

# 21st Century Newton

Tai Wang, Hao Zhou, Rui Feng  
Zhejiang University

February 10, 2018

## **Abstract**

Winner winner chicken dinner

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Theoretical Analysis</b>	<b>3</b>
2.1	Assumptions and Model Framework	3
2.2	Representing the Two Factors	4
2.3	Probabilistic Fragility Measure	4
2.4	Modeling the Interaction between Variables	6
	Direct Effect of Environmental Factors	6
	Indirect Effect of Environmental Factors	7
2.5	Temporal Model	8
<b>3</b>	<b>Experiments</b>	<b>9</b>
3.1	Data Preparation	9
<b>4</b>	<b>Related Work</b>	<b>9</b>
<b>5</b>	<b>Conclusion</b>	<b>9</b>

# 1 Introduction

Quantifying the fragility of states based on human factors, such as **[\*\* List here some example human factors \*\*]** has been extensively studied by researchers and institutions, such as in **[\*\* cite some important works here. \*\*]** However, these models do not consider the impact of environmental factors. For example, deterioration in natural environment may contribute to regional instability and violence **[\*\* some references required here \*\*]**. As environmental factors are important in determining a state or a region's sustainability, merely considering human factors is clearly insufficient.

## 2 Theoretical Analysis

In this part we propose a theoretical framework for our analysis of the impact of climate change on *state fragility*.

### 2.1 Assumptions and Model Framework

We propose two natural assumptions, based on which we derive the basic framework of our model.

1. The *state fragility*, a concept to estimate the sustainability of states, is dependent on and only on *human factors* and *environmental factors*.
2. The environmental factors and human factors interact with each other.

The assumptions are natural. Assumption 1 requires to quantify the fragility which considers both human factors and environmental factors. We propose a novel framework to quantify fragility by incorporating the human and environmental factors into a probabilistic framework.

Assumption 2 requires a more sophisticated analysis of the two factors, including their respective and joint effects on the fragility, and the interaction between them. We are especially interested in the effects of environmental factors, which include *direct* effect, which is the influence on fragility directly imposed by environmental factors; and *indirect* effect, which is the influence on fragility imposed by environmental factors indirectly through human factors. Two hypothesis models to explain the indirect effect are visualized in Figure 1(a) and 1(b).

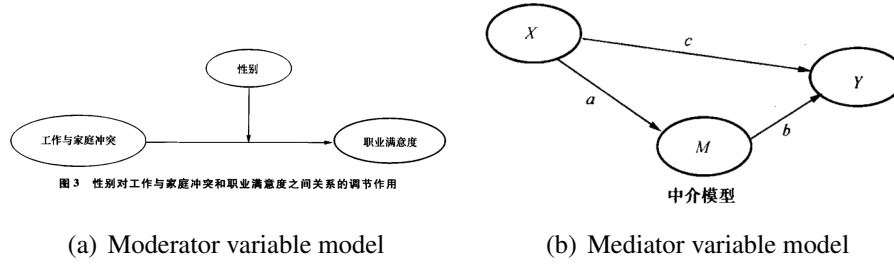


Figure 1: Hypothesis models for indirect effect of environmental factors.  $E$  represents environmental factors,  $H$  represents human factors, and  $F$  represents state fragility. The fragility is jointly determined by  $E$  and  $H$ , by two hypothesis approach.

The direct effect of environmental factors is measured by the effect of environmental factors on fragility score, with an unbiased estimation of the effect obtained by propensity score matching **[\*\* citation \*\*]**. The indirect effect of environmental factors is explained by two models: the *moderator variable* model, as shown in Figure 1(a); and the *mediator variable* model, as shown in Figure 1(b).

In the following of this section, we are dedicated to enrich our model.

