Master thesis Econometrics and management science

Predicting the state of local markets after natural disasters to inform humanitarian cash programs

Author:

Auke Tas 400701

Supervisors:

Prof.Dr. P.H. Franses Erasmus University Rotterdam

M. van der Veen 510

Second assessor:

Date:

October 2, 2019

Contents

1 Literature review

Disaster relief

This research concerns the functioning of markets after a disaster has struck the area. The topic of how to respond to a disaster has been extensively studied in the literature already, particularly in the past few decades. There have been numerous case studies concerning specific disasters (see Maxwell and Fitzpatrick 2012, Horwich 2000, Selcuk and Yeldan 2001, Vigdor 2008, Mertens et al. 2016 among others) but also many studies examining the general consequences of disasters and how best to combat them. These are more interesting for this research as its conclusions should be applicable to any disaster that might strike in the future. Research on disaster management has become more and more important over the years as the frequency of disasters increases. In fact, De Haen and Hemrich 2007 project that between now and 2050 the global cost of disasters will increase 5-fold, primarily due to climate change and the large concentration of the world population in vulnerable habitats. The link between climate change and the increasing number of natural disasters has been examined more extensively, among others by E. A. Cavallo and Noy 2009, Kellenberg and Mobarak 2011 and Strömberg 2007 for specific weather-related disaster types like floods and droughts. Moreover, Kellenberg and Mobarak 2011 find that even though the number of disasters is increasing the number of deaths and people affected per disaster is decreasing, indicating that the amount of research and investment over the years has enabled people to better counter the effects of disasters.

Evidently, it is not always possible for a country to combat the effects of a disaster on its own. Apart from the fact that some countries are inherently more vulnerable to disasters, as will be elaborated on later, disasters can cause serious damage to infrastructure that is needed to get disaster relief up and running. As such, it oftentimes falls upon other countries to provide relief in an area where a disaster has struck. The World Bank describes how in the 60's and 70's lots of high income countries heavily subsidized domestic agriculture. This led to high production surpluses which were donated as in-kind aid to developing countries facing food shortages. By in-kind aid we mean the sending of material aid like food, clothes or medicines as opposed to cash. Over the 80's and 90's most countries started to disentangle in-kind aid from agricultural targets. The decline in food surpluses available for disaster relief allowed local food production to flourish, leading to an increase in locally procured in-kind commodities from 9% to 33%

between 1990 and 2005. This is a good example of positive boosts that local economies can receive from disaster relief. Further research on these positive boosts for local economies is done by Hallegatte, Hourcade, and Dumas 2007. These effects are especially important as contrary to popular belief, oftentimes humanitarian aid is not a one-off expenditure but a long-term project. In fact, Barder et al. 2015 find that 89% of humanitarian aid goes to places that have required funding for more than 3 years. Much research focuses on how to better streamline humanitarian aid to meet the needs of people that have been affected by disasters. Allaire 2018 argues that humanitarian aid focuses too much on compensating property damage even though this might not serve the needs of all affected parties. Darcy and Hofmann 2003 find that inefficiencies in humanitarian aid are usually not down to a lack of information, but to a lack of structure in information streams and reluctance to act on this information. The view that organizations are sometimes unwilling to act upon available information is supported by Besley and Burgess 2002 and Eisensee and Strömberg 2007, who find that political systems are more inclined to protect areas against the consequences of natural disasters if their work is more 'visible', or if they are held accountable for example by the press or upcoming elections. This seems to echo the sentiments of Sen 1982 that "Starvation is the characteristic of some people not having enough to eat. It is not the characteristic of there being not enough food to eat".

Vulnerability

As alluded to before, the consequences of a disaster are not shared equally across the globe. Some areas seem inherently more exposed to natural disasters occurring, and less able to deal with them. One of the recurring themes in the literature on this subject is the fact that developing countries are hit relatively hard as argued by both Ludwig et al. 2007 and De Haen and Hemrich 2007. Albala-Bertrand 1993 confirms this view and argues that the reason is that developing countries are more dependent on agriculture than other countries, which is generally the sector that is hit the hardest by disasters. Also within countries the poorer households suffer more from the consequences of natural disaster than richer ones, as shown by Peacock, Gladwin, and Morrow 2012, Fothergill and Peek 2004 and Benjamin Wisner et al. 2004 among others.

If some areas are more exposed to natural disasters, this suggests that it could be worthwhile to take preventive measures against disasters, so before they have occurred. This as opposed to most disaster relief that we have discussed before that only takes place after a disaster has hit.

In the literature this distinction is usually referred to as 'ex ante' (before disaster strikes) and 'ex post' (after disaster strikes) relief, as further elaborated upon by Kunreuther and Michel-Kerjan 2008. Ex ante disaster relief is supported by the idea that disaster damage is not just a function of the severity of the disaster but also socio-economic factors. This was not a very popular idea up until a few decades ago as discussed by Benjamin Wisner et al. 2004. Kates 1971 and O'Keefe, Westgate, and Ben Wisner 1976 were among the first to argue that societal factors had a role to play in the shaping of disasters in the 70's, a view much more common now and aired in Noy and Yonson 2018 among others. An example is a comparison between the earthquake in Haiti in January 2010 and the earthquake in Chili in February 2010. The disaster in Haiti claimed between 200.000 and 250.000 lives, while the earthquake in Chili claimed less than 100. This is odd as the earthquake in Chili was physically stronger and hit a densely populated area. This suggests that other factors than the force of the earthquake played in role in the severity of the damage. E. Cavallo, Noy, et al. 2011 even go as far as to say that disasters are mainly economic events due to the big influence of economic factors on the effects of disasters. Schumacher and Strobl 2011 argue that the areas where disasters are likeliest to strike are usually the ones where the consequences are least severe due to spending on precautionary measures. In this research the researcher focused mainly on direct losses but it would be interesting to see if this spending also makes local markets more robust against disasters. The idea of resilience of a community against disasters has been expanded greatly over the years. Dacy and Kunreuther 1969 and Albala-Bertrand et al. 1993 both explored the idea of economic resilience against disasters as opposed to social or geographical resilience and Rose 2007 distinguished between 'static' and 'dynamic' economic resilience. He defined static resilience as the efficient allocation of resources and dynamic resilience as quick reconstruction of capital stock after a disasters. In fact, people have been studying the idea of resilience with such granularity that F. Miller et al. 2010 argue that all these different angles have led to differences in understanding, if not confusion, about the concept. Multiple attempts have also been made to quantify vulnerability and disaster risk, such as in Peduzzi 2006 and De Groeve, Poljansek, and Vernaccini 2014. Most of the research mentioned suggests that risk profiles are a promising avenue to better understand why disasters hit some areas harder than others and in this research we will definitely take this into account when looking at how disasters affect local markets.

Cash transfer programs

To provide disaster relief in affected areas this research focuses specifically on cash transfer programs. Cash transfer programs are a form of humanitarian aid that transfers money directly to people in areas that have been affected by disasters. Other forms of humanitarian aid are called 'in-kind', and mainly refer to the distribution of food, medicines etc. As these commodities tended to be resold rather than used by beneficiaries, cash transfer programs got more popular as they enable people that have been affected by a disaster to decide for themselves what commodities would help them most efficiently. The use of cash transfer programs has increased tremendously over the past 15 years. Having constituted only 1% of all humanitarian aid in the world in 2004 that number had gone up to 15% by the end of 2017 according to the Red Cross. Initially cash transfers were mainly used by governments in Latin America and later Africa but oftentimes to fight enduring poverty rather than for disaster relief. The first time humanitarian organizations started using cash transfers for disaster relief on a large scale was in the aftermath of the 2004 tsunami in the Indian ocean. After this the use of cash transfers has increased rapidly and earned rave reviews such as from the secretary general of the world health summit, who in 2016 recommended the use of "cash-based programming as the preferred and default method of support". Other endorsement of cash transfer programs can be found in Staunton 2011 who remarks that in a cast study in Zimbabwe "the injection of cash to very poor households had a much more significant positive impact on the market than distributions of food rations" and J. Miller 2002 who notes in a case study in Mozambique that "The money was spend mainly near local distribution points, and thus remains in the region, stimulating sales and job creation by retail traders". Van den Berg and Cuong 2011 on the other hand, find in a case study in Vietnam that cash transfers only marginally help against poverty and inequality as most cash transfers go to non-poor households, though it can be argued this poor targeting is not a fundamental criticism of cash transfer programs. Cash transfers can be divided in so-called conditional and unconditional transfers. Conditional cash transfers are only handed out when people meet certain criteria. These criteria can be anything from enrolling kids in school to receiving vaccinations, but are generally used to promote behaviour that the organization that hands out these cash transfers considers desirable. Unconditional cash transfers are handed out to people without any such restrictions. Barder et al. 2015 examine this distinction and find that conditional cash transfers are often used in long-term social protection programs but are rarely used for disaster relief. Davis et al. 2016 argue that cash transfer programs in Sub Saharan Africa (SSA) should be unconditional and that cash transfer programs in these countries can have multiplier effects in local economies of between 1.5 and 2.5. Cash transfers have exhibited a number of advantages over in-kind aid. Some will be mentioned here but a more thorough review can be found in Bastagli et al. 2016. Carter and Barrett 2006 suggest that cash transfers can decrease the skew of low-income households towards low-risk, low-income activities that inhibit growth and investment, possibly releasing them from poverty traps. Margolies and Hoddinott 2015 found that cash transfers are more cost efficient than food transfers and Smith and Mohiddin 2015 argue that they are less susceptible to corruption than other forms of humanitarian aid. Another advantage, as outlined by Asfaw, Davis, and Dewbre 2011, is that cash transfers can force people to alter their behaviour in a way to fight climate change, though this primarily applies to conditional cash transfers. Barder et al. 2015 note that cash transfers also bring an extra degree of transparency to humanitarian aid as for in-kind aid surprisingly little is known about what share of donated funds ends up with beneficiaries.

This is not to say no concerns have been expressed in the literature over the use of cash transfers. We mentioned before that Van den Berg and Cuong 2011 raised doubts over where the donated funds will end up. These doubts are echoed by Barrientos and DeJong 2006, who find that in three cash transfers designed to support poor households between 20 and 40% of recipients were actually among the non-poor. Lloyd-Sherlock 2006 finds in a case study in Brazil that when conditional cash transfers are contribution-based they tend to exclude people without substantial employment history which decreases coverage among the poor. Another issue which is described by Smith and Mohiddin 2015 is that it is difficult to argue when to stop cash transfer programs and that ending them now is often due to budget limitations rather than them not being effective anymore. Ikiara 2009 and Moore 2009 express fear that sudden distribution of cash will lead to increased consumption of unnecessary goods like alcohol and tobacco, though Evans and Popova 2014 debunk this fear by reviewing 19 separate studies on cash transfers and finding no signs of increased spending on alcohol or tobacco. Another possible downside is that sudden large influxes of cash in local communities can cause inflation, which would be very unhelpful to local communities right after a disaster has struck. Bailey and Pongracz 2015 argue that cash transfer programs don't cause noteworthy inflation because cash flows are relatively small compared to other cash flows and Hedlund et al. 2013 quote a 110 million dollar cash transfer program in

Somalia where food prices actually decreased, though this was probably due to global food prices decreasing at the same time. As this danger still exists, in this research we will account for the possibility that cash transfers cause inflation when examining market functioning.

Market functioning

As we can see, numerous studies have established that even though the popularity of cash transfer programs is rising rapidly it is not the best idea in every situation. Specifically, there are some conditions that need to be met in order for cash transfers to work properly. Some examples of these conditions are that people most in need have to be defined (though this applies for most humanitarian aid) and there needs to be a system in place to efficiently distribute cash in a local community. The condition necessary for cash transfer programs to function that this research focuses on is that markets need to be functioning properly. If markets don't function properly anymore this can mean that even when given an influx of cash people in a local community still cannot buy the commodities they need most. In this situation it can be more efficient to provide these commodities directly to the area that has been affected by a disaster. There are multiple reasons why market functioning might be impacted by a disaster. One that has been alluded to before is that inflation might kick in. Other reasons include that the marketplace might not be physically accessible anymore due to damage to infrastructure such as roads or that damage on the supply side of a local economy means a market does no longer have the capacity to supply everyone in a local community.

As this research focuses on countries in SSA, it is worthwhile to have a look at market functioning in this area without the influence of disasters as this can differ substantially from other geographical regions. The body of literature on this topic is not very extensive but Good 1973 finds that because of the volatile climate, weather can have a big influence on the functioning of markets, even when this weather does not qualify as a disaster. As climate change has only increased the volatility of the climate this point is even more relevant nowadays. Dercon 2004 found in rural Ethiopia that one-off weather events can have long-term effects on consumption and that even though consumption grew in the area as a whole there were notable differences between different sub-areas. Market functioning is particularly vital because the prevalence of poverty is relatively high in SSA. As Benson and Clay 2004 show, SSA contains the largest concentration of high-risk countries that are increasingly exposed to climatic disasters, so the influence that these disasters have on market functioning is very relevant. De Janvry, Fafchamps, and Sadoulet 1991 argue that particular market failures can severely constrain peasants' ability to respond to external shocks in prices, which could be eradicated by cash transfers. Additionally, malfunctioning markets are often overlooked when modelling damage assessments. This last issue has been targeted in recent years by using economic models instead of engineering. Engineering models focus on stock value which is usually too limited to capture all economic damage. economic models that focus on so-called flow loss are more adequate, as explained by Okuyama, Hewings, and Sonis 2004 and Allaire 2018.

The topic of market functioning in disaster affected areas has been studied before and numerous tools have been devised for this purpose. Specifically, the following tools were all built for the purpose of estimating market functioning or food security at some point after a disaster:

- Rapid Assessment of Markets [1]
- Emergency Mapping Market Analysis (Albu 2010)
- Market Analysis Guidance [2]
- Market Information and Food Insecurity Response Analysis (Barrett et al. 2009)
- Pre-Crisis Market Mapping and Analysis [3]
- 48 hour assessment tool [4]
- Price Monitoring, Analysis and Response Kit [5]

There are some downsides to all of these methods that we try to improve upon in this research. First of all, most of these methods are all relatively time-consuming which can seriously hinder their practical usefulness. Disaster relief is often a very time sensitive operation and having to wait a few days to know if a market is functioning or not can already be too late to help some people in the affected areas. Therefore, in this research we will only make use of data

[1] https://www.icrc.org/eng/assets/files/publications/icrc-002-4199.pdf
[2] https://www.icrc.org/eng/assets/files/publications/icrc-002-4200.pdf
[3] https://rescue.app.box.com/jc003zroe4pjzft5n83s
[4] https://foodsecuritycluster.net/sites/default/files/48%20Hour%20Assessment%20Tool.pdf
[5] https://www.crs.org/sites/default/files/tools-research/markit-price-monitoring-analysis-response-kit.pdf

that is available more or less in real time and even use the concept of vulnerability to try and give an indication of how likely a market is to malfunction if a disaster were to happen before it actually has. Another issue is that most of these methods require people on the ground in disaster affected areas to conduct surveys or collect other sources of information. Not only is this restricting the resources that organizations have available for other purposes, we have also established that disasters can affect infrastructure to the extent that it can limit the physical accessibility of this area. This means that sometimes it is not even possible for humanitarian aid organizations to visit affected areas to collect the necessary information to use the tools. We try to solve this shortcoming by using only open data which ensures that collecting information necessary to use the tool will cost virtually no time or expenses. All in all, this research hopes to provide a purely quantitative way of measuring market functioning after a disaster has struck to complement the more qualitative tools that are already available.

Data analysis in disaster relief

The use of data analysis in disaster relief is relatively new as data availability has grown exponentially over the past few decades, a phenomenon not exclusive to disasters. This increase in data enables the more quantitative approach that we described in the previous section. Specifically, this increase in data has also opened the door for the Bayesian approach that we will use in this research. Olives and Pagano 2010 were one of the first to use Bayesian statistics in humanitarian aid and argued that it would be a suitable approach for many more studies being conducted at the time. Possible applications include logistics planning for disaster response (Liu 2014) and hurricane forecasts (Taskin and Lodree Jr 2011). A particular advantage of a Bayesian approach is its suitability to work with small sample sizes, see Heudtlass, Guha-Sapir, and Speybroeck 2018, which is a very useful feature for our research as we will allude to later. A review of the use of more general data analysis methods is given by Arslan et al. 2017 who highlight the construction of early warning systems as a valuable product of big data. He also warns that the use of data is still very much in its infancy in the context of disaster relief and highlights data consistency, accuracy and completeness as some of the most important issues that still need to be tackled. His worries about the quality were voiced much earlier by Janis and Mann 1977 who argued that quantity and quality of data played a big part in potential applications of data analysis for disaster relief. Choi and Bae 2015 show how real time data analysis can help local governments manage disasters. As disaster aid is a very time-sensitive enterprise, we will attempt to use real-time data in this research as well. The use of data by governments to help disaster management is further explored in Horita et al. 2017 in a case study in Brazil. Yusoff et al. 2015 advocate the use of data for flood prediction but with the footnote that it is crucial to integrate different data sources. Over the years an increasingly varied collection of data sources has been used for disaster management. One source that has been increasingly exploited is social media data. In recent years a multitude of studies have used this source, including but not limited to Torkildson, Starbird, and Aragon 2014, Graham, Avery, and Park 2015, Grolinger et al. 2013, Cen et al. 2011, Ragini, Anand, and Bhaskar 2018, Gurman and Ellenberger 2015 and Zielinski et al. 2013. Tapia et al. 2011 temper this enthusiasm, though, and argue that even though it's very timely and accessible concerns over the reliability of the data has made humanitarian organizations reluctant to use social media data. Another frequently used source is geo-spatial data, as thoroughly reviewed by Carrara and Guzzetti 2013. Its applications include optimizing the locations of disaster shelters (Tsai et al. 2008) and modelling human mobility patterns during disasters (Song et al. 2015). Abdalla 2016 and Cutter 2003 discuss some of the challenges that still exist when using geo-spatial data, most notably accuracy of data, access restrictions, processing time and lack of familiarity with its application. There is also a number of studies using household surveys, among which De Silva and Kawasaki 2018, Henry, Takigawa, and Meguro 2017, Hoddinott 2006, Barbier 2008 and Morris et al. 2002. Because it is difficult to execute market assessments at scale, so visiting all relevant markets, this data source is difficult to use as input variable in this research. Market assessments could, however, be used to verify predictions of the model. Other less exploited data sources are crowd sourced web-based mapping services (Zook et al. 2010), though these can only cover small areas, and imagery from unmanned aerial vehicles (Ofli et al. 2016). De Albuquerque et al. 2015 and Goswami 2018 echo Yusoffs view that it is always beneficial to combine multiple data sources, which is the approach we will take in this research.

