For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

Team Control Number
1924678

Problem Chosen
C

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

2019
MCM/ICM
Summary Sheet

Some Mathematical Models for Countering the Opioids Crisis

Summary

The United States is experiencing a national crisis regarding the use of synthetic and non-synthetic opioids. We analyze the data about the the opioids crisis and construct several mathematical models.

For part 1, we first preprocess the initial data. We add **longitude and latitude data** for every specific counties. Next, we construct the **Combined Spread Model** for part1. Through analyzing the characteristics of the data, we construct first model derived from **Smoke Diffusion Model**. Then we build up another model named **Holt-Winter exponential smoothing model** to describe the time varying property of the data. We combined the two models together into one **Combined Spread Model**.

Using our **Combined Spread Model**, we find the spread and characteristics of the opioids cases. Cuyahoga (OH), Hamilton (OH), Allegheny (PA) and Philadelphia (PA) are four counties that almost have the largest number of both heroin cases and synthetic opioids cases. Then we use **Extreme Value Model** to find the starting locations. We put forward **three concerns** and corresponding **three threshold levels**: **critical threshold level, alarm threshold level and diffusion threshold level**. We use the **Combined Spread Model** to predict and obtain the places and time for every specific opioid.

For part 2, We see data in part 2 as **long panel data** and set **dummy variables**. We use **LSDV, stepwise regression, best-subsets regression** to analyze the initial data. Then we perform **heteroscedasticity test** and **cross sectional correlation test**. We find 8 factors and 6 factors that are significant to synthetic opioids and heroin, respectively. We use **PCA** and **linear combination** to modify our model in part 1 separately. Then we combine the two modified models together to form a **Modified Combined Spread Model**.

For part 3, According to the factors found in part 2, we set the following strategy: Improve the **education** level of the five states. We also test the effectiveness of our strategy, and the results show that we succeed.

Keywords: Combined Spread Model; Extreme Value Model; Threshold levels; long panel; LSDV; Regression; PCA; Linear combination

Contents

MEMO

From: Team 1924678, MCM2019

To: the Chief Administrator, DEA/NFLIS Database

Date: January 28, 2019

Subject: Some insights and advices for countering the opioids crisis

Dear Chief Administrator, we are honored to inform some insights and results we identify during our modeling effort.

The United States is experiencing a national crisis regarding the use of synthetic and non-synthetic opioids. We analyze the data about the the opioids crisis and construct **Combined Spread Model** to describe the spread the spread and characteristics of synthetic opoids and heroin cases in and between five states: Ohio, Kentucky, West Virginia, Virginia and Pennsylvania, and their counties.

Using our **Combined Spread Model**, we can see that in general the number of heroin cases increase from 2010 to 2017, however, if we look at certain counties, the number of heroin cases decrease 2017. But for synthetic opioids, most of them increase dramatically from 2016 to 2017. For both heroin and synthetic opioids, there are only several counties that have tremendous quantities of cases, and most counties possess a relatively small amount of cases. **Cuyahoga (OH), Hamilton (OH), Allegheny (PA) and Philadelphia (PA)** are four counties that almost have the largest number of both heroin cases and synthetic opioids cases, while **Montgomery (OH)** counties has a great amount of synthetic opioids cases but the heroin cases are relatively less. And we can also know that **Ohio and Pennsylvania** are two states that drug abuse situations are the most serious.

We put forward three concerns for U.S. government to pay attention to:

(1) For counties that only begin to abuse drugs, we should kill this phenomenon in the cradle;

(2) For counties that the number of drug reports increase rapidly, we should limit their acceleration and do not let them goes too swiftly;

(3) For counties that already have a great number of drug reports, we should isolate them and do not let them spread drugs to other counties.

Corresponding to the three concerns above, we set three threshold levels:

(1) **Critical threshold level**: the threshold whether one county begins to abuse drugs;

(2) **Alarm threshold level**: the threshold whether one county gets to the fastest increasing speed;

(3) **Diffusion threshold level**: the threshold whether one county arrives at its peak number of drug reports.

After we adopt several statistical methods to analyze the U.S. Census Bureau socio-economic data, we conclude the following variables that are significant to the spread of opioids, shown in table (**0-1**).

We analyze the variables above, and set the following strategy for countering the opioid crisis: Improve the **education** level of these five states, make sure that

Table **0-1**: Significant Variables Opoids Cases

Number	Variables
1	Estimate; HOUSEHOLDS BY TYPE - Nonfamily households - Householder living alone - 65 years and over
2	Estimate; RESIDENCE 1 YEAR AGO - Abroad
3	Estimate; GRANDPARENTS - Who are married
4	Estimate; SCHOOL ENROLLMENT - College or graduate school
5	Estimate; EDUCATIONAL ATTAINMENT - 9th to 12th grade, no diploma
6	Estimate; EDUCATIONAL ATTAINMENT - Associate's degree
7	Estimate; EDUCATIONAL ATTAINMENT - Percent high school graduate or higher
8	Estimate; LANGUAGE SPOKEN AT HOME - Language other than English - Speak English less than "very well"
9	Estimate; ANCESTRY German
10	Estimate; DISABILITY STATUS OF THE CIVILIAN NON INSTITUTIONALIZED POPULATION - With a disability
11	Estimate; FERTILITY - Number of women 15 to 50 years old who had a birth in the past 12 months
12	Estimate; HOUSEHOLDS BY TYPE - Households with one or more people 65 years and over
13	Estimate; VETERAN STATUS - Civilian veterans
14	Estimate; RELATIONSHIP - Other relatives

all the students (even adults) accept high school education (if accepted, college education), and popularize the knowledge of synthetic opioids and heroin. The effectiveness test shows that our strategy is successful.

We sincerely hope that by the efforts of U.S. government and all American citizens, one day the opioids crisis will be put through.

Please contact us if you have any problems.

1 Introduction

1.1 Background

Between 2012 and 2016, the number of U.S. overdose cases involving synthetic opioids increased by nearly 640 percent. Over 42,000 Americans died from overdosing on opioids in 2016 alone, and preliminary data for 2017 is even more alarming[1]. On October 26, 2017, President Trump announced that his Administration was declaring the opioid crisis a national Public Health Emergency under federal law, effective immediately. The opioid crisis is in a grim situation and needs to be solved urgently.

1.2 Problem Restatement

For part 1, we should build a mathematical model to demonstrate the spread and characteristics of synthetic opioid and heroin incidents separately. The range of spread and characteristics are not only in Ohio, Kentucky, West Virginia, Virginia and Pennsylvania as well as their counties, but also between these states and their counties. Next we will use this model to identify the possible starting locations of each specific opioid. In addition, we will use this model to predict the future spread of opioids abuse and set threshold levels to identify some concerns' occurrence that the U.S. government should pay attention to. We will also predict when and where these concerns are going to occur.

For part 2, we will analyze which parameters in U.S. Census socio-economic data have significant influence on the use or trends-in-use of opioid. Then take these parameters into consideration and modify the model in part 1.

For part 3,we will figure out a strategy for fighting the opioid crisis and use our model to test whether this strategy is effective. During the testing process, we will also try to find which parameters have critical impact on whether the strategy is successful.

Finally, we will write a memo to Chief Administrator, DEA/NFLIS Database to summarize the results and significant findings during the course of mathematical modeling.

1.3 Assumptions

- Assume that the county location data are correct as provided in NFLIS Data;
- The development of the forecast objective things belongs to the gradual type without jumping change;
- The factors that influence the development of objective things in the past and at present also determine the development of future things;
- Since we will use ordinary least square algorithm (OLS) to process the data, we make the following assumptions to guarantee that OLS is able to apply[2].
    (1) Assume that the cases of opioid is linear to the variables in U.S. Census Bureau Socio-economic Data;

(2) In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others;
(3) For each year, the expected value of the error, given the explanatory variables for all time periods, is zero;
(4) Conditional on all variables, the variance of error is the same for all years;
(5) Conditional on all variables, the errors in two different time periods are uncorrelated.

1.4 Nomenclature

We list the main notations here, and some of the notations will be explained through the modeling process.

Table **1-1**: Nomenclature

Number	Symbol	Meaning
1	E	estimated value in U.S. Census Bureau Data
2	E_m	estimate margin of error in U.S. Census Bureau Data
3	E_u	upper limit of estimated value
4	E_l	lower limit of estimated value
5	P	percent value in U.S. Census Bureau Data
6	P_m	percent margin of error in U.S. Census Bureau Data
7	P_u	upper limit of percent value
8	P_l	lower limit of percent value
9	α	data smoothing factor
10	β	trend smoothing factor
11	γ	periodic smoothing factor
12	$\vec{s}$	smoothing value vector
13	$\vec{x}$	Data vector, number of cases in a particular county
14	$\vec{t}$	trend value vector
15	$\vec{p}$	periodic value vector
16	Z	the number of cases in all five states
17	x	the longitude of each county
18	y	the latitude of each county
19	t	time or year
20	Q	a parameter used in the model
21	k	a parameter used in the model
22	Z_w	the whole model for the first model built in part 1
23	Z_c	the combined model for the first model built in part 1
24	Z_{cm}	the modified model for the first model built in part 1
25	X	the results matrix of time series model
26	α	a parameter determined by PCA

2 Data Processing

2.1 NFLIS Data Processing

Firstly, we split data for synthetic opioids and heroin from MCM_NFLIS_Data.xlsx file. We learn that opioids are typically divided into three categories: non-synthetic opioids, semi-synthetic opioids and synthetic opioids. Since the examples of non-synthetic opioids in problem C's glossary contain heroin - a semi-synthetic opioid, we categorize semi-synthetic opioids as non-synthetic opioids. As far as we could learn from pharmaceutical and biological resources, we believe that non-synthetic opioids include only Morphine, Codeine, Thebaine, Heroin, Oxymorphone (Opana), Hydrocodone (Vicodin, Lortab, Lorcet), Oxycodone (OxyContin, Oxecta, Roxicodone) and Hydromorphone (Dilaudid, Exalgo).

Then we split data for heroin by screen "SubstanceName" column to find which row of the "SubstanceName" column is equal to "Heroin". To split data for synthetic opioids, we screen "SubstanceName" column which row of the "SubstanceName" column is not equal to the any one of the substances of non-synthetic opioids listed above.

Secondly, we sum the reports of synthetic opioids for each counties and states by year. Since we do not need to sum the reports of heroin, we merely sum the reports of heroin for each of the five states.

Thirdly, in order to analyze and display the reports of drugs in and between all five states and their counties geographically, we search for the **longitude and latitude data** for each county and add them to heroin data files and synthetic opioids data files.

