skill. Each of these measures falls in the intermediate range – from 0.34-0.41 – indicating some predictive skill. But none of the resilience measures predicts appreciably better than the naïve approach of using the most recent observed value of wellbeing to predict future wellbeing. It is therefore unclear what the added complexity of these measures buys analysts in terms of predictive/targeting performance.

VI. Summary and Recommendations

The comparative analyses presented here establish that three increasingly widely-used resilience measurement approaches – FAO's RIMA, TANGO's method, and the Cissé and Barrett (C&B) method – generate measures that differ in substantive ways. They depict markedly different distributions of resilience within populations. They order households differently by resilience, meaning that they do not classify the same households as "resilient" or not. This is also true if one focuses just on those households classified as least resilient, the subpopulation on which resilience building interventions presumably most want to focus. It is also true at subpopulation-level, using any of seven commonly used indicators for household-level targeting (e.g., female headship, landlessness). The RIMA and TANGO methods also exhibit an important internal inconsistency between the methods used to generate a resilience capacity index (RCI) and to classify households as resilient or not. Within the existing set of resilience measures available for application in standard (LSMS-style) household survey data, the C&B method outperforms the RIMA and TANGO measures in groupwise matching with FCS measures and in out-of-sample prediction, but not by much. Moreover, none of the measures consistently outperforms the far simpler approach of using the most recent value of the relevant wellbeing measure to predict the future value of that same variable. So the choice of resilience measure matters to the descriptive and predictive conclusions one would make about stochastic wellbeing dynamics, and the predictive gains from these measures appear negligible.

We offer a few caveats to our comparative analysis. First, the present comparison is neither comprehensive nor ideal. The data are not optimally suited to all methods. More could likely be learned and with greater precision from (considerably more expensive) "fit-for-purpose" data that allows each method to be implemented exactly as prescribed. But part of the task in resilience measurement is generating tolerably good measures using the data national statistical agencies already generate, without having to incur significant additional data collection costs. So testing in publicly available data seems a reasonable first step at comparative performance evaluation of different measures.

[19] The R measure doesn't lend itself to cross-sectional out-of-sample prediction since one would not observe prior periods in the unobserved testing subsample.

Second, we could not examine how each method performs in evaluating the dynamic wellbeing impacts of a discrete stressor, shock or policy. Such analyses have been undertaken with C&B (Cissé and Ikegami 1996; Phadera et al. 2019; Premand and Stoeffler 2020) but are more difficult with RCI measures meant as explanatory rather than outcome variables.

But perhaps the chief takeaway from the analyses reported here is the at-best-mediocre performance of these resilience measures. It remains unclear what these really measure nor what descriptive, inferential, or predictive benefits they yield.

We close with a few conclusions about future directions in resilience measurement, and some modest recommendations for best practices. First, resilience measures that do not directly integrate a wellbeing outcome indicator are unlikely to accomplish what they intend. If we focus entirely on household or community characteristics to measure resilience, we risk presupposing relationships that we would otherwise like to examine, such as, *which* of these factors affects or engenders resilience? This issue is of particular importance for policy actors, whose operational questions often relate to which specific, actionable variables, such as specific assets (e.g., a livestock transfer), products (e.g., insurance or credit), or infrastructure (e.g., irrigation, roads, or health clinics) build resilience. Index-based approaches that use data reduction (e.g., factor analysis) methods to combine all of these features into composite capacity measures obscure the differences among components of any pillar or capacity and make it exceedingly difficult to do rigorous causal inference around any one component of an index.

Second, a simplistic measure of resilience that uses only observed change over time in the wellbeing outcome, without explicitly integrating the level of the measure, is also insufficient. Those with relatively high levels of consumption may suffer a loss, but if they are still quite well off following that loss it may not be appropriate to classify them as non-resilient. By contrast, if one classifies as resilient those who have little or nothing to start with, and so did not suffer a loss – e.g., a homeless person cannot lose a house to an earthquake or tsunami or fire – then the resilience measure fails a simple 'pro-poor' test (Béné et al. 2017). We show that a realized resilience measure based only on a non-negative change in observed wellbeing does not correlate at all with more thoughtfully developed resilience measures, the RCIs or RS, nor with actual wellbeing outcomes. Resilience measures must be positively related to levels of wellbeing, not just to changes in those levels over time (Barrett and Constas 2014; Béné et al. 2017).

If the development and humanitarian communities are to harness the potential of the currently-popular resilience concept, we need a measure (or measures) that allows us to accurately and reasonably inexpensively identify those most likely to suffer from shocks or stressors in the absence of intervention, to be able to rigorously estimate impacts of interventions on the resilience measure(s), and to ensure that any such change reflects improvement in the shock-and-stress-proofing of wellbeing over time. While existing resilience measures have made some progress in those directions, the development community is clearly not there yet. For example, while the C&B approach has slightly greater predictive power than the other two measures and has been used effectively in rigorous impact evaluation (e.g., Phadera et al. 2019, Premand and Stoeffler 2020), it does not consistently add much value as compared to the far simpler method of just using the most recent wellbeing measure available to predict future resilience.

Resilience has become a popular and influential concept in development and humanitarian policy and practice. While some are tempted to say that debates around resilience measurement are behind us, simply because agencies have invested in and mandated the use of one or another measure, our analysis shows that the approaches presently in play are all, at best, imperfect, and at worst deeply flawed. Researchers can play an important role in advancing resilience measurement to support operational programming, in part through rigorous ongoing scrutiny of the measures we employ. Perhaps the underlying data being collected and used in standard surveys is insufficient for the task, and higher frequency measures will prove better suited to this inherently dynamic concept (Headey and Barrett 2015, Knippenberg et al. 2019). Perhaps the workhorse measures need reconceptualization and methodological refinement. Whatever the source, our testing suggests that existing resilience measures do not yet get us to the point of reliable measurement to guide and evaluate development resilience interventions.

References

Alinovi, Luca, Marco d'Errico, Erdgin Mane, and Donato Romano (2010). Livelihoods strategies and household resilience to food insecurity: an empirical analysis to Kenya. Background paper to the *European Development Report 2010*. Fiesole: European University Institute.

Alinovi, Luca, Erdgin Mane and Donato Romano (2008). Towards the measurement of household resilience to food insecurity: An application to Palestinian households. In Sibrian (Ed.), *Deriving food security information from national household budget surveys. Experiences, achievement, challenges* (pp. 137-152). Rome, FAO.

Alloush, Mo (2019). Resilience and Psychological Wellbeing. Dissertation Chapter, University of California at Davis.

Ansah, Isaac Gershon Kodwo, Cornelis Gardebroek, and Rico Ihle (2019). "Resilience and household food security: a review of concepts, methodological approaches and empirical evidence." *Food Security* 11: 1-17.

Barrett, Christopher B. (2010). Measuring food insecurity. *Science* 327(5967): 825-828.

Barrett, Christopher B. and Mark A. Constas (2014). Toward a Theory of Resilience for International Development Applications. *Proceedings of the National Academy of Sciences* 111(40): 14625-14630.

Barrett, Christopher B., Teevrat Garg and Linden McBride (2016). Well-Being Dynamics and Poverty Traps. *Annual Review of Resource Economics* 8: 303-327.

Barrett, Christopher B., Kate Ghezzi-Kopel, John Hoddinott, Nima Homami, Elizabeth Tennant, Joanna Upton, and Tong Wu (2021). A Scoping Review of the Development Resilience Literature: Theory, Methods and Evidence. World Development 146: 105612.

Béné, Christophe, Fahim S. Chodhury, Mamun Rashid, Sabbir A. Dhali, and Ferdous Jahan (2017). Squaring the Circle: Reconciling the need for rigor with the reality on the ground in resilience impact assesssment. *World Development*, 97: 212 – 231.

Brück, Tilman, Marco d'Errico and Rebecca Pietrelli (2019). The effects of violent conflict on household resilience and food security: Evidence from the 2014 Gaza conflict. *World Development,* 119:

203-223.

Cissé, Jennifer D. and Christopher B. Barrett (2018). Estimating development resilience: A conditional moments-based approach. *Journal of Development Economics, 135*(2018): 272-284.

Cissé, Jennifer D. and Munenobu Ikegami (2016). Does insurance improve resilience? Measuring the impact of index-based livestock insurance on development resilience in Northern Kenya. International Livestock Research Institute working paper.

Constas, Mark, Tim Frankenberger and John Hoddinott (2014). Resilience measurement principles: Toward an agenda for measurement design. Resilience Measurement Technical Working Group Technical Series 1. Rome: Food Security Information Network.

Dang, Hai-Anh H., and Peter F. Lanjouw. "Welfare dynamics measurement: Two definitions of a vulnerability line and their empirical application." *Review of Income and Wealth* 63, no. 4 (2017): 633-660.

Dercon, Stefan and Pramila Krishnan (1998). Changes in poverty in rural Ethiopia 1989-1995: Measurement, Robustness Tests, and Decomposition. Center for Economic Studies, Discussion Paper Series. DPS 98.19.

DFID. (2014) *Methodology for reporting against KP14 – Number of people whose resilience has improved as a result of project support*. London: Department for International Development. Available on line at: https://www.gov.uk/government/uploads/
system/uploads/attachment_data/file/328254/BRACED-KPI4-methodology-June2014.pdf
(accessed June 30, 2019).

D'Errico, Marco and Lisa Smith (2019). Comparison of FAO and TANGO measures of household resilience and resilience capacity. Working paper.

D'Errico, Marco, Donato Romano and Rebecca Pietrelli (2018a). Household resilience to food insecurity: evidence from Tanzania and Uganda. *Food Security, 10*(4): 1033-1054.

D'Errico, Marco and Stefania Di Guiseppe (2018b). Resilience mobility in Uganda: A dynamic analysis. *World Development, 104*: 78-96.

D'Errico, Marco, Francesca Grazioli, and Rebecca Pietrelli (2017). Cross-country evidence of the relationship between resilience and the subjective preception of wellbeing and social inclusion: Evidence from the regions of Matam (Senegal) and the Triangle of Hope (Mauritania). *Journal of International Development, 30*(8): 1339-1368.

Food and Agriculture Organization (2016). Reseilience Index for Measurement and Analysis – II: Analysing resilience for better targeting and action. FAO, 2016.

Food and Agriculture Organization (2015). Analysing resilience for better targeting and action: Resilience analysis in Niger, 2011. FAO Resilience Analysis No. 3.

Foster, James, Joel Greer and Erik Thorbecke (1984). A class of decomposable poverty measures. *Econometrica, 52*(3): 761-766.

Garbero, Alessandra Bezawit Beyene Chichaibelu (2019). IFAD Impact Assessment – Participatory Small Irrigation Development Programme I (Pasidp I): Results from a high-frequency data collection – Ethiopia. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3389303

Grimm, M., Waibel, H. and Klasen, S. eds. (2016). *Vulnerability to poverty: Theory, measurement and determinants, with case studies from Thailand and Vietnam*. Berlin: Springer.

Hauke, Jan and Tomasz Kossowski (2011). Comparision of values of Pearson's and Spearman's correlation coefficients on the same set of data. *Quaestiones geographicae, 30*(2: 87-93.

