A stochastic dominance approach to program evaluation with an application to child nutritional status in Kenya

Felix Naschold* and Christopher B. Barrett**

4 July 2020

*Corresponding author: University of Wyoming, Department of Economics
1000 E. University Avenue, Laramie, WY 82071
+1 307 766 2574
fnaschol@uwyo.edu

**Cornell University, Dyson School of Applied Economics and Management
340D Warren Hall, Ithaca, NY 14853

JEL classifications: I32, O22, H51
Keywords: Program evaluation, stochastic dominance, child nutrition, Kenya

We would like to thank Nancy Johnson and Ayago Wambile for inspiration and comments, Xu Claire Cheng for excellent research assistance and the Arid Lands Resource Management Project of the Government of Kenya for making the data available. Financial support for this work was provided by the International Livestock Research Institute. Any views expressed and any remaining errors are ours alone.
Abstract:

Most existing program evaluation methods examine the average impact of a program. This necessarily overlooks the potential for different program impacts over different parts of the distribution of the variable of interest. To overcome this limitation we develop a novel methodology for program evaluation which combines stochastic dominance with difference-in-differences methods. We use this new method to evaluate the impact of a large decentralization program in Kenya on changes in child nutritional status, where one’s primary concern is about sharp adverse (i.e., negative) changes. Using standard difference-in-differences regression as a baseline we find no statistically or practically significant mean impact. In contrast, our stochastic dominance estimations reveal that project expenditures have had different impacts on different parts of the distribution. In particular, they are associated with less deterioration in children's nutritional status among the worst-off children, indicating that the program effectively functions as a nutritional safety net.
**Introduction**

Most existing program evaluation methods, such as difference-in-differences estimators, regression discontinuity, or propensity score matching, examine the average impact of a program. By design they can only identify changes in a particular summary statistic of an outcome indicator, most commonly the mean or the median. However, we are often interested not only in the mean impact of an intervention, or the average treatment effect, but also the differential impact on different subpopulations such as the rich and the poor, the well-nourished and the malnourished, or some finer disaggregation of the welfare domain. In principle, one could explore heterogeneous program impact on various subpopulations by applying existing program evaluation techniques on smaller and smaller subsamples of the data. In practice, this approach faces three main problems. First, it can be cumbersome both for carrying out the analysis and for interpreting the results. Second, one faces arbitrary choices of how to split the sample. And third, increasing the number of subgroups leads smaller sample sizes and wider confidence intervals in the regression estimates. Few studies are sufficiently well powered to study impacts throughout the distribution of the full sample. To circumvent these problems we suggest a novel approach to program evaluation combining stochastic dominance testing with difference-in-differences methods.

The program evaluation literature has evolved separately from the stochastic dominance literature. Reviews of the state-of-the-art in program evaluation (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2016; Imbens & Wooldridge, 2009) do not contain any reference to stochastic dominance. In this paper we argue that there is a way to incorporate stochastic dominance techniques into program evaluation. Advances made by the stochastic dominance literature (see, e.g., Andersen 1996, Davidson and Duclos 2000, Barrett and Donald 2003) can provide another dimension to evaluating the success of programs and projects. These may be especially useful if one’s concern lies primarily in the lower tail of a distribution, as might be the case when trying to ameliorate the most severe cases of malnutrition or other well-being indicators. These techniques enable looking beyond changes in single summary
statistics, which are the focus of traditional impact evaluations, and instead examining the entire distribution of changes.

Verme (2010) showed how stochastic dominance techniques can be used for program evaluation. He uses simulated income data to show that a program can have no average treatment effect while impacting the rich and the poor quite differently. Drawing on the analogies between poverty and stochastic dominance orderings (Foster & Shorrocks, 1988) he proposes a simple method for program evaluation for the case of randomized assignment of intervention. This article extends the method to difference-in-differences evaluation to make it applicable to cases where intervention and control populations do not share the same initial distribution. It also provides the first empirical application of this technique, highlighting the importance to look beyond average treatment effects.¹

To illustrate this method we use a unique, large data set from arid and semi-arid Kenya to compare changes in acute child malnutrition, measured by the Mid-Upper Arm Circumference (MUAC) and evaluate how effective newly decentralized public expenditure was in reducing child malnutrition. We compare the differences in changes in nutritional status between areas that did and did not benefit from additional, decentralized public expenditures through the second phase of the World Bank-funded Arid Lands Resource Management Project (ALRMP II). This article appears to be the first to evaluate welfare changes over time in a stochastic dominance framework. It is also the first study to use stochastic dominance analysis for MUAC data.

¹ We recently discovered another paper (Van de gaer, Vandenbossche, & Figueroa, 2013) which cites an earlier version of our paper and pursues a somewhat similar line of evaluation, based on propensity score matching followed by stochastic dominance analysis.
Acute malnutrition remained pervasive in arid and semi-arid Kenya over the 2005-2009 period we study. Using standard difference-in-differences regression as a baseline we find no statistically or practically significant mean impact of ALRMP II expenditures on child malnutrition. In contrast, our stochastic dominance estimations reveal that project expenditures have had different impacts on different parts of the distribution. In particular, they are associated with a positive impact on child nutritional status at the lower end of the distribution. They may have prevented the nutritional status of the worst-off children from worsening and, thus, may have functioned as a nutritional safety net. While data limitations limit our ability to assess the robustness of these findings, the association between program investments and nutritional declines averted in the part of the distribution policymakers most care about is clear.

These findings highlight the importance of looking beyond average impacts. Looking beyond averages has entered the mainstream in poverty analysis and has yielded more nuanced insights (Ravallion, 2001). The stochastic dominance based difference-in-differences technique proposed in this article suggests a way for doing the same in program evaluation.

**Existing program evaluation approaches**

The fundamental problem of program evaluation is that we cannot observe a person $i$’s outcomes in two states: treatment and non-treatment. Let $x$ be the outcome of interest and subscripts $T$ and $C$ denote treatment and non-treatment, respectively. In our application below this will be a malnutrition indicator for children, but $x$ could equally be income, consumption, mortality or any other welfare indicator or any other continuous measure relevant for program evaluation. We would like to evaluate the program impact $\beta_i$

$$\beta_i = x_{iT} - x_{iC}$$

but cannot because we only observe either $x_{iT}$ or $x_{iC}$ but not the corresponding counterfactual.
One standard way to overcome this problem is to look at differences across people rather than the unobservable differences for \(i\) over states. When treatment assignment is randomized then the distribution of the outcome variable should be the same for the subpopulation that benefited from a program (the ‘treatment group’) and those that did not participate in the program (the ‘control group’). We can then look at single differences to compare the difference in outcomes. In the case of means, the average program impact, \(\beta\), is equal to

\[
E[\beta] = E[x_T] - E[x_C]
\]  