## 2.2 Representing the Two Factors

Notation	Description
$H$	random variable of human factors
$E$	random variable of environmental factors
$F$	binary random variable of fragility
$s_f$	fragility score

## 2.3 Probabilistic Fragility Measure

**[\*\* Need to emphasize that it is logistic regression. \*\*]** In this part, we derive a novel fragile score,  $s_f$ , which incorporates both environmental and human factors. The score  $s_f$  is based on probabilistic intuitions, and is therefore called the *probabilistic fragility score*, or the fragility score for convenience.

Without loss of generality, we refer to regions, sovereign states, and other concerned geographical entities as states.

We assume that a state either fragile or stable, described by a binary random variable  $F$ , where  $F = 1$  if the considered state fragile, and  $F = 0$  if it is stable.  $H$  and  $E$

are random variables describing the human and environmental factors of the state. For convenience, we further assume that  $E$  is binary, i.e.  $E = 1$  when the state's environment is sustainable, and  $E = 0$  when it is not.

Consider the probability of a state being fragile, given its human and environmental factors:

$$\mathbb{P}(\mathbf{F} = 1 | E = e, \mathbf{H}) \quad (e = 0, 1) \quad (1)$$

The probability given in 1 quantifies the extent of fragility of the state, given certain environmental and human factors. It is higher when the state is more vulnerable. However, the conditional distribution is hard to estimate. We then factorize it into a more easily calculated form:

$$\mathbb{P}(\mathbf{F} = 1 | E = e, \mathbf{H}) = \frac{\mathbb{P}(E = e, \mathbf{F} = 1 | \mathbf{H})}{\mathbb{P}(E = e | \mathbf{H})} = \frac{\mathbb{P}(\mathbf{Z} = 1 | \mathbf{H})}{\mathbb{P}(E = e | \mathbf{H})} \quad (e = 0, 1) \quad (2)$$

In which we defined a new random variable  $\mathbf{Z} = 1$  if  $E = e, \mathbf{F} = 1$  and  $\mathbf{Z} = 0$  otherwise. Eqn. 2 allows us to only estimate the conditional probability of two binary random variables given human factors  $\mathbf{H}$ .

For convenience of calculation, we assume linear relationships:

$$\begin{aligned} \log \frac{\mathbb{P}(\mathbf{Z} = 1 | \mathbf{H})}{\mathbb{P}(\mathbf{Z} = 0 | \mathbf{H})} &= \mathbf{W}_1 \mathbf{H} + \mathbf{e}_1 \\ \log \frac{\mathbb{P}(E = e | \mathbf{H})}{\mathbb{P}(E = 1 - e | \mathbf{H})} &= \mathbf{W}_2 \mathbf{H} + \mathbf{e}_2 \end{aligned} \quad (3)$$

Where  $\mathbf{W}_i$  are parameters,  $\mathbf{e}_i$  are Gaussian errors,  $i = 1, 2$ .

Using the linear assumption and the logistic form in Eqn 3, we obtain the estimate of the probabilities respectively:

$$\begin{aligned} \hat{p}_z &= \frac{\exp(\mathbf{W}_1 \mathbf{H})}{1 + \exp(\mathbf{W}_1 \mathbf{H})} \\ \hat{p}_e &= \frac{\exp(\mathbf{W}_2 \mathbf{H})}{1 + \exp(\mathbf{W}_2 \mathbf{H})} \end{aligned} \quad (4)$$

In order to make the estimated probability distribution resemble the true distribution, we estimate parameters  $\mathbf{W}_1, \mathbf{W}_2$  by minimizing the cross entropy loss, which is equivalent to minimizing the KL divergence **reference needed** between the estimated distribution and the empirical distribution. **specific form omitted.**

Finally, probabilities in Eqn. 2 is replaced by the estimates given in Eqn. 4, yielding the *probabilistic fragility score*:

$$s_f = \frac{\hat{p}_z}{\hat{p}_e} \quad (5)$$

**Notes on the fragility score.** The fragility score,  $s_f$ , is derived based on probabilistic intuitions and linear assumptions. Higher  $s_f$  indicates higher risks of being fragile. However, the score  $s_f$  can be larger than one, and is, therefore, not in form of probability. However, it does not hurt its applicability: if the estimated  $\hat{p}_z$  is larger than  $\hat{p}_e$ , we have even more reasons to believe that the considered state is fragile.