Headey, Derek, and Christopher B. Barrett (2015). Opinion: Measuring development resilience in the world's poorest countries. *Proceedings of the National Academy of Sciences* 112(37): 11423-11425.

Henly-Shepard, S. and Sagara, B. (2018). *Resilience Measurement Practical Guidance Note Series: An Overview*. Produced by Mercy Corps as part of the Resilience Evaluation, Analysis and Learning (REAL) Associate Award.

Jones, Lindsey and Marco d'Errico (2019). Whose resilience matters? Like-for-like comparison of objective and subjective evaluations of resilience. *World Development* 124: 104632.

Jones, Lindsey and Thomas Tanner (2017). 'Subjective resilience': Using perceptions to quantify household resilience to climate extremes and disasters. *Regional Environmental Change,* 17, 229-243.

Jöreskog, Karl G. and Arthur S. Goldberger (1974). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*(351a): 631-639.

Kamanou, Gisele and Jonathan Morduch (2004). Measuring Vulnerability to Poverty in Stefan Dercon, ed., *Insurance Against Poverty*. Oxford: Oxford University Press.

Knippenberg, Erwin, Nathan Jensen and Mark Constas (2019). Quantifying household resilience with high-frequency data: Temporal dynamics and methodological options. *World Development, 121*: 1-15.

Knippenberg, Erwin and John Hoddinott (2017). Shocks, social protection, and resilience: Evidence from Ethiopia, ESSP working papers 109, International Food Policy Research Institute.

Ligon, Ethan, and Laura Schechter (2003). Measuring vulnerability. *Economic Journal* 113(486): C95-C102.

Lipton, Michael, and Martin Ravallion (1995). Poverty and policy. In Jere Behrman and T.N. Srinivasan, eds., *Handbook of Development Economics* Volume 3, Part B. Amsterdam: Elsevier.

Maxwell, Daniel, Bapu Vaitla, and Jennifer Coates (2014). How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia. *Food Policy* 47: 107-116.

Phadera, Lokendra, Hope Michelson, Alex E. Winter-Nelson, and Peter D. Goldsmith (2019). Do asset transfer build household resilience? *Journal of Development Economics, 138*: 205-227.

Premand, Patrick and Quentin Stoeffler (2020). Do cash transfers foster resilience? Evidence from rural Niger. World Bank working paper.

Ravallion, Martin (2019). On measuring global poverty. National Bureau for Economic Research working paper 26211.

Smerlak, Matteo and Babu Vaitla (2017). A non-equilibrium formulation of food security resilience. *Royal Society Open Science,* 4(1):160874.

Smith, Lisa and Tim Frankenberger (2018). Does resilience capacity reduce the negative impact of shocks on household food security? Evidence from the 2014 floods in Bangladesh. *World Development 102*(2018): 358-376.

Smith, Lisa, Tim Frankenberger, Ben Langworthy, Stephanie Martin, Tom Spanggler, Suzanne Nelson, and Jeanne Downen (2015). Ethiopia Pastoralist Areas Resilience Improvement and Market Expansion (PRIME) Project Impact Evaluation. Feed the Future FEEDBACK project report, USAID.

Twigg, J. (2009) *Characteristics of a Disaster-Resilient Community. A Guidance Note.* NGO Inter-agency group. Available on line at: http://community.eldis.org/.59e907ee/Characteristics2EDITION.pdf (accessed June 30, 2019).

UNDP (2013) *Community Based Resilience Assessment (CoBRA) Conceptual Framework and Methodology.* United Nations Development Programme (UNDP) Drylands Development Centre. Available on line at: http://www.disasterriskreduction.net/fileadmin/user_upload/drought/docs/CoBRA%20Conceptual%20Framework%20and%20Methodology%20-%20 Post%20Arusha%20-%204%20April%202013.pdf (accessed June 30, 2019).

Upton, Joanna B., Jennifer D. Cissé and Christopher B. Barrett (2016). Food security as resilience: reconciling definition and measurement. *Agricultural Economics, 47*(2016): 135-147.

USAID. (2013) The Resilience Agenda: Measuring Resilience in USAID. Available on line at: https://www.usaid.gov/sites/default/files/documents/1866/Technical%20Note_Measuring%20Resilience%20in%20USAID_June%202013.pdf (accessed June 30, 2019).

Vaitla, Bapu, Jennifer Coates, Laura Glaeser, Christopher Hillbruner, Preetish Biswal, and Daniel Maxwell. 2017. The measurement of household food security: Correlation and latent variable analysis of alternative indicators in a large multi-country dataset. *Food Policy* 68: 193-205.

Vaitla, Bapu, Jennifer Denno Cissé, Joanna Upton, Girmay Tesfay, Nigussie Abadi, and Daniel Maxwell (2020). How the choice of food security indicators affects the assessment of resilience – an example from northern Ethiopia. *Food Security* 12(1): 137–150

Vollenweider, Xavier (2015). Measuring climate resilience and vulnerability: A case study from Ethiopia. Final Report, FEWS.NET TSC Project, United States Agency for International Development. October 2015.

Watts, Michael. (2016). From vulnerability to resilience: Hans-Georg Bohle's scholarship and contemporary political ecology. *Die Erde (Journal of the Geographical Society of Berlin) 147(4):* 252–265.

Webb, Patrick, Jennifer Coates, Edward A. Frongillo, Beatrice Lorge Rogers, Anne Swindale, and Paula Bilinsky (2006). Measuring household food insecurity: why it's so important and yet so difficult to do. *Journal of Nutrition* 136(5): 1404S-1408S.

Wiesmann, Doris, Lucy Bassett, Todd Benson, and John Hoddinott (2009). *Validation of the World Food Programme's food consumption score and alternative indicators of household food security*. International Food Policy Research Institute discussion paper 00870.

World Bank (2017a). Living Standards Measurement Surveys: Microdata Catalogue. Available on line at: http://microdata.worldbank.org/index.php/catalog/lsms (accessed May 4, 2017).

World Bank (2017b). Living Standards Measurement Study – Integrated Surveys on Agriculture. Available on line at: http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:23512006~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html (accessed May 4, 2017).

Tables and Figures

Table 1. Food Security Status (across waves)

	NIGER				ETHIOPIA		
	Wave 1, 2012-2013		Wave 2, 2014-2015		Wave 1	Wave 2	Wave 3
	PP	PH	PP	PH			
Food Consumption Score	57.67	63.02	57.92	65.49	41.26	42.51	42.04
	(25.47)	(23.66)	(22.36)	(24.33)	(17.11)	(17.05)	(17.06)
Share Food Insecure (FCS ≤ 35)	0.21	0.14	0.18	0.11	0.33	0.32	0.32
	(0.41)	(0.34)	(0.38)	(0.31)	(0.47)	(0.47)	(0.47)
Observations	3969	3969	3621	3621	3969	3776	3699

Notes: Standard errors shown in parentheses

Table 2. Index Components, TANGO and RIMA (final wave)

	NIGER		ETHIOPIA	
	Mean	SD	Mean	SD
TANGO Model - Capacities				
ABSORPTIVE CAPACITY	*46.43*	*13.25*	*4,79*	*4.49*
Bonding Social Capital	17.73	18.08	4.01	3.89
Asset Ownership	3.64	4.07	7.15	3.57
Safety Nets	1.61	5.2	5.68	5.83
Disaster Preparedness	16.4	15.82	60.67	27.23
ADAPTIVE CAPACITY	*30.64*	*24.05*	*21.15*	*16.84*
Bridging Social Capital	6.12	5.52	7.15	17.74
Linking Social Capital	2.02	8.57	0.06	0.25
Assets (same as above)	3.64	4.07	7.15	3.57
Aspirations	32.75	28.29	-	-
Livelihood Diversity (# of activities)	1.24	0.59	0.14	0.37
Human Capital	23.16	13.11	24.74	17.06
Access to Information	49.31	40.18	0.79	0.94
TRANFORMATIVE CAPACITY	*41.89*	*11.9*	*27.04*	*12.81*
Bridging Social Capital (same as above)	6.12	5.52	4.01	3.89
Linking Social Capital (same as above)	2.02	8.57	0.06	0.25
Market Access				
Distance to permanent market (km)	27.02	39.86	7.66	16.99
Distance to periodic market (km)	10.18	18.16	-	-
Access to Services				
Number of key services within village	4.2	2.56	3.53	1.17
Women's Empowerment	25.01	31.25	38.64	18.72
RIMA Model - Pillars				
Access to Basic Services (ABS)	30.56	23.24	21.87	21.63
Asset (ASS)	37.29	5.8	34.53	10.58
Social Safety Net (SSN)	1.84	5.82	4.22	4.43
Adaptive Capacity (AC)	26.97	14.5	29.94	18.28
Observations	3621		3699	

Table 3. Resilience Indices (Final Wave)

	NIGER				ETHIOPIA			
	Mean	SD	Min	Max	Mean	SD	Min	Max
RCI, TANGO method	55.87	11.98	0	100	20.99	9.32	0	66.62
RCI, RIMA method	60.78	6.56	0	93.22	23.28	12.42	0	68.89
$\tilde{R}$, realized resilience continuous measure	49.81	14.35	4.87	100	47.26	13.21	2.38	100
C&B Resilience Score	86.27	12.63	3.52	100	64.85	17.89	9.51	99.43
Number of observations	3621				3699			

All indices are re-scaled from 0-100 for comparative purposes. The RIMA RCI is constructed including both FCS and RAEC as outcome variables for the SEM model, as prescribed by FAO. C&B and $\tilde{R}$ are constructed only using FCS. For C&B, the threshold used for FCS is 35 (following Weisman et al. (2009)).

Table 4. Resilience Capacity Indices, Scores, and Raw Outcome Variables, by Binary Scores (Final Wave)

		C&B Binary			I		
		Not Resilient	Resilient	P-value	Not Resilient	Resilient	P-value
RCI - TANGO	Niger	N/A			*57.08*	*55.44*	*1.00*
	Ethiopia				20.65	21.35	0.02
RCI - RIMA	Niger				*61.28*	*60.98*	*0.89*
	Ethiopia				23.18	23.37	0.32
$\tilde{R}$	Niger				*35.84*	*58.95*	*0.00*
	Ethiopia				37.12	57.41	0.00
C&B RS	Niger	*73.28*	*94.92*	*0.00*	N/A		
	Ethiopia	50.06	79.57	0.00			
FCS	Niger	*53.28*	*73.38*	*0.00*	*51.63*	*74.62*	*0.00*
	Ethiopia	35.81	47.29	0.00	35.12	48.74	0.00
Observations	Niger	*1263*	*1895*		*1213*	*1854*	
	Ethiopia	1,487	1,649		1,820	1,879	

P-values are for t-tests of the null that the resilient and not resilient (column) values are equal for that measure (identified in the leftmost column) and country, versus the one-tailed alternate hypothesis that the value is higher for the resilient subsample.

