A stochastic dominance approach to program evaluation with an application to child nutritional status in arid and semi-arid Kenya

Felix Naschold and Christopher B. Barrett

Cornell University
Department of Applied Economics and Management
313a Warren Hall
Ithaca, NY 14853
fn23@cornell.edu

30 April 2010

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2010

AAEA & ACCI Joint Annual Meeting, Denver, Colorado, July 25-27, 2010

JEL classifications: I32,O22, H51

Keywords: Program evaluation, stochastic dominance, child nutrition, Kenya

Word Count: 5390 words

We would like to thank the Arid Lands Resource Management Project of the Government of Kenya for making available the data. Financial support for this work was provided under the ILRI-Cornell Collaborative Work Agreement. Any views expressed and any remaining errors are ours alone.

Introduction

Existing program evaluation methods such as difference-in-difference estimators are designed to examine the overall impact of a program. By design they can only examine changes in a particular summary statistic of an outcome indicator, most commonly the mean or the median or a particular quantile. 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 examine the 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 is 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 to sample size issues in the regressions. To circumvent these problems this paper suggests a novel approach to program evaluation which combines stochastic dominance with difference-in-difference methods.

We apply this new method to 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). In particular, we focus on the differences in changes in nutritional status between areas that have benefited from additional public expenditures through the second phase of the Arid Lands Resource Management Project (ALRMP II) and areas that have not. This paper is 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.

Acute malnutrition remains pervasive in arid and semi-arid Kenya between 2005 and 2009. Using standard difference-in-difference 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 correlated 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.

These findings highlight the importance of looking beyond average impacts. Looking beyond averages has become more mainstream in poverty analysis and has yielded more nuanced insights ((Ravallion 2001), the increasing use of higher order P-alpha indices). The stochastic dominance based difference-in-difference technique proposed in this paper 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 Δ

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

but cannot because we only see 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 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 Δ is equal to

$$\Delta = E\left[x_T\right] - E\left[x_C\right] \qquad (1)$$

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

$$\Delta = E\left[x_{T,t} - x_{T,t-1}\right] - E\left[x_{C,t} - x_{C,t-1}\right] \qquad (2)$$

The key shortcoming of any of the 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. To look beyond the average treatment effect we need a different evaluation method. This paper proposes one such method based on stochastic dominance.

Using stochastic dominance for difference-in-difference estimation

Stochastic dominance analysis takes account of entire distributions or sub-ranges of distributions. There are two ways in which it superior and more robust for making welfare comparisons across space or across time.

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. Though 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 on. Since the location of a poverty line *z* is arbitrary it is often contentious. Instead, it is often much easier to agree on a range in which the poverty line should be set such that $z \in \left[z_{\min}, z_{\max}\right]$ where $z_{\min}$ and $z_{\max}$ 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. Minus 1 and minus 2 are often regarded as the cut-off points for mild and severe malnutrition. 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 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 really any alternative indicators to the Z-score based malnutrition measure. However, when evaluating material poverty there is often disagreement on 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.

Definitions of orders of dominance

Let $\mathfrak{F}$ denote a set of probability density functions of a random variable *x* defined on a closed interval [x_{min}, x_{max}]. Further, let $f_A(x) \in \mathfrak{F}$ and $f_B(x) \in \mathfrak{F}$. 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 $z \in [x_{\min}, x_{\max}]$ if and only if (iff) $F_B(x) - F_A(x) \geq 0 \ \forall x \in [x_{\min}, z]$, 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* denote the order of integration. Then $F^s(x) = \int_{x_{\min}}^{x_{\max}} F^{s-1}(z)\,dz$. Therefore, distribution *A* s^{th}-order dominates distribution *B* iff

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

These standard stochastic dominance criteria can be applied directly to program evaluation if treatment and control populations share the same initial distribution.

SD, poverty orderings and social welfare orderings

Program evaluations that focus on poverty and social welfare impacts can exploit some convenient symmetries between stochastic dominance and poverty orderings. Poverty indicators, here, are loosely defined as any (quasi-)continuous measure of well-being including consumption, assets or anthropometric measures such as the MUAC Z-scores used in our application below.

