Team Control Number

For Office Use Only	**1919458**	For Office Use Only
T1 _____		F1 _____
T2 _____	**Problem Chosen**	F2 _____
T3 _____	**C**	F3 _____
T4 _____		F4 _____

2019 MCM/ICM Summary Sheet

A Time Series Model for Studying the Variation Trend of Drug Report

Summary

Before the model constructing, we preprocess the given NFLIS data, selecting synthetic opioids and heroin drug report from various drug crime incidents, regardless of other non-synthetic opioids.

Firstly, in order to discuss the characteristics of drug report spread over time, we decide to use ARIMA model to describe the time series by considering two aspects, adjacent counties' influence and history records. For the adjacent counties' influence, we import SIS model to solve, which was usually used in infectious problem.

Secondly, we use the socio-economic data of the US census to optimize the model. Due to the huge amount of data and the complexity of the data types, we decided to select the indicators that are most likely to affect the model.We use Correlation and T-test to decide Top-Five factors, then we use Partial Least Square (PLS) method to combines the factors, and modifies our original model.

Finally, considering the relevant factors we discuss beyond, we provide some strategies to limit the drug abuse, and evaluate the influence of each policy, analysis whether the strategy is effective so that giving advice to U.S.government to achieve the goal.

Keywords: Drug report, synthetic opioids, SIS, ARIMA, PLS

Content

Memorandum for Chief administrator of the DEA/NFLIS database

Dear chief administrator,

It’s a nice exprience for us to use NFLIS Database records to do model training and data using. With your various kinds of data, we can be aware of what society factors are now changing, and how they inflect the drug report event number.

However, we are here to find that, in your provide excel table, there are several NaN value in the locations that have to be a number, and I actually want to make a advice for your data field providing, aiming that you can reduce some likely field into one, avoind repeatation. (That’s because I find several same field in ANN datas).

To deal with hundreds of indicators which are from the U.S.Census socio-economic data, we spent much time looking up their meaning. The work is heavy and we decided to only consider the estimated value, but discarded error margin and its percentage value to lighten our burden.

In the last part, we offered some creative strategies to control the number of drug reports, based on the related indications beyond the discuss.

1 Introductions

1.1 Problem Background

The United States is experiencing a national crisis regarding the use of synthetic and nonsynthetic opioids, either for the treatment and management of pain (legal, prescription use) or for recreational purposes (illegal, non-prescription use). Federal organizations such as the Centers for Disease Control (CDC) are struggling to "save lives and prevent negative health effects of this epidemic, such as opioid use disorder, hepatitis, and HIV infections, and neonatal abstinence syndrome."1 Simply enforcing existing laws is a complex challenge for the Federal Bureau of Investigation (FBI), and the U.S. Drug Enforcement Administration (DEA), among others.

There are implications for important sectors of the U.S. economy as well. For example, if the opioid crisis spreads to all cross-sections of the U.S. population (including the college-educated and those with advanced degrees), businesses requiring precision labor skills, high technology component assembly, and sensitive trust or security relationships with clients and customers might have difficulty filling these positions. Further, if the percentage of people with opioid addiction increases within the elderly, health care costs and assisted living facility staffing will also be affected.

1.2 Our Goal

Based on comprehension of the points, our purpose is as follows:

- Using the provided NFLIS data to find out the number of synthetic opioid and heroin incidents in each state and county in the next year.
- Establish a time series model (ARIMA) between 2010 and 2016 to describe the regularity of synthetic opioid and heroin drug report over time in each state, and predict the possible location of any particular opioid drug that has been used.
- By comparing the correlation coefficient and the significant difference, we select the social indicators that are most likely to affect the change rule of drug report, and add them to the model to make it more perfect;

- Giving strategies to limit the drug abuse, and evaluate whether effective.

2 Assumptions

We make the following assumptions about the drug reports' changing over time in this paper.