Table 5. Subpopulation Resilience Capacity Indices, Scores, and Raw Outcome Variables, by Binary Scores (Final Wave)							
		C&B Binary		**TANGO/RIMA Binary**		**FCS**	
		Not resilient	Resilient	Not resilient	Resilient	Food insecure	Food secure
Female head	Niger	23.99	13.88	16.25	18.47	18.9	17.04
		p-value =0.000		*p-value* =0.116		*p-value* =0.202	
	Ethiopia	34.10	20.38	25.05	28.74	29.86	25.52
		p-value = 0.000		*p-value* = 0.012		*p-value* = 0.005	
Dependency ratio above sample median	Niger	52.68	48.38	50.63	50.99	51.32	50.30
		p-value =0.020		*p-value* =0.847		*p-value* =0.601	
	Ethiopia	56.64	60.22	59.59	57.49	59.80	57.92
		p-value = 0.047		*p-value* = 0.206		*p-value* = 0.292	
HH head has no formal education	Niger	89.87	66.54	74.03	76.21	87.31	73.33
		p-value =0.000		*p-value* =0.170		*p-value* =0.000	
	Ethiopia	74.31	54.85	63.11	65.94	67.39	63.19
		p-value = 0.000		*p-value* = 0.072		*p-value* = 0.012	
Num. of TLU owned below sample median	Niger	44.52	55.79	53.47	48.60	52.67	50.74
		p-value =0.000		*p-value* =0.008		*p-value* =0.456	
	Ethiopia	90.67	88.07	88.78	89.30	89.69	88.74
		p-value = 0.022		*p-value* = 0.618		*p-value* = 0.400	
HH has no access to ag. land	Niger	11.88	59.16	40.68	39.25	25.30	43.29
		p-value =0.000		*p-value* =0.432		*p-value* =0.000	
	Ethiopia	12.71	15.76	15.12	17.68	13.29	17.92
		p-value = 0.015		*p-value* = 0.036		*p-value* = 0.000	
HH did not receive remittances	Niger	55.74	49.39	54.13	53.13	47.62	54.42
		p-value =0.000		*p-value* =0.590		*p-value* =0.000	
	Ethiopia	88.43	79.68	84.16	83.33	86.38	82.54
		p-value = 0.000		*p-value* = 0.532		*p-value* = 0.003	
Members of community migrate seasonally	Niger	94.93	83.91	87.87	90.77	92.27	88.42
		p-value =0.000		*p-value* =0.010		*p-value* =0.000	
	Ethiopia	84.73	69.56	74.34	74.88	75.48	74.20
		p-value = 0.000		*p-value* = 0.706		*p-value* = 0.403	

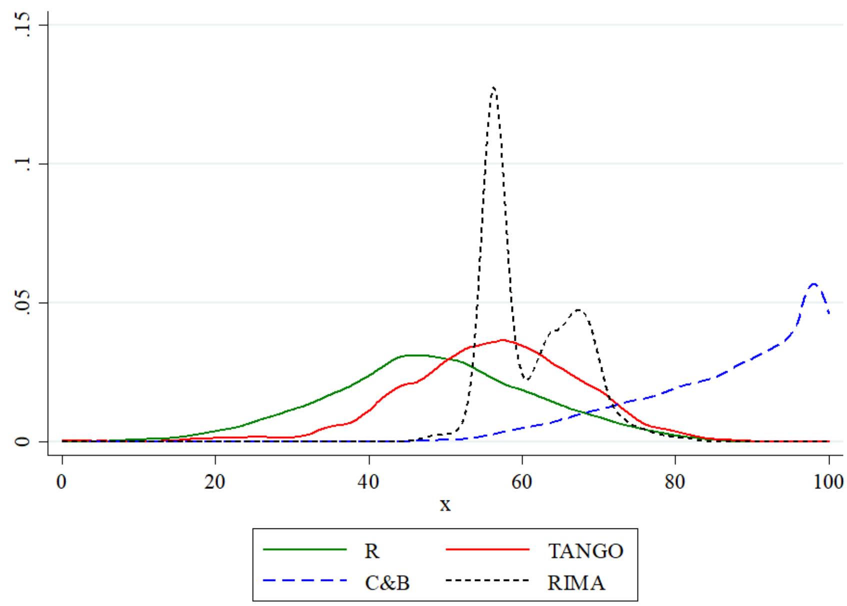
Note: Summary statistics in each country refer to last survey wave. All values correspond to the percentage of the population that falls into the specific demographic category, classified by resilience and food security status. The p-values refer to the t-test of equality of proportions among resilient/not resilient and food secure/insecure groups based on that targetable indicator.

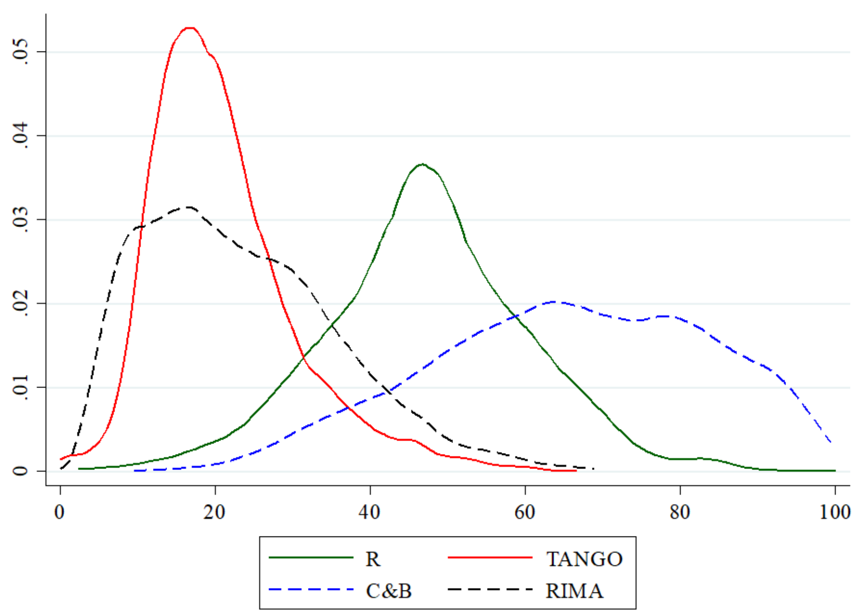
Table 6. Out-of-sample well-being predictive accuracy of resilience measures (RMSE)

	Niger		Ethiopia	
Cross-Sectional Tests (Regression of FCS on out-of-sample RCI / RS)				
	RMSE	*Ratio to sample mean*	*RMSE*	*Ratio to sample mean*
RCI TANGO	23.52	0.36	16.42	0.39
RCI RIMA	22.33	0.34	16.41	0.39
C&B RS	22.03	0.34	15.38	0.37
Forward-looking Tests (Regression of Final Wave outcome on prior wave RCI / RS, or outcome)				
RCI TANGO	23.84	0.36	17.02	0.40
RCI RIMA	22.54	0.34	17.30	0.41
R	24.29	0.37	16.91	0.40
C&B RS	22.44	0.34	15.76	0.37
FCS	22.3	0.34	15.92	0.38
Sample Average (final wave)	65.49		42.04	

Notes: In each case, we simulate ten times to obtain the average RMSE across simulations. For Niger, the differences between all combinations are statistically significant at the 1% level in both the cross-sectional and forward-looking tests. For Ethiopia, the differences between all combinations are statistically significant at the 1% level, except for the cross-sectional difference between the TANGO and RIMA RCI, which is just significant at the 5% level, and the forward-looking regression of FCS on prior period R, which is not significant at any conventional significance level.

Figure 1. RCI, RS, and $\tilde{R}$ Density Functions, Niger (top) and Ethiopia (bottom)
Underlying outcome = FCS





Note: Kolmogorov-Smirnov tests reject the null of equal distributions for every combination of measures in both countries at the 1% level.

Figure 2. Spearman's Rank Correlation Coefficients (Final Wave)

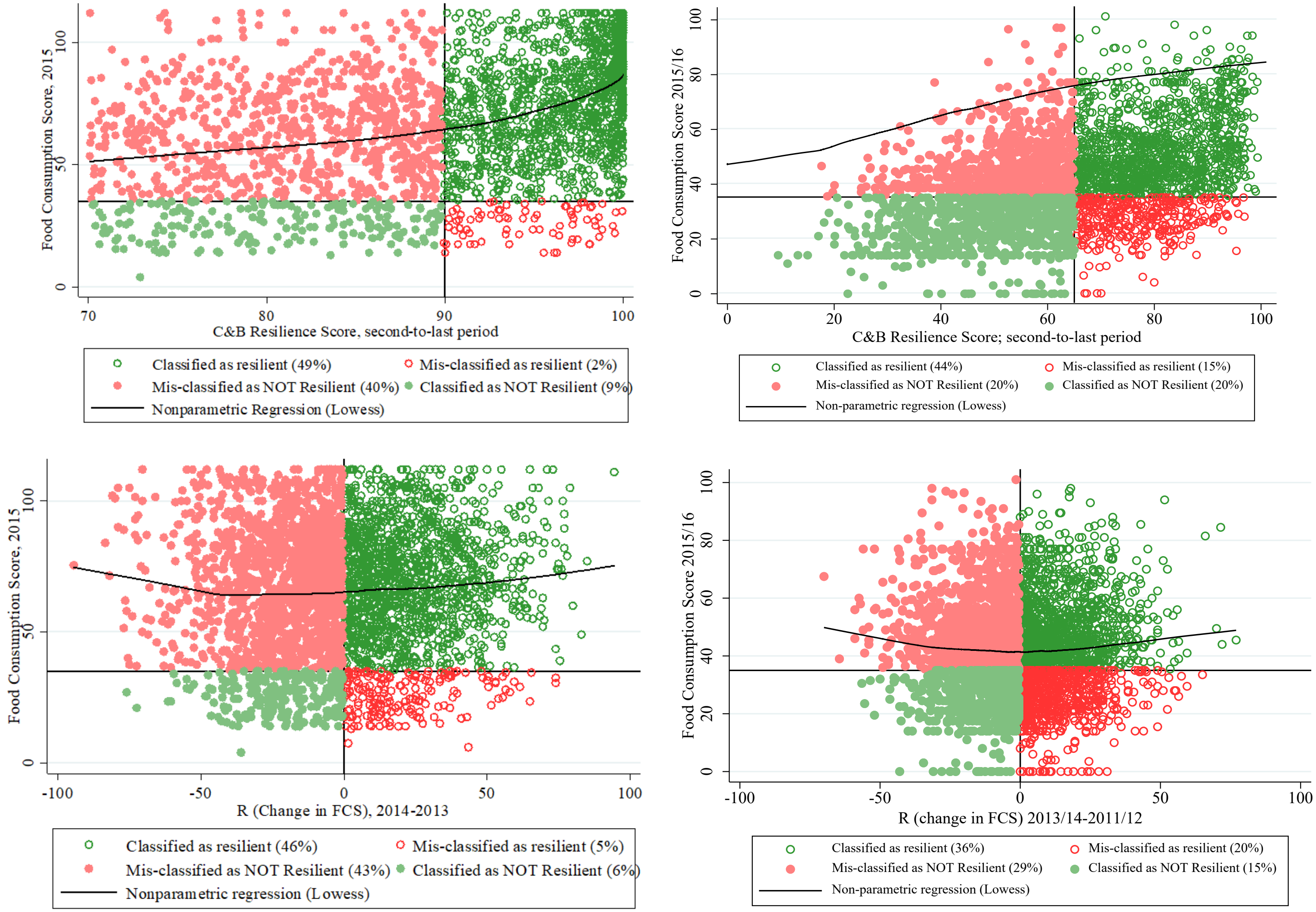
		TANGO	RIMA	$\tilde{R}$	C&B RS
RIMA	*Niger*	*0.519**			
	Ethiopia	0.702*			
$\tilde{R}$	*Niger*	*-0.064**	*-0.017*		
	Ethiopia	0.041*	0.024		
C&B RS	*Niger*	*0.465**	*0.697**	*0.068**	
	Ethiopia	0.393*	0.441*	-0.373*	
FCS	*Niger*	*0.250**	*0.404**	*0.560**	*0.482**
	Ethiopia	0.221*	0.229*	0.515*	0.412*

* Indicates statistical significance, all at the 1% level (others not significant, even at 10% level).