(2)

When the assignment of intervention has been non-random and intervention and control groups differ systematically the estimated \(E[\beta]\) is biased. Instead, we can then test for a treatment effect by comparing differences over time between intervention and control groups. If we have repeated observations over time at \(t\) and \(t-1\) for each \(i\) the average treatment effect \(\beta\) can be estimated through differences-in-differences (DD)

\[
E[\beta] = E[x_{T,t} - x_{T,t-1}] - E[x_{C,t} - x_{C,t-1}]
\]  

(3)

The key shortcoming of most existing approaches to program evaluation is that they are limited to focusing on the impact of an intervention on a particular moment of the distribution, typically the mean. Yet, policy makers often want to know about the distributional impacts of programs, beyond the average treatment effect as the impact of any policy change likely has heterogeneous impacts with gains from participation varying across the population (Ravallion, 2019). Moreover, any estimated average treatment effect in the sample is unlikely to fully represent the average treatment effect across the whole population (Deaton & Cartwright, 2018). Analysis by subgroups – including by quantile regression – is the de facto ‘state of the art’ to incorporate heterogeneity in average treatment effects.
(Subramanian, Kim, & Christakis, 2018). While this can provide some insights beyond the mean effect, the usefulness is limited by the need to define the subgroups a priori, and potentially by sample size.

Improved methods for estimating heterogeneous treatment effects are emerging. First, regression tree methods based on machine learning (Athey & Imbens, 2016) can define subgroups that differ in the size of their treatment effect based on the data themselves. Second, local instrumental variables (Heckman, Urzua, & Vytlacil, 2006) can account for heterogeneity in responses to treatment, when people’s participation in programs depends on their own, partial knowledge of their idiosyncratic response to the program treatment. Local IV can help estimate the marginal treatment effects over the entire distribution of being treated. Third, one can treat residual variances as different nuisance parameters for treatment and control groups (Deaton & Cartwright, 2018); or even incorporate heterogeneity as a population characteristic of intrinsic interest by including separate residual terms for treatment and control groups (Subramanian et al., 2018).

These promising new methods that look beyond the average treatment effect have not yet seen many empirical applications, however. This article proposes another method to examine heterogeneous program impacts. It is based on stochastic dominance and can estimate treatment effects across the entire distribution of the outcome variable.

**Using stochastic dominance for difference-in-differences estimation**

Stochastic dominance analysis takes account of entire distributions or sub-ranges of distributions. In the context of welfare comparisons across space or across time this makes stochastic dominance methods superior to and more robust than traditional difference-in-differences techniques in the following two ways.
First, it expands welfare comparisons beyond a single, arbitrary cut-off point. We use the term ‘poverty line’, denoted by \( z \), as a shorthand for this cut-off. Note that this ‘poverty line’ could be an actual consumption poverty line or any similar metric such as the negative standard deviation of an anthropometric index that we use in our application later. Since the location of a poverty line \( z \) is arbitrary, it is often contentious. It is often much easier to agree on a range in which the poverty line should be set such that \( z \in [z_{\text{min}}, z_{\text{max}}] \) where \( z_{\text{min}} \) and \( z_{\text{max}} \) are the lowest and highest poverty lines that are considered reasonable. Stochastic dominance techniques accommodate ranges of poverty lines and, thus, can make welfare comparisons robust to the choice of poverty line.

As an example consider the evaluation problem in our application below. It is not clear where to set the malnutrition poverty line expressed as standard deviations from the mean of Mid-Upper Arm Circumference (MUAC) Z-score measures for small children. Z-scores of -1, -2, and -3 are often regarded as the cut-off points for mild, moderate, and severe acute malnutrition, respectively. However, one can easily make the case for other ‘poverty lines’. The entire range of ‘reasonable’ MUAC poverty lines is probably spanned by, say, \( z \in [-3, 0] \). Then, if one distribution consistently has less malnutrition than another over that range of poverty lines then the former distribution is strictly preferable to the latter.

Second, stochastic dominance can be used to make comparisons for broad classes of welfare indicators. In our analysis below there aren’t any credible alternative indicators to the Z-score based malnutrition measure. However, when evaluating material poverty there is often disagreement about which indicator to use. In practice this can matter as different poverty indicators can yield different results. Stochastic dominance analysis can consolidate the conclusions as they are valid for a range of poverty measures that satisfy some basic common properties such as additive separability which is satisfied by the class of P-alpha measures, the Watts index and the Clark-Hemming-Ulph indicator (Fields, 2001).
Definitions of orders of dominance

Let $F$ denote a set of probability density functions of a random variable $x$ defined on a closed interval $[x_{\text{min}}, x_{\text{max}}]$. Further, let $f_A(x) \in F$ and $f_B(x) \in F$. Denote the respective cumulative density functions (cdf) by $F_A(x)$ and $F_B(x)$.

Distribution $A$ first order stochastically dominates (FOD) distribution $B$ up to poverty line $z \in [x_{\text{min}}, x_{\text{max}}]$ if and only if (iff) $F_B(x) - F_A(x) \geq 0 \ \forall \ x \in [x_{\text{min}}, z]$, that is, iff $F_A(x)$ lies nowhere above $F_B(x)$.

Higher orders of stochastic dominance are defined on higher order integrals of the cdf. Let $s \geq 2$ denote the order of integration. Then $F^s(x) = \int_{x_{\text{min}}}^{x_{\text{max}}} F^{s-1}(z) \, dz$. Therefore, distribution $A$ $s^{\text{th}}$-order dominates distribution $B$ iff

$$F_B^s(x) - F_A^s(x) \geq 0 \ \forall \ x \in [x_{\text{min}}, z]$$

(4)

These standard stochastic dominance criteria can be applied directly to program evaluation if intervention and control populations share the same initial distribution. This same initial distribution criterion is important and implies a need for either randomized treatments or effective control for other covariates that affect the initial distributions of the intervention and control populations.

---

2 While this definition of FOD always holds in a statistical sense, whether the statistically dominant distribution is preferable to the dominated distribution in reality depends on whether $x$ is a ‘good’ such as income or MUAC or a ‘bad’ such as deprivation.
Stochastic dominance for difference-in-differences impact evaluation

Many, perhaps most, policy or program evaluations rely on non-(quasi-)experimental data. Intervention and control groups are not randomly selected and are, thus, likely to differ in some intrinsic characteristics. Therefore, we cannot look at the simple difference in outcomes between the two groups but need to examine the difference-in-differences (DiD) in outcomes across time and across subgroups, controlling for lots of other covariates that might simultaneously impact outcomes and perhaps be correlated with group membership. The DiD approach can be applied in a stochastic dominance context. Much of the discussion of stochastic dominance on simple differences from above carries straight over but there are important difference in interpretation and usefulness of higher order dominance tests.