Stochastic dominance of order α is directly related to P-α poverty measures (Foster and Shorrocks 1988) in the following way. Let $\boldsymbol{SD_s}$ denote stochastic dominance of order *s* and $\boldsymbol{P_\alpha}$ stand for poverty ordering ('has less poverty'). Let $\alpha = s - 1$ .Then,

A $\boldsymbol{P_\alpha}$ B iff A $\boldsymbol{SD_s}$ B.

The poverty ordering is the same as the stochastic dominance ordering. And poverty orderings are nested in the same way as stochastic dominance orderings. Let → denote 'implies. Then, for stochastic dominance orderings A $\boldsymbol{SD_1}$ B → A $\boldsymbol{SD_2}$ B → A $\boldsymbol{SD_3}$B. Similarly, for poverty orderings A $\boldsymbol{P_1}$ B → A $\boldsymbol{P_2}$ B → A $\boldsymbol{P_3}$ B. If one welfare distribution has unambiguously less welfare according the headcount ratio (P_1), then it also has less welfare according to the gap (P_2) and the gap squared (P_3) indices.

Foster and Shorrocks (1988) show how these orderings can be expanded to social welfare functions. Let U(F) be the class of symmetric utilitarian welfare functions. Then,

$$A \ \boldsymbol{P_\alpha} \ B \text{ iff } A \ \boldsymbol{U_\alpha} \ B.$$

Define U_1 as the subset of U for which $u'>0$. U_1 represents the monotonic utilitarian welfare functions. Less malnutrition is better, regardless for whom.[1] Let U_2 be a subset of U_1 such that $u''<0$. This subset of social welfare functions represents equality preference in that a mean preserving progressive transfer increases U_2. Finally, define U_3 as the subset of U_2 for which $u'''>0$. U_3 contains the transfer sensitive social welfare functions which value a transfer more highly the lower in the distribution it occurs.

Thus, using stochastic dominance analysis on welfare data we can identify social welfare changes for nested classes of welfare measures. Also, in the context of comparing levels of welfare it makes sense to test for at least third order stochastic dominance as transfer sensitivity is generally desirable.

The program evaluation literature has evolved separately from the stochastic dominance literature. Reviews of the state-of-the-art in program evaluation (Todd 2008) and best practice guides (Baker 2000) do not contain any reference to stochastic dominance. To date Verme (2010) is the only study that has started to show 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. On the basis of the Foster and Shorrocks results he proposes a simple method for program evaluation for the case of randomized assignment of treatment.

This paper extends the method to difference-in-difference evaluation to make it applicable to cases where treatment and control populations do not share the same distribution. It also provides the first empirical application of this technique highlighting the importance to look beyond average treatment effects.

Stochastic dominance for difference-in-difference impact evaluation

In the majority of evaluation problems the available data is not based on experimental or quasi-experimental data. Treatment and control groups are not randomly selected and are, thus, likely to differ in their intrinsic characteristics. Therefore, we cannot look at the simple difference in outcomes between the two groups but need to examine differences-in-differences in outcomes across time and across subgroups. The difference-in-difference approach can be applied in a stochastic dominance context. Much of the discussion on stochastic dominance on simple differences from above carries straight over but there are important difference in interpretation and usefulness of higher order dominance tests.

[1] Again, with a nutritional indicator this is only defensible up to a certain point, but certainly up to x_{max}.

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

Further, let $g_A(\Delta) \in \mathfrak{G}$ and $g_B(\Delta) \in \mathfrak{G}$. Denote the respective cumulative density functions (cdf) by $G_A(\Delta)$ and $G_B(\Delta)$. Then, distribution *A* first order stochastically dominates (FOD) distribution *B* iff $G_B(\Delta) - G_A(\Delta) \geq 0 \ \forall \Delta \in \left[\Delta_{\min}, \Delta_{\max}\right]$, 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 as the location of such a 'poverty line' in differences is even more subjective than a poverty line in levels. The use of such a cut-off point depends on the particular focus of the evaluation. For instance, we could focus on negative changes to determine whether the treatment or control group had fewer negative changes. The corresponding FOD condition would be

$G_B(\Delta) - G_A(\Delta) \geq 0 \ \forall \Delta \in \left[\Delta_{\min}, z\right]$. 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* denote the order of integration. Then $G^s(\Delta) = \int_{\Delta_{\min}}^{\Delta_{\max}} G^{s-1}(z)\,dz$. Therefore, distribution A s^{th} order dominates distribution B iff $G_B^s(\Delta) - G_A^s(\Delta) \geq 0 \ \forall \Delta \in \left[\Delta_{\min}, \Delta_{\max}\right]$.