Figure 3. Resilience Ranking - Correspondence of the "Bottom 20%" - Final Wave

			ALSO in bottom 20% of ranking by…			
			TANGO	RIMA	$\tilde{R}$	C&B RS
In Bottom 20% of ranking by …	RIMA	*Niger*	*38%*			
		Ethiopia	60%			
	$\tilde{R}$	*Niger*	*14%*	*21%*		
		Ethiopia	19%	17%		
	C&B RS	*Niger*	*33%*	*47%*	*22%*	
		Ethiopia	40%	42%	6%	
	FCS	*Niger*	*23%*	*34%*	*50%*	*37%*
		Ethiopia	28%	30%	46%	40%

Figure 4. Prior period Resilience and "future" FCS – C&B (top panel) and R (bottom panel), Niger (left) and Ethiopia (right)



Note: Type 1 ("Mis-classified as resilient") and Type 2 ("Mis-classified as NOT resilient") errors indicated in legend. These are calculated by comparing prior period resilience classification with the subsequent period outcome; i.e., Type 1 = 1 if the prior period classified the household as resilient, while in the subsequent period the household did *not* in fact fall above the critical threshold.

Appendix to *Caveat utilitor*:
A comparative assessment of resilience measurement approaches

September 2021 revision

LSMS Data: Maps, Sampling & Descriptive Statistics

Figure A1. Map of EAs in Ethiopia, coded by Livelihood Zone (excluding Urban)

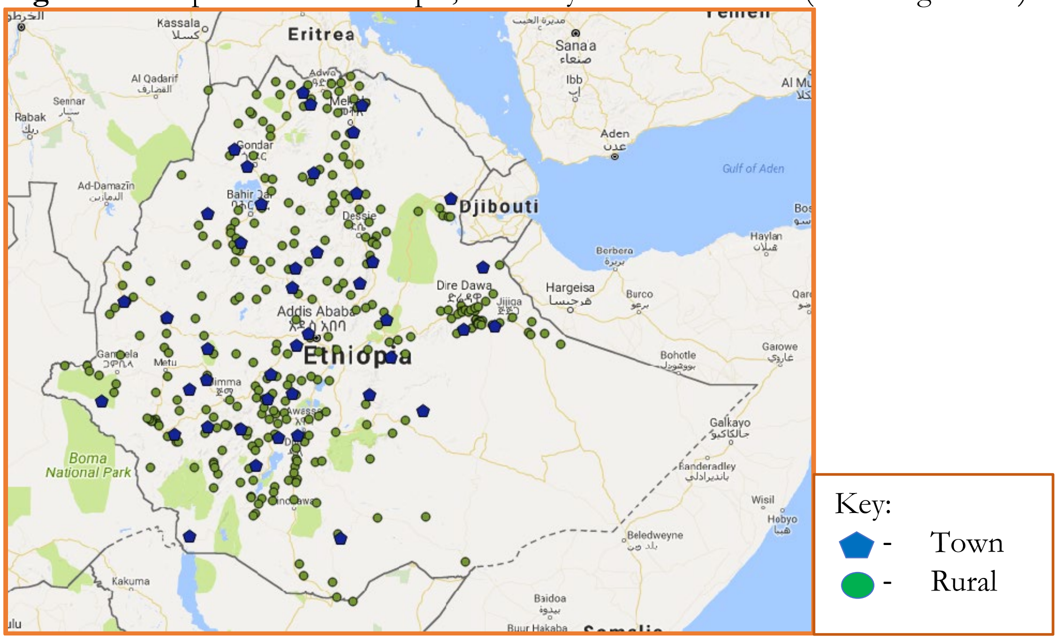


Figure A2. Map of ZDs in Niger, coded by Livelihood Zone

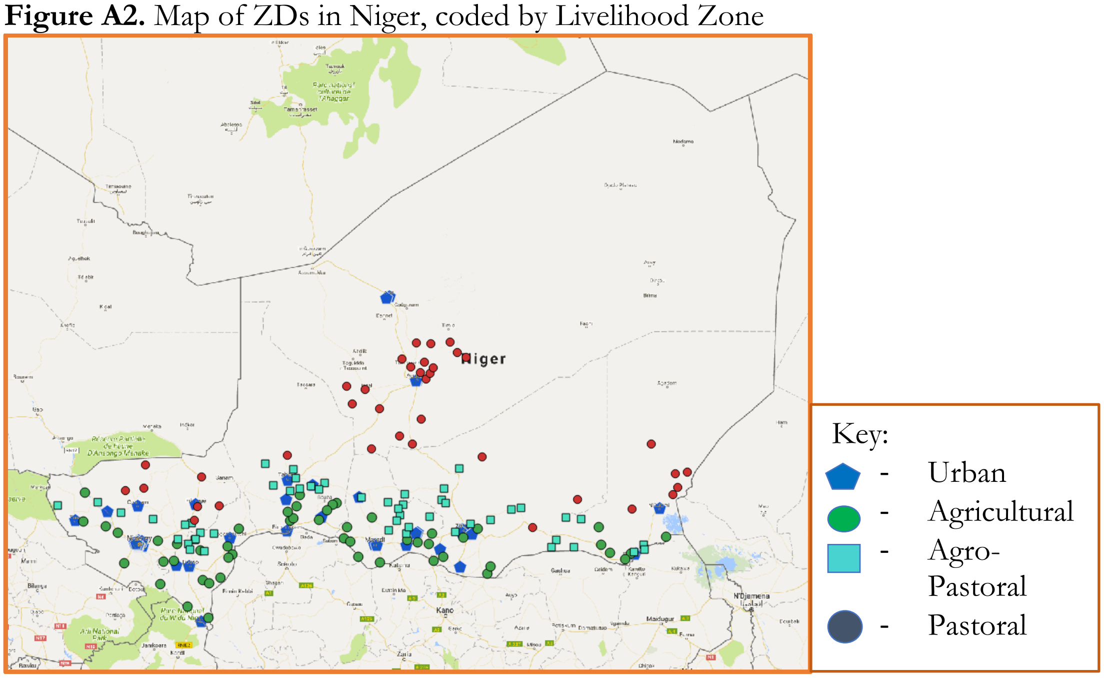


Table A1: Sampling Design (Wave 1)

Ethiopia (Excluding Urban and Addis)

Region	Pop Share	Rural EAs	Rural HHs	Town EA	Town HH	Total EA	Total HH
National	100%	290	3466	43	503	333	3969
Tigray	6.6%	30	360	4	48	34	408
SNNP	20.8%	74	885	10	119	84	1004
Amhara	26.6%	61	726	10	115	71	841
Oromiya	37.6%	55	658	10	113	65	771
Other Regions*							
Afar	1.7%	10	120	2	24	12	144
Somali	4.5%	20	237	3	36	23	273
Benis.- Gumuz	1.0%	10	120	1	12	11	132
Gambela	0.4%	10	120	1	12	11	132
Harari	0.3%	10	120	1	12	11	132
Dire Dawa	0.5%	10	120	1	12	11	132

Niger

Region	Pop Share (2001)	Ag ZD	Ag HH	Ag-Past ZD	Ag-Past HH	Pastoral ZD	Pastoral HH	Urban ZD	Urban HH	Total	
National	100%	51	918	51	918	37	666	131	1572	270	4074
Agadez	10.0%	0	0	0	0	20	360	7	84	27	444
Diffa	8.5%	6	108	8	144	6	108	3	36	23	396
Dosso	9.6%	13	234	7	126	0	0	6	72	26	432
Maradi	11.1%	9	162	10	180	0	0	11	132	30	474
Tahoua	10.7%	9	162	9	162	1	18	10	120	29	462
Tillabery	9.3%	7	126	7	126	7	126	4	48	25	426
Zinder	11.9%	7	126	10	180	3	54	12	144	32	504
Niamey	28.9%	0	0	0	0	0	0	78	936	78	936

* Ethiopia sampling is representative for the four major regions and all smaller regions combined.

Sources:

Ethiopia: World Bank (2017). Ethiopia Socioeconomic Survey, Wave Three (2015/2016), Basic Information Document. February 2017.

Niger: INS Niger (2011), 2011 National Survey on Household Living Conditions and Agriculture, Basic Information Document. October 2013.

Table A2. Household Characteristics (Wave 1)

	NIGER (2011)		Ethiopia (2012)	
	Mean	SD	Mean	SD
Head Characteristics				
Age of HH Head (years)	45.45	14.54	45.71	15.49
Female household head	0.14	0.34	0.26	0.44
Head has primary school education	0.12	0.32	0.22	0.41
Head has secondary education or higher	0.14	0.35	0.09	0.29
HH head in ag labor	0	0.05	-	-
HH head in non-ag labor	0.22	0.42	-	-
HH head self employed (trade, arts)	0.14	0.35	-	-
HH head does other employment	0.01	0.07	-	-
HH Demographics				
HH Size	6.33	3.42	4.93	2.38
% members male <=15 yrs	0.24	0.19	0.15	0.16
% members male 16-65 yrs	0.23	0.18	0.24	0.2
% members female 16-65 yrs	0.26	0.15	0.28	0.2
% members male >65 yrs	0.02	0.07	0.03	0.1
% members female >65 yrs	0.01	0.08	0.03	0.15
Observations	3,968		3,912	

Underlying Variables used in RCIs and RS

Table A3: Underlying RHS Variables for all models, in Niger (Wave 4 / Post-Harvest Wave 2)

