5-2011

Empirical Forecasting of Slow-Onset Disasters for Improved Emergency Response: An Application to Kenya's Arid North

Andrew G. Mude
Cornell University

Christopher B. Barrett
Cornell University

John G. McPeak
Syracuse University

Robert Kaitho
Texas A & M University and Center for Natural Resource Information Technology

Patti Kristjanson
CCAFS

Empirical Forecasting of Slow-Onset Disasters for Improved Emergency Response: an Application to Kenya's Arid North

Andrew Mude[1], Christopher B. Barrett,

John G. McPeak, Robert Kaitho and Patti Kristjanson

November 2006

Abstract:

Mitigating the negative welfare consequences of crises such as droughts, floods, and disease outbreaks, especially in highly vulnerable areas insufficiently equipped to prevent food and livelihood security crisis in the face of adverse shocks, is a major challenge in many areas of the world. Given the finite resources allocated for emergency response, and an the expected increase in incidences of humanitarian catastrophe due to changing climactic patterns, there is a need for the development of rigorous and efficient methods of early warning and emergency needs assessment. In this paper we develop an empirical model, based on a relatively parsimonious set of regularly measured variables from communities in Kenya's Arid North, that generates sufficiently accurate forecasts of the likelihood of famine with at least three months lead time. While several early warning and emergency needs assessment guides exist, our empirical forecasting method has the advantage of demonstrable statistical rigor and out-of-sample performance. Such a forecasting model is an invaluable tool for emergency awareness and response needs, offering rigorous, cost-effective and practical early warning capacity.

---

[1] Corresponding author. Email: amude@cgiar.org

1. INTRODUCTION

The ability to forecast, with reasonable accuracy, the onset, duration and severity of events such as droughts, floods, and disease outbreaks, especially in terms of their prospective human welfare effects, is critical to the design of rapid, timely and cost-effective emergency response systems that can minimize the suffering of populations adversely affected by such disasters. As the consensus on climate change and its consequences grows, there is an increasing worry that climate variability is growing and that the frequency of such disasters will continue to rise (IPCC, 2001). Incidences of humanitarian catastrophe and the ensuing demand for emergency response are therefore widely expected to increase.

Given the finite resources allocated for emergency response initiatives, there is growing demand worldwide for the development of rigorous, efficient and practical methods of early warning and emergency needs assessment. The famine early warning systems network (FEWS-NET), funded by USAID, is a comprehensive effort to provide timely assessments on the impending status of food insecurity threats in 20 African countries, as well as in Afghanistan and Haiti. They are also tasked with strengthening regional and national early warning and food security organizations through networking and capacity development. The USAID Global Livestock Collaborative Research Support Program (GL-CRSP) Livestock Early Warning System (LEWS) project and its successor, the Livestock Information Network and Knowledge System (LINKS) project, represent a complementary effort to provide pastoralists in East Africa with tools to track climate and forage conditions so as to inform herd management decisions and mitigate the consequences of impending and current drought. Meanwhile, the World Food Program's emerging project on Strengthening Emergency Needs Assessment Capacity (SENAC), for example, is currently engaged in the identification of methods and tools for emergency assessments, and for evaluating the effects of different types and combinations of shocks on various livelihood groups. Such early warning systems are increasingly recognized as essential to effective emergency response (Barrett and Maxwell, 2005).

There could be great humanitarian and economic value to developing accurate forecasting tools aimed at the needs of populations vulnerable to slow-onset disasters

such as drought – which are quite distinct from rapid-onset events such as those associated with many other natural disasters, including earthquakes, hurricanes or tsunamis – and the identification of where and when to intervene and at what scale. In this paper, we aim to develop such a forecasting tool and demonstrate its practical value. We focus on the arid lands of northern Kenya, largely populated by nomadic pastoralists and particularly vulnerable to covariate shocks in the form of droughts and floods.

Our primary objective is to make use of household data collected over several years by the Arid Lands Resource Management Project (ALRMP) of the Government of Kenya, and the spatially explicit data on forage conditions, rainfall and normalized differenced vegetation index[2] (NDVI) generated by the LEWS/LINKS team to develop an empirical forecasting model that can predict the expected human impact of covariate shocks and thereby provide a useful statistical method for early warning and emergency needs assessment in the arid and semi-arid lands (ASAL) of northern Kenya.

We explore the village-level joint dynamics of: (i) herd size, pastoralists' primary asset (Lybbert et al., 2004, McPeak, 2004); (ii) lactation rates, which drive pastoralists' primary income source (McPeak, 2004); (iii) climate and forage conditions, the key exogenous environmental determinants of productivity dynamics in pastoralist systems (Coppock et al., 1986; Ellis, 1994; Ellis and Swift, 1988); and (iv) child nutritional status, a key outcome variable of humanitarian and economic interest. The idea is to investigate whether patterns in climate variables, herd size and other key system variables can predict changes in village-level child nutritional indicators. We expect the model we develop to contribute to improved early warning capacity and accuracy, especially for increased precision in targeting areas for emergency assistance with sufficient lead time to mount an appropriate response.

The potential value of such a model is best recognized in light of the deficiencies and constraints of current systems. First, the resources available to respond to emergencies are extremely scarce. As Barrett and Maxwell (2006) report, only 16% of consolidated appeals for emergencies in Sub-Saharan Africa are funded. Given such scarcity, available resources need to be used wisely. Second, even when response is

---

[2] NDVI is a numerical indicator, typically generated from satellite data, which describes the relative density of green vegetation within a target area. The measure of greenness is calculated by observing the ratio of distinct wavelengths of visible and near-infrared sunlight that are giving off by plants (Clevers, 1988).

initiated, numerous bureaucratic and logistical obstacles cause significant delays in emergency food aid delivery that can substantially reduce its effectiveness. Third, while existing targeting of food aid in Kenya has been closely related to rainfall, Lentz and Barrett (2005) show that rainfall shocks are surprisingly weakly correlated with household-level shocks to income and assets.

Given tight resource constraints, inevitable response delays and the limited effectiveness of rainfall-based targeting, there is much potential value in improved methods of multi-month early warning approaches that offer greater accuracy and a longer forecasting horizon. By forecasting a key outcome variable, the nutritional health of children in the affected population, our model improves on existing early warning systems that merely forecast the likelihood of a drought event, which is imperfectly correlated with the human outcomes of interest to policy makers and relief agencies. It is important to clarify that the clientele for such a forecasting tool is *not* the subject populations, who often have their own, traditional methods of forecasting droughts and initiating personal and community level response strategies (Luseno et al., 2003), but rather government officials, operational agency managers and others tasked with responding to prospective food crises with external resources. Our objective is to illustrate how a relatively parsimonious set of regularly measured variables can generate reasonably accurate forecasts with at least three months' lead time and relatively low rates of Type I errors (i.e., false negatives, in this setting implying that the model predicts no crisis but one nonetheless occurs).

The rest of the paper is structured as follows. Section 2 summarizes the data we use. In section 3, we briefly describe the methodology we apply to develop our forecasting model. Descriptive statistics of key model variables are presented and discussed in section 4. In section 5 we present forecasting results and test the performance of our forecasts. Section 6 concludes by exploring the potential uses of the model and its policy implications.

2. DATA

The World Bank-funded ALRMP seeks to address the vulnerability of populations living in Kenya's arid lands and to improve their ability to manage the risks

they often face (World Bank Group, 2003). As part of the project, repeated cross-sectional data have been collected in various communities across Kenya's arid districts since 1996. Data relevant for our objective are sourced from household level surveys that contain detailed information on livestock including herd sizes, mortality rates, lactation rates, and managed off-take rates. Critically, child nutritional data in the form of mid-upper arm circumference (MUAC) was also collected.