There is an important difference in interpreting the results from SD on welfare levels versus on changes. 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 of individuals. Indeed, the largest negative changes are likely to be from people who were relatively well off at *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. 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.

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 of the domain regardless of the welfare level. Hence, the attributes of second and third order dominance, namely equality preference and transfer sensitivity, no longer apply in the same way as for the stochastic 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 in MUAC Z-score summary statistics are concentrated at the lower end of the distribution of *changes*. Third order SD tests, however, are not really meaningful. 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.

Methodologically, changes in these MUAC Z-score summary statistics are analogous to changes in incomes. Hence, we can draw on the literature on economic mobility. However, in this literature the term 'economic mobility' is implicitly or explicitly defined in at least six different ways (Fields 2001; Fields 2007). The mobility definition that is most appropriate for analyzing MUAC Z-score changes is that of directional MUAC movement[2], as we want to capture both the magnitude and the direction of MUAC Z-score changes over time, and capture them in absolute, not relative terms, that is, irrespective of what happened to other changes in MUAC Z-scores of other locations.

There is no meaningful range of sensible 'poverty lines' expressed 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 difference-in-difference estimators. In the standard regression program evaluation approach we can include other covariates as right hand side variables. In practice this doesn't matter if we are primarily interested in whether the program has had an effect or not. Furthermore, our stochastic dominance method can be used to evaluate program impact net of other covariates; it just can't do it simultaneously with estimating the program impact. To account for covariates we first run a regression of the outcome variable on the desired covariates before using the residuals, which represent the variation in the outcome variable net of observables, in the stochastic dominance estimation. In the application below we use this method to strip out the effect of drought on child malnutrition by using the residuals of a regression of MUAC Z-scores on NDVI as our variable in the stochastic dominance analysis.

The setting and data

To illustrate the use of stochastic dominance for difference-in-difference evaluation we use a unique, large dataset of child nutrition from arid and semi-arid lands (ASALs) Kenya. These areas are characterized by livestock production and highest incidences of poverty in Kenya. Over 60% of the population live below the poverty line and levels of access to basic services are very low. Infant mortality rates are high, in some districts more than double the (already high) national average. Child malnutrition levels in Kenyan ASALs are generally declining but are still above emergency threshold levels, worsened by recurrent droughts, high poverty rates, and HIV/AIDS (UNICEF, xx). In the North Eastern Province, for example, 23.2 per cent of children under five suffering from acute malnutrition and infant and under-five mortality rates are rising (ibid).

The data we use in our illustration below were collected by the Kenyan government under the second phase of the Arid Lands Management Project (ALRMP II), a community-based drought management initiative that provided additional, decentralized financial resources to 28 arid and semi arid districts in Kenya from 2003 to 2010. The project sought to improve the effectiveness of emergency drought

[2] Other economic mobility concepts relate to movements in ranks, in shares, and in symmetric income. For our MUAC analysis we are not concerned with these.

response while at the same time reducing vulnerability, empowering local communities, and raising the profile of arid and semi-arid areas in national policies and institutions.

Since one of the objectives of ALRMP II was to reduce the levels of child malnutrition the project's monitoring strategy included the collection of information on child nutritional stats. The specific anthropometric indicator collected was the Mid-Upper Arm Circumference (MUAC) measurement 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 (Shakir and Morley 1974; Shakir 1975). In addition is considered more appropriate than other measures for children in pastoral areas (REF).

We use MUAC Z-scores rather than absolute MUAC measures as they 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 nutrition status. For some reason, perhaps inertia from when MUAC Z-scores were difficult to calculate, even recent 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 (Gernaat *et al.* 1996; de Onis *et al.* 1997; Mei *et al.* 1997).

