Predicting the Impacts of Natural Disasters in Canada

MATH 6627: Case Study

Consultant: Marcos Cardenas-Zelaya

April 1, 2020

Introduction

- Canadian Disaster Database is a data set that keeps records of all major Canadian natural disasters since 1900.
- The events include those that caused significant economic damage to communities from which they could not recover on their own, or events that led to the loss of life or displaced or physically harmed Canadians
- The Canadian Goverenment has a duty to prevent and mitigate future disasters from impacting Canadians.
- Finding what predicts the impacts of natural disasters is of upmost concern for the economy and for the protection of Canadians.

Objective

1. Determine what predicts the impacts of natural disasters.
2. Determine whether these characteristics vary by province and time.
3. Predict the future impacts of natural disasters in Canada.

Data Source

- We used data from the Canadian Disaster Database (CDD).
- The data contains descriptions of natural disasters that have occurred in Canada since 1900.
- There are over 1000 different natural disasters in the data such as biological epidemics, earthquakes, and floods.

The CDD tracks significant disaster events which conform to the Emergency Management Framework for Canada definition of a disaster and meet one or more of the following criteria:

- 10 or more people killed;
- 100 or more people affected/injured/infected/evacuated or homeless;
- an appeal for national/international assistance;
- historical significance;
- significant damage/interruption of normal processes such that the community affected cannot recover on its own,

Data Source

- The data describes when and which provinces/territories the event took place in, the number of injuries, evacuations, and fatalities, as well as an estimate of the costs.
- The data also displays cost data in the dollar amount of the year that the event took place or the year a specific payment was made.
- As many of the events took place across multiple provinces, the data set provides only the aggregate economic costs and human displacements/loss. As such, events that took place across multiple provinces/territories had their economic cost and human loss transformed to account for each provinces/territories population size during the events year.

Data Source

- If an event spanned several years, we denoted that event year as an average of all the years it occurred over.
- Each provinces population by year was web scraped from each provinces/territories Wikipedia page from 1900 on wards. However, the population dataset was taken at decade intervals from 1900 to 1951 and the every 6 years after. As such, we joined the population dateset with the CDD datasets on the events year (and rounded an events year to the closest year in the population data) to get the population of each province/territory during each event.
- As the population datasets only went up to 2014, all events years then take place between 1900 and 2014.

Response

In this analysis, we used two measures of natural disaster impacts; the economic and human costs.

- We decided to separate these two predictors as we believe that these two variables are very different and no one measure can capture both of them adequately. Further, due to changes in the size of governments, demographics and science will cause natural disasters to effect each one differently.
- For example, as science improve we are able to reduce biological pandemics which impacts human life more directly than economic costs; in contrast a flood that destroys an evacuated town (as we become better at predicting floods) will have more economic than human costs.

Economic Impact

We composed a composite popularity score using the following variables:

- The most quantifiable measurement of a disasters impact is its economic cost.
- This includes damages to property and businesses, forgone wages, and increase in investments to replace damaged capital.
- As we transformed the estimated total cost variable to 2014 prices using the canadian CPI measure, we control for inflation which allows us to compare costs of each event across time.
- If a natural event occurred in multiple provinces, then the total cost for the event in one of those provinces needs to be standardized by its population. The reason being that the impact of the natural disaster is not spread evenly across provinces and hence neither will the total costs in the data.

Economic Impact

- For example, a flood that effects both Quebec and New Brunswick will incur different costs. Our reasoning is that since Quebec is a more capital intensive economy than New Brunswick, it will incur more costs in damaged physical capital, thus incurring a larger economic loss in absolute terms.
- Thus, the datasets does not reflect this asymmetry in losses and provides just the total Federal loss to the event and not the provincial loss.
- As such, we standardize the losses by multiplying the loss by each provinces proportional population size to get the cost per province.

Human Impact: A Composite Score

- The data set provides three variables that are related to the direct impact to humans; fatalities, injured/infected, evacuated and number of people that loose access to utility services.
- We transform all the human impact measures into percentages; we divide the total human loss by the sum of all provinces population involved to get what percentage of each provinces population is effected. We assume that every province involved suffers the same amount of human loss in percentages.
- By transforming the human loss into percentage this allows us to control for increases in population over time which makes comparing losses over time easier.