While data have been collected monthly from 1996 and across 10 different districts, poor data organization and storage unfortunately resulted in significant quantities of lost data that rendered many areas too patchy for any rigorous analysis. Consequently, our effective data set, while still substantial, constitutes a mere subset of what was collected.

Furthermore, no authoritative document on the collection procedures and sampling methodologies employed exists. As such, while we know that community sites were purposively selected to take into account population density and spatial distribution across a district, we are only aware that enumerators were asked to randomly select 30 households per community without being clear on the method of randomization used or whether it was enforced. Nevertheless, while this forces us to be cautious about (mis)representing our results as statistically representative of any single place, the data seem sufficiently rich to shed important light on dynamic processes that are as yet not well understood and for which quantitative evidence of any sort is distressingly scarce. In Figure 1, we show the four districts, Baringo, Marsabit, Samburu and Turkana, as well as the sample communities within each district, for which a sufficient number of continuous observations were available and therefore used to estimate our models.

[Figure 1 about here]

Although surveyed households were theoretically revisited each month for a year before a new sample was generated, insufficient record keeping prevents individual households from being linked across periods. As such, we create a community-level pseudo-panel by generating community-level summary statistics of the pertinent variables. Despite the unavoidable loss of information, such a pseudo-panel is well suited to investigating the impacts of covariate, community-level shocks whose dynamics

are more relevant for external emergency interventions than are idiosyncratic, household-level shocks (Deaton 1985, McKenzie 2004).

We supplement the ALRMP data with a rich source of climate and forage availability data collected and produced by LEWS/LINKS researchers. LEWS/LINKS has developed a set of technologies and models that provide high-resolution, high-frequency estimates of livestock forage availability in the pastoralist-dense regions of East Africa (Kaitho et al. 2003, Stuth et al. 2003). These data provide key variables for our model. Most major covariate shocks that hit pastoralist communities are a function of adverse forage and water conditions associated with climate fluctuations (Ellis and Swift 1988, Ellis 1994, Galvin et al. 2001). As changes in livestock fertility, mortality and lactation are closely related to forage and water quality and availability, access to the detailed dynamics of these variables greatly enhances the value and precision of models generated to investigate and predict the human impact of the climate shocks that frequently destabilize pastoralist communities. To the best of our knowledge, no available forecasting tools integrate these biophysical, socioeconomic and anthropometric data for early warning purposes.

3. METHODOLOGY

As adverse covariate shocks come in various forms and have varied effects on the affected populations, emergency assessments often focus on the response, or predicted response, of critical indicator variables. Proxies of food insecurity such as the real price of key staples or the level of food production, availability, or expenditure can yield estimates of the population's vulnerability to hunger and starvation. ALRMP data include sample readings of the Mid-Upper Arm Circumference (MUAC) for selected community children. Among indicators of nutritional wellbeing, MUAC is particularly well-suited for our purposes. As a measure of wasting, MUAC is capable of capturing short-term fluctuations in the presence of nutritional stress and can thus serve as a gauge of the human impact of various shocks. Furthermore, MUAC is easier and less costly to collect than weight-for-height (W/H), the most commonly used and most documented anthropometric measure for wasting. Indeed, several studies have shown MUAC to be a far better predictor of child mortality than W/H (Chen et al. 1980, Alam et al. 1989, Vella

et al. 1994). As such, MUAC is a particularly appropriate indicator of the welfare impacts of a humanitarian crisis.

As we are primarily concerned with generating the most precise forecasts possible, our objective is to maximize the forecasting accuracy of the model. While estimating unbiased regression coefficients would allow us to reasonably infer the relationship between key model variables and MUAC, attempting to correct for endogeneity by using instruments, or imposing limits on the error structure, would necessarily reduce efficiency and thus the accuracy of forecasts. Consequently, we leave the potentially interesting inference-based estimation for further study and in this paper focus exclusively on optimizing the forecast capabilities of our model.

As MUAC trends are likely to display a high degree of persistence – that is, current values are highly correlated to recent historical values - it would be logical to include lagged values of MUAC as an explanatory variable to capture the inevitable dynamic component. We therefore estimate a dynamic model for our unbalanced panel[3] that takes the following form:

$$y_{it} = \sum_{s=H}^{S} \lambda_s y_{i,t-s} + \sum_{s=H}^{S} X'_{i,t-s}\beta_{t-s} + u_i + \varphi_t + v_{it} \quad \begin{pmatrix} i = 1,...,N \\ t = 1,...,T \end{pmatrix},$$

where y_{it} denotes is an observation on the dependent variable for unit i (community level MUAC means in our case) at time t, with $y_{i,t-s}$ the s-period lagged MUAC community specific means that capture the dynamic nature of the relationship via the unknown coefficients, λ_s. H defines the most recent lagged (or contemporaneous if H=0) value included while S>H denotes the most distant lag included. $X_{i,t-s}$ is a vector of severally lagged explanatory variables that vary across space and time. In this case, $X_{i,t-s}$ would be all those variables thought to have some effect on MUAC. These include variables such as herd size, herd mortality, lactation and livestock sales and slaughters that may affect MUAC directly via the provision of nutrients and calories in milk and meat, or indirectly through changes in income and wealth due to managed offtake. Policy variables such as the provision of food aid, which occurs with some regularity across our

---

[3] Recall that the panel is unbalanced. As such, some observations will be missing in the interval $t \in [\,0,T\,]$.

sample, are also to be included in $X_{i,t-s}$ as are biophysical variables such as rain, NDVI and forage conditions that indirectly affect MUAC via herd compositions, food prices etc. The parameter u_i denotes the community effect capturing unobserved time-invariant heterogeneity, φ_t are time dummies that capture additional seasonality not controlled for by temporal variation in $X_{i,t-s}$, and v_{it} is the unobserved error term. The dynamic model characterized above is quite general and, as implied by the inclusion of a non-specified limit on the lag term s,, can accommodate multiple lags of both the dependent and independent variables. Lag structure is generally chosen based on the assumed data generating process, the estimators chosen and the researcher's objective. As our objective is to maximize forecasting efficiency, we will use the Root Mean Square Error (RMSE) criterion to select for optimal lag structure.

Our exclusive focus on forecasting performance considerably simplifies our estimation. One key problem with estimating a dynamic model in general is that the presence of the lagged dependent variable introduces endogeneity that causes the least squared dummy variable (LSDV) estimator to become biased and inconsistent.[4] While several estimators have been developed that deal with this problem (Anderson and Hsiao 1982, Arellano and Bond, 1991, Kiviet, 1995, Hansen 2001), they trade-off efficiency for unbiasedness and have greater variance than the LSDV estimator. The inclusion of other predetermined or endogenous explanatory variables, such as food aid, would also result in biased coefficients if not appropriate instrumented. As we are not interested in the coefficient estimates generated by the estimation, we can abstract away from these concerns and simply use the LSDV estimator which is well known to be of least variance.

4. DESCRIPTIVE STATISTICS

In what follows we briefly describe the data used to estimate our models and present graphical representations of the trends of key variables. The data structure, shown in Table 1, lists the number of observations per district for which we have the full set of pertinent variables in the specified time periods. Recall that the observations are all community level aggregates, largely community means or functions of means. Though

---

[4] The LSDV estimator is simply an Ordinary Least Squared, Fixed-Effects Panel estimator (Hansen, 2001)

each survey round was to sample 30 households per community, smaller household specific-samples are frequently observed in the data. To strike a balance between throwing out observations and taking averages over unreasonably small sample sizes, we only include the 55 communities for which 15 or more households were sampled in a given time period.

