

SCALIBILITY OF REMOTELY SENSED LIVESTOCK INDEX INSURANCE IN EAST
AFRICA

A Thesis

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Master of Science

by

Hira Channa

August 2013

ABSTRACT

This thesis presents a scalable approach for an index based livestock insurance product in Kenya. The index is formulated using remote sensing data from the Modis satellite data to capture the systemic risk associated with livestock mortality. A spatial econometric approach is taken to model the response functions associated with the insurance. A key research question that it addresses is the level of aggregation at which to estimate the response function. Through the examination of marginal impacts of parameter estimates and using out of sample log likelihood techniques, it is determined that the optimal level of aggregation for the vegetation effects is at the district level, with division level intercepts/fixed effects. The thesis also examines the impact of using Tropical Livestock Units versus animal specific contracts. It determines that the use of Tropical Livestock Units might lead to mispricing of the contract due to the differences in adaptation capacity of the various animals.

BIOGRAPHICAL SKETCH

Hira Channa completed her Bachelors in Business Administration from the Institute of Business Administration Karachi. She then worked for a year in the banking industry in Karachi, Pakistan. She then came as a Fulbright scholar to the Charles Dyson School of Applied Economics for her MS in Applied Economics.

This thesis is dedicated to my amazing and supportive parents, and my equally amazing siblings
Appe, Yamunana and Hajju.

ACKNOWLEDGMENTS

I would really like to appreciate the support and guidance provided by my advisors Dr Joshua Woodard and Dr Calum Turvey. They remained engaged with me throughout the process and were readily available whenever I needed help.

Table of Contents

Table of Contents.....	vi
Chapter 1-Introduction.....	1
1.1Problem Statement	1
1.2 Research Objectives	2
1.3Background of the Area	3
1.4 Research Partners	5
1.5 Organization of Thesis	5
Chapter 2-Literature Review	6
2.1 Index Based Insurance	6
2.2 Challenges with Traditional Insurance	8
2.3The Flip side: Basis Risk	11
2.4 Remotely Sensed Vegetation Data	14
Chapter 3- Data.....	17
3.1 Data Availability	17
3.2 Index Data	17
3.3 Mortality Data	20
3.4 Missing Data	21
Chapter 4-Model Discussion	25

4.1 Functional Form	25
4.2 Initial Analysis	25
4.3 Spatial Lag model.....	29
4.4 Maximum Likelihood Estimation	31
4.5 Spatial Model Steps.....	32
Chapter 5-Results.....	34
5.1 Response Functions.....	34
5.2 Model Permutations	36
5.3 Non-Spatial Models.....	38
5.4 Spatial Models.....	42
5.5 Comparing non-spatial and spatial models	45
5.6 Using out of sample log likelihood	46
5.7 Weight Matrices	48
Chapter 6-Tropical Livestock Units	51
Chapter 7- Verifying the model through simulation.....	54
Chapter 8-Final Contract Design	57
Chapter 9-Conclusion	59
Bibliography	60
Appendix.....	64

List of Figures

Figure 1.1-Map of the Arid and Semi-arid Lands in Kenya	3
Figure 3.1-Examining seasonal nature of contract	19
Figure 5.1 –Division versus District Level Response Functions for non-spatial LRLD model ...	39
Figure 5.2 – Division versus District for non-spatial SRSD model.....	41
Figure 5.3-Examining District and Division Level Model Results-LRLD.....	43
Figure 5.4-Examining District Level and Division Model Results-SRSD.....	44
Figure 5.5-Comparing Spatial Model and Non-Spatial Model Results-LRLD	45
Figure 6.1- Animal Specific Contract versus TLU	53
Figure 7.1 Plotting out parameters across simulations	55
Figure 9.1-Livestock Production Systems	64
Figure 9.2-Plotting Different Kinds of Response Functions	65
Figure 9.3 Comparing seasonal differences in mortality	65
Figure 9.4-Examining different ways of plotting response functions.....	66
Figure 9.5-Remaining plotted tables for simulation exercise	66

List of Tables

Table 4.1 -Description of Variables Used.....	25
Table 4.2- Initial Analysis of LRLD Data	27
Table 4.3- Initial Analysis of SRSD Data.....	28
Table 5.1-Model results of aggregation in non-spatial models.....	38
Table 5.2- Results of aggregation in non-spatial models.....	40
Table 5.3-Comparing aggregation in spatial models-LRLD	42
Table 5.4-Examining Aggregation Results in Spatial Model-LRLD	44
Table 5.5-Comparing different weight matrices-LRLD	50
Table 6.1-Examining Animal Specific Functions-LRLD.....	52
Table 7.1-Setting Parameters for simulation-LRLD.....	54
Table 9.1-Percent Losses per District-.....	58
Table 9.1 Percentage Distribution of Magnitude of Shock on Households by Region	67
Table 9.2 Missing LRLD TLU Data.....	68
Table 9.3-District Codes and Names	68
Table 9.4-Livestock Density.....	68
Table 9.5- Figures for Total Livestock Averaged Across Seasons.....	69
Table 9.6-Division Specific Losses	72

Chapter 1-Introduction

1.1 Problem Statement

The pastoralist lifestyle is extremely vulnerable to extreme weather conditions. Drought in particular can be devastating, with the potential to destroy the asset base of entire communities, making them dependent on external food aid for survival. Because such weather related disasters impact the entire community, local mechanisms of handling such risks, such as self-insurance or social insurance, are typically ineffective in extreme events.

For example if a community member experiences the loss of some of his livestock, it is likely that social insurance mechanisms will kick in and protect him from total devastation. This is done with the expectation of reciprocity. Severe weather incidents like droughts, however, destroy a community's own capacity to manage without outside help.

The perception that pastoralist communities should be absorbed into the larger sedentary communities has now given way to the realization that pastoralism is actually an efficient utilization of resources. There is, however, also the reality that devastation wrought by serious droughts and continually degrading rangelands, which are the lifeline for pastoralists, has seen a drop in numbers in pastoralist communities in many regions within Africa. Many previously pastoralist households have moved to urban and peri-urban areas seeking temporary employment and living on the goodwill of relatives. This migration has also resulted in entire communities becoming heavily dependent on external food aid. This highlights the importance of effective risk transfer mechanisms which allow communities to manage systemic risks. This thesis is

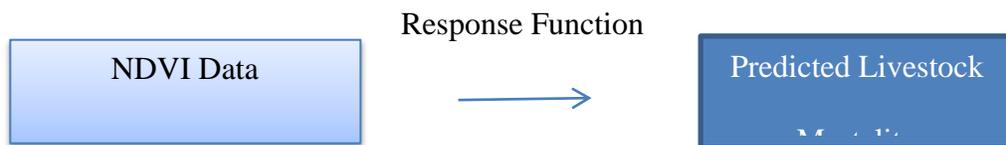
focused on the design and development of an index based insurance product which protects pastoralist livestock producers.

1.2 Research Objectives

The key objective is to utilize available livestock mortality and remote sensing NDVI (Normalized Difference Vegetation Index) data to design an index based livestock insurance to hedge weather-related risk that cause livestock mortality. The objectives can be described in detailed as follows,

- 1) To design a method that allows us to define the most unbiased and efficient relationship possible with the available data between livestock mortality and remote sensing data.

This is accomplished by estimating response functions livestock mortality to NDVI.



- 2) To investigate the prospect of designing animal-type specific contracts, as opposed to aggregate animal-type contracts.

Livestock pastoralism is a way of life for survival in the arid and semi-arid lands (ASALS) for nearly three million Kenyans.¹ It mainly involves the breeding and management of livestock such as goats, camels, sheep, and cattle. The livestock is used for multiple purposes ranging from milk production for household consumption to the trading of livestock. In essence, the livestock is the key asset of the household and is essential for their survival. There has been a push to better understand the challenges faced by communities living in the ASALs in Kenya and to design systematic processes which enable communities to better manage the risks they face. This was partly motivated by the devastation caused by the recent droughts in 2008-2009.

Results from the Kenyan Household Survey, carried out in 2005 (GoK 2005), indicate that communities in the arid and semi-arid areas tend to face high levels of systemic risk. For example 74.5% of responses from Tana River district indicated that the shocks experienced by their households were also felt by all the other households in the community. While this figure was lower for some arid districts the average of 49% when compared to the overall Kenyan average of 23% suggests that communities in these regions are overall more susceptible to covariate risk.² These figures further highlight the vulnerability of groups living in these districts to external climatic events.³

¹ Arid areas are defined as those with Rainfall between 0-500mm and a growing season of <90 days

² District Specific information available in appendix. The Ijara district was carved out post this survey from the Garissa District. In the context of this research work this survey the Garissa district has data from Ijara and Garissa

³ The survey asked about recent shocks that communities had faced

1.4 Research Partners

This thesis is a part of a much larger project which is a collaborative effort involving the International Livestock Research Institute with partners at Cornell University, Syracuse University, and the BASIS Research Program. The purpose behind the broader project is to design a scalable and sustainable index-based insurance market that protects pastoralists in the Arid and Semi Arid Lands in Kenya and Ethiopia. A pilot was launched in the Marsabit District in Kenya in 2010. The insurance in the pilot product was provided by UAP Insurance Limited and Equity Bank Limited directly engaged the clients. For the purposes of this thesis, the International Livestock Research Institute provided livestock mortality data from the ALRMP dataset.

1.5 Organization of Thesis

The first section of this thesis reviews previous research conducted on Index Based Livestock Insurance (IBLI) and the benefits and challenges of working with index based insurance products. The next section discusses the nature of the data that and specifically the challenge posed by the lack of livestock mortality data in some regions/years. The next examines the modeling procedures used to estimate response function and the methods developed to address missing data. Finally, an examination of tropical livestock units versus animal specific contracts is explored.

Chapter 2-Literature Review

2.1 Index Based Insurance

Risk Management Techniques amongst Pastoralists

Communities who live with the burden of this kind of systemic risk without access to adequate risk management tools are likely to invest less in livestock, their key asset, because of the high risk of mortality in bad weather which can result in overall lower accumulated savings.

Exposure to this kind of systemic risk also leaves commercial lenders wary and unwilling to lend, further depriving communities to access to potential sources to credit. (Barnett, Barrett et al., 2008)

Traditionally growth theories suggest the presence of a unique equilibrium. The presence of such an equilibrium suggests that the effects of shocks are temporary because it implies that individuals and communities will eventually fall back into the unique upward moving equilibrium. The concept of “multiple dynamic equilibria” indicates, however, that there is no one unique growth path or natural equilibrium. This leads to the conclusion that shocks do not have only temporary effects and can in fact throw households into a lower income equilibrium.

It is also important to note that the complex and multi-dimensional role that livestock plays in the lives of pastoralists. Livestock is not only the key source of income it also plays a key role in terms of providing food security. Milk specifically is a significant source of nutrition especially amongst children.

The issue of systemic Risk

There is a great deal of covariate risk in agricultural activity, which is the main economic engine in almost all rural areas. Standard insurance works on the principle that a large number of uncorrelated observations reduce variance thereby reducing the risk. However in agriculture, weather events such as drought are highly spatially correlated. Natural disasters such as these also damage a community's ability to survive shocks, and since everyone in the community is affected the local support network breaks down.

While local money lenders, family and friends, and semi-formal setups such as ROSCA's (Rotating and Saving Credit Associations) can be an effective tool to deal with idiosyncratic risks that farmers or pastoralists experience, the very local nature of such setups means that they cannot spatially diversify enough to deal with covariate risk. This is also true for self-insurance techniques such as selling semi-liquid assets, because surplus might bring down prices. A review report examining pastoralism when talking about risk management techniques amongst pastoralists by Hatfield (2006) points out that:

“.. the safest place to store wealth is on the hoof, where the returns to investment are substantial during good times, but with the inevitable downfall that its value is eroded at exactly the time when it is most useful.”

The high level of systemic risk in agricultural is one of the key reasons why even in countries with highly developed financial systems, agricultural insurance is usually heavily subsidized by the government. In a well-known paper which highlighted the role of systemic risk in agricultural insurance in the USA the authors write:

“The correlation induced among individual losses by these systemic weather effects raises the risk associated with holding a portfolio of individual crop insurance contract liabilities above what it would otherwise be if individual losses were stochastically independent... (Miranda and Glauber, 1997)”

Running simulations on actual yield data the authors estimate that the absence of stochastic independence in yields across farms translates into insurance portfolios which are 50 times riskier. This would imply extremely high premiums when translated into a financially viable product for insurers, and would eventually make it unaffordable and unattractive for most farmers.

2.2 Challenges with Traditional Insurance

Countries with lesser developed financial systems and overall infrastructure tend to face other challenges which compound together to make delivery of formal financial services such as insurance products difficult. These have been well documented in the microfinance literature with some of the key ones including lack of contract enforcement, adverse selection, moral hazard, and high transaction costs.

Adverse Selection leads to a situation where because of asymmetric information, providers of the financial service do not correctly assess the risk of the purchasers of the financial product, and the buyer knows better their true risk. Generally the issuers of the insurance will underestimate the risk of the insured party and end up underpricing the product for some, and overpricing for others. This incorrect pricing leads to higher losses than predicted as underpriced buyers will tend to purchase more than overpriced buyers. This, in turn, will lead to an increase in the prices

charged by the insurer in later periods, which will further attract a higher risk customer base. Eventually asymmetry in information will lead to a market failure. Adverse selection is more apparent in developing countries because a less developed infrastructure system makes information more expensive and difficult to acquire. It is easy to see that this issue is further compounded when working with semi-nomadic communities in rural areas.

Moral Hazard is more specific to insurance products where because of the insurance product the client engages in risky behavior which the insurance provider had not accounted into their pricing beforehand. This again leads to mispricing of the product which will eventually lead to the market failure of the product. In addition to these challenges, the logistics of tracking and verifying claims associated with insuring individual livestock can lead to prohibitively high transaction costs for insurance products.

Index Based Insurance for farmers and pastoralists

The concept behind index based insurance is to design a product to better manage covariate or systemic risk in agriculture. While most frequently applied to crops it has occasionally been used within the context of insuring livestock as with a World Bank Project in Mongolia (Mahul, Belete et al., 2009), and in a more recent project that is the subject of this thesis, in the ASALs of Kenya (Chantarat, 2012; Woodard, 2013), among many others.

The key difference between Index based insurance and other insurance products is that index based insurance products payouts are based on the movement of a certain index which should be correlated with overall loss in a specified region but independent of the actual loss that an individual might experience (or at least not able to be influenced by individual action). Index

based products have been used in many sectors outside of agriculture for as well, for example in electricity markets in the form of weather derivatives as risk management tools. There is an emerging opinion that index based products can be an effective risk management tool for agriculturalists in developing countries as well (Jerry R. Skees, 2006).

Index Based Products are linked to an independent index ensuring that there are no concerns related to adverse selection and moral hazard. This in turn also insures that the index based products have lower transaction costs when compared to traditional insurance products, a major concern when considering small-holder farmers in developing countries. One of the key determining factors behind the successful implementation of index based insurance products is the design of an index which is highly correlated with the covariate risk that the insurance product is covering (Smith and Watts, 2009). The lack of an effective index basically indicates that the product has failed to capture the risk which it was designed to manage. This failure is likely to lead to dissatisfaction amongst clients and act as a deterrent towards repurchase. Also, despite the many recorded benefits of index based insurance projects in developing countries, few have moved beyond the pilot stage.

The largest in terms of number of farmers is the index insurance provided by the Agriculture Insurance Company in India. The original version of this crop insurance scheme initiated in 1999 and covered nearly 22 million farmers in 2010, and uses an area-yield as an index. This program, which was primarily a government run program focused in crops, placed a cap on the premiums that farmers paid with the central and state government subsidizing the rest at the time of claim payment. This program faced many challenges not the least of which was that farmers received 3.5 times in claims the amount that they paid in premiums. Perhaps even more problematically

the yield estimation calculation used has proven to be inaccurate, resulting in ‘arbitrary allocation of government subsidies’ (Mahul and Verma, 2012).

There is now a modified version of this insurance product ‘Modified National Agriculture Insurance Scheme’ launched in 2010 which combines weather and yield data for the development of the index. This product has also been opened to the private sector with the government stepping in by providing premium subsidies instead of covering the balance at the time of claim settlement.

2.3 The Flip side: Basis Risk

One of the key challenges associated with index based products is the element of basis risk, which does not typically exist in traditional insurance products (Woodard and Garcia, 2008b). This kind of risk implies that those suffering losses (in terms of yield loss or livestock mortality) might not receive indemnity payments, while those farmers who have not experienced any losses might end up receiving payments. The risk of this error depends a great deal on the index being used and its correlation with the loss experienced by farmers.

Average area yields are frequently used as an index for crop insurance. As discussed earlier India’s yield based index insurance program is perhaps one of the most extensive in the developing world. However research indicates that the use of 3-5 year averages for determining product yield has led to inappropriate risk classification of crops and adverse selection. The challenges faced by the NAIS product reflects on the importance of designing an accurate index.

Alternate index options include precipitation and temperature. Other relatively newer include those of vegetation indices, of which the Normalized Difference Vegetation Index (NDVI) is the most popular.