Human Impact: A Composite Score

- We composed a composite human impact score using these variables. We use equal weightings in the construction of the composite variable.
- Since all of these variables are on vastly different scales of magnitude, we normalize the variables and add them to create our composite human impact score.

Predictors

- **Event Type**: We include the event type as the main predictor as we predict that different types of events will result with differing effects on economic and human costs. For example, a biological pandemic will create more human loss than direct economic costs, whereas a flood will cause more economic damage than human loss.
- **Earthquakes Magnitude**: We assume that stronger earthquakes will cause greater economic and human loss.
- **Events Duration**: We predict that as an events duration increases, both measures will increase.

Predictors

- **Number of Provinces Involved**: We included this as we posit that as more provinces are involved the costs will increase as all levels of government are strained and not able to focus on mitigating costs. Indeed, as more provinces are involved, then it is likely that more government and hence more bureaucratic slog will slow down responses to crisis management.
- **Province of Event**: The data indicates in which provinces an event occurred. We would like to see how costs are ranked by provinces.
- **Year of Event**: We would like to see how costs evolve over time.

Variable	**Levels**	**n**	**%**	$\sum$ **%**
EVENT.TYPE	Avalanche	17	1.5	1.5
	Cold Event	68	6.1	7.6
	Drought	127	11.4	19.0
	Earthquake	5	0.5	19.5
	Epidemic	87	7.8	27.3
	Flood	311	27.9	55.2
	Geomagnetic Storm	1	0.1	55.3
	Heat Event	24	2.1	57.5
	Hurricane / Typhoon / Tropical Storm	49	4.4	61.9
	Infestation	2	0.2	62.0
	Landslide	39	3.5	65.5
	Storm - Unspecified / Other	17	1.5	67.1
	Storm Surge	11	1.0	68.0
	Storms and Severe Thunderstorms	125	11.2	79.3
	Tornado	43	3.9	83.1
	Tsunami	3	0.3	83.4
	Wildfire	93	8.3	91.8
	Winter Storm	92	8.3	100.0
	all	1114	100.0	

:

Statistical Analysis

- As these events span over 100 years, and if we assume that climate change is causing more frequent and destructive natural events, then there is reason to believe that there exists variation in the data over time.
- Further, due to obvious geographical and social demographic difference among provinces, there is variation in the data among provinces.
- As such, we use mixed models to capture this time and space variation.

Statistical Analysis With Mixed Models: Time Variation

- To determine whether the impact of natural disasters has changed over time, we divide time by the 12 decades from 1900 to 2014.
- **Random intercept:** Provinces evolve tremendously in every passing decade, in particular as global warming changes the environment, we posit that the average response variable will be different across time within a province. Also, due to advances in technology, such as river level controls in the Don Valley River, the average cost of these events will drop as we become better at managing crisis' over time.
- **Random slope:** To account for potential differences in how the different types of events can effect the responses over time. Indeed, a flood in the early 1900s was more destructive when houses were made of logs and hay, compared to todays more resilient buildings.

Statistical Analysis With Mixed Models: Time Variation

We apply Mixed Models to the Linear regression to model the variation in time for the normalized economic cost. Economic Cost_{ij} denotes the normalized economic impact.

$$\begin{aligned}\text{Economic Cost}_{ij} &= \beta_0 + b_{0j} + (\beta_1 + b_{1j})\text{Event Type}_{ij} \\ &+ (\beta_2 + b_{2j})\text{Event Duration}_{ij} + (\beta_3 + b_{3j})\text{Magnitude}_{ij} \\ &+ (\beta_4 + b_{4j})\text{Num.Prov. Involved}_{ij} + \varepsilon_{ij}\end{aligned} \tag{1}$$

$$i = \{1, ..., n_j\}, j = \{1910, ..., 2008, 2014\}$$

In this model we have b's as the random slope/intercept for i observations from each time group j.

Statistical Analysis With Mixed Models: Time Variation

We apply Mixed Models to the Linear regression to model the variation in time for the composite human costs score Human Impact$_{ij}$ denotes the human costs score.

$$\begin{aligned}\text{Human Cost}_{ij} &= \beta_0 + b_{0j} + (\beta_1 + b_{1j})\text{Event Type}_{ij} \\ &+ (\beta_2 + b_{2j})\text{Event Duration}_{ij} + (\beta_3 + b_{3j})\text{Magnitude}_{ij} \\ &+ (\beta_4 + b_{4j})\text{Num.Prov. Involved}_{ij} + \varepsilon_{ij}\end{aligned} \tag{2}$$

$$i = \{1, ..., n_j\}, j = \{1910, ..., 2008, 2014\}$$

In this model we have b's as the random slope/intercept for i observations from each time group j.