[Table 1 about here]

As previously noted, poor data entry and storage protocols render the ALRMP panel data quite unbalanced. All time periods suffer the loss of several community level observations, often across multiple districts. In certain cases, all observations from a particular district are missing. No district has observations in each time period under consideration, nor can any districts claim a complete set of community level observations for the time periods for which observations are available. As we would lose too many observations if we conditioned our estimates on a balanced sub-sample, we use the unbalanced panel as is.[5] Fortunately, most of the oft-used dynamic panel estimators are capable of estimating unbalanced panels. Moreover, it has been shown that given the same number of effective lagged observations, a highly unbalanced panel does not negatively affect the bias or Root Mean Squared Error (RMSE) of estimators (Bruno, 2005).

To get a sense of the dynamics of key variables in our model, we estimated kernel regressions of the variables across time. As MUAC was collected for children aged 6-59 months residing within sample households or in close proximity (for a maximum of 5 children per household), we transformed our MUAC data into standardized Z-scores.[6] We used the internationally recognized 1978 CDC/WHO growth chart whose reference population is American children sampled in the 1977 National Center for Health

---

[5] We attempted to artificially balance the panel by imputing the missing variables. Using the *hotdeck* command in STATA, which applies the Bayesian bootstrapping method of Rubin and Shenker (1986) to estimate the missing values, we created multiple imputed data sets. Estimates based on the imputed data sets were significantly inferior to those on the unbalanced panel across all specifications that we tried. This is likely because the number of missing observations is quite large and imputing all the variables of the missing observations adds a considerable amount of error in the variables limiting the accuracy and consistency of estimates based on them.

[6] While standardization is condition on both age and sex, we only had age data. As such, we generated Z-scores for both males and females and took the average.

Statistics survey. Figure 2 depicts the trend of one of our dependent variables, village-level mean child MUAC z-score, across our sample districts.

[Figure 2 about here]

The generally low mean Z-scores show that even during relatively good periods, the average MUAC value of the sample population is much less than the reference population. Nonetheless, within our sample, significant temporal fluctuations in the Z-scores are suggestive of the influence exogenous shocks exert on nutritional wellbeing. 2000 was a particularly bad year with MUAC levels falling drastically to their lowest levels in the sample period. Rapid gains in MUAC were registered early in 2001, with sustained improvement until mid-2002 when a gradual decline set in, which generally flattened out by early 2003. While the trend in sample averages are somewhat mirrored by the district means, there are also clear inter-district differences, both in levels and trends.

Trends in rainfall and forage availability rates (Figure 3) may partly explain the MUAC dynamics. As can be inferred from the figure, 2000 was a year of drought, with the main rains, which generally fall during the months of March through May, failing. Lack of rain results in shrinking forage availability for livestock. Largely habited by pastoralists who generate income from their animals, the drought may account for the low MUAC levels witnessed in 2000. However, strict agroecological determinism of child nutrition status seems implausible. Other factors must also influence MUAC, as is evident with falling MUAC rates around mid-2002 despite relatively high forage availability.

[Figure 3 about here]

The increasing incidence of drought and subsequent widespread severe wasting in Kenya's arid north in the past decade or so has catalyzed a significant food aid response whose intensity varies across time, depending on the degree of the crisis affecting recipient populations as well as other socioeconomic and political factors. Figure 4 shows the trends in the fraction of households that received Unimix (a micronutrient fortified corn soy blend specifically formulated for undernourished children) and the

fraction receiving food aid in the form of cereals from 2000 to 2005. While Unimix is not as widely distributed as cereals, its pattern of distribution across time is generally similar despite a notable exception in the early months of 2000 where Unimix distribution took longer to respond to the MUAC crisis than did cereal provision. The rapid scaling back of both Unimix and cereal food aid support in 2002 may help explain why, despite better than average forage conditions, MUAC Z-scores began to fall from mid-2002.

[Figure 4 about here]

Livestock represent the main stock of assets and a significant source of nutrition for pastoralists. As such, herd dynamics may also contain an important source of information pertaining to the health and wellbeing of pastoralist populations. For example, declining lactation rates would limit the availability of a key nutrient source among pastoralists which, if severe enough, will translate to lower MUAC. Increasing incidence of sales or slaughter on the other hand, may signal coping behaviors triggered by negative shocks. In Figure 5, we present herd sizes aggregated into Total Livestock Units (TLU) and compare their dynamics across districts.[7]

[Figure 5 about here]

While there is a general upward trend in herd sizes across the sample period, the drought of 2000 resulted in a noticeable decline as can be clearly seen in the TLU trend. TLU trends follow a similar pattern across districts, with the exception of Baringo, where a significant fall in herd sizes occurs toward the end of 2002. Much of this decline can be explained by the rate of managed off-take (sales and slaughter) in Baringo, which was significantly higher than the rest of the sample at this time. In Figure 6, we show the trends in lactation, mortality, sales and slaughter rates. As expected, the drought period in 2000 is characterized by a high degree of mortality across the four animals and significantly lower lactation rates among cattle and camels.[8] Trends in sales and

---

[7] TLUs allow for a comparison of livestock quantities across species. One TLU is equivalent to 1 head of cattle, 0.7 camels, 10 goats or 11 sheep.
[8] Lactation rates are calculated as daily means per herd and thus include both female and male of the species as well as young and old. This could partly explain the generally low rates of lactation posted.

slaughter rates do not seem to follow a pattern that can be readily interpreted as either post-shock coping behavior or ex-ante risk mitigation. Nonetheless, sales and slaughter, especially among small stock are noticeably higher during the drought period.[9]

[Figure 6 about here]

5. ESTIMATION RESULTS

5.1 Estimating the Forecasting Model

Armed with a sense of the dynamics of key model variables, we now move on to estimate our model. Recall that our primary objective is to estimate a model that generates the most accurate forecasts of MUAC levels possible. As we are not interested in making inferences on the relationship between our explanatory variables and MUAC levels, we do not concern ourselves with estimating unbiased, consistent coefficients. Instead, we focus exclusively on trying to maximize the predictive efficiency of our model. As such, we include as many explanatory variables as possible, the only condition being that they contribute to an increase in the (degrees of freedom adjusted) Root Mean Squared Error (RMSE).[10] We use the LSDV estimator which, as discussed in section 3, is the most efficient estimator for this purpose. After estimating numerous specifications of various combinations of the variables, including squared terms, cross-products, and varying lag lengths, the specification that offered the best RMSE is shown in Table 2.

[Table 2 about here]

---

[9] As previously mentioned, the large increase in sales rates starting from the end of 2002 is largely driven by sales in Baringo. One possible explanation is that the exceptionally favorable forage availability in Baringo in early to mid 2002 created a particularly healthy stock of animals that fetched relatively high prices at the market. The *El nino* floods that then hit Baringo especially hard in late 2002 may also explain the spike in sales, as the destruction caused by floods increased the demand for cash to cope with the unexpected catastrophe. Because increased livestock mortality due to floods would also affect the denominator for the sales rate variable, this would also increase sales per livestock holdings.

[10] RMSE is the most commonly used criterion for comparing the performance of forecasting models. Other popular criteria include the asymptotically efficient Akaike Information Criterion (AIC), and the asymptotically consistent Schwarz Information Criterion (SIC). The main difference between these three model selection criteria is that they penalize degrees of freedom differently. For the various specifications we estimated we also calculated the resulting AIC and SIC. In no situation did any of the three criteria disagree as to which model was best, probably because degrees of freedom represent a large fraction of the total available observations. For simplicity, we thus only present the RMSE.