Closer examination of the rainfall index, which has been a frequently used index in insurance products, reveals a great deal of variability in correlation between the index (rainfall) and the yields that are being insured (Woodard and Garcia, 2008a). Depending also on specific crops this correlation can vary from a low of 0.2 to a high of 0.82. A paper examining basis risk and how it varies with correlation between rainfall and farm yields uses a simple concept where it accounts for correlation between yield and precipitation at the farm near the weather station and then multiplies it with the correlation for farms further away from the weather station. Their calculations indicate that even where there are high correlations of 0.82 between the index and the yield there is an estimated 0.168 probability that farmers will receive no indemnity payment when they are actually experiencing severe losses (Smith and Watts 2009). While the valuation of this probability will vary extremely depending on the part of the world, the index being used and the spatial aggregation of the index, it serves to highlight that basis risk can be a serious deterrent towards the success of index based insurance products.

The increasing movement towards using of satellite based data is an attempt to design indices which are more highly correlated with yields in order to reduce basis risk, or to reduce transaction costs where weather station data or local level yield data do not exist or are too expensive to be collected. Some potential challenges with indices using remotely sensed data is that they make contracts more complicated. Also, more research is still needed to determine their performance in terms of reducing basis risk.

Additional Challenges with index based products

Another concern that emerges is that of willingness to pay amongst target consumers. This is a major concern in projects like those in the Kenyan ASALs where the target customers are livestock farmers with relatively small herds and few other sources of income. Therefore, affordability and liquidity are serious concerns. Research on the rainfall based Index Based Insurance Product in India initiated by BASIS has indicated that take-up of the product is extremely sensitive to the pricing. Evidence from Malawi also indicates that take-up of the product was highly correlated with household wealth (Giné, Townsend et al., 2007)

This issue can be dealt with in a number of ways, the simplest of which is direct government subsidies. Some of the largest projects in index based agriculture insurance in USA and India have been heavily subsidized by the government. Other methods used include requiring farmers to work on community projects in exchange for insurance. Yet another alternate could be collaborating with others in the supply chain, for example fertilizer and seed suppliers, to reduce the eventual cost of the premium for farmers.

Research has indicated that more information and training provided to consumers increases take-up of the product, however, such training comes at an increased cost to the provider and might eventually end up being reflected as high premiums (Chantarat, et al., 2012) It is also important that the insurance is provided by organizations which the customer base trusts. Also key are in-country human resource capacity and a local regulatory framework. Initial results from games amongst pastoralists in the ASALs in Kenya on demand for Index Based Livestock Insurance

indicate that having other sources of income outside of livestock and subjective assessment of loss are key factors that affect choice of initial coverage (i.e., percentage of livestock that the household head chooses to insure initially). The research also finds that, conditional on the coverage chosen, willingness to pay is perhaps well below the commercially viable rate. Also, contrary to earlier hypotheses, those with smaller herds prove to be less price elastic, and overall more willing to pay the market premium than those with larger herds.

2.4 Remotely Sensed Vegetation Data

The goal of remote sensing projects is to give some idea of the extent and type of vegetation on the surface. Vegetation Indices are generally modified in order to account for atmospheric and surface differences across regions in order to draw more accurate pictures of what is actually happening on ground. Vegetation Indices are computed using multiple spectral bands. They can either be in the form of a ratio or in the form of linear equations. Vegetation indices are based on the concept that light reflectance is different for bare soils as opposed to areas with vegetation, which in turn implies that the value of the index can be an expression for the density of the vegetation on the area.

A key benefit of using remotely sensed data is that in the context of many developing countries it is available at much greater spatial density when compared to ground weather data like precipitation. This means that it can be used independently of weather data or enhance sparse weather observations that are available.

This paper utilizes the Normalized Difference Vegetation Index, which is one of the most commonly used vegetation indices. NDVI is calculated as follows,

$$NDVI = NIR - VIS / NIR + VIS$$

where 'VIS' denotes the reflection of visible light, and NIR denotes reflection of near infrared light. Generally healthy vegetation will absorb more of the visible light and reflect back more of the infrared light when compared to less healthy vegetation. NDVI can have values ranging from -1.0 to +1.0. Higher values indicate high concentrations of healthy vegetation while extremely low/negative values indicate that the area is either barren or possibly a water body (Ceccato et al., 2013)

Increased interest in NDVI for the purposes of monitoring forage and crop growth amongst other factors has resulted in improvements in NDVI satellite data that are processed and available for use. NDVI data is now available at extremely high spatial resolutions, below 10 meters in certain areas, a big step forward from the initial 8km by 8km resolution Advanced Very High Resolution Radiometer (AVHRR) data. Improved algorithms can also better account for elements like cloud cover to which NDVI values can be very sensitive. One of the most commonly used techniques is called "Maximum Value Compositing" which takes into account only the maximum NDVI value within a specific time set in order to reduce the effect of cloud cover. The data that we use for the purposes of this research (MODIS) uses a more complex algorithm which is designed to minimize the impact of cloud cover on readings and get the readings that are recorded at the 'nadir' (horizon) (Carroll et al., 2013)

The United States Department of Agriculture (USDA) has designed an index insurance product which uses NDVI data as well as rainfall data to measure forage conditions in specific areas. The Agriculture Insurance Company of India is considering using NDVI data specifically to design a

product for wheat growers. Other projects which plan to use NDVI data include crop insurance projects in Ethiopia which involve creation of indices which measure crop growth and monitor droughts.

However, in a paper which closely examines prior research on NDVI data and its correlation with weather variables and vegetation biomass, the authors point out that a number of factors such as cloud cover and solar zenith angles amongst others can affect the correlation between NDVI and biomass (Turvey and McLaurin, 2012). Additionally while NDVI data preprocessing has improved considerably over time, these factors can be extremely complex and variable over locations making it difficult to devise a standard way in which to deal with them. That work also examines USA crop yields, NDVI data, and weather variables for temperature and precipitation. Their results show variability in significance and sign depending on districts and the form of the weather variable used. This complexity is further highlighted in a paper which examines crop yields and NDVI data in 9 districts in Zimbabwe for the purposes of offering index insurance for drought protection. Their results indicate a great deal of spatiotemporal variability with correlations varying from -0.11 in the early season to values of 0.71 in the late season for certain districts. Correlation between NDVI data and maize and cotton are also found to be different (Makaudze and Miranda, 2010).

A key conclusion that can be drawn from this research is that when using NDVI data as a predictor for vegetation growth, it is critical to examine localized relationships between variables and to examine closely what levels of spatial aggregation appear to be the most reasonable.

Chapter 3- Data

3.1 Data Availability

During 2013 the political division system in Kenya changed to a system with 47 distinct counties. However, since the data that this paper utilizes is prior to 2013 we will be utilizing the previous administrative system which uses districts.⁴ Under this system, Kenya was divided into eight provinces, which were divided into 46 districts, which in turn were divided into 262 divisions.⁵ There are also further sub-administrative boundaries which are referred to as ‘locations’ and ‘sub-locations’.

The MODIS NDVI data is available at a high pixel density of 250 meters for the entire region which we are considering. The mortality data is household survey data collected by the Kenyan government. The smallest administrative division at which the households are tagged is at the ‘sub-location’ level. The common data that we had for livestock mortality and NDVI data is for a total span of 11 years. Detailed data information are provided below.

3.2 Index Data

Processing

As noted, this work employs Normalized Difference Vegetation Index from the MODIS Satellite for the creation of the index.⁶ It has been downloaded from the Global Land Cover Facility

⁴ Post 2010 the division numbering also changed.

⁵ Excluding Nairobi.

⁶ MODerate-resolution Imaging Spectroradiometer

(GLCF) website, where it is available at a pixel density of 250 meters for every 16 days (NASA, 2011). It is available in ‘raster’ format, with one file for each of the specific time frames. The initial step after the download of the data involves creating a boundary raster file of the same size as the files with the unprocessed NDVI data. We then standardize NDVI data across location and time to account for physiological variations across the region.

The data are aggregated across each division, season, and year, and a z-score calculated as follows,

$$zndvi_{pdt} = \frac{ndvi_{pdt} - E_{pd}(ndvi_{pdt})}{S_{pd}(ndvi_{pdt})}$$

where p is the pixel in a specific location, t refers to the year, while d refers to the specific 16-day time slot in the year. This value can then be averaged across pixels and across time to determine the cumulative ($czndvi$) value for a specific location. Separate values for each of two seasons are calculated.

Another method investigated to standardize values is to aggregate the NDVI values for all the pixels for each division, and then calculate the standardized cumulative value. A hypothesized benefit of this could be potentially that since each of the divisions have a different number of pixels, the variance of the final measure could differ simply due to the size of the division, thus complicating estimation. Aggregating all the pixels first before standardization could then result in a better and more consistent fit in a spatial model. A closer examination of the method indicates that data from each of the methods are highly correlated (0.95).

from March-September. Potential customers can purchase a contract for the next season in the last two months of the present season and is in effect for two seasons. For example, a contract for the SRSD season can be brought until the month of September.

An exploratory analysis involving plotting out of the quadratic function of the two different seasons and the aggregated annual data indicates that the aggregated annual data behaves differently from both the seasonal graphs. This is because livestock mortality is lower in the SRSD season than in the LRLD season typically, and although the two seasonal graphs are correlated, the aggregated graph tends towards predicting lower mortality for the same *czndvi* values. We continue to assume two different response functions in all our models for the two seasons.

3.3 Mortality Data

The livestock mortality data available for the estimation of the functions and the eventual design of this contract were from the Arid Land Resource Management Project (ALRMP) survey conducted by the Kenyan government. The dataset is a household level survey with a total of 900,000 data points from the arid and semi-arid districts in Kenya. The number of data points for each divisions (a sub-administrative category to district) varies across division. Although the data extends from 1992-2012 it does not follow the same households across the 21 years.

The data are initially aggregated for each division across households. The next step is to sum up the data for each of the months that we have according to the specific season and year to which it belongs. After this stage, this value is divided by the highest value of mortality from the initial monthly aggregated value as,

$$M_{l,s} = \sum \bar{H}_{mort,m} / \text{Max}_{(m \in s)}(\bar{H}_{begm}) \quad .$$

3.4 Missing Data

There is a great deal of literature in the world of econometrics on how to deal with and work around the problem of missing data. Generally ‘missingness’ (Schafer, 1999) is not the key focus of the research and it is generally treated as a problem that needs to be managed to conduct further analysis. In the empirical research world, the extent of missing data and the processes behind it can be extremely varied.

The data that used to calculate mortality values for specific regions was not originally collected for this purpose. It is household level data that was collected to gain information about the challenges faced by pastoralist communities living in the arid and semi-arid areas in Kenya and the progress of these communities in terms of development indicators across time. For this research, missing data is in those areas for which no data was originally recorded.

One of the most common approaches in dealing with missing data amongst researchers in dealing with research is that of ignoring observations which possess missing data through methods such as list-wise deletion. However in the context of this case, the practical necessity of designing contracts for each of these regions and the spatial nature of the models employed means the divisions for which no data exist cannot just be ignored.

Multiple imputation is a process in which predicted values are generated as a substitute for missing data (Fichman and Cummings, 2003). Broadly speaking, multiple imputation uses available data to provide some information about the missing data. The concept of generating

multiple values for the same missing data point is to incorporate the concept of uncertainty and variability in the data.

A technique that can be considered a kind of multiple imputation is the 'Markov Chain Monte Carlo' method. This process basically involves imputing multiple values through a simulation process which recalculates the means, variances, and covariances of the missing data. Generally the multiple imputation process requires the running of a large number of iterations, where the variables for the next iteration are randomly drawn from the error distribution from the previous model. The prior data can generally be non-informative and should not affect the imputed variables. A drawback perhaps of this technique for dealing with missing data is that it requires the use of an analysis model which requires making assumptions about the distribution of the missing data, or may not converge (Fichman and Cummings, 2003).

The literature on dealing with missing spatial data is still evolving and somewhat sparse. This literature is almost non-existent in the context of dealing with applied problems. In the spatial econometric literature, LeSage and Pace (2004) investigate the problem of missing data in the lag version of a spatial model. The paper seeks to refine the model used for housing prices by incorporating information from unsold properties. The prices of unsold properties are treated as missing data.

The initial stage requires the estimation of the missing observation based on the expectation of the available data. The paper conducts this by first estimating the parameters of a spatial panel model using the maximum likelihood estimates from the data that were actually available. The next step is to calculate the missing values using the parameters estimated in the step before. The

eventual step is to estimate the full log-likelihood which includes the process of the sold properties and the properties of the unsold properties.

In addition to the challenge posed by missing data, closer screening indicated that the data at some points were potentially erroneous. The reason behind this concern was that even with initial inspection of the data it was evident for some points that high or extremely low NDVI values were associated with extremely low and implausible mortality values. To minimize the potential inaccuracies in the data, two techniques were used to systematically identify these data points and to treat them as missing. Initially extremely low mortality values below the 10th percentile were identified and then examined to see if any of these mortality values corresponded with extremely high (higher than the 85th percentile) or extremely low (below the 15th percentile) previous and within season *czndvi* values. If the mortality values did not correspond with any of these four categories, then those values were deemed to be highly likely to be incorrect and were treated as missing data. Manual inspection of these data indicated that most of the values dropped were unreasonably low (less than 0.01) and appeared to be due to missing household values being treated as zeros in the ALRMP data.

The mortality data that are used for the design of this contract are aggregated across a specific season in a year for one particular division. To ensure that the mortality values used are reflective of the entire division, divisions for which less than 10 data points were available for aggregation are dropped from the data and treated as missing.

When dealing with climate data and mortality of livestock owned by pastoralists, there is good reason to expect that spatial patterns and relations exist between the data. The rationale behind

the process used to estimate the missing data and the response functions is the utilization of the information within these spatial patterns. The model that used to deal with the challenge of missing data makes the assumption that spatial patterns of livestock mortality are similar for regions with missing and non-missing data. This would appear to be a reasonable assumption specifically when concerned with the impact of climatic conditions on mortality. More importantly, because when working with livestock owned by pastoralists it is expected that mobility, which could involve mobility across borders, is a key part of their lifestyle.

The model that used for this paper combines elements of both these techniques. It has been designed specifically keeping in mind the nature and availability of the data. A detailed description of the model employed is presented below.

Chapter 4-Model Discussion

4.1 Functional Form

Table 4.1 provides a summary of the explanatory variables explored in this study.

Table 4.1 -Description of Variables Used

Pre_czndvi	Czndvi values for the previous season. For example the czndvi value for the prior SRSD season when we are looking at a LRLD season.
Czndvi	Czndvi values for the present season.
X	This is the longitude of the centroid of the polygon measured in degrees constituting a particular division measured.
Y	This is the latitude of the centroid of the polygon measured in degrees constituting a particular division.

The squared value for the response functions were included because on-ground information indicates that increased wetness (reflected in high positive czndvi values) also result in increased mortality due to a combination of disease and insects. We use the latitude and longitude coordinate data for the Cassetti spatial expansion models.

4.2 Initial Analysis

Initially, before we move into spatial models, we examine the data using non-spatial models. One of the simplest models would be to estimate response functions using an Ordinary Least Squares Regression. This would involve running an initial regression using the data that is available. The simplest model would be a regression using the *czndvi* value for the present season and the prior season. Another option to add further complexity to the model would be to employ fixed-effects. This would allow the intercept of each division to vary separately.

Another model which would allow the introduction of different impacts for different divisions is the Casetti Expansion Model (Casetti, 1972). The implementation of this model involves multiplying each of the *czndvi* values with the latitude and longitude coordinates as follows,

$$X = X_0 + pX_0$$
$$X = X_0 + qX_0$$

The expansion is of the form as given in the above equation where the *X* matrix constitutes the *czndvi* values for the present season and the previous season and their squared values and the '*p*' value constitutes the *X* coordinates and the '*q*' value constitutes the *Y* coordinates. The significant coefficients for the expansion with the *czndvi* value for the previous season indicate that spatial expansion effects are present.