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

$$Z(MUAC_{ijt}) = \frac{MUAC_{ijt} - \overline{MUAC(reference\ population)}}{\sigma_{MUAC(reference\ population)}}$$

where $MUAC_{it}$ is child *i*'s MUAC at time *t* in location *j* and σ indicates the standard deviation. The reference population is taken from the WHO/NCHS (de Onis *et al.* 1997).

Over 602,000 individual child MUAC measurements were taken in 128 sublocations in 10 arid and semi-arid ALRMP II districts between June 2005 and August 2009. Table 1 shows the sample size by financial year and district. There is some variation in coverage across districts. Turkana accounts for around a quarter of all observations, while there are only 27,000 observations for Mandera, including none for 2007/08.

Table 1 Sample size by financial year (July-June) and district

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

Table 2 Median MUAC Z-score by financial year (July-June) and district

year	Garissa	Kajiado	Laikipia	Mandera	Marsabit	Mwingi	Narok	Nyeri	Tharaka	Turkana
2005/06	-1.51	-1.06	-.66	-1.53	-1.32	-1.23	-1.4	-.66	-.97	-1.34

2008/09	-.76	-1.21	-.76	-1.17	-1.22	-1.04	-1.18	-.66	-.77	-1.36

Tables 2 depicts the high prevalence of malnutrition by showing the median MUAC Z-score by district. Moreover, according to the median Z-score nutritional status does not seem to change perceptibly over time. The severity of malnutrition is evident from tables 3 and 4 which present the 10^{th} and 25^{th} percentile MUAC Z-score for each district and year. With the exception of Nyeri, and possibly Laikipia, 10 percent of children have a MUAC of less than -2 standard deviations indicating severe malnutrition. Even the 25^{th} percentile figures from table 6 are closer to the -2 cut-off point than the -1 standard deviation level that indicates mild malnutrition.

Table 3 10^{th} percentile MUAC Z-score – whole sample

Year	Garissa	Kajiado	Laikipia	Mandera	Marsabit	Mwingi	Narok	Nyeri	Tharaka	Turkana
2005/06	-2.4	-2.14	-1.75	-2.65	-2.33	-2.36	-2.55	-1.67	-1.87	-2.26
2008/09	-1.88	-2.22	-2.1	-2.13	-2.29	-2.14	-2.35	-1.54	-1.74	-2.25

Table 4 25^{th} percentile MUAC Z-score – whole sample

year	Garissa	Kajiado	Laikipia	Mandera	Marsabit	Mwingi	Narok	Nyeri	Tharaka	Turkana
2005/06	-1.97	-1.67	-1.16	-2.06	-1.79	-1.84	-1.96	-1.2	-1.45	-1.85
2008/09	-1.45	-1.76	-1.4	-1.69	-1.69	-1.68	-1.76	-1.15	-1.28	-1.86

To estimate changes over time and compare them between intervention and control sublocations 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.

For the 128 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. 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 severe malnutrition.[3]

[3] 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

These particular summary statistics are sensible truncations of the MUAC Z-scores distribution since we want to focus on undernourished children. This right-truncation in these summary statistics is analogous to the focus axiom in poverty measurement. We can safely ignore level and changes at higher levels of MUAC Z-scores since high MUAC observations and large positive changes at the upper tail of the distribution are not necessarily desirable or positive. Unlike income or consumption, in the context of child nutrition more is not always better.

We classified sublocations into intervention and control groups according to the cumulative ALRMP II investment data provided by the ALRMP district data managers.[4] The distribution of project investments suggests a natural cut-off point with sublocations without any sublocation specific investment forming the control locations and sublocations with some investment the intervention locations.

Results

We present program evaluation results for both the difference-in-difference regressions and for stochastic dominance to highlight the potential practical importance of looking beyond the average treatment effect.

Regression results

The difference-in-difference estimator in equation 2 was estimated as

$$\Delta MUAC\_SS_{j} = \gamma_0 + \gamma_1 D_j + \gamma_2 NDVI_j + \gamma_3 NDVI_j^2 + \sum_{l=2}^{L} \delta_l L_l + \varepsilon_j$$

where $\Delta MUAC\_SS_{jt}$ is the change in a MUAC summary statistic for sublocation *j*, NDVI is the normalized difference vegetation index, and L_l are district dummy variables to capture regional variation. The first row of p-values in table 5 shows that none of the five MUAC Z-score summary statistics had a significant average treatment effect. This suggests that on average there has been no impact of ALRMP expenditure on child nutrition levels.

with Z-score below 0, -1, and -2; the Z-score gap of children with Z-score below -1 and -1; and the squared Z-score gap of children with Z-score below -1 and -2.Results for the additional indicators are available on request.