The first set of explanatory variables in Table 2, lagged MUAC Moments, control for the distribution of lagged MUAC within a community. By more accurately describing the distribution of past MUAC Z-Scores across a community, including higher moments of MUAC affords more precise forecasts of the means. Each MUAC moment variable, including values of the lagged dependent variable, is introduced with one, two, and three month lags. Including additional lags reduced the resulting RMSE. This most likely due to the unbalanced nature of the ALRMP panel data which causes an increasing number of observations to be dropped as additional lags are introduced.

The herd dynamics variables were disaggregated into two different categories (measured in TLUs): one for large stock (cattle and camels), and the other for small stock (goats and sheep).[11] As Figure 6 shows, large stock trends generally follow a similar pattern across locations while small stock trends often move in tandem. As food aid receipts within sample households are quite considerable, and as they may impact MUAC levels, we include four food aid-related variables; the fraction of households receiving UNIMIX, the fraction receiving regular cereals, the amount (in kilograms) of UNIMIX received per recipient household, and the amount of cereal received per recipient household. Both the herd dynamic variables and the food aid variables were introduced with one and two month lags.

The biophysical variables, mean monthly rain (mm), forage availability (kg/hectare) and NDVI were the only variables introduced with squared terms in order to capture potential nonlinearities in their effect on MUAC. In addition, each variable was included with five lags (from a two through six-month lag).[12] To account for other sources of seasonal variation not captured by these biophysical variables, we include monthly dummy variables.

Note that we do not introduce any of the explanatory variables contemporaneously. Despite the likelihood that this would increase RMSE, we leave out contemporary variables in order to allow for a legitimate forecast. This specification enables one-month ahead forecasts of MUAC means. Given the lags in emergency food

---

[11] Including separate variables for each of the four animal species reduced the RMSE.

[12] Starting from a one-month lag, or extending to a seven-month lag does not improve results. As the effect of biophysical variables on MUAC likely occurs with a lag, the fact that including a one-month lag did not improve the forecast accuracy suggests that the effect of these variables on MUAC is felt with more than a month lag. Conversely, the effect is likely to wear off after six-months.

aid response caused by bureaucratic and other constraints (Barrett and Maxwell, 2005), a one-month forecast is admittedly short, leaving little leeway for aid workers to make effective use of the forecasts. As readers will note, the inverse relationship that exists between forecasting horizon and forecast precision involves a delicate tradeoff between timely provision of the forecast and improved accuracy of forecasts, both of which determine the usefulness of the forecasting model to potential users of this information. As an alternative, we also estimate a three-month ahead forecasting model. We utilize the same specification presented in Table 2, the difference being that each variable is lagged by an additional two months in this case.[13]

We present key statistics from both the one-month ahead and the three-month ahead forecast models in Table 3. We also include a few statistics on the dependent variable. However, as both models introduce a considerable number of explanatory variables, and as we are focusing on inference or the precision of estimated coefficients, we do not include coefficient estimates or t-statistics.[14]

[Table 3 about here]

The results are quite encouraging. The RMSE are 0.1666 and 0.1986 for the one and three month models, respectively. These RMSEs, which define the mean error deviation of within-sample forecasts from the actual, are relatively small, especially given that the variable being forecast has a sample standard deviation of 0.40. This gives an indication that out-of-sample forecast performance should be quite good. Furthermore, the models fit the data surprisingly well, generating high r-squared values for both within community temporal variation and between community spatial variation. The between r-squares are strikingly high at 0.98 and 0.88 for the one and three month models respectively. While the within r-squares, which are also relatively impressive at 0.53 and 0.34, are more relevant for forecasting, the extremely high between r-squares suggest that our explanatory variable explain the bulk of spatial variation in community level MUAC means across the sample area for a given time period. This may also explain the good model fit despite the highly unbalanced nature of the panel. If spatial

---

[13] So, for example, the lag structure for the MUAC Moments variables, including the lags of the dependent variable, now becomes L(3,4,5). For the biophysical variables the lag structure now becomes L(4,5,6,7,8). This is necessary to allow for legitimate three-month forecasts.

[14] Interested readers can obtain the full set of results from the authors.

14

variation is so well captured by the explanatory variables, the absence of observations for certain communities will not have as much of a negative effect on model fit and forecasting performance as missing a full set of observations for particular time periods. As we can see from Table 1, the unbalanced nature is fortunately more a product of a few and changing set of communities missing across numerous time periods rather than whole sets of communities missing at a given time.

The next step is to operationalize the model and test its out-of-sample forecast performance. This allows us to ask how food security managers or the relevant policy makers can best make use of the forecasts. Note that the forecast is a point estimate of predicted community MUAC means. Given a forecast estimate, one might want to know, for example, what the resulting 90% confidence interval for the true value is. Alternatively, one might want to know the degree of confidence with which a person using this forecast can claim that the true value lies below a certain critical threshold of policy interest.

In either case, if the objective is to gauge the intensity or magnitude of food insecurity, or to estimate the severity of child malnutrition, a potential weakness in our forecast is revealed. Our model forecasts the mean MUAC Z-score for sample community children. The distribution of MUAC is likely to reflect within and between community inequalities in calorie and nutrient intake and may therefore vary between and within communities. MUAC distributions could also vary across time due to differential capacities of households to support their children's nutritional wellbeing during times of food stress. Children of particularly vulnerable households are likely to suffer higher rates of nutritional deficiencies and wasting in the face of shocks. This would show up in the data as a lengthening in the left-tail of the MUAC distribution, quite apart from a general leftward shift in the entire distribution that would result if the adverse nutrition effects of the shock were equally felt by all community children. As emergency response to widespread hunger is largely conditioned by the degree and prevalence of gross malnutrition, often defined by the proportion of children whose anthropometric conditions fall below a certain 'unacceptable' threshold (Howe and Devereux, 2004; UN Sub-Committee on Nutrition, 1999; World Food Programme, 2000), forecasting MUAC means may not be the most suitable strategy.

We therefore re-estimate our model with an alternative dependent variable that more precisely specifies the left-tail density of the MUAC distribution: the fraction of children in each community who have MUAC Z-scores less than -2. We set -2 as our threshold to be consistent with the benchmark often employed by key emergency relief agencies to define various levels of food stress and famine (Howe and Devereux, 2004; UN Sub-Committee on Nutrition, 1999; World Food Programme, 2000).[15] Figure 7 presents full-sample and district level kernel regression of the fraction of children per community with MUAC Z-scores below -2 across time.

[Figure 7 about here]

Not surprisingly, the proportion of children with MUAC Z-scores below -2 tells a largely similar story of the degree of food stress as that suggested by the results on MUAC means (Figure 2). 2000 was a time of widespread distress, as witnessed by the considerable proportion of children so malnourished that their MUAC Z-scores fell below 2 standard deviations of the reference mean. According to the scale of famine intensity devised by Howe and Devereux (2004), during the 2000 drought the vast majority of our sample experienced 'famine conditions', which they define as 20% or more of children with Z-scores<-2. Indeed, in some areas of Turkana, the condition could be classified as a 'severe famine' (40% or more with Z-scores<-2).[16] Differences across districts also mirror the pattern revealed by MUAC means, with Baringo and Samburu relatively better off. Under the Howe-Devereux classification scale, Marsabit sample communities experienced 'famine conditions' from early 2000 to late 2004. By offering a universal scale by which we can interpret our results in relation to the severity and intensity of food insecurity faced by the target population, focusing on the proportion of

---

[15] Howe and Devereux (2004) offer a useful framework by which to define the intensity and magnitude of famine. The scaling system they develop uses various levels of wasting, defined by the proportion of children with Z-Score less than -2, as a key indicator variable (though the anthropometric measure they use is weight-for-height, standardizing into Z-Scores should make the use of MUAC a valid substitute). Such a scaling system would be useful in conjunction with forecast estimates as a means to classify the estimated severity of food insecurity.