The next step in this process would be to estimate data for the missing values. We possess 11 years of complete *czndvi* data for the present and previous season for the 108 divisions. If we do not have a corresponding mortality value for that division for that year for the season we treat that as a missing value. We utilize the slope intercepts that we estimated using the available data earlier. Since our initial missing values are estimated with a fixed-effect model we replace the intercepts for divisions for which we could not estimate intercepts by first replacing the missing values with the average of the intercepts which we did estimate. In order to utilize the spatial information that we possess, we could replace these averages by spatially weighted averages by

Table 4.2- Initial Analysis of LRLD Data

Parameters	Sparse OLS	OLS Model	OLS Model with fixed effects	Casetti model with X and Y coordinates	Casetti model with X and Y coordinates and fixed effects
Intercept	0.14140*** 20.09253	0.10411***	0.10534	0.10548 9.6995	0.09574
Pre_czndvi	0.00088 0.62296	-0.00219 -1.39741	-0.00159 -1.09135	0.07675*** 2.48331	0.07992*** 2.55309
Czndvi	- 0.00676*** -5.72167	- 0.00676*** -5.80043	- 0.00696*** -6.52475	-0.03368 -1.4075	-0.02403 -1.0884
Pre_czndvi ^2		0.00057*** 3.54852	0.00048*** 3.05151	-0.00755** -1.97668	-0.00894** -1.99636
Czndvi ^2		0.00035*** 2.86874	0.00040*** 3.4675	0.00262 1.14995	0.00082 0.31361
Pre_czndvi *X				-0.00195*** -2.42758	-0.00205*** -2.51458
Czndvi *X				0.0007 1.09847	0.00045 0.76106
(Pre_czndvi ^2)*X				0.00020** 1.98294	0.00024** 2.05426
(Czndvi ^2)*X				-0.00006 -0.98622	-0.00001 -0.08867
Pre_czndvi *Y				-0.00212** -2.24381	-0.00183** -1.97002
Czndvi *Y				0.00011 0.12672	0.00014 0.17643
(Pre_czndvi ^2)*Y				0.00025*** 2.48421	0.00014 1.26043
(Czndvi ^2)*Y				0 -0.00939	-0.00011 -1.25352
R-squared	0.0601	0.0926	0.3687	0.1341	0.3903
Sum of Squared Errors	0.0299	0.0289	0.0234	0.0280	0.0299

Notes-The above table presents the results of the models estimated using available data. This consists of 627 observations. We report the tstat for each statistic below the value. Those figures significant at a confidence level of 0.95 or above are in bold

All these results are only for the LRLD season

***Significant at a confidence level of 0.99, **Significant at a confidence level of 0.95

Table 4.3- Initial Analysis of SRSD Data

Parameters	Sparse OLS Model	OLS Model	OLS Model with fixed effects	Casetti model with X and Y coordinates	Casetti model with X and Y coordinates and fixed effects
Intercept	0.10599*** <i>18.60316</i>	0.09981*** <i>11.58693</i>	0.09757	0.10004*** <i>11.64107</i>	0.09600
Pre_czndvi	0.00071 <i>0.80916</i>	0.00080 <i>0.88424</i>	0.00041 <i>0.46742</i>	0.01351 <i>0.71830</i>	0.00810 <i>0.43813</i>
Czndvi	-0.00227** <i>-2.14855</i>	-0.00283*** <i>-2.43863</i>	-0.00259*** <i>-2.26918</i>	0.03414 <i>1.52061</i>	0.02199 <i>0.94722</i>
Pre_czndvi ^2		0.00002 <i>0.21294</i>	0.00004 <i>0.38326</i>	-0.00179 <i>-1.02052</i>	0.00002 <i>0.01127</i>
Czndvi ^2		0.00017 <i>1.18095</i>	0.00016 <i>1.18008</i>	-0.00032 <i>-0.65970</i>	-0.00548** <i>-1.97977</i>
Pre_czndvi *X				-0.00093 <i>-1.60338</i>	-0.00020 <i>-0.41843</i>
Czndvi *X				0.00004 <i>0.92953</i>	-0.00064 <i>-1.06050</i>
(Pre_czndvi ^2)*X				0.00010 <i>1.56495</i>	0.00000 <i>-0.01565</i>
(Czndvi ^2)*X				-0.00042 <i>-0.74866</i>	0.00014** <i>1.98015</i>
Pre_czndvi *Y				-0.00064 <i>-0.96440</i>	0.00000 <i>0.00383</i>
Czndvi *Y				0.00013** <i>2.48478</i>	-0.00004 <i>-0.05686</i>
(Pre_czndvi ^2)*Y				0.00014** <i>1.98889</i>	0.00003 <i>0.47721</i>
(Czndvi ^2)*Y					0.00011 <i>1.40318</i>
R-square	0.00900	0.01140	0.30810	0.04920	0.31760
Sum of Squared Errors	0.0183	0.0183	0.0151	0.0179	0.0151

Notes-The above table presents the results of the models estimated using available data.

The total number observations is 575

We report the tstat for each statistic below the value except where the values are aggregated figures

All these results are only for the LRLD season

***Significant at a confidence level of 0.99

4.3 Spatial Lag model

A spatial lag model (LeSage and Pace 2009)⁷ adds a lagged value of the dependent value on the right hand side of the equation. This particular model can be used when there is indication that value of ‘y’ from region ‘i’ are influenced not only by variables from its own region but also from regions neighboring ‘i’. In this case it can be hypothesized that livestock mortality is correlated across administrative divisions. Livestock mortality is a function of climatic and geographical factors which would be similar across areas close to each other. When dealing with semi-nomadic pastoralists it can also be assumed that pastoralists move across administrative boundaries in search of forage areas. Additionally not accounting for spatial relationships when they do exist will result in biased estimates. This is because the assumption of independence in the error terms will not hold valid.

$$y = \rho W y + X \beta + \varepsilon$$
$$Y = (I_n - \rho W)^{-1} (X \beta + \varepsilon)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

In the above stated spatial lag model, also referred to as the spatial autoregressive model (SAR), if the ρ value tends towards zero then the estimates are the same as that of a least squares model.

⁷ Our description of the spatial lag model and the maximum likelihood estimation procedures uses this reference heavily.

The value of ρ can be considered a measure of the strength of the spatial relationship between the dependent variables. A ρ value tending towards 1 indicates an extremely strong spatial relation between variables and vice versa.

The weight matrix defines the network relationship between the dependent variables. This is defined before the model estimation which implies that its formulation requires certain theoretical assumptions.⁸ The general consensus on the choice of weight matrices is that they should be chosen on a previously hypothesized model-based approach.

Weight matrices can be defined in many different ways from contiguity based matrices which give equal rank to all neighbors (defined as those with contiguous borders) to distance based matrices. There can also be more complex formulations of weight matrices which may be based on other geographical or non-geographical information. As in most situations when dealing with applied data there are no strong motivations to pick any one weight matrix. The initial estimation for this paper uses a first order queen contiguity matrix, which is quite standard. This matrix simply assigns a value of '1' to a neighbor (shares borders or vertices) and a '0' to every other division.

The value of $\rho W y$ referred to as the spatial lag in the model for our data is

$y_i = \rho \sum_{j=1}^{n=108} W_{ij} y_j + \varepsilon_i$. We can describe this value as a spatially weighted average. The weight is determined by the 'W' matrix. In order to form a linear combination of values we can normalize the W matrix so that the row sums equal to 1,

⁸ GeoDa is used to generate the Queen Contiguity Matrix.

$$\mathbf{w}_{ij}^i = \mathbf{w}_{ij} / \sum_j \mathbf{w}_{ij}$$

The standardization process creates a relative measure.

The interpretation of the model parameters is much more complex compared to the marginal impacts estimated from a least squares model. This is because each observation not only has an impact on itself but also on neighboring observations. The derivative of the i th y observation with respect the j th' X observation for the r th X explanatory variable is presented as follows.

$$\frac{\partial E(y_i)}{\partial X_{jr}} = S_r(\mathbf{W})_{ij}, S_r(\mathbf{W}) = (\mathbf{I}_n - \rho \mathbf{W})^{-1} \beta_r$$

where \mathbf{W} is a N -by- N matrix. The results that are estimated for β_r can be considered as summary measures of the marginal impact and are calculated as follows.

$$\begin{aligned} \beta_0 &= (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Y}) \\ \beta^\wedge &= \beta^0 - \rho (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{W} \mathbf{Y}) \end{aligned}$$

In order to account for the multiplier or feedback effect that we know exists in the spatial lag model, the β_r values are reported here. In effect the value that we report can be considered to be the expected value of the marginal impact created by the entire system.

$$\beta = \beta_{direct} (\mathbf{I} - \rho)^{-1}.$$

4.4 Maximum Likelihood Estimation

The presence of the weight matrix with a lagged variable on the right-hand side of the equation leads to biased and inconsistent Ordinary Least Squares estimates. This is why maximum likelihood estimation is used. The log likelihood function is

$$LnL = \frac{-n(\ln(2\pi))}{2} - \frac{n(\ln(\sigma^2))}{2} - \frac{(y - \rho Wy - X\beta)'(y - \rho Wy - X\beta)}{2\sigma^2} + \ln |I - \rho W|.$$

The actual estimation process is carried out using a concentrated log-likelihood function which is estimated as,

$$\begin{aligned} \ln(L(\rho)) &= \frac{n(\ln(2\pi))}{2} + \ln |I - \rho W| - n/2 \ln(S(\rho)) \\ S(\rho) &= e(\rho)'(\rho) = e_0'e_0 - 2\rho e_0'e_d + \rho^2 e_0'e_d \\ e(\rho) &= e_0 - \rho e_d, e_0 = y - Z\delta_0 \\ e_0 &= y - Z\beta_0 \\ e_d &= Wy - Z\delta_d \\ \delta_0 &= (Z'Z)^{-1}Z'y \\ \delta_d &= (Z'Z)^{-1}Z'Wy \\ Z &= [\tau_n X], \tau_n = n \times 1 \end{aligned}$$

Spatial Panel Lag Model

In the spatial panel model we run a pooled model by stacking the X and Y observations by time,

$$Y_{TN \times 1} = \rho(I_T \otimes W)Y + X_{TN \times K} + \varepsilon.$$

We use the fixed coefficients model in order to allow the intercepts and slopes of various spatial units vary.

4.5 Spatial Model Steps

This section of the paper covers the processes that we ran in order to deal with the challenge of missing data and to estimate response functions describing the relationship between NDVI and livestock mortality.

The first step in the process is to run an Ordinary Least Squares Regression (ols) using the data points that are available.⁹ An intercept matrix allows for each division to have its own intercept allowing for more flexibility in the regression.

The initial model used at this stage is as follows,

$$X_{cndvi} = [\text{previous}, \text{previous}^2, \text{present}, \text{present}^2, \text{previous} * X, \text{previous}^2 * X, * \text{present} * X, \text{present}^2 * X, \text{previous} * Y, \text{previous}^2 * Y, \text{present} * Y, \text{present}^2 * Y.]$$

In the next step of the process the spatial panel lag model is used in order to utilize the spatial information that this paper theorizes exists within the data. The first iteration uses the fitted values from the earlier OLS regression in place of the missing data points.¹⁰

However, each iteration continues to utilize the actual data for the points for which data is available, whereas the missing values are replaced by their former fitted value,

$$Y_{\text{Values for missing data in iteration } i} = Y_{\text{fitted values from iteration } i-1}.$$

A total of 30 iterations of this process are conducted. The results from this iterative process allows us to gain an average values for the parameter estimates and the spatial dependence estimates. The next step involves the use of a boot strapping process to obtain an estimate of the standard errors for each of the estimated values.

⁹ At this point we have incomplete data for 85 divisions.

¹⁰ We discuss the process of estimation of the fitted values in the earlier section.

Chapter 5-Results

An initial comparison is conducted between results for specific spatial and non-spatial models. For the spatial models with iterations the results presented in the table are the expected values from the output of the various iterations. The results of the response function that are presented have also been averaged across districts or divisions depending on the model that we use.¹¹

This paper will present the results of the two seasons SRSD (Short Rain Short Dry) and LRLD (Long Rain Long Dry) separately, since as discussed earlier it is assumed that two separate response functions exist for the seasons.

A key point to be noted regarding the results is that for the models where an in iterative process is used for estimation, it is hard to determine appropriate Goodness-of-Fit measures. The results indicate as expected an upward bias in the standard Goodness-of-Fit measures such as R-squared and Log-Likelihood.

In order to acquire a clearer picture of the pattern of these response functions in terms of determining livestock mortality, this paper provides graphical comparisons of the response functions across different levels of aggregation for a few chosen districts.

5.1 Response Functions

The response functions plot out the marginal impact of the *czndvi* values for the present season while holding all the other variables constant.

¹¹ Response functions of specific districts are presented in the appendix.

There are a few different ways in which the marginal impacts of the *czndvi* variables can be plotted out. The simplest method would be simply to plot out the $X * \beta$ values. The advantage of this model is that it would be simple to plot and not require making assumptions about the *czndvi* values (for present season and previous season) of other divisions when plotting out the response functions for the spatial lag models. However ignoring the spatial feedback effects is likely to lead to biased predicted values.

Another option would be to use the historical *czndvi* values for the present season (previous season) and take an average across all the *czndvi* values for the previous season (present season) for all the divisions. This method would allow for the inclusion of the effects of the spatial filter, which is a key part of the spatial lag model.

$$\Delta Y = (I - \rho W)^{-1} (X \beta)$$

Yet another method could be to draw *czndvi* values for the present season (previous season) for the other divisions from the pool of historical data by giving each value an equal probability of being selected. The same values are retained for the divisions of interest. This allows for a closer examination of the effect of our assumptions about the *czndvi* values for the present season (previous season) for other divisions. To contain the difference in values we draw the 108 values for the different divisions from a smaller pool of *czndvi* values for each of the 11 years. This insures that the values across divisions are similar and also similar to the values of the district for which we are plotting the response functions.

A quick examination of each of the methods involves plotting response functions for each of the models. Plotting out various response functions indicates that the three response functions are behaving similarly.

For the presentation of results of spatial models we use the second method that we discussed, which we refer to as the historical data method.

5.2 Model Permutations

Spatial Aggregation

As discussed earlier, the key outcome from this process is the design of a response function that will link remotely sensed NDVI data with livestock mortality. Ideally the response function would capture most of the weather related livestock mortality. One of the most serious questions that this design process poses is the level of aggregation to use when estimating the response function.

Described by Anselin (1988) as the Modifiable Area Unit Problem, it is clear that estimates are sensitive to the level of aggregation chosen for the parameterization. Heterogeneity in the underlying spatial processes of a region classified as one spatial unit can also lead to biased estimates.

There are many potential variations that we could use when estimating the response functions. They could range from treating all of the 11 districts as one, while alternatively giving each sub-location a separate response function. If we chose an extremely restrictive model, for example, that of imposing one response function in all of the 11 districts, the parameters are likely to be biased because we are ignoring the underlying heterogeneity in the system. However, choosing

the least restrictive option (that is with each sub-location receiving its own response function), is likely to lead to over-fitting. There is an underlying trade-off in this scenario between over-fitting and inefficiency. We initially examine this trade-off by looking at the behavior of the response functions from different models.

This paper will primarily discuss variations of district based models and the division based model. The various levels of aggregation are achieved by the assignment of dummy variables to each administrative level over which we wish to aggregate.

1) Country Level Model.

At this stage one model is run for the entire 11 districts. This would mean theoretically that there is one response function for the entire region.

2) District Level Model.

For this particular model response functions are aggregated across districts

3) District Level Intercepts with Division level coefficients.

This is a variation of the above model but which allows more flexibility in the variation of the response function.

4) Division Level Intercepts with District level coefficients

5) Division Level Model

This is a model with division level intercepts and division level coefficients.

5.3 Non-Spatial Models

Long Rain Long Dry –LRLDT

The following table presents the different results from non-spatial model that was discussed in the earlier section. After the estimation of the missing values response functions are estimated at the level of aggregation indicated using dummy variables. The following parameters are estimated post the imputation of the missing values. OLS is employed for all models. The models are otherwise the same except for varying levels of aggregation in the parameterization.

Table 5.1-Model results of aggregation in non-spatial models

	Country Level Model	District Level Model	Division Level Intercepts with District level coefficients	District Level Intercepts with Division level coefficients	Division Level Model
Intercept	0.09723*** <i>16.81556</i>	0.09551	0.0973	0.09103	0.09669
Pre_czndvi	-0.00156* <i>-1.64308</i>	-0.00137	-0.00004	-0.00238	- 0.00184
Czndvi	0.00052*** <i>5.06326</i>	0.00028	0.00031	0.00058	0.00045
Pre_czndvi ^2	-0.00732*** <i>-10.36153</i>	-0.00716	0.00046	-0.00576	- 0.00618
Czndvi ^2	0.00041*** <i>5.60114</i>	-0.00716	0.00158	0.00034	0.00043
R-squared	0.1413	0.2903	0.5861	0.6782	0.7686

Notes-Tstats are only provided for the first model because the others have been aggregated across districts/divisions.

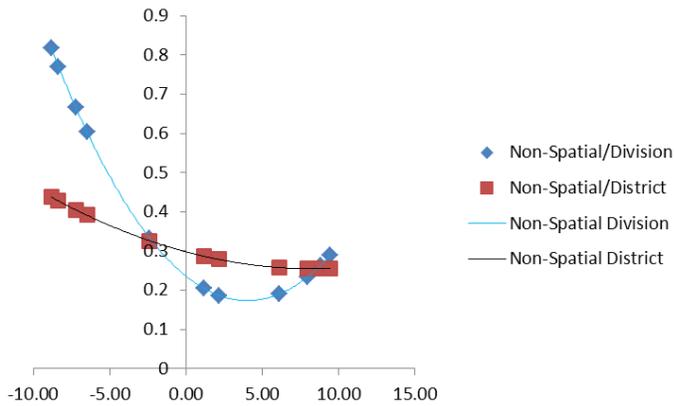
***Significant at a confidence level of 0.99

** Significant at a confidence level of 0.95

According to the following graph (Figure 5.1 the graph for the Bura division), the division level response function is predicting a significantly higher mortality than the response function

estimated at the district level. The district level model parameters are estimated using all of the available data points for that district.

Figure 5.1 –Division versus District Level Response Functions for non-spatial LRLD model



Notes-The graph includes only data from the LRLD season of the Bura Division in Tana River District . This response function is plotted holding $czndvi$ for the previous season constant at a value of -0.79 and then plotting the marginal impact of $czndvi$ values for the present season from a range of -17 to +20.

Short Rain Short Dry-SRSD

A similar exercise as conducted with the long rain long dry season data is repeated with short rain short dry season. For the SRSD season, the *czndvi* values of the previous season are not significant.

Table 5.2- Results of aggregation in non-spatial models

	Country Level Model	District Level Model	Division Level Intercepts with District level coefficients	District Level Intercepts with Division level coefficients	Division Level Model
Intercept	0.096169*** <i>19.368764</i>	0.089612	0.095277	0.090974	0.098406
Pre_czndvi	0.000249 <i>0.4584</i>	0.000215	0.000143	-0.001377	-0.000986
Pre_czndvi ^2	0.000045 <i>0.729572</i>	0.000068	0.000051	0.000077	0.000018
Czndvi	-0.002867*** <i>-4.28047</i>	-0.002461	-0.002384	-0.001896	-0.002768
Czndvi ^2	0.000099 <i>1.219383</i>	0.000105	0.000111	0.000145	-0.000079
R-squared	0.0176	0.1535	0.4473	0.5479	0.6574

Tstats are only provided for the first model because the others have been aggregated across districts/divisions.