Let $\Delta$ denote the difference in a random variable $x$ between time $t$ and $t-1$ defined on the closed interval $[\Delta_{\text{min}}, \Delta_{\text{max}}]$ such that $\Delta = x_t - x_{t-1}$. Further, let $G$ denote the set of probability density functions of $\Delta$.

Further, let $g_A(\Delta) \in G$ and $g_B(\Delta) \in G$. Denote the respective cdfs by $G_A(\Delta)$ and $G_B(\Delta)$. Then, distribution $A$ first order stochastically dominates (FOD) distribution $B$ iff

$$G_B(\Delta) - G_A(\Delta) \geq 0 \quad \forall \Delta \in [\Delta_{\text{min}}, \Delta_{\text{max}}],$$

that is, iff $G_A(\Delta)$ lies nowhere above $G_B(\Delta)$.

Note that unlike in the case of stochastic dominance between two outcome levels this definition does not refer to a poverty line. Rather the ‘poverty line’ cut-off is expressed in differences and, therefore, determining its location is even more subjective than a poverty line in levels. The use of such a cut-off point for changes in the variable of interest depends on the particular focus of the evaluation. For instance, we could focus on negative changes to determine whether the intervention or control group had fewer negative changes. The corresponding FOD condition would be

$$G_B(\Delta) - G_A(\Delta) \geq 0 \quad \forall \Delta \in [\Delta_{\text{min}}, z].$$

Conversely, we could look at only positive changes, or any other partial range of welfare changes.
Higher orders of stochastic dominance of welfare differences are defined on higher order integrals of the cdf. Let \( s \geq 2 \) denote the order of integration. Then \( G^s_s (\Delta) = \int_{\Delta_{\min}}^{\Delta_{\max}} G_{x-1}^s (z) \, dz \). Therefore, distribution A \( s^{th} \) order dominates distribution B iff \( G^s_B (\Delta) - G^s_A (\Delta) \geq 0 \) \( \forall \Delta \in [\Delta_{\min}, \Delta_{\max}] \).

There is an important difference in interpreting the results from SD on welfare levels versus on changes or differences. Stochastic dominance analysis is based on cdfs, which order the variable of interest from smallest to largest. In the case of welfare levels, the lowest values pertain to the worst-off individuals and welfare levels are always positive. In contrast, welfare changes can be negative and the largest negative changes are not necessarily associated with the worst-off individuals. Indeed, under any regression-to-the-mean process, the largest negative changes are more likely from people who were relatively well off in period \( t-1 \) and, thus, had farther to fall. In any event, the cdfs of welfare changes are ‘poverty blind’. This difference in interpretation of stochastic dominance results of welfare levels vs. welfare changes matters most if we are concerned about the poor, or in our case, malnourished children. To partially overcome the ‘poverty blindness’ we can run stochastic dominance on differences on the subset of people that were poor at \( t \), at \( t-1 \) or in both periods. Alternatively, we can link levels and changes in welfare by weighting changes for individuals, for instance, according to their starting positions.

The difference in interpretation between SD results on levels and changes is also relevant as we move from first to higher order stochastic dominance. The smallest welfare changes appear at the lower end

---

3 We have replicated the analysis weighting the MUAC Z-scores in two different ways: by the inverse of the starting MUAC value (which is negative for all sublocations) and by the inverse of the cdf. The weighting function does not change the results substantively but, since the weights lie between zero and 1, they compress the support of MUAC Z-score changes.
of the domain regardless of the welfare level. Hence, the conventional poverty analysis attributes of
second and third order dominance, namely equality preference and transfer sensitivity, no longer apply
in the same way as for the analysis of levels. First order SD tests sensibly check for differences in
distributions of changes between intervention and control sublocations. Second order SD tests assess
the extent to which one distribution’s changes are concentrated at the lower end of the distribution of
changes. Third order SD tests on differences distributions, however, are not really meaningful in any
policy sense we can determine. The lower end of this distribution, that is, the most negative changes in
nutritional status, do not necessarily represent the most malnourished sublocations and it would make
little sense to give additional weight to the lower end of the distribution, which is what third order SD
testing would do. We therefore suggest applying SD test on changes data only up to order s=2.

There is no meaningful range of sensible malnutrition cut-off values in terms of changes in MUAC Z-
scores. Therefore, we test for stochastic dominance over the entire domain rather than the typical right-
truncated domain used in consumption or income poverty analysis.

Our stochastic dominance based method for program evaluation has one potential disadvantage
compared to regression-based DiD estimators. In the standard regression program evaluation approach
one can include other covariates as righthand side variables in a single estimation step. Our stochastic
dominance method can likewise be used to evaluate program impact net of other covariates, as we
demonstrate; it just cannot do it simultaneously with estimating the program impact. To account for
covariates we run a first stage regression of the outcome variable on the desired covariates; then we use
the residuals, which represent the variation in the outcome variable net of control variables, in the
stochastic dominance estimation. The potential statistical error in this first stage is analogous to the
error term in standard regressions, including DiD estimation. In the application below we use this
method to strip out the effects of drought on child malnutrition by using the residuals of a regression of
MUAC Z-scores on the Normalized Difference Vegetation Index (NDVI) – a regionally-appropriate indicator of drought stress – in the stochastic dominance analysis. This makes the stochastic dominance analysis conditional on NDVI and, thus, comparable to conventional regression-based DiD analysis.

**The setting and data**

To illustrate the use of stochastic dominance for DiD evaluation we use a unique, large dataset of child nutritional status from arid and semi-arid lands (ASALs) in Kenya. These areas are characterized by extensive livestock production and the highest incidences of poverty in Kenya. Over 60% of the population lives below the poverty line and levels of access to basic services are very low (Kenya, 2006). Child malnutrition levels in Kenyan ASALs are generally declining over time, but are still above emergency threshold levels, and regularly worsened by recurrent droughts and disease outbreaks. In 2005, the starting year of this evaluation, stunting levels for children in arid and in semi-arid districts were 46.5% and 43.0%, respectively, while 15.0% and 21.2%, respectively, suffered from wasting.

The data were collected by the Kenyan government under the second phase of the Arid Lands Management Project (ALRMP II), a World Bank-financed community-based drought management initiative that provided additional financial resources to 28 ASAL districts from 2003 to 2010. The overall aim of ALRMP II was to foster economic growth and reduce poverty within the framework of Kenya’s Poverty Reduction Strategy Program. This was to be achieved by focusing on a number of development objectives including reducing livelihood vulnerability, increasing access to basic services, enhancing food security, empowering local communities, and raising the profile of ASAL areas in national policies and institutions. As the first seven-year phase of ALRMP had already been completed in 2003, the implementation of ALRMP II was not rolled out but started simultaneously in all project districts. ALRMP II sought to achieve its development objectives through improved drought and natural resource management.
management, community driven development, and greater decentralization and enhanced support to local government.