- In order to simplify the model, we divide the counties into three status: low spreading, middle spreading, and high spreading, according to its drug reports' numbers for special medicine.
- An county has a recovery rate, which reduce the drug reports' number, and the recovery rate has negative correlation with drug reports' number.
- An county has a infectious rate, which increase its adjacent counties' drug reports, and the infectious rate has positive correlation with drug reports' number.
- Only the adjacent counties can infect each. We only consider three factors: adjacent countries' influence, history records' influence, and some indicators of U.S Census socio-economic data.

3 Notations

For the definition of problem notations, we make the following notation table to record every sign used in our model, including Susceptible Infected Susceptible(SIS) model and ARIMA model, and the modified-improved model for Part 2.

Table 1.Model Notation table

S(i)	The i-th state we considered in problem
C(S(i), j)	The j-th county we considered in special state *S(i)*
D(t)	The t-th drug kind we considered
Y(p)	The p-th year we considered
Adj(S(i))	The two-dimension adjency matrix for state *S(i)*

Beta(k)	In SIS Model, the probability of one county k to infect its neighbor counties
Gamma(k)	In SIS Model, the probability of one county k to cure and reduce its own severity of drug transmission
DrugReport(C(S(i), j), D(t), Y(p))	The total number of the special drug 's report in the county *C(S(i), j)* in year *Y(p)*
Neigh(y)	The neighbors of the county -- y
SIS_predict(C(S(i), j), D(t), Y(p))	The predict result of SIS Model, output one county's future drug report of *D(t),* in the next year of *Y(p)*
ARIMA_predict(*C(S(i), j), D(t), Y(p)*)	The predict result of ARIMA Model, output one county's future drug report of *D(t),* in the next year of *Y(p)*
NeighSum(C(S(i), j), D(t), Y(p))	Sum of drug reports of all the neighbor counties of *C(S(i), j),* for the special drug kind *D(t)* in year *Y(p)*

4 Data Preprocessing

For data-analysis problem, there are usually some incomplete and abnormal data in the large amount of raw data, which may seriously affect the efficiency of modeling and the accuracy of conclusions. So it is quite important to preprocess the data.

Firstly, we discuss a question of the regularity about the variation trend of synthetic opioids and heroin in Part 1. There are 69 substance names in NFLIS data corresponding to medicine names, but not all of them belongs to synthetic opioids and heroin. After selecting from the NFLIS data, we get only 14 drugs satisfied the requirement, which displays as a below.

Table 4.1 synthetic opioids and heroin in NFLIS data

substancename
Propoxyphene
Methadone
Heroin
Dextropropoxyphen
Buprenorphine
Fentanyl
Tramadol
Pethidine
Pentazocine
Butorphanol
Nalbuphine
Remifentanil
Mitragynine

Secondly, to discuss the influence of geographical factors in the variation process, we should consider the adjacent geographical positions, since a county having high drug reports record would easily influence its neighbors. In order to simplify the model, we regard each county as nodes, constructing undirected graph, and build an one-to-one match between nodes and FIPS_Combine as a below figure 4.2.

Figure 4.2 Ohio - Match between FIPS_Combine and geographical position[3]

5 The Model

5.1 Predict Model with NFLIS Data

5.1.1 Overview

To describe the variation trend of drug reports for each county, with the help of NFLIS Data, we both consider the influence from history records of itself and other adjacent counties.

5.1.2 Susceptible Infected Susceptible Model[1]

Since having had a hypothesis of that the adjacent counties which have drug reports will have an effect on spreading illegal medicine for each county, so that increases the number of drug reports, we now in tend to calculate the influence from its neighbors.

To deal with the problem, we decide to use the SIS[1] model, which is usually used in infectious diseases problem, and modify it to adapt our model in county level. We assume each county has a recovery rate which decreases the the number of drug reports recorded in last year, and an infectious rate which increases the adjacent counties' drug reports.

With the rise of drug reports, the recovery rate will drop but the infectious rate will increase, as a result of more and more people become involved in illegal medical use. The sheet of recovery rate and infectious rate is as follow.