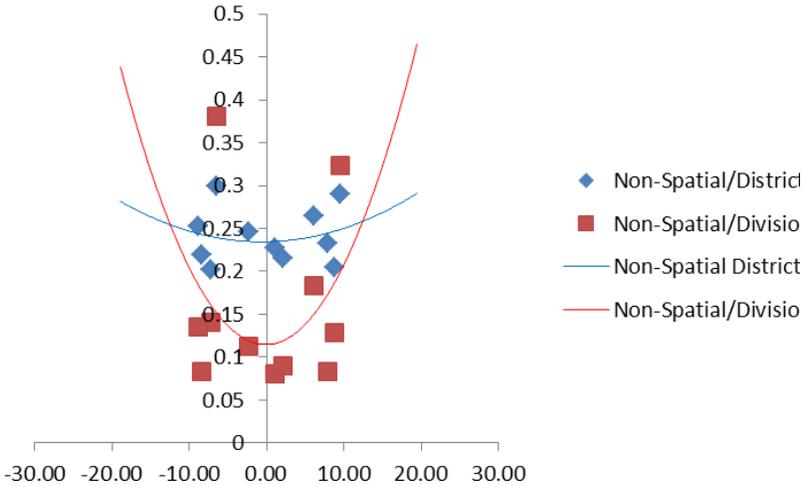
***Significant at a confidence level of 0.99

** Significant at a confidence level of 0.95

The above table plots out the response functions of non-spatial models at different levels of aggregation. The graph includes only data from the SRSD season. This response function is plotted by hold *czndvi* for the previous season constant at a value of -0.79 and then plotting the marginal impact of *czndvi* values for the present season from a range of -17 to +20. We plot out the response function for District 1 which is the Tana River District. We aggregate the coefficients and intercept values to achieve a district level function

The following graph provides the division level and district level response functions for a specific division in the above district.

Figure 5.2 – Division versus District for non-spatial SRSD model



Notes-The graph includes only data from the LRLD season for the Bura Division in the Tana River District. This response function is plotted by hold *czndvi* for the previous season constant at a value of -0.79 and then plotting the marginal impact of *czndvi* values for the present season from a range of -17 to +20.

We observe similar trends in terms of variation with aggregation with the LRLD season. The division level function is similar to the function with division level intercepts. From this graph it would appear that other division in this district should have lower mortalities, which is why the district level function appears to be indicating no rise in mortality across the *czndvi* values.

The OLS estimation does not incorporate any spatial relationships. The next section examines the effect of aggregation when using spatial models which systematically incorporate spatial feedback effects between divisions. The next section continues onto compare spatial and non-spatial models.

5.4 Spatial Models

Comparing Levels of Aggregation

The following section examines spatial panel models run at different level of aggregation. The hypothesis behind using the spatial model is that when concerned with livestock mortality and its correlation with vegetation as determined by NDVI, it is expected that spatial factors will come into play. We first present results for the Long Rain Long Dry Season and then results for the Short Rain Short Dry season.

It is important to note that even at higher levels of aggregation it is assumed that the pricing will be at the division level.

Long Rain Long Dry (LRLD)

The following provides an aggregated comparison of the parameter values from the different models.

Table 5.3-Comparing aggregation in spatial models-LRLD

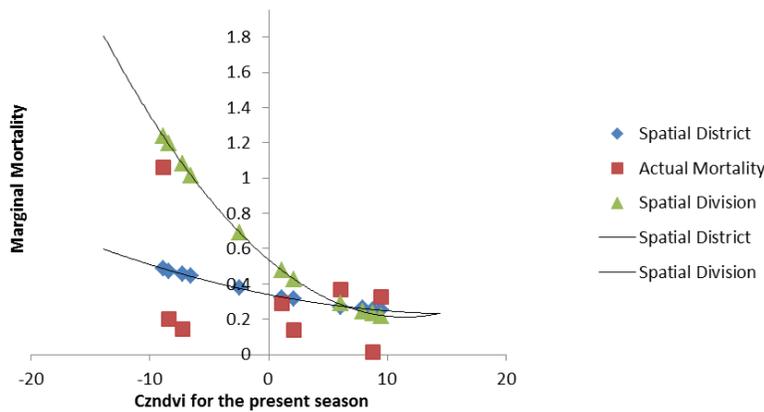
Parameter	Country Level Model	District	District coefficient	Division Coefficient	Division Level Model
Intercept	0.125026992	0.109970407	0.275952603	0.092072393	0.150305366
Pre_czndvi	-0.00162168	-	-	-0.00689288	-0.006480447
Pre_czndvi ^2	0.000456949	0.000153435	9.96E-05	0.001285448	0.000381509
Czndvi	-0.006185291	-	-	-0.002278549	-0.001162735
Czndvi ^2	0.000301533	0.001327045	0.001390544	0.001573041	0.001198349
ρ	0.230051757	0.032098253	0.127939421	0.159843541	0.285931776
ρ (std deviation)	0.022581105	0.025510395	0.026470754	0.033315429	0.032213253

Notes-The table above provides the aggregated values of the coefficients across iterations and across districts/divisions. We divide the aggregated values with(1- ρ) to account for the spatial feedback effects

The graph gives us a quick look into the behavior of the response at the different levels of aggregation. The graph indicates that there is variation between the functions especially at the extremes.

The following graph plots out the response function for the District Level Model and the Division Level Model Response Function. It also plot outs the predicted values from each model and also the actual mortality values available for that district.

Figure 5.3-Examining District and Division Level Model Results-LRLD



Notes-The graph plots out the response functions for the Bura Division in the Tana River District under the District Level and the Division Level Model. It also presents the actual data available for the district and the predicted mortalities from the models. The graph includes only data from the LRLD season. This response function is plotted by hold *czndvi* for the previous season constant at a value of -0.79 and then plotting the marginal impact of *czndvi* values for the present season from a range of -17 to +20.

Short Rain Short Dry (SRSD)

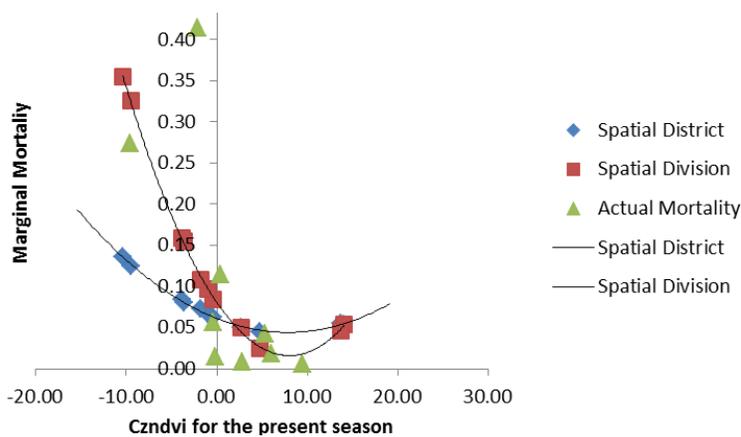
Table 5.4- Spatial Models Results-LRLD

Parameter	Country Level Model	District	District coefficient	Division Coefficient	Division Level Model
Intercept	0.110547199	0.099960902	0.199515687	0.070125392	0.157585042
Pre_czndvi	0.001281738	0.001079061	0.000946808	-0.004452167	-0.003442711
Pre_czndvi ^2	1.48E-05	7.94E-05	5.44E-05	0.001277146	0.000563374
Czndvi	-0.00346589	-0.002459827	-0.002721948	-0.000583377	-0.001280721
Czndvi ^2	0.000164817	1.40E-04	0.000192752	0.001100359	0.000165638
ρ	0.372690684	0.316338466	0.41146067	0.279786991	0.331941856
ρ (std deviation)	0.021794046	0.021950127	0.0215688	0.026615684	0.028735669

Notes -The table above provides the aggregated values of the coefficients across iterations and across districts/divisions. We divide the aggregated values with(1- ρ) to account for the spatial feedback effects

A comparison of the two different seasonal models indicates that the ρ values for this season are slightly higher than the ρ values for the LRLD season. This could be due to the fact that the SRSD is over a shorter span of time which means less temporal heterogeneity in the data. The next graph takes a closer look at the district level and division level response function for the same division. It is also presenting the predicted values from each of the models

Figure 5.4-Examining District Level and Division Model Results-SRSD



Notes-The graph plots out the response function for the entire district using the District Level Model. It also presents the actual data available for the division. The graph includes only data from the SRSD season. This responsefunction

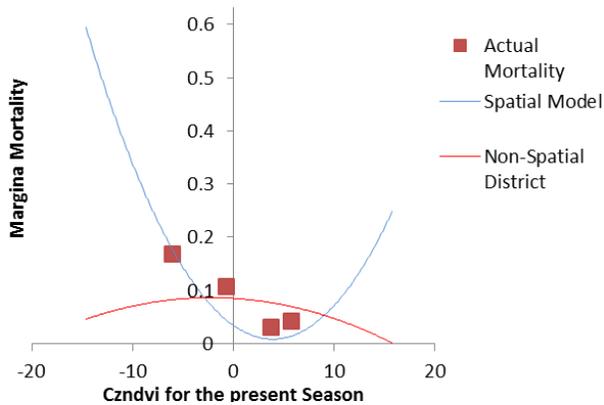
is plotted by hold czndvi for the previous season constant at a value of -0.79 and then plotting the marginal impact of czndvi values for the present season from a range of -17 to +20.

The division level response function veers upward at the positive extreme. The predicted values appear to be similar for each of the models except for when the ndvi values are extremely negative. Here the division level model is giving a much higher level of mortality than the district level model. It appears that division level response functions are more sensitive to outliers.

5.5 Comparing non-spatial and spatial models

This following section compares district level model of the spatial model and the non-spatial OLS estimation.

Figure 5.5-Comparing Spatial Model and Non-Spatial Model Results-LRLD



Notes-The graph includes only data from the LRLD for the Buna Division in Wajir District and it uses the district level model. It plots out the response using the spatial model and the non-spatial model. It also shows the actual mortality values for this division. This response function is plotted by hold czndvi for the previous season constant at a value of -0.79 and then plotting the marginal impact of czndvi values for the present season from a range of -17 to +20.

The non-spatial OLS estimation for this division shows a lower marginal impact when czndvi values are extremely negative and positive.

In the next section out-of-sample likelihood methods are used to examine the different models and the levels of aggregation systematically in order to arrive at the model which provides most optimum response function.

5.6 Using out of sample log likelihood

In the earlier section the paper examined the performance of the various models on the basis of the behavior of the response function. However while response functions allow us to check if the index is behaving properly this section takes a closer look at the models performance in out of sample estimation since this is a key purpose of this model.

The model is estimated while withholding an observation (or set of observations) and then evaluating the model on the observations which were not included in the estimation process.

$$L_k^{out} = \prod_{i=1}^N \int_{k,i} (y_i, \theta_k^*(y_{-i}))$$

The value of $\int_{k,i} (y_i, \theta_k^*(y_{-i}))$ denotes the evaluation of model k at the out of sample observations.

This paper uses a cross-validation technique which involves dropping the observations from year 'i' and then estimating the observations with the remaining years. The model is then evaluated at the observations of the remaining years.

This paper adapts a criterion which chooses a model 'k' which has greater out of sample likelihood when compared to other models. It employs a mixing weight methodology that

assigns weights to models that optimizes the out of sample likelihood (Woodard and Sherrick, 2011) where weights are defined as follows,

$$W = \{w_1, w_2, \dots, w_n \mid \sum_{i=1}^n w_i = 1, w_i \geq 0 \forall i\}.$$

Two models are compared at a time assigning weights to each of the models that maximizes the out-of-sample likelihood. The weights are basically optimized based on the values that minimize the difference between the predicted values (calculated from parameters using separate values) and the initial livestock mortality values (which includes prior values and the actual mortality values), as follows,

$$w^* = \arg \max(L_{out}^{Mix}(W, Y)).$$

Six different models are compared

- 1) Spatial District Level Model
- 2) Spatial District Level Coefficient Model
- 3) Spatial Division Level Coefficient Model
- 4) Spatial Division Level Response Function
- 5) Non Spatial Division Level Response Function

When the district level model is compared to the spatial division level model it gets assigned a weight of 0.68 and a weight of 0.73 when compared to the non-spatial division level model.

However comparing the district level model to other variations, the spatial model with district level coefficients and division level intercepts outperforms the district level model.

The information from this model is used as indication that the model with the district level coefficients and division level intercepts performs the best out-of-sample.

5.7 Weight Matrices

This section examines the sensitivity of the response functions to the different weight matrices.

The previous models used are formulated using the Queen Contiguity Matrix. This section of the paper will examine the sensitivity of the response functions to the weight matrix chosen.

This paper hypothesizes that distance based matrices will distort the ρ values. This is because specific divisions have very different areas. For example the Lasamis Division in the Marasbit District has an area of 8673 square km while the Central Mandera division in the Mandera district has an area of 99 square km. Since the distance is measured between polygon centroids distance based matrices would place more weight on neighbors of smaller divisions, and may have little theoretical basis.

Queen Matrix

This matrix is simply one that assigns a '1' value to a neighboring division with which it shares a border and a '0' value otherwise. An example [3,3] matrix could look like the matrix shown below.

$$w = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Inverse Distance Weighting Matrix

This is a weighting¹² matrix which is based on the intuition that the impact of the weighted variable decreases with increasing distance. For this matrix a threshold distance of 150 km is used. The threshold distance simply serves as the cutoff distance for defining neighbors. The distance is measured between the centroids of each division.

$$w = \begin{matrix} 0 & 1/(d_{(1,2)})^2 & 1/(d_{(1,3)})^2 \\ 1/(d_{(2,1)})^2 & 0 & 0 \\ 1/(d_{(3,1)})^2 & 1/(d_{(3,2)})^2 & 0 \end{matrix}$$

Inverse Distance Weighting Matrix –No threshold distance

The response function using another variation of the inverse distance weighting matrix is also examined . For this variation no threshold distance is specified. This means that each division will have 107 neighbors. The exponent is increased to ‘3’ in order to create a more sharply decaying function

$$w = \begin{matrix} 0 & 1/(d_{(1,2)})^3 & 1/(d_{(1,3)})^3 \\ 1/(d_{(2,1)})^3 & 0 & 1/(d_{(2,3)})^3 \\ 1/(d_{(3,1)})^2 & 1/(d_{(3,2)})^3 & 0 \end{matrix}$$

In order to examine the sensitivity of the response functions¹³ to the weight matrices used the following table will present the results comparing the different weight matrices using the district coefficient model for the LRLD season.

¹² These matrices were generated using ArcMap

¹³ Figure in Appendix

Table 5.5-Comparing different weight matrices-LRLD

Parameter	Queen Contiguity	Inverse Distance 1	Inverse Distance 2
Intercept	0.275952603	0.176	0.182
Pre_czndvi	-0.005411741	-0.0005	-0.0005
Pre_czndvi ^2	9.96E-05	0.0002	0.0002
Czndvi	-0.001718496	-0.0072	-0.0074
Czndvi ^2	0.001390544	0.0004	0.0004
ρ	0.127939421	0.06007	0.1099
ρ (std deviation over iterations)	0.026470754	0.03628	0.001

Notes-This table provides comparison between using different weight matrices. We use the LRLD model which uses division level intercepts and district level coefficients. The beta values have been aggregated across iterations and intercepts. We divide the aggregated values with $(1-\rho)$ to account for the spatial feedback effects

The distance specified weight matrices appear to have much lower ρ values when compared to the contiguity based one. As hypothesized the distance based matrices are sensitive to the areas of the divisions which are quite varied. A closer examination of the data also reveals that for the first version of the inverse distance weighting matrix, which specifies a threshold distance of 150 km, translates into some larger divisions such as North Horr and Laisamis not having all of their neighbors being included in the weight matrix. The plotting out of each of the response function for a chosen district however indicates that the function is not sensitive to the formulation of the weight matrices. The response functions are similar for the three different weight matrices.

Chapter 6-Tropical Livestock Units

The initial pilot for IBLI in Marasbit in Kenya and Borozana district in Ethiopia utilized the following equivalency for the design of the contract (Chantararat, 2012)

TLU = Camels 1.0; Cattle 0.7; Sheep/Goats: 0.1.

While the use of TLUs might simplify the process of contract design it might not be a truly valid measure in the context of livestock mortality in arid and semi-arid regions. The TLU system is deemed to be an effective unit in the scenario when each animal is being given the same feed. However different animals have different grazing patterns and behaviors which make comparison difficult in this context. For example sheep and goats would survive better when feed density is lower when compared to cattle. Camels might be able to reach higher ground for grazing while goats might be able to reach between rocks which would be inaccessible for camels.

A study based in the Turkana district in Kenya examining the impact of the droughts and extreme rainfall that affected the region from 1980-1985 found different patterns in livestock mortality between different animals. Their sample herds saw the loss of approximately 47% of their cattle population 30% of their goat and sheep population and 26% of their camel population due to disease and starvation. While these patterns might vary across regions and for different weather conditions, we hypothesize that cattle mortality would also be higher in bad weather when compared to other animals in other arid districts in Kenya (McCabe, 1987).