[16] Note that the figures display averages over communities. During 2000, 35 out of 56 of our sample communities would have been classified as experiencing severe famine conditions for one or more months. Fourteen of these suffered severe famines for 6 or more months of 2000.

the sample who fall below a critically defined threshold, rather than on mean MUAC, may be of greater use.

Table 4 presents key statistics for the forecasting model with proportion MUAC Z-score<-2 as its dependent variable. Other than simply substituting the relevant dependent variables and their lags, no other difference exists between the specification we employed in this case and that presented in Table 2. Again, we employed an LSDV estimator for both a one-month ahead and a three-month ahead forecast.[17]

While the overall R-squares for both the one-month and the three-month forecasts are slightly lower for the proportion model presented above, RMSE, the more relevant within-sample proxy for out-of-sample forecast performance, favors the mean forecast. Though absolute RMSE, at 0.1205 and 0.1395 for the one-month and three-month models respectively, are lower for the proportion forecast, we must keep in mind that RMSE are presented in units of the dependent variable. As such, with MUAC means ranging from -2.73 to 0.27 across the sample period, as opposed to the zero to one range of MUAC proportions, it would seem that the forecasting performance for the means model is superior.

[Table 4 about here]

5.2 Testing Forecast Performance

There are other methods of evaluating the forecast performance of our models than using RMSE. To do so, we move on to generate a series of rolling one and three month ahead forecasts for both the mean and proportions models in order to more precisely compare and estimate their forecast performance. Here, we use lagged information to generate a forecast of a future event in a given time period, then assess how accurate the forecast turned out to be using information in the data set on what actually happened in the future time period for which the prediction was generated. We begin by generating forecasts for January 2004. As such, one-month ahead forecasts are

---

[17] Note that in theory the dependent variable as a proportion is doubly centered at zero from below and one from above. In practice we find that left censoring is particularly prevalent with 27% zero-observations (Less than 1% of the observations are right-censored). This suggests using a panel tobit estimator. We fit a random-effects tobit model (a sufficient statistic allowing for the fixed effects to be conditioned out of likelihood dos not exist) to our data but the resulting RMSE was larger. We thus opted for the more efficient LSDV estimator.

based on model coefficients estimated on the subset of the data ending December 2003 and the three-month forecasts are estimated on data up to October 2003. We then fold in an extra month of data, estimating one-month forecasts for February 2004 using data truncated at January 2004 and three-month forecasts on the subset of data ending on November 2004. We continue in this manner, generating monthly forecasts until May 2005. We can thus gauge forecast performance by comparing out-of-sample forecasts with the actual observed value at the comparison period. In Figure 8, we show the resulting one and three month forecasts for MUAC means in conjunction with the actual means. We first present the forecasts superimposed on the full period smoothed actual means and then highlight the forecasting errors by focusing only on the forecasted period and the unsmoothed actual means. Figure 9 does the same for the proportions model.

<div align="center">

[Figure 8 about here]

[Figure 9 about here]

</div>

Two key points emerge from both Figure 8 and 9. First, it seems that there is a tendency for forecasts to systematically overstate the degree of malnutrition. For the means model, the vast majority of both one and three month forecasts indicate lower (recalling that lower MUAC measures indicate higher levels of malnutrition) than actual means. Similarly, the proportions forecast consistently predict a larger fraction of children with under -2 MUAC Z-scores. One explanation could be that the high degree of persistence in the model, implied by the modest deterioration in forecast accuracy as the horizon increases from one to three months, causes forecasts to respond slower to a marked turnaround in MUAC levels.[18] As such, the consistent decline in MUAC means witnessed from mid 2002 until late 2003 that taper off just as we begin forecasting in January 2004 could explain the initially large overshooting of the forecasts which, due to the autocorrelated structure of the model, continue in the trajectory of past trends for a few months before correcting. This is of significance, as it suggests that forecasting accuracy may diminish at particularly critical periods such as, for example, when a severe shock hits, causing a rapid deterioration in food security as was the case in early 2000.

---

[18] Moreover, the three lagged dependent variables for all four specifications were both statistically significant and had relatively large coefficients, especially for the earlier lags. This supports the hypothesis of considerable persistence.

For emergency response and food security managers, such a time is when accurate information would be most valuable.

We take a look at how serious this problem may be by estimating forecast bias. This is done by regressing the actual, observed value of the dependent variables on the forecasted value in a simple linear model of the form, $X^a = \alpha + \beta X^f + \varepsilon$ where X^a is the actual value and X^f is the forecasted value. Testing the null hypotheses $H_0 : \alpha = 0, H_0 : \beta = 1$, is thus a test for an unbiased forecast. Rejection of that null implies bias. The results for the mean and proportions models for both forecasting horizons are presented in Table 5.

[Table 6 about here]

We can only reject a null of an unbiased forecast for the 3-month means model. The 3-month means model is the only one for which both α and β are estimated significantly (thus narrowing the range of the 95% confidence interval) and are both significantly different from the values consistent with an unbiased forecast. This implies that overall, the forecasting models performs quite well as unbiased and reasonably accurate predictors of both mean MUAC levels (for the one-month forecast) and proportions of children suffering severe wasting (for both forecasting horizons).

As there have only been a few abrupt changes in MUAC trends over the sample period, and the sample period is relatively short, the is likely to 'learn' and improve as more data are folded into it, especially data bracketing relatively sharp changes in MUAC. Indeed, the second salient point revealed in Figures 7 and 8 is that the models seem to learn quite fast as more data is folded into this. In both cases, and for both forecast horizons, it is clear that the forecasts track actual trends closer as we roll forward in time. While this could be a feature of the reasonably stable trends across the latter part of the forecasting period, a portion of the improved forecasts may be explained by learning. This highlights the value, not only of adding more data to the model, but also of assuring that data are continuous as the highly unbalanced nature of the ALRMP panel as it currently stands undoubtedly reduces the model's performance.

Another revealing feature of these forecasts is that despite the fact that the RMSE (when normalized for units) suggests that the mean model has superior forecast performance relative to the proportions model, this does not come across in comparing

Figures 8 and 9. From Figure 8, the largest differences between actual and forecasted values, which seem to occur in March 2004, are about 0.07 and 0.14 Z-scores for the one and three month forecasts respectively. From Figure 8, forecast errors, which also appear to occur in March 2004, are approximately 2-7 percentage points for the one and three month forecasts respectively. Moreover, the 3-month means model was the only one shown to be affected by forecasting bias.

While this does not provide conclusive support for the relative performance of either model, especially as these full sample average results mask much of the structure of community level forecast errors, it nevertheless calls into question the superiority of the means model over the proportion model and calls for further investigation. Consequently, we further test the models by gauging how they perform based on the likelihood of correct responses that the models provide to questions they might be called upon to answer.

For practical purposes, suppose that a food security manager wants to know the likelihood that a condition of 'famine' will prevail in a particular area. Given a particular forecast, he or she would then need to know the confidence with which the actual proportion of children with MUAC Z-score<-2 was greater or equal to 20%.[19] We generate such confidence levels for the one and three month forecasts of both the proportion and the means models. As none of the forecasting residuals for any of the models were normally distributed,[20] we employed percentile counts of the in-sample distribution of forecast residuals to generate our confidence levels. So, for example, if 10% of the sample residuals for the one-month proportion forecast had an error of greater than 0.15 percentage points, and the forecasted one-month ahead proportion was 0.05, then the confidence with which one could claim that in one months time, 20% or more children will register MUAC Z-scores<-2, would only be 10%.