Table 5.1. Recovery Rate and Infectious Rate

Status	Recovery Rate	Infectious Rate
Low	0.3	0.1
Middle	0.2	0.2
High	0.1	0.3

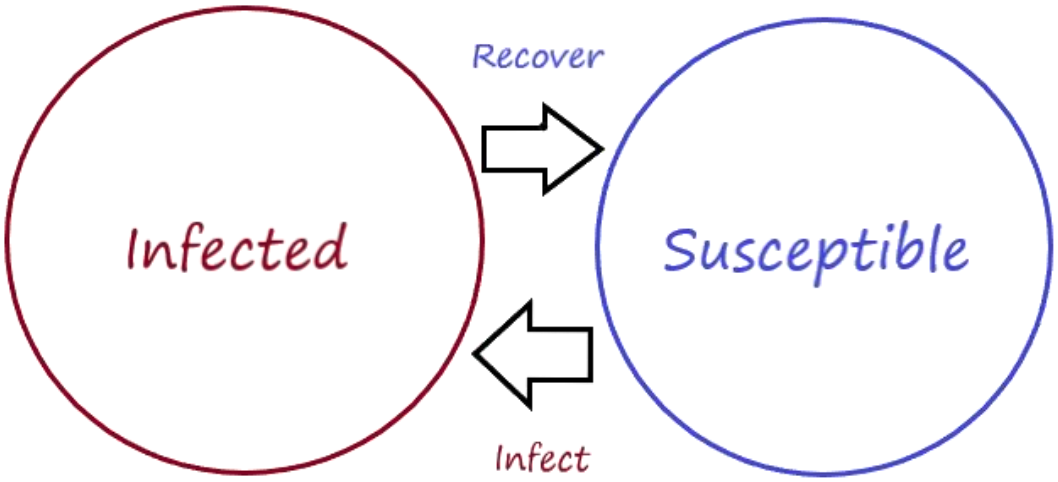
Basically, the SIS model divide every individuals into two classes -- *S* for Susceptible class, *I* for Infected class. Using the dividing coditions below, we define three drug transmission states inside the county. What you have to be aware of is that we calculate the county states for different drug kinds -- *D(t)*, and for diffreent year *Y(p)* of it.

Table 5.2. Healthy, slightly infected, strongly infected county state definition

Status	The special drug *D(t)'s* total reports number in year *Y(p)*
Healthy	0 ~ 10
Slightly Infected	10 ~ 30
Strongly Infected	> 30

The SIS model emphasizes the county individuals transmission between S-class and I-class, for this part we use infectious disease ordinary differential equation (will be shown later) to simulate the growing and changing drug reports number of one county, with the constraints below:

1. For every county inside a special state *S(i)*, one county *C(S(i), j)* only have ability to infect all its neighbor counties *Neigh(C(S(i), j))*.
2. One county may infect its neighbor with its own infectious rate *Beta(k)* for one special drug *D(t)* in the year *Y(p)*.
3. With the society policy implementation, one county may recover its own drug abuse, with the recovery rate *Gamma(k)*.



The formula of ordinary differential equation about the relationship between drug reports and adjacent nodes influence is below.

$$Con = C(S(i), j)$$

$$SIS\_predict(Con, D(t), Y(p)) =$$

$$NeighborSum(Con, D(t), Y(p)) * Beta(Con) - Drugreport(Con, D(t), Y(p)) * Gamma(Con)$$

Explain of Formula upon:

1. Firstly, we select the target county *Con* inside one state, to do its SIS-Predict.

2. Then it's necessary to know all the neighbor of the target county, with the Infectious rate of the target county itself, we collect all neighbor report sum of one drug in one year, to do infection prediction of *Con.*

3. Considering the recovery stage, we use recover rate to make sure the county's drug transmission inside can be reduced, to do recovery prediction of *Con*.

5.1.3 ARIMA[2]

ARIMA[2] is a widely used time serials analysis model. To describe the change regulation of drug reports for each counties, we should consider both adjacent counties' influence and history drug reports. The adjacent counties' influence is solved by the SIS model, so the next step is to synthesize the two aspects, which we use auto regressive integrated moving average model (ARIMA) to settle.

The step is as follow:

1、Stationary Processing of Non-stationary Time Series Data

2、Establish the corresponding time series model according to the identified features.

3、Estimation of parameters and test statistical significance.

4、Analyze and predict the time series.