These differences in patterns raise the question of whether the TLU equivalency method works within the context of livestock mortality for animals whose feed is predominantly from foraging.

This section examines this question by comparing response functions generated using the TLU equivalency versus animal specific contracts.

The following table presents the response functions generated using the District Coefficient model (division intercepts and district level coefficients) using TLU and animal specific data.

Table 6.1-Examining Animal Specific Functions-LRLD

Parameter	TLU/Animal Specific Contracts				
	TLU	Sheep	Goat	Camel	Cattle
Intercept	0.109970407	0.1819316	0.142993821	0.106304814	0.11765789
Pre_czndvi	-0.005075527	0.0035985	0.001844281	-0.00493365	-0.003581
Pre_czndvi ^2	0.000153435	0.0003088	0.000205122	0.000150791	4.36E-05
Czndvi	-0.001702257	-0.0087034	-0.00463554	-0.00165734	-0.0063745
Czndvi ^2	0.001327045	0.0004479	0.000152553	0.001285747	0.00048634
ρ	0.032098253	-0.042638	-0.00316384	0.032606176	-0.000466708
ρ (std deviation)	0.025510395	0.0279846	0.029253962	0.025969095	0.029734159

Notes-This table provides comparison between using mortality in Tropical Livestock Units versus animal specific contracts. We use the LRLD model which uses division level intercepts and district level coefficients. The beta values have been aggregated across iterations and intercepts. We divide the aggregated values with $(1-\rho)$ to account for the spatial feedback effects.

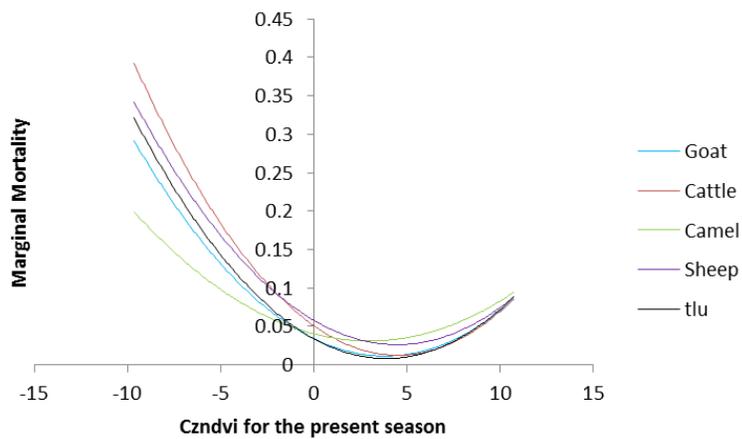
A closer examination spread of livestock indicates that different types of livestock tend to be concentrated in different areas. This would imply that if neighboring districts possessed different types of livestock then the spatial pattern in term of livestock mortality would be weaker.¹⁴ The negative ρ value for the camel model is attributable to the sparse camel data available to us from our dataset.

Plotting out a response function for a chosen district highlights the problems that using a TLU equivalency might create. Previous research and evidence indicates that cattle are more

¹⁴ Data available in Table 5.

susceptible to poorer weather. This is reflected in the response function where predicted mortality for cattle is higher in poorer weather (czndvi values tending towards -20). Predicted mortality for camels on the other hand is much lower when compared to all of the animals. This difference implies that contracts using TLU equivalency might result in higher basis risk when compared to animal specific contracts. However as the two alternate graph below shows us it is important to note that this relation does not stand true for each of the divisions. The decision on whether to use TLU as opposed to animal specific contracts needs to take into account the animal density in each district.

Figure 6.1- Animal Specific Contract versus TLU



Notes-This graph provides comparison between using mortality in Tropical Livestock Units versus animal specific contracts for the Buna Division in the Wajir District. We use only LRLD data for these response functions for different animals. This response function is plotted by hold czndvi for the previous season constant at a value of -0.79 and then plotting the marginal impact of czndvi values for the present season from a range of -17 to +20

Chapter 7- Verifying the model through simulation

In order to test that the imputation model provides unbiased estimates, the next section investigates its performance in a simulated framework. The parameters used are similar to that of the system level model.

Table 7.1-Setting Parameters for simulation-LRLD

Parameters	Country Level Model
Intercept	0.12503
Pre_czndvi	-0.00162
Pre_czndvi ^2	0.00046
Czndvi	-0.00619
Czndvi ^2	0.00030
ρ	0.23005
ρ (std deviation)	0.02258
Sum of Squared Residuals	0.02193

Notes-These are the parameter values which constitute our “true” model for the simulations. The spatial country level model was employed.

The predicted values for mortality are generated using the parameters from this model. In the next step Gaussian residuals are simulated with a variance equivalent to the sum of squared errors in the spatial model given above.

$$y = (I - \rho w)^{-1} (X * \beta + \varepsilon)$$

$$\varepsilon \sim N(0, 0.02193)$$

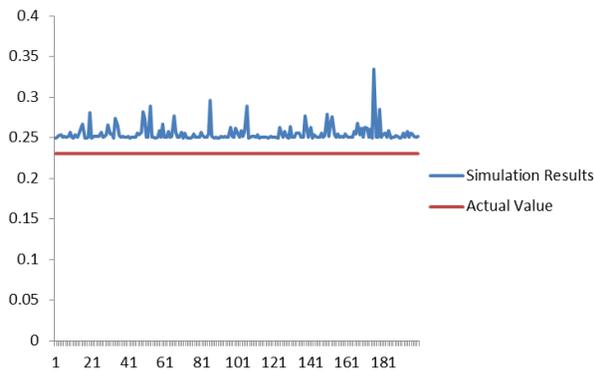
Since this section is examining the effectiveness of the model discussed in this paper in estimating unbiased missing values up to 40% of the data is randomly dropped since this constitutes approximately the missing data in our LRLD season when working Tropical

Livestock Units. These randomly dropped values are replaced with the mean of the available values. This step is to test the assumption that the chosen prior does not have any impact on the final imputed values.

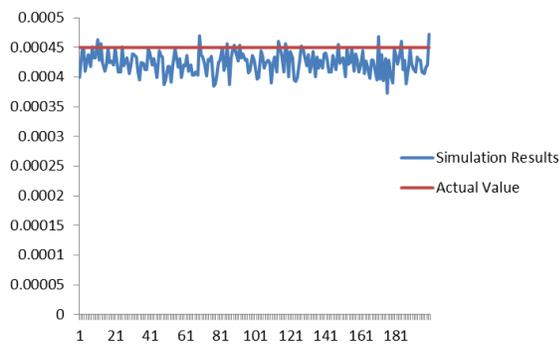
The next step is to run through a number of iterations of the imputation process. This process is repeated a number of times and it allows for the examination of the behavior of the parameters in comparison to the true value across iterations.

Figure 7.1 Plotting out parameters across simulations

ρ (Rho)Parameter



Parameter for squared *czndvi* value for previous season



Notes –The graphs above plot out the mean of the simulated parameter for the 200 different models across the 30 iterations of each of those 200 models

The graphs above allow for the examination of the performance of the parameters by giving a sense of how they are moving across the iterations. The graphs plot out the mean of the simulated parameter value for the 200 different models and plotted across the 30 different iterations of each of these models. It can be seen that there is only a small amount of bias in the ρ value, which indicates that the model is behaving as expected.

Chapter 8-Final Contract Design

The final model that is determined as providing the most efficient and least biased response functions after analyzing the models through the plotting of response functions and using the out of sample likelihood maximization method is the district level spatial model. This model is then utilized to calculate the losses that a contract designed using this response function would produce.

The contract design is based on the design of a European call option.

$$\text{Indemnity}_{t,s,d} = \text{Max}[M_{t,s,d}(\text{NDVI}) - (\bar{M} + \text{StrikeRate}), 0]$$
$$\text{Premium} = \int \text{Max}[(M_{t,s,d}(\text{NDVI}) - (\bar{M} + \text{Strike})), 0]^* \rho * \text{TLU}$$

$$f(V) = \text{pdf of NDVI Variables}$$

The contract basically works as follows. The response function that provides a predicted mortality figure based on the NDVI values for that particular season and the prior season. This value will then be subtracted from the average predicted mortality across the previous years plus the strike rate that is decided on. If the predicted mortality is higher than the trigger, then the contract will have a payout. Two different contracts will be sold across the year for the two different seasons.

To calculate the losses, initially the mean for each division is calculated for each of the simulations. Then the division mean along with the chosen strike rate, which for the following results is 0.1, is subtracted from each of the predicted mortalities. The losses that appear in each of the simulations are averaged and then aggregated across divisions and then across districts.

The table below provides a snapshot of the losses in the SRSD and the LRLD season when using the spatial model at the district level of aggregation.

Table 9.1-Percent Losses per District-

Loss Cost Percent by District		
District	LRLD	SRSD
TANA RIVER	1.916178	1.610003
BARINGO	0.9573205	0.6263798
GARISSA	2.1518871	1.3965007
IJARA	1.7660344	2.5285311
ISIOLO	2.3874054	1.2129465
MANDERA	2.3152045	1.969394
MARASBIT	2.4669913	3.3576058
MOYALE	3.5776775	2.5312362
SAMBURU	5.0878179	3.0638026
TURKANA	2.6059465	1.2890637
WAJIR	4.8033431	1.0255208

Notes- The table provides the losses for the LRLD and SRSD calculated using a district level response function.

Chapter 9-Conclusion

The key objective of this research was to estimate a response function which mapped out remotely sensed data onto livestock mortality to design an insurance contract. The foremost challenge in meeting this objective was the missing mortality data.

A spatial model is employed to meet this objective. The rationale behind using a spatial model is that animals owned by pastoralists move across administrative borders in search of pasture and water, especially during times of poorer weather. Several models are examined, and the results indicate the model with the best out-of-sample performance was that which included district level coefficients and division level intercepts.

A secondary objective was also to examine the use of Tropical Livestock Units for estimation of the model and investigate animal-type specific contracts. We examine the response functions of animal specific contracts using the same techniques and determine that different animals have different responses to poor weather. This indicates that there may be some motivation to estimate separate contracts for specific animals, as opposed to an aggregated TLU measure.

Bibliography

Anselin, L. (1988). Spatial Econometrics: Methods and Models, Kluwer Academic Publishers.

Barnett, B. J., C. B. Barrett, et al. (2008). "Poverty Traps and Index-Based Risk Transfer Products." World Development **36**(10): 1766-1785.

Barret, B. C. and G. A. Mude (2012). Index-Based Livestock Insurance in Northern Kenya: An Analysis of the Patterns and Determinants of Purchase. 10th International Conference on the Ethiopian Economy, Addis Ababa.

Carroll, M., C. DiMiceli, (2013) MODIS Data Compositing: Lessons Learned at the SCF. College Park, Department of Geography, University of Maryland.

Casetti, E. (1972). "Generating Models by the Expansion Method: Applications to Geographical Research*." Geographical analysis **4**(1): 81-91.

Ceccato, P., M. Brown, (2013) Workshop: Technical Issues in Index Insurance Topic 8: Remote Sensing - Vegetation.

Chantarat, S., G. A. Mude, et al. Willingness to pay for Index Based Livestock Insurance; Results from a field experiment in Northern Kenya.

Chantarat, S. M., Andrew G. ; Barrett, Christopher B. ; Carter, Michael R. (2012). "Designing Index-Based Livestock Insurance for Managing Asset Risk in Northern Kenya." Journal of Risk and Insurance: no-no.

FAO. "Tropical Livestock Units (TLU)." from

<http://www.fao.org/ag/againfo/programmes/en/lead/toolbox/Mixed1/TLU.htm>.

Fichman, M. and J. N. Cummings (2003). "Multiple imputation for missing data: Making the most of what you know." Organizational Research Methods **6**(3): 282-308.

Giné, X., R. Townsend, et al. (2007). "Statistical Analysis of Rainfall Insurance Payouts in Southern India." American Journal of Agricultural Economics **89**(5): 1248-1254.

GoK (2005). Arid and Semi Arid Lands (ASAL) :National Vision and Strategy, PricewaterhouseCoopers, ASAL National Vision and Strategy.

GoK (2005). Kenya Integrated Household Budget Survey. K. N. B. o. Statistics. Kenya, Kenyan Government.

Hatfield, R. D., Jonathan (2006). Global Review of the Economics of Pastoralism, World Initiative for Sustainable Pastoralism.

Jerry R. Skees, B. J. B. (2006). "Enhancing microfinance using index-based risk-transfer products." Agricultural Finance Review **66**(2): 235-250.

LeSage, J. and R. K. Pace (2009). Introduction to Spatial Econometrics, CRC Press.

LeSage, J. P. and R. K. Pace (2004). "Models for spatially dependent missing data." The Journal of Real Estate Finance and Economics **29**(2): 233-254.

Mahul, O., N. Belete, et al. (2009). Index-based Livestock Insurance in Mongolia, International Food Policy Research Institute (IFPRI).

Mahul, O. and N. Verma (2012). "Improving farmers' access to agricultural insurance in India." World Bank Policy Research Working Paper(5987).

Makaudze, E. M. and M. J. Miranda (2010). "Catastrophic drought insurance based on the remotely sensed normalised difference vegetation index for smallholder farmers in Zimbabwe." Agrekon **49**(4): 418-432.

McCabe, J. T. (1987). "Drought and Recovery: Livestock Dynamics among the Ngisonyoka Turkana of Kenya." Human Ecology **15**(4): 371-389.

Miranda, M. J. and J. W. Glauber (1997). "Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets." American Journal of Agricultural Economics **79**(1): 206-215.

NASA (2011). MODIS 16-Day Composite Imagery. College Park, Maryland, Global Land Cover Facility, University of Maryland.

Schafer, J. L. (1999). "Multiple imputation: a primer." Statistical methods in medical research **8**(1): 3-15.

Smith, V. and M. Watts (2009). Index Based Agricultural Insurance in Developing Countries: Feasibility, Scalability and Sustainability.

Turvey, C. G. and M. K. McLaurin (2012). "Applicability of the Normalized Difference Vegetation Index (NDVI) in Index-Based Crop Insurance Design." Weather, Climate, and Society **4**(4): 271-284.

Wikipedia (2013). "Divisions of Kenya." from http://en.wikipedia.org/wiki/Divisions_of_Kenya.

Woodard, J. D. and B. J. Sherrick (2011). "Estimation of Mixture Models using Cross-Validation Optimization: Implications for Crop Yield Distribution Modeling." American Journal of Agricultural Economics **93**(4): 968-982.

Woodard, J.D., G.D. Schnitkey, B.J. Sherrick, N. Lozano-Gracia, and L. Anselin, "A Spatial Econometric Analysis of Loss Experience in the U.S. Crop Insurance Program," *Journal of Risk and Insurance* (2012).

Woodard, J.D., and P. Garcia "Basis Risk and Weather Hedging Effectiveness," *Agricultural Finance Review* (spring, 2008).

Woodard, J.D., and P. Garcia, "Weather Derivatives, Spatial Aggregation, and Systemic Insurance Risk: Implications for Reinsurance Hedging," *Journal of Agricultural and Resource Economics* (April, 2008).

"Woodard, J.D., "A Spatial Econometric Approach to Developing Scalable Livestock-NDVI Index Insurance with Missing Data," International Agricultural Risk, Finance, and Insurance Conference (IARFIC), Central University of Finance and Economics, "Agricultural Risk, Finance, Insurance and Actuarial Science", Beijing, China (June 14-16, 2013).