2.2 U.S. Census Bureau Socio-economic Data Processing

Firstly, we split data from seven files named ACS_xx_5YR_DP02_with_ann.csv into 28 (4 kinds of data $\times$ 7 years) subfiles by selecting the following columns: Estimate, Estimate margin of error, Percent and Percent margin of error. Then we use these files to calculate the upper limit and lower limit of estimated value and percent value, separately. Denote estimated value, estimate margin of error, percent value, percent margin of error as E, E_m, P and P_m, respectively. We use equation (2-1) to calculate the upper limit and the lower limit:

$$\begin{cases} E_u = E + E_m \\ E_l = E - E_m \\ P_u = P + P_m \\ P_l = P - P_m \end{cases} \tag{2-1}$$

where the subscript u represents the upper limit and subscript l represents the lower limit. Then we get 42 (6 kinds of data $\times$ 7 years) data files for subsequent analysis.

Secondly, we take each state as a whole and sum the data of its counties for every year.

Finally we obtain all the data required for further analysis, with formation convenient for programming.

3 Part 1

3.1 Initial Data Analysis

To get some intuitive images about the distribution and characteristics of synthetic opioids and heroin incidents in and between the five states and their counties, we plot the number of each county's drug reports with respect to its longitude and latitude. Since the data is discrete, when we draw the figures, we adopt linear interpolation to make the surface smooth. For example, the following figure (**3-1**) and figure (**3-2**) are the distribution of heroin incidents in five states in 2010 and 2017. Figures of other years and distribution of synthetic opioids cases are presented in appendix A.

From figure (**3-1**) to figure (**3-2**), figure (**A-1**) to figure (**A-6**) and figure (**A-7**) to figure (**A-14**), we can immediately learn that there are several sources of heroin cases and synthetic opioids cases. The sources are probably the counties that correspond to the peak of each extreme value in the figures.

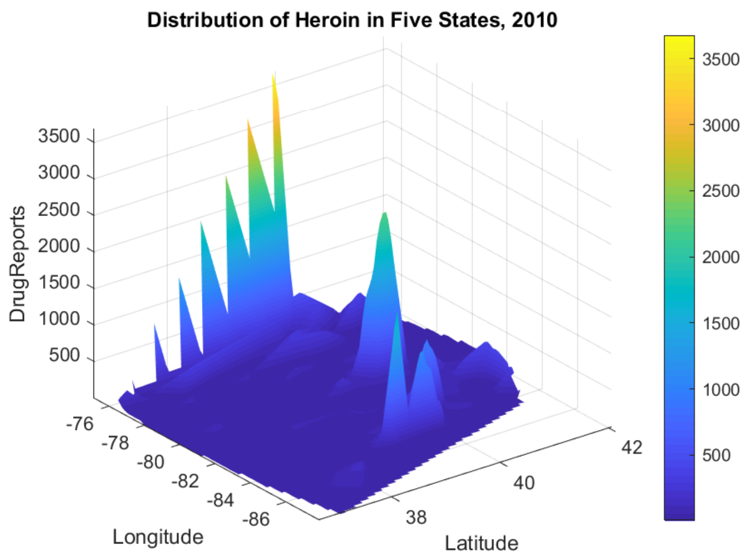

Figure **3-1**: Distribution of heroin cases in five states, 2010

3.2 Spread Model Construction

In order to describe the spread and characteristics of the reported synthetic opioid and heroin incidents in and between five states and their counties, we use the combination of two models to set up the spread model of opioids.

Firstly, we assume that the spread of opioids is like the diffusion of smoke during explosion, and derive a formula from **Smoke Diffusion Model** that fits this problem[3]. Let the time when there is one case in a particular county be 0,

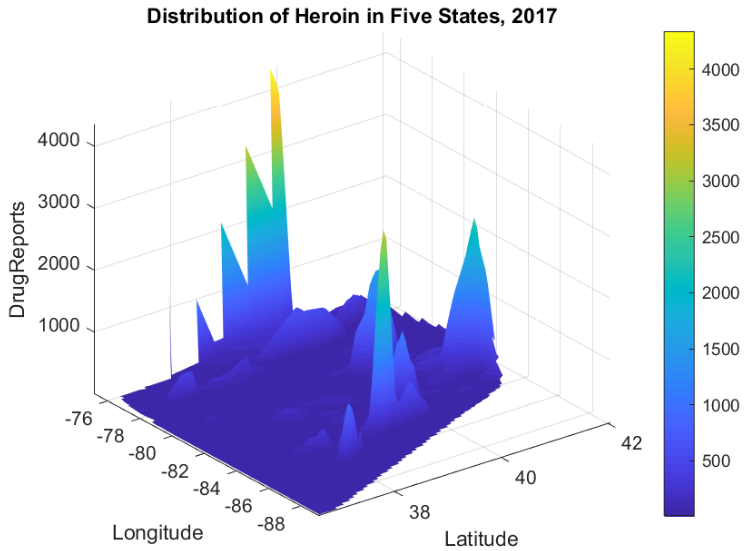

Figure **3-2**: Distribution of heroin cases in five states, 2017

denoted $t = 0$. The position of this county is denoted as (x_0, y_0), $x_0,\ y_0$ represent longitude and latitude, respectively. At time t, the number of cases in all five states is denoted as $Z(x, y, t)$. Analogous to the diffusion equation for smoke, we derive the following formula (3-1):

$$Z(x, y, t) = \frac{Q}{(kt)^{3/2}} \cdot \exp\left[-\frac{(x - x_0)^2 + (y - y_0)^2}{kt}\right] \tag{3-1}$$

where $Q,\ k$ are two parameters determined by the practical data.

For more than one source, namely n sources, we denote each starting location as $(x_{is}, y_{is}),\ i = 1, 2, \cdots, n$. For each source, the spread equation is:

$$Z_i(x, y, t) = \frac{Q_i}{(k_i \cdot t)^{3/2}} \cdot \exp\left[-\frac{(x - x_{is})^2 + (y - y_{is})^2}{k_i \cdot t}\right], \quad i = 1, 2, \cdots, n. \tag{3-2}$$

Therefore, the whole spread model of the opioids is explained by equation (3-3):

$$\begin{aligned} Z_w &= \sum_{i=1}^{n} Z_i(x, y, t) \\ &= \sum_{i=1}^{n} \frac{Q_i}{(k_i \cdot t)^{3/2}} \cdot \exp\left[-\frac{(x - x_{is})^2 + (y - y_{is})^2}{k_i \cdot t}\right]. \end{aligned} \tag{3-3}$$

To calculate Q_i and k_i parameters, we first look at equation (3-2). Take logarithm of both sides of the equation (3-2), we get

$$\ln Z_i(x, y, t) = \ln Q_i - \frac{3}{2} \ln k_i \cdot t - \frac{(x - x_{is})^2 + (y - y_{is})^2}{k_i \cdot t}. \tag{3-4}$$

We first find starting locations of every specific opioid. The method how to find resources will be explained in section (3.4). Substituting $x = x_{is}$ and $y = y_{is}$, we immediately get a relationship between Q_i and k_i. Then we use points around the source to get auxiliary equations. Solving those equations, we can fit to get Q_i and k_i parameters and equations for every specific source. Summing all the equations, we can obtain a whole model for demonstrating spread and characteristics of heroin and synthetic opioids cases in and between five states and their counties.

Secondly, Since the main problem need to be solved in Part1 is to describe the time-varying relationship between the propagation and characteristics of reported synthetic opioids and heroin events in various counties, time series is considered as another model to further analyze the location and future situation.

Considering the trend and periodicity of the data in the model, the third exponential smoothing model, **Holt-Winter exponential smoothing model**, is used[4].

The cubic exponential smoothing algorithm is based on the recursive relationship between the first and second exponential smoothing. The first exponential smoothing algorithm is based on the following recursive relationship:

$$s_i = \alpha \cdot x_i + (1-\alpha) \cdot s_{i-1} \tag{3-5}$$

Where $\vec{s}$ is the smoothing value vector, $\vec{x}$ is data vector, number of cases in a particular county. The subscription i represents the ith value of the vector. The equation (3-5) shows that as α goes to 1, the smoothed value goes to the data value of the current time and the data is more uneven. And the closer α is to 0, the smoother the data is.

Quadratic exponential smoothing retains the trend information and adds a new variable to represent the trend after smoothing. The recurrence relationship is as follows:

$$\begin{cases} s_i = \alpha \cdot x_i + (1-\alpha) \cdot (s_{i-1} + t_{i-1}) \\ t_i = \beta \cdot (s_i - s_{i-1}) + (1-\beta) \cdot t_{i-1} \end{cases} \tag{3-6}$$

Where $\vec{t}$ is trend value vector.

The cubic exponential smoothing retains the periodic information on the basis of the quadratic exponential smoothing, which makes it possible to predict the time series with periodicity. The cumulative cubic exponential smoothing is used here.

$$\begin{cases} s_i = \alpha \cdot (x_i - p_{i-1}) + (1-\alpha) \cdot (s_{i-1} + t_{i-1}) \\ t_i = \beta \cdot (s_i - s_{i-1}) + (1-\beta) \cdot t_{i-1} \\ p_i = \gamma \cdot (x_i - s_i) + (1-\gamma) \cdot p_{i-1} \end{cases} \tag{3-7}$$

where $\vec{p}$ is periodic value vector; $\alpha,\ \beta,\ \gamma \in [0,1]$, and they are decided by the data provided, and can be adjusted according to practical situation.

The prediction equation is:

$$x_{i+h} = s_i + h \cdot t_i + p_{i+h-1}, \quad h = 1, 2, \cdots \tag{3-8}$$

Since the memory ability of exponential smoothing method is very short, the influence of initial value will become very weak. We assign the following initial value:

$$\begin{cases} s_1 = x_1 \\ t_1 = x_2 - x_1 \\ p_1 = 0 \end{cases} \tag{3-9}$$

Then we can use the latitude, longitude and year corresponding to $\vec{x}$ to make up a matrix $X(x, y, t) = \vec{x}(t)$. Now since we have constructed two models, we combine them into one **Spread Model** using equation (3-10) to get more accurate and convincing results.

$$Z_c = \frac{Z_w + X}{2} \tag{3-10}$$

3.3 Spread and Characteristics

Since plotting the number of reports of drugs with respect to latitude, longitude and time requires four dimensions, which is not able to bring an intuitive image of how opioids spread in and between the five states, we plot the number of drug reports with respect to longitude and latitude separately, shown in figure (**3-3**) to figure (**3-6**).