One of the key indicators chosen to monitor the success of ALRMP II was child malnutrition normalized by the severity of drought. Child malnutrition is a major concern in arid and semi-arid Kenya. ALRMP II was designed to improve child nutrition through various channels. First, participating sublocations and districts were to develop an effective drought cycle management system to minimize the need for emergency food aid as well as to enhance the effectiveness of food aid when it is needed. Second, improved natural resources management was intended to reduce the vulnerability of agricultural and pastoral households to drought and to improve their livelihoods through sustainable management of rangeland, water and other natural resources. Third, decentralizing decision making power to the local level was hypothesized to empower communities to identify their development priorities and enable them to use resources where and how they are most needed, particularly in the areas of health and education. Fourth, delegating fiscal authority to the local level was intended to make government expenditures more responsive to local needs in order to improve the delivery of essential health services. The child malnutrition indicator, MUAC, was to be ‘normalized by drought’ in an attempt to strip out the variation in child nutrition caused by exogenous changes in agro-ecological conditions and to focus instead on the impact of ALRMP II. The program aimed to mitigate increases in the prevalence of mild, moderate and severe acute malnutrition among children through a suite of locally-customized interventions.

The project’s monitoring strategy therefore included the collection of information on child nutritional status. The specific anthropometric indicator collected was Mid-Upper Arm Circumference (MUAC) for children younger than 60 months. MUAC is a reliable and relatively cheap-to-collect indicator for child nutrition status. It is also closely correlated with clinical and other anthropometric indicators of
nutritional status (A Shakir, 1975; A. Shakir & Morley, 1974). MUAC is also considered a better predictor of severe acute malnutrition than weight-for-height (Briend, Maire, Fontaine, & Garenne, 2012; Myatt, Khara, & Collins, 2006) and than other anthropometric measures in identifying acute malnutrition as MUAC measures are less affected by acute dehydration (Mwangome, Fegan, Prentice, & Berkley, 2011). Overall, MUAC is widely considered more appropriate than other measures for children in the pastoral areas that cover much of arid and semi-arid Kenya (Mude, Barrett, McPeak, Kaitho, & Kristjanson, 2009).

We use MUAC Z-scores rather than absolute MUAC measures as z-scores allow a direct comparison across age and gender of children. Z-scores for weight-for-age or height-for-age are routinely used to measure child nutritional status. For some reason, perhaps inertia from when MUAC Z-scores were difficult to calculate, most studies (Ritmeijer, 1998) and the current 2006 WHO Child Growth Standards for emergency nutrition programs still use raw MUAC measures in centimeters, despite clear evidence that Z-scores are the preferable measure of nutritional status (de Onis, Yip, & Mei, 1997; Gernaat, Dechering, & Voorhoeve, 1996; Mei, Grummer-Strawn, De Onis, & Yip, 1997). One reason may be that converting MUAC measurements into Z-scores may not be necessary if the only objective is to identify children with high mortality risk (Rasmussen et al., 2012).

The raw MUAC measures for all children from 0-59 months old were converted into z-scores as follows

$$Z(MUAC_{ijt}) = \frac{MUAC_{ijt} - MUAC(\text{reference population})}{\sigma_{MUAC(\text{reference population})}}$$  \hspace{1cm} (5)$$

where $MUAC_{ijt}$ is child $i$'s MUAC at time $t$ in location $j$ and $\sigma$ indicates the standard deviation. The reference population values are taken from the 2006 WHO/NCHS data (de Onis et al., 1997) and are disaggregated by age and gender.
Between June 2005 and August 2009 over 602,000 individual child MUAC measurements were taken in 115 sublocations in the 10 arid and semi-arid ALRMP II districts encircled in Figure 1. A sublocation is the smallest administrative unit defined under ALRMP II and corresponds to a geographically cohesive rural area. Average sublocation population was 5-10,000 people during this period, indicating that sublocations are the size of large rural villages.

We classified sublocations into intervention and control groups according to the cumulative ALRMP II investment received by each sublocation. These data were provided by the ALRMP district data managers. Investment per capita varied across intervention locations. 52 sublocations did not receive any sublocation specific ALRMP II funds. We use these sublocations that received no ALRMP II Investment funds as the control group. 63 sublocations received ALMRP II money. These amounts varied across sublocations. However, we would expect to see a positive impact in all intervention sublocations given that these sublocation i) received additional funds with the specific aim to reduce child malnutrition and ii) were simultaneously delegated authority to spend these funds based on local needs for basic health services.\(^4\)

\(^4\) As a robustness check we also used the district project managers’ own classification of sublocations in their districts. Their classifications were based on i) whether at least 2 out of 3 ALRMP II program components were implemented and ii) whether there was substantial ALRMP II investment. However, managers only provided classifications for 93 of the 115 sublocations. For these 93 sublocations, managers’ classifications agreed with investment data-based classification for 64 sublocations. Managers classified a further 23 as intervention when the investment data showed no investment. This is probably because managers identified sublocations as intervention when at least two out of three ALRMP II program components were implemented, even if sublocations did not receive any additional investment funding. In any case no substantive differences emerge among the two classification schemes. For brevity and to preserve full sample size we focus on the investment-based treatment and
We use the 2005/06 round of data as our pre-treatment starting point as this represent the baseline data point before ALRMP II policies and expenditures were implemented. While there was a first phase of ALRMP between 1996 and 2003 this prior program had little effect on the baseline of this evaluation. Unlike ALRMP II, the first phase was not focused on reducing child malnutrition or on decentralization of public expenditure. ALRMP I was also small in scope with a total expenditure of USD 25.1million versus the USD 142.85 million budget of ALRMP II. Moreover, ALRMP I was only administered in ten arid districts. 18 new, semi-arid districts were added in ALRMP II.

< Figure 1 ALRMP II Project Districts HERE>

To estimate changes over time and compare them between intervention and control sublocations more rigorously we need to construct a panel. Our child-level observations are unsuitable for this for three reasons. First, individual child identifiers are not consistent across time in the data set. Second, MUAC data are not available for all children in all months. Third, and most importantly, the sample of children will necessarily change over time. A large proportion of MUAC observations is lost over the three year period from 2005/06 to 2008/09 as many children observed in the early years have exited the 6-59 month age group and children born since 2005/06 were added to the sample. We therefore treat the sublocations as units of analysis.