Statistical Analysis With Mixed Models: Provincial Variation

- To determine whether the impact of natural disasters is different among provinces, we set the random levels as the 13 provinces and territories.
- We chose to use provinces over geographical regions (Maritimes, Prairies etc.) because we believe that including more levels will allow a finer grain analysis, and also we posit that there is large variation within each region.
- For example, the prairies includes Alberta, Saskatchewan and Manitoba which despite sharing similar geographies and hence similar types of natural disasters, vary in their economies. These economic differences will lend themselves to differences in how the province prevents/mitigates natural disasters. Indeed, in oil rich Alberta, the government might have more resources to mitigate the harms of floods than Manitoba which in effect will result with lower economic and human losses. We chose to include random intercepts and slopes.

Statistical Analysis With Mixed Models: Provincial Variation

- **Random intercept:** Provinces vary greatly by geography and thus the type of events that effect them. More destructive events such as floods are concentrated in certain provinces, most notably the prairies. This concentration of events by geography will cause each province to have different average economic and human costs.
- **Random slope:** To account for potential differences in how the different types of events can effect the response variables among provinces. Indeed, a flood in an area accustomed to floods with existing measures to cope will likely incur less damage than, say, Southern Ontario where floods are rare and will cause havoc as they are not prepared.

Statistical Analysis With Mixed Models: Provincial Variation

We apply Mixed Models to the Linear regression to model the variation among provinces for the normalized economic cost. Economic Cost_{ij} denotes the normalized economic impact.

$$\begin{aligned}\text{Economic Cost}_{ij} = \beta_0 + b_{0j} + (\beta_1 + b_{1j})\text{Event Type}_{ij} \\ + (\beta_2 + b_{2j})\text{Event Duration}_{ij} + (\beta_3 + b_{3j})\text{Magnitude}_{ij} \\ + (\beta_4 + b_{4j})\text{Num.Prov. Involved}_{ij} + \varepsilon_{ij}\end{aligned} \quad (3)$$

$$i = \{1, ..., n_j\}, j = \{Alberta, ..., Ontario, Quebec\}$$

In this model we have b's as the random slope/intercept for i observations from each province j.

Statistical Analysis With Mixed Models: Provincial Variation

We apply Mixed Models to the Linear regression to model the variation among provinces for the composite human costs score Human Impact_{ij} denotes the human costs score.

$$
\begin{aligned}
\text{Human Cost}_{ij} = \beta_0 + b_{0j} &+ (\beta_1 + b_{1j})\text{Event Type}_{ij} \\
&+ (\beta_2 + b_{2j})\text{Event Duration}_{ij} + (\beta_3 + b_{3j})\text{Magnitude}_{ij} \\
&+ (\beta_4 + b_{4j})\text{Num.Prov. Involved}_{ij} + \varepsilon_{ij}
\end{aligned} \tag{4}
$$

$$i = \{1, ..., n_j\}, j = \{Alberta, ..., Ontario, Quebec\}$$

In this model we have b's as the random slope/intercept for i observations from each province j.

Results: Linear Regressions

: Economic and Human Cost Linear Regression

	Dependent variable:	
	NORMALIZED.TOTAL.COST	human_cost_comp_s
	Economic Cost(Millions)	Human Cost
Cold Event	-50.496(115.809)	0.108^{***} (0.039
Drought	44.972(99.813)	-0.076^{**} (0.034
Earthquake	-175.112(356.707)	0.005(0.120)
Epidemic	-43.554(116.558)	0.023(0.039)
Flood	28.715(94.346)	-0.007(0.032)
Geomagnetic Storm	-34.053(384.312)	0.075(0.129)
Heat Event	23.283(128.400)	0.150^{***} (0.043
Hurricane	27.184(108.548)	0.006(0.037)
Infestation	-38.530(279.883)	-0.005(0.094)
Landslide	38.164(109.167)	-0.010(0.037)
Storm Surge	0.264(147.492)	0.002(0.050)
Storms	10.218(97.846)	-0.004(0.033)