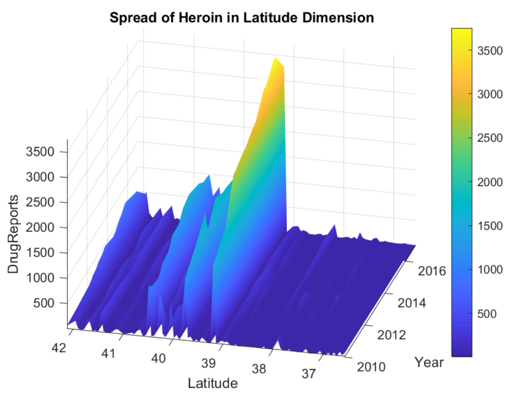

Figure **3-3**: Heroin Spread, Latitude

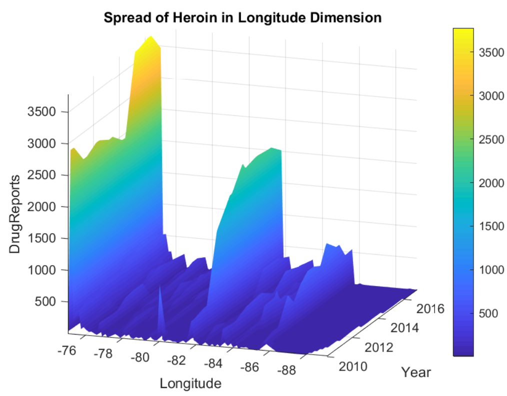

Figure **3-4**: Heroin Spread, Longitude

Taking some counties as the center, it gradually spread outward to more counties. At the beginning, there are more heroin reports in some counties (more than 100 cases). Then it is found that the number of heroin reports in neighboring counties will increase year by year, and the situation of synthetic opioids is similar: some counties start drug abuse earlier and neighbors are very likely to follow.

We can see that in general the number of heroin cases increase from 2010 to 2017, however, if we look at certain counties, the number of heroin cases decrease 2017. But for synthetic opioids, most of them increase dramatically from 2016 to 2017. For both heroin and synthetic opioids, there are only several counties that have tremendous quantities of cases, and most counties possess a relatively small amount of cases. **Cuyahoga (OH), Hamilton (OH), Allegheny (PA) and**

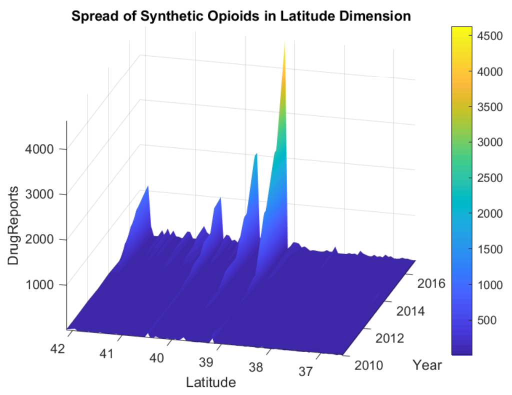

Figure **3-5**: Opioids Spread, Latitude

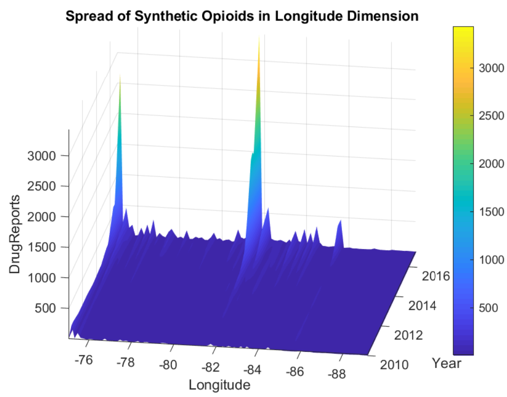

Figure **3-6**: Opioids Spread, Longitude

Philadelphia (PA) are four counties that almost have the largest number of both heroin cases and synthetic opioids cases, while **Montgomery (OH)** counties has a great amount of synthetic opioids cases but the heroin cases are relatively less. And we can also know that **Ohio and Pennsylvania** are two states that drug abuse situations are the most serious.

3.4 Identify Starting Locations

We construct a **Extreme Value Model** to find the starting locations[5]. Let Z be the number of cases of heroin or synthetic opioids, $Z = Z(x, y)$ is a function of $x,\ y$ at a specific year. $x,\ y$ are latitude and longitude of each county respectively. We see $Z(x_i, y_i)$ as a source if it satisfies the following equation (3-11):

$$\begin{cases} Z(x_i, y_i) \geqslant Z(x_i, y_{i+1}) \\ Z(x_i, y_i) \geqslant Z(x_i, y_{i-1}) \\ Z(x_i, y_i) \geqslant Z(x_{i+1}, y_i) \\ Z(x_i, y_i) \geqslant Z(x_{i-1}, y_i) \\ Z(x_i, y_i) \geqslant Z(x_{i+1}, y_{i+1}) \\ Z(x_i, y_i) \geqslant Z(x_{i-1}, y_{i+1}) \\ Z(x_i, y_i) \geqslant Z(x_{i+1}, y_{i-1}) \\ Z(x_i, y_i) \geqslant Z(x_{i-1}, y_{i-1}) \end{cases} \tag{3-11}$$

We use this method to find sources (x_i, y_i) year by year for a specific opioid, and integration all the sources found in one or more years, making up a set $\{X, Y\}$ that contains all the possible starting locations.

The starting locations of heroin and synthetic opioids are shown in figure (**3-7**) and (**3-8**), respectively.

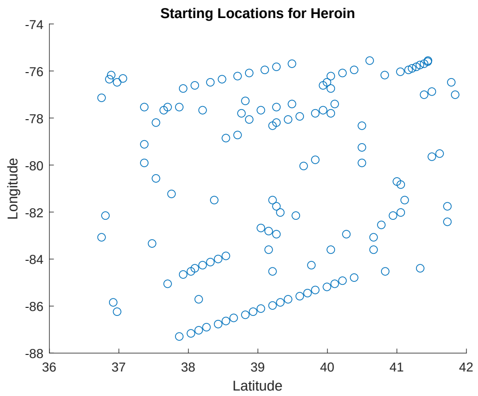

Figure **3-7**: Sources of Heroin

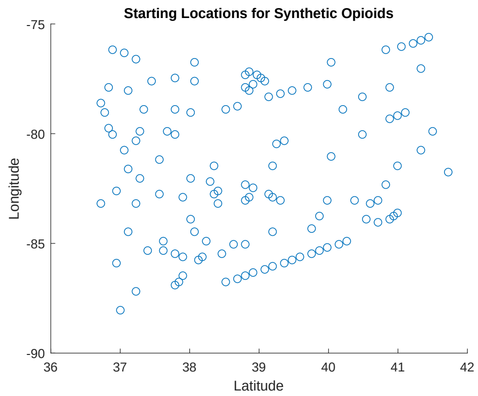

Figure **3-8**: Sources of Opioids

3.5 Prediction

3.5.1 Government Concerns

Observing the trends of drug reports in figure (**3-3**) to figure (**3-6**), we find that there exist many counties that the number of drug reports are increasing year by year and there are even some counties that the number increasing dramatically during recent years. Besides, there are some counties that the number of reports are too large that exceed several thousand. In the meantime, there are also many counties that just begin to have drug reports.

By the analysis above, we put forward three concerns for U.S. government to pay attention to:

(1) For counties that only begin to abuse drugs, we should kill this phenomenon in the cradle;

(2) For counties that the number of drug reports increase rapidly, we should limit their acceleration and do not let them goes too swiftly;

(3) For counties that already have a great number of drug reports, we should isolate them and do not let them spread drugs to other counties.

3.5.2 Threshold Levels

Corresponding to the three concerns above, we set three threshold levels:

(1) **Critical threshold level**: the threshold whether one county begins to abuse drugs;

(2) **Alarm threshold level**: the threshold whether one county gets to the fastest increasing speed;

(3) **Diffusion threshold level**: the threshold whether one county arrives at its peak number of drug reports.

We use equation (3-1) to calculate the three threshold levels. We first find out the **diffusion threshold level**. The condition is that the increasing speed is zero.

Let the derivative of equation (3-1) over t be zero, we get

$$\begin{aligned}\frac{\partial Z}{\partial t} &= \frac{\partial}{\partial t}\left[\frac{Q}{(kt)^{3/2}}\mathrm{e}-\frac{(x-x_0)^2+(y-y_0)^2}{kt}\right] \\ &= -\frac{3Qk}{2\,(kt)^{5/2}}\mathrm{e}^{-\frac{(x-x_0)^2+(y-y_0)^2}{kt}} + \frac{Q\left[(x-x_0)^2+(y-y_0)^2\right]}{(kt)^{3/2}\,kt^2}\mathrm{e}^{-\frac{(x-x_0)^2+(y-y_0)^2}{kt}} \\ &= 0\end{aligned} \tag{3-12}$$

Solving equation (3-12), we get

$$t_1 = \frac{2}{3}\cdot\frac{(x-x_0)^2+(y-y_0)^2}{k} \tag{3-13}$$

Then we decide the **alarm threshold level**. calculate the second derivative of equation (3-1):

$$\begin{aligned}&\frac{\partial^2}{\partial t^2}\left[\frac{Q}{(kt)^{3/2}}\mathrm{e}^{-\frac{(x-x_0)^2+(y-y_0)^2}{kt}}\right] \\ &= \frac{15\,Qk^2}{4\;(kt)^{7/2}}\mathrm{e}^{-\frac{(x-x_0)^2+(y-y_0)^2}{kt}} \\ &-3\,\frac{Q\left[(x-x_0)^2+(y-y_0)^2\right]}{(kt)^{5/2}\,t^2}\mathrm{e}^{-\frac{(x-x_0)^2+(y-y_0)^2}{kt}} \\ &-2\,\frac{Q\left[(x-x_0)^2+(y-y_0)^2\right]}{(kt)^{3/2}\,kt^3}\mathrm{e}^{-\frac{(x-x_0)^2+(y-y_0)^2}{kt}} \\ &+\frac{Q\left[(x-x_0)^2+(y-y_0)^2\right]^2}{(kt)^{3/2}\,k^2t^4}\mathrm{e}^{-\frac{(x-x_0)^2+(y-y_0)^2}{kt}}\end{aligned} \tag{3-14}$$

Let equation (3-14) be zero, we get

$$t_2 = 2\cdot\left(\frac{1}{3}\pm\frac{1}{15}\right)\cdot\frac{(x-x_0)^2+(y-y_0)^2}{k} \tag{3-15}$$

The time t first arrives at smaller value, thus the bigger value is unrealistic. Adopt minus sign, we get

$$t_2 = \frac{8}{15}\cdot\frac{(x-x_0)^2+(y-y_0)^2}{k} \tag{3-16}$$

Finally, we figure out the **critical threshold level**. Here in our models, Z from 0 to 1 means that one county begin to abuse opioids. Let $Z = 1$ in equation (3-1), we get

$$0 = \ln Q - \frac{3}{2}\ln k\cdot t - \frac{(x-x_0)^2+(y-y_0)^2}{k\cdot t}. \tag{3-17}$$