Appendix

Additional Figures

Figure 9.1-Livestock Production Systems

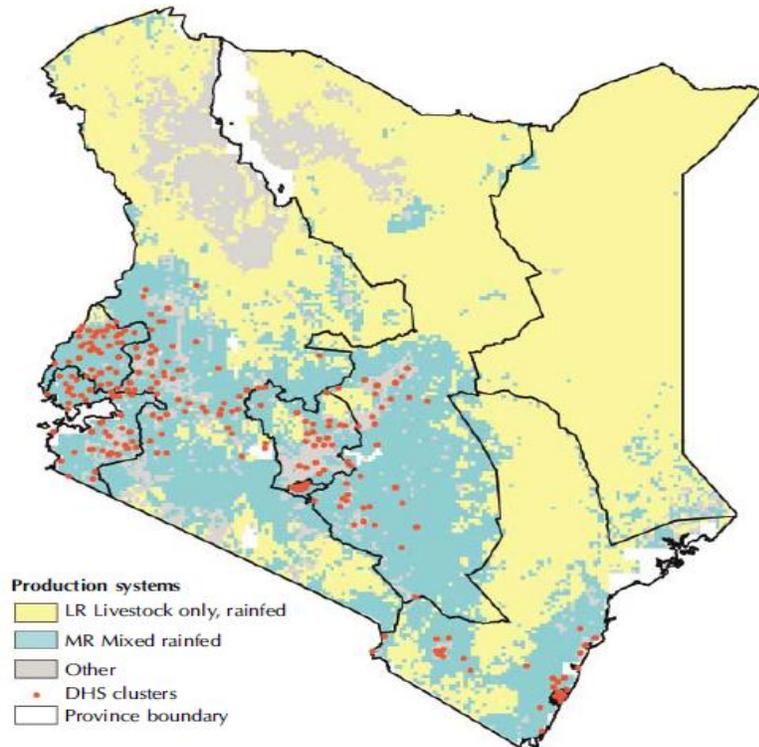


Figure 9.2-Plotting Different Kinds of Response Functions

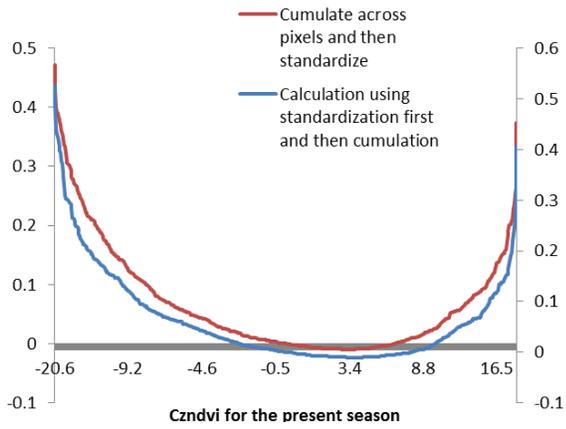


Figure 9.3 Comparing seasonal differences in mortality

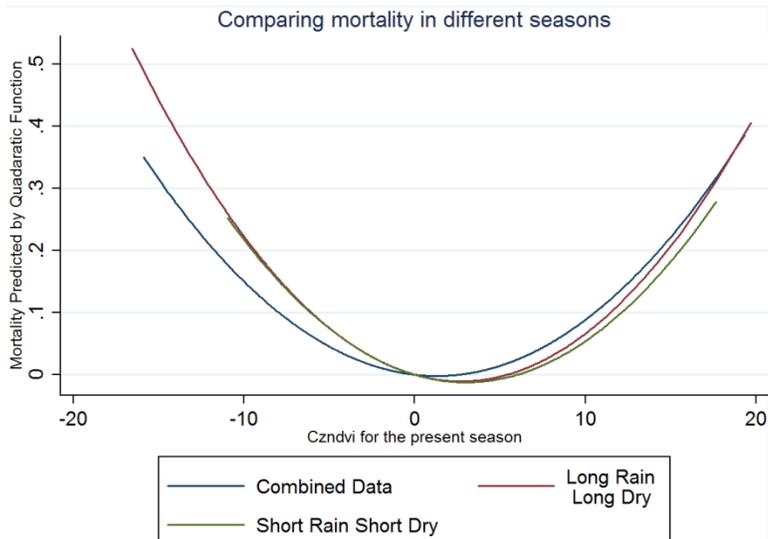


Figure 9.4-Examining different ways of plotting response functions

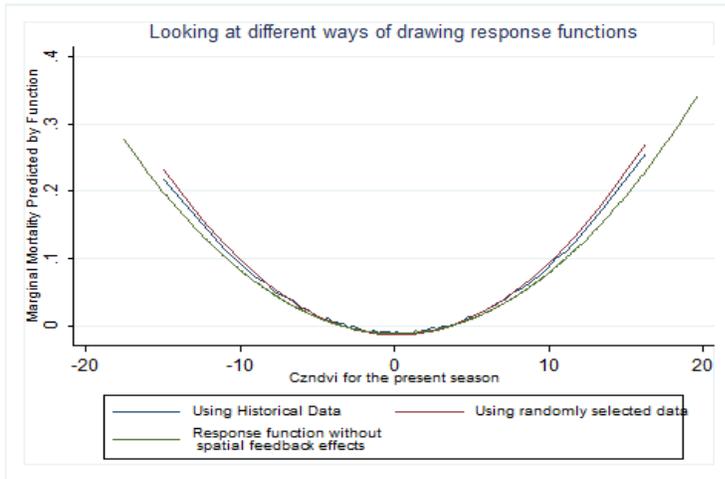
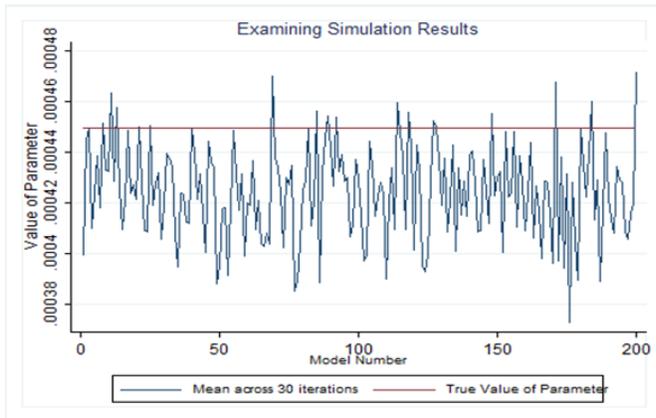
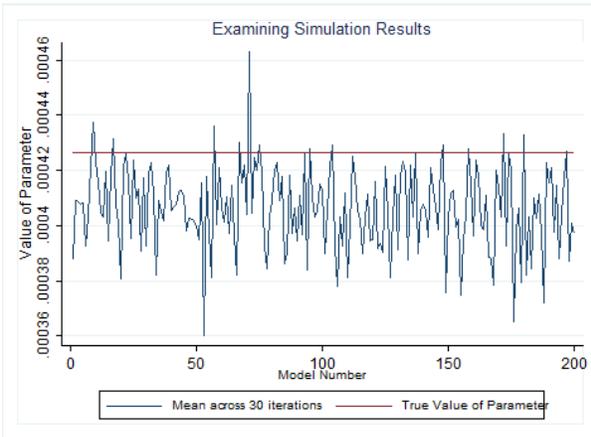
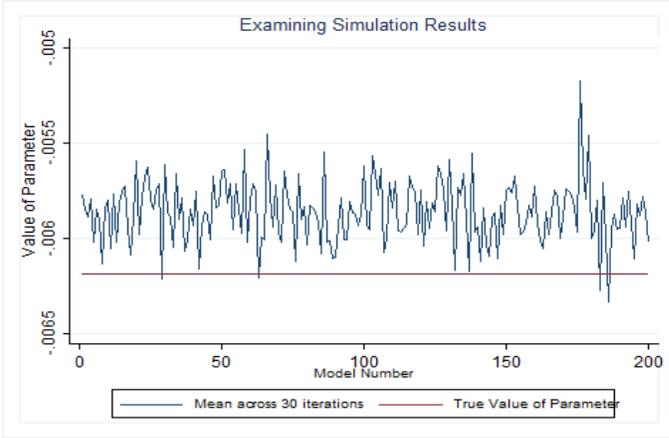


Figure 9.5-Remaining plotted tables for simulation exercise





Additional Tables

Table 9.1 Percentage Distribution of Magnitude of Shock on Households by Region

Region	Own HH Only	Some other HH too	Most HH in community	All HH in Community	Total Count
Kenya	35.6	24.0	17.5	22.8	5,777,253
Rural	34.4	22.6	18.8	24.1	4,502,177
Urban	40.7	29.7	12.2	17.3	1,275,076
Nairobi	40.2	40.5	13.5	5.9	477,448
Tana River	14.7	2.9	7.9	74.5	39,660
Isiolo	18.4	10.8	14.3	56.5	17,844
Marsabit	14.9	15.5	23.5	46.2	23,498
Moyale	7.3	13.4	5.0	74.1	10,388
Garissa	22.3	47.2	13.7	16.9	49,053
Mandera	8.4	10.8	17.9	62.9	41,871
Wajir	15.7	10.7	18.3	55.3	63,666

Baringo	18.7	17.5	39.3	24.2	54,060
Samburu	9.0	2.3	42.4	46.2	32,720
Turkana	24.5	7.8	36.0	31.8	82,501

Table 9.2 Missing LRLD TLU Data

Missing Data ALRMP	
Total Number of Divisions	108
Total No of Years	11
Total Data Points	1188
Missing Data Points	526
Divisions with no data	23
Divisions with complete data	9

Table 9.3-District Codes and Names

Number	Name
1	TANA RIVER
2	BARINGO
3	GARISSA
4	IJARA
5	ISIOLO
6	MANDERA
7	MARASBIT
8	MOYALE
9	SAMBURU
10	Turkana
11	Wajir

Table 9.4-Livestock Density

Sheep Density	Goat Density	Cattle Density	District ID
9.001837373	11.33040704	9.205688316	1
52.46003586	175.6155164	55.01820442	2

1.966557657	8.282860202	10.48164789	3
10.70851244	10.41717553	2.895257746	4
8.622625636	17.73907187	8.125909578	5
7.093509933	7.814287762	3.428947125	6
1.918814972	3.076363263	0.740838416	7
30.51883566	37.56323384	12.17284242	8
15.42074663	35.05057816	2.340225168	9
6.096518663	3.204276364	4.292600146	10
19.75346662	33.08870768	13.09793769	11

Table 9.5- Figures for Total Livestock Averaged Across Seasons

Division ID	Sheep	Goat	Camel	Cattle
1	5.113264	24.7287	0.056948	5.474476
2	-99	-99	-99	-99
3	-99	-99	-99	-99
4	-99	-99	-99	-99
5	3.835407	36.75377	0.035122	1.431783
6	4.895727	28.54834	0.910224	7.729913
7	-99	-99	-99	-99
8	-99	-99	-99	-99
9	6.893794	41.98175	1.772699	5.350541
10	8.688169	22.48899	0.032546	4.10379
11	4.367676	34.20755	0.489752	7.312296
12	6.634126	11.54485	0.003338	5.740092
13	10.85217	12.71013	3.854004	12.3695
14	-99	-99	-99	-99
15	6.925373	34.8054	0.312859	7.212399
16	34.08702	42.73717	1.529468	3.070342
17	1.081964	3.145775	.4571486	1.390381
18	2.792631	46.18464	0.256367	5.647387
19	34.51701	15.21592	0.922028	9.095965
20	5.047529	32.74741	2.410689	6.383349
21	23.92873	24.08623	3.49888	6.68589
22	4.059267	14.03895	2.732897	8.575639
23	20.90291	23.75405	4.106739	1.671521
24	17.90121	44.52859	1.852686	1.549971

25	0.700405	21.391	0.864947	13.42206
26	20.26262	15.04093	0.036153	15.78206
27	13.01048	13.25044	0.027899	10.253
28	6.678996	5.013102	0.017877	7.84703
29	17.68201	19.87998	0.005618	16.5715
30	8.633314	12.33931	0.200772	4.656328
31	16.2149	10.66683	1.254921	10.17144
32	14.4396	10.48967	1.861758	7.397402
33	11.15329	8.346154	0.08906	4.972123
34	11.38443	3.988317	0.094796	5.769455
35	13.95552	38.98743	1.291012	6.315591
36	15.51229	66.7933	13.26339	4.883799
37	4.170966	7.835592	4.223437	6.383472
38	0.679879	2.355019	0.032314	0.219733
39	1.36453	3.159101	0.231633	0.297614
40	9.348597	28.9271	0.910898	5.672071
41	1.781719	3.90814	0.034106	2.096407
42	11.31314	29.82503	0.708838	1.719605
43	3.010913	9.140377	7.548611	2.636409
44	3.415826	12.75349	7.494574	11.02483
45	5.276431	14.45829	0.151309	3.455868
46	9.112562	32.04502	3.80009	3.166141
47	6.876662	34.69188	1.224668	1.007336
48	5.002145	21.9437	5.532976	2.613941
49	7.741907	23.56985	5.866519	5.427051
50	8.71081	16.5124	0.814226	8.437997
51	18.1198	39.16639	2.351082	0.818636
52	3.314919	7.710309	0.163984	3.113963
53	3.45966	9.056263	3.089172	3.805732
54	1.729915	3.457423	0.101999	5.239171
55	8.184366	12.69006	1.441694	2.600859
56	25.2607	20.68101	4.024559	2.740302
57	15.52186	20.06195	3.148231	7.96586
58	-99	-99	-99	-99
59	29.01472	21.47771	6.860465	1.575635
60	-99	-99	-99	-99
61	1.562302	6.420432	2.004029	4.905612
62	2.478539	7.565135	0.365211	5.557229
63	1.77673	7.318396	0.929245	5.254717

64	4.788153	13.15597	1.64257	2.085133
			.5680228	
65	10.60903	11.67446	0	3.88624
66	5.257691	19.90886	0.799062	4.238373
67	4.839719	14.14355	1.307597	6.018417
68	15.4632	13.02921	0.116433	3.860704
69	9.084415	7.723645	0.146972	4.265289
70	-99	-99	-99	-99
71	1.78687	2.747859	1.046051	1.684872
72	4.002392	25.17907	1.871749	8.318386
73	0.476428	3.183979	0.037755	0.393528
74	2.229369	4.991196	0.029797	4.220212
75	7.766684	7.798215	0.054536	13.58053
76	1.777815	2.412155	0.008635	1.821322
77	5.640953	4.137129	0.030168	6.040088
78	14.52733	43.18148	2.292943	1.05692
79	8.505397	24.99003	0.595995	0.243202
80	-99	-99	-99	-99
81	-99	-99	-99	-99
82	-99	-99	-99	-99
83	-99	-99	-99	-99
84	7.396134	64.4479	2.107531	0.723705
85	9.773597	35.7069	2.555231	1.130769
86	18.62169	42.77661	1.621271	2.758493
87	-99	-99	-99	-99
88	-99	-99	-99	-99
89	-99	-99	-99	-99
90	-99	-99	-99	-99
91	-99	-99	-99	-99
92	0.750416	4.508736	1.172213	5.544509
93	-99	-99	-99	-99
94	8.970769	30.33375	1.440519	1.545627
95	2.655875	3.901463	1.906258	4.433229
96	4.183171	4.597743	0.755259	3.015393
97	0.589281	2.738037	1.634947	2.923708
98	3.107116	8.150265	2.443603	4.14383
99	2.3	2.761404	0.562281	2.378947
100	5.637403	10.24508	5.227252	3.538637
101	8.986184	10.80049	5.099959	4.358391
102	1.665342	9.048565	1.318322	3.947903

103	15.04167	19.52827	5.329613	4.256696
104	18.30007	15.07534	4.547987	8.836946
105	20.35807	22.8965	4.762021	5.099511
106	-99	-99	-99	-99
107	-99	-99	-99	-99
108	18.2109	12.60964	7.114721	3.536585

Table 9.6-Division Specific Losses

Division Number	Losses/Division	
	LRLD	SRSD
	District Level Model	District Level Model
1	0.874296503	87.42965033
2	1.032019077	103.2019077
3	0.892021371	89.20213709
4	0.761397849	76.13978487
5	0.920169694	92.01696939
6	0.921112194	92.11121939
7	1.016454295	101.6454295
8	1.021114415	102.1114415
9	8.84E-01	88.39957297
10	1.127157421	112.7157421
11	1.009922518	100.9922518
12	1.171954544	117.1954544
13	0.792131551	79.21315507
14	0.978739779	97.87397788
15	2.440025727	244.0025727
16	2.700985831	270.0985831
17	2.169029673	216.9029673
18	1.668304869	166.8304869
19	1.514397046	151.4397046
20	1.652834693	165.2834693
21	1.755424286	175.5424286
22	2.389460703	238.9460703
23	2.552600004	255.2600004
24	2.309287123	230.9287123
25	2.518408536	251.8408536
26	1.756236604	175.6236604
27	1.70360228	170.360228

28	1.823161981	182.3161981
29	1.781136928	178.1136928
30	2.43E+00	243.0836204
31	2.428024223	242.8024223
32	2.507492122	250.7492122
33	2.11934396	211.934396
34	2.173905038	217.3905038
35	2.664830915	266.4830915
36	1.95E+00	195.4506817
37	1.815516479	181.5516479
38	3.116850945	311.6850945
39	2.379342203	237.9342203
40	2.295191935	229.5191935
41	2.737871826	273.7871826
42	2.536555438	253.6555438
43	2.40E+00	240.1556244
44	2.045923261	204.5923261
45	2.382104375	238.2104375
46	2.185737885	218.5737885
47	1.83E+00	183.2564063
48	2.474842232	247.4842232
49	1.932147521	193.2147521
50	2.918423522	291.8423522
51	2.202349183	220.2349183
52	2.898403837	289.8403837
53	1.563793758	156.3793758
54	1.871524384	187.1524384
55	1.374510057	137.4510057
56	1.76823466	176.823466
57	1.79211362	179.211362
58	4.94831429	494.831429
59	3.05E+00	304.7250997
60	3.701613848	370.1613848
61	3.156645083	315.6645083
62	4.105095071	410.5095071
63	3.347355912	334.7355912
64	4.344664692	434.4664692
65	6.607643674	660.7643674
66	4.606534333	460.6534333

67	4.299791635	429.9791635
68	5.863506434	586.3506434
69	4.804766426	480.4766426
70	2.235744474	223.5744474
71	2.05E+00	205.1569395
72	2.454035096	245.4035096
73	1.741933509	174.1933509
74	2.336948468	233.6948468
75	1.363694806	136.3694806
76	1.487993049	148.7993049
77	1.657504927	165.7504927
78	3.122055052	312.2055052
79	2.242122725	224.2122725
80	2.877628634	287.7628634
81	3.045913267	304.5913267
82	3.16192889	316.192889
83	1.818957904	181.8957904
84	2.128782343	212.8782343
85	1.962392792	196.2392792
86	3.360739911	336.0739911
87	1.984305143	198.4305143
88	2.741482561	274.1482561
89	3.589072672	358.9072672
90	2.377285213	237.7285213
91	2.506470832	250.6470832
92	2.734480867	273.4480867
93	2.146331343	214.6331343
94	2.501141165	250.1141165
95	5.686213207	568.6213207
96	4.355683022	435.5683022
97	5.286415442	528.6415442
98	4.343093801	434.3093801
99	5.127260064	512.7260064
100	6.061588755	606.1588755
101	4.109085073	410.9085073
102	5.238550084	523.8550084
103	4.455722155	445.5722155
104	4.876655211	487.6655211
105	4.322270122	432.2270122