[4] As a robustness check we also asked district project managers to classify sublocations in their districts into treatment and control groups. For brevity we focus on the investment-based treatment and control classifications.

Table 5 Diff-in-diff Panel Regression: Sublocation Summary Statistics of MUAC z-score

VARIABLES	(1) median of MUAC Z <0	(2) 10th percentile	(3) 25th percentile	(4) median of MUAC Z <-1	(5) median of MUAC Z <-2
intervention dummy based on ALRMP investment	0.0735	0.0832	0.0661	0.0793	0.0531
	(0.248)	(0.316)	(0.371)	(0.188)	(0.155)
change in NDVI 2005/06-08/09	1.308*	2.611***	2.058***	0.927*	0.768*
	(0.0545)	(0.00294)	(0.00754)	(0.0997)	(0.0767)
squared change in NDVI 2005/06-08/09	-12.91**	-8.672	-12.70*	-0.954	1.924
	(0.0293)	(0.136)	(0.0510)	(0.802)	(0.479)
Constant	0.501***	0.892***	0.839***	0.203***	0.120***
	(2.99e-07)	(1.40e-08)	(8.70e-09)	(0.000133)	(0.00114)
Observations	114	114	114	114	106
R-squared	0.319	0.299	0.297	0.249	0.280

Robust p-values in parentheses

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

District dummy variables included.

To test whether the lack of significance might be due to only having 93 differenced observations in the sublocation pseudo panel we ran the following difference-in-difference estimation on the individual MUAC Z-scores

$$Z(MUAC_{ijt}) = \beta_0 + \beta_1 T_t + \beta_2 D_j + \beta_3 (T_t * D_j) + \beta_4 NDVI_{jt} + \varepsilon_{ijt}$$

This increased the sample size to more than 270,000 but still shows no statistically significant average treatment effect as shown by the 'diff-in-diff' p-value.

Table 6 Difference in difference regression of individual MUAC z-scores

2005/06-2008/09 Diff-in-diff Regression - Dependent Variable: individual MUAC Z-score

VARIABLES	
time dummy (=1 for 2008/09)	0.0785
	(0.290)
control - intervention by investment	-0.0576
	(0.425)
Diff in diff	0.0245
	(0.782)
Normalized Difference Vegetation Index	1.029***
	(6.25e-07)
Constant	-1.391***
	(0)
Observations	271061
R-squared	0.033

Robust p-values in parentheses

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

Categorization of treatment and control sublocations by investment data.

Stochastic dominance results

Our stochastic dominance analysis proceeds in three steps.[5] The first two steps represent standard single-difference stochastic dominance tests. First, we test for SD within control and treatment over time. If outcomes in the intervention sites improved while those in the control sites worsened this would present evidence for a positive ALRMP II program effect. However, we find no difference

[5] Full results for the first two steps testing for stochastic dominance within treatment and control groups across time and across treatment and control sublocations at each point in time are available on request.

between trends in intervention and control sublocation. Both have improved slightly, so on the basis of these tests we cannot conclude that ALRMP II has had no impact. Second, we compare intervention and control sublocations before and after ALRMP II. Control sublocations dominate in most cases and interventions sites never dominate. Again, this does not indicate any program effect. Third, we apply the difference-in-difference method outlined above to test for stochastic dominance between changes in intervention and changes in control sublocations. Results are summarized in table 7.

Table 7 Summary table of Stochastic Dominance Results – Difference in intervention vs. differences in control sublocations

	Median MUAC of obs < 0			***% below -1 SD***		
	Dominance	Which*	Signif.	Dominance	Which**	Signif.
FOSD	N	-	NS	N	-	NS
SOSD	Y?	Intervention	NS	Y	Intervention	NS

* Lower curves to the right are dominate for these indicators for which a greater number indicates 'better'
**For changes from 2005/06-2008/09 in part III. larger positive changes are better, so lower curves to the right dominate.