Solving this equation, we get

$$t_3 = \frac{k \ln Q \pm \sqrt{\Delta}}{3k \ln k}, \tag{3-18}$$

where

$$\Delta = (k \ln Q)^2 - 6k \ln k \cdot \left[(x - x_0)^2 + (y - y_0)^2\right]. \tag{3-19}$$

Here we should choose the smaller t_3, thus using minus sign.

$$t_3 = \frac{k \ln Q - \sqrt{\Delta}}{3k \ln k}, \tag{3-20}$$

The threshold levels are

$$\begin{cases} Z = Z(x, y, t_1), & \text{diffusion threshold level,} \\ Z = Z(x, y, t_2), & \text{alarm threshold level,} \\ Z = Z(x, y, t_3), & \text{critical threshold level,} \end{cases} \tag{3-21}$$

where

$$\begin{cases} t_1 = \dfrac{2}{3} \cdot \dfrac{(x - x_0)^2 + (y - y_0)^2}{k}, \\ t_2 = \dfrac{8}{15} \cdot \dfrac{(x - x_0)^2 + (y - y_0)^2}{k}, \\ t_3 = \dfrac{k \ln Q - \sqrt{\Delta}}{3k \ln k}. \end{cases} \tag{3-22}$$

Then for every county, we substitute longitude and latitude into equation (3-22) to get the time $t_i,\ i = 1,\ 2,\ 3$. Substituting $x,\ y,\ t_i$ into equation (3-21), we obtain the threshold levels, respectively. For example, the alarm threshold levels of heroin and synthetic opioids are shown in figure (**3-9**) and figure (**3-10**).

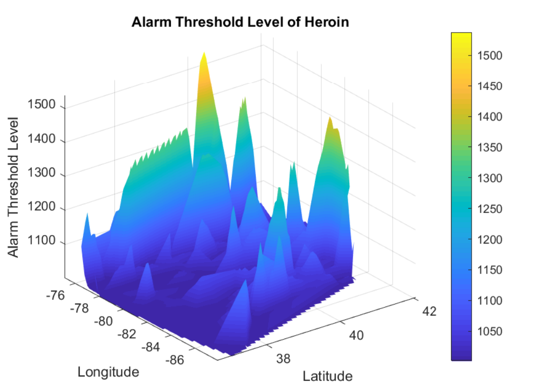

Figure **3-9**: Alarm Threshold, Heroin

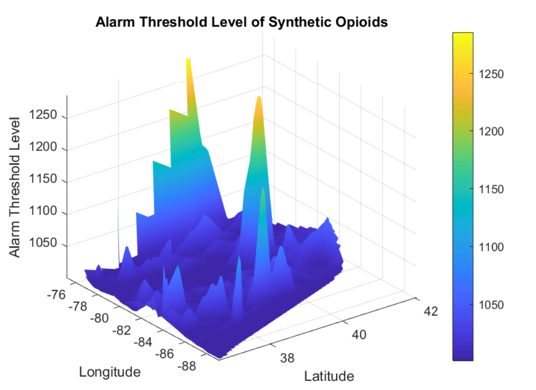

Figure **3-10**: Alarm Threshold, Opioids

3.5.3 Places and Time

Using equation (3-21), equation (3-22) and equation (3-10), we are able to obtain the time and places of all specific opioids for diffusion threshold level, alarm threshold level and critical threshold level. For example, the time and places of heroin and synthetic opioids cases for alarm threshold level are shown in figure (**3-11**) and figure (**3-12**).

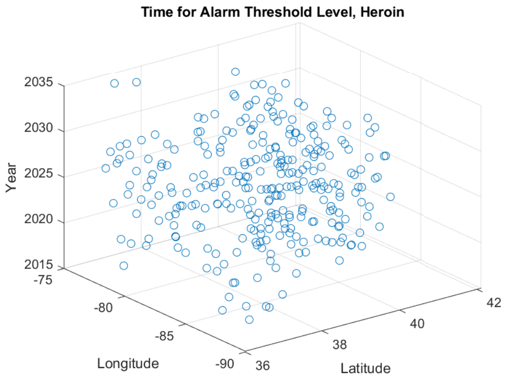

Figure **3-11**: Time and Places, Heroin

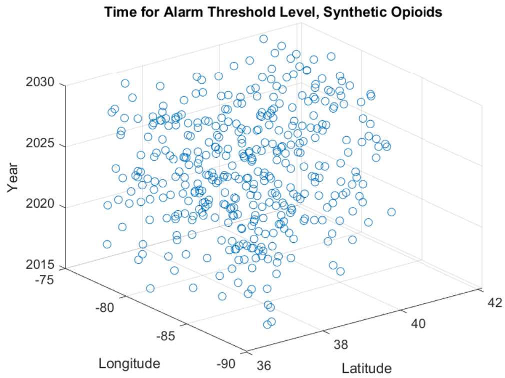

Figure **3-12**: Time and Places, Opioids

4 Part 2

4.1 Initial Data Analysis

In the data set provided for part 2, there are some percentage data have some error of input, we delete them all.

4.2 Model Construction

Firstly, the type of the data is long panel, since the number of states n is less than the number of years T. In order to control individual effects, generate the dummy variables of state[7]:

$$\text{State 2} = \begin{cases} 1, & \text{State OH;} \\ 0, & \text{State KY} \end{cases} \tag{4-1}$$

$$\text{State 3} = \begin{cases} 1, & \text{State PA;} \\ 0, & \text{State KY} \end{cases} \tag{4-2}$$

$$\text{State 4} = \begin{cases} 1, & \text{State VA;} \\ 0, & \text{State KY} \end{cases} \tag{4-3}$$

$$\text{State 5} = \begin{cases} 1, & \text{State WV;} \\ 0, & \text{State KY} \end{cases} \tag{4-4}$$

Secondly, T is a only little bigger than n, so T can not provide enough information to estimate autoregressive coefficient of every panel, so we assume they are all equal[8]. Use time trend variable t to consider about the time effect.

Thirdly, we choose variables. The provided U.S. Census Bureau socio-economic data has a lot of variables which makes analysis of these data sophisticated. Observing all the variables, we find that every four columns data refers to one variable except for the first three variables. Then we subtract one fourth data of different variables to analyze. By the meaning of each variable, we divide all variables into 16 categories, shown in table (**4-1**).

Table **4-1**: Variable Categories in U.S. Census Bureau Data

Number	Variable Categories
1	Geography
2	Households by type
3	Relationship
4	Marital status
5	Fertility
6	Grandparents
7	School enrollment
8	Educational attainment
9	Veteran status
10	Disability status of the civilian non institutionalized population
11	Residence 1 year ago
12	Place of birth
13	U.S. citizenship status
14	Year of entry
15	World region of birth of foreign born
16	Language spoken at home
17	Ancestry

In addition, the variables also include point estimate, percent estimate, the standard error of point estimate and the standard error of percent estimate. There are so many variables, so we use three steps to screen variables, the steps are as follows.

(1) The model of estimating two-way fixed effects using LSDV[9]

Consider all the data of economic and social census into regression as variables. It's a fixed effect model not a stochastic effect model, since we consider the dummy variables. Omit variables which have obvious multicollinearity or have very small regression estimation parameters. We use clustering robustness standard coefficient (considering autocorrelation of different disturbance terms of the same state), but there are too many variables, we can't calculate t value and p value.

(2) Stepwise Regression

Using Stepwise Regression carrying out surplus variablesthe backward elimination procedure begins with a model that includes all the independent variables. It then deletes one independent variable at a time using the same procedure as stepwise regression. Set AIpha = 0.05.

(3) Best-Subsets Regression[10]

Using Best-Subsets Regression to carry out remain variables. Consider both reducing variables and making R-Sq bigger to get final variables.

Fourthly, we test long panel data about heteroscedasticity and autocorrelation:

(1) Test of heteroscedasticity between groups

We first make an assumption: $H_0 : \sigma_i^2 = \sigma^2,\ i = 1, 2, \cdots, n$. Using Wald test[11], the result strongly rejects the assumption of homogeneity between groups, illustrating that different states have different situations.

(2) Intra group autocorrelation

Under the original assumption that there do not exist intra group autocorrelation, the variance and autocovariance of the disturbance term $\Delta\epsilon_{it}$ are:

$$Var(\Delta\epsilon_{it}) = Var(\epsilon_{it} - \epsilon_{i,t-1}) = Var(\epsilon_{it}) + Var(\epsilon_{i,t-1}) = 2\sigma_\epsilon^2 \tag{4-5}$$

$$\begin{aligned} Cov(\Delta\epsilon_{it}, \Delta\epsilon_{i,t-1}) &= Cov(\epsilon_{it} - \epsilon_{i,t-1}, \epsilon_{i,t-1} - \epsilon_{i,t-2}) \\ &= -Cov(\epsilon_{i,t-1}, \epsilon_{i,t-1}) \\ &= -Var(\epsilon_{i,t-1}) = -\sigma_\epsilon^2 \end{aligned} \tag{4-6}$$

Then the autocorrelation coefficient is:

$$Corr(\Delta\epsilon_{it}, \Delta\epsilon_{i,t-1}) = \frac{Cov(\Delta\epsilon_{it}, \Delta\epsilon_{i,t-1})}{Var(\Delta\epsilon_{it})} = \frac{-\sigma_\epsilon^2}{2\sigma_\epsilon^2} = -0.5 \tag{4-7}$$

Denote the sample value of $\Delta\epsilon_{it}$ as e_{it}, then carry out first order autoregression upon e_{it}:

$$e_{it} = \rho e_{i,t-1} + error_{it},\ i = 1, \cdots, n;\ t = 3, \cdots, T. \tag{4-8}$$

Then we perform Wald test upon "$H_0 :\ \rho = -0.5$". The results strongly reject the original hypothesis that there is no first order intra group autocorrelation.

(3) Cross sectional correlation test

Consider the original hypothesis "there is no component cross-section correlation". If this hypothesis holds, then the correlation coefficients between individual perturbation terms calculated by residual error should be close to zero. If we arrange there correlation coefficients to a matrix, that is, correlation matrix of residuals, then the non-principal diagonal elements are close to 0.

With insufficient evidence to reject H_0, considering the second error is enough small, we accept H_0, illustrating individual disturbance terms are independent, indicating that the disturbance factors of drug use in different states are exogenous and almost unrelated.