			TANGO Model		RIMA Pillar	C&B
	Mean	SD	Capacity Index	Sub-Index		
Received any private transfer (12 mos)	0.47	0.5	Absorptive	Bonding Social Capital		Y
Number of private transfers received	0.81	1.12	Absorptive	Bonding Social Capital		
Gave any private transfer (12 mos)	0.33	0.47	Absorptive	Bonding Social Capital		
Number of private transfers given	0.5	0.85	Absorptive	Bonding Social Capital		
Durable Asset Index	0	2.82	Absorptive, Adaptive	Asset Ownership	ASS	Y
HH has access to agricultural land	0.61	0.49	Absorptive, Adaptive	Asset Ownership	ASS	
Value of Land Owned (estimate, Birr)	122900	370655	Absorptive, Adaptive	Asset Ownership	ASS	
Livestock Holdings (TLU)	1.98	8.48	Absorptive, Adaptive	Asset Ownership	ASS	
Number of agricultural cooperative in this village	1.32	0.79	Absorptive	Safety Nets	SSN	Y
Number of transfers received from within village	0.2	0.58	Absorptive	Safety Nets	SSN	
Value of transfers received from within village	9721	63542	Absorptive	Safety Nets	SSN	
Distance to Cereal Bank (km)	11.89	31.23	Absorptive	Disaster Preparedness		Y
Distance to ambulance service, km	22.15	31.88	Absorptive	Disaster Preparedness		
Distance to hospital, km	34	40.88	Absorptive	Disaster Preparedness		
Members of community migrate seasonally	0.89	0.31	Absorptive	Disaster Preparedness		
Any hh member traveled outside of village	0.59	0.49	Adaptive, Transformative	Bridging Social Capital		Y
Total hh income from remittances	62861	261117	Adaptive, Transformative	Bridging Social Capital		

Number of agricultural cooperative in this village	1.32	0.79	Adaptive, Transformative	Bridging Social Capital		
Any HH member in government	0.07	0.25	Adaptive, Transformative	Linking Social Capital		Y
Any HH member in traditional or religious role	0.01	0.09	Adaptive, Transformative	Linking Social Capital		
Believes will have success in life (male in hh)	0.11	0.31	Adaptive	Aspirations		Y
Believes will have success in life (female in hh)	0.08	0.28	Adaptive	Aspirations		
Any hh member traveled outside of village	0.59	0.49	Adaptive	Aspirations		
Woman in hh traveled outside of village	0.36	0.48	Adaptive	Aspirations		
Believes she could move if work available	0.19	0.39	Adaptive	Aspirations		
Livelihood Diversity (# of activities)	**1.24**	**0.59**	Adaptive	Livelihood Diversity	AC	Y
No. of hh members literate	1.87	2.2	Adaptive	Human Capital	AC	Y
No. of hh members with primary education	0.44	0.79	Adaptive	Human Capital	AC	
No. of hh members with secondary ed. or higher	7.02	3.7	Adaptive	Human Capital	AC	
No. of hh members sick, past two weeks	1.57	1.96	Adaptive	Human Capital	AC	
No. of hh members with long-term illness	0.1	0.34	Adaptive	Human Capital	AC	
Any hh member traveled outside of village	0.59	0.49	Adaptive	Access to Information	AC	Y
Woman in hh traveled outside of village	0.36	0.48	Adaptive	Access to Information	AC	
Member of hh owns mobile phone	0.71	0.45	Adaptive	Access to Information	AC	
Member of hh made call in past month	0.3	0.46	Adaptive	Access to Information	AC	
Member of hh used internet in past month	0.1	0.3	Adaptive	Access to Information	AC	
Distance to permanent market (km)	27.02	39.86	Transformative	Market Access	ABS	Y
Distance to periodic market (km)	10.18	18.16	Transformative	Market Access	ABS	Y
Number of key services within village (of 13)	4.2	2.56	Transformative	Access to Services	ABS	Y
Woman in hh traveled outside of village	0.36	0.48	Transformative	Women's Empowerment		Y
Believes she could move if work available	0.19	0.39	Transformative	Women's Empowerment		
Observations	3621					

Table A4: Underlying RHS Variables for all models in Ethiopia (Wave 3)

	Mean	SD	Tango Model: Capacity Index	Tango Model: Sub-Index	RIMA Pillar	C&B
Probability of borrowing from a close circle	0.15	0.18	Absorptive	Bonding Social Capital		
% of hh that received informal transfers	0.17	0.19	Absorptive	Bonding Social Capital		Y
Value of transfers received (hh)	93.21	609.85	Absorptive	Bonding Social Capital		
Durable Asset Index	0	2.09	Absorptive, Adaptive	Asset Ownership	ASS	
HH has access to agricultural land	0.8	0.4	Absorptive, Adaptive	Asset Ownership	ASS	Y
Productive Asset Index	0	1.29	Absorptive, Adaptive	Asset Ownership	ASS	
Livestock Holdings (TLU)	4.16	17.51	Absorptive, Adaptive	Asset Ownership	ASS	
Presence of cooperative in this village	0.18	0.39	Absorptive Capacity	Safety Nets	SSN	
% of hh that received informal transfers	0.17	0.19	Absorptive Capacity	Safety Nets	SSN	Y
Value of transfers received (com)	92.62	242.76	Absorptive Capacity	Safety Nets	SSN	
% of hh that received assistance	0.19	0.3	Absorptive Capacity	Safety Nets	SSN	
Distance to health center, km	1.02	4.09	Absorptive Capacity	Disaster Preparedness		Y
Members of community migrate seasonally	0.73	0.44	Absorptive Capacity	Disaster Preparedness		
Total hh income from remittances	138.66	951.96	Adaptive, Transformative	Bridging Social Capital		
Presence of cooperative in this village	0.18	0.39	Adaptive, Transformative	Bridging Social Capital		Y
Share of males participate in coop	3.19	12.23	Adaptive, Transformative	Bridging Social Capital		
Share of females participate in coop	1.74	6.65	Adaptive, Transformative	Bridging Social Capital		Y
Any HH member in gov't/political party	0.06	0.24	Adaptive, Transformative	Linking Social Capital		
Livelihood Diversity (# of income sources)	**0.13**	**0.37**	Adaptive	Livelihood Diversity	AC	Y
No. of hh members literate	1.97	1.79	Adaptive	Human Capital		Y
Max educational attainment (all members)	4.92	4.06	Adaptive	Human Capital	AC	

No. of hh members with impairing disability	0.36	0.73	Adaptive	Human Capital	AC	
Access to Information (hh ownership of tv, phone, radio, cell-phone)	0.79	0.94	Adaptive	Access to Information	AC	Y
Distance to periodic market (km)	7.66	16.99	Transformative	Access to Services	ABS	Y
Distance to primary school (km)	1.17	10.41	Transformative	Access to Services	ABS	Y
Distance to secondary school (km)	12.99	24.52	Transformative	Access to Services	ABS	Y
Distance to pharmacy (km)	6.61	19.96	Transformative	Access to Services	ABS	Y
Distance to bus (km)	17.61	30.81	Transformative	Access to Services	ABS	Y
Distance to paved road (km)	43.12	58.83	Transformative	Access to Services	ABS	Y
Distance to extension agent (km)	1.04	7.95	Transformative	Access to Services	ABS	Y
% of hh with enterprises owned by a woman	0.39	0.38	Transformative	Women's Empowerment		Y
% of hh w/ loans woman decides over	0.21	0.3	Transformative	Women's Empowerment		
% of hh receive income controlled by woman	0.34	0.34	Transformative	Women's Empowerment		Y
% of hh own ag land & woman owns all/part	0.29	0.16	Transformative	Women's Empowerment		
Needs for which community asks government	5.42	2.8	Transformative	Quality of governance		Y
Needs for which gov't organizes community	5.56	2.78	Transformative	Quality of governance		
Mean level of need addressed	3.59	1.19	Transformative	Quality of governance		Y
Observations	3699					

RCI Factor Loadings

Table A5. TANGO factor loadings for all indices and components, Niger

NIGER		
Index, sub-Index, and Indicator	**Tango FA Loading^**	**Regular FA Loading**
ABSORPTIVE CAPACITY	0.344	0.358
Bonding Social Capital	0.32	-0.249
Received any private transfer (12 mos)	0.256	0.256
Number of private transfers received	0.269	0.269
Gave any private transfer (12 mos)	0.357	0.357
Number of private transfers given	0.369	0.369
Asset Ownership	-0.136^	0.27
Durable Asset Index	-0.244^	-0.29
HH has access to agricultural land	0.417	0.403
Value of Land Owned (estimate, Birr)	0.258	0.207
Livestock Holdings (TLU)	-0.003^	-0.002
Safety Nets	0.266^	-0.195
Number of agricultural cooperative in this village	0.042	0.042
Number of transfers received from within village	0.382	0.382
Value of transfers received from within village	0.379	0.379
Disater Preparedness	-0.212	0.279
Distance to Cereal Bank (km)	0.137	0.139
Distance to ambulance service, km	0.378	0.374
Distance to hospital, km	0.471	0.474
Members of community migrate seasonally	0.058	0.057
ADAPTIVE CAPACITY	0.174	-0.161
Bridging Social Capital	0.067	0.031
Any hh member traveled outside of village	0.143	0.14
Total hh income from remittances	0.137	0.117
Number of agricultural cooperative in this village	-0.057^	-0.065
Linking Social Capital	0.019	0.036
Any HH member in government	0.107	0.056
Any HH member in traditional or religious role	0.107	0.056
Asset Ownership (*same as above*)	0.463^	0.359
Aspirations	0.062	0.062
Believes will have success in life (male in hh)	0.107	0.107
Believes will have success in life (female in hh)	0.417	0.417
Any hh member traveled outside of village	0.416	0.416

Woman in hh traveled outside of village	0.093	0.093
Believes she could move if work available	0.029	0.028
Livelihood Diversity (# of activities)	0.009	-0.029
Human Capital	0.034^	0.036
No. of hh members literate	0.356^	0.356
No. of hh members with primary education	0.249^	0.249
No. of hh members with secondary education or higher	0.21^	0.21
No. of hh members sick, past two weeks	0.23^	0.23
No. of hh members with long-term illness	0.107^	0.107
Access to Information	0.458	0.585
Any hh member traveled outside of village	0.42	0.421
Woman in hh traveled outside of village	0.424	0.422
Member of hh owns mobile phone	0.094	0.091
Member of hh made call in past month	0.034^	0.032
Member of hh used internet in past month	0.071^	0.077
TRANFORMATIVE CAPACITY	-0.437	0.431
Bridging Social Capital (*same as above*)	-0.109	-0.077
Linking Social Capital (same as above)	-0.072	-0.071
Market Access		
Distance to permanent market (km)	0.39	0.39
Distance to periodic market (km)	0.265	0.279
Access to Services		
Number of key services within village (of 13)	-0.252	-0.258
Women's Empowerment	-0.12	-0.1
Woman in hh traveled outside of village	0.159	0.159
Believes she could move if work available	0.159	0.159
Observations	3621	15,180

^ NOTE: For Version 1 (Tango Method), factor loadings are calculated on baseline data and applied in the index to endline data, where variables are available in both periods. Components are dropped from index if the sign on the factor loading is in the "wrong" direction (counter to expectation), and/or if the KMO is < 0.5. Those dropped are indicated with a "^".The regular FA loadings are computed using multiple survey rounds and do not drop any components, per more standard statistical practice.