For the 115 sublocations we constructed a two period panel for 2005/06 and 2008/09 of sublocation-specific MUAC z-scores by summarizing the child-level MUAC z-scores in sublocation summary statistics. This provides a baseline-to-endline evaluation of the ALRMP II program impacts. One could in principle compare earlier distributions with the baseline – or indeed, compare between any two distributions control classifications. Note that in the event of spillovers from sublocations with ALRMP II investments to those without, our estimates are biased towards finding no impact.
over the evaluation period – but the longer interval minimizes noise associated with measurement error and transitory shocks (Naschold and Barrett 2011) and the program target was results by endline. We therefore restrict comparisons to those two periods. To focus primarily on malnourished children the results presented below are based on summary statistics that focus on that subpopulation such as the median Z-score of all children with Z-scores below zero, or the proportion of children with MUAC Z-scores below -1 and -2 standard deviations, focusing on standard cut-off levels to capture the prevalence of mild or moderate malnutrition.5

These particular summary statistics are sensible truncations of the MUAC Z-scores distribution since we want to focus on acutely undernourished children. This right-truncation in these summary statistics is analogous to the focus axiom in poverty measurement. We can safely ignore level of and changes among higher MUAC Z-scores since high MUAC observations and large positive changes at the upper tail of the distribution are not necessarily desirable or positive. Although overweight and obesity are rare in these Kenya ASAL population, in the context of child nutrition more is not always better. In any event, this truncation does not affect the panel estimates as none of the sublocations had positive average MUAC Z-scores.

Although our analysis is at the aggregate, sublocation level, for which there is no attrition nor missing data, these do reflect panels and repeated cross-sections of children within sublocations. As with any panel, attrition can be a concern for two reasons. First, children may drop out of the sample due to

5 In total we constructed annual means for 14 monthly sublocation-specific MUAC Z-score summary statistics. These summary statistics include the median MUAC Z-score for children with Z-scores below 0, -1, and -2; the mean MUAC Z-score; the median Z-score of children with Z-scores below 0, -1, and -2; the percentage of children with Z-score below 0, -1, and -2; the Z-score gap of children with Z-score below -1 and -2; and the squared Z-score gap of children with Z-score below -1 and -2. Results for the additional indicators are available on request.
spatial mobility. Households might migrate especially in relatively bad years such as the period spanned by this panel. This could affect observed changes in malnutrition summary statistics either because the best- or worst-fed children leave their place of residence. We don’t have specific data to investigate systematically if either of these two sources of attrition are prevalent. However, anecdotal evidence from ALRMP staff and local collaborators that helped collect data in the ALRMP II project districts suggest that there was not much spatial mobility across sublocations.

A second potential source of panel attrition can be child deaths. In our panel of MUAC measures this was less of a concern as all MUAC Z-score summary statistics improved, if only slightly. If indeed there had been a significant increase in child deaths then those children would drop out of the sample. However, if conditions were so bad as to have led to more child deaths then we would also expect the whole sublocation distribution of MUAC measurements to shift to the left. Malnutrition levels overall would increase as all children would be expected to be affected by worsening food availability and access associated with deteriorating agro-ecological conditions. The fact that we did not see such a pattern strongly suggests an absence of significant differential child mortality across intervention and control sublocations.

Results

Table 1 presents the sample sizes and selected MUAC Z-score statistics for intervention and control sublocations for the two years used in the analysis. Overall levels of nutrition are very low. All summary statistics lie below minus one standard deviations, indicating mild malnutrition on average in this population. The severity of malnutrition is particularly evident when looking at the 25th and 10th percentile of the distribution. For example, more than 10 percent of children fall below the threshold of moderate malnutrition of minus two standard deviations. Children in intervention sublocations are worse off than those in control sites. These differences between intervention and control MUAC Z-
scores in table 1 are statistically significant at the 1% level at the median, 25th and 10th percentile of the distribution in both years. These differences between intervention and control sublocations indicate a placement effect of the ALRMP interventions. The difference-in-differences method helps control for some of this non-randomness and avoids the negative bias of simple evaluation methods that fail to control for initial differences between control and intervention groups. These summary statistics suggest that child nutrition improved from 2005/06 to 2008/09 with larger improvements in intervention communities and among the least well off.

Table 1: Sample size and selected MUAC Z-score statistics for intervention and control sublocations

The remainder of this section presents the program evaluation results for both the difference-in-differences regressions and for stochastic dominance to highlight the potential practical importance of looking beyond the average treatment effect.

Regression results

The difference-in-differences estimator in equation 3 was estimated as

\[
\Delta MUAC_{SS_j} = \gamma_0 + \gamma_1 D_j + \gamma_2 NDVI_j + \gamma_3 NDVI_j^2 + \sum_{l=2}^{L} \delta_l L_l + \epsilon_j \tag{6}
\]

where \(\Delta MUAC_{SS_j}\) is the change in a MUAC summary statistic for sublocation \(j\), \(D_j\) is a dummy variable equal to one for intervention sites, NDVI is the normalized difference vegetation index, and \(L_l\) are district dummy variables to capture unobserved regional variation, for example, in the effectiveness of

---

6 We used quantile regression tests for equality of percentiles between two distributions following Conroy (2012).
administration as well as geography and remoteness. The NDVI is a measure of agro-ecological conditions which are an important determinant of agricultural and pastoral outcomes and, hence, of child nutritional status in the ASAL. NDVI is an outcome variable that captures the variation arising from a range of variables such as rainfall, soil types, grazing pressure, and other agro-ecological factors that determine the level of vegetation, including the yields of important staple crops. These, in turn, are a key determinant of child nutritional status directly, through the availability of self-produced food and indirectly, through lower household incomes. Indeed, NDVI has been shown to be highly correlated with child nutritional status in Sub-Saharan Africa (Sedda et al., 2015) and that in the ASAL regions of Northern Kenya a one standard deviation increase in NDVI z-score decreases the probability of child malnourishment by 12-16 percent (Bauer & Mburu, 2017). We did not include other control variables that may influence health more directly, such as time and expense needed to reach health clinics as no reliable measurements for these are available. Similarly, we could not include common household survey variables such as economic well-being or education as these were not collected in the ALMRP dataset. Our NDVI data correspond to the sublocation level and reflect agro-ecological conditions at each locality at each relevant point in time.

\[X_{jt} = \alpha_0 + \alpha_1 T + \alpha_2 P2 + \alpha_3 T \times P2,\]

where \(T\) is a dummy variable indicator for being in the treatment group, and \(P2\) is a dummy variable indicator for the treatment period. The parameter \(\alpha_3\) from this levels specification is equivalent to \(\gamma_1\) from equation 6, and parameter \(\alpha_2\) is equivalent to \(\gamma_0\) from equation 6. The differences specification sacrifices degrees of freedom for control for all time-invariant unobservables, which seems a desirable tradeoff given our sample size and the non-random assignment of interventions to sublocations.
In arid and semi-arid Kenya changes in agro-ecological conditions over the year introduce seasonal fluctuations in child nutrition. To control for this seasonality we use MUAC data from the same time of year in both the starting and the ending periods of the panel. The NDVI data are from the same months as the MUAC measures.