Fifthly, we carry out regression. we deal with Intra group autocorrelation's and group synchronization's FGLS at the same time. We find that different individual disturbance terms are relevant but have different variance. There are some autocorrelations having the same autocorrelation coefficients in the same group.

4.3 Model Results

The significant variables for synthetic opoids are shown in table (**4-2**).

Table **4-2**: Significant Variables for Synthetic Opoids Cases

Number	Variables
1	Estimate; HOUSEHOLDS BY TYPE - Nonfamily households - Householder living alone - 65 years and over
2	Estimate; RESIDENCE 1 YEAR AGO - Abroad
3	Estimate; GRANDPARENTS - Who are married
4	Estimate; SCHOOL ENROLLMENT - College or graduate school
5	Estimate; EDUCATIONAL ATTAINMENT - 9th to 12th grade, no diploma
6	Estimate; EDUCATIONAL ATTAINMENT - Associate's degree
7	Estimate; EDUCATIONAL ATTAINMENT - Percent high school graduate or higher
8	Estimate; LANGUAGE SPOKEN AT HOME - Language other than English - Speak English less than "very well"

Then we use the upper limits and the lower limits $E_u,\ E_l,\ P_u,\ P_l$ to repeat the above steps. We find that the results are relevant. Since the standard error/estimate is small, we ignore it.

For heroin cases, we perform the same procedures as synthetic opioids. The significant variables for heroin cases are shown in table (**4-3**).

Since the order of magnitude between other variables and dummy variables is large, resulting that the coefficients of dummy variables are largeso we consider using the percent estimate.

4.4 Modify Model in Part 1

Firstly, we use **Principal Components Analysis (PCA)** to get a modified parameter α for the first model built in part 1.

(1) The main explanatory variables obtained by regression are analyzed by principal component factor analysis, and the factors whose eigenvalue are less than 1 are removed.

(2) Since the factors are correlated, the factors are rotated obliquely.

Table **4-3**: Significant Variables for Heroin Cases

Number	Variables
1	Estimate; ANCESTRY German
2	Estimate; DISABILITY STATUS OF THE CIVILIAN NON INSTITUTIONALIZED POPULATION - With a disability
3	Estimate; FERTILITY - Number of women 15 to 50 years old who had a birth in the past 12 months
4	Estimate; HOUSEHOLDS BY TYPE - Households with one or more people 65 years and over
5	Estimate; VETERAN STATUS - Civilian veterans
6	Estimate; RELATIONSHIP - Other relatives

(3) Using factor scores to standardize each variable to have a zero mean and variance equals to 1. Perform weighted summation upon factor scores coefficients, formating linear composites.

(4) Calculate each factor and use combination weighting approach to get the coefficient α_i for each state at a specific year.

Then we get the first modified spread model:

$$\begin{aligned} Z_{wm} &= \sum_{i=1}^{n}(1+\alpha_i)\cdot Z_i(x,y,t) \\ &= \sum_{i=1}^{n}(1+\alpha_i)\cdot \frac{Q_i}{(k_i\cdot t)^{3/2}}\cdot \exp\left[-\frac{(x-x_{is})^2+(y-y_{is})^2}{k_i\cdot t}\right]. \end{aligned} \tag{4-9}$$

Secondly, we modify the second model built in part 1. We have known from part 2 that the influence factors of synthetic opioids and their correlation coefficients $w_j,\ j=1,2,\cdot,n$, n is the total number of correlation coefficients. We have known from the model of part 1 that the assumption is that the factors that influence the development of objective things in the past and at present also determine the development of future things. In other words, when these influence factors change, it will influence the direction of the whole model. According to the model of part 2, we simplify this question and assume that these changes only influence the value of the trend vector, so the second modified model is as following:

$$\begin{cases} s_i = \alpha\cdot(x_i-p_{i-1})+(1-\alpha)\cdot(s_{i-1}+t_{i-1}) \\ t_i = \beta\cdot(s_i-s_{i-1})+(1-\beta)\cdot t_{i-1} \\ p_i = \gamma\cdot(x_i-s_i)+(1-\gamma)\cdot p_{i-1}+\sum_{j=1}^{n} w_j\Delta m_j \\ x_{i+h} = s_i+h\cdot t_i+p_{i+h-1}, \quad h=1,2,\cdots \end{cases} \tag{4-10}$$

Where w_j means the correlation coefficients of the j^{th} influence factor, and Δm_j means the variation of j^{th} influence factor.

Then we can use the latitude, longitude and year corresponding to $\vec{x}$ to make up a modified matrix $X_m(x, y, t) = \vec{x}(t)$. Now since we have constructed two modified models, we combine them into one **Modified Spread Model** using equation (4-11) to get more accurate and convincing results.

$$Z_{cm} = \frac{Z_{wm} + X_m}{2} \tag{4-11}$$

5 Part 3

5.1 Identify A Possible Strategy

According to part 2, we have known that there are many factors that will influence the values. Among them there are more factors are concerning about education. Therefore we set the following strategy: Improve the **education** level of these five states, make sure that all the students (even adults) accept high school education (if accepted, college education), and popularize the knowledge of synthetic opioids and heroin.

5.2 Effectiveness Test and Significant Parameter Bounds

In the section, we use the number of drug reports as the significant parameter. And we think that when the drug reports number of five states in a year reduce 2% comparing with that without the strategy, we success. That is to say, we set the significant parameter bounds to be 2%. When the strategy becomes effective, it will influence the following four factors:

- Estimate; SCHOOL ENROLLMENT - College or graduate school;
- Estimate; EDUCATIONAL ATTAINMENT - 9th to 12th grade;
- Estimate; EDUCATIONAL ATTAINMENT - Associate's degree;
- Estimate; EDUCATIONAL ATTAINMENT - Percent high school graduate or higher.

Assume that the four factors change similarly, we can get the following figure (**5-1**) to figure (**5-3**).

The X-axis represents the sum of changing values of these four factors, and the Y-axis represents the reducing percentage of the drug reports compared with that without the strategy. It's obvious that as time goes on, the reducing percentages are bigger, thus the influence degree is increasing. In 2018, when the changing values exceed 15, the percentage exceed 2%. In 2019, the critical value is 10 and in 2020 the value is 5. Therefore, according to the judgment basis, this strategy successes.

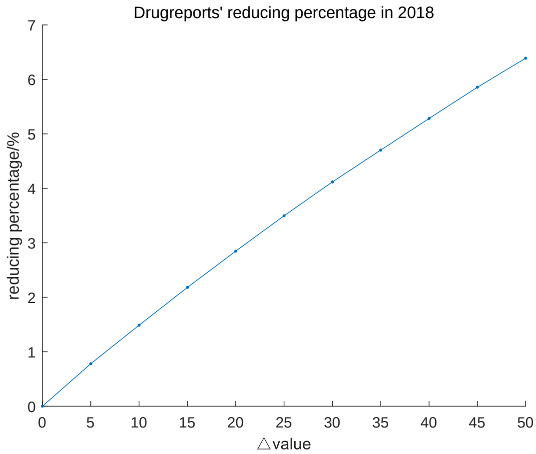

Figure **5-1**: Drug Reports' Reducing Percentage, 2018

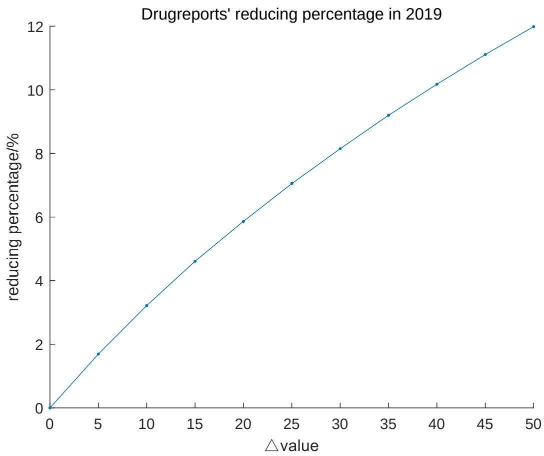

Figure **5-2**: Drug Reports' Reducing Percentage, 2019

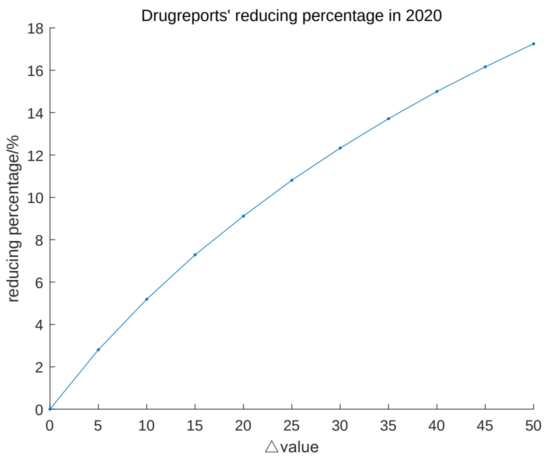

Figure **5-3**: Drug Reports' Reducing Percentage, 2020

6 Evaluation of Models

6.1 Strengths

- We perform **initial data processing**, add **longitude and latitude data** to every county, categorize data into different types, which are beneficial for successive analysis and results display;
- We construct the **Spread Model** with two different models, which is more accurate and convincing;
- We use **Extreme Value Model** to find the starting locations of cases, which is convenient and easy to perform;
- We put forward **three concerns** that U.S. government should concern and corresponding **three threshold levels**: critical threshold level, alarm threshold level and diffusion threshold level;
- We offer a **detailed calculating process** for figuring out the three threshold levels, which is easy to understand and carry out;
- We use the **combination of two models** to predict. The results are closer to the practical data;
- We see data in part 2 as **long panel data** and set **dummy variables**, which increases the accuracy of the analysis;
- We use **LSDV, stepwise regression, best-subsets regression** to analyze the initial data. Then we perform **heteroscedasticity test** and **cross sectional correlation test**. These procedures ensure that what we get from this regression analysis is correct;
- We use **PCA** and **linear combination** to modify our model in part 1 separately. Then we combine the two modified models together to form a **Modified Combined Spread Model**. We take the new factors into considerate thinking.

6.2 Weaknesses

- Without considering all provided data, our model is a simplification of the real situation;
- Our models are a little complicated, which are not easy for others to adopt;
- We make many assumptions during modeling process, which makes our model deviate from the practical situations.

References

[1] U.S. Department of State website, *https://www.state.gov/j/inl/opioid/index.htm*, accessed 27 January 2019.

[2] Jeffrey M. Wooldridge, Michigan State University, *Introductory Econometrics*, six edition, 315-317.

[3] Konstantinos Kotzakoulakis,Simon C. George. Predicting the weathering of fuel and oil spills: *A diffusion-limited evaporation model*[J]. Chemosphere,2018,190.