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 1. However, below 0.2 interventions 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. This suggests that ALRMP intervention sites were more effective in preventing a worsening of nutritional status, even if in absolute nutrition levels intervention sites still lag behind control sites. Above 0.2 the two cdfs are fairly close and intersect repeatedly indicating that treatment and control sites had roughly equal proportions of sites that experienced equal improvements child nutritional levels over time.

Figure 2 indicates that the partial first order stochastic dominance is statistically significant around zero and almost significant below zero. Given the small size of the sublocation pseudo panel, the short time period and the relatively modest investments the lack of greater statistically significance is not surprising.[6]

[6] For all of the other stochastic dominance tests shown in the appendix table where we could use the individual data to complement the sublocation pseudo panel, the results of the panel and the results of the individual data always matched with the latter always statistically significant. This suggests that sample size is the limiting factor in the

Figure 1

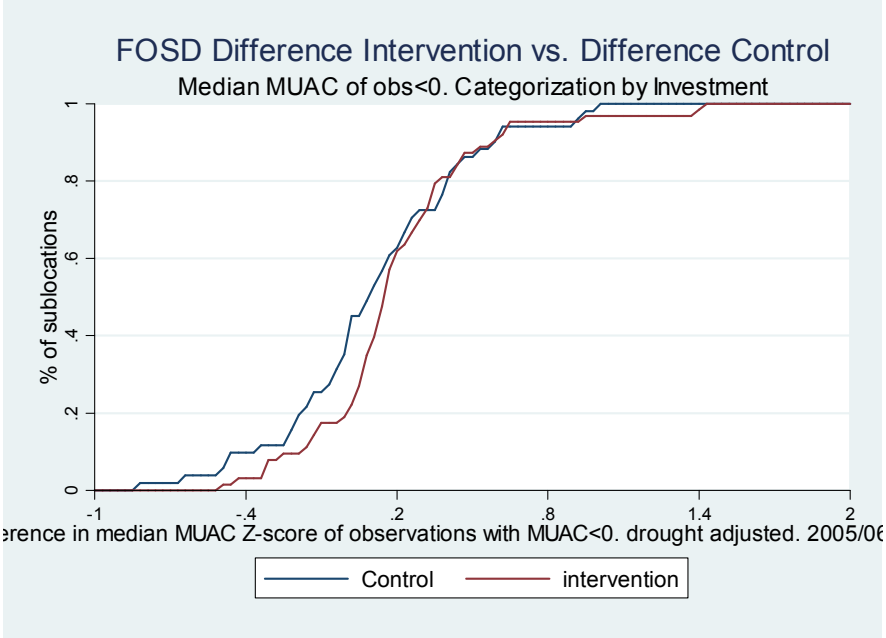
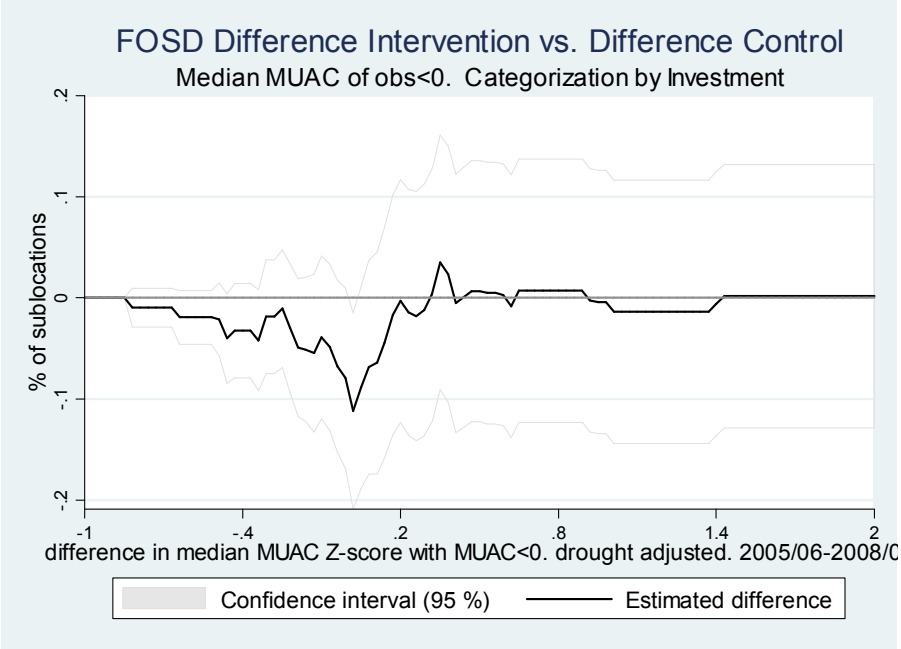