Reliability data after disasters

An additional challenge when using data analysis for disaster relief is that disasters tend to complicate reliable data collection. Disasters can cause substantial damage to infrastructure that is used for data collection and management and for obvious reasons reliable data collection is not the first concern in areas that have just been struck by a disaster. Additionally, as we do not know where a disaster might strike next it is difficult to argue where to set up monitoring systems as tracking all available countries and markets is simply not a feasible effort. Allaire 2018 argues that disaster loss data can be inflated because replacement values are reported instead of depreciated values and people tend to overestimate the damages caused to them. On other occasions disaster loss data might be under reported as as some areas do not report any numbers of the damage. Allaire also points attention to the fact that oftentimes disaster loss data is only available at national level, though in some cases data is only released on a national level but collected more locally. Anttila-Hughes and Hsiang 2013 add that death numbers are often under reported because infant deaths in subsequent years after a disaster are due to economic losses suffered in the disaster. These can be serious issues as Morris et al. 2002 show in a case study on hurricane Mitch that data unreliability can lead different researchers to distinctly different conclusions. This risk is magnified because this research considers disasters in Ethiopia, Kenya and Malawi. Reliable data collection in SSA is notoriously difficult, even without taking disasters into consideration. Onyancha 2016 even shows that only 0.03% of the worlds research data is generated in SSA. Osuteye, Johnson, and Brown 2017 show that this hampers robust analysis for disaster relief and Alegana et al. 2018 elaborate on the medical issues that this lack of reliable data collection causes. Shaffer et al. 2018 argues that this poor data management in this area is mainly down to limited internet, computer resources and trained staff.

2 Data

In order to predict market functioning the characteristics of any possible future disaster are compared to those of previous disasters in the same region. Only droughts, floods and storms are considered for this research. As mentioned before, we will only look at disasters in Ethiopia, Kenya and Malawi. This research only focuses on disasters that occurred in the last 10 years as we hypothesized that it would be very hard to get reliable data from before that time frame in this geographical area. In order to come up with a list of disasters in this time frame we consulted the EM-DAT database from the Centre for Research on the Epidemiology of Disasters at the Université catholique de Louvain (see Guha-Sapir et al. 2011 for an explanation of how this database was constructed). This database is used extensively in the literature, for example by Dilley et al. 2005 and Schumacher and Strobl 2011. The database defines 56 disasters that we are interested in in the specific time frame and geographic region. It has to be noted that there are only two storms in the dataset, both of them in Malawi. Another issue is that droughts are usually hard to pin down as a one-off event with a start and end date. They are more protracted events that slowly develop over a longer period of time. The EM-DAT database contains the geographical location and time of the disasters along with some estimates of the total damage. As these damage assessments usually take time they are not useful to include in the model as it is intended to provide an assessment of market functioning directly after a disaster has struck. This means that for the data used in this research we will combine multiple other data sources that are more accessible and informative right after a disaster has struck.

The dependent variable in this research is how well a market is functioning to find out if cash transfers are an appropriate response to a disaster. Because market functioning is a rather ambiguous concept, it will be split up into three parts:

- Accessibility
- Capacity
- Inflation

The first part is about accessibility of a local market. After a disaster roads can be blocked or flooded or the means of transport that people use can be severely limited. This can mean that people are no longer able to get to markets to buy the goods that they need. If this is the case there is not much point in giving them cash as their ability to get to a market where they can spend it is limited. The second part is about capacity. When a disaster hits this can cause serious damage to the supply side of a local market. When there is a shortage of food, clothes, medicines or another commodity, it is probably better to supply more of this commodity to the market than to give people more cash to spend on something of which there is not enough anyway. The third part focuses on inflation. This aspect has already been discussed in the literature review and can be another reason why it is not a good idea to pump more cash into a local economy. Note that both these three aspects of market functioning and the way they are influenced by the dependent variables are oftentimes related to each other. If a market is hard to access this usually hits the supply side of a market as well limiting the capacity and if the capacity of a market is decreased this can lead to an increase in prices. Each disaster under consideration will be scored on all three aspects with a zero or a one, zero indicating that this aspect of market functioning was affected by the disaster and one indicating that it was not. Unfortunately the systematic collection of quantitative data is not at all commonplace yet in the field of humanitarian aid. As as consequence, there is currently no dataset available that scores the functioning of markets after disasters on any of these three counts. The disasters were scored by hand using qualitative market assessments and reports conducted after the disasters had struck. To clarify how exactly this was done we will give an example of how we arrived at the score for a disaster. An EMMA report about the 2012 drought in Ethiopia reported that "A market based food security response in the form of cash or vouchers however is unlikely to trigger significant price increases as traders across the board reported that they had never experienced supply bottlenecks that had led to price increases", therefore, this disasters scores a one for inflation as this did not appear to hinder market functioning. A report by the Malawian government on the floods in January 2015 states that "A number of roads and bridges were destroyed and firms reported temporary stoppage of production. Market places and shops in the flood areas were closed during the period of heavy rains and flooding. Business operations were disrupted due to the heavy rains and the floods. Business time was lost afterwards when some of the buildings were reconstructed, while in some cases the business operations completely stopped.", hence this disasters scored a zero for accessibility as this report suggests that lack of access to markets severely hindered its functioning. We are very aware that this way of scoring the disasters based on qualitative reports imposes a level of uncertainty onto the results but unfortunately we see no real alternative.

The explanatory variables can be split into two parts: variables that form part of the risk profile and variables that we call state variables. Variables that form part of the risk profile don't need to have the same temporal or spatial granularity as state variables. They can be collected at any time, also before a disaster strikes, and provide an indication of how well-prepared a market is to deal with a disaster. The variables that are part of the risk profile in this research and how they might affect market functioning are:

- *Hazard*, rating for a countries exposure to hazards, measured by INFORM[6]. If a country is more exposed this might mean that markets are more prepared to maintain functioning when a disaster strikers.

- *Vulnerability*, rating for a countries socio-economic vulnerability in case a disaster strikes, measured by INFORM. If a country is more socio-economically vulnerable for a disaster this might also indicate that the functioning of markets decreases.

- *CopingCapacity*, rating for a countries institutional and infrastructural capacity to cope with a disaster, measured by INFORM. Institutional and infrastructural capacity play a role in getting markets up and running again after a disaster.

- *Corruption*, how a country scores on the Corruption Perceptions Index released by Transparency International. If a society is more corrupt this could also apply to humanitarian aid which was intended to aid market functioning.

- *BusinessEase*, how a country scores on the Ease of Doing Business Index released by the World Bank. The entrepreneurial spirit in countries where it's easier to do business might also lead to better functioning markets after a disasters

- *DietaryEnergy*, the average dietary energy supply adequacy as measured by the United Nations. If a market is not able to provide the necessary dietary energy to people without taking disasters into account, this could mean it is even less equipped to do so after a disaster.

All of these variables are aggregated nationally and available on a yearly basis. Next to the variables that make up the risk profile we use a state variable. This variable is supposed to give an indication of the functioning of a market after a disaster has hit. These variables are harder to collect as they need to be available almost in real-time straight after a disaster has occurred and need to be indicative of market functioning. State variables can also be characteristics of the disaster that just hit. The main limitation is that it has to be information that is available straight away, so total number of deaths, for example, is not an option as it usually takes some

[6] INFORM is a collaboration of the Inter-Agency Standing Committee Reference Group on Risk, Early Warning and Preparedness and the European Commission. The European Commission Joint Research Centre is the technical lead of INFORM. see http://www.inform-index.org/

time for that number to be available. It is also up for debate to what extent number of deaths is indicative of market functioning. In this research we will use the following state variable:

- $FoodPrice$, increase in food prices per local market from the World Food Programme that covers food such as maize, rice, beans, fish and sugar

Summary Statistics

Figure 1 shows a histogram of the explanatory variables in this research. Note that even though the histogram shows the variables on their original scale, they were standardized for the analysis for reasons we will explain later. The first three variables, the ratings of the risk management index, are all on a scale of 1 to 10 where on all three counts the best situation is a low rating. In this dataset they are on average 4.6, 6 and 6.5 respectively. The fact that the first rating is considerably lower is because of a collection of observations with a rating much lower than the rest, which can be seen in the top left histogram. This is because the Hazard rating of Malawi has been substantially lower over the last 10 years than the other two countries under consideration. The variable $Corruption$ is on a scale of 0 to 100 where 0 is very corrupt and 100 is very clean. The worldwide average in 2018 was 43 but the average in our dataset is substantially lower at 28. Moreover, not one observation in our dataset features an observation higher than the worldwide average, with the highest value being 37. The variable $BusinessEase$ tells what ranking a country has in a specific year among all countries in the world with respect to the ease of doing business index by the world bank. The latest ranking featured 188 countries with the observations in the dataset coming in between place 61 and 171 with an average of 119. The variable $DietaryEnergy$ measures a country's average dietary energy requirement normalized by the average supply of calories for food consumption. As the supply can exceed the energy requirement this measure can be greater than 1. In fact, the average of our dataset is slightly greater than 1 at 1.009. The variable $FoodPrice$ gives the increase in prices for selected foods such as maize, beans and bread right after a disaster occurred. On average food prices increased by little over 5% in our dataset. It was not possible to recover changes in food prices for all disasters in our dataset which means some values are missing. In the methodology section we will elaborate upon how we plan to deal with this issue.

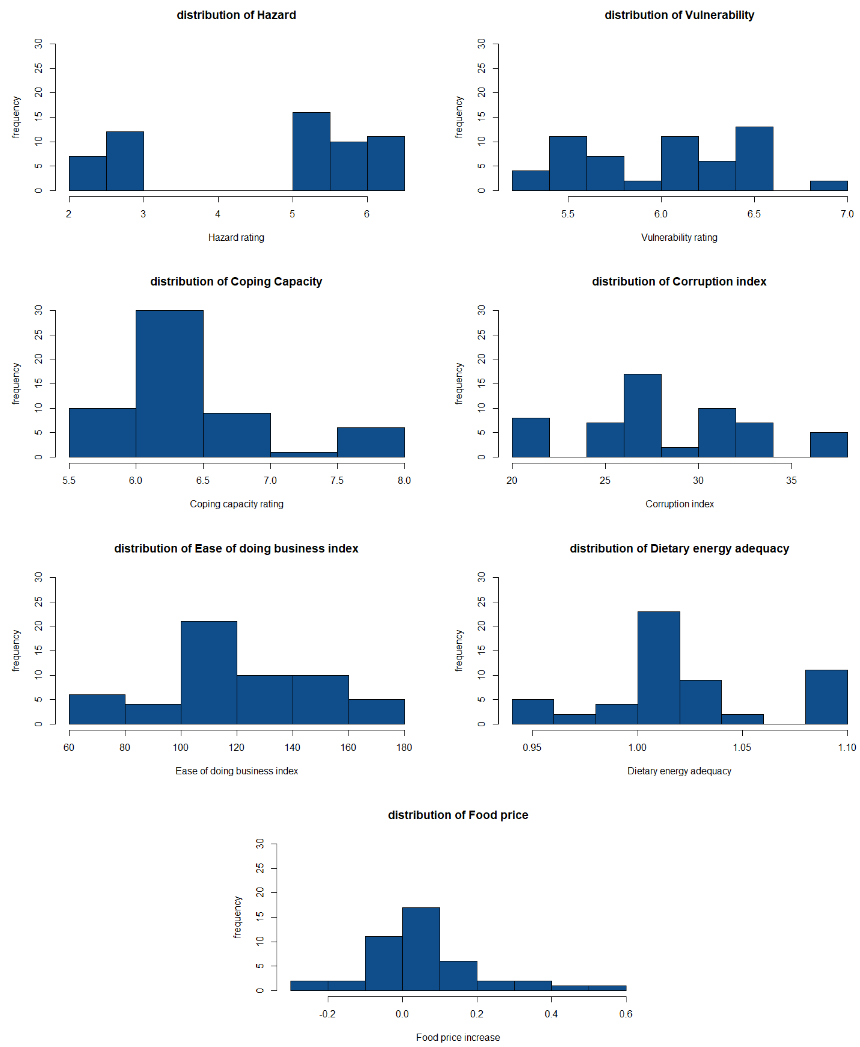

Figure 1: Distribution of explanatory variables, N=56

3 Methodology

Logistic Regression

The variable that we wish to predict is whether or not a market is still functioning after a disaster. Binary variables with only two possible outcomes can easily be modelled using logistic regression. This methodology can be extended to model any discrete number of possible outcomes but we do not need to do this here. The two possible outcomes in a logit are coded as zero and one. The model then tries to predict a probability of the outcome being one that varies between 0 (definitely a zero) and 1 (definitely a one), hence the labelling. The probability of the outcome being one is given by:

$$Pr(Y_i = 1) = \frac{e^{\mathbf{X}'\beta}}{1 + e^{\mathbf{X}'\beta}} \tag{1}$$

Where $\mathbf{X}$ is the matrix containing the explanatory variables and β the vector of coefficients to be estimated. The parameters β can be interpreted as the first derivative of the log odds ratio with respect to the corresponding explanatory variable. The log odds ration is given by:

$$log(\frac{Pr(Y_i = 1)}{Pr(Y_i = 0)}) = \mathbf{X}'\beta \tag{2}$$

Bayesian Econometrics

As mentioned before the dataset that will be used for this research features only 56 observations. Most traditional frequentist methods struggle to produce useful results with this little observations as they are based on asymptotic properties so for this research we will make use of a Bayesian approach. A Bayesian approach can have numerous advantages over a traditional frequentist approach, some of which will be discussed here. A particular advantage for this research is that a Bayesian approach is more suitable to model small datasets as discussed by Zhang et al. 2007. Kim et al. 2013 found that a Bayesian approach can help in model identification and Depaoli 2013 showed that Bayesian approach can produce more accurate parameter estimates. Where frequentist methods focus on deriving a test statistic that is a function purely of the available data, Bayesian methods require the researcher to formulate some prior expectation on the parameter that is to be estimated and evaluate to what degree the data gives a reason to deviate from this prior expectation. This allows the researcher to monitor evidence as data accumulates, as shown by Rouder 2014. A Bayesian approach also allows to distinguish between "evidence of absence" or "absence of evidence" with regards to any effect under investigation, as argued by Dienes 2014. In mathematical terms, when we try to estimate a parameter vector θ using data y, in a frequentist approach the estimated parameter $\hat{\theta}$ is the one that maximizes $p(y|\theta)$, so the value of θ that is most likely to have generated the data y. In a Bayesian approach the researcher first formulates a prior expectation given by $p(\theta)$ of how he or she expects the parameter θ to be distributed. The data and model generate a likelihood function $p(y|\theta)$. together these give the posterior distribution of parameter θ given by Bayes rule:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \propto p(\theta)p(y|\theta) \tag{3}$$

where $\propto$ means 'is proportional to'. In this case this applies because $p(y)$ does not depend on parameter vector θ, so the shape of the posterior distribution can be calculated using $p(\theta)$ and $p(y|\theta)$ and just needs to be divided by a constant to make sure it integrates to 1.

The posterior distribution $p(\theta|y)$ gives a distribution of the parameter(s) of interest, and from this distribution point and spread estimates such as the mean and the variance can be calculated. Ideally, multiplying $p(\theta)$ and $p(y|\theta)$ gives a posterior density function of the same family as the prior and likelihood function. In this case the prior density is called conjugate and in this case the posterior can be calculated analytically. In a lot of practical cases, however, this is not the case. If the posterior density cannot be calculated analytically one has to resort to numerical methods like Monte Carlo simulation. This way if we can find a way to sample values $\theta^{(m)}$ from the posterior density $p(\theta|y)$ for $m = 1, ..., M$ we can use $\frac{1}{M}\sum_{m=1}^{M} h(\theta^{(m)})$ as an approximation for any statistic $h(\theta)$ that depends on θ. Sampling error decreases with $\sqrt{M}$ for independent draws. The critical part of this procedure is that it requires the researcher to be able to make draws $\theta^{(m)}$ of the posterior density $p(\theta|y)$.

Gibbs sampler

One way of getting draws of the posterior density $p(\theta|y)$ is called the Gibbs sampler. The Gibbs sampler started receiving wide attention in the literature after the work of S. Geman and D. Geman 1987 but can be traced back as far as Metropolis et al. 1953. The Gibbs sampler starts from a random initialization of estimates for k-dimensional parameter vector θ, denoted by $\theta^{(0)}$ and initializes m to be zero. After this the algorithm iterates over the following two steps:

1. - Simulate $\theta_1^{(m+1)}$ from $p(\theta_1|\theta_2^{(m)}, \theta_3^{(m)}, ..., \theta_k^{(m)}, y)$
   - Simulate $\theta_2^{(m+1)}$ from $p(\theta_2|\theta_1^{(m+1)}, \theta_3^{(m)}, ..., \theta_k^{(m)}, y)$
   - Do the same for $\theta_3^{(m+1)}, ..., \theta_k^{(m+1)}$

2. set m to $m + 1$ and return to step 1.

After a while, say at $m = m*$, the distribution of the draws converges to the conditional posterior distribution and subsequent draws $\theta^{(m*+1)}, \theta^{(m*+2)}$, etc can be used as draws from the joint posterior distribution $p(\theta|y)$. These draws can be used to calculate any desired statistic, for example the true parameter vector θ can be estimated by the mean of the draws $\hat{\theta} = \frac{1}{M}\sum_{m=1}^{M}\theta^{(m)}$. When the draws $\theta^{(m)}$ are independent the variance of the estimator can be shown to be equal to $\frac{\sigma_\theta^2}{M}$. However, subsequent draws $\theta^{(m)}$ and $\theta^{(m+1)}$ are correlated when using a Gibbs sampler, meaning the variance of the estimator is given by $\frac{\sigma_\theta^2}{M}(1 + 2\sum_{j=1}^{M} \rho_j) \geq \frac{\sigma_\theta^2}{M}$ where ρ_j is the correlation between $\theta^{(m)}$ and $\theta^{(m-j)}$. To reduce the variance of the estimator we can either increase the number of draws M or reduce the correlation between subsequent draws ρ_j. One way of doing this is by only keeping every k'th value and deleting all the others, called thinning. If we choose a suitable value for k the subsequent draws will almost be independent. Because of the loss of information incurred from throwing away the other draws this is mainly done when it is not possible to store all draws because of memory issues. As we have no such issue here we will not use any thinning.

Model convergence

As stated before, after some time the resulting draws from the Gibbs sampler converge to the conditional posterior distribution of the parameter of interest. All draws up to this point $\theta_1, ..., \theta_m*$ are called the burn-in sample and are discarded as they are not draws from the conditional posterior distribution that we are aiming to obtain draws from. An important issue is the number $m*$ of draws that it takes for the draws to converge. In this research we will use the Gelman-Rubin statistic to estimate whether or not the draws have converged. This statistic was first introduced by Gelman, Rubin, et al. 1992 and later slightly modified by Brooks and Gelman 1998 to account for sampling variability. The Gelman-Rubin statistic is based on the comparison of multiple chains starting from different random starting values. By a chain we mean the collection of subsequent Gibbs draws from some starting point. The idea is that in the beginning of the chain the values can be different due to the different starting values but as the chain progresses closer to convergence the distribution of the different chains will edge closer towards each other and towards the conditional posterior distribution. Suppose we construct P different chains for a parameter of interest θ of length N. Note that the Gelman-Rubin statistic does not require all chains to be of equal length, but as this is the case in our research the notation hereafter will assume that they are. Let $\hat{\theta_p}$ be the sample posterior mean of the p'th sample and $\hat{\sigma_p}^2$ the sample posterior variance. Furthermore we define the overall sample posterior mean $\hat{\theta}$ to be $\frac{1}{P}\sum_{p=1}^{P} \hat{\theta}_p$. The between-chains variance B and within-chain variance W are given respectively by:

$$B = \frac{N}{P-1}\sum_{p=1}^{P}(\hat{\theta}_p - \hat{\theta})^2 \tag{4}$$

$$W = \frac{1}{P}\sum_{p=1}^{P}\hat{\sigma_p}^2 \tag{5}$$

Under certain stationarity conditions an unbiased estimator of the marginal posterior variance of θ is given by the pooled variance given by:

$$\hat{V} = \frac{N-1}{N}W + \frac{P+1}{PN}B \tag{6}$$

The potential scale reduction factor (PSRF) is then given by:

$$PSRF = \sqrt{\frac{\hat{d}+3}{\hat{d}+1}\frac{\hat{V}}{W}} \tag{7}$$

where $\hat{d}$ is the degrees of freedom estimate of a t-distribution. The PSRF gives an estimate of the potential decrease in B with respect to W. If the PSRF is large one can expect that further simulations will either decrease B or increase W as the chains have not converged yet. Brooks and Gelman 1998 suggest that if the PSRF is less than 1.2 for all model parameters one can be fairly sure that the chains have converged and subsequent draws are sampled from the conditional posterior distribution.