106	4.344915318	434.4915318
107	4.715548235	471.5548235
108	4.323803146	432.3803146

Matlab Code

```

% running code for regression with actual data
%used this to create initial results and prepare data for spatial models
W=sparse(queenraw(:,1),queenraw(:,2),queenraw(:,3)); %build weight matrix
[N,T]=size(mortwmissing_sheep2);
xcoordMat=repmat(xykenya(:,1),1,T);
ycoordMat=repmat(xykenya(:,2),1,T);
xcoordStack=xcoordMat(:);
ycoordStack=ycoordMat(:);
preNDVIStack=(czndvi(:,2));
NDVIStack=(czndvi(:,3)-czndvi(:,4));
%stack if not NAN
MortMissStack=mortwmissing_sheep2(:);
MortStack=MortMissStack(~isnan(MortMissStack));

%stack and pick out NANs for other variables
areaIDstack=areaIDmat(:);
areaID=areaIDstack(~isnan(MortMissStack));
xcoordStackMiss=xcoordStack(~isnan(MortMissStack));
ycoordStackMiss=ycoordStack(~isnan(MortMissStack));
[bMiss,catchIDsMiss]=buildInterceptUnbal(areaID); %build intercept:NOTE, will not be in order
preNDVIStackMiss=preNDVIStack(~isnan(MortMissStack));
NDVIStackMiss=NDVIStack(~isnan(MortMissStack));
%running the simplest regression
NDVIsPAR=[ preNDVIStackMiss NDVIStackMiss ];
b=[1];
b=repmat(b,575,1);
regressor=[b,NDVIsPAR];
%add on other variables
respar=ols(yinit,regressor);

NDVIs=[ preNDVIStackMiss NDVIStackMiss preNDVIStackMiss.^2 NDVIStackMiss.^2];
regressor1=[b,NDVIs];
yinit=MortStack;
%add on
respar1=ols(yinit,regressor1);
%get initial x and y and fitted variables
NDVIs=[ preNDVIStackMiss NDVIStackMiss preNDVIStackMiss.^2 NDVIStackMiss.^2];
xvar=[bMiss NDVIs];
res=ols(yinit,xvar);
results=mean(res.beta(1:87),1);
xvarInit=[b NDVIs matmul(NDVIs,xcoordStackMiss) matmul(NDVIs,ycoordStackMiss)];
resinit=ols(yinit,xvarInit);

```

```

xvarInitFE=[bMiss NDVIs matmul(NDVIs,xcoordStackMiss) matmul(NDVIs,ycoordStackMiss)];
resinitFE=ols(yinit,xvarInitFE);
results2=mean(resinitFE.beta(1:87),1);
%calc aggregation functions for ols functions
mIntC=size(xintC,2);
mIntD=size(xintD,2);
Xvar=[czndvi(:,2),czndvi(:,2).^2,(czndvi(:,3)-czndvi(:,4)),...
      (czndvi(:,3)-czndvi(:,4)).^2];
%   czndvi(:,4).*czndvi(:,2),...neg pre interaction
%   czndvi(:,4).*czndvi(:,3),...%inter seas vol
%   ]; %6 varibl
[N,T]=size(mortwmissing_tlu2);
czndvopreavg=mean(czndvi(:,2),1);
response=[1,1,czndvopreavg,(czndvopreavg^2)];
response= repmat(response,N*T,1);
response(:,4)=Xvar(:,3);
response(:,5)=Xvar(:,4);
b=[1];
b= repmat(b,N*T,1);
Xvarsimple=[b,Xvar];
simpleols=ols(y,Xvarsimple);
betas=simpleols.beta;
responseall=response*betas;
responseone=[response(:,4),responseall];
responseone=sortrows(responseone);
xDistrict=[xintD,xintD.*(Xvar(:,1)*ones(1,mIntD)),...
           xintD.*(Xvar(:,2)*ones(1,mIntD)),...
           xintD.*(Xvar(:,3)*ones(1,mIntD)),...
           xintD.*(Xvar(:,4)*ones(1,mIntD))];

olstdistrict=ols(y,xDistrict);
DistrictBeta=reshape(olstdistrict.beta,11,5);
for i=1:11;
    responsealldist(:,i)=response*(DistrictBeta(i,:));
end
responsedist=[response(:,4),responsealldist(:,1)];
responsedist=sortrows(responsedist);
DistrictBeta=mean(DistrictBeta,1);
xDistrict3=[xintC, xintD.*(Xvar(:,1)*ones(1,mIntD)),...
           xintD.*(Xvar(:,2)*ones(1,mIntD)),...
           xintD.*(Xvar(:,3)*ones(1,mIntD)),...
           xintD.*(Xvar(:,4)*ones(1,mIntD))];
olsmodel3=ols(y,xDistrict3);
DistrictBeta3=reshape(olsmodel3.beta(109:152),11,4);
Intercepts=grpstats(olsmodel3.beta(1:108),DistID);
betafull=[Intercepts,DistrictBeta3];
for i=1:11;
    responsealldist3(:,i)=response*(betafull(i,:));
end
DistrictBeta3=mean(DistrictBeta3,1);
Intercept3=mean(olsmodel3.beta(1:108),1);
xDistrict4=[xintD, xintC.*(Xvar(:,1)*ones(1,mIntC)),...
           xintC.*(Xvar(:,2)*ones(1,mIntC)),...

```

```

    xintC.*(Xvar(:,3)*ones(1,mIntC)),...
    xintC.*(Xvar(:,4)*ones(1,mIntC));
olsmodel4=ols(y,xDistrict4);
DistrictBeta4=reshape(olsmodel4.beta(12:443),108,4);
for i=1:4;
    beta4(:,i)=grpstats(DistrictBeta4(:,i),DistID);
end
betaall4=[olsmodel4.beta(1:11),beta4];
for i=1:11;
    responsealldist4(:,i)=response*(betaall4(i,:));
end
DistrictBeta4=mean(DistrictBeta4,1);
Intercept4=mean(olsmodel4.beta(1:11),1);
xDivision=[xintC, xintC.*(Xvar(:,1)*ones(1,mIntC)),...
    xintC.*(Xvar(:,2)*ones(1,mIntC)),...
    xintC.*(Xvar(:,3)*ones(1,mIntC)),...
    xintC.*(Xvar(:,4)*ones(1,mIntC))];
olsmodel5=ols(y,xDivision);
DistrictBeta5=reshape(olsmodel5.beta(1:540),108,5);
for i=1:5;
    betaall5(:,i)=grpstats(DistrictBeta5(:,i),DistID);
end
for i=1:11;
    responsealldist5(:,i)=response*(betaall5(i,:));
end
DistrictBetamean5=mean(DistrictBeta5,1);
%plotting out responsefunctions
frstata=[response(:,4),responseall,responsealldist(:,1),responsealldist3(:,1),responsealldist4(:,1),responsealldist5(:,1)]
;
frstata=sortrows(frstata);
%plotting functions for specific divisions
%for district specific and all district is same
% for model3
DistrictBeta3=reshape(olsmodel3.beta(109:152),11,4);
beta1=olsmodel3.beta(71);
Divmodel3=[beta1,DistrictBeta3(1,:)];
responsediv3=response*(Divmodel3');
%for model4
DistrictBeta4=reshape(olsmodel4.beta(12:443),108,4);
Divmode4=[olsmodel4.beta(1),DistrictBeta4(71,:)];
responsediv4=response*(Divmode4');
responsedistrict4model=[response(:,4),responsediv4];
responsedistrict4model=sortrows(responsedistrict4model);
%for divlevel model
DistrictBeta5=reshape(olsmodel5.beta(1:540),108,5);
responsediv5=response*(DistrictBeta5(71,:));
%trying to plot out all of them
% will have to delete first 69 columns
for i=70:77;
    responsediv(:,i)=response*(DistrictBeta5(i,:));
end
frstataalldivs=[response(:,4),responsediv(:,70),responsediv(:,71),responsediv(:,72),responsediv(:,73),responsediv(:,74),responsediv(:,75),responsediv(:,76),responsediv(:,77)];

```

```

frstataalldivs=sortrows(frstataalldivs);
responsedivisionmodel=[response(:,4),responsediv5];

responsedivisionmodel=sortrows(responsedivisionmodel);
frstata=[response(:,4),responsealldist(:,1),responsediv3,responsediv4,responsediv5];
frstata=sortrows(frstata);
%spatial analysis
%get intercepts that correspond to the correct locations using catchIDs
inters=vlookup([1:N]',catchIDsMiss,resinitFE.beta(1:size(catchIDsMiss,1)),ones(size(catchIDsMiss,1),1));
inters1=inters;
%initialize the vals with the missing intercepts with the avg, then we'll
%loop through replacing only those that were missing with their spat wgt
%avg
inters1(inters==0)=mean(inters(inters~=0)); %replace the missing locations with the avg intercept val
for i=1:10 % just loop until converges...should be after a few loops
Winters=W*inters1; %calc spatial weighted avg
inters1(inters==0)=Winters(inters==0); %update those that were originally missing with their spat wgt'd avg
end

%build full intercept mat
[b,catchIDs]=buildInterceptUnbal(areaIDstack);
% Get fitted values for all obs including missin areas
NDVIFit=[ preNDVISTack NDVISTack preNDVISTack.^2 NDVISTack.^2];

xfitInitFE=[b NDVIFit matmul(NDVIFit,xcoordStack) matmul(NDVIFit,ycoordStack)];

BetaFitInit=[inters1 ; resinitFE.beta(size(catchIDsMiss,1)+1:end)];

yFitInitFE=xfitInitFE*BetaFitInit;

Yinit=MortMissStack;
Yinit(isnan(MortMissStack))=yFitInitFE(isnan(MortMissStack));
% Yinit is the initial vals we'll use in the spatial regression model

mortRaw=reshape(Yinit,N,T);
%Run spatial model interative procedure
iN=eye(size(mortRaw,1),size(mortRaw,1)); %identity matrix for yields
mortYears=[1:size(mortRaw,2)'];
iT=eye(size(mortRaw,2),size(mortRaw,2)); % %identity matrix for time to create panel weight matrix
Wt=kron(iT,W); % create panel weight matrix
iNT=kron(iT,iN);
mortvec=mortRaw(:); %stacks data into vector: yieldsRaw must be crosstab of yields

y=mortvec; % just calling y yields for smiplicity

DIVids=[1:size(mortRaw,1)']; % 108 divs
div_idmat= repmat(DIVids,size(mortRaw,2),1); %stack an id matrix for resorting later

[nDivs,nYears]=size(mortRaw); %total years=11
[nObs,~]=size(mortvec);
[nDists,~]=size(DistNum);

```

```
%% Build district and county xint and xtrends
```

```
xint_C=buildIntercept(DIVids,[DIVids,DIVids]);  
xint_repC=repmat(xint_C,nYears,1); %replicate T tilings  
xintC=sparse(xint_repC); %sparsify, xint for counties  
xint_D=buildIntercept(DistID,[DistNum,[1:nDists]]);  
xint_repD=repmat(xint_D,nYears,1); %replicate T tilings  
xintD=sparse(xint_repD); %sparsify, xint for districts
```

```
trendVec=[1:nYears]';  
xtrend_C=buildtrend(DIVids,[DIVids,DIVids],trendVec);  
xtrendC=sparse(xtrend_C); %sparsify, County trends
```

```
xtrend_D=buildtrend(DIVids,[DistNum,[1:nDists]],trendVec); %build it using the yield div addresses and DivNum  
match  
xtrendD=sparse(xtrend_D); %sparsify, District trends
```

```
panelinfo.rmin=-1; %just to speed it up  
panelinfo.rmax=1; %just to speed it up  
panelinfo.model = 0; %default  
panelinfo.lflag=1; %default  
panelinfo.order =50; %default  
panelinfo.iter =30; %default
```

```
ytempLogit=y;
```

```
NumMissing=sum(isnan(MortMissStack));  
Xvar=[czndvi(:,2),czndvi(:,2).^2,(czndvi(:,3)-czndvi(:,4)),...  
      (czndvi(:,3)-czndvi(:,4)).^2];  
%different levels of aggregation  
mIntC=size(xintC,2);  
mIntD=size(xintD,2);  
% for calculating the simplest of response functions  
response=[1,-0.79,1.58];  
response=repmat(response,nObs,1);  
response(:,4)=Xvar(:,3);  
response(:,5)=Xvar(:,4);  
b=[1];  
b=repmat(b,nObs,1);  
Xvarsimple=[b,Xvar];
```

```
xDistrict=[xintD,xintD.*(Xvar(:,1)*ones(1,mIntD)),...  
          xintD.*(Xvar(:,2)*ones(1,mIntD)),...  
          xintD.*(Xvar(:,3)*ones(1,mIntD)),...  
          xintD.*(Xvar(:,4)*ones(1,mIntD))];
```

```
xDistrict3=[xintC, xintD.*(Xvar(:,1)*ones(1,mIntD)),...  
          xintD.*(Xvar(:,2)*ones(1,mIntD)),...  
          xintD.*(Xvar(:,3)*ones(1,mIntD)),...  
          xintD.*(Xvar(:,4)*ones(1,mIntD))];
```

```
xDistrict4=[xintD, xintC.*(Xvar(:,1)*ones(1,mIntC)),...
```

```

xintC.*(Xvar(:,2)*ones(1,mIntC)),...
xintC.*(Xvar(:,3)*ones(1,mIntC)),...
xintC.*(Xvar(:,4)*ones(1,mIntC));

xDivision=[xintC, xintC.*(Xvar(:,1)*ones(1,mIntC)),...
xintC.*(Xvar(:,2)*ones(1,mIntC)),...
xintC.*(Xvar(:,3)*ones(1,mIntC)),...
xintC.*(Xvar(:,4)*ones(1,mIntC))];

for i=1:5000
i
% resC_sheep(i)=sar_panel(ytempLogit,xDK,W,nYears,panelinfo);
resC_sheep(i)=sar_panel(ytempLogit,xDistrict,W,nYears,panelinfo);

% yhat=((iNT-resC_sheep(i).p*Wt)^-1)*(xDK*resC_sheep(i).beta);
yhat=((iNT-resC_sheep(i).p*Wt)^-1)*(xDistrict*resC_sheep(i).beta);
ResidNonMiss=resC_sheep(i).resid(~isnan(MortMissStack));
SigeNonMiss=std(ResidNonMiss);
sigeStd=std(resC_sheep(i).resid);
ResidSimMissing=resC_sheep(i).resid;
ResidSimMissing(isnan(MortMissStack))=norminv(rand(NumMissing,1),0,sigeStd);
% ysim_sheep=((iNT-resC_sheep(i).p*Wt)^-1)*(xDK*resC_sheep(i).beta+ResidSimMissing);
ysim_sheep=((iNT-resC_sheep(i).p*Wt)^-1)*(xDistrict*resC_sheep(i).beta+ResidSimMissing);
ytempLogit(isnan(MortMissStack))=ysim_sheep(isnan(MortMissStack));
ytempLogit(isnan(MortMissStack))=max(0.01,min(.99,ysim_sheep(isnan(MortMissStack))));
% yhat(isnan(MortMissStack))=max(0.01,min(.99,ysim_sheep(isnan(MortMissStack))));
ysim_sheep(ysim_sheep<0)=0.01;
ysim_sheep2(:,i)=ysim_sheep;
MissValsLogit_sheep(:,i)=yhat(isnan(MortMissStack));
yhat(yhat<0)=0.01;
yhat_full(:,i)=yhat;
% MissValsLogit_sheep(:,i)=yhat(isnan(MortMissStack));
BetasSim_sheep(:,i)=resC_sheep(i).beta;
PSim_sheep(:,i)=resC_sheep(i).p;
TstatSim_sheep(:,i)=resC_sheep(i).tstat;
RsqrSim_sheep(:,i)=resC_sheep(i).rsqr;
R2Sim_sheep(:,i)=resC_sheep(i).r2;
% SurfaceNeg10(:,i)=((iN-resC(i).p*W)^-1)*(ndviIntersNeg10*resC(i).beta(109:end,1));
% SurfaceNeg20(:,i)=((iN-resC(i).p*W)^-1)*(ndviIntersNeg20*resC(i).beta(109:end,1));

% SurfacePerc10(:,i)=((iN-resC(i).p*W)^-1)*(ndviIntersperc10*resC(i).beta(109:end,1));
%
% % SurfaceNeg10(:,i)=((iN-resC(i).p*W)^-1)*([xint_C,ndviIntersNeg10]*resC(i).beta);
% % SurfaceNeg20(:,i)=((iN-resC(i).p*W)^-1)*([xint_C,ndviIntersNeg20]*resC(i).beta);
% SurfacePerc10(:,i)=((iN-resC(i).p*W)^-1)*([xint_C,ndviIntersperc10]*resC(i).beta);
% SurfInterOnly(:,i)=((iN-resC(i).p*W)^-1)*([xint_C,ndviInterszero]*resC(i).beta);
end
%sorting spatial data for record and plotting
%code for cleaning out the spatial data
%first of all similar code
expectedbetas=mean(BetasSim_tlu,2);
p=mean(PSim_tlu,2);
%sorting out response functions per dist and then aggregate over dist