[4] Leung Dennis H,Kapoor Yash,Alleyne Candice,Walsh Erika,Leithead Andrew,Habulihaz Bahanu,Salituro Gino M,Bak Annette,Rhodes Timothy. *Development of a Convenient In Vitro Gel Diffusion Model for Predicting the In Vivo Performance of Subcutaneous Parenteral Formulations of Large and Small Molecules*.[J]. AAPS PharmSciTech,2017,18(6).

[5] Kotzakoulakis Konstantinos,George Simon C. Predicting the weathering of fuel and oil spills: *A diffusion-limited evaporation model*.[J]. Chemosphere,2018,190.

[6] R. E. Jones,F. S. Gittleson,J. A. Templeton,D. K. Ward. *A Simple Model for Interpreting the ReactionDiffusion Characteristics of Li-Air Batteries*[J]. Journal of the Electrochemical Society,2017,164(1).

[7] Elmansouri, Rachid,Elbeqqali, Omar,Ziyati, Elhoussaine. Normed principal components analysis: *A new approach to data warehouse fragmentation*[P]. ,2013.

[8] Ramahaleomiarantsoa, J.F.,Sambatra, E.J.R.,Heraud, N.,Razafimahenina, J.M.. Faults *diagnosis of wind energy conversion chain based on doubly fed induction generator by principal components analysis method*[P]. ,2013.

[9] Sankar, D Sandeep Vara,ARoy, Lakshi Prosad. *Principal component analysis (PCA) approach to segment primary components from pathological phonocardiogram*[P]. Communications and Signal Processing (ICCSP), 2014 International Conference on,2014.

[10] Lan Yu,Xia Liu. *Debris flow risk assessment using principal components analysis and rough set techniques*[P]. Control Conference (CCC), 2014 33rd Chinese,2014.

[11] Gaskin Cadeyrn J,Happell Brenda. On exploratory factor analysis: *A review of recent evidence, an assessment of current practice, and recommendations for future use*[J]. International journal of nursing studies,2013,51(3).

Appendices

Appendix A Graphs

The figures for the distribution of heroin cases in five states from 2011 to 2016 are shown below.

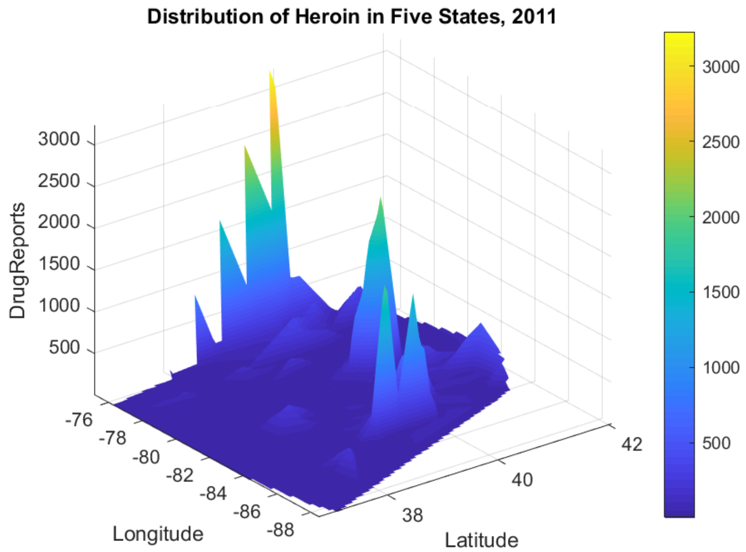

Figure **A-1**: Distribution of heroin cases in five states, 2011

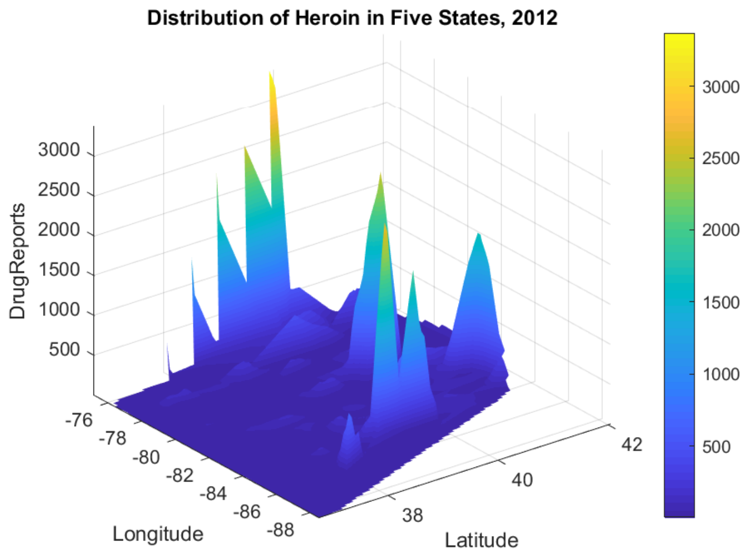

Figure **A-2**: Distribution of heroin cases in five states, 2012

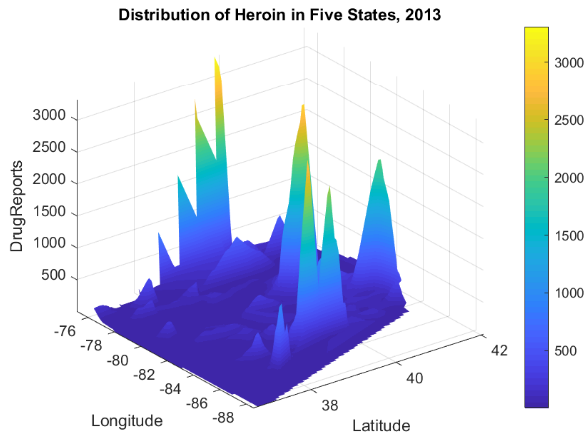

Figure **A-3**: Distribution of heroin cases in five states, 2013

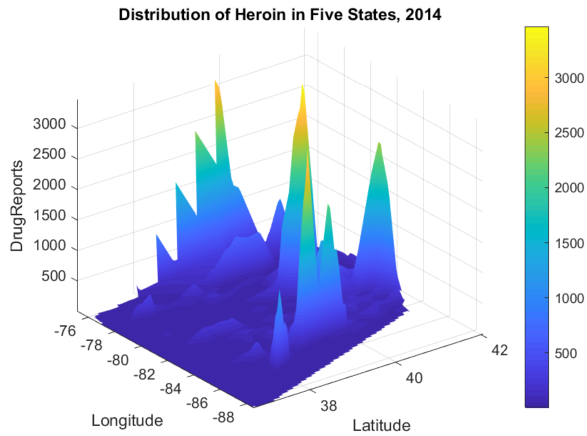

Figure **A-4**: Distribution of heroin cases in five states, 2014

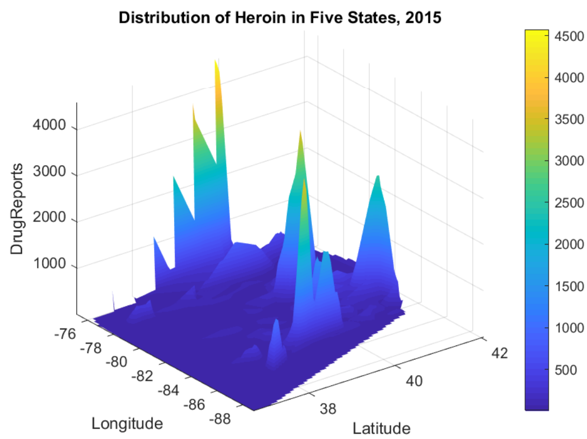

Figure **A-5**: Distribution of heroin cases in five states, 2015

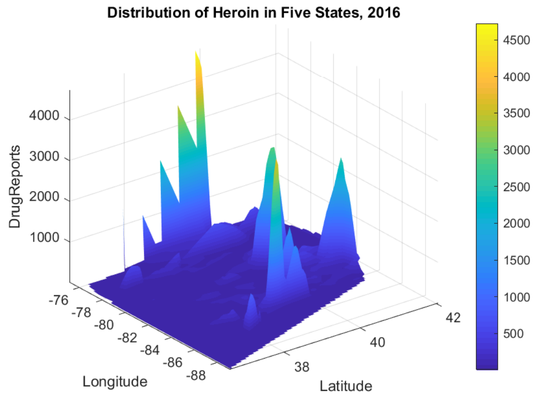

Figure **A-6**: Distribution of heroin cases in five states, 2016

The figures for the distribution of synthetic opioids cases in five states from 2010 to 2017 are shown below.