Data augmentation

Data augmentation is a technique to incorporate latent variables in a model and still be able to do Bayesian analysis. The presence of latent variables in a model can often lead to a posterior distribution of unknown form that is also very hard to sample numerically. We consider a model with parameter vector θ and latent variable z that generated data y. We denote the density of the data by $p(y|\theta; z)$ and the density of the latent variable by $p(z|\theta)$. For analysis both may depend on the parameter vector θ. The likelihood function $p(y|\theta)$ can be obtained by $\int p(y|\theta; z)p(z|\theta)dz$. By adding a prior $p(\theta)$ we get the posterior given by $p(\theta|y) \propto p(\theta)p(y|\theta) = \int p(\theta)p(y|\theta; z)p(z|\theta)dz$. Solving this integral with respect to z is oftentimes difficult to do analytically. A solution is to not just sample parameter vector θ but simultaneously sample the latent variable z. We can derive the full conditional posterior distributions $p(\theta|z,y)$ and $p(z|\theta,y)$ directly from the joint posterior $p(\theta,z|y)$. The Gibbs sampler thus repeats the following steps:

1. draw $z^{(m)}$ from $p(z|\theta^{(m)},y)$
2. draw $\theta^{(m+1)}$ from $p(\theta|z^{(m)},y)$

The draws $\theta^{(m)}$ can then be used to calculate any statistic of the parameter vector, such as the mean.

Hierarchical Bayes

In the previous cases we have always defined a prior specification for the parameter vector θ given by $p(\theta)$. This prior is generally a distribution that reflects prior beliefs about the values of the parameters. These prior beliefs can stem from previous research, but it is also possible to define a prior that reflects complete ignorance about the possible values of the parameters. Generally, $p(\theta)$ is a distribution that can depend on some other parameter vector. For example, one can say that the prior for θ is a normal distribution with mean μ and covariance matrix $\mathbf{\Sigma}$. More generally speaking we say that the prior of θ, $p(\theta)$, can depend on a parameter vector γ for which a value is usually chosen. Another option is to define another prior for γ and make it dependent on another parameter vector δ. Off course it is possible to define another prior for δ and make it dependent on another parameter vector and so forth. Latent variables in analysis like the so-called hyperpriors in this case can lead to a highly nonlinear integral and cause serious issues when trying to analytically estimate the posterior distribution. By using a Gibbs sampler with data augmentation we can avoid solving these integrals analytically and sample γ along with θ.

A model where you define a prior for another prior is called a hierarchical Bayesian model, and it offers some advantages over regular Bayesian analysis. One advantage of hierarchical Bayesian models is that they offer more flexibility in the way effects are correlated across observations than non-hierarchical models. In a non-hierarchical model one has to choose between either estimating a separate coefficient for each observation or pooling some observations together and estimating the same coefficient for all these observations. Both approaches might not be suitable as it is possible that effects are not the same for different individuals but are still related. In a hierarchical Bayesian setting we can reflect this by putting a prior on the parameters that might be related that distributes them around the same mean. For example, by defining the following model:

$$\beta_1 \sim N(\beta, \sigma^2)$$

$$\beta_2 \sim N(\beta, \sigma^2)$$

$$\beta_3 \sim N(\beta, \sigma^2)$$

The individual parameters β_1, β_2 and β_3 are not restricted to be the same yet they are drawn from the same distribution. MacLehose et al. 2007 shows that this hierarchical structure can help models converge when predictors are correlated. An added source of flexibility he points out is that by tuning the parameter σ^2 the researcher can change how strongly the different parameters β_1, β_2 and β_3 are shrunk towards each other. By making σ^2 a random variable and defining a prior for it it is even possible to let the data indicate to some extent how much shrinkage is appropriate. This is what we will do in this research, as will be specified later. Additionally, Gelman 2006 points out that using hierarchical structures can improve prediction which is the goal of this research.

In this research it is not possible to estimate separate coefficients for each observation, as we only have one measurement per variable per disaster. Instead, we are going to build a hierarchical model to link the effects on each of the three dependent variables: accessibility, capacity and inflation. As mentioned before, it is likely that the effects of an explanatory variable on each of these three aspects are related. Therefore we will shrink these parameters towards each other. That means that if $\beta_{Hazard1}$, $\beta_{Hazard2}$ and $\beta_{Hazard3}$ are the parameters for the variable $Hazard$ in the models to predict accessibility, inflation and capacity, their priors are defined as followed (Note that this is only one example, the full model specifications can be found in the appendix):

$$\beta_{Hazard1} \sim N(\beta_{Hazard}, \sigma^2_{Hazard})$$

$$\beta_{Hazard2} \sim N(\beta_{Hazard}, \sigma^2_{Hazard})$$

$$\beta_{Hazard3} \sim N(\beta_{Hazard}, \sigma^2_{Hazard})$$

$$\beta_{Hazard} \sim N(0, 0.5)$$

$$\sigma^2_{Hazard} \sim N(0, 100)I(> 0)$$

These specific priors for β_{Hazard} and σ^2_{Hazard} were based on previous research. Exactly why they were chosen will be explained later. In order to avoid confusion between the two, from now on the parameters $\beta_{Hazard1}$, $\beta_{Hazard2}$ and $\beta_{Hazard3}$ that estimate the effect of an explanatory variable (in this case *Hazard*) on a dependent variable are called parameters. A 1 as a suffix means the parameter is used in the model to predict accuracy, 2 for inflation and 3 for capacity. The parameters of their prior distribution, so β_{Hazard} and σ^2_{Hazard} will be referred to as hyper-parameters. Note that for every explanatory variable the prior structure is set up excactly the same, with one exception. The parameters for the variable *CopingCapacity* were sampled in a slightly different manner. The software used in this research uses precision (1/variance) as the second parameter of the normal distribution. This means that after sampling σ^2 it has to be inverted to be used to sample the three model-specific parameters. Sometimes $\sigma^2_{CopingCapacity}$ was very small which caused numerical issues when sampling the normal distribution in terms of the precision (which in turn became very large). To solve this issue the following changes were made to the prior structure:

- $\tau_{CopingCapacity} = 10/\sigma^2_{CopingCapacity}$ instead of $1/\sigma^2_{CopingCapacity}$
- $\beta_{CopingCapacity} \sim$ N(0,5) instead of $\sim$N(0,0.5)

Where τ is the precision as the software defined the second parameter of the normal distribution. Note that throughout this research when we say $\sim$ N(μ, σ^2) by σ^2 we mean to the variance.

The decrease in variance from the first change and the increase in the second offset each other to make sure that while solving the sampling issues the effects on the posterior distribution are minimal. To ensure this we have applied the same transformation to other parameters in the model and compared the posterior distribution to the one obtained without the transformation. Figure 2 shows the effect of the posterior distribution for one randomly chosen variable: the effect of *Hazard*. As we can see the changes, if any, resulting from this transformation are much too small to influence the results.

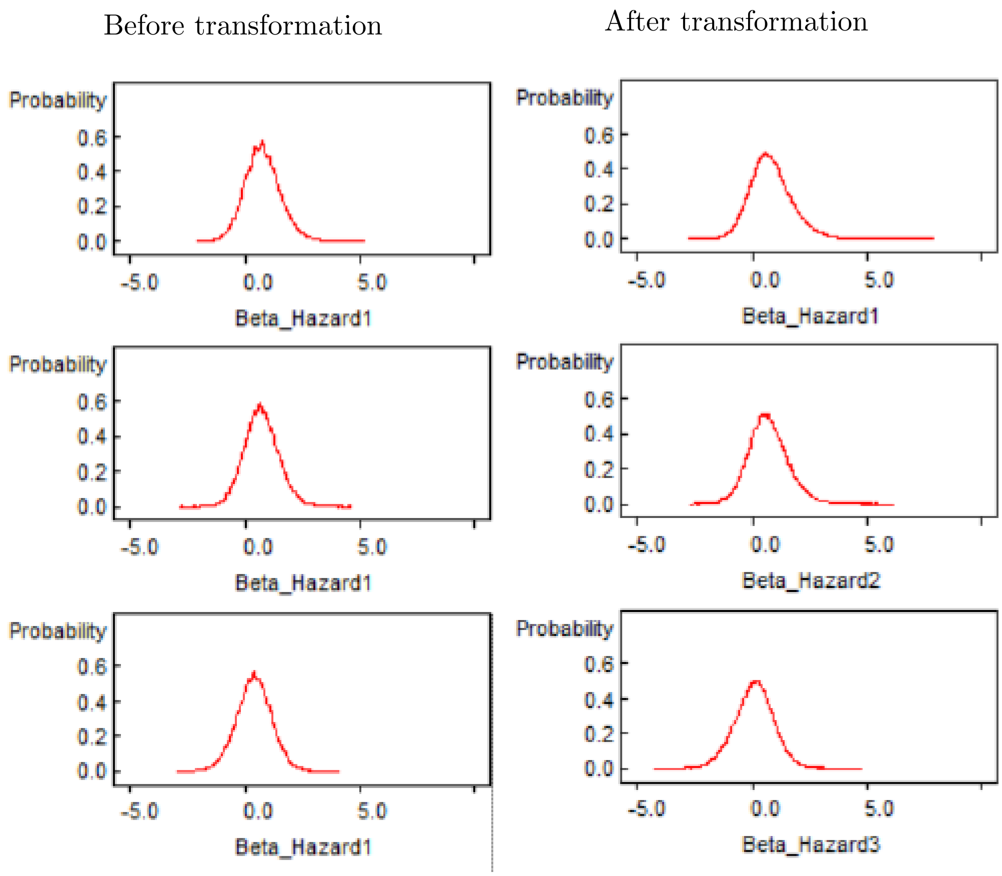

Figure 2: Comparison of posterior before and after transformation

Next we discuss the choices of the prior specifications. Priors are an essential part of Bayesian analysis. They are supposed to reflect all prior beliefs of a researcher that are not necessarily captured by the data and will be part of the research. Prior specifications that are intended to reflect explicit beliefs about the parameter values are called informative. Multiple information sources can be used to formulate these priors, such as previous research, expert opinion and common sense. Previous research can, when available, be a very useful source to base the prior on, especially if this research has used the same variables (Zondervan-Zwijnenburg et al. 2017). Unfortunately, as the functioning of markets after disasters is very much uncharted territory there is no previous research on which to base our prior beliefs. Though it can be helpful to include expert opinion, O'Hagan et al. 2006 point out that there are some issues when using this source of information. Specifically, they argue it can be difficult for experts to formulate prior beliefs on a parameter level, even more so when these parameters do not have a straightforward interpretation like in a logit model. This problem is aggravated because Bayesian analysis requires the prior specification of the parameters to be a probability distribution rather than just one estimate. Without any other obvious sources of prior information, any informative prior that we formulate will be unreliable and might bias results. When it it not possible to formulate an informative prior studies often revert to so-called uninformative priors. Uninformative priors such as a uniform distribution over the whole parameter space or normal distribution with very large variance reflect ignorance about the possible parameter value. Formulating uninformative priors can still be helpful, for example by limiting the admissible parameter space to only values within the range of measurement scale or to avoid convergence issues (Zondervan-Zwijnenburg et al. 2017). This latest advantage is especially prevalent when using small samples like the one in used in this research but is not exclusive to uninformative priors. The use of completely uninformative priors is discouraged by Lemoine 2019. He argues that formulating an uninformative prior for a parameter can still transmit a lot of information as it is hard to keep track of the effects of a prior after non-linear transformations. More importantly, Lemoine 2019 argues that by using uninformative priors, one gives up most of the advantages of Bayesian analysis and suffers from some of the same deficiencies as one would when using traditional frequentist methods. A deficiency he mentions that is particularly worrying in the context of this research is the low statistical power in small samples. Lemoine proposes to mitigate this risk by using a so-called weakly informative prior. The idea for this prior is that it imposes some regularization on the parameters. The weakly informative prior will shrink the effect sizes towards zero unless the data provides sufficient evidence that an effect is present. This is particularly useful in our research as not much is known yet about the functioning of markets after a disaster or the influence of the explanatory variables on this. Therefore we would like to have sufficient evidence from the data, especially considering the size of the available sample, to be sure of an estimated effect. Additionally, McElreath 2018 shows that regularization by weakly informative priors, even modest, can reduce type 1 errors and improve out-of-sample prediction. Lemoine 2019 proposes to use a $N(0, 0.5)$ distribution as a weakly informative prior for coefficients in logistic regression, which is why we will use this prior for the β hyperparameters. In order to make sure the shrinkage of the parameters does not depend on the scale of the variable all variables are standardized, as the research by Lemoine suggests. Because the α parameters do not really reflect effects that we wish to regularize their prior is a normal distribution with mean zero and very large standard deviation to reflect our ignorance.

The variance parameter in a hierarchical model represents the degree to which the different parameters following the same distribution are shrunk towards one another. As it is hard to quantify to what degree this should be done it is difficult to formulate an informative prior for this parameter. This applies even more in this case as to the best of our knowledge none of the variables under consideration here have been estimated in a hierarchical model together before. In practice it is oftentimes difficult to formulate an expectation on the amount of shrinkage which is why the use of uninformative priors for variance parameters is quite common (see Browne, Draper, et al. 2006). Gelman et al. 2006 discuss at length the choice of priors for variance parameters in hierarchical models. They discourage the use of the inverse gamma distribution which is often used, for example in the examples given by the BUGS software (Spiegelhalter, Thomas, et al. 1996). They argue it is not a good choice as the posterior distribution of the parameter is highly dependent on the parameter choice. This is backed up by the results from our model when using the inverse gamma distribution as a prior for the variance parameters. The results can be seen in figure 3 and clearly show that for respectively the inverse gamma(1,1), (1,2) and (7.5,1) prior the posterior distribution very closely resembles the prior. Gelman et al. 2006 instead advise to use a truncated normal distribution with large variance or a uniform prior bounded between zero and some large value. In this research we will use the truncated normal distribution with mean zero and variance 100 as prior specification for the variance hyperparameters. Although we believe the priors specified here offer some advantages, for future research it could definitely be beneficial to have more problem-specific information available to base the priors on.

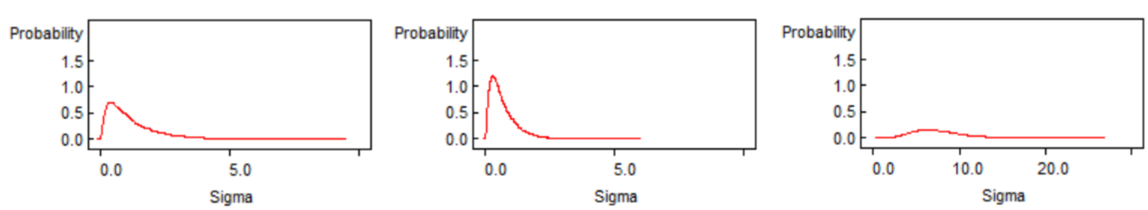

Figure 3: Posterior distrubution of variance with inverse gamma prior

Missing data

As alluded to before the dependent variable was constructed using qualitative reports on the disasters in the dataset. Unfortunately, these reports were not available for every disaster that we consider, and not every report that we could find elaborated on the state of the local markets after the disaster. This leads to a number of missing values in our dataset. Missing data is a common problem that has been extensively studied in the literature. Having missing data means that the dataset $\mathbf{Z}$ can be divided into $\mathbf{Z_{miss}}$ and $\mathbf{Z_{obs}}$ which refers to respectively the missing and the observed part of the dataset. Note that for now the missing part of $\mathbf{Z}$ is assumed to be the dependent variable unless it is explicitly mentioned that this is not the case. Most statistical methods require a full dataset, so with no missing points. There are numerous methods that are designed specifically to deal with missing data and make sure the researcher ends up with a full dataset that can be used for analysis. The choice of which one to use mainly depends on which assumptions seem reasonable with regards to the mechanism that led to the data being missing. Little and Rubin 2019 identified three different mechanisms that could lead to the missingness of data. The first and most convenient case is called Missing Completely At Random (MCAR). If this is the case every observation in the dataset has the same probability to go missing, regardless of the actual value of the variables. In mathematical terms this means that $P(\mathbf{Z_{miss}}|\mathbf{Z}) = P(\mathbf{Z_{miss}})$. No underlying mechanism can be defined that led to the data going missing and that can be used to model the missingness of the data. An example is when data comes from a household questionnaire and some questionnaires got lost in the mail. The second mechanism is called Missing At Random (MAR) and constitutes that whether or not data from an observation is missing depends only on the observed data from this observation. Mathematically this means that $P(\mathbf{Z_{miss}}|\mathbf{Z}) = P(\mathbf{Z_{miss}}|\mathbf{Z_{obs}})$. In this case the observed data can be used to model the missing data mechanism. An example is that males are less likely to fill in surveys on depression but after correcting for gender this refusal to fill in the survey is not related to actual depression levels. The third missing data mechanism is called Missing Not At Random (MNAR). This means that whether or not data from an observation is missing depends on the value that is missing. Mathematically this means that $P(\mathbf{Z_{miss}}|\mathbf{Z}) = P(\mathbf{Z_{miss}}|\mathbf{Z_{miss}}, \mathbf{Z_{obs}})$. This is the most problematic case of the three as the missing data mechanism depends on data that cannot be observed, hence it is harder to model the missing data. An example of this mechanism is when poor people are less likely to disclose their income in a survey.