Comparability between intervention and control sites is as important in stochastic dominance impact evaluation as in traditional types of impact evaluation. The main control variable used in the analysis below is the NDVI. Balance tests show that in 2005/06 there was no statistically significant difference in NDVI between intervention and control sublocations, although the average NDVI in intervention sites was slightly lower. In 2008/09 the difference was statistically significant (see online appendix 2), as mean NDVI remained constant in control sites but fell in intervention sublocations. These patterns suggest that drought dynamics might actually attenuate the statistical power of our main results on MUAC Z-score changes.

In the absence of other quantitative data at the sublocation level we attempted to elicit any systematic differences between intervention and control locations using two qualitative approaches. First, the comparability of intervention and control sublocations was discussed at an ALRMP II workshop in Nairobi. Managers from all ALRMP II project districts attended. Collectively, they did not feel that there was any substantive bias or difference between intervention and control sublocations ex ante. Second, this assessment was confirmed through interviews with local collaborators who as part of this research project had visited the districts and talked with district officials.

Table 2 reports the regression estimates from equation 6. The first row shows that there was no statistically significant average treatment effect in any of the five MUAC Z-score summary statistics. Based on traditional difference-in-differences estimation we would conclude that there has been no average impact of ALRMP expenditure on child nutrition levels. Of course, that is just an estimate of
average treatment effect. Although that magnitude is of interest, there is greater interest in assessing whether the program was associated with reducing sharp negative changes in child nutritional outcomes, i.e., did ALRMP II effectively mitigate adverse effects of drought on child nutrition. In order to answer that question effectively, we need to focus attention on the lower tail of the MUAC change distribution.

Table 2 Difference-in-differences Panel Regression

<HERE>

**Stochastic dominance results**

The first order stochastic dominance results comparing the changes in intervention with the changes in control sublocations are plotted in figures 2-5. For the median MUAC of all MUAC observations below zero there is no full first order dominance between changes in intervention and changes in control locations. The cdfs cross at a positive change in the MUAC Z-score of around 0.2 as shown in figure 2. However, below Z=0.2 intervention sites FOD control sites, indicating that a smaller percentage of intervention sites had negative changes in the drought adjusted median MUAC Z-score of all observations with MUAC less than zero. For instance, around 45% of control sites had negative changes in their Z-score compared to only around 20% of intervention sites. Note that figure 2 compares changes in median MUAC Z-scores which for both intervention and control sublocations are between -1

---

8 The stochastic dominance tests were run in STATA using the *cfgt* command from the DASP package (Abdelkrim & Duclos, 2007).

9 Figures 2-5 show a dotted line at MUAC Z-score equal to zero. To the right and left of that line changes in MUAC Z-score are positive and negative.
and -1.2 (table 1), far enough from any biological boundary that would limit how negative any changes in MUAC could be.

Analogous second order stochastic dominance results are provided in online appendix 3. Table 3 contains a summary of the results discussed in the previous paragraph and, as a robustness check, results for another MUAC Z-score summary statistics, the proportion of sublocations below minus 1 standard deviation, which display the same pattern of dominance.

These stochastic dominance results suggest that ALRMP intervention sites were more effective in preventing a worsening of nutritional status, even if in terms of absolute MUAC Z-scores intervention sites still lag behind control sites (see Table 1). At the very low levels of nutrition characteristic of the arid and semi-arid lands of Kenya, any further drop in child nutrition levels has serious consequences. This adds practical significance to the finding that intervention sublocations have fared less badly than their control counterparts. Above 0.2 the two cdfs are fairly close and intersect repeatedly indicating that intervention and control sites had roughly equal proportions of sites that experienced equal improvements child nutritional levels over time. Hence the statistically insignificant average treatment effect reflected in Table 2; the main effects are concentrated in the lower tail of the change distribution, which is itself an outcome of note for a program such as ALRMP II.

Table 3 Summary of Stochastic Dominance Results – Difference in intervention vs. control sublocations

Figure 2

Figure 3
Figure 3 displays the difference between the two curves in figure 2 including the 95% confidence bands. It indicates that the partial first order stochastic dominance between intervention and control groups is statistically significant around zero and almost significant below zero. Given the small size of the sublocation panel, the short time period and the relatively modest investments, the lack of greater statistical significance is not surprising.\footnote{To cross-check our results we also used the individual data wherever possible to complement the sublocation panel data. The results of the panel and the results of the individual data always matched with the latter always being statistically significant. This suggests that sample size is the limiting factor in the sublocation panel analysis. Intuitively, the close correspondence of results of the SD test where we can use both datasets might let one put a bit more confidence in the significance of the panel result.} As an additional test for the statistical strength of the dominance result we calculated Somers’ D statistics. Below changes in MUAC Z-scores of 0.2 (the crossing point of the cdfs in figure 2) intervention sites are 12% more likely to have experienced a more positive/less negative change in MUAC Z-scores. Below changes of 0 (the vertical line of no change in MUAC Z-score) this rises to 23% more likely.

Figures 4 and 5 show the results for changes in MUAC Z-scores for the 25\textsuperscript{th} and 10\textsuperscript{th} percentile, respectively. As we focus on smaller and smaller percentiles of the distribution (from the median to the 25\textsuperscript{th} to the 10\textsuperscript{th} percentile) the analysis concentrates increasingly on the worst-off sublocations. In all cases the intervention sites seem to have succeeded in preventing negative changes in MUAC Z-scores relative to the control sites. For the 25\textsuperscript{th} percentile subsample in figure 4, cdfs cross at around 0.25 indicating that there were fewer negative changes for the intervention sites than for the control sites. Similarly, for the 10\textsuperscript{th} percentile in Figure 5, the cdfs cross near 0.3. Around 15% of intervention sublocations had a negative change in MUAC Z-scores of -0.1 whereas around 30% of control sublocations had the same negative change. In addition, at the 10\textsuperscript{th} percentile there were also fewer
smaller positive changes. Again, these results are not statistically significant, likely a result of the small sample size of the sublocation panel dataset. However, the pattern of results from figures 4 and 5 mirror the results for the median MUAC Z-scores in figure 2.