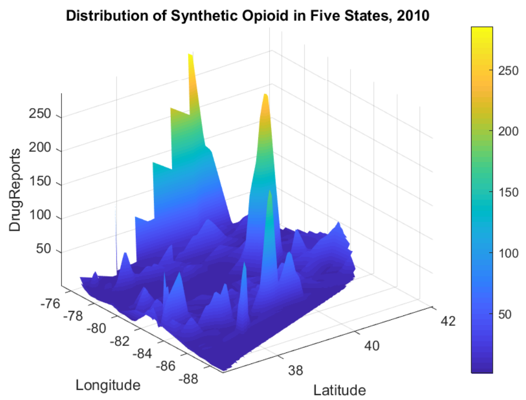

Figure **A-7**: Distribution of synthetic opioids cases in five states, 2010

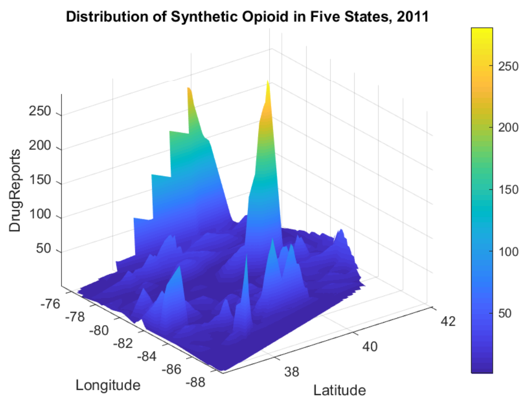

Figure **A-8**: Distribution of synthetic opioids cases in five states, 2011

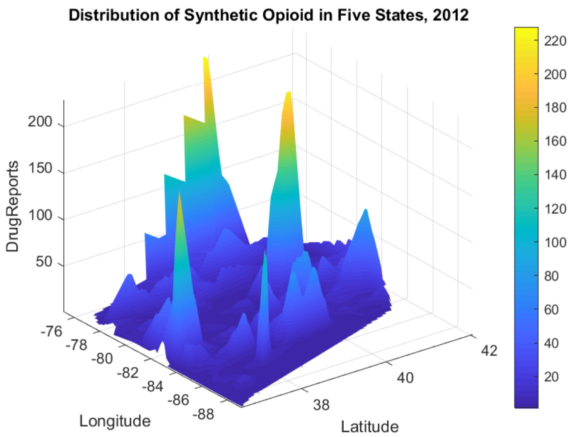

Figure **A-9**: Distribution of synthetic opioids cases in five states, 2012

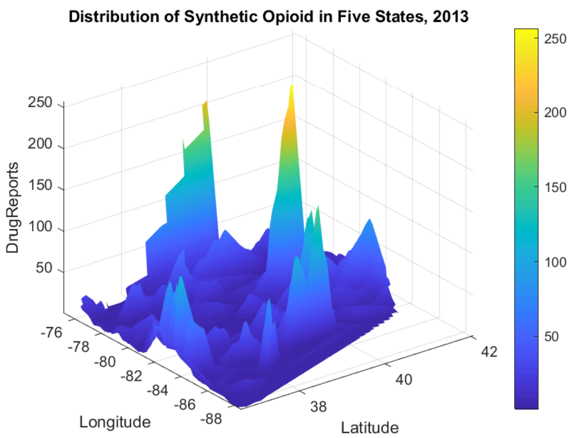

Figure **A-10**: Distribution of synthetic opioids cases in five states, 2013

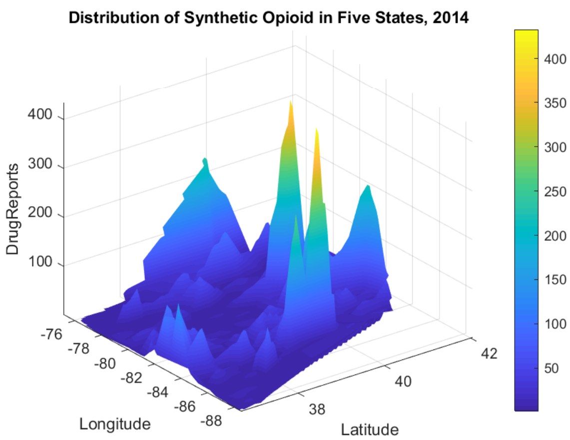

Figure **A-11**: Distribution of synthetic opioids cases in five states, 2014

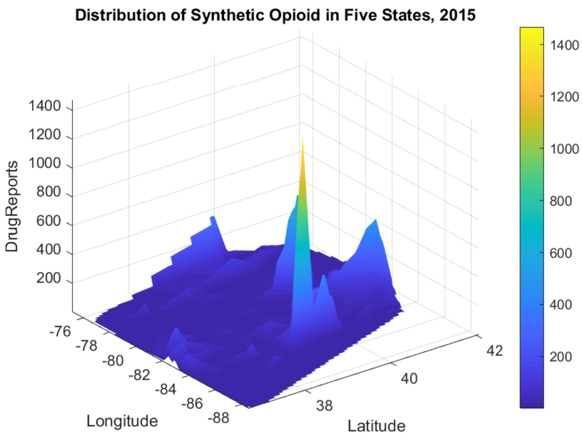

Figure **A-12**: Distribution of synthetic opioids cases in five states, 2015

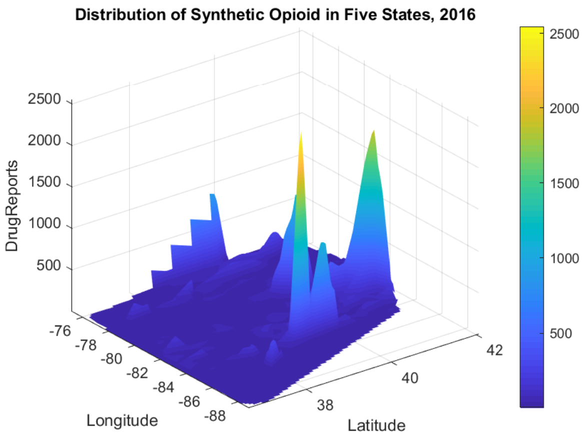

Figure **A-13**: Distribution of synthetic opioids cases in five states, 2016

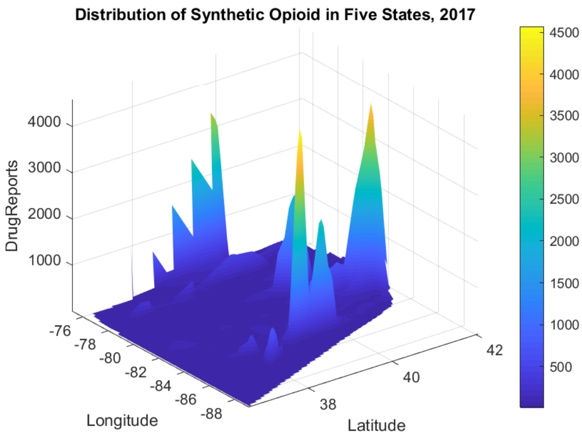

Figure **A-14**: Distribution of synthetic opioids cases in five states, 2017

Appendix B Codes

Here are programs used for our models.

Input matlab source:

```
A = MCMopioidsdata;
As = MCMopioidsdata1;
[m,n] = size(A);
%A1 = {2010 2011 2012 2013 2014 2015 2016 2017};
%A2 = {'KY';'OH';'PA';'VA';'WV'};
%import position variables
KYx = KY(:,9);
KYy = KY(:,10);
[mk,nk] = size(KY);
OHx = OH(:,9);
OHy = OH(:,10);
[mo,no] = size(OH);
PAx = PA(:,9);
PAy = PA(:,10);
[mp,np] = size(PA);
VAx = VA(:,9);
VAy = VA(:,10);
[mv,nv] = size(VA);
WVx = WV(:,9);
WVy = WV(:,10);
[mw,nw] = size(WV);

mtot = [mk,mo,mp,mv,mw];
mtotal = [0,mk,mk+mo,mk+mo+mp,mk+mo+mp+mv];
Stotal = [KY;OH;PA;VA;WV];
[ms,~] = size(Stotal);
Xtotal = [KYx;OHx;PAx;VAx;WVx];
Ytotal = [KYy;OHy;PAy;VAy;WVy];
Sheet = 2010:1:2017;

%strcmpi ignore case
for i = 1:m
   for j = 1:ms
      if isequal(As(i,6),Stotal(j,2))
         As(i,11) = Xtotal(j);
         As(i,12) = Ytotal(j);
      end
   end
end

xlswrite('synthetic_opioids_data.xlsx',As);
```