## 2.4 Modeling the Interaction between Variables

We then begin to model the relationship between the three variables: environmental factors  $E$ , human factors  $H$ , and the fragility score  $s_f$ .

### Direct Effect of Environmental Factors

In order to measure the effect of environmental factors on fragility score, a naive approach would be to sample states of both sustainable environment and unsustainable environment, and compare their average fragility score. Formally, write  $s_f^0$  as the average fragility score of the sustainable group, and  $s_f^1$  as the average fragility score of the unsustainable group. One then compares the difference  $s_f^* = s_f^0 - s_f^1$ .

The above approach gives, however, biased estimation, because the apparent difference between these two groups may be depend on human factors that affected whether or not a state's environment is sustainable, instead of the environmental status per se. For example, the approach might compare scores of the states that are environmentally unsustainable and in political turmoil, with the scores of the states that are environmentally sustainably and politically stable. The difference of political status results in unbiased estimate of the effect of environmental factors.

To control for the differences of human factors between the sustainable group and the unsustainable group, we use the propensity score matching, a statistical technique that attempts to estimate unbiasedly the effect of a variable, in this case, the environmental status.

Formally, the propensity score of a certain state is defined as the conditional probability of the environmental status given its human factors,

$$p = \mathbb{P}(E = 1 | \mathbf{H}) \quad (6)$$

**[\*\* environmental status? environmental factors? they are probably different. \*\*]**

In order to estimate the probability, we adopt similar procedures used in Section 2.3, using the score of logistic regression,  $\hat{p}$ , of human factors  $\mathbf{H}$  against environmental status  $E$ . Then we match each of the unsustainable states to one sustainable state on propensity score, by using *Nearest Neighbor Matching*: each unsustainable state is matched to the sustainable state whose propensity score is the closest. As such, a new data set in which the sustainable group and the unsustainable group and their propensity scores are balanced, is obtained.

Based on the newly obtained data set, we calculate the adjusted score difference:

$$\hat{s}_f^* = \hat{s}_f^0 - \hat{s}_f^1. \quad (7)$$

Where  $\hat{s}_f^0$ ,  $\hat{s}_f^1$  are average scores of the sustainable and unsustainable group, drawn from the data set obtained by PSM.

### Indirect Effect of Environmental Factors

In order to measure the indirect effect of environmental factors  $E$ , we propose two candidate models: the moderator variable model, and the mediator variable model.

**[\*\* is the difference clearly explained? \*\*]** The moderator variable model assumes that the environmental factors influence the fragility score, and the relationship is calibrated by the effect of human factors. In this case, the human factors  $\mathbf{H}$  are called *the moderator variable*, as illustrated in Figure 1(a). **[\*\* reference. \*\*]**

The mediator variable model assumes that the environmental factors and human factors jointly influence the fragility score. Furthermore, the environmental factors influence the fragility score both directly, and indirectly through the human factors. The illustration of this model is as in Figure 1(b). **[\*\* reference. \*\*]**

**Moderator variable model.** The model is written as

$$s_f = \mathbf{W}_1 \mathbf{H} + W_2 E + \sum_{i=1}^p \beta_i h_i E \quad (8)$$

where  $\mathbf{H} = [h_1, \dots, h_p]'$ , and  $h_i$  is the  $i$ th component of  $\mathbf{H}$ , indicating a specific factor.  $\mathbf{W}$  and  $\beta_i$  are parameters.