Table A6. TANGO factor loadings for all indices and components, Ethiopia		
ETHIOPIA		
Index, sub-Index, and Indicator	**Tango FA Loading^**	**Regular FA Loading**
ABSORPTIVE CAPACITY	**0.254**	**0.236**
Bonding Social Capital	0.824	0.823
Probability of borrowing from a close circle	0.197	0.295
% of hh that received informal transfers	0.347	0.414
Value of transfers received (hh)	0.255	0.266
Asset Ownership	-0.165^	-0.165
Durable Asset Index	-0.184^	-0.214
HH has access to agricultural land	0.530	0.512
Productive Asset Index	0.534	0.028
Livestock Holdings (TLU)	0.033	0.567
Safety Nets	0.824	0.828
Presence of cooperative in this village	0.246	0.107
% of hh that received informal transfers	0.562	0.615
Value of transfers received (com)	0.53	0.561
% of hh that received assistance	0.209	0.242
Disater Preparedness	-0.020^	-0.020
Distance to health center, km	0.097	0.055
Members of community migrate seasonally	-0.097	0.055
ADAPTIVE CAPACITY	**0.534**	**0.514**
Bridging Social Capital	0.214	0.225
Total hh income from remittances	0.075	0.049
Presence of cooperative in this village	0.727	0.693
Share of males participate in coop	0.742	0.733
Share of females participate in coop	0.696	0.733
Linking Social Capital		
Any HH member in government/political party	0.443	0.512
Asset Ownership (*same as above*)	-0.183^	-0.321
Livelihood Diversity (# of activities)	0.083	0.083
Human Capital	0.606	0.563
No. of hh members literate	0.689	0.689
Maximum level of ed. Attainment (all hh members)	0.691	0.690
No. of hh members with impairing disability	-0.04	-0.032
Access to Information	0.625	0.627
TRANFORMATIVE CAPACITY	**0.561**	**0.536**
Bridging Social Capital (*same as above*)	0.491	0.543

Linking Social Capital (same as above)	0.16	0.157
Market Access		
Distance to permanent market (km)	-0.126	-0.091
Access to Services	0.484	0.316
Number of key services within village (of 13)		
Women's Empowerment	0.330	0.490
% of hh with enterprises where it is owned by a woman	0.377	0.287
% of hh with loans where woman decides over it	0.314	0.346
% of hh that receive income where a woman decides over it	0.405	0.296
% of hh that own ag land where woman owns all/part	0.459	0.282
Share of females participate in coop	0.118	0.207
Quality of governance	0.226	0.140
Needs for which community asks government	0.805	0.832
Needs for which government organizes community	0.828	0.833
Mean level of need adressed	0.357	0.072
Observations	3969	11,262

^ NOTE: For Version 1 (Tango Method), factor loadings are calculated on baseline data and applied in the index to endline data, where variables are available in both periods. Components are dropped from index if the sign on the factor loading is in the "wrong" direction (counter to expectation), and/or if the KMO is < 0.5. Those dropped are indicated with a "^".The regular FA loadings are computed using multiple survey rounds and do not drop any components, per more standard statistical practice.

Table A7. RIMA factor loadings for all indices and components, Niger	
Pillar and Underlying Indicator	*Loading*
ACCESS TO BASIC SERVICES (ABS)	
Number of basic services in village	0.45
Distance to permanent market (km, inverse)	0.36
Distance to periodic market (km, inverse)	0.17
ASSETS (ASS)	
Durable Asset Index	-0.28
HH has access to agricultural land	0.41
Value of land owned	0.22
Livestock Holdings (TLU)	0
SOCIAL SAFETY NETS (SSN)	
Presence of cooperative in this village	0.06
Percentage of hhs that received transfer	0.4
Value of transfer received (comm)	0.39
ADAPTIVE CAPACITY (AC)	
Livelihood Diversity (# of activities)	0.1
No. of hh members literate	0.34
No. of hh members with primary education	0.12
No. of hh members with secondary education or higher	0.1
No. of hh members sick, past two weeks	0.13
No. of hh members with long-term illness	0.04
Any hh member traveled outside of village	0.21
Woman in hh traveled outside of village	0.23
Member of hh owns mobile phone	0.16
Member of hh made call in past month	0.06
Member of hh used internet in past month	0.11
Observations	15180

Table A7. RIMA factor loadings for all indices and components, Ethiopia	
Pillar and Underlying Indicator	*Loading*
ACCESS TO BASIC SERVICES (ABS)	
Inverse distance to periodic market (km)	0.354
Inverse distance to primary school (km)	0.000
Inverse distance to secondary school (km)	0.543
Inverse distance to pharmacy (km)	0.389
Inverse distance to health center (km)	-0.027
Inverse distance to bus (km)	0.584
Inverse distance to paved road (km)	0.339
Inverse distance to extension agent (km)	-0.175
ASSETS (ASS)	
Durable Asset Index	-0.215
HH has access to agricultural land	0.567
Productive Asset Index	0.512
Livestock Holdings (TLU)	0.028
SOCIAL SAFETY NETS (SSN)	
Presence of cooperative in this village	0.004
% of hh that received informal transfers	0.434
Value of transfers received (com)	0.239
% of hh that received assistance	0.352
ADAPTIVE CAPACITY (AC)	
Maximum level of ed. Attainment (all hh members)	0.601
No. of hh members with impairing disability	-0.082
Livelihood Diversity (# of activities)	0.076
Access to Information (hh ownership of tv, phone, radio, cell-phone)	0.613
Observations	11262

Regression results for Cissé and Barrett method

Table A8. Cissé and Barrett RS Construction, Estimating Conditional Mean and Variance, Niger

	Consumption Expenditure		FCS	
	Mean	Variance	Mean	Variance
Lagged Outcome	0.272***	400.389***	0.215	-0.235
	(0.075)	(104.050)	(0.177)	(0.815)
Lagged Outcome - Squared	0		0	
	(0.000)		(0.003)	
Lagged Outcome - Cubed	0		0	
	(0.000)		(0.000)	
Drought (NDVI, z-score if <-1.5)	-287.553	-1260580	-6.498***	104.054***
	(220.267)	(907176.199)	(1.069)	(27.325)
Age of HH Head (years)	-8.139*	-5345.939	0.025	0.91
	(4.866)	(20251.728)	(0.031)	(0.579)
HH Size	-196.859***	-954899.5***	0.118	-3.345
	(14.080)	(207150.220)	(0.167)	(6.252)
Female household head	-49.243	225832.225	-3.205**	17.7
	(169.135)	(346871.232)	(1.245)	(15.650)
Head has primary school education	-22.824	1061148.72	-1.781	15.691
	(163.921)	(821677.783)	(1.141)	(29.007)
Head has secondary education or higher	231.735**	1904616.38**	-1.137	1.595
	(117.320)	(762556.967)	(0.975)	(24.297)
HH head in ag labor	433.924	-2955728.66**	-0.339	-109.989
	(784.073)	(1328724.830)	(2.493)	(84.316)
HH head in non-ag labor	196.128	-442434.7	2.895***	9.907
	(134.617)	(570282.498)	(1.052)	(15.231)
HH head self employed (trade, arts)	249.019**	-11863.553	2.437**	51.147**
	(113.661)	(334295.440)	(1.119)	(24.726)
HH head does other employment	-1191.085	-2456098.9	-14.009**	77.453
	(895.530)	(2392363.297)	(5.604)	(72.709)
% members male <=15 yrs	-150.325	-1105622.8	0.671	-1.376
	(223.836)	(962212.031)	(2.162)	(31.678)
% members male 16-65 yrs	1477.834**	2908000***	-0.112	79.847***
	(623.329)	(1056541.765)	(2.560)	(14.199)
% members female 16-65 yrs	2293.950***	1894856.06	1.664	-23.834
	(351.399)	(1216043.793)	(3.173)	(84.794)
% members male >65 yrs	518.592	1452733.68	-10.838**	12.469

	(855.810)	(1653102.456)	(5.242)	(56.897)
% members female >65 yrs	2126.071***	1875765.32	-4.842	-106.542
	(553.294)	(1441180.860)	(4.367)	(94.647)
Bonding Social Capital	12.390***	21774.424*	0.100***	-1.805***
	(3.822)	(11999.631)	(0.029)	(0.652)
Asset Ownership Index	-137.821***	-32910.958	-0.919***	11.238***
	(18.398)	(131566.755)	(0.111)	(2.910)
Access to Informal Safety Nets	-15.551*	-10988.812	-0.039	3.102***
	(8.998)	(39482.874)	(0.084)	(0.761)
Availability of Disaster Prep	8.550*	18473.729	0.03	-0.138
	(4.735)	(26657.462)	(0.064)	(1.086)
Bridging Social Capital	35.366	60563.878	0.04	1.908
	(24.411)	(78918.057)	(0.208)	(1.917)
Linking Social Capital	37.151***	66194.845	0.150**	0.295
	(8.271)	(41757.651)	(0.066)	(0.702)
Aspirations	-26.502***	-36240.67	-0.138**	-0.825
	(6.545)	(41330.195)	(0.059)	(1.207)
Number of livelihoods (hh-level)	-72.89	-1011950***	1.909***	-36.994***
	(55.684)	(207794.051)	(0.599)	(11.489)
Human Capital	14.070***	94535.660**	0.089**	0.316
	(3.843)	(42902.797)	(0.037)	(1.171)
Access to Information	21.378***	29511.974*	0.119***	0.412
	(4.519)	(16521.433)	(0.040)	(0.898)
Distance to permanent market (km)	-2.238	-20720.9***	0	-0.429
	(1.563)	(7522.224)	(0.013)	(0.424)
Distance to periodic market (km)	5.43	14636.02	-0.008	-0.093
	(5.072)	(17110.085)	(0.042)	(0.361)
Number of services within village	78.364***	139565.231	0.477*	13.858***
	(22.663)	(138075.670)	(0.261)	(4.323)
Women's Empowerment	2.819	10954.467	0.026	0.226
	(2.658)	(17028.017)	(0.023)	(0.428)
Constant	7658.789***		67.526***	
	(722.933)		(6.258)	
No. of Households	9427	9427	9391	9391
R-Squared	0.419		0.184	

*, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

Table A9. Cissé and Barrett RS Construction, Estimating Conditional Mean and Variance, Ethiopia