Input matlab source:

```
%  MCM_C PART1

clear;
clc;

%  set the correlation coefficent of each possible factor of
   synthetic
%  opioid
wo = 0.1006094;    % HOUSEHOLDS BY TYPE - Households with one or
   more people under 18 years
wcj = 0.0428767;  % PLACE OF BIRTH - Total population
wbb = -0.069315;  % PLACE OF BIRTH - Total population
wbh = -0.1604505; % EDUCATIONAL ATTAINMENT - Population 25
   years and over
wbk = -0.0565836; % EDUCATIONAL ATTAINMENT - High school
   graduate (includes equivalency)
wbn = 0.1024029;  % EDUCATIONAL ATTAINMENT - Bachelor's degree
wbr = 0.1051412;  % VETERAN STATUS - Civilian population 18
   years and over
wdl = -0.1203898; % EDUCATIONAL ATTAINMENT - Some college, no
   degree
wec = 0.0144962;  % ANCESTRY ÍC German
wbu = 0.000114;   % DISABILITY STATUS OF THE CIVILIAN
   NONINSTITUTIONALIZED POPULATION - With a disability
wal = 0.0423417; % FERTILITY - Number of women 15 to 50 years
   old who had a birth in the past 12 months
wp = -0.0307825;  % HOUSEHOLDS BY TYPE - Households with one or
   more people 65 years and over
wbs = -0.0529627; % VETERAN STATUS - Civilian veterans
ww = 0.0643631;   % RELATIONSHIP - Other relatives

%  set the change of these factors
deltao = 0;
deltacj = 0;
deltabb = 0;
deltabh = 0;
deltabk = 15;
deltabn = 0;
deltabr = 0;
deltadl = 15;
deltaec = 0;
deltabu = 0;
deltaal = 0;
deltap = 0;
deltabs = 0;
deltaw = 0;

%  data input
```

```
%  data for Heroin
data_h = xlsread('Part1_Heroin_data_combined', 'Sheettotal');
data_h_10 = xlsread('Part1_Heroin_data_combined', 'Sheet2010');
data_h_11 = xlsread('Part1_Heroin_data_combined', 'Sheet2011');
data_h_12 = xlsread('Part1_Heroin_data_combined', 'Sheet2012');
data_h_13 = xlsread('Part1_Heroin_data_combined', 'Sheet2013');
data_h_14 = xlsread('Part1_Heroin_data_combined', 'Sheet2014');
data_h_15 = xlsread('Part1_Heroin_data_combined', 'Sheet2015');
data_h_16 = xlsread('Part1_Heroin_data_combined', 'Sheet2016');
data_h_17 = xlsread('Part1_Heroin_data_combined', 'Sheet2017');
%  data for synthetic opioids
data_so = xlsread('synthetic_opioids_data_total', 'Sheettotal');
data_so_10 = xlsread('synthetic_opioids_data_total',
   'Sheet2010');
data_so_11 = xlsread('synthetic_opioids_data_total',
   'Sheet2011');
data_so_12 = xlsread('synthetic_opioids_data_total',
   'Sheet2012');
data_so_13 = xlsread('synthetic_opioids_data_total',
   'Sheet2013');
data_so_14 = xlsread('synthetic_opioids_data_total',
   'Sheet2014');
data_so_15 = xlsread('synthetic_opioids_data_total',
   'Sheet2015');
data_so_16 = xlsread('synthetic_opioids_data_total',
   'Sheet2016');
data_so_17 = xlsread('synthetic_opioids_data_total',
   'Sheet2017');
%  time
year = [2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017];

%  deal with heroin's data
FIPSIN_h = data_h(:, 8);
FIPSIN_h_10 = data_h_10(:, 8);
FIPSIN_h_11 = data_h_11(:, 8);
FIPSIN_h_12 = data_h_12(:, 8);
FIPSIN_h_13 = data_h_13(:, 8);
FIPSIN_h_14 = data_h_14(:, 8);
FIPSIN_h_15 = data_h_15(:, 8);
FIPSIN_h_16 = data_h_16(:, 8);
FIPSIN_h_17 = data_h_17(:, 8);
num = length(FIPSIN_h);
%  get all the county's FIPS
FIPS(1) = FIPSIN_h(1);
for i = 2:num
   pos = find(FIPS == FIPSIN_h(i));
   if isempty(pos)
      FIPS(end + 1) = FIPSIN_h(i);
   end
```

```
end

%  sort all the county's FIPS
FIPS = FIPS';
FIPS = sort(FIPS);
countynum = length(FIPS);
%  get a matrix that only have number of cases
for i = 1:countynum
   pos = find(FIPSIN_h_10 == FIPS(i));
   if isempty(pos)
      reph(i, 1) = 0;
   else
      reph(i, 1) = data_h_10(pos, 5);
   end
   pos = find(FIPSIN_h_11 == FIPS(i));
   if isempty(pos)
      reph(i, 2) = 0;
   else
      reph(i, 2) = data_h_11(pos, 5);
   end
   pos = find(FIPSIN_h_12 == FIPS(i));
   if isempty(pos)
      reph(i, 3) = 0;
   else
      reph(i, 3) = data_h_12(pos, 5);
   end
   pos = find(FIPSIN_h_13 == FIPS(i));
   if isempty(pos)
      reph(i, 4) = 0;
   else
      reph(i, 4) = data_h_13(pos, 5);
   end
   pos = find(FIPSIN_h_14 == FIPS(i));
   if isempty(pos)
      reph(i, 5) = 0;
   else
      reph(i, 5) = data_h_14(pos, 5);
   end
   pos = find(FIPSIN_h_15 == FIPS(i));
   if isempty(pos)
      reph(i, 6) = 0;
   else
      reph(i, 6) = data_h_15(pos, 5);
   end
   pos = find(FIPSIN_h_16 == FIPS(i));
   if isempty(pos)
      reph(i, 7) = 0;
   else
      reph(i, 7) = data_h_16(pos, 5);
```

```
    end
    pos = find(FIPSIN_h_17 == FIPS(i));
    if isempty(pos)
       reph(i, 8) = 0;
    else
       reph(i, 8) = data_h_17(pos, 5);
    end
end

%  set the parameter of the model
alpha = 0.4;
beta = 0.7;
gamma = 0.6;
%  calculate the factor change's value of heroin
fcvh = wec * deltaec + wbu * deltabu + wal * deltaal + wp *
    deltap + ...
    wbs * deltabs + ww * deltaw;

for i = 1:countynum
    data = reph(i, :);
    s(1) = data(1);
    t(1) = data(2) - data(1);
    p(1) = 0;
    for j = 2:8
       if j == 8
          s(j) = alpha * (data(j) - p(j-1)) + (1 - alpha) *
              (s(j-1)+t(j-1));
          t(j) = beta * (s(j) - s(j-1)) + (1-beta) * t(j-1) +
              fcvh;
          p(j) = gamma * (data(j) - s(j)) + (1-gamma) * p(j-1);
       else
          s(j) = alpha * (data(j) - p(j-1)) + (1 - alpha) *
              (s(j-1)+t(j-1));
          t(j) = beta * (s(j) - s(j-1)) + (1-beta) * t(j-1);
          p(j) = gamma * (data(j) - s(j)) + (1-gamma) * p(j-1);
       end
    end
    for j = 9:11
       data(j) = s(j-1) + t(j-1) + p(j - 2);
       data(j) = round(data(j));
       if data(j) < 0
          data(j) = 0;
       end
       s(j) = alpha * (data(j) - p(j-1)) + (1 - alpha) *
           (s(j-1)+t(j-1));
       t(j) = beta * (s(j) - s(j-1)) + (1-beta) * t(j-1) + fcvh;
       p(j) = gamma * (data(j) - s(j)) + (1-gamma) * p(j-1);
       reph(i, j) = data(j);
    end
```

```
end

%  find the source point and the dangerous point
reph_10 = reph(:, 1);
reph_17 = reph(:, 8);
reph_18 = reph(:, 9);
reph_19 = reph(:, 10);
reph_20 = reph(:, 11);
souh(1) = 0;
for i = 1:length(reph_10)
   if reph_10(i) > 100
      souh(end+1) = FIPS(i);
   end
end
souh(1) = [];
danh(1) = 0;
for i = 1:length(reph_17)
   if reph_17(i) > 100
      pos = find(danh == FIPS(i));
      if isempty(pos)
         danh(end+1) = FIPS(i);
      end
   end
end
for i = 1:length(reph_18)
   if reph_18(i) > 100
      pos = find(danh == FIPS(i));
      if isempty(pos)
         danh(end+1) = FIPS(i);
      end
   end
end
for i = 1:length(reph_19)
   if reph_19(i) > 100
      pos = find(danh == FIPS(i));
      if isempty(pos)
         danh(end+1) = FIPS(i);
      end
   end
end
for i = 1:length(reph_20)
   if reph_20(i) > 100
      pos = find(danh == FIPS(i));
      if isempty(pos)
         danh(end+1) = FIPS(i);
      end
   end
end
%  output the data
```

```
danh(1) = [];
xlswrite('Heroin_pre.xlsx', FIPS, 1, 'A1');
xlswrite('Heroin_pre.xlsx', reph, 1, 'B1');
clear FIPS;

%  deal with synthetic opioid's data
FIPSIN_so = data_so(:, 6);
FIPSIN_so_10 = data_so_10(:, 6);
FIPSIN_so_11 = data_so_11(:, 6);
FIPSIN_so_12 = data_so_12(:, 6);
FIPSIN_so_13 = data_so_13(:, 6);
FIPSIN_so_14 = data_so_14(:, 6);
FIPSIN_so_15 = data_so_15(:, 6);
FIPSIN_so_16 = data_so_16(:, 6);
FIPSIN_so_17 = data_so_17(:, 6);
%  get all the county's FIPS
num = length(FIPSIN_h);
FIPS(1) = FIPSIN_so(1);
for i = 2:num
   pos = find(FIPS == FIPSIN_so(i));
   if isempty(pos)
      FIPS(end + 1) = FIPSIN_so(i);
   end
end
%  sort all the county's FIPS
FIPS = FIPS';
FIPS = sort(FIPS);
countynum = length(FIPS);
%  get a matrix that only have number of cases
for i = 1:countynum
   pos = find(FIPSIN_so_10 == FIPS(i));
   if isempty(pos)
      repso(i, 1) = 0;
   else
      repso(i, 1) = data_so_10(pos, 8);
   end
   pos = find(FIPSIN_so_11 == FIPS(i));
   if isempty(pos)
      repso(i, 2) = 0;
   else
      repso(i, 2) = data_so_11(pos, 8);
   end
   pos = find(FIPSIN_so_12 == FIPS(i));
   if isempty(pos)
      repso(i, 3) = 0;
   else
      repso(i, 3) = data_so_12(pos, 8);
   end
   pos = find(FIPSIN_so_13 == FIPS(i));
```

```
    if isempty(pos)
       repso(i, 4) = 0;
    else
       repso(i, 4) = data_so_13(pos, 8);
    end
    pos = find(FIPSIN_so_14 == FIPS(i));
    if isempty(pos)
       repso(i, 5) = 0;
    else
       repso(i, 5) = data_so_14(pos, 8);
    end
    pos = find(FIPSIN_so_15 == FIPS(i));
    if isempty(pos)
       repso(i, 6) = 0;
    else
       repso(i, 6) = data_so_15(pos, 8);
    end
    pos = find(FIPSIN_so_16 == FIPS(i));
    if isempty(pos)
       repso(i, 7) = 0;
    else
       repso(i, 7) = data_so_16(pos, 8);
    end
    pos = find(FIPSIN_so_17 == FIPS(i));
    if isempty(pos)
       repso(i, 8) = 0;
    else
       repso(i, 8) = data_so_17(pos, 8);
    end
end

%  set the parameter of the model
alpha = 0.4;
beta = 0.75;
gamma = 0.7;
%  calculate the factor change's value of heroin
fcvso = wo * deltao + wcj * deltacj + wbb * deltabb + wbh *
    deltabh + ...
    wbk * deltabk + wbn * deltabn + wbr * deltabr + wdl *
       deltadl;

for i = 1:countynum
    data = repso(i, :);
    s(1) = data(1);
    t(1) = data(2) - data(1);
    p(1) = 0;
    for j = 2:8
       if j == 8
          s(j) = alpha * (data(j) - p(j-1)) + (1 - alpha) *
```

```
            (s(j-1)+t(j-1));
         t(j) = beta * (s(j) - s(j-1)) + (1-beta) * t(j-1) +
            fcvso;
         p(j) = gamma * (data(j) - s(j)) + (1-gamma) * p(j-1);
      else
         s(j) = alpha * (data(j) - p(j-1)) + (1 - alpha) *
            (s(j-1)+t(j-1));
         t(j) = beta * (s(j) - s(j-1)) + (1-beta) * t(j-1);
         p(j) = gamma * (data(j) - s(j)) + (1-gamma) * p(j-1);
      end
   end
   for j = 9:11
      data(j) = s(j-1) + t(j-1) + p(j - 2);
      data(j) = round(data(j));
      if data(j) < 0
         data(j) = 0;
      end
      s(j) = alpha * (data(j) - p(j-1)) + (1 - alpha) *
         (s(j-1)+t(j-1));
      t(j) = beta * (s(j) - s(j-1)) + (1-beta) * t(j-1) + fcvso;
      p(j) = gamma * (data(j) - s(j)) + (1-gamma) * p(j-1);
      repso(i, j) = data(j);
   end
end

%  find the source point and the dangerous point
repso_10 = repso(:, 1);
repso_17 = repso(:, 8);
repso_18 = repso(:, 9);
repso_19 = repso(:, 10);
repso_20 = repso(:, 11);
souso(1) = 0;
for i = 1:length(repso_10)
   if repso_10(i) > 50
      souso(end+1) = FIPS(i);
   end
end
souso(1) = [];
danso(1) = 0;
for i = 1:length(repso_17)
   if repso_17(i) > 100
      pos = find(danso == FIPS(i));
      if isempty(pos)
         danso(end+1) = FIPS(i);
      end
   end
end
for i = 1:length(repso_18)
   if repso_18(i) > 100
```

```
      pos = find(danso == FIPS(i));
      if isempty(pos)
         danso(end+1) = FIPS(i);
      end
   end
end
for i = 1:length(repso_19)
   if repso_19(i) > 100
      pos = find(danso == FIPS(i));
      if isempty(pos)
         danso(end+1) = FIPS(i);
      end
   end
end
for i = 1:length(repso_20)
   if repso_20(i) > 100
      pos = find(danso == FIPS(i));
      if isempty(pos)
         danso(end+1) = FIPS(i);
      end
   end
end
%  output the data
danso(1) = [];
xlswrite('SynO_pre.xlsx', FIPS, 1, 'A1');
xlswrite('SynO_pre.xlsx', repso, 1, 'B1');
```