Figure 4
<HERE>

Figure 5
<HERE>

Conclusions and implications

Existing approaches to program evaluation are designed to examine the average treatment effect. In practice, however, we are often interested not just in the mean impact but also, or instead, in impacts over a particular interval, often the lower tail of a distribution. Indeed, as in the case of child nutrition we are particularly interested in program impacts on the worst off children and locations and in the largest declines in nutritional status. This article has proposed a new method to evaluate program impacts across the entire distribution – or any interval – of outcomes or changes in outcomes. The method does not require experimental data as it applies stochastic dominance estimation to conditional (i.e. pre-filtered) differences-in-differences across subgroups and time.

Our empirical results highlight the practical added value of this method. Standard difference-in-differences regressions find no statistically significant average treatment effect of public expenditures through ALMRP II on child malnutrition levels in Kenya. Our stochastic dominance difference-in-
differences estimation allows us to look beyond the mean impact to tease out program effects that differ across the distribution of nutrition changes. For all MUAC Z-scores summary statistics, intervention sublocations had fewer negative changes over time than the control sublocations. While the data do not enable us to identify causality cleanly the results suggest that additional public expenditures under the ALRMP II project were associated with less deterioration in children’s nutritional status among the worst-off children, thus, effectively functioning as a nutritional safety net.

Beyond the specific Kenyan context we study, our stochastic dominance in differences method could generate substantive new insights on the effects of various interventions on orderable outcome indicators virtually anywhere in the world, not just of decentralized government investments on child acute malnutrition. These could be health and nutrition outcomes of the sort we study, or poverty outcomes of various sorts. Replication of this analysis in other countries similarly pursuing decentralization of interventions aiming to improve public health sector could also be beneficial, as evaluations of such programs to date have focused predominantly on the decentralization process itself, such as standards and regulations, rather than changes in health outcomes resulting from decentralization (Zon, Pavlova, Drabo, & Groot, 2017).

Moreover, our new method could prove useful more broadly whenever one is concerned with the distributional implications of a policy or policy change or, more specifically, how it affects the least well-off that are often of particular concern and the explicit target of anti-poverty policies. It can also be used to examine the heterogeneity of impacts of RCTs or natural policy experiments, including whether treatments may harm some individuals. For example, are school feeding programs (primarily) reaching the neediest? Does index-based agricultural insurance help protect the human capital of the poorest farmers? Do targeted or universal food price subsidies protect the most nutritionally vulnerable
children? Any evaluation where one is especially concerned about those who are least well off at baseline could in principle apply a variant of this stochastic dominance in differences method.
References


Online Appendix 1 – Sample size and MUAC Z-score statistics by district

Table A1 Sample size by financial year (July-June) and district

<table>
<thead>
<tr>
<th>Year</th>
<th>Garissa</th>
<th>Kajiado</th>
<th>Laikipia</th>
<th>Mandera</th>
<th>Marsabit</th>
<th>Nyeri</th>
<th>Mwingi</th>
<th>Narok</th>
<th>Tharaka</th>
<th>Turkana</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/06</td>
<td>16,517</td>
<td>9,974</td>
<td>15,243</td>
<td>17,437</td>
<td>10,921</td>
<td>14,805</td>
<td>19,165</td>
<td>4,837</td>
<td>18,607</td>
<td>36,626</td>
</tr>
<tr>
<td>2008/09</td>
<td>4,623</td>
<td>13,541</td>
<td>8,184</td>
<td>3,042</td>
<td>8,079</td>
<td>15,044</td>
<td>11,091</td>
<td>10,880</td>
<td>7,767</td>
<td>42,979</td>
</tr>
</tbody>
</table>

Table A2 Median MUAC Z-score by financial year (July-June) and district

<table>
<thead>
<tr>
<th>Year</th>
<th>Garissa</th>
<th>Kajiado</th>
<th>Laikipia</th>
<th>Mandera</th>
<th>Marsabit</th>
<th>Mwingi</th>
<th>Narok</th>
<th>Nyeri</th>
<th>Tharaka</th>
<th>Turkana</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/06</td>
<td>-1.51</td>
<td>-1.06</td>
<td>-.66</td>
<td>-1.53</td>
<td>-1.32</td>
<td>-1.23</td>
<td>-1.4</td>
<td>-.66</td>
<td>-.97</td>
<td>-1.34</td>
</tr>
<tr>
<td>2008/09</td>
<td>-.76</td>
<td>-1.21</td>
<td>-.76</td>
<td>-1.17</td>
<td>-1.22</td>
<td>-1.04</td>
<td>-1.18</td>
<td>-.66</td>
<td>-.77</td>
<td>-1.36</td>
</tr>
</tbody>
</table>

Table A3 10th percentile MUAC Z-score – whole sample

<table>
<thead>
<tr>
<th>Year</th>
<th>Garissa</th>
<th>Kajiado</th>
<th>Laikipia</th>
<th>Mandera</th>
<th>Marsabit</th>
<th>Mwingi</th>
<th>Narok</th>
<th>Nyeri</th>
<th>Tharaka</th>
<th>Turkana</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/06</td>
<td>-2.40</td>
<td>-2.14</td>
<td>-1.75</td>
<td>-2.65</td>
<td>-2.33</td>
<td>-2.36</td>
<td>-2.55</td>
<td>-1.67</td>
<td>-1.87</td>
<td>-2.26</td>
</tr>
<tr>
<td>2008/09</td>
<td>-1.88</td>
<td>-2.22</td>
<td>-2.10</td>
<td>-2.13</td>
<td>-2.29</td>
<td>-2.14</td>
<td>-2.35</td>
<td>-1.54</td>
<td>-1.74</td>
<td>-2.25</td>
</tr>
</tbody>
</table>

Table A4 25th percentile MUAC Z-score – whole sample

<table>
<thead>
<tr>
<th>Year</th>
<th>Garissa</th>
<th>Kajiado</th>
<th>Laikipia</th>
<th>Mandera</th>
<th>Marsabit</th>
<th>Mwingi</th>
<th>Narok</th>
<th>Nyeri</th>
<th>Tharaka</th>
<th>Turkana</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/06</td>
<td>-1.97</td>
<td>-1.67</td>
<td>-1.16</td>
<td>-2.06</td>
<td>-1.79</td>
<td>-1.84</td>
<td>-1.96</td>
<td>-1.20</td>
<td>-1.45</td>
<td>-1.85</td>
</tr>
<tr>
<td>2008/09</td>
<td>-1.45</td>
<td>-1.76</td>
<td>-1.40</td>
<td>-1.69</td>
<td>-1.69</td>
<td>-1.68</td>
<td>-1.76</td>
<td>-1.15</td>
<td>-1.28</td>
<td>-1.86</td>
</tr>
</tbody>
</table>
**Online Appendix 2**

Balance tests of NDVI, the main covariate used in the differences-in-differences regression and in the stochastic dominance analysis.