Figure 2



Figures 3 and 4 show the results for changes in MUAC Z-scores for the 25th and 10th percentile, respectively. As we focus on smaller and smaller percentiles of the distribution (from the median to the 25th to the 10th percentile) the analysis concentrates increasingly on the worst-off kids.

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 pseudo panel result.

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 25th percentile subsample in figure 3 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 10th percentile in Figure 4 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 10th percentile there were also fewer smaller positive changes. Again, these results are not statistically significant[7], likely a result of the small sample size of the sublocation panel dataset.

[7] Difference figures with confidence bands (in the style of figure 2) omitted for brevity. I am considering to take out the discussion of the 25th and 10th percentile, anyway.

Figure 3

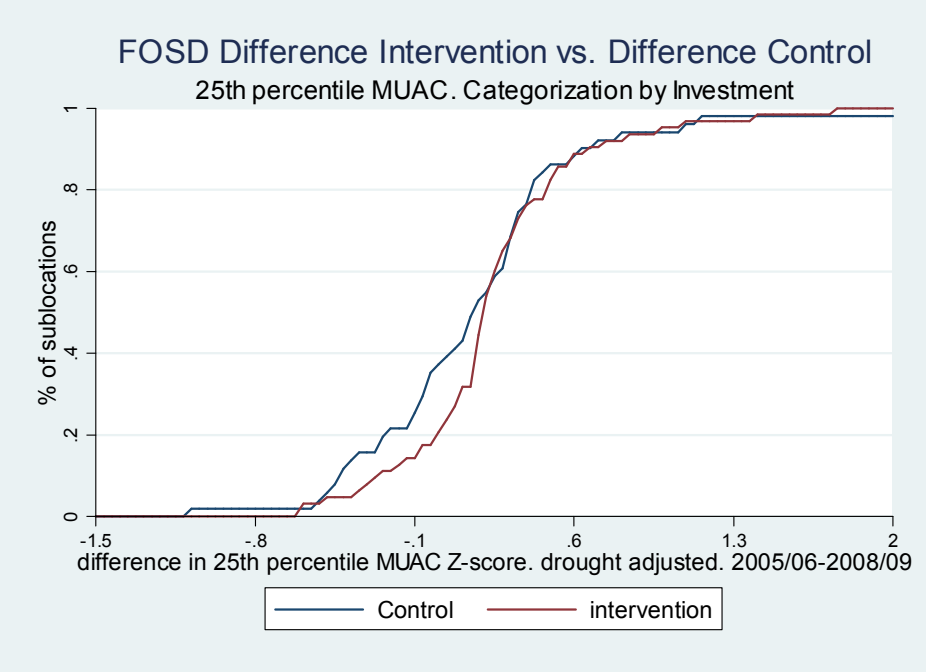
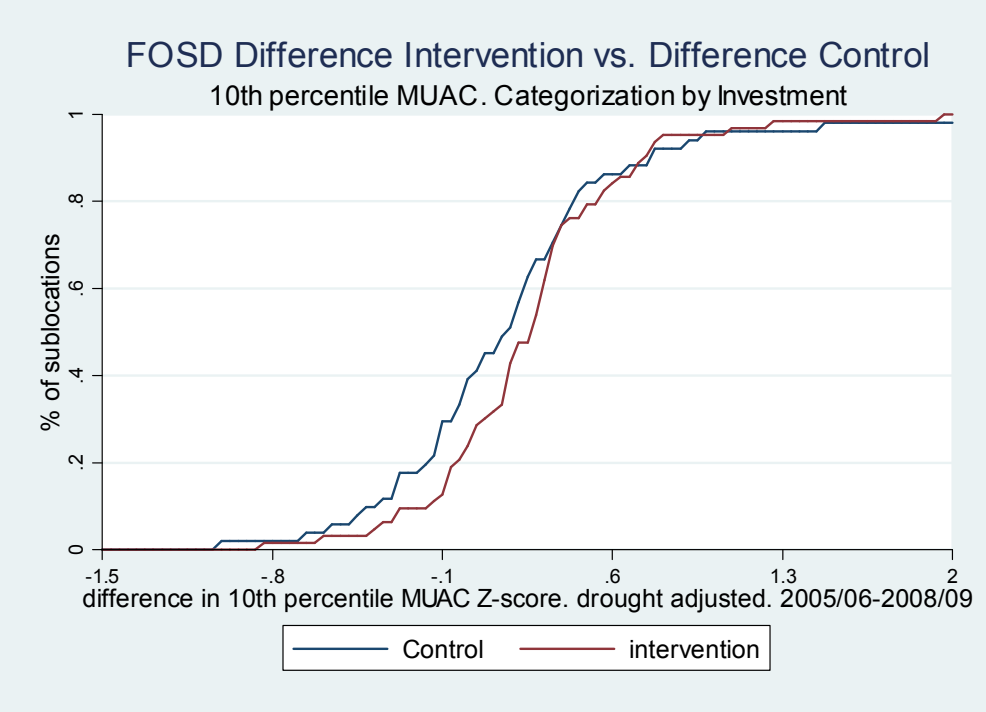