Extracting confidence levels allows for defining trigger points that, when crossed, could set in motion a series of predetermined response mechanisms (Barrett, 1997).

---

[19] Following Howe and Devereux (2004), we define famine to be the condition at which twenty percent or more children have MUAC Z-scores<-2. For purposes of testing the performance of the means model, we employ an alternative MUAC means-based definition of famine that we arbitrarily set as the condition at which the mean level of MUAC Z-scores is less than -1.8. The model's performance should be robust to threshold chosen to define famine.

[20] Shapiro-Wilks tests for normality of a distribution rejected normality of the forecasting residuals in each of the for cases.

Thus, for example, one could imagine a set of rules such that if famine is expected to arise with 66% confidence, food aid and other forms of emergency support are immediately mobilized for the affected areas. Since such a coordinated response inevitably takes time as logistical, financial and bureaucratic concerns are sorted out, a dual-trigger system could be devised in which, at the crossing of a lower confidence band (say 33%), readiness measures are taken in preparation for a deployment should the second trigger be crossed.

We use such a hypothetical decision making system to offer another measure of the performance of the forecasts. First, we generate confidence levels for each forecasting horizon for both the means and proportions models. We then set three arbitrary trigger points, one each at the 75%, 66% and 50% confidence levels. The trigger points can be thought of as the minimum confidence level that a policy maker requires to initiate a famine emergency response. Defining emergency response when there is actually a famine, or no response when there is no famine, as a 'correct' decision, we calculate the fraction of correct decisions that would result from utilizing the different models. Table 5 presents the results. We also calculate the fraction of wrong decisions that are Type 1 errors - the proportion of total wrong decisions that result from failing to respond when a famine actually occurs.

[Table 5 about here]

The results clearly reveal the superiority of the proportions model in generating correct decisions. The magnitudes of the differences in performance are quite striking. For all confidence thresholds, and both forecasting horizons, decisions based on the proportions model are about 15% more likely to be correct. Interestingly enough, the three-month proportions forecast also outperforms the one-month means forecast by around 8% under each confidence threshold.

Further underscoring the results from Figures 8 and 9, the fairly small depreciation in performance as we increase our forecasting horizon is promising and shows that the models can be used fairly accurately to give policy makers a reasonable, three-month early warning window to mitigate the consequences of impending shocks. The two-month gain in early warning is arguably worth the modest loss of accuracy.

Requiring higher degrees of confidence to guide the release of emergency resources for famine support does not necessarily result in a significant additional number of correct decisions. Though higher confidence thresholds do reduce the likelihood of taking costly measures to mitigate the consequences of a famine that does not materialize, it increases the frequency of making the arguably more costly mistake of failing to initiate the requisite response when famine actually occurs. As can be seen in Table 5, which presents the fraction of total error (wrong decisions) that are Type 1 (no action is initiated when a famine even actually occurs), all four forecasting regimes err on the side of responding to a wrongly declared famine (a Type 2 error), rather than committing the mistake of failing to act when famine is real.

Deciding how to distribute incorrect decisions between Type 1 and Type 2 errors is a normative judgment best left to the policymakers. As this is nonetheless an important decision requiring the crucial tradeoff between ineffectively distributing scarce resources (committing a Type 2 error) and failing to offer emergency respite to disaster-stricken populations (a Type 1 error), we investigate this issue further. In Figure 10, we graph the relationship between the fraction of correct decisions and the desired confidence threshold, as well as the expected fraction of Type 1 errors by confidence threshold.

[Figure 10 about here]

The key feature of Figure 10 is to show that the fraction of correct decisions is relatively stable, and highest, along a wide range of confidence thresholds. What changes most, especially for the proportions model, is the relationship between expected Type 1 errors committed and the confidence threshold. Lower confidence thresholds for decision making necessarily result in a smaller fraction of Type 1 errors. The observation that the proportions model results in substantially more correct decisions over the entire range of confidence thresholds supports use of the proportions model over the means model. One decision making strategy might be to choose the trigger confidence threshold that yields the maximum fraction of Type 1 errors that emergency responders are willing to accept; another might be to choose the threshold level that maximizes the difference between the proportion of correct decisions and the fraction of Type 1 errors; another might be the

threshold that maximizes the proportion of correct decisions. Such normative decisions are the domain of policymakers.

6. CONCLUSION AND POLICY IMPLICATIONS

Using a pseudo-panel of community level cohorts collected from primarily pastoralist communities selected across four Districts in Kenya's arid north, we set out to develop an empirical forecasting model that can predict, with reasonable accuracy, the expected welfare impact of impending drought. We find that the joint dynamics of herd composition and herd management, climate and forage availability and food aid flows are able to forecast child nutritional status dynamics with impressive precision. Forecasting the proportion of children that fall below a critical nutritional threshold (proxied by a threshold MUAC value, a relevant parameter for assessments of the intensity and severity of a humanitarian crisis), yielded particularly accurate predictions. Moreover, offering policy-makers more response leeway by forecasting three months into the future only marginally reduced forecast performance relative to a one-month ahead forecast.

The policy implications are immediately clear. It is possible to generate accurate forecasts of a key human welfare indicator that is particularly sensitive to the shocks that occasionally affect large numbers of the target population simultaneously. Moreover, the forecasts were generated from a relatively small set of variables that are not overly difficult or costly to collect. In addition, reasonable accuracy is achieved despite deficiencies in the data used. Improvements to data collection through a more systematic sampling procedure and more careful data handling would likely further improve forecast performance. This underscores the need to develop standardized data collection procedures and failsafe methods for entering, identifying and storing data.

While several early warning and emergency needs assessment guides exist, our empirical forecasting method has the advantage of demonstrable statistical rigor and out-of-sample performance. Once developed, the model can be easily and regularly updated with new information, each time quickly re-estimating the relevant parameters in a learning process that results in improved performance. Such a forecasting model is an

invaluable tool for emergency awareness and response needs, offering rigorous, cost-effective and practical early warning capacity.

REFERENCES

Alam A., Wojtyniac B., and Rahaman M. (1989), "Anthropometric indicators and risk of death", *American Journal of Clinical Nutrition* 49, 884-888

Anderson, T. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics* 18, 47-82

Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and An Application to Employment Equations", *Review of Economic Studies* 58, 277-294

Barrett, C. (1997), "Heteroscedastic Price Forecasting for Food Security Management in Developing Countries", *Oxford Development Studies* 25 (2), 225-236

Barrett, C. and D. Maxwell, (2005), *Food Aid After Fifty Years: Recasting its role*. Routeledge: London, New York

Chen L., Chowdhury, A. and S. Huffman (1980), "Anthropometric assessment of energy protein malnutrition and subsequent risk of mortality among preschool aged children", *American Journal of Clinical Nutrition*. 33: 1806-1845

Clevers, J. G. P. W. (1988), "The derivation of a simplified reflectance model for the estimation of leaf area index", *Remote Sensing of Environment*, 35: 53-70.

Coppock, D.L., Ellis, J., and D. Swift (1986), "Livestock feeding ecology and resource utilization in a nomadic pastoral ecosystem", *Journal of Applied Ecology* 23(2), 573-583.

Deaton, A. (1985), "Panel data from time series cross-sections", *Journal of Econometrics* 30, 109-126.