Most frequentist methods that deal with missing data define a way to impute missing data and use the resulting dataset for analysis. One of the advantages of Bayesian analysis is that the missing data mechanism can be explicitly modelled and using data augmentation we can sample the missing data as we can any other unobserved quantity such as latent model parameters. In a Bayesian setting the missing data mechanism is called 'ignorable' if the following conditions are met:

- Data is MCAR or MAR
- Parameters of the missing data mechanism and the analysis model are distinct
- Prior of parameters in missing data mechanism and analysis model are independent

Where by the analysis model we mean the model that estimates the relationship between the response variable and the covariates. If the missing data mechanism is ignorable it does not have to be explicitly modelled. Instead estimating the missing values $\mathbf{Z}_{\mathbf{miss}}$ is the same as posterior prediction from the model fitted to the observed data. Therefore imputation of the missing data will not affect the model parameters. It is important to note that an ignorable missing data mechanism does not necessarily mean that we can discard the observations with missing values, only that we do not have to explicitly model the mechanism that led to the missing data. If the missing data mechanism is not ignorable this means that the mechanism has to be explicitly modelled in order not to bias the draws from the posterior distribution of the parameters. Ignorability of the data is also a different concept from the missing data mechanisms. Missing data mechanisms describe how missing cells are introduced into the dataset and ignorability describes whether this is a problem during analysis. What missing data mechanism is chosen is largely based on assumptions and cannot conclusively be derived from the available dataset. It is possible to check how sensitive results are to the assumptions made on the missing data mechanism but literature review as well as common sense are also needed to arrive at a good model for the missing data.

Aside from the response data such as our scores on market functioning, it is also possible for covariates to contain missing data points. This poses some additional challenges to the analysis. The main difference is that the response variable is treated as a stochastic quantity even when it is complete, so if the missing data mechanism is assumed to be ignorable we can sample from its posterior distribution without specifying any additional models. The covariates, however, are assumed to be fixed values. This means that no matter what assumptions we make about the missing data mechanism we will need to set up an additional model that we can sample from for covariates that contain missing values. This additional model can be either a prior we place on the covariate or a regression model. If we define a prior for the covariate that contains missing data we can choose any distribution to fit the pattern we expect from the covariate as we could with any other stochastic quantity in the model. If a binary variable U has missing values u_i for some individuals i we could define a prior saying that $u_i \sim Bernoulli(q_i)$ where $q_i \sim Beta(1, 1)$. In this case the imputed U's will depend on the posterior distribution of q estimated from the observed U's and the relationship between U and outcome variable Y that was estimated in the analysis model of interest. Note that in a Bayesian setting, for missing covariates we never have to explicitly include a relationship with the outcome variable in the missing data model as this relationship is already included in the main analysis model. This is as opposed to frequentist methods like multiple imputation where the researcher has to explicitly include this relationship. Another option is to specify a separate regression model for the missingness of the covariate. If we look at the same variable U this means we again define $U_i \sim Bernoulli(q_i)$ but this time we say that $q_i = \frac{e^{\theta * x_i}}{1+e^{\theta * x_i}}$ where x_i is a completely observed covariate that is assumed to be indicative of the missingness of U_i and θ is the associated parameter in the imputation model for which we define another prior as this is again an unknown quantity. Again we do not have to explicitly include any relationship with the response variable in the imputation model. Note that the prior or regression model that for the covariate containing missing values has to be defined regardless of the assumptions we make about the missing data mechanism. If we assume the missing data mechanism to be informative we still have to set up an extra model for the missingness indicator the same way as for missing response data.

In models where an explicit missingness model has been included we also need to define priors for the parameters in this model. As stated before all explicit missing data models are logit models with the value of the corresponding variable as the only explanatory variable in the missingness model. As the analysis model was also a logit model we will use the same rationale for the priors. This means that we include a constant with a normal prior with mean zero and large variance. The coefficient that reflects the influence of the value of a variable on it's likelihood of going missing is given a normal prior with mean 1 and standard deviation 0.5. Note that this specification is used for all missing data models, both for dependent and explanatory variables.

Model selection

Because assumptions about the missing data mechanism cannot be quantitatively checked from the data we will formulate multiple models and compare them to see which one best fits the data that we have. Bayesian researchers often use odds ratio's or Bayes factors to compare possible models but their numerical computation is oftentimes difficult, especially for hierarchical models like the ones we consider in this research. In practice, model selection is usually done using information criteria. Information criteria are a way of model selection that take into account both how well a model fits, as well as a measure of the complexity of the model to impose a penalty on possible overfitting. A common measure of the complexity of the model is the number of parameters required to estimate it. Two information criteria that use this measure and are very popular in the literature are the Akaike information criterion (AIC) proposed by Akaike 1974 and the Bayesian information criterion (BIC) proposed by Schwarz 1978. These information criteria are given by:

$$AIC = 2k - 2ln(\hat{L}) \tag{8}$$

$$BIC = log_n(k) - 2ln(\hat{L}) \tag{9}$$

where k is the number of parameters to be estimated, $ln(\hat{L})$ is the optimized value of the log likelihood of the model and n is the number of observations used to estimate the model. A common criticism of these criteria is that the number of parameters is a somewhat simplistic measure for model complexity. Additionally, in a hierarchical Bayesian setting the number of parameters k that should be used to calculate the AIC and BIC is not entirely clear. For example, imagine three parameters $\beta_{1.1}, \beta_{1.2}$ and $\beta_{1.3}$ being normally distributed around common mean parameter β_1 with variance σ^2. If σ^2 is really small the parameters will be restricted to be very similar so the number of effective parameters in practice ≈ 1. If σ^2 is really large, the parameters can have very different values and the number of effective parameters ≈ 3. There are no clear rules in this setting for how many parameters need to be taken into account when calculating the AIC or BIC. For this reason they are not applied very often to hierarchical Bayesian models. A more popular choice is the Deviation information criterion (DIC) designed by Spiegelhalter, Best, et al. 2002, and this is also the information criterion that we will use in this research. The formula for the DIC is given by:

$$DIC = -2ln(p(y|\bar{\theta})) + 2p_{DIC} \tag{10}$$

where $\bar{\theta}$ is the posterior mean given by $E_{\theta|y}(\theta) = \int \theta p(\theta|y)d\theta$ and p_{DIC} is a penalty term given by:

$$p_{DIC} = 2(ln(p(y|\bar{\theta})) - E_{\theta|y}ln(p(y|\theta))) \tag{11}$$

where $E_{\theta|y}ln(p(y|\theta))$ is given by $\int ln(p(y|\theta))p(\theta|y)d\theta$. To calculate the value of the DIC we use draws $\theta^{(m)}$ from the posterior distribution and estimate $\bar{\theta}$ by $\frac{1}{M}\sum_{m=1}^{M}\theta^{(m)}$ and $E_{\theta|y}ln(p(y\theta))$ by $\frac{1}{M}\sum_{m=1}^{M} ln(p(y|\theta^{(m)}))$ where M is the number of available draws from the posterior distribution of θ. The penalty term p_{DIC} can be interpreted as the difference between the log likelihood value at the posterior mean and the posterior mean of the log likelihood. If these quantities differ by much, this implies a lot of parameter uncertainty which is a sign that the model used might be to complex for the task at hand and we might be overfitting. As this is not what we want from a model, a bigger difference will inflate the penalty term and therefore increase the DIC. As a smaller log likelihood will also increase the DIC we generally choose the model that produces the smallest DIC. Note that all observations will be used to calculate the DIC, also the ones with missing values. This is done to make sure that the effects of different missingness assumptions are reflected in the DIC values. Though there is no scientific literature on exactly what constitutes a significant difference in DIC between models the MRC biostatistic unit in Cambridge, UK [7] argues that if models differ by less than 5 it could potentially be misleading to report only one, so we will follow their advice.

4 Results

As explained in the methodology section some choices have to be made regarding the missingness mechanisms of the data. As both response data and covariates are missing and it is not intuitively clear which mechanism is appropriate we constructed multiple models each corresponding to different assumptions about the missing data. By comparing these models we hope to find out which assumptions are most appropriate for our dataset and to model the response variable as best we can. Specifically, for the missing responses we make one of two assumptions. Either we

[7]http://users.jyu.fi/ hemipu/itms/DIC%20web%20site%20from%20BUGS%20project.pdf

assume ignorable missingness, meaning the missing responses are sampled from their posterior distribution that was estimated using the available data, or we estimate a logit model for the missingness of the response variable where the value of the missing variable is the only covariate. For the missing covariates we compare three different models. As explained before even if the missingness of a covariate is ignorable a prior has to be specified as the missing data makes the covariate stochastic. We define the prior either as a normal distribution or as a logistic regression model using the variables $Hazard$, $Vulnerability$ and $CopingCapacity$ as covariates. The third option is that the missingness of the covariates is non-ignorable. In this case we define a logit model for the missingness of the covariates, again with the value as the only explanatory variable. This results in 2 x 3 = 6 models that are compared using the DIC. All models were estimated using 5000 draws as a burn-in sample and 50.000 consequent draws that were used to calculate statistics of interest. All models were estimated using the WinBUGS [8] software. The results for all models is shown in table 1 where Y and X refer to response data and covariates respectively. As said before we consider a difference between DIC values of different models to be significant if it is five or larger. Thus, we will look at the model with the best DIC and all models with DIC that differ from it by less than five. Doing this, we can clearly see that setting up a separate model for the missingness of the response variable doesn't lead to enough improvement in the fit of the model to justify the added complexity, at least according to the DIC. Hereafter we will discuss the best models in order of DIC, so we will begin with the model that explicitly includes a model for missing covariate data, next the model using a normal prior and we will conclude with the model using a logit model as prior. They will hereafter be referred to as model 1, model 2 and model 3 in this order. We will conclude by examining whether a combination of all the models we created predicts better than any of the individual models. Note that in the appendix one can also find the model specifications of the models that we do not further discuss hereafter.

		Missingness X		
		Non-ignorable	Ignorable-Logit prior	Ignorable-Normal prior
Missingness Y	Ignorable	156.9	160.9	158.2
	Non-ignorable	166.9	167	166.2

Table 1: Deviance information criterion for various models

[8] https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/

Model 1

In model 1 all the dependent variables Y follow a Bernouilli distribution with parameter Mu_i which is given by:

$$Mu_i = \frac{e^{\alpha+\beta_1*X1i+\beta_2*X2i+\beta_3*X3i+\beta_4*X4i+\beta_5*X5i+\beta_6*X6i+\beta_7*X7i}}{1+e^{\alpha+\beta_1*X1i+\beta_2*X2i+\beta_3*X3i+\beta_4*X4i+\beta_5*X5i+\beta_6*X6i+\beta_7*X7i}} \tag{12}$$

Where the priors of α and the respective β parameters are as stated in the methodology section. As the variable $FoodPrice$ ($X7i$ in the graphic) contains some missing values we have to define a prior for it. In this example the prior is a normal distribution with parameters μ_x and σ_x. μ_x has a standard normal distribution as the data was rescaled to have mean zero and standard deviation 1. The parameter σ follows an inverse gamma distribution with both parameters equal to 1. As this model explicitly models the missingness of $FoodPrice$ we added an extra dummy to the model indicating whether the corresponding value of $FoodPrice$ was missing. This dummy is modelled using a logit specification including a constant ($a.X$) and the value of the variable $FoodPrice$ with corresponding coefficient $b.X$. Following the reasoning that led to the priors in the analysis model $a.X$ has a normal prior with mean zero and a very large variance (10^6) and $b.X$ has a normal prior with mean zero and variance 0.5.

First we investigate how sensitive the posterior distributions are to the choices that we made regarding the prior. In order to do this we change the parameter values of the priors and see to what extent they change the posterior distribution of the corresponding parameters. To keep this report concise we restrict the following parameters to have the same prior when we change them:

- all α parameters corresponding to the different dependent variables
- all β hyperparameters corresponding to the coefficients of the different explanatory variables
- all σ^2 hyperparameters corresponding to the coefficients of the different explanatory variables

As the variance of the priors of the α parameters is so large the distribution is almost flat it does not make much sense to change the mean of this prior but we will change the variance to examine how a more informative prior changes the posterior. For the prior of the β hyperparameters we will examine both the effect of changing the mean and the variance. For the prior of the σ^2 hyperparameters the argument goes as well that the large variance flattens the distribution to such as extent that changing the mean does not really change it anymore. Therefore we will only change the variance of the prior to see how the posterior changes as more probability mass in the prior is dragged towards zero. Note that all distributions that we consider as a prior for σ^2 are truncated at zero to ensure we do not sample any negative variances. For priors of the $a.X$ and $b.X$ parameters we will follow the same logic as the α and β (hyper)parameters respectively. This leads to the following prior specifications that we will compare:

$$\alpha \sim N(0, \sigma*), \sigma* \in (10^6, 10, 5, 2)$$

$$\beta \sim N(\mu*, \sigma*), \mu* \in (-1, 0, 1), \sigma* \in (10, 0.5, 0.2)$$

$$\sigma^2 \sim N(0, \sigma*), \sigma* \in (100, 10, 1)$$

$$a.X \sim N(0, \sigma*), \sigma* \in (10^6, 100, 5, 2)$$

$$b.X \sim N(\mu*, \sigma*), \mu* \in (-1, 0, 1), \sigma* \in (5, 0.5, 0.1)$$

Note that the priors we use in this research were defined before seeing the data and the above values are therefore not chosen to discover a somehow optimal parameter setting for the priors. We just choose these values to get a sense of how changing the priors would affect the results. Sometimes when we discuss the prior sensitivity of parameters we will provide plots of the posterior distribution of parameters belonging to certain variables under different priors in the appendix. Unless stated otherwise these variables are chosen randomly and the behaviour in the graphs is representative of that of the parameters corresponding to the other variables.

Figure A1 shows the results when we change the parameters of the priors of the α parameters. The main takeaway from this figure is that the posterior distribution changes very little when we make the prior more informative, as we see that the distribution is almost the same from left to right. As we wanted to make sure we do not capture any information with this prior this is not a problem, it just shows us that there are more priors that would have had the same effect. This probably means that the posterior distributions are very much dominated by the data which is good given we did not have much knowledge to base this prior on.

When examining the effect of changes in the mean of the β hyperparameters we see that they have quite a pronounced effect on the posterior distribution. This means that when we try to shrink the parameters to some other value than zero it does excactly that, which is encouraging. Especially the β hyperparameter towards which the individual parameters are shrunk shifts by almost the amount by which we change the mean of the prior. This means that this hyperparameter is heavily influenced by the prior and less by the data. This could be because the hierarchical structure makes it hard for the data to influence this hyperparameter as only the individual parameters are directly used to estimate the data. Although this hyperparameter resembles the prior very closely in its shape, the posterior of some individual parameters looks markedly different. The posterior distribution of the parameters is also less influenced by changes in the prior, sometimes it seems not at all. This seems to suggest that the data has more influence on the posterior distribution of these parameters. When there are changes in these parameters it seems to be tilted towards the value that we chose as mean of the prior, which is an encouraging sign that some shrinkage took place when the data did not provide enough evidence. To illustrate, figure A2 shows the posterior distribution of the (hyper)parameters according to the variable $Corruption$.

There is one exception to the way the priors influence the posterior distribution of the parameters. The parameters of the variable $CopingCapacity$ hardly seem to be influenced by the change in the prior. The corresponding parameters can be seen in figure A3. In this case, the change in the hyperparameter towards which the individual coefficients are shrunk is much smaller than the change in the prior. This suggests that for this hyperparameter the data provides strong evidence which nullifies the effect of the prior. This can also be seen by the lack of change in the posterior distribution of the individual parameters.

Next we look at the effect of the variance of the prior of β. Note that this does not signal the degree to which the individual coefficients are shrunk towards the β hyperparameter, just how densely the prior of the β parameter is distributed around zero. The prior variance seems to have an effect on the spread of the posterior distribution, though not necessarily proportional to the change in the prior. This is shown in figure A4. We see that the posterior distribution is concentrated more densely around the mean as the variance of the prior gets smaller. This effect is more visible for the mean β hyperparameter than the individual coefficients and again not close to proportional to the change in the prior. This seems to suggest that the spread of the posterior distribution is both influenced by the prior and the data that we use to update it.

The hyperparameter that does influence to what degree the individual coefficients are shrunk towards each other is the variance σ^2 in the higher level of the hierarchical model. We defined a truncated normal prior for this hyperparameter, and in this case we are interested in finding out to what extent changing the variance of this distribution affects the posterior distribution. As stated before, the variance for the coefficient of $CopingCapacity$ is not sampled in the same way as the others but its prior changes by the same factor as the others to ensure the effect on the variance is the same. As we can clearly see in figure A5, changing the prior has a huge effect on the spread of the posterior distribution. Even though we hoped that this prior would be reasonable uninformative, it appears that it exerts more influence on the posterior distribution than the data. We saw for the β hyperparameters as well that the posterior of higher-level parameters in the hierarchical model is less influenced by the data and more by the priors so this might be the reason for the results we see here.

Next we examine the sensitivity to the priors of the parameters in the missingness model. As we can clearly see in figure A6, the posterior hardly changes if at all. This means that the prior that we defined has a very small influence on the posterior which is mostly shaped by the data. As we wanted to define an uninformative prior for this parameter this is encouraging.

Next we look at the parameter $b.X$ of the value of the variable $FoodPrice$. The results are shown in figure A7. From the figure we see that changing the mean has a clear effect on the posterior distribution, changing the mean by as much, if not more, as the change in the prior. The data seems to skew the posterior distribution towards more negative values, an effect that is more pronounced when the prior has higher probability mass here. The idea of the prior was to shrink the parameter value towards zero. If the prior works well, changing the mean should mean we shrink the parameter value towards -1 or 1 respectively. By looking at the changes in the posterior this seems to be happening indeed.

When we change the variance of the prior of $b.X$ we change the level of shrinkage we impose on the parameter. The results when we do this can be seen in figure A8. The figure clearly indicates that not only the spread of the posterior increases as the prior variance decreases, also the shape is more flexible. This is because the prior imposes less shrinkage and the shape of the posterior is more freely influenced by the data. Again, the data seems to skew the posterior towards negative values when more spread assigns higher probability mass in the prior to these values. This again seems to support that this prior works the way we intended it to.

In order for the draws from the posterior distribution that we consider to be reliable we need to make sure the draws have converged to the conditional posterior distribution from the parameter of interest. As outlined before, we do this using the Gelman-Rubin statistic. Figure 4 displays the plots of the Gelman-Rubin statistic for a couple of parameters included in the model. Specifically, the figure first shows the hyperparameter $\beta_{Vulnerability}$ and below that from left to right the individual coefficients in the models for Accessibility, Inflation and Capacity. As the model includes a large number of parameters we do not display them all but the convergence of the other parameters follows the same pattern. The Gelman-Rubin was estimated using two chains. The green and the blue lines represent the pooled and the within variance respectively. The red line represents the parameter of interest that we described in the part about model convergence. Note that the pooled and within variance are normalized so that all quantities are roughly on the same scale which makes them easier to display together. Using the rule we established before, we can consider the chain to have converged when the statistic of interest is lower than 1.2. As the graphics show, the chains clearly converge well before the 5000th draw, which is the amount of draws we use as a burn in sample. This indicates that we can safely assume that all subsequent parameter draws that we consider come from the posterior conditional distribution of the parameter of interest.