```

```

% for model 1
betamodel1=expectedbetas;
%for model2
betamodel2=reshape(expectedbetas(1:55),11,5);
betamodel2mean=mean(betamodel2,1);
% for model 3
intercept=expectedbetas(1:11);
beta=reshape(expectedbetas(109:152),11,4);
betamodel3=[intercept,beta];
betamodel3mean=mean(betamodel3,1);
% for model 4
betalook=reshape(expectedbetas(12:443),108,4);
meanbeta4=grpstats(betalook(:,4),DistID1);
for i=1:4;
    beta4(:,i)=grpstats(betalook(:,i),DistID);
end
betaall4=[expectedbetas(1:11),beta4];
betall4mean=mean(betaall4,1);
%for model5
betamodel5=reshape(expectedbetas(1:540),108,5);
for i=1:5;
    beta5(:,i)=grpstats(betamodel5(:,i),DistID);
end
betamodel5mean=mean(beta5,1);

% writing out code for building kind of graph josh wanted
%plotting functions using historical data method
expectedbetas=mean(BetasSim_cattle,2);
rho=mean(PSim_cattle,2);
pstd=std(PSim_cattle,1,2);% animal versus tlu
betamodel2=reshape(expectedbetas(1:55),11,5);
betamodel2mean=mean(betamodel2,1);
[N,T]=size(mortwmissing_cattle2);
czndvisort=czndvi(:,5);% calculate if its not there
czndvisort=sortrows(czndvisort);
betafinal=betamodel2;% use whatever model you want to use here
Dist1=95;
Dist2=108;
d=11;% depends on dist
mortdist1=mortwmissing_cattle2(Dist1:Dist2,:);
mortdist1=mortdist1(:);

czndviindivorder=reshape(czndvi(:,5),N,T);
czndvidist1=czndviindivorder(Dist1:Dist2,:);
czndvidist1=czndvidist1(:);
dist1stack=[mortdist1,czndvidist1];
dist1stack(any(isnan(dist1stack),2),:)=[];
reps=nObs/18;
czndviforsamp=reshape(czndvisort,18,reps);

czndvireplace=mean(czndviforsamp,1);
czndvishape=reshape(czndvi(:,5),N,T);
czndvishape= repmat(czndvishape,1,6);

```

```

for i=Dist1:Dist2;
czndvishape(i,:)=czndvireplace;
end

czndvopreavg=mean(czndvi(:,2),1);
czndvihist=[1,czndvopreavg,(czndvopreavg^2)];
czndvihist= repmat(czndvihist,(N*T*6),1);
czndvihist(:,4)=czndvishape(:);

czndvihist(:,5)=czndvihist(:,4).^2;
xhathist(:,1)=(czndvihist*(betafinal(d,:)))';
N=108;
T1=reps;
% taking rho to be whatever value the model gives=
p=rho;
for t=1:T1
    t1=1+(t-1)*N;t2=t*N;
    yhathist([t1:t2],1)=(speye(N) - p*sparse(W))\xhathist([t1:t2],1);
end
% will change depending on what district I am plotting out
yhatdist=reshape(yhathist,108,reps);
yhatdist1=yhatdist(Dist1:Dist2,:);
yhatdist1=mean(yhatdist1,1);
yhatdist1=yhatdist1(:);
czndvidist=reshape(czndvihist(:,4),108,reps);
czndvidist1=czndvidist(Dist1:Dist2,:);
czndvidist1=mean(czndvidist1,1);
czndvidist1=czndvidist1(:);
plohist=[czndvidist1,yhatdist1];
plohist=sortrows(plohist);
plot(plohist(:,1),plohist(:,2),dist1stack(:,2),dist1stack(:,1),'!');
yhats=mean(yhat_full,2);
yhatstd=std(yhat_full,1,2);
yhatpos=yhats+yhatstd;
yhatneg=yhats-yhatstd;
yhatdiv=reshape(yhats,N,T);
yhatstddiv=reshape(yhatstd,N,T);
yhatposdiv=reshape(yhatpos,N,T);
yhatnegdiv=reshape(yhatneg,N,T);
yhatdivsp=yhatdiv(Dist1:Dist2,:);
yhatpossp=yhatposdiv(Dist1:Dist2,:);
yhatnegsp=yhatnegdiv(Dist1:Dist2,:);
yhatstdsp=yhatstddiv(Dist1:Dist2,:);
yhatdivsp=yhatdivsp(:);
yhatpossp=yhatpossp(:);
yhatnegsp=yhatnegsp(:);
yhatstdsp=yhatstdsp(:);
czndviactual=czndvi(:,5);
czndviactual=reshape(czndviactual,N,T);
czndviactualDIV=czndviactual(Dist1:Dist2,:);

```

```

czndviaactualDIV=czndviaactualDIV(:);
frstata=[czndviaactualDIV,yhatdivsp,yhatstdsp,yhatpossp,yhatnegsp];
frstata=sortrows(frstata);
frstata2=[plothist];
frstata3=[dist1stack];
%to generate per div level functions replace dist1:dist2 with division over
%which you want to plot
%maximizing weights
%run oos loop
matlabpool open local 2;
startTime = clock;

for i=1:nYears
    disp(i)
    elapsedTime = etime(clock, startTime);
    fprintf('Elapsed time of iteration %2.5f', i)
    fprintf(' is %2.1f seconds\n', elapsedTime)
    %Splice out only obs without year t = i
    ytemp=xcutouti(y,i,nDivs,nObs);
    xtempDD=xcutouti(xDD,i,nDivs,nObs);
    xtempDK=xcutouti(xDK,i,nDivs,nObs);
    xtempDc=xcutouti(xDK,i,nDivs,nObs);
    %Now get out of sample fitted values
    xtempDDout=xpluckouti(xDD,i,nDivs);
    xtempDKout=xpluckouti(xDK,i,nDivs);
    xtempDcout=xpluckouti(xDK,i,nDivs);
    %Estimate models
    %resulttempCRobust=robustfit(xtempC,ytemp,'huber',1.345,'off');
    resulttempDK=ols(ytemp,xtempDK);
    resulttempDD=sar_panel(ytemp,xtempDD,W,nYears-1,panelinfo);
    resulttempDc=sar_panel(ytemp,xtempDc,W,nYears-1,panelinfo);
    %Now calc yhat out-of-sample
    YDDoos(:,i)=((iN-resulttempDD.ρ*W)^-1)*xtempDDout*resulttempDD.beta;
    YDKoos(:,i)=xtempDKout*resulttempDK.beta;
    YDcoos(:,i)=((iN-resulttempDc.ρ*W)^-1)*xtempDcout*resulttempDc.beta;
end

matlabpool close;

YDDoos=min(max(YDDoos,0),1);
YDKoos=min(max(YDKoos,0),1);
YDcoos=min(max(YDcoos,0),1);
YDDoosStack=YDDoos(:);
YDKoosStack=YDKoos(:);
YDcoosStack=YDcoos(:);

%% Maximize OSLL

%set objective function handle
fcwstar=@(x_)(-1.*outTrendModel_likeModel(YDDoosStack,YDKoosStack,x_(1,1), y));
fcwstarK=@(x_)(-1.*outTrendModel_likeModel(YDcoosStack,YDKoosStack,x_(1,1), y));
fcwstarD=@(x_)(-1.*outTrendModel_likeModel(YDcoosStack,YDDoosStack,x_(1,1), y));

```

```

%set initial weight value
wnot=0.5;

%set constraints for opt
lb(1,1)=0;
ub(1,1)=1;

%set options for opt
minconOpts=optimset('fmincon');
minconOpts=optimset(minconOpts,'Algorithm','sqp','MaxIter',1000,'TolFun',...
    .000001,'TolCon',.000001,'TolX',.000001,'UseParallel','always');

%Now solve for wstar (i.e., optimal model weight)
[wstar,fval]=fmincon(fcwstar,wnot,[],[],[],[],lb,ub,[],minconOpts);
[wstarcK,fval]=fmincon(fcwstarcK,wnot,[],[],[],lb,ub,[],minconOpts);
[wstarcD,fval]=fmincon(fcwstarcD,wnot,[],[],[],lb,ub,[],minconOpts);
% calculation losses
Divisions= repmat(DIVids,11,1);
for i=1:5000;
    yhatmeans(:,i)=grpstats(yhat_full(:,i),Divisions);
    yhatmean(:,i)=yhatmeans(:,i);

    final1(:,i)=repmat(yhatmean(:,i),11,1);

    predictedmodel(:,i)=yhat_full(:,i)-(final1(:,i)+0.1);
end

yhats=mean(yhat_full,2);
predictedmodel(predictedmodel<0)=0;
avg=mean(predictedmodel,2);
ypred=reshape(avg,108,11);
ypredbydiv=mean(ypred,2);
ypredbydist=grpstats(ypredbydiv,DistID);

```

Stata Code just used for plotting

```

// writing this code for just plotting graphs was struggling in matlab to produce good quality
//code will run after I copy variables in to stata
//this is for non-spatial aggregation model District 1 copy variable named dist1
//have not named variables but are in order of increasing level of freedom
tway (line var2 var1) (line var3 var1) (line var4 var1) (line var5 var1) (line var6 var1), ytitle(Marginal Mortality
Predicted by Function,size(small)) xtitle(Czndvi for the present season,size(small)) ///
title( Looking at impact of different levels of aggregation in Non-Spatial Models,size(medium)) ///
legend( size(small) label( 1 "One model for all areas" )label( 2 "One model per District") label( 3 "Division Level
Intercepts" "and District Level Coefficients" ) label( 4 "District Level Intercepts and ""Division Level Coefficients")
label( 5 "Division Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
/// code for plotting seasonal graph first is all then lrlld then srsd
tway (line var9 var8) (line var11 var10) (line var13 var12), ytitle( Mortality Predicted by Quadaratic
Function,size(small)) xtitle(Czndvi for the present season,size(small)) ///
title( Comparing mortality in different seasons,size(medium)) legend( label( 1 "Combined Data") label( 2 "Long
Rain"" Long Dry") label(3 "Short Rain Short Dry")) ///
graphregion(fcolor(white) ifcolor(white))

```

```

//comparing different levels of aggregation
//first is the aggregated level of aggregation // second is the non aggregated data
tway ( line var15 var14) (line var17 var16),ytitle( Mortality Predicted by Quadaratic Function,size(small))
xtitle(Czndvi for the present season,size(small)) ///
title( Comparing czndvi calculated through different methods ,size(medium)) legend( size(small) label( 1 "Cumulate
across pixels and"" then standardize") label( 2 "Calculation using standardization"" first and then cumulation")) ///
graphregion(fcolor(white) ifcolor(white))

//this is for division level comparing non-spatial ols
//first is czndvi and the rest are in increasing order of freedom

tway(line var19 var18)(line var20 var18)(line var21 var18)(line var22 var18)(line var23 var18), ///
title( "Looking at impact of different levels"" of aggregation in Non-Spatial Models for Divisions", size(medium))
ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small) label( 1 "One model for all areas" )label( 2 "One model per District") label( 3 "Division Level
Intercepts" "and District Level Coefficients" ) label( 4 "District Level Intercepts and ""Division Level Coefficients")
label( 5 "Division Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
//now adding data for czndi different kinds of response functions
//historical data, simulated data, then without responsefunction

tway(line var25 var24)( line var27 var26) (line var29 var28) , title(Looking at different ways of drawing
response functions, size(medium)) ///
xtitle(Czndvi for the present season,size(small)) ytitle(Marginal Mortality Predicted by Function,size(small)) ///
legend( size(small) label( 1 "Using Historical Data" )label( 2 "Using randomly selected data") label( 3 "Response
function without"" spatial feedback effects" ) ) ///
graphregion(fcolor(white) ifcolor(white))

//combining div and dist graphs
tway(line var20 var18)(line var23 var18), ///
title( "Looking at impact of different levels"" of aggregation in Non-Spatial Models for Divisions", size(medium))
ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "One model per District") label( 2 "Division Level Model")) ///
graphregion(fcolor(white) ifcolor(white))

//adding an variables for srsd aggregated /district
tway (line var32 var31)(line var33 var31) (line var34 var31) (line var35 var31) (line var36 var31), ///
title( "Looking at impact of different levels of aggregation in Non-Spatial Models"" SRSD",size(medium)) ///
legend( size(small) label( 1 "One model for all areas" )label( 2 "One model per District") label( 3 "Division Level
Intercepts" "and District Level Coefficients" ) label( 4 "District Level Intercepts and ""Division Level Coefficients")
label( 5 "Division Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
//adding on variables for division
tway (line var38 var37) (line var39 var37) (line var40 var37) ( line var41 var37), ///
title( "Looking at impact of different levels"" of aggregation in Non-Spatial Models for Divisions""SRSD",
size(medium)) ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "One model per District") label( 2 "Division Level Intercepts" "and District Level
Coefficients" ) label( 3 "District Level Intercepts and ""Division Level Coefficients") label( 4 "Division Level
Model")) ///
graphregion(fcolor(white) ifcolor(white))

```

```

//now adding on response function for each divison in district to compare
tway (line var43 var42) (line var44 var42) (line var45 var42) (line var46 var42)(line var47 var42) (line var48
var42) (line var49 var42) (line var50 var42), ///
title( "Looking at response functions of specific Divisions"" in Tana River District""SRSD", size(medium))
ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "TSAVO E N .PARK") label( 2 "BANGALE" ) label( 3 "BURA") label( 4
"MADOGO") label( 5 "GALORE") label( 6 "WENJE") label( 7 "KIPINI") label( 8 "GARSEN" ) ///
graphregion(fcolor(white) ifcolor(white))

//now adding spatial district function
//first adding the response function and add then adding on the actual values and then adding on pred values and
then their error(positive and then negatiive0

tway (line var52 var51) (scatter var53 var54) (rcap var57 var58 var55, sort), ///
title( "Looking at response functions of Tana River District""LRLD", size(medium)) ytitle(Marginal Mortality
Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "District Level Response Function") label( 2 "Actual Data Points" ) label( 3 "Variation
in Predicted Values from District Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
//comparing ols and spatial
tway (line var52 var51) (scatter var53 var54) (line var34 var31), ///
title( "Comparing non-spatial and spatial District Level Models""LRLD", size(medium)) ytitle(Marginal Mortality
Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "Spatial District Level Response Function") label( 2 "Actual Data Points" ) label( 3
"Non-Spatial District Level Response Function")) ///
graphregion(fcolor(white) ifcolor(white))

//using new file to build all level functions
tway(line var2 var1)( line var4 var3)(line var6 var5)(line var8 var7)(line var10 var9), ///
title( "Looking at response functions of Tana River District""Spatial Model""LRLD", size(medium)) ytitle(Marginal
Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small) label( 1 "One model for all areas" )label( 2 "One model per District") label( 3 "Division Level
Intercepts" "and District Level Coefficients" ) label( 4 "District Level Intercepts and ""Division Level Coefficients")
label( 5 "Division Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
//
//adding on div level effects for tana district so 7 effects
tway (line var12 var11)(line var14 var13) (line var16 var15) (line var18 var17)(line var20 var19) (line var22
var21)(line var24 var23)(line var26 var25), ///
title( "Looking at response functions of specific Divisions"" in Tana River District""Spatial Model""LRLD",
size(medium)) ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "TSAVO E N .PARK") label( 2 "BANGALE" ) label( 3 "BURA") label( 4
"MADOGO") label( 5 "GALORE") label( 6 "WENJE") label( 7 "KIPINI") label( 8 "GARSEN" ) ///
graphregion(fcolor(white) ifcolor(white))
//
//now adding srds functions spatial
tway (line var28 var27) (line var30 var29)(line var32 var31)(line var34 var33)(line var36 var35), ///

```

```

title( "Looking at response functions of Tana River District""Spatial Model""SRSD", size(medium)) ytitle(Marginal
Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small) label( 1 "One model for all areas" )label( 2 "One model per District") label( 3 "Division Level
Intercepts" "and District Level Coefficients" ) label( 4 "District Level Intercepts and ""Division Level Coefficients")
label( 5 "Division Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
///
//detaileddistrict level model
tway (line var43 var42 ) (scatter var44 var45) (rcap var40 var41 var37) , ///
title( "Looking at response functions of Tana River District""Spatial""SRSD", size(medium)) ytitle(Marginal
Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "District Level Response Function") label( 2 "Actual Data Points" ) label( 3 "Variation
in Predicted Values from District Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
//adding lrd for a specific div 71 and then adding srsd
//lrd for specifi div
tway (line respfunction var51) (scatter actualmort var54) (rcap yhatneg hatpos czndvi) , ///
title( "Looking at response functions of the Bangale Division in ""the Tana River District""LRLD", size(medium))
ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "District Level Response Function") label( 2 "Actual Data Points" ) label( 3 "Variation
in Predicted Values from District Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
//srsd
tway (line var61 var60) (scatter var62 var63) (rcap var58 var59 var55), ///
title( "Looking at response functions of the Bangale Division in ""the Tana River District""SRSD", size(medium))
ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small)label( 1 "District Level Response Function") label( 2 "Actual Data Points" ) label( 3 "Variation
in Predicted Values from District Level Model")) ///
graphregion(fcolor(white) ifcolor(white))
//adding tlu data
//just for resp function
// camel,cattle ,goat tlu sheep
tway (line var65 var64)(line var67 var66)(line var69 var68)(line var71 var70) (line var73 var72), ///
title( "Looking at response functions of animal specific "contracts" the Wajir District""LRLD", size(medium))
ytitle(Marginal Mortality Predicted by Function,size(small)) ///
xtitle(Czndvi for the present season,size(small)) ///
legend( size(small) label( 1 "Camel" )label( 2 "Cattle") label( 3 "Goat" ) label( 4 "Tropical Livestock Units") label(
5 "Sheep")) ///
graphregion(fcolor(white) ifcolor(white))
//now trying to plot comparisons between OLS and TLU

```