The term,  $\sum_{i=1}^p \beta_i h_i$ , represents the effect of each human factor on the relation between the environmental factor and the fragility score, since by taking partial derivative, we observe that

$$\frac{\partial s_f}{\partial E} = W_2 + \sum_{i=1}^p \beta_i h_i$$

The derivative indicates that the effect of environmental factor  $E$  comes from both itself, described by  $W_2$ , and each human factor  $h_i$ , described by  $\beta_i$ . If the factor  $h_i$  has no effect on the relation, then  $\beta_i$  should be close to zero. Consider, therefore, the following statistical test:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p \quad (9)$$

The rejection of  $H_0$  shows that the moderator effect of  $\mathbf{H}$  is significant. Furthermore, we wish to specifically investigate the effect of each factor. Hence in the following test,

$$H_0 : \beta_i = 0 \quad (10)$$

if  $H_0$  is rejected, we are confidence to say that the moderator effect of  $h_i$  is statistically significant.

If the moderator effect is indeed significant, then the extent of the effect of  $\mathbf{H}$  is then quantified by  $\beta = [\beta_1, \dots, \beta_p]'$  and respective components.

**Mediator variable model. [\*\* Human factors are multivariate. Maybe consider human factors only as FSI, a single variable. \*\*]**

We first declare the following values:

- $b_1$ : the coefficient of the linear regression model, in which  $E$  predicts  $s_f$ .
- $b_2$ : the coefficient of the linear regression model, in which  $\mathbf{H}$  predicts  $s_f$ .
- $b_3$ : the coefficient of  $E$  in the linear regression model in which  $E$  and  $\mathbf{H}$  jointly predicts  $s_f$ .
- $b_4$ : the coefficient of  $\mathbf{H}$  in the linear regression model in which  $E$  and  $\mathbf{H}$  jointly predicts  $s_f$ .

First, we need to verify by statistical tests that  $b_1$  and  $b_2$  are significantly nonzero.

If  $\mathbf{H}$  indeed acts as a mediator variable, by model assumption,  $b_4$  is significantly nonzero. Furthermore, since  $E$  indirectly influences  $s_f$  through  $\mathbf{H}$ , the explaining power of  $E$  alone should be reduced once  $\mathbf{H}$  is introduced. In this case,  $b_3 < b_1$ .

If all the above four conditions hold significantly, we are confident to say that human factors act as mediator variables, through which the environmental factors influence the fragility score.

## 2.5 Temporal Model

In order to better model climate change, we further consider time as a variable, and investigate how climate change would have influenced the fragility of states.



Specifically, we choose China as an example state, and illustrate how and when it would reach a tipping point. We build the following temporal model:

$$E = T + S + C + I \quad (11)$$

in which  $E$  is the environmental factor,  $T$  represents long-term tendency,  $S$  represents seasonal fluctuation,  $C$  represents periodic fluctuation with an unknown period, and  $I$  represents random fluctuation. This is a traditional temporal model; for specifics, readers may refer to **[\*\* reference \*\*]**.

The long-term tendency  $T$  is the crucial ingredient. In the following analysis, we consider the temporal model on a yearly basis, and therefore ignore the seasonal fluctuation  $S$ . The periodic fluctuation in climate change is majorly manifested in the periodic movement of the solar system **[\*\* reference \*\*]**, and the period is long enough to ignore in our study which considers only several decades of years. The random fluctuation  $I$  is assumed in our analysis to be the white noise.

Therefore, our model is expressed as

$$E_t = \alpha + \delta t + \epsilon_t \quad (12)$$

where  $t$  indicates time,  $E_t$  is the environmental factor at time  $t$ ,  $\epsilon_t \sim N(0, \sigma^2)$  is a white noise. To estimate  $\alpha$  and  $\delta$ , we utilize tools of traditional temporal analysis, the OLS estimate, to obtain the asymptotic distributions of parameters  $\alpha$  and  $\delta$ .

## 3 Experiments

### 3.1 Data Preparation

We prepared several representative datasets for the variables we need:  $E$ ,  $H$ , and  $F$ .

## 4 Related Work

## 5 Conclusion