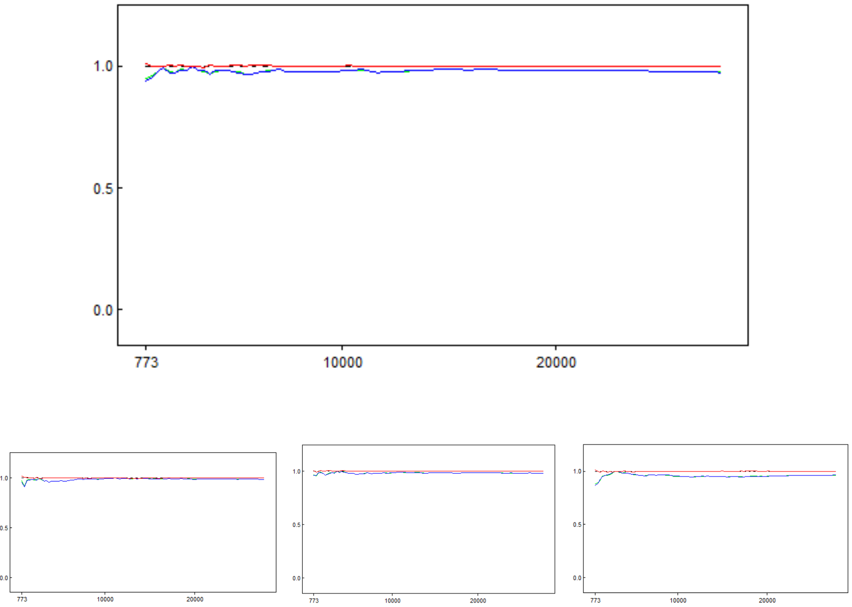

Figure 4: Convergence of $\beta_{Vulnerability}$ model 1

Table 2 shows the parameter estimates resulting from the first model. The table shows summary statistics from the posterior distributions of each of the coefficients in each of the three analysis models as well as the coefficients from the missingness model. The table first displays the mean, standard deviation and median of the draws from the posterior distributions. The column with header "MC error" displays the Monte Carlo sampling standard error for the mean of each coefficient that was briefly touched upon in the section about the Gibbs sampler. The column with header "percentile zero" displays what percentage of the draws from the posterior distribution are larger than zero. If the value in this column is 50 that means excactly the same amount of positive and negative values were drawn. 0 means only negative values were drawn and 100 that only positive values were drawn.

	Variable	Mean	Standard deviation	Median	MC error	Percentile zero
Accessibility	α	-1.47	0.61	-1.4	0.005	0.42
	β_{Hazard}	0.45	0.55	0.43	0.006	79.79
	$\beta_{Vulnerability}$	0.07	0.61	0.10	0.006	56.5
	$\beta_{CopingCapacity}$	0.68	0.37	0.67	0.007	97.04
	$\beta_{Corruption}$	0.50	0.60	0.49	0.007	80.33
	$\beta_{BusinessEase}$	-0.21	0.56	-0.22	0.007	34.92
	$\beta_{DietaryEnergy}$	-0.03	-0.03	-0.03	0.01	48.77
	$\beta_{FoodPrice}$	0.0015	0.64	-0.01	0.01	49.22
Inflation	α	0.25	0.56	0.25	0.003	67.39
	β_{Hazard}	0.51	0.51	0.50	0.006	84.23
	$\beta_{Vulnerability}$	0.62	0.58	0.60	0.005	86.01
	$\beta_{CopingCapacity}$	0.78	0.40	0.76	0.008	98.11
	$\beta_{Corruption}$	0.67	0.61	0.65	0.008	86.81
	$\beta_{BusinessEase}$	-0.46	0.56	-0.45	0.007	20.13
	$\beta_{DietaryEnergy}$	-0.12	0.90	-0.11	0.01	45.21
	$\beta_{FoodPrice}$	0.02	0.69	0.01	0.01	49.43
Capacity	α	0.66	0.62	0.65	0.004	85.92
	β_{Hazard}	0.45	0.53	0.44	0.006	80.32
	$\beta_{Vulnerability}$	0.88	0.66	0.84	0.006	91.84
	$\beta_{CopingCapacity}$	0.60	0.40	0.60	0.008	94
	$\beta_{Corruption}$	0.07	0.64	0.09	0.008	56.08
	$\beta_{BusinessEase}$	-0.69	0.60	-0.67	0.008	11.72
	$\beta_{DietaryEnergy}$	-0.13	0.90	-0.12	0.01	44.68
	$\beta_{FoodPrice}$	0.009	0.69	-0.01	0.01	49.19
Missingness	$a.X$	-1.63	0.53	-1.56	0.007	0
	$b.X$	-0.22	0.79	-0.16	0.016	42.66

Table 2: Parameter estimates from model 1

The column 'percentile zero' is supposed to give an indication of how reliable any estimated effect is. In a frequentist setting this is done using p-values, with a significance level of 5% usually employed in the literature. In the column in table 2 this means that we look for values both close to zero and one, where values close to zero indicate a likely negative effect and values close to one indicate a likely positive effect. The closer the value in this column is to 50, the more equal the share of positive and negative draws and the less likely it is that the parameter has an effect either way.

Keeping the cutoff point of 5% in mind we see that only very few parameters contain meaningful information. However, due to the fact that we had very few observations available and specified a prior in order to shrink the parameters towards zero, we will discuss all parameter values while still taking their reliability into account. The posterior distributions of the parameters $\beta_{DietaryEnergy}$ and $\beta_{FoodPrice}$ are very close to 50%. This means that we observed roughly the same number of positive and negative draws for this parameter which is an indication that there might not be any effect present. Especially the fact that $FoodPrice$ does not help to model Inflation is a surprise. A possible explanation is that food is such a necessity that rising prices would not stop people from accessing markets to purchase it. The variable $BusinessEase$ seems to have a slight negative effect on market functioning. When compared with the aforementioned variables the posterior distribution of $\beta_{BusinessEase}$ is clearly less balanced. Nevertheless, question marks remain over its reliability, especially when modelling Accessibility. A clear positive effect is observed for the variables $CopingCapacity$ and to a lesser extent $Hazard$. For the other two variables the effect is markedly different in the different models for Accessibility, Capacity and Inflation. $Vulnerability$ has hardly any effect on Accessibility but seems to have a positive effect on Inflation and Capacity. $Corruption$ has a positive effect on Accessibility and Inflation but no meaningful effect on Capacity. The fact that these effects are so different can be interpreted as support for splitting the dependent variable in three parts, as there are clear differences in the way the explanatory variables contribute to each of the three. When we look at the estimates of the missingness model it appears that small values have a slightly larger probability to go missing, although this effect is very unreliable. For the intercept in the missingness model $a.X$ we only sample negative values. The intercept can be interpreted as the log odds ratio when the value of $FoodPrice$ is zero. The odds ratio in this case is the probability of a value to be missing divided by the probability that it is observed. The log odds ratio is therefore positive only if the probability of an observation to be missing is larger than the probability that it is observed. As missing observations are very rare for this variable it makes sense that we only sample negative values.

Because the aim of the model is to predict market functioning after a disaster we will look at how well the model predicts. In order for the model to predict each of the observed scores we first took the score we wanted to predict out of the dataset. This was done in order not to bias the predictions because the score we want to predict is used to train the model. For every draw from the posterior distribution of the parameters we also draw a value for the score that we want to predict, either a zero or a one. After observing 50.000 draws from the posterior distribution of Y we calculate the percentage of predicted ones and zeros, and use the percentage of predicted ones as the predicted probability that the respective market is functioning according to the model. We assess these predicted probabilities using the Mean Absolute Error (MAE). The MAE is defined as the mean of the absolute differences between the estimated probabilities and the observed outcomes. To estimate how good the model is predicting we compare it to a simple benchmark model for which we use the majority class. This means that in all cases we predict the most frequent outcome. As the dataset is perfectly balanced with ones and zeros (of the observations for which we could derive such a score) the majority class algorithm is excactly right half the time and 100% wrong the other half, leading to a MAE of 0.5. The MAE of model 1 is 0.43. Although this is better than the benchmark algorithm it is still somewhat high considering we hope to use the model to inform humanitarian cash programs. Table 3 shows the confusion matrix of the model where we say that we predicted that a market was still functioning if more than 50% of the draws from the posterior distribution said so and that we predicted a market was not functioning otherwise. We can see that the performance of the model does not depend much on whether a market is actually functioning or not. That is, it has about the same reliability when predicting ones or zeros.

		Predicted market functioning	
		Yes	No
Actual market functioning	Yes	21	13
	No	11	23

Table 3: Confusion matrix model 1

In order to improve the predictions of the model, we see what happens if we take some variables out. As table 2 suggests that some of the variables we have included have no meaningful effect on market functioning the out of sample predictions might actually improve if we take them out. This is because any effect they estimate might be due to overfitting, or the estimating of effects in a sample used to train a model that does not generalize well to the wider population. We begin by taking out the variable $FoodPrice$ as table 2 shows that its coefficient is closest to zero and has nearly the same amount of positive and negative draws in each of the three models. This change did slightly improve the model. Compared to a MAE of 0.43 for the full model, the reduced one achieves a MAE of 0.41. Table 4 shows the confusion matrix for the reduced model. As we can see, the reduced model predicts two cases correctly that the full model got wrong, further outlining how the reduced model is a minor improvement.

		Predicted market functioning	
		Yes	No
Actual market functioning	Yes	23	11
	No	11	23

Table 4: Confusion matrix model 1

Taking $DietaryEnergy$ out of both models causes negligible improvement in MAE and no changes in the confusion table. In a final effort to improve predictions we eliminated the variable $Vulnerability$ from the first model and $Corruption$ from the final one, which decreased the MAE to 0.4 and again did not alter the confusion table.

Model 2

The difference between model 2 and model 1 is that we no longer explicitly model the missingness of $Foodprice$. Instead, we just define a normal prior for it with parameters $Mu.X$ and $Sigma.X$. Other than that the model and priors are defined in the same way.

We begin by discussing the priors. The first thing to note is that as this model is to a large degree similar to the previous one, so there will be some similarities in the way the prior influences the respective posterior distributions. To keep this report concise, we will not discuss sensitivity of all the priors as comprehensively as we did for the previous model. We will focus on the parameters with different prior sensitivities compared to the previous model and where this is not the case this will be stated. Note that we do not examine whether the posterior distributions look the same, just whether they are influenced in the same way by changes in the prior.

There is not much change in the way the posteriors of the α parameters are influenced by the choice of prior and also the mean of the β hyperparameters is not much different. There is some difference in the change in the posteriors when we change the variance of the β hyperparameters. In the previous model we commented that these changes caused some change in the posterior, though nowhere near proportional to the change in the prior. When we change the variance of the prior in this model, we see more change in the posterior. This is demonstrated in figure A9.

Another difference between this and the previous model is the sensitivity to the prior of σ^2, the degree to which the individual coefficients are shrunk towards their common mean. Before we saw that any change in this prior massively altered the posterior. Though the change is still clear to see in this model, it is not as pronounced as it was before. To illustrate this figure A10 shows σ^2 for the variables $DietaryEnergy$ and $FoodPrice$.

Convergence is not an issue for this model either as Gelman-Rubin statistics have converged to one well before the 5000th draw that we use as burn-in sample. This means that we can be assured that the draws we use for the statistics of interest are taken from the conditional posterior distribution of the parameters. The parameter estimates are largely similar to those of model 1. The biggest differences are in the coefficients of $DietaryEnergy$ and $FoodPrice$. That the coefficients of $FoodPrice$ are slightly different makes sense as the only change we made was in the modelling of the missingness of this variable. The new estimates are shown in table 5. As we can see the changes are not very large and in any case the coefficients are still very evenly distributed around zero. The full table with coefficients can be found in the appendix.

	Variable	Mean	Standard deviation	Median	MC error	Percentile zero
Accessibility	$\beta_{DietaryEnergy}$	0.007	1.09	0.001	0.01	50.05
	$\beta_{FoodPrice}$	0.03	0.76	0.02	0.01	50.83
Inflation	$\beta_{DietaryEnergy}$	-0.17	1.09	-0.16	0.01	43.75
	$\beta_{FoodPrice}$	0.12	0.73	0.11	0.009	55.97
Capacity	$\beta_{DietaryEnergy}$	-0.17	1.09	-0.15	0.01	43.97
	$\beta_{FoodPrice}$	0.12	0.75	0.10	0.009	55.41

Table 5: Parameter estimates from model 2

In terms of prediction we again calculated the MAE to assess how well the model performed. It turns out this model predicts marginally better with a MAE of 42% but the difference is not very substantial. Moreover, the confusion matrix looks exactly the same as it did for model 1. This is because for all observations model 2 predicts the same outcome as model 1. All in all we can conclude that in terms of both model inference and prediction the differences between model 1 and 2 are negligible and we do not loose any meaningful information by only using one of the two.

Model 3

The third and final model is very similar to model 2 that we discussed before. This time we use a different prior model for the missing variable $FoodPrice$, namely a logit model using $Hazard$,

$Vulnerability$ and $CopingCapacity$ as explanatory variables. The parameters γ_1, γ_2 and γ_3 are their respective coefficients in the logit model.

As we saw that model 1 and model 2 were very similar in almost every respect we will again not go too deep into all aspects of model 3 but rather we focus on the differences that stand out with the other two models. As before, first we examined the sensitivity to the priors using the same specifications that were described before. It turns out that the differences in prior sensitivity between this model and the previous two were very minor, if any, and therefore not worth discussing here. Also in this model parameter convergence was not a problem and the Gelman-Rubin statistics showed that the parameters have converged well before the 5000 draws that we use as a burn-in sample.

When we look at the actual parameter estimates we see some differences between model 3 and the others. Because the main differences appear in the model for Capacity descriptive statistics of the parameters of this model are shown in table 6. The parameters of the other models behave similar to the ones shown for model 1 unless mentioned otherwise. The full table with all parameter estimates can be found in the appendix. The thing that perhaps stands out most is the fact that all standard deviations are inflated compared to model 1. This holds for almost all parameters, not just the ones shown in table 6. With regard to the effect of the variables, we saw in model 1 that the posterior distributions of the coefficients of $DietaryEnergy$ and $FoodPrice$ were very much centered around zero. In table 6 we see that in this model this is less obvious and it appears the variables might have some effect on Capacity. In earlier models we saw too that $Hazard$, $Vulnerability$ and $CopingCapacity$ had pronounced positive effects on all three counts of market functioning. In this model for Capacity we see that only $Vulnerability$ has a clear positive effect. The parameter of $Hazard$ and $CopingCapacity$ have apparently been shrunk towards zero because these variables were also used in the prior for $FoodPrice$. The effects of $Corruption$ and $BusinessEase$ are very similar to the ones estimated in model 1 and 2.

	Variable	Mean	Standard deviation	Median	MC error	Percentile zero
Capacity	α	0.88	0.77	0.86	0.005	87.92
	β_{Hazard}	0.09	0.86	0.10	0.008	55.04
	$\beta_{Vulnerability}$	1.76	1.12	1.64	0.012	96.42
	$\beta_{CopingCapacity}$	-0.06	0.73	-0.05	0.007	47.04
	$\beta_{Corruption}$	-0.07	1.116	-0.02	0.010	49.32
	$\beta_{BusinessEase}$	-1.52	1.12	-1.43	0.012	7.05
	$\beta_{DietaryEnergy}$	-0.86	2.51	-0.66	0.025	37.18
	$\beta_{FoodPrice}$	0.60	1.06	0.52	0.011	71.35

Table 6: Parameter estimates from model 3

With regards to prediction, model 3, as was the case with model 2, performed very similar to model 1. It has a marginally worse MAE at little over 0.43 and the confusion matrix is again excactly the same as the one showed in table 3. Again it seems fair to conclude that there are is little to no substantial difference with model 1. Therefore we will mainly refer to model 1 in our conclusions as the other two models that we considered did not appear to change the model output significantly with respect to inference or prediction.

In a final effort to increase the predictions we employ ensemble learning to combine the predictions of the different models into a final prediction. Ensemble learning refers to the concept where multiple weak learners or models are combined into one stronger learner. There are some important caveats about the use of ensemble learning that we need to discuss first. Ensemble learning is based on the idea that different relatively weak learners each predict certain features relatively well and others less so. By combining these learners the overall prediction will be much more accurate than those from any individual model. In other words, the success of ensemble learning relies on exploiting the different features of each model (Kuncheva and Whitaker 2003). As we have seen in the discussion of the three models we have treated so far, the differences were relatively few and far between. We do not know how similar the three models are that we did not discuss, but it would likely improve the quality of the predictions of our ensemble model if they are more different than the three models we have discussed so far. In general, when using ensemble learning it is a good idea to use fundamentally different models and combine them (Adeva, Beresi, and Calvo 2005). In this research we did not consider other types of models because they are not as adept at dealing with small data sets and high frequency of missing data as the models we used here. Another issue when using ensemble learning is that to avoid overfitting it is better to fit the ensemble model on a different dataset than was used to train

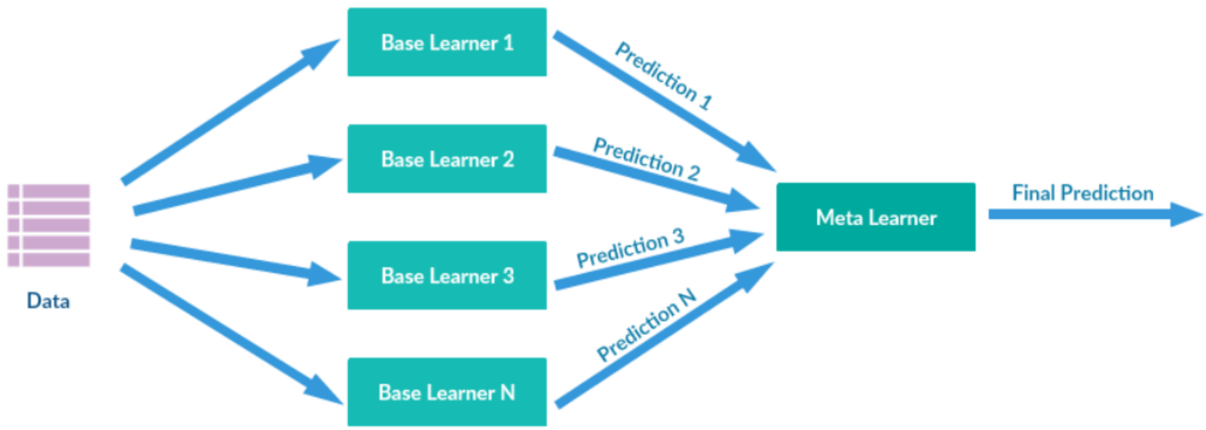

Figure 5: Graphical representation of stacking

the separate models that comprise the ensemble model. Because our dataset is already very small, we decided not to split it up any further and therefore used the same dataset to train the individual models and the ensemble model that combines them.

In this research we specifically used stacking to combine the predictions of the different models into a final prediction. Stacking is a relatively simple method that has been used extensively in the literature (see for example Deng and Platt 2014 and Wang et al. 2011). The idea of stacking is that of all the individual models we use only the final predictions. These predictions are used as the inputs to a new model and the output of this model is the final prediction. A graphical depiction of this idea can be found in figure 5. The base learners in our case are the models that were compared in table 1. Their predictions are used as input for a final model, called the 'meta learner' in the figure. As we predict a binary variable the meta learner will be a logit model.

		Missingness X		
		Non-ignorable	Ignorable-Logit prior	Ignorable-Normal prior
Missingness Y	Ignorable	3.36 (0.39)	0.94 (0.85)	2.85 (0.61)
	Non-ignorable	8.70 (0.39)	15.91 (0.19)	-29.2 (0.09)

Table 7: Deviance information criterion for various models

Table 7 shows the coefficients of the predictions of the separate models in the final logit model.