Results: Linear and Regressions

: Economic and Human Cost Linear Regression

	Dependent variable:	
	NORMALIZED.TOTAL.COST Economic Cost(Millions)	human_cost_comp_score Human Cost
Duration (days)	0.008(0.043)	0.001*** (0.00001)
Magnitude	36.119(55.734)	-0.003(0.019)
Num.Provinces	5.274(5.992)	-0.013*** (0.002)
BC	-55.778(48.583)	-0.009(0.016)
MB	-28.391(46.412)	0.006(0.016)
NB	3.381(52.596)	-0.009(0.018)
NL	-85.908(56.492)	0.014(0.019)
NS	-83.971(59.272)	-0.011(0.020)
NT	-65.142(77.596)	0.022(0.026)
NU	-63.944(108.701)	-0.002(0.037)
ON	-19.376(43.658)	-0.008(0.015)
Observations	1,114	1,114

Results: Linear and Regressions

: Economic and Human Cost Linear Regression

	Dependent variable:	
	NORMALIZED.TOTAL.COST Economic Cost(Millions)	human_cost_comp_score Human Cost
ON	-19.376(43.658)	-0.008(0.015)
PE	-103.227(68.038)	-0.006(0.023)
QC	-13.482(46.461)	-0.012(0.016)
SK	-33.949(48.191)	-0.005(0.016)
YT	-62.270(88.978)	-0.001(0.030)
Year	1.190** (0.465)	-0.0004*** (0.0002)
Intercept	-2,327.665** (925.329)	0.887*** (0.311)
Observations	1,114	1,114
R^2	0.037	0.697
Adjusted R^2	0.007	0.688
Note:		*p<0.1; **p<0.05; ***p<0.01

Results: Linear Regressions

- **Event Type**: Both models do not agree on any of the type of events.Recall, each $\beta_{eventtype}$ is the difference between the average response for that event type and the the average response for the base event type 'Avalanche'. Since all the events, except for winter storms, are not significant, this suggests that strongest predictors of costs is anything snow related. However, this is not the case for human costs. For human costs; cold and heat events are among the most impact full events.
- **Earthquakes Magnitude**: Both models agree that earthquake magnitude is not significant.
- **Events Duration**: Only the human costs model suggests that the event duration is significant and positive. We posit that many of the events occur over a short period of time, over a day or two, so durration is not a strong predictor for economic costs, however epidemics last a very long time and thus will incur more impacts to humans than psychical capital.

Results: Linear Regressions

- **Num. Provinces Involved**: Only the human costs model indicates it as significant. Although what is interesting is that it suggests that as more provinces are involved, the human impact decreases. We posit that this might be because as more provinces are involved, the loss is "diluted" across the provinces.
- **Province of Event**: Both models find the provinces to not be significant, which suggests that losses are felt equally among provinces.

Results: Linear Regressions

- **Year of Event**: What is interesting is that the economics model suggests that costs have increased over time whereas human costs have decreased over time. We posit that with advancements in science we are better equipped to forecast events and thus evacuate people thus reducing human losses. However, the increase in costs can be for several reasons; increases "strength" of natural disasters over time, increases in the cost of items damaged. A flood that wreaks havoc in 1920 Southern Ontario, even when adjusting for inflation, will incur less absolute economic costs than a flood in 2020 Southern Ontario.
- **Adjusted** R^2: The economic costs model has an adjusted R^2 of almost 0, where as the human costs is almost .7. This suggests that these variables do not predict the economic impacts well, where as it does well for the human impact.

Model Selection: Testing Random Intercept

We now test the random intercepts and random slopes for each model. That is, we test whether there exists difference in the average responses across time and provinces, and whether there exists differences in how the predictors effect the response across time and provinces.

Table: Testing Random Intercept

	df	AIC
Economic Cost Povince null	22.00	47151.83
Economic Cost Povince full	23.00	46376.05
Economic Cost Time null	22.00	47151.83
Economic Cost Time full	23.00	46357.20
Human Cost Province null	22.00	-1447.57
Human Cost Province full	23.00	-1307.21
Human Cost Time null	22.00	-1447.57
Human Cost Time full	23.00	-1394.44

Model Selection: Testing Random Intercept

- **Economic Cost Model with Random Intercept for Provinces**: The AIC for the mixed model with random intercept is lower than the null, so it is preferred. This suggest that there exists differences in the average economic cost across provinces.
- **Economic Cost Model with Random Intercept for Time**: The AIC for the mixed model with random intercept is lower than the null, so it is preferred. This suggest that there exists differences in the average economic cost across time.
- **Human Cost Model with Random Intercept for Provinces**: The AIC for the mixed model with random intercept is lower than the null, so it is preferred. This suggest that there exists differences in the average composite human cost score across provinces.
- **Human Cost Model with Random Intercept for Time**: The AIC for the mixed model with random intercept is lower than the null, so it is preferred. This suggest that there exists differences in the average composite human cost score across time.