Figure 4



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 in the impact across various parts of the distribution. This paper has proposed a new method to evaluate program impacts across the entire distribution of outcomes. The method does not require experimental data as it applies stochastic dominance estimation to differences-in-differences across subgroups and time.

Our empirical results highlight the practical added value of this method. Standard difference-in-difference regressions find no statistically significant average effect of additional public expenditures on child malnutrition levels. The stochastic dominance difference-in-difference estimation allowed us to look beyond the mean impact and 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 allow us to identify causality the results suggest that additional public expenditures under the ALRMP II project may have prevented nutritional status from worsening for the worst-off children, thus, effectively functioning as a nutritional safety net.

References

Baker, J. L. (2000). Evaluating the impacts of development projects on poverty - A handbook for practitioners. Washington, DC, World Bank.

de Onis, M., R. Yip and Z. Mei (1997). "The development of MUAC-for-age reference data recommended by a WHO Expert Committee." Bulletin of the World Health Organization **75**(1): 11-18.

Fields, G. S. (2001). Distribution and Development: a new look at the developing world. New York, Russell Sage Foundation.

Fields, G. S. (2007). Income Mobility. Ithaca, ILR Cornell University**:** 12.

Foster, J. E. and A. F. Shorrocks (1988). "Poverty Orderings." Econometrica **56**(1): 173-177.

Gernaat, H., W. Dechering and H. Voorhoeve (1996). "Absolute values or Z scores of mid-upper arm circumference to identify wasting? Evaluation in a community as well as a clinical sample of under fives from Nchelenge, Zambia." Journal of Tropical Pediatry **42**(1): 27-33.

Mei, Z., L. Grummer-Strawn, M. De Onis and R. Yip (1997). "The development of a MUAC-for-height reference, including a comparison to other nutritional status screening indicators." Bulletin of the World Health Organization **75**(4): 333-341.

Ravallion, M. (2001). "Growth, inequality and poverty: looking beyond averages." World Development **29**(11): 1803-1815.

Ritmeijer, K. (1998). "Finding the right MUAC cut-off to improve screening efficiency." Field Exchange **4**: 24-25.

Shakir, A. (1975). "The surveillance of protein-calorie malnutrition by simple and economical means." Journal of Tropical Pediatry **21**: 69-85.

Shakir, A. and D. Morley (1974). "Measuring malnutrition." The Lancet **1**(7860): 758-759.

Todd, P. (2008). Evaluating social programs with endogenous program placement and selection of the treated. Handbook of Development Economics Vol 4. T. P. Schultz and J. Strauss. North Holland, Elsevier. **4**.

Verme, P. (2010). "Stochastic dominance, poverty and the treatment effect curve." Economics Bulletin **30**(1): 365-373.