The p-value is shown between brackets. Surprisingly, the models where the missingness of Y was explicitly modelled have more significant coefficients in the final solution even though their DIC was markedly lower, though one of them has a negative coefficient. This is probably because the ensemble model does not add a penalty for complexity of the separate model whereas the DIC does. This give some hope that the models are sufficiently different for an ensemble model to improve the predictions. Unfortunately, the predictions of the ensemble model are only marginally better than the separate models. The MAE decreases by only 0.01 with respect to the earlier models and the confusion matrix is shown in table 8. As we can see the confusion matrix is worse than it was when we took some variables out of the model, further showing that the ensemble model is only a very slight improvement over the original models. A likely explanation for the lack of improvement is that, as posited before, the separate models that are combined in the ensemble model are too similar to yield substantially better predictions when added together.

		Predicted market functioning	
		Yes	No
Actual market functioning	Yes	22	12
	No	12	22

Table 8: Confusion matrix ensemble

5 Conclusion

In this paper we have studied the functioning of markets in areas where a natural disaster has struck. This information is important for the Red Cross to inform humanitarian cash transfer projects as these are not very useful if there are no functioning markets for people to spend cash. The models that we set up predicted better than random, although the difference was not very large. The predictions of the model slightly improved after simplifying the model somewhat. Some variables turned out to have a clear effect on market functioning even though the priors were designed to have a regularizing effect, which is a promising sign. More specifically, higher Hazard, Coping Capacity and to a lesser extent Vulnerability ratings clearly have a positive effect on all three counts of market functioning that we examined. A higher Hazard rating indicates that a country is more exposed to natural hazards. The results suggest that this leads markets to be better adapted to keep functioning after disasters. The Vulnerability and Coping Capacity ratings measure socio-economic and infrastructural resilience when a disaster strikes respectively. For these indices, it is harder to see why they positively impact market functioning after a disasters as a higher rating indicates less resilience. Also surprisingly, the effect of food prices on market functioning was minimal, as well as that of the average dietary energy adequacy. The corruption and ease of doing business indices have mixed effects for the different kinds of market functioning that we examined in this research, which supports our decision to split market functioning into three parts. Corruption seems to particularly hurt accessibility of markets and cause inflation but we found no effect on capacity worth mentioning. In countries where it's easier to do business see less inflation and capacity problems after a disaster, but we did not find as strong an effect on accessibility.

Limitations

With regards to discussion and recommendations for future research it would be remiss not to begin with the biggest issue of this research: the quality and availability of the data. Not only is at least one of the three dependent variables missing for a lot of observations, even when we do have one it is derived from a qualitative report rather than an objective measurement. Although on each occasion we have reason to believe that we have correctly observed whether a market was functioning, having an objective measure would definitely be an improvement when trying to model it. Moreover, on a lot of occasions we couldn't even find a qualitative indication of market functioning. Because the number of disasters in a specified area and timeframe is naturally limited, having a sparse dataset is an even bigger issue. In addition to these challenges in collecting the dependent variables, there were some issues with the explanatory variables as well. The research focuses on Ethiopia, Kenya and Malawi and as the literature review suggested, finding reliable data in these countries can be challenging. The few data sources that openly publish data that is relevant for our research often have poor temporal and spatial resolution. Indeed, all bar one of the explanatory variables in this research are only measured yearly and on national level, and that one variable turned out to have no meaningful effect on market functioning. Because disaster relief is a very time-sensitive business, variables measured on a yearly basis can only get you so far in predicting something that we wish to be available as soon as a disaster has occurred. As we wish to predict the functioning of local markets, the spatial granularity of the data is also an issue. Variables that are measured on a national level tell us nothing about how different markets within the country react to a shock like a disaster, or even which ones have been affected by the shock. As a consequence of these issues, the result should probably be interpreted as more of a risk profile rather than an up-to-date picture of how well a market is functioning right after a disaster. Possible promising sources of data to improve this research include migration records from phone or social media usage, satellite imagery and social media data, but for none of these complete historical data is available at present.

One promising avenue to improve the results of this research lies in changing the priors. The role that priors play in Bayesian research has been extensively discussed before but since the topic of this research is rather new it is difficult to formulate informative choices for the priors. Because of this lack of information we formulated priors that were intended to have a regularizing effect on the parameters. This approach payed off as it helped us define some variables that do not seem to have any effect on market functioning, and when we examined the sensitivity of the models to the choice of prior they generally seemed to have the effect we intended them to. However, if more information were to become available in the future on market functioning after disasters it would be a good idea to incorporate this knowledge into more informative priors that shape the results in a more pronounced way.

The different models that were compared in this research corresponded to the different assumptions about the missingness of the data that we considered. Inevitably, these are only a limited number of all the possible assumptions that led to the missingness of the data. In general, we tried to keep all missingness models and priors for missing data as simple as possible. We did this because the Deviance Information Criterion punishes the added model complexity which puts the models at a disadvantages when we have no indication that more complicated models would achieve a better fit. Despite this, it is very well possible that there are more complicated missingness models and/or priors that better fit the data that we have not considered. We hope that as more data becomes available the role of missingness models in future research diminishes but when richer data sources become available this can also lead to a better understanding of why and how data goes missing. In this case, better modelling the missingness of the data can improve the results of the models. Aside from the modelling of the missing data we only considered one analysis model in this research. The resulting models were too similar to be improved by combining them into one superior predictor. Again, the choice of model in this research was primarily guided by the availability and quality of the data but if this problem diminishes in the future this would allow others to design other, more effective models to model market functioning. As these models become more diverse the literature suggests that they can be improved by combining them into one model that can be assumed to predict better than the first attempt we have drafted up in this research.

References

Abdalla, Rifaat (2016). “Evaluation of spatial analysis application for urban emergency management”. In: *SpringerPlus* 5.1, p. 2081.

Adeva, Juan José Garcıa, U Beresi, and R Calvo (2005). “Accuracy and diversity in ensembles of text categorisers”. In: *CLEI Electronic Journal* 9.1, pp. 1–12.

Akaike, Hirotugu (1974). “A new look at the statistical model identification”. In: *Selected Papers of Hirotugu Akaike*. Springer, pp. 215–222.

Albala-Bertrand, Jose-Miguel (1993). “Natural disaster situations and growth: A macroeconomic model for sudden disaster impacts”. In: *World Development* 21.9, pp. 1417–1434.

Albala-Bertrand, Jose-Miguel et al. (1993). “Political economy of large natural disasters: with special reference to developing countries”. In: *OUP Catalogue*.

Albu, Mike (2010). *Emergency market mapping and analysis toolkit*. Oxfam GB.

Alegana, Victor A et al. (2018). “National and sub-national variation in patterns of febrile case management in sub-Saharan Africa”. In: *Nature communications* 9.1, p. 4994.

Allaire, Maura (2018). “Socio-economic impacts of flooding: A review of the empirical literature”. In: *Water Security* 3, pp. 18–26.

Anttila-Hughes, Jesse and Solomon Hsiang (2013). “Destruction, disinvestment, and death: Economic and human losses following environmental disaster”. In: *Available at SSRN 2220501*.

Arslan, Muhammad et al. (2017). “A Review on Applications of Big Data for Disaster Management”. In: *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, pp. 370–375.

Asfaw, Solomon, Benjamin Davis, and Joshua Dewbre (2011). “Cash transfer programs in sub-Saharan Africa: measuring the impact on climate change adaptation”. In: *4th meeting of the Wye City Group on Statistics on Rural Development and Agriculture Household Income, Rio de Janeiro, Brazil*, pp. 9–11.

Bailey, Sarah and Sophie Pongracz (2015). “Humanitarian Cash Transfers–Cost, Value for money and economic impact”. In: *Background note for the High Level Panel on Humanitarian Cash Transfers (https://www. odi. org/sites/odi. org. uk/files/odi-assets/publications-opinionfiles/9731. pdf)*.

Barbier, Edward B (2008). “In the wake of tsunami: Lessons learned from the household decision to replant mangroves in Thailand”. In: *Resource and Energy Economics* 30.2, pp. 229–249.

Barder, Owen et al. (2015). "Doing cash differently: How cash transfers can transform humanitarian aid". In: *Report of the High Level Panel on Humanitarian Cash Transfers, Overseas Development Institute, London, September.*

Barrett, Christopher B et al. (2009). "Market information and food insecurity response analysis". In: *Food Security* 1.2, pp. 151–168.

Barrientos, Armando and Jocelyn DeJong (2006). "Reducing child poverty with cash transfers: A sure thing?" In: *Development policy review* 24.5, pp. 537–552.

Bastagli, Francesca et al. (2016). "Cash transfers: what does the evidence say? A rigorous review of programme impact and of the role of design and implementation features". In: *London: ODI.*

Benson, Charlotte and Edward Clay (2004). *Understanding the economic and financial impacts of natural disasters.* The World Bank.

Besley, Timothy and Robin Burgess (2002). "The political economy of government responsiveness: Theory and evidence from India". In: *The Quarterly Journal of Economics* 117.4, pp. 1415–1451.

Brooks, Stephen P and Andrew Gelman (1998). "General methods for monitoring convergence of iterative simulations". In: *Journal of computational and graphical statistics* 7.4, pp. 434–455.

Browne, William J, David Draper, et al. (2006). "A comparison of Bayesian and likelihood-based methods for fitting multilevel models". In: *Bayesian analysis* 1.3, pp. 473–514.

Carrara, Alberto and Fausto Guzzetti (2013). *Geographical information systems in assessing natural hazards.* Vol. 5. Springer Science & Business Media.

Carter, Michael R and Christopher B Barrett (2006). "The economics of poverty traps and persistent poverty: An asset-based approach". In: *The Journal of Development Studies* 42.2, pp. 178–199.

Cavallo, Eduardo A and Ilan Noy (2009). "The economics of natural disasters: a survey". In: *IDB working paper series.*

Cavallo, Eduardo, Ilan Noy, et al. (2011). "Natural disasters and the economy—a survey". In: *International Review of Environmental and Resource Economics* 5.1, pp. 63–102.

Cen, Jianhu et al. (2011). "Developing a disaster surveillance system based on wireless sensor network and cloud platform". In: *IET International Conference on Communication Technology and Application*.

Choi, Seonhwa and Byunggul Bae (2015). "The real-time monitoring system of social big data for disaster management". In: *Computer science and its applications*. Springer, pp. 809–815.

Cutter, Susan L (2003). "GI science, disasters, and emergency management". In: *Transactions in GIS* 7.4, pp. 439–446.

Dacy, Douglas C and Howard Kunreuther (1969). "economics of natural disasters; implications for Federal policy". In: *The National Agricultural Library*.

Darcy, James and C Hofmann (2003). "According to need?: needs assessment and decision-making in the humanitarian sector". In: *The National Agricultural Library*.

Davis, Benjamin et al. (2016). *From evidence to action: the story of cash transfers and impact evaluation in sub Saharan Africa*. Oxford University Press.

De Albuquerque, Joao Porto et al. (2015). "A geographic approach for combining social media and authorative data towards identifying useful information for disaster management". In: *international journal of geographical information* 29.4, pp. 667–689.

De Groeve, Tom, K Poljansek, and L Vernaccini (2014). "Index for Risk Management–InfoRM". In: *Concept and Methodology Version*.

De Haen, Hartwig and Günter Hemrich (2007). "The economics of natural disasters: implications and challenges for food security". In: *Agricultural economics* 37, pp. 31–45.

De Janvry, Alain, Marcel Fafchamps, and Elisabeth Sadoulet (1991). *Peasant Household Behavior with Missing Markets: Some Paradoxes Explain*. Tech. rep.

De Silva, M.M.G.T. and Akiyuki Kawasaki (2018). "Socioeconomic vulnerability to disaster risk: a case study of flood and drought impact in a rural Sri Lankan community". In: *Ecological Economics* 152, pp. 131–140.

Deng, Li and John C Platt (2014). "Ensemble deep learning for speech recognition". In: *Fifteenth Annual Conference of the International Speech Communication Association*.

Depaoli, Sarah (2013). "Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation." In: *Psychological methods* 18.2, p. 186.

Dercon, Stefan (2004). "Growth and shocks: evidence from rural Ethiopia". In: *Journal of Development Economics* 74.2, pp. 309–329.

Dienes, Zoltan (2014). “Using Bayes to get the most out of non-significant results”. In: *Frontiers in psychology* 5, p. 781.

Dilley, Maxx et al. (2005). *Natural disaster hotspots: a global risk analysis*. The World Bank.

Eisensee, Thomas and David Strömberg (2007). “News floods, news droughts, and US disaster relief”. In: *Quarterly Journal of Economics* 122.2.

Evans, David K and Anna Popova (2014). *Cash transfers and temptation goods: a review of global evidence*. The World Bank.

Fothergill, Alice and Lori A Peek (2004). “Poverty and disasters in the United States: A review of recent sociological findings”. In: *Natural hazards* 32.1, pp. 89–110.

Gelman, Andrew (2006). “Multilevel (hierarchical) modeling: what it can and cannot do”. In: *Technometrics* 48.3, pp. 432–435.

Gelman, Andrew et al. (2006). “Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)”. In: *Bayesian analysis* 1.3, pp. 515–534.

Gelman, Andrew, Donald B Rubin, et al. (1992). “Inference from iterative simulation using multiple sequences”. In: *Statistical science* 7.4, pp. 457–472.

Geman, Stuart and Donald Geman (1987). “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”. In: *Readings in computer vision*. Elsevier, pp. 564–584.

Good, Charles M (1973). “Markets in Africa: A Review of Research Themes and the Question of Market Origins (Marchés africains: recension des thèmes de recherche et de la question de l’origine des marchés)”. In: *Cahiers d’études africaines*, pp. 769–780.

Goswami, Saptarsi (2018). “A review on application of data mining techniques to combat natural disasters”. In: *Ain shams engineering journal* 9.3, pp. 365–378.

Graham, Melissa W, Elizabeth J Avery, and Sejin Park (2015). “The role of social media in local government crisis communications”. In: *Public Relations Review* 41.3, pp. 386–394.

Grolinger, Katarina et al. (2013). “Knowledge as a service framework for disaster data management”. In: *2013 Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*. IEEE, pp. 313–318.

Guha-Sapir, Debarati et al. (2011). “Annual disaster statistical review 2010”. In: *Centre for Research on the Epidemiology of Disasters*.

Gurman, Tilly A and Nicole Ellenberger (2015). "Reaching the global community during disasters: findings from a content analysis of the organizational use of Twitter after the 2010 Haiti earthquake". In: *Journal of health communication* 20.6, pp. 687–696.

Hallegatte, Stéphane, Jean-Charles Hourcade, and Patrice Dumas (2007). "Why economic dynamics matter in assessing climate change damages: illustration on extreme events". In: *Ecological economics* 62.2, pp. 330–340.

Hedlund, Kerren et al. (2013). "Final evaluation of the unconditional cash and voucher response to the 2011–12 crisis in southern and central Somalia". In: *UNICEF, New York*.

Henry, M., I. Takigawa, and K. Meguro (2017). "The impact of income disparity on vulnerability and information collection: an analysis of the 2011 Thai flood". In: *Journal of flood risk management* 10, pp. 338–348.

Heudtlass, Peter, Debarati Guha-Sapir, and Niko Speybroeck (2018). "A Bayesian hierarchical model for mortality data from cluster-sampling household surveys in humanitarian crises". In: *International journal of epidemiology* 47.4, pp. 1255–1263.

Hoddinott, John (2006). "Shocks and their consequences across and within households in rural Zimbabwe". In: *The journal of development studies* 42, pp. 301–321.

Horita, Flávio EA et al. (2017). "Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil". In: *Decision Support Systems* 97, pp. 12–22.

Horwich, George (2000). "Economic lessons of the Kobe earthquake". In: *Economic development and cultural change* 48.3, pp. 521–542.

Ikiara, Gerrishon K (2009). "Political economy of cash transfers in Kenya". In: *Overseas Development Institute*.

Janis, Irving L and Leon Mann (1977). "Emergency decision making: a theoretical analysis of responses to disaster warnings". In: *Journal of human stress* 3.2, pp. 35–48.

Kates, Robert W (1971). "Natural hazard in human ecological perspective: hypotheses and models". In: *Economic Geography* 47.3, pp. 438–451.

Kellenberg, Derek and A Mushfiq Mobarak (2011). "The economics of natural disasters". In: *Annual Reviews of Resource Economics* 3.1, pp. 297–312.

Kim, Su-Young et al. (2013). “Single and multiple ability estimation in the SEM framework: A noninformative Bayesian estimation approach”. In: *Multivariate behavioral research* 48.4, pp. 563–591.

Kuncheva, Ludmila I and Christopher J Whitaker (2003). “Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy”. In: *Machine learning* 51.2, pp. 181–207.

Kunreuther, Howard and Erwann Michel-Kerjan (2008). “A framework for reducing vulnerability to natural disasters: Ex ante and ex post Considerations”. In: *Commissioned by the Joint World Bank-UN project on the Economics of Disaster Risk Reduction. Wharton Risk Management and Decision Processes Center. The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, USA*.

Lemoine, Nathan P (2019). “Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses”. In: *Oikos*.

Little, Roderick JA and Donald B Rubin (2019). *Statistical analysis with missing data*. Vol. 793. Wiley.

Liu, Nan (2014). “Yong Ye. Humanitarian logistics planning for natural disaster response with Bayesian information updates”. In: *Journal of Industrial & Management Optimization* 10.3, pp. 665–689.

Lloyd-Sherlock, Peter (2006). “Simple transfers, complex outcomes: The impacts of pensions on poor households in Brazil”. In: *Development and Change* 37.5, pp. 969–995.

Ludwig, Fulco et al. (2007). *Climate change impacts on developing countries-EU accountability*. Tech. rep. European Parliament.

MacLehose, Richard F et al. (2007). “Bayesian methods for highly correlated exposure data”. In: *Epidemiology* 18.2, pp. 199–207.

Margolies, Amy and John Hoddinott (2015). “Costing alternative transfer modalities”. In: *Journal of Development Effectiveness* 7.1, pp. 1–16.

Maxwell, Daniel and Merry Fitzpatrick (2012). “The 2011 Somalia famine: Context, causes, and complications”. In: *Global Food Security* 1.1, pp. 5–12.

McElreath, Richard (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Mertens, Kewan et al. (2016). “The direct impact of landslides on household income in tropical regions: A case study from the Rwenzori Mountains in Uganda”. In: *Science of the Total Environment* 550, pp. 1032–1043.

Metropolis, Nicholas et al. (1953). “Equation of state calculations by fast computing machines”. In: *The journal of chemical physics* 21.6, pp. 1087–1092.

Miller, Fiona et al. (2010). “Resilience and vulnerability: complementary or conflicting concepts?” In: *Ecology and Society* 15.3, pp. 1–25.

Miller, J (2002). “Mozambique 1999–2000 Floods, Impact Evaluation: Resettlement Grant Activity”. In: *Maputo: United States Agency for International Development.*

Moore, Charity (2009). *Nicaragua’s Red de Protección Social: an exemplary but short-lived conditional cash transfer programme.* Tech. rep. Country Study.