	Consumption Expenditure		FCS	
	Mean	Variance	Mean	Variance
Lagged Outcome	0.420***	0.000147	0.320*	0.0137
	(0.0982)	(9.90e-05)	(0.179)	(0.0130)
Lagged Outcome - Squared	-7.34e-06	0	-0.00110	-0.000272
	(1.65e-05)	(1.28e-08)	(0.00419)	(0.000269)
Lagged Outcome - Cubed	-5.87e-10	0	1.12e-05	1.97e-06
	(7.60e-10)	(0)	(2.86e-05)	(1.69e-06)
Drought (NDVI, z-score if <-1.5)	63.93	-0.0628	-0.251	-0.0729
	(65.83)	(0.0834)	(1.198)	(0.0527)
Age of HH Head (years)	1.534	-0.00178	-0.0143	-0.00144
	(3.262)	(0.00413)	(0.0275)	(0.00190)
HH Size	-160.3***	-0.0689**	0.507**	0.0298*
	(20.65)	(0.0276)	(0.212)	(0.0155)
Male household head	269.5***	0.0711	-0.114	-0.0713
	(86.93)	(0.107)	(0.800)	(0.0575)
Head has primary school education	139.0**	-0.106	-0.0687	0.0140
	(70.00)	(0.0856)	(0.847)	(0.0575)
Head has secondary education	528.4***	0.172	1.053	0.0789
	(115.7)	(0.161)	(1.299)	(0.0805)
Head has college education	1,301***	0.551	2.466	0.185
	(332.5)	(0.394)	(2.179)	(0.148)
% members male <=4 yrs	-211.9	0.293	0.277	0.162
	(425.2)	(0.446)	(3.359)	(0.223)
% members male 5-15 yrs	-608.6	0.0266	-2.328	0.00968
	(426.3)	(0.414)	(2.967)	(0.215)
% members female 5-15 yrs	-229.5	0.320	-1.050	0.170
	(426.5)	(0.353)	(3.000)	(0.208)
% members male 16-65 yrs	8.978	0.759**	-2.650	0.238
	(469.9)	(0.377)	(3.290)	(0.212)
% members female 16-65 yrs	209.1	0.574	-3.577	0.0828
	(428.1)	(0.478)	(3.048)	(0.216)
% members male >65 yrs	185.4	0.953*	-0.390	0.117
	(541.4)	(0.558)	(3.932)	(0.334)
% members female >65 yrs	-298.3	0.379	-1.825	0.168
	(463.4)	(0.487)	(3.457)	(0.248)

Wealth Quintile	178.7***	0.150***	1.162***	-0.00804
	(38.25)	(0.0381)	(0.297)	(0.0219)
Bonding Social Capital	-3.882	-0.0666	0.923	0.103
	(148.2)	(0.132)	(1.344)	(0.101)
Bridging Social Capital	-42.73	-0.0298	-0.546	0.00951
	(59.26)	(0.0425)	(0.558)	(0.0344)
Linking Social Capital	450.7	-0.354	3.306*	0.0208
	(279.9)	(0.325)	(1.719)	(0.108)
Asset Ownership Index	206.2*	-0.0252	1.427*	-0.0791
	(116.1)	(0.0868)	(0.734)	(0.0520)
Number of livelihoods (hh-level)	19.40	-0.0631	-2.005**	-0.122*
	(84.90)	(0.102)	(0.783)	(0.0630)
Number of key services within village	21.48	-0.0528	0.680	0.0236
	(54.97)	(0.0397)	(0.425)	(0.0297)
Human Capital	-128.4*	-0.140*	-0.704	-0.0841*
	(68.29)	(0.0751)	(0.656)	(0.0485)
Access to Information	350.0***	0.122**	1.229***	0.0620**
	(55.49)	(0.0519)	(0.420)	(0.0280)
Women's Empowerment	-152.6	-0.0562	-0.283	0.104
	(121.0)	(0.0837)	(1.106)	(0.0671)
Quality of governance	-74.55	-0.0451	0.267	-0.0133
	(63.32)	(0.0459)	(0.589)	(0.0371)
Access to Informal Safety Nets	150.5	0.233	3.379**	-0.139
	(159.0)	(0.144)	(1.339)	(0.0980)
Availability of Disaster Prep	-952.4	0.574	-23.96***	-0.183
	(731.6)	(0.539)	(5.779)	(0.374)
No. of Households	6,137	6,137	6,460	6,460
R-Squared	0.295		0.207	

*, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively

Equivalence Scales for Household Consumption Expenditure

Table A10. Conversion Factors for Household Equivalency

Age Range in completed years	Men	Women
0	0.30	0.30
1	0.46	0.46
2	0.54	0.54
3 - 4	0.62	0.62
5 - 6	0.74	0.70
7 - 9	0.84	0.72
10 - 11	0.88	0.78
12 - 13	0.96	0.84
14 - 15	1.06	0.86
16 - 17	1.14	0.86
18 - 29	1.04	0.80
30 - 59	1.00	0.82
60 or more	0.84	0.74

Results using real adult equivalent consumption expenditure (RAEC)

The main manuscript reports all results using just a standard measure of food security, the Food Consumption Score (FCS). Here we report results from all the same tests, now using consumption expenditure measures on which poverty estimates typically rely. These measures are deflated to real terms for comparison across space and time, and normalized by household size using the equivalence scales reported in Table A10.

The expenditures measure includes weekly or monthly recall of detailed expenditures, and annual recall of larger expenditures. In the LSMS data, food consumption expenditure is annualized from seven day recall data on food consumption collected in the post-harvest period. It is based on the reported expenditure on food purchases, as well as the value of reported quantities consumed from own production, gifts, and other sources. For non-food items, recall was 30 days or 12 months, also collected only in the post-harvest period, and includes various categories of non-food items and services, including stimulants such as khat (or in the case of Niger, cola nuts and tobacco). Prices are adjusted for spatial variation using a Fisher price index, using price data from both household purchases and the market price survey. Expenditure is adjusted for temporal variation in prices using the consumer price index (CPI), constructed from information on monthly inflation from tradingeconomics.com, and expressed in constant prices (base period February 2012). The result is a real adult equivalent consumption (RAEC) measure that is broadly comparable to a poverty line.

For Niger, food consumption expenditure is based on the self-reported value of consumption from all sources over the past seven days, and is collected in both the post-harvest and post-planting periods. The value of food consumption is adjusted for spatial variation in prices using a Fisher food price index, constructed from the price data available in each survey using the market price survey. Food expenditure is likewise adjusted for temporal variation in prices using the CPI (also constructed from information on monthly inflation from tradingeconomics.com). For non-food items, recall varied from seven days to 12 months, with shorter recall items (7 and 30 days) collected in the post-harvest period and longer-recall items (6 months and 12 months) collected in the post-planting period, and based on self-reported values. No extensive spatial correction can be made in Niger for non-food items, but the overall value is adjusted for temporal variation in prices using the CPI, and expressed in constant prices (base period August 2011). Note that in Niger—but not Ethiopia—seven day food consumption data were collected in both post-planting and post-harvest seasons, along with a subset of short-recall non-food consumption items. In order to leverage these multiple periods and facilitate our estimation, we use the weekly measure for Niger as opposed to the annualized measure.

Within each data set, a poverty line was constructed, based on a food basket of 2700 calories – the average number consumed by poor-to-middle income households – with a non-food poverty line then designated according to the non-food consumption expenditure of households close to that poverty line.[1] The food poverty line and poverty line are expressed as Birr (Ethiopia) or CFAF (Niger) per adult equivalent per year. We note that these measures do not compare directly to poverty lines such as the World Bank global extreme poverty line or those calculated by national governments, due to differences in what exactly is collected in expenditure modules, as well as potentially measurement differences associated with different recall and aggregation periods. These

[1] Specifically, we use the Ravallion method, which averages real non-food expenditure for those households whose food consumption is within x% of the food poverty line, for x $\subset$ [1,10].

differences have some implications for the interpretation of our results, which are noted below. But because the measures are uniform across methods within the data they cannot explain any of the differences we report across resilience measures and their performance and correspondence within a given data set.

Table A11 shows the RAEC across waves for both countries, along with the percentage classified as poor. Average weekly expenditure in Niger ranges between 5,357 and 5,768 FCFA, roughly equivalent to $12/week or $1.75/day, with about a quarter of the population classified as poor. In Ethiopia, real annual consumption expenditure declines over the period from 4612 to 3302 Birr, or about $0.74 to $0.53 in base year February 2012 prices, with first 51% and then 65% classified as poor.[2] There are clearly measurement differences in consumption expenditure (as noted above); it seems counterintuitive that the true poverty rate in Ethiopia is much or at all higher than in Niger nor that poverty rates rose over the period in Ethiopia. These discrepancies reflect differences in the content of the expenditure modules, possibly measurement differences due to varying recall periods and levels of aggregation, and problems with the Ethiopia deflator that are difficult to overcome given data constraints. But as we do not pool data across countries, this is irrelevant to the comparisons of resilience measures that are this paper's purpose. Nonetheless, they provide another reason we prioritize the FCS measures in the main body of the paper.

Table A12 reports the descriptive statistics on the resilience measures based on RAEC for each country, equivalent to Table 3 in the main paper.

Figure A3 then displays the distribution of each of the resilience measures when RAEC is the wellbeing variable of interest, comparable to Figure 1 in the main paper. Figures A4 then reports the rank correlation coefficients for the full sample, comparable to Figure 2 in the main paper. And Figure A5 shows the correspondence between households ranked in the lowest 20% by each measure., comparable to Figure 3

Table A13 is the RAEC equivalent to Table 4 in the main paper, offering a test of internal consistency between binary resilience classification (indicator) variables and the underlying continuous measures.

Figures A6 then shows the out-of-sample predictive relationship between the different resilience measures and RAEC in the final survey period, comparable to Figures 4 and 5 in the main paper.

Finally, Table A14 reports the out-of-sample predictive performance with respect to RAEC outcomes for the different resilience measures, comparable to Table 5 in the main paper.

As the tables and figures reveal, the core qualitative findings from the main paper, based on using the food security measure FCS as the wellbeing outcome of interest, all carry through when we instead use real expenditures as the wellbeing measure.

[2] This trend in poverty in Ethiopia runs contrary to most poverty estimates in the country and appears an artefact of the spatial and temporal deflator used. Since that is a scaling issue that does not vary across methods, however, the comparisons undertaken here are invariant to any corrections one might make to the deflator method. So we do not claim to be reflecting poverty rates accurately, especially not in the Ethiopia data, just studying correspondences and predictive performance that are invariant to the scaling.

Table A11. Consumption Expenditure and Poverty Status (across waves)

	NIGER				ETHIOPIA		
	Wave 1, 2012-2013		Wave 2, 2014-2015		Wave 1	Wave 2	Wave 3
	PP	PH	PP	PH			
Real per adult equivalent consumption expenditure^	5357.22	5430.82	5767.68	5564.81	4,612.22	3,916.50	3,302.94
	(4183.94)	(3700.83)	(4262.20)	(3811.94)	(2,741.37)	(2,373.81)	(2,138.57)
Poor	0.22	0.2	0.26	0.21	0.48	0.51	0.62
	(0.42)	(0.20)	(0.26)	(0.21)	(0.49)	(0.51)	(0.48)
Observations	3969	3969	3621	3621	3969	3776	3699

Notes: Standard Errors shown in parentheses

RAEC for Niger is weekly, expressed in F CFA, Base period August 2011 (446 CFA = 1 USD);

RAEC for Ethiopia is annual, expressed in Birr, Base period February 2012 (17 Birr = 1 USD)

Table A12. Resilience Indices (Final wave)

All measures scaled [0,100]	NIGER				ETHIOPIA			
	Mean	SD	Min	Max	Mean	SD	Min	Max
RCI, TANGO method	55.87	11.98	0	100	20.99	9.32	0	66.62
RCI, RIMA method	60.78	6.56	0	93.22	23.28	12.42	0	68.89
$\tilde{R}$ realized resilience	48.58	6.62	11.15	96.64	47.19	6.92	0	100
C&B resilience score	74.56	21.7	2.17	100	40.77	22.03	1.29	100
Number of observations	3621				3699			

The RIMA Index is constructed including both FCS and RAEC as outcome variables for the SEM model, as prescribed by FAO; while C&B and $\tilde{R}$ are constructed only using RAEC. For C&B, the threshold used for RAEC is the consumption poverty line, as indicated in Table A11.