Ellis, J.E. (1994), "Climate variability and complex ecosystem dynamics: Implications for pastoral development" in Scoones I., editor, *Living with Uncertainty: New directions in pastoral development in Africa* (London: Intermediate Technology Publication, 1994)

Ellis, J.E., and D. Swift (1988), "Stability of African pastoral ecosystems: Alternate paradigms and implications for development," *Journal of Range Management* 41(6): 450-469

Funk, C., J. Michaelsen, J.Verdin et al. (2003), "The Collaborative Historical African Rainfall Model: Description and Evaluation." *International Journal of Climatology*, 23: 47-66

Galvin, K.A., Boone R., Smith N., and S. Lynn (2001), "Impacts of climate variability on East African pastoralists: Linking social science and remote sensing." *Climate Research* 19: 161-172

Hansen, G.(2001), "A Bias-Corrected Least Squares Estimator of Dynamic Panel Models," *Allgemeines Statistisches Archiv* 85: 127-140

Howe, P. and S. Devereux (2004), "Famine Intensity and Magnitude Scales: A Proposal for an Instrumental Definition of Famine", *Disasters* 28(4):353-372

Intergovernmental Panel on Climate Change, 2001, *Climate Change 2001: Synthesis Report,* IPCC/Cambridge University Press.

Kaitho, R., J. Stuth, J. Angerer et al. (2003), "A New System to Forecast Near-Term Forage Conditions for Early Warning Systems in Pastoral Regions of East Africa", Research Brief 03-02-LEWS

Lentz, E., and C.B. Barrett (2005), "Food Aid Targeting, Shocks and Private Transfers Among East African Pastoralists," Cornell University Working Paper

Luseno, W. K., McPeak, J. G., Barrett, C. B., Little, P. D., & Gebru, G. (2003), "Assessing the value of climate forecast information for pastoralists: Evidence from Southern Ethiopia and Northern Kenya". *World Development, 31*(9), 1477-1494.

Lybbert, T., C. Barrett, S. Desta and D. Coppock (2004), "Stochastic Wealth Dynamics and Risk Management Among A Poor Population", *The Economic Journal* 114 (498): 750-777.

McKenzie, D. (2004), "Asymptotic theory for heterogeneous dynamic pseudo-panels", *Journal of Econometrics* 120: 235-262

McPeak, J. (2004), "Contrasting Income Shocks with Asset Shocks: Livestock Sales in Northern Kenya." *Oxford Economic Papers* 56: 263-284.

Nicks, A.D. (1974), "Stochastic generation of the occurrence, pattern, and location of maximum amount of daily rainfall". p. 154–171. *In* Proc. Symp. on Statistical Hydrol., Tucson, AZ. Misc. Publ. 1275. USDA, Washington, DC.

Rubin, D.B., and N. Schenker (1986). "Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse", *Journal of the American Statistical Association*, 81 (394): 366-374

Stuth J., R. Kaitho, J. Angerer et al. (2003), "Integrating Information and Communication Technology for the Livestock Early Warning System (LEWS) in East Africa" Research Brief 03-01-LEWS

UN Sub-Committee on Nutrition (UNSCN) (1999), "Refugee Nutrition Information System (RNIS), No. 26, Report on the Nutrition Situation of Refugee and Displaced Populations; see http://www.unsystem.org/scn/archives/rnis26

Vella V., Tomkins A., Ndiku J., Marshal T., Cortinovis I. (1994), "Anthropometry as a predictor for mortality among Ugandan children, allowing for socio-economic variables", *European Journal of Clinical Nutrition* 48:189-197

World Bank Group (2003), "KENYA: Arid Lands Resource Management Project", Project Information Document No. AB39

World Food Programme (2000), *Food and Nutrition Handbook*. WFP, Rome

TABLES

Table 1: Available Observations by District, Across Time

YYYYMM	BARINGO	MARSABIT	SAMBURU	TURKANA	YYYYMM	BARINGO	MARSABIT	SAMBURU	TURKANA
Feb-00		9	13		Oct-02	12	9	13	19
Mar-00		7	13		Nov-02	12	8	12	19
Apr-00		9	14	16	Dec-02	11	9	12	19
May-00			14	19	Jan-03	11	9	13	19
Jun-00		8	14	17	Feb-03		9	13	18
Jul-00		9	14	8	Mar-03	11	9	14	18
Aug-00			14	19	Apr-03	12	9	12	19
Sep-00		7	13	19	May-03	11	9	11	19
Oct-00		9	13	19	Jun-03	10	9	14	19
Nov-00		8	14	18	Jul-03	10	9	14	18
Dec-00		7	14	18	Aug-03	10	9	13	19
Jan-01	12	8	14	20	Sep-03	7	9	13	19
Feb-01	11	5	14	19	Oct-03	7	9	13	19
Mar-01	12	6	13	20	Nov-03		7	11	19
Apr-01	12	8	12	18	Dec-03	1	8	13	19
May-01	12	8	13	19	Jan-04	12	9	13	19
Jun-01	11	8	13	17	Feb-04	12	9	13	20
Jul-01	12	8	12	19	Mar-04	12	9	14	19
Aug-01	12	8	11	17	Apr-04	12	9	12	18
Sep-01	12	8	12	18	May-04	12	9	8	18
Oct-01	11	9	13	18	Jun-04	12	9	13	19
Nov-01	12	9	13	19	Jul-04	10	9	11	18
Dec-01	12	9	11	19	Aug-04	12	9	14	20
Jan-02	12	9	14	19	Sep-04	10	9	13	19
Feb-02	12	8	13	18	Oct-04	11	9	14	19
Mar-02	12	9	13	20	Nov-04	11	9	10	18
Apr-02	12	9	14	19	Dec-04	12	8	12	18
May-02	12	9	13	18	Jan-05	11	9	14	16
Jun-02	12	9	14	19	Feb-05	12	7	13	18
Jul-02	12	8	13	19	Mar-05	12	9	14	19
Aug-02	12	9	13	18	Apr-05	11	9		18
Sep-02	12	9	13	19	May-05	12	9		18

Table 2: Model Specification

Dependent Variable		
MUAC Means (Z-Score)		
Explanatory Variables		
Group	**Lag Length**	**Variables**
MUAC Moments	L(1,2,3)	MUAC Mean, MUAC std. dev., MUAC Skewness, MUAC Kurtosis
Herd Dynamics (Large Stock)	L(1,2)	Size, Mortality, Sales, Slaughters
Herd Dynamics (Small Stock)	L(1,2)	Size, Mortality, Sales, Slaughters
Food Aid	L(1,2)	UNIMIX (Kgs and Fraction), Regular (Kgs and Fraction)
Biophysical	L(2,3,4,5,6)	Rain, Rain Squared, Forage, Forage Squared, NDVI, NDVI Squared
Seasonality		Monthly Dummies

Table 3: Forecasting MUAC Means: Model Statistics

		One-Month Ahead	Three-Month Ahead
R-squared:	**within**	0.5297	0.3430
	between	0.9800	0.8806
	overall	0.7848	0.6195
RMSE		0.1666	0.1986
# of Observations		2290	2185
# of Groups		53	53
Observations per group:	**Min**	19	17
	Avg	43.2	41.2
	Max	58	56
Dependent Variable Statistics			
Mean: -1.59	**Std. dev**: 0.40	**Min**: -2.73	**Max**: 0.27

Table 4: Forecasting proportion with MUAC Z-score < -2: Model Statistics

		One-Month Ahead	Three-Month Ahead
R-squared:	**within**	0.5095	0.3391
	between	0.981	0.9022
	overall	0.7489	0.6054
RMSE		0.1205	0.1395
# of Observations		2290	2185
# of Groups		53	53
Observations per group:	**Min**	19	17
	Avg	43.2	41.2
	Max	58	56