Morris, Saul S et al. (2002). “Hurricane Mitch and the livelihoods of the rural poor in Honduras”. In: *World development* 30.1, pp. 49–60.

Noy, Ilan and Rio Yonson (2018). “Economic Vulnerability and Resilience to Natural hazards: A survey of concepts and measurements”. In: *Sustainability* 10.8, p. 2850.

Ofli, Ferda et al. (2016). “Combining human computing and machine learning to make sense of big (aerial) data for disaster response”. In: *Big data* 4.1, pp. 47–59.

O’Hagan, Anthony et al. (2006). *Uncertain judgements: eliciting experts’ probabilities.* John Wiley & Sons.

O’Keefe, Phil, Ken Westgate, and Ben Wisner (1976). “Taking the naturalness out of natural disasters”. In: *Nature* 260, pp. 566–567.

Okuyama, Yasuhide, Geoffrey JD Hewings, and Michael Sonis (2004). “Measuring economic impacts of disasters: interregional input-output analysis using sequential interindustry model”. In: *Modeling spatial and economic impacts of disasters.* Springer, pp. 77–101.

Olives, Casey and Marcello Pagano (2010). “Bayes-LQAS: classifying the prevalence of global acute malnutrition”. In: *Emerging themes in epidemiology* 7.1, p. 3.

Onyancha, Omwoyo Bosire (2016). “Open research data in Sub-Saharan Africa: a bibliometric study using the Data Citation Index”. In: *Publishing Research Quarterly* 32.3, pp. 227–246.

Osuteye, Emmanuel, Cassidy Johnson, and Donald Brown (2017). “The data gap: An analysis of data availability on disaster losses in sub-Saharan African cities”. In: *International journal of disaster risk reduction* 26, pp. 24–33.

Peacock, Walter Gillis, Hugh Gladwin, and Betty Hearn Morrow (2012). *Hurricane Andrew: Ethnicity, gender and the sociology of disasters*. Routledge.

Peduzzi, Pascal (2006). "The disaster risk index: overview of a quantitative approach". In: *Measuring vulnerability to natural hazards: towards disaster resilient societies*, pp. 171–181.

Ragini, J Rexiline, PM Rubesh Anand, and Vidhyacharan Bhaskar (2018). "Big data analytics for disaster response and recovery through sentiment analysis". In: *International Journal of Information Management* 42, pp. 13–24.

Rose, Adam (2007). "Economic resilience to natural and man-made disasters: Multidisciplinary origins and contextual dimensions". In: *Environmental Hazards* 7.4, pp. 383–398.

Rouder, Jeffrey N (2014). "Optional stopping: No problem for Bayesians". In: *Psychonomic bulletin & review* 21.2, pp. 301–308.

Schumacher, Ingmar and Eric Strobl (2011). "Economic development and losses due to natural disasters: The role of hazard exposure". In: *Ecological Economics* 72, pp. 97–105.

Schwarz, Gideon (1978). "Estimating the dimension of a model". In: *The annals of statistics* 6.2, pp. 461–464.

Selcuk, Faruk and Erinc Yeldan (2001). "On the macroeconomic impact of the August 1999 earthquake in Turkey: a first assessment". In: *Applied Economics Letters* 8.7, pp. 483–488.

Sen, Amartya (1982). *Poverty and famines: an essay on entitlement and deprivation*. Oxford university press.

Shaffer, Jeffrey G et al. (2018). "Development of a data collection and management system in West Africa: challenges and sustainability". In: *Infectious diseases of poverty* 7.1, p. 125.

Smith, Gabrielle and Lili Mohiddin (2015). "A review of evidence of humanitarian cash transfer programming in urban areas". In: *International Institute for Environment and Development (IIED), London http://pubs. iied. org/10759IIED.*

Song, Xuan et al. (2015). "A simulator of human emergency mobility following disasters: Knowledge transfer from big disaster data". In: *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Spiegelhalter, David J, Nicola G Best, et al. (2002). "Bayesian measures of model complexity and fit". In: *Journal of the royal statistical society: Series b (statistical methodology)* 64.4, pp. 583–639.

Spiegelhalter, David J, Andrew Thomas, et al. (1996). "BUGS: Bayesian inference using Gibbs sampling". In: *Version 0.5,(version ii) http://www. mrc-bsu. cam. ac. uk/bugs* 19.

Staunton, Cormac (2011). "Hard cash in hard times: A social accounting matrix multiplier analysis of cash transfers and food aid in rural Zimbabwe". In: *prepared for the International Conference ""Social Protection for Social Justice" Institute of Development Studies, UK*, pp. 13–15.

Strömberg, David (2007). "Natural disasters, economic development, and humanitarian aid". In: *Journal of Economic perspectives* 21.3, pp. 199–222.

Tapia, Andrea et al. (2011). "Seeking the trustworthy tweet: can microblogged data fit the information needs of disaster response and humanitarian relief organizations". In: *Proceedings of the 8th international ISCRAM conference*, pp. 1–10.

Taskin, Selda and Emmett J Lodree Jr (2011). "A Bayesian decision model with hurricane forecast updates for emergency supplies inventory management". In: *Journal of the Operational Research Society* 62.6, pp. 1098–1108.

Torkildson, Megan K, Kate Starbird, and Cecilia Aragon (2014). "Analysis and visualization of sentiment and emotion on crisis tweets". In: *International Conference on Cooperative Design, Visualization and Engineering*. Springer, pp. 64–67.

Tsai, Chung-Hung et al. (2008). "Application of geographic information system to the allocation of disaster shelters via fuzzy models". In: *Engineering Computations* 25.1, pp. 86–100.

Van den Berg, Marrit and Nguyen Viet Cuong (2011). "Impact of public and private cash transfers on poverty and inequality: evidence from Vietnam". In: *Development Policy Review* 29.6, pp. 689–728.

Vigdor, Jacob (2008). "The economic aftermath of Hurricane Katrina". In: *Journal of Economic Perspectives* 22.4, pp. 135–54.

Wang, Gang et al. (2011). "A comparative assessment of ensemble learning for credit scoring". In: *Expert systems with applications* 38.1, pp. 223–230.

Wisner, Benjamin et al. (2004). *At risk: natural hazards, people's vulnerability and disasters*. Psychology Press.

Yusoff, Aziyati et al. (2015). "Big data analytics for Flood Information Management in Kelantan, Malaysia". In: *2015 IEEE Student Conference on Research and Development (SCOReD)*. IEEE, pp. 311–316.

Zhang, Zhiyong et al. (2007). "Bayesian analysis of longitudinal data using growth curve models". In: *International Journal of Behavioral Development* 31.4, pp. 374–383.

Zielinski, Andrea et al. (2013). "Social media text mining and network analysis for decision support in natural crisis management." In: *ISCRAM*.

Zondervan-Zwijnenburg, Mariëlle et al. (2017). "Where do priors come from? Applying guidelines to construct informative priors in small sample research". In: *Research in Human Development* 14.4, pp. 305–320.

Zook, Matthew et al. (2010). "Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake". In: *World Medical & Health Policy* 2.2, pp. 7–33.

A6 Appendix

Model specification

In this section we will specify the exact setup of the models we have considered. As explained in the main text of this paper we considered six different models corresponding to six different assumptions about the missing data. Table A1 shows what name will refer to what model hereafter. Note that model 1, 2 and 3 refer to the same model as in the main text of this paper.

		Missingness X		
		Non-ignorable	Ignorable-Logit prior	Ignorable-normal prior
Missingness Y	Ignorable	Model 1	Model 2	Model 3
	Non-ignorable	Model 4	Model 5	Model 6

Table A1: Overview of different models

We start with the main part of the model which is the same for all different models. After that we denote for each model what additional part is added to deal with the missing data. In describing the models we use some notation to keep this part relatively concise. First the names of the explanatory variables are no longer written out. Rather, they are referred to as X_i with $i \in 1:7$. X_3 is the only variable with a slightly different specification in the analysis model. This refers to the variable *CopingCapacity*, as explained in the Methodology section. Variable X_7 refers to $FoodPrice$ which is the only explanatory variable with missing values. With regard to the distributions used to define priors, $N()$ will refer to the normal distribution and $IG()$ to the inverse gamma distribution. At times the function $logit()$ will be used. $logit(a)$ refers to $\frac{a}{1+a}$ and is written in this way to make the model specification easier to read. When the number inf is used as parameter of a distribution (typically the scale parameter) it refers to the number one million. When after a function we add $I(>0)$, this means that only values greater than zero were sampled from the specified distribution to avoid sampling negative variances.

$$Accessibility \sim Bernouilli(P_1)$$

$$logit(P_1) = \alpha_1 + \beta_{1.1} * X_1 + \beta_{2.1} * X_2 + \beta_{3.1} * X_3 + \beta_{4.1} * X_4 + \beta_{5.1} * X_5 + \beta_{6.1} * X_6 + \beta_{7.1} * X_7$$

$$Inflation \sim Bernouilli(P_2)$$

$$logit(P_2) = \alpha_2 + \beta_{2.1} * X_1 + \beta_{2.2} * X_2 + \beta_{2.3} * X_3 + \beta_{2.4} * X_4 + \beta_{2.5} * X_5 + \beta_{2.6} * X_6 + \beta_{2.7} * X_7$$

$$Capacity \sim Bernouilli(P_3)$$

$$logit(P_3) = \alpha_3 + \beta_{3.1} * X_1 + \beta_{3.2} * X_2 + \beta_{3.3} * X_3 + \beta_{3.4} * X_4 + \beta_{3.5} * X_5 + \beta_{3.6} * X_6 + \beta_{3.7} * X_7$$

Priors:

- $\alpha_1 \sim N(0, inf)$
- $\alpha_2 \sim N(0, inf)$
- $\alpha_3 \sim N(0, inf)$
- $\beta_{1.1} \sim N(\beta_1, \sigma_{b1}^2)$
- $\beta_{1.2} \sim N(\beta_1, \sigma_{b1}^2)$
- $\beta_{1.3} \sim N(\beta_1, \sigma_{b1}^2)$
- $\beta_1 \sim N(0, 0.5)$
- $\sigma_{b1}^2 \sim N(0, 100)I(>0)$
- $\beta_{2.1} \sim N(\beta_2, \sigma_{b2}^2)$
- $\beta_{2.2} \sim N(\beta_2, \sigma_{b2}^2)$
- $\beta_{2.3} \sim N(\beta_2, \sigma_{b2}^2)$
- $\beta_2 \sim N(0, 0.5)$
- $\sigma_{b2}^2 \sim N(0, 100)I(>0)$
- $\beta_{3.1} \sim N(\beta_3, \sigma_{b3}^2)$
- $\beta_{3.2} \sim N(\beta_3, \sigma_{b3}^2)$
- $\beta_{3.3} \sim N(\beta_3, \sigma_{b3}^2)$
- $\beta_3 \sim N(0, 5)$
- $\sigma_{b3}^2 \sim N(0, 10)I(>0)$

- $\beta_{4.1} \sim N(\beta_4, \sigma^2_{b4})$
- $\beta_{4.2} \sim N(\beta_4, \sigma^2_{b4})$
- $\beta_{4.3} \sim N(\beta_4, \sigma^2_{b4})$
- $\beta_4 \sim N(0, 0.5)$
- $\sigma^2_{b4} \sim N(0, 100)I(> 0)$
- $\beta_{5.1} \sim N(\beta_5, \sigma^2_{b5})$
- $\beta_{5.2} \sim N(\beta_5, \sigma^2_{b5})$
- $\beta_{5.3} \sim N(\beta_5, \sigma^2_{b5})$
- $\beta_5 \sim N(0, 0.5)$
- $\sigma^2_{b5} \sim N(0, 100)I(> 0)$
- $\beta_{6.1} \sim N(\beta_6, \sigma^2_{b6})$
- $\beta_{6.2} \sim N(\beta_6, \sigma^2_{b6})$
- $\beta_{6.3} \sim N(\beta_6, \sigma^2_{b6})$
- $\beta_6 \sim N(0, 0.5)$
- $\sigma^2_{b6} \sim N(0, 100)I(> 0)$
- $\beta_{7.1} \sim N(\beta_7, \sigma^2_{b7})$
- $\beta_{7.2} \sim N(\beta_7, \sigma^2_{b7})$
- $\beta_{7.3} \sim N(\beta_7, \sigma^2_{b7})$
- $\beta_7 \sim N(0, 0.5)$
- $\sigma^2_{b7} \sim N(0, 100)I(> 0)$
- $a \sim N(0, inf)$
- $b \sim N(0, 0.5$

- $X_7 \sim N(\mu_X, \sigma_X^2)$
- $\mu_X \sim N(0, 1)$
- $\sigma_X^2 \sim IG(1, 1)$

Model 1

$$Miss_X \sim Bernouilli(P_{miss})$$

$$logit(P_{miss}) = a + b * X_7$$

Priors:

- $a \sim N(0, inf)$
- $b \sim N(0, 0.5)$
- $X_7 \sim N(\mu_X, \sigma_X^2)$
- $\mu_X \sim N(0, 1)$
- $\sigma_X^2 \sim IG(1, 1)$

Model 2

$$X_7\ Bernouilli(P_X)$$

$$logit(P_X) = \gamma_1 * X1 + \gamma_2 * X2 + \gamma_3 * X3$$

priors:

- $\gamma_1 \sim N(0, 1)$
- $\gamma_2 \sim N(0, 1)$
- $\gamma_3 \sim N(0, 1)$

Model 3

priors:

- $X_7 \sim N(\mu_X, \sigma_X^2)$

- $mu_X \sim N(0,1)$
- $\sigma_X^2 \sim IG(1,1)$

Model 4

$$Miss_Y \sim Bernouilli(P_{MissY})$$

$$logit(P_{MissY}) = a_1 + b_1 * Accessibility + b_2 * Inflation + b_3 * Capacity$$

$$Miss_X \sim Bernouilli(P_{MissX})$$

$$logit(P_{MissX}) = a_2 + b_4 * X_7$$

priors:

- $a_1 \sim N(0, inf)$
- $b_1 \sim N(0, 0.5)$
- $b_2 \sim N(0, 0.5)$
- $b_3 \sim N(0, 0.5)$
- $a_2 \sim N(0, inf)$
- $b_4 \sim N(0, 0.5)$

Model 5

$$Miss_Y \sim Bernouilli(P_{MissY})$$

$$logit(P_{MissY}) = a_1 + b_1 * Accessibility + b_2 * Inflation + b_3 * Capacity$$

$$X_7\ Bernouilli(P_X)$$

$$logit(P_X) = \gamma_1 * X1 + \gamma_2 * X2 + \gamma_3 * X3$$

priors:

- $a_1 \sim N(0, inf)$
- $b_1 \sim N(0, 0.5)$

- $b_2 \sim N(0, 0.5)$
- $b_3 \sim N(0, 0.5)$
- $\gamma_1 \sim N(0, 1)$
- $\gamma_2 \sim N(0, 1)$
- $\gamma_3 \sim N(0, 1)$

Model 6

$$Miss_Y \sim Bernouilli(P_{MissY})$$

$$logit(P_{MissY}) = a_1 + b_1 * Accessibility + b_2 * Inflation + b_3 * Capacity$$

priors:

- $a_1 \sim N(0, inf)$
- $b_1 \sim N(0, 0.5)$
- $b_2 \sim N(0, 0.5)$
- $b_3 \sim N(0, 0.5)$
- $X_7 \sim N(\mu_X, \sigma_X^2)$
- $\mu_X \sim N(0, 1)$
- $\sigma_X^2 \sim IG(1, 1)$