Model Selection: Testing Random Slope

Similarly, we test the random slopes. We test these random slopes using χ^2 model selection.

Table: Testing Random Slopes

	Df	Chisq	Chi Df	Pr(>Chisq)
Economic Cost Povince simple	23			
Economic Cost Povince full	253	0.00	230	1.0000
Economic Cost Time simple	23			
Economic Cost Time full	253	935.65	230	0.0000
Human Cost Province simple	23			
Human Cost Province full	253	47.41	230	1.0000
Human Cost Time simple	23			
Human Cost Time full	253	1719.02	230	0.0000

Model Selection: Testing Random Slope

- **Economic Cost Model with Random Slope for Provinces**: The p-value is large, so the model without random slope is an adequate simplification of the full model. This suggest that there does not exist differences in how the predictors effect economic costs across provinces.
- **Economic Cost Model with Random Slope for Time**: The p-value is very small, so the model without random slope is an adequate simplification of the full model. This suggest that there exists differences in how the predictors effect economic costs over time.
- **Human Cost Model with Random Slope for Provinces**: The p-value is large, so the model without random slope is an adequate simplification of the full model. This suggest that there does not exist differences in how the predictors effect human costs across provinces.
- **Human Cost Model with Random Slope for Time**: The p-value is very small, so the model without random slope is not an adequate simplification of the full model. This suggest that there exists differences in how the predictors effect human costs over time.

Model Selection

- The best models include both random intercepts and slopes for both human and economic costs when time is the random effects. However, when provinces is the random effects, we see that the best model is only that with random intercepts.
- These two findings suggest that the average impacts of natural disasters vary across provinces and over time. However, how the various variables predict the impacts is the same across provinces, but varies over time.
- The variation across provinces by average impacts is quite intuitive as provinces that experience more natural disasters will have higher impacts. Further, the lack of random slopes for provinces suggest that all provinces are effected the same by natural disasters.
- We would like to mention that although the best models have random intercepts, the AIC values are all quite close together which questions how strong this result is.

Prediction of Future Impacts

- What is interesting is the significance of the random slopes for time random effects for both cost models. This suggests that most of the variation in the data can be explained by changes over time. This suggests that although all provinces are effected the same by natural disasters, as a country we will all experience more increases in costs as time increases, as the linear models above showed.
- Although, the impacts over time will decrease for human costs (which is good), but increase for economic costs, although not by much (about 1.2 million CAD$ a year) is substantial and worthy of further investigation.
- We would like to point out that the increase in costs over time can be for several reasons such as the data not accurately capturing the changes in the absolute amount of physical goods in an economy. Even after adjusting for inflation, 1 million in damages in 1900s Ontario meant a world of difference than 1 million dollars in damages in 2020. Most notably the ability of modern economies to bounce back from natural disasters makes comparison across time difficult.

Conclusion

- In this analysis we used data from the CDD and web scraped data from the Demographics Wikipedia pages for each province to determine the impacts of natural disasters in Canadian provinces.
- We measured impact by using an economics cost and a composite human costs score that encompasses several intuitive measures of human loss.
- We used linear mixed models with random intercepts and slopes to exploit variation in events across time and provinces.
- We found that the majority of variation in the data can be explained by variation across time and provinces. In particular, we find that there is weak evidence to support random intercepts for both cost models with time and provincial random effects (similar AIC scores).

Conclusion

- Further, we find that there is little evidence that suggests that the variables that predict costs vary by provinces, however we find that the random slopes vary over time. We posit that although all provinces are effected the same by natural disasters, the impacts will increase for all provinces over time.
- We also pointed out that we are unable to determine what causes the increase in costs over time. This could be due to general increases in the amount of physical capital that can be damaged, a measure that adjusting for inflation cannot capture, or whether this is because of the "strength" of natural disasters is increasing. Further analysis would benefit from including the magnitudes of each type of event.