Table 5:Model Performance in Generating Correct Decision for Famine Response

Model Type	Forecast Horizon	Confidence Threshold		
		75%	**66%**	**50%**
		Fraction of Correct Decisions		
Proportions	One Month	0.777	0.786	0.785
	Three Month	0.753	0.756	0.758
Means	One Month	0.626	0.631	0.638
	Three Month	0.600	0.608	0.612
		Fraction of Errors that are Type 1		
Proportions	One Month	0.328	0.275	0.214
	Three Month	0.341	0.288	0.212
Means	One Month	0.215	0.180	0.120
	Three Month	0.227	0.194	0.134

Table 6: Testing the Null Hypotheses for Forecasting Bias

Model	Forecasting Horizon	α	β	Prob > F	
				$H_0 : \alpha = 0$	$H_0 : \beta = 1$
Mean	1 month	-0.039	0.981	0.159	0.242
	3 months	-0.146	0.918	0.000	0.000
Proportions	1 month	-0.001	1.019	0.901	0.331
	3 months	0.002	1.001	0.785	0.950

FIGURES

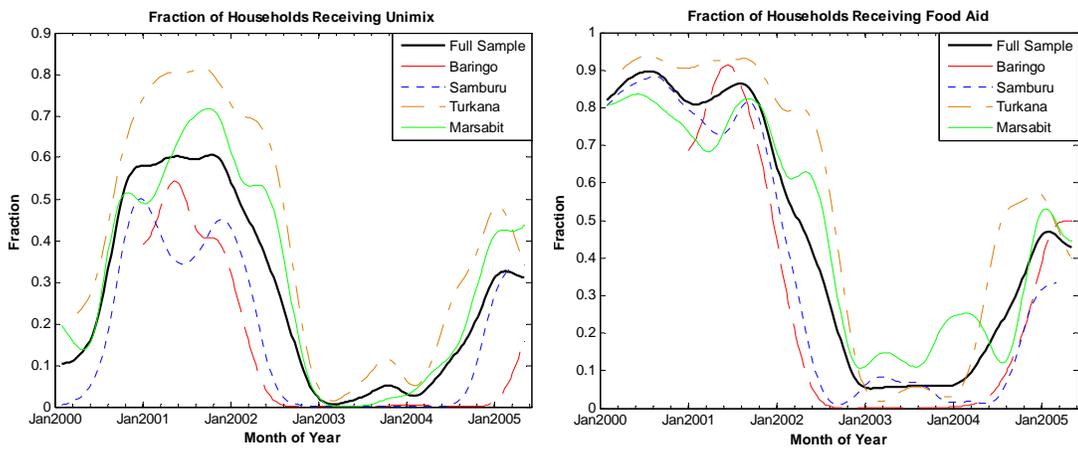## Figure 1: Sample Community Districts and Community Sites

Figure 2

Figure 3

Mean District Monthly Rainfall(mm):

Mean District Forage Availability (kg/ha)

Figure 4

Fraction of Households Receiving Unimix

Fraction of Households Receiving Food Aid
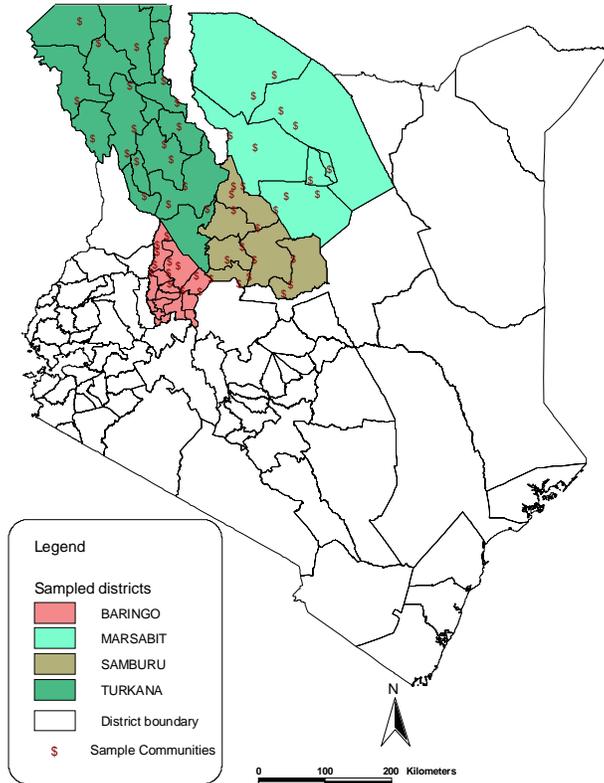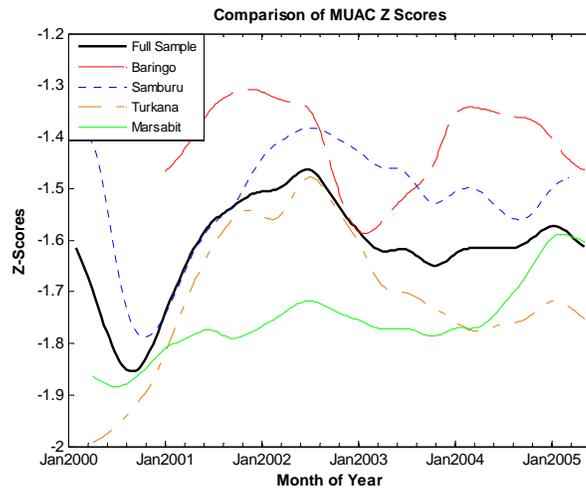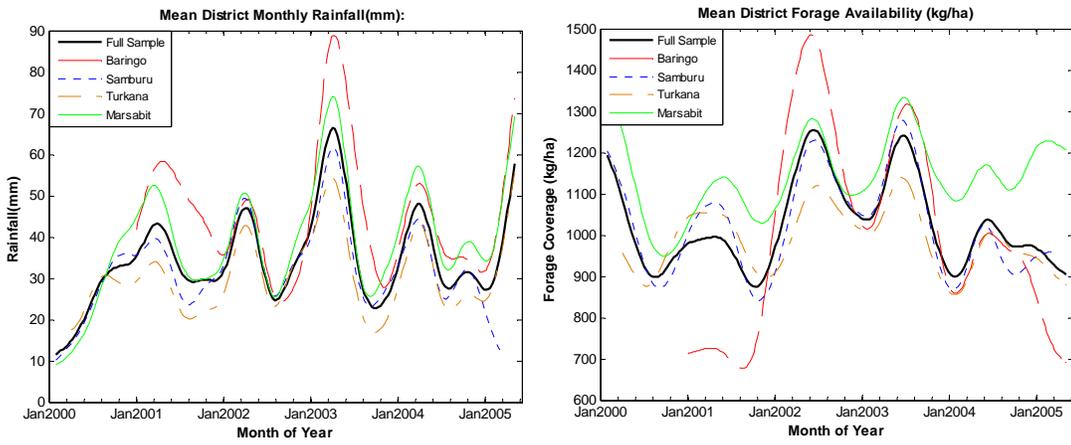
Figure 5

Figure 6

Figure 7

Fraction of MUAC Z-Scores<-2

Figure 8

Mean MUAC:One and Three Month Rolling Forecasts

Figure 9

Fraction of Children with Z-Score Less than -2:One and Three Month Rolling Forecasts

Fraction of Children with Z-Score Less than -2:One and Three Month Rolling Forecasts

Figure 10

Fraction of Correct Decisions by Confidence Threshold

Fraction of Type 1 Errors by Confidence Threshold