Prior sensitivity plots

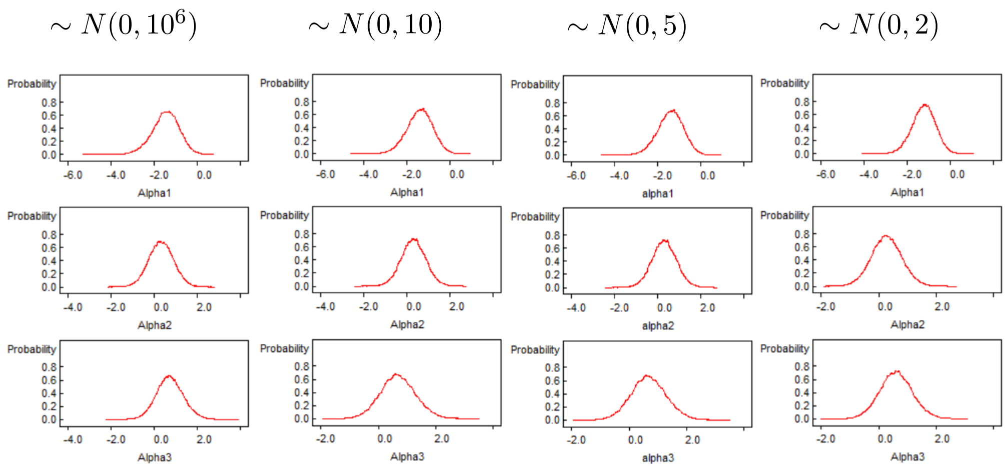

Figure A1: Prior sensitivity α model 1

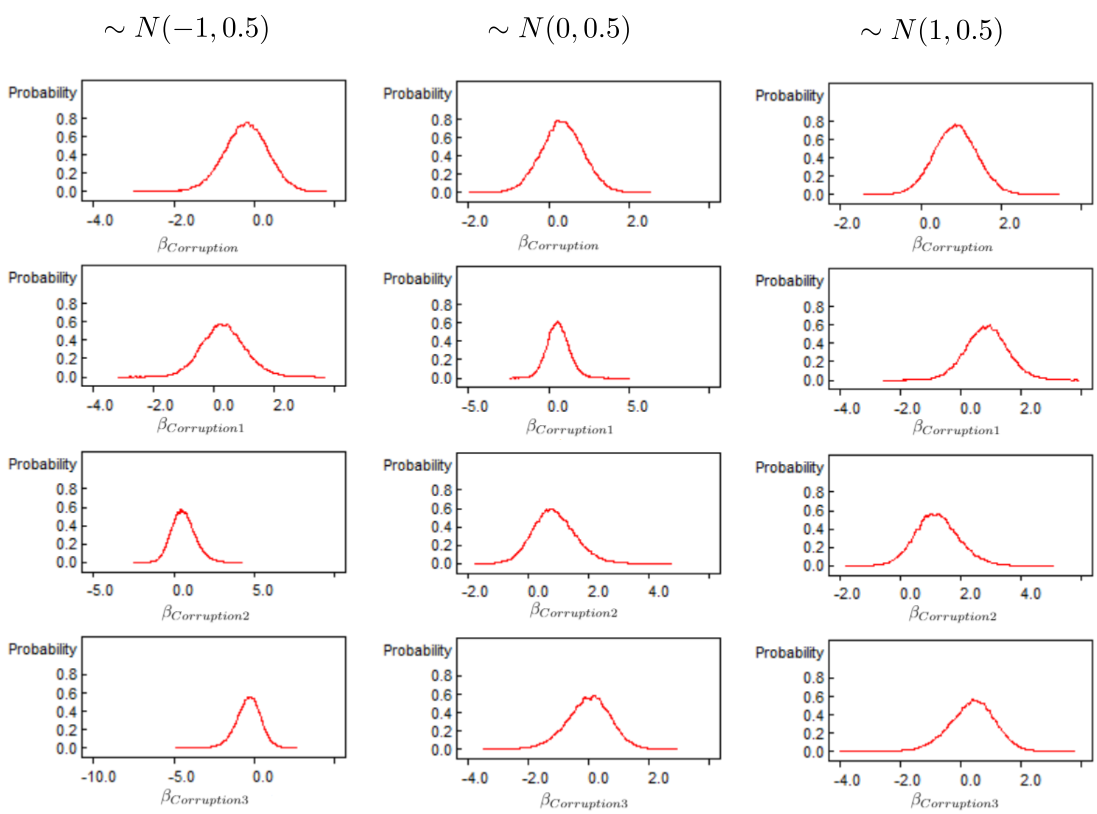

Figure A2: Prior sensitivity $\beta_{Corruption}$ model 1

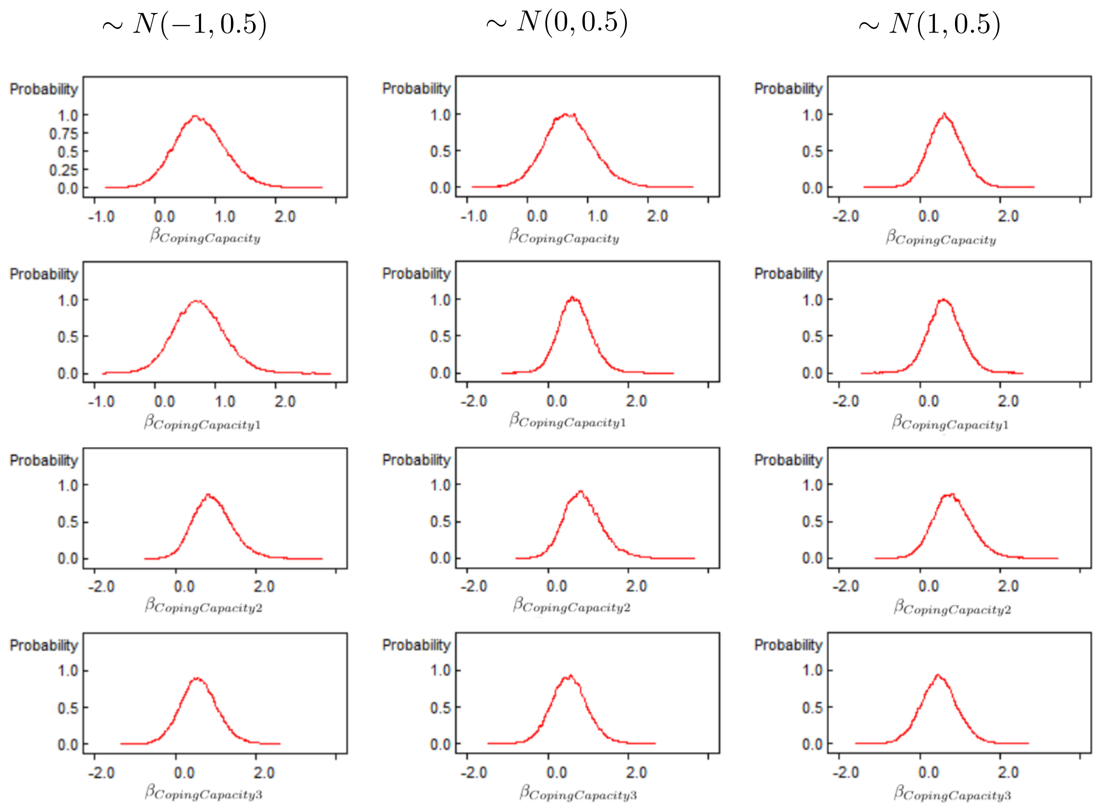

Figure A3: Prior sensitivity $\beta_{CopingCapacity}$ model 1

$\sim N(0, 10)$ $\sim N(0, 0.5)$ $\sim N(0, 0.2)$

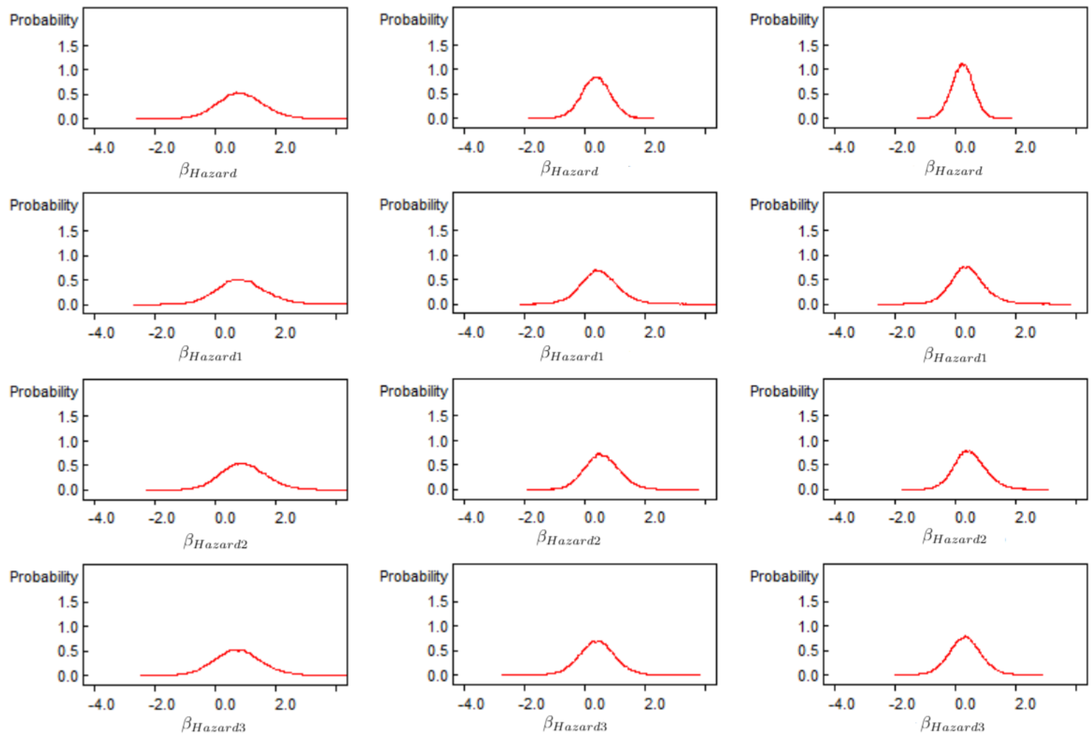

Figure A4: Prior sensitivity β_{Hazard} model 1

$\sim N(0, 100)I(>0)$ $\sim N(0, 10)I(>0)$ $\sim N(0, 1)I(>0)$

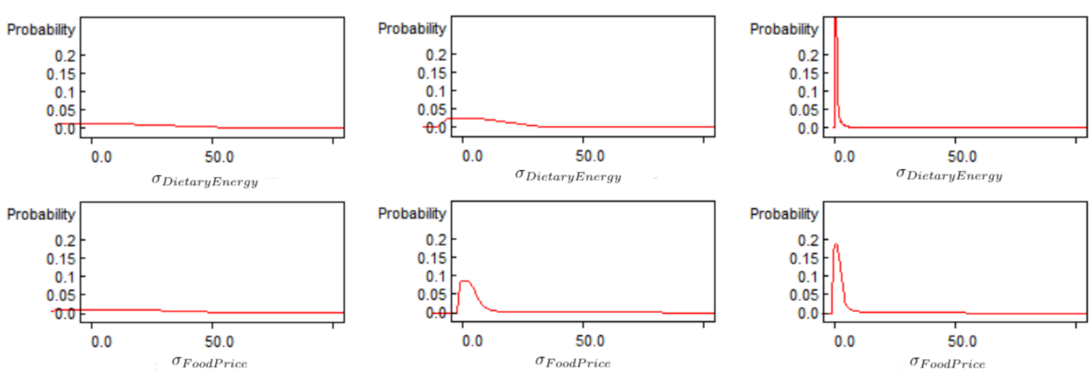

Figure A5: Prior sensitivity σ^2 model 1

$\sim N(0, 10^6)$ $\sim N(0, 100)$ $\sim N(0, 5)$ $\sim N(0, 2)$

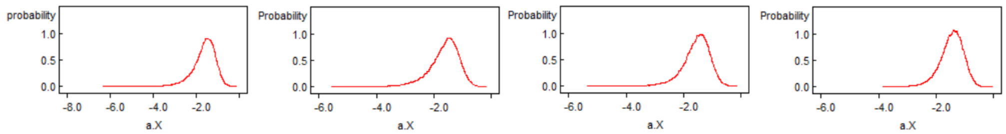

Figure A6: Prior sensitivity $a.X$ model 1

$\sim N(-1, 0.5)$ $\sim N(0, 0.5)$ $\sim N(1, 0.5)$

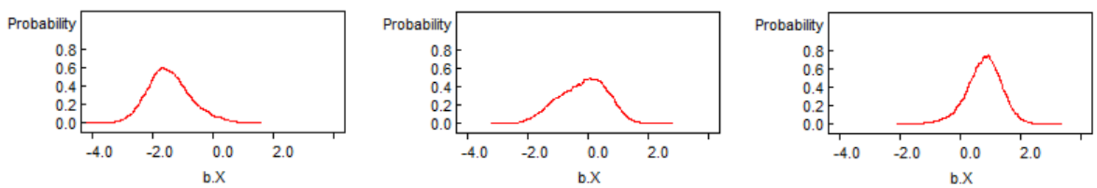

Figure A7: Prior sensitivity $b.X$ model 1

$\sim N(0.0.1)$ $\sim N(0, 0.5)$ $\sim N(0, 2)$

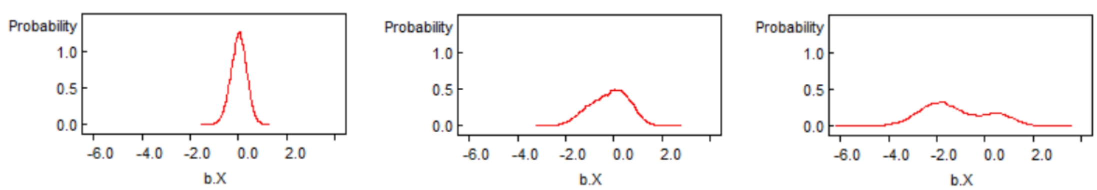

Figure A8: Prior sensitivity $b.X$ model 1

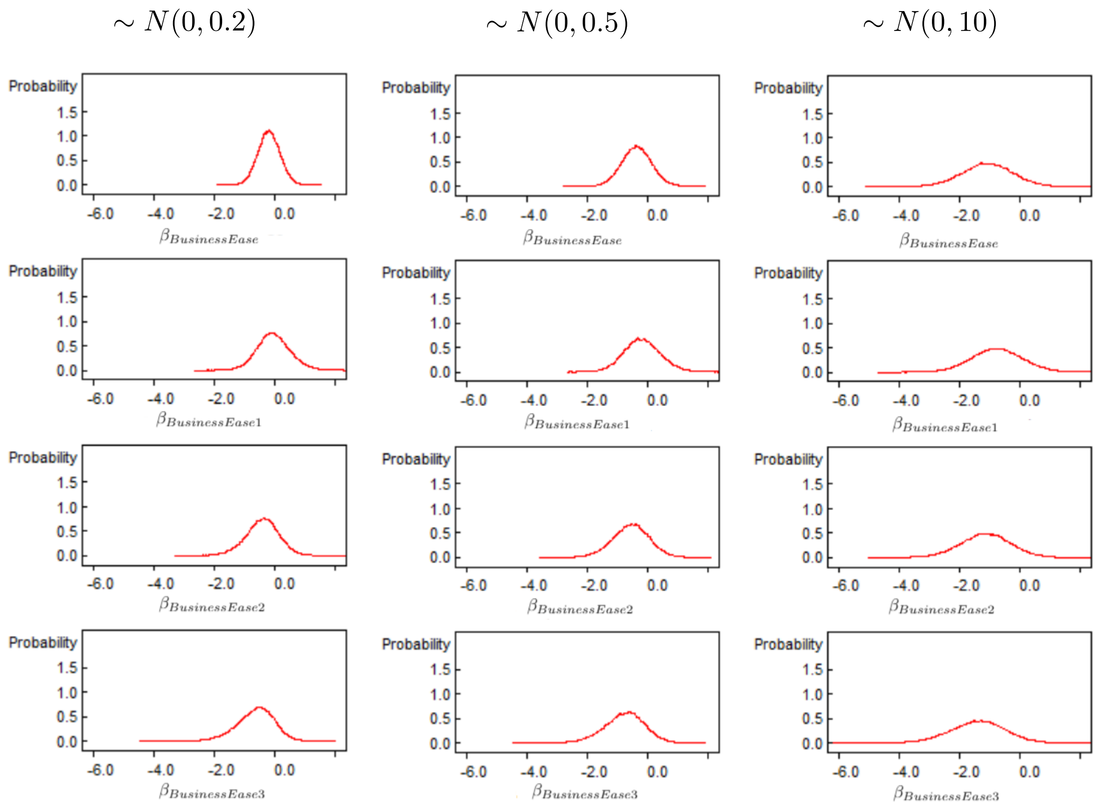

Figure A9: Prior sensitivity $\beta_{BusinessEase}$ model 2

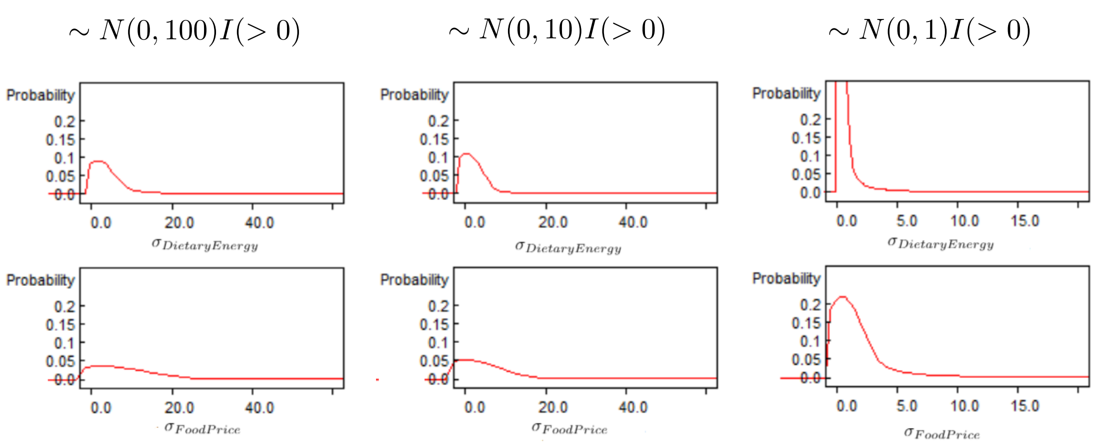

Figure A10: Prior sensitivity σ^2 model 2

Model estimates

	Variable	Mean	Standard deviation	Median	MC error	Percentile zero
Accessibility	α	-1.45	0.60	-1.43	0.005	0.43
	β_{Hazard}	0.45	0.47	0.44	0.008	83.44
	$\beta_{Vulnerability}$	0.11	0.61	0.14	0.006	59.05
	$\beta_{CopingCapacity}$	0.68	0.40	0.66	0.006	95.86
	$\beta_{Corruption}$	0.55	0.68	0.54	0.007	79.71
	$\beta_{BusinessEase}$	-0.18	0.61	-0.20	0.006	37.04
	$\beta_{DietaryEnergy}$	0.007	1.09	0.001	0.01	50.05
	$\beta_{FoodPrice}$	0.03	0.76	0.02	0.01	50.83
Inflation	α	0.34	0.57	0.34	0.003	72.32
	β_{Hazard}	0.48	0.46	0.47	0.008	85.34
	$\beta_{Vulnerability}$	0.66	0.58	0.65	0.005	88.06
	$\beta_{CopingCapacity}$	0.89	0.47	0.85	0.007	98.09
	$\beta_{Corruption}$	0.80	0.68	0.77	0.007	88.77
	$\beta_{BusinessEase}$	-0.56	0.61	-0.55	0.006	17.27
	$\beta_{DietaryEnergy}$	-0.17	1.09	-0.16	0.01	43.75
	$\beta_{FoodPrice}$	0.12	0.73	0.11	0.009	55.97
Capacity	α	0.72	0.62	0.71	0.003	87.97
	β_{Hazard}	0.44	0.47	0.43	0.008	82.58
	$\beta_{Vulnerability}$	0.90	0.65	0.86	0.007	92.89
	$\beta_{CopingCapacity}$	0.53	0.45	0.52	0.007	88.54
	$\beta_{Corruption}$	0.002	0.68	0.03	0.007	51.67
	$\beta_{BusinessEase}$	-0.78	0.67	-0.75	0.007	11.08
	$\beta_{DietaryEnergy}$	-0.17	1.09	-0.15	0.01	43.97
	$\beta_{FoodPrice}$	0.12	0.75	0.10	0.009	55.41

Table A2: Parameter estimates from model 2

	Variable	Mean	Standard deviation	Median	MC error	Percentile zero
Accessibility	α	-1.61	0.76	-1.58	0.007	1.01
	β_{Hazard}	0.87	0.97	0.77	0.011	83.15
	$\beta_{Vulnerability}$	-0.29	0.87	-0.25	0.009	38.21
	$\beta_{CopingCapacity}$	0.69	0.62	0.68	0.006	87.57
	$\beta_{Corruption}$	0.64	0.98	0.61	0.009	74.73
	$\beta_{BusinessEase}$	0.15	0.89	0.13	0.010	55.84
	$\beta_{DietaryEnergy}$	-0.005	2.45	0.003	0.021	50.04
	$\beta_{FoodPrice}$	0.18	1.13	0.15	0.014	55.91
Inflation	α	0.62	0.81	0.59	0.006	77.86
	β_{Hazard}	0.71	0.86	0.66	0.008	80.43
	$\beta_{Vulnerability}$	1.25	0.91	1.18	0.009	93.09
	$\beta_{CopingCapacity}$	2.02	1.13	1.88	0.012	98.78
	$\beta_{Corruption}$	1.41	1.12	1.31	0.013	91.58
	$\beta_{BusinessEase}$	-1.26	0.97	-1.20	0.010	8.47
	$\beta_{DietaryEnergy}$	0.05	2.52	0.03	0.023	50.57
	$\beta_{FoodPrice}$	0.34	1.06	0.32	0.010	62.72
Capacity	α	0.88	0.77	0.86	0.005	87.92
	β_{Hazard}	0.09	0.86	0.10	0.008	55.04
	$\beta_{Vulnerability}$	1.76	1.12	1.64	0.012	96.42
	$\beta_{CopingCapacity}$	-0.06	0.73	-0.05	0.007	47.04
	$\beta_{Corruption}$	-0.07	1.116	-0.02	0.010	49.32
	$\beta_{BusinessEase}$	-1.52	1.12	-1.43	0.012	7.05
	$\beta_{DietaryEnergy}$	-0.86	2.51	-0.66	0.025	37.18
	$\beta_{FoodPrice}$	0.60	1.06	0.52	0.011	71.35

Table A3: Parameter estimates from model 3