Table A13. Resilience Capacity Indices, Scores, and raw Outcome Variables (Final Wave)

		C&B Binary			I		
		Not Resilient	Resilient	P-value	Not Resilient	Resilient	P-value
RCI - TANGO	*Niger*	N/A			*56.38*	*55.84*	*0.24*
	Ethiopia				23.24	23.32	0.87
RCI - RIMA	*Niger*				*61.33*	*60.95*	*0.13*
	Ethiopia				20.96	21.04	0.82
R	*Niger*				*43.56*	*52.66*	*0.00*
	Ethiopia				43.45	53.14	0.00
C&B RS (RAEC)	*Niger*	*54.75*	*90.76*	*0.00*	N/A		
	Ethiopia	22.89	59.83	0.00			
FCS	*Niger*	*57.21*	*72.44*	*0.00*	*60.47*	*69.67*	*0.00*
	Ethiopia	39.77	44.38	0.00	40.32	44.31	0.00
Observations	*Niger*	*1263*	*1895*		*1213*	*1854*	
	Ethiopia	1,577	1,479		2,107	1,592	

Table A14. Out-of-sample well-being predictive accuracy (RMSE)

	Niger	Ethiopia
Cross-Sectional Tests (Regression of FCS on out-of-sample RCI / RS)		
RCI TANGO	3801.87	2051.95
RCI RIMA	3423.17	1961.16
C&B RS	3381.45	1824.62
Panel Tests (Regression of Final Wave outcome on prior wave RCI / RS, or outcome)		
RCI TANGO	3764.97	2070.95
RCI RIMA	3444.30	2098.55
$\tilde{R}$	3790.10	2119.81
C&B RS	3374.57	1832.51
RAEC	2903.23	1915.9

Note: In the cross-sectional case, we simulate multiple times to obtain the average RMSE across simulations. We can then test the significance of difference between the estimated coefficients.
For Niger, the differences between all combinations are significant at the 1% level or smaller, with the exception of the difference between RS C&B and RCI RIMA, which is significant at the 5% level.
For Ethiopia, the differences between all combinations are significant at the 1% level or smaller.

In the panel case, all differences statistically significant at the 1 % level other than the regression of FCS on prior period $\tilde{R}$ score, which is not significant.

Figure A3. RCI, RS, and R Density Functions, Niger (top) and Ethiopia (bottom), Underlying outcome = RAEC

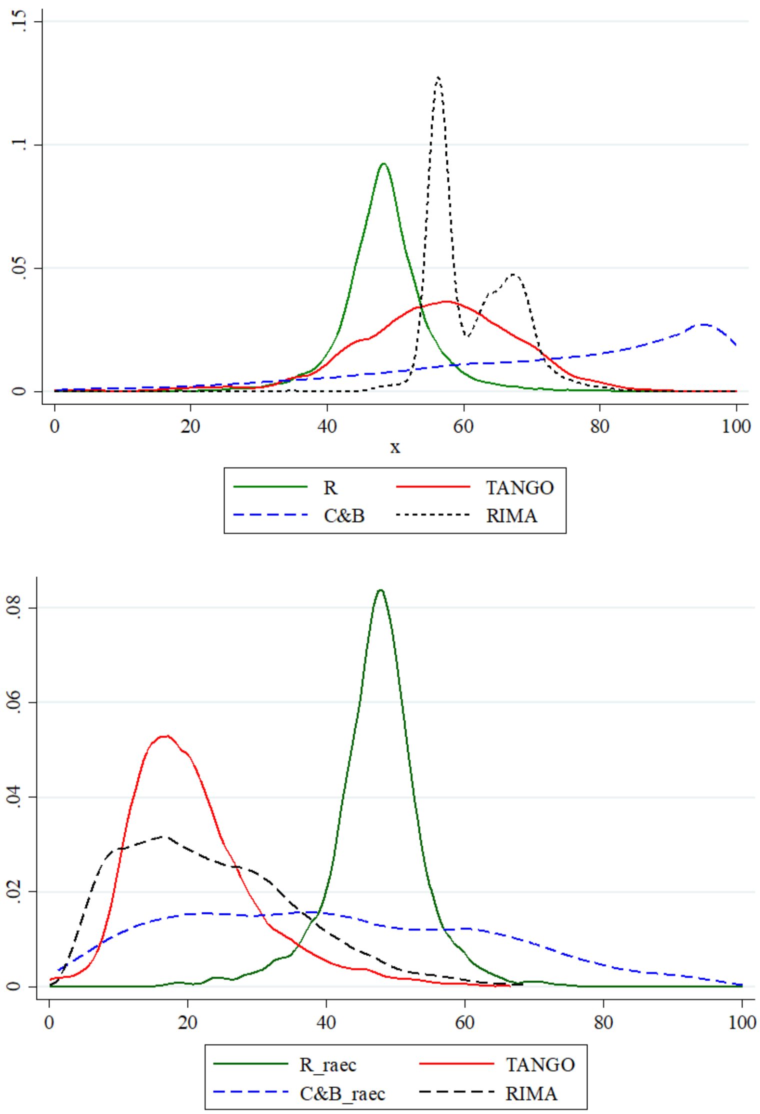


Figure A4. Spearman's Rank Correlation Coefficients (Final Wave)

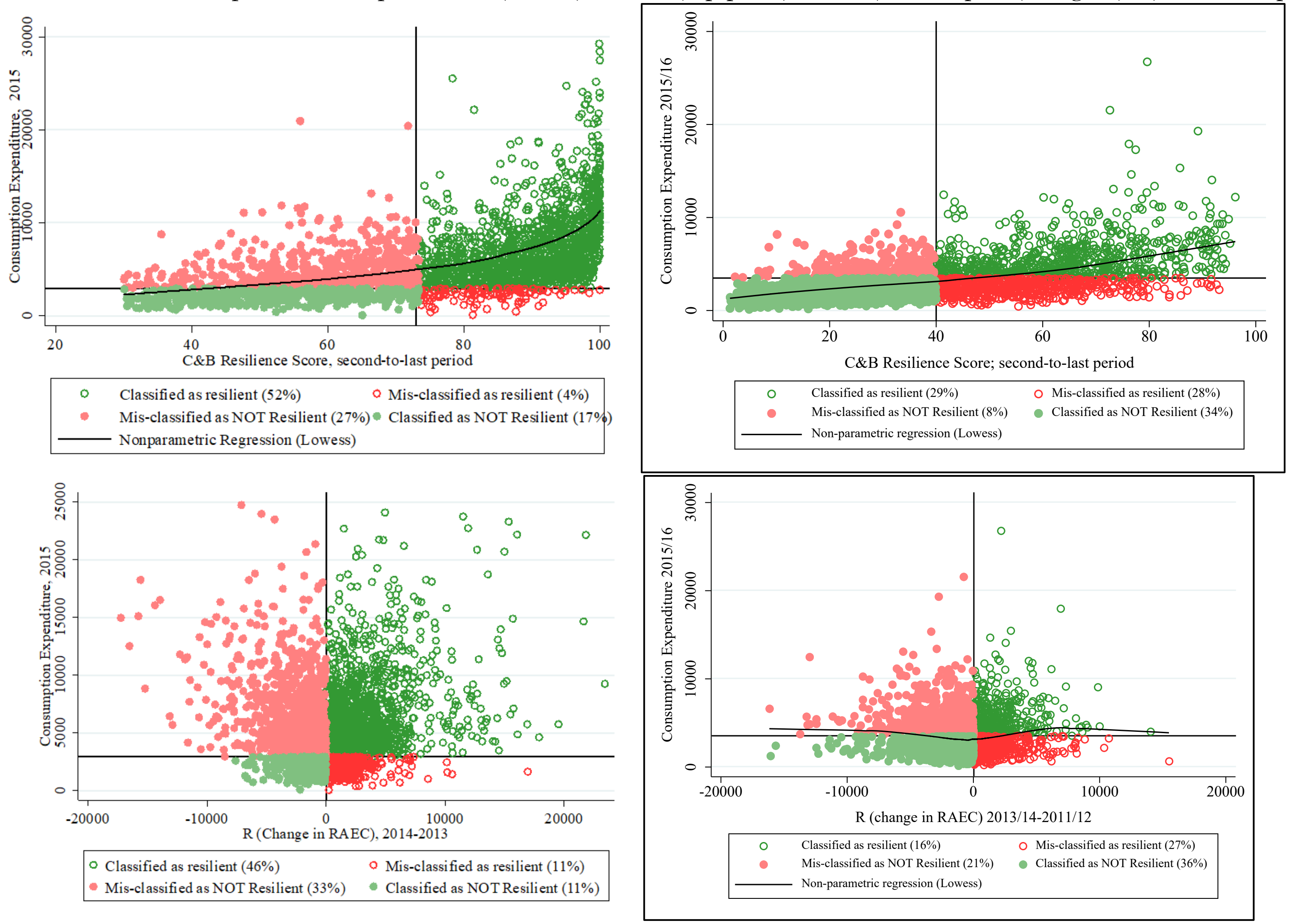
		TANGO RCI	RIMA	$\tilde{R}$_RAEC	C&B RS, RAEC
RIMA	*Niger*	*0.519*			
	Ethiopia	0.702			
$\tilde{R}$_RAEC	*Niger*	*-0.023*	*0.012*		
	Ethiopia	0.017	0.019		
C&B RS, RAEC	*Niger*	*0.354*	*0.716*	*0.093*	
	Ethiopia	0.393	0.426	-0.257	
RAEC	*Niger*	*0.221*	*0.420*	*0.428*	*0.648*
	Ethiopia	0.196	0.269	0.423	0.551

* All correlations are statistically significant at the 99% level, except the correlation between $\tilde{R}$ (realized resilience) and the TANGO and RIMA RCIs in both countries.

Figure A5. Resilience Ranking - Correspondence of the "Bottom 20%" - Final Wave

			ALSO in bottom 20% of ranking by…			
			TANGO	RIMA	R	C&B RS
In Bottom 20% of ranking by …	RIMA	*Niger*	*38%*			
		Ethiopia	51%			
	R	*Niger*	*14%*	*19%*		
		Ethiopia	18%	18%		
	C&B RS	*Niger*	*31%*	*45%*	*14%*	
		Ethiopia	26%	24%	3%	
	RAEC	*Niger*	*21%*	*32%*	*28%*	*46%*
		Ethiopia	23%	26%	24%	38%

Figure A6. Resilience and next period real expenditures (RAEC) – C&B (top panel) and R (bottom panel), Niger (left) and Ethiopia (right)



Note: Type 1 ("Mis-classified as resilient") and Type 2 ("Mis-classified as NOT resilient") errors indicated in legend. These are calculated by comparing prior period resilience classification with the subsequent period outcome; i.e., Type 1 = 1 if the prior period classified the household as resilient, while in the subsequent period the household did *not* in fact fall above the critical threshold.