<table>
<thead>
<tr>
<th></th>
<th>Control sublocations</th>
<th>Intervention sublocations</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean/SE</td>
<td>N</td>
</tr>
<tr>
<td><strong>NDVI</strong></td>
<td><strong>2005/06</strong></td>
<td>52</td>
<td>0.271</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.020]</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>2008/09</strong></td>
<td>52</td>
<td>0.271</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.020]</td>
<td></td>
</tr>
</tbody>
</table>

*** indicates statistical significance at the 1% level.
Online Appendix 3 - Second order stochastic dominance tests

This online appendix provides second order dominance (SOSD) curves to complement the first order stochastic dominance (FOSD) in figures 2-5 in the paper.

SOSD graphs require extra care as the variable of interest is MUAC Z-scores changes. Since changes can be negative (in contrast to levels of Z-scores) and positive this has implications for the plotting of second order dominance curves. Using the same data of MUAC Z-score changes that was used for FOSD in figures 2-3 the SOSD graph looks like this and is a bit harder to interpret:

SOSD graph that corresponds to FOSD in figure 2

An alternative is to transpose the distribution of MUAC Z-score changes so that all of them are positive. The figure below replicates the figure just above, except +1 is added to all MUAC Z-scores (the biggest negative change in the data was -0.841 in the figure above; and is +0.159 in the figure below). The figures below show SOSD in a more familiar manner. However, the cut-off point between positive and negative MUAC Z-scores changes is now +1 (rather than 0 in the figure above).
SOSD graph (with MUAC Z-score distribution shifted by +1) that corresponds to FOSD graph in figure 2

SOSD Difference Intervention vs. Difference Control
2005/06-2008/09. drought adjusted.

SOSD graph (with MUAC Z-score distribution shifted by +1) that corresponds to FOSD graph in figure 3

SOSD Difference Intervention vs. Difference Control
2005/06-2008/09. drought adjusted.
SOSD graph (with MUAC Z-score distribution shifted by +1) that corresponds to FOSD graph in figure 4

![SOSD Difference Intervention vs. Difference Control](image1)

SOSD graph (with MUAC Z-score distribution shifted by +1) that corresponds to FOSD graph in figure 5

![SOSD Difference Intervention vs. Difference Control](image2)

In sum, the above figures suggest there is SOSD (as also reported in table 3) even though it is not statistically significant, which is not necessarily surprising given the small size of the sublocation panel, the short time period and the relatively modest investments.
Figures and Tables
Figure 1 ALRMP II Project Districts
Figure 2

FOSD Difference Intervention vs. Difference Control
2005/06-2008/09. drought adjusted.

% of sublocations

-1 -0.5 0 0.5 1 1.5 2
difference in median MUAC Z-score (for all observations with MUAC<0)

-1 -0.5 0 0.5 1 1.5 2

Difference for Control
Difference for Intervention

Figure 3

FOSD Difference Intervention vs. Difference Control
2005/06-2008/09. drought adjusted.

% of sublocations

-1 -0.5 0 0.5 1 1.5 2
difference in median MUAC Z-score (for all observations with MUAC<0)

Confidence interval (95 %) Estimated difference
Figure 4

**FOSD Difference Intervention vs. Difference Control**

2005/06-2008/09. drought adjusted.

<table>
<thead>
<tr>
<th>% of sublocations</th>
<th>Difference for Control</th>
<th>Difference for Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5

**FOSD Difference Intervention vs. Difference Control**

2005/06-2008/09. drought adjusted.

<table>
<thead>
<tr>
<th>% of sublocations</th>
<th>Difference for Control</th>
<th>Difference for Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Difference for Control

Difference for Intervention
### Table 1 Sample size and selected MUAC Z-score statistics for intervention and control sublocations

<table>
<thead>
<tr>
<th>Year</th>
<th>Sample size</th>
<th>Median MUAC Z-score</th>
<th>25th percentile MUAC Z-score</th>
<th>10th percentile MUAC Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>intervention</td>
<td>control</td>
<td>intervention</td>
<td>control</td>
</tr>
<tr>
<td>2005/06</td>
<td>83,678</td>
<td>69,925</td>
<td>-1.22</td>
<td>-1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-1.80</td>
<td>-1.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-2.31</td>
<td>-2.14</td>
</tr>
<tr>
<td>2008/09</td>
<td>63,037</td>
<td>54,421</td>
<td>-1.15</td>
<td>-1.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-1.70</td>
<td>-1.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-2.16</td>
<td>-2.12</td>
</tr>
</tbody>
</table>

Note: All differences between intervention and control MUAC Z-scores are significant at the 1% level.

### Table 2 Difference-in-differences Panel Regression

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) median of MUAC Z &lt;$0$</th>
<th>(2) 10th percentile</th>
<th>(3) 25th percentile</th>
<th>(4) median of MUAC Z &lt; -1</th>
<th>(5) median of MUAC Z &lt; -2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention dummy based on ALRMP investment</td>
<td>0.0735 (0.0594)</td>
<td>0.0832 (0.0784)</td>
<td>0.0661 (0.0701)</td>
<td>0.0793 (0.0557)</td>
<td>0.0531 (0.0342)</td>
</tr>
<tr>
<td>Change in NDVI 2005/06-08/09</td>
<td>1.308* (0.592)</td>
<td>2.611*** (0.647)</td>
<td>2.058*** (0.600)</td>
<td>0.927* (0.505)</td>
<td>0.768* (0.384)</td>
</tr>
<tr>
<td>Sq change in NDVI 2005/06-08/09</td>
<td>-12.91** (4.986)</td>
<td>-8.672 (5.296)</td>
<td>-12.70* (5.647)</td>
<td>-0.954 (3.686)</td>
<td>1.924 (2.606)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.501*** (0.0374)</td>
<td>0.892*** (0.0469)</td>
<td>0.839*** (0.0418)</td>
<td>0.203*** (0.0319)</td>
<td>0.120*** (0.0257)</td>
</tr>
</tbody>
</table>

Observations: 114, 114, 114, 114, 106

R-squared: 0.319, 0.299, 0.297, 0.249, 0.280

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

District dummy variables included.
Table 3 Summary of Stochastic Dominance Results – Difference in intervention vs. control sublocations

<table>
<thead>
<tr>
<th>Median MUAC of obs &lt; 0</th>
<th>% below -1 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dominance across entire domain</td>
</tr>
<tr>
<td></td>
<td>Dominance across entire domain</td>
</tr>
<tr>
<td>FOSD</td>
<td>No</td>
</tr>
<tr>
<td>SOSD</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Median MUAC of obs &lt; 0</th>
<th>% below -1 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dominance across entire domain</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>