The ARIMA model bases on the data sequence itself, input one influence factor's value sequence with time, constrcut the predict model to get the future time result. The Basic equation of ARIMA model is shown below:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$$

1. ARIMA model consists of three part -- AR, I, MA.The autocorrelation coefficient of the AR(p) model exhibits exponential decay or oscillating attenuation with increasing k. The specific attenuation form depends on the coefficient of the lag term of the AR(p) model; the partial autocorrelation of the AR(p) model The coefficients are c-order truncated. Therefore, the order p of the AR(p) model can be determined by identifying the number of partial autocorrelation coefficients of the AR(p) model.

2. The autocorrelation coefficient of the MA(q) model is truncated after step q. The partial autocorrelation coefficient of the MA(q) model must exhibit a trailing attenuation form.

3. The ARMA(p,q) model is a combined model of AR(p) model and MA(q) model, so the autocorrelation coefficient of ARMA(p,q) is the autocorrelation coefficient of AR(p) autocorrelation coefficient and MA(q). a mixture of coefficients.

Since we only have drug reports in [2010 , 2017], the related parameters -- p, q should not be too large. After a comparison between different p and q, and based on the minimum-AIC guideline, we repeatly change p and q in range [1, 5], to find the best p,q pair that mostly suit every state -- making AIC-sum of every county model in every state the minimum.

Table 5.3. Average AIC-Sum of ARIMA model with different p,q pair

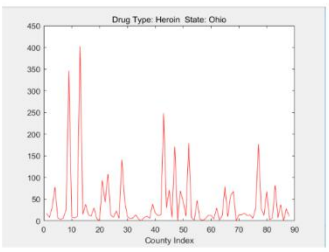
	Q=1	Q=2	Q=3	Q=4	Q=5
P=1	59.1	79.3	95.21	128.03	176.33
P=2	67.77	91.2	176.3	211.5	298.74
P=3	70.98	103.87	157.3	204.4	267.3
P=4	82.34	201.6	264.12	308.54	344.51
P=5	98.41	251.4	309.51	404.77	509.2

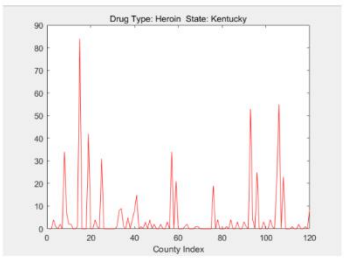
Based on the upon table, we decide the best p-q pair will be: $p = 1, q = 1$

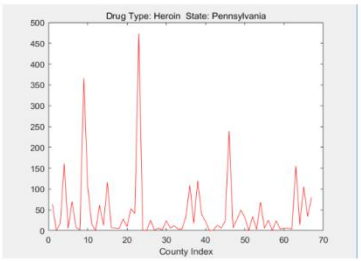
5.1.4 Results of the Predictive Model

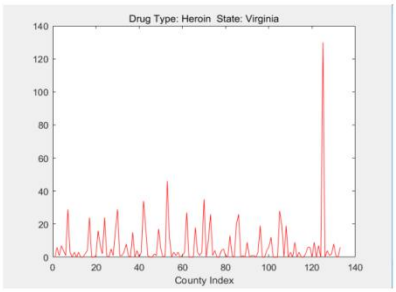
Based on the predictive model provided, we can obtain different data about different counties' drug reports in 2017.

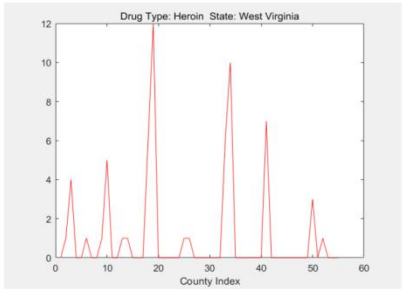
Firstly we introduce our SIS-ARIMA model's prediction data of year 2018, here we select opioids -- Heroin and make chart of five state's counties, showing the drug report of Heroin that will happen inside each state.











Analysing from above charts, we can decide that Ohio and Pennsylvania is mostly influenced by Heroin drug, cause they both have max heroin drug report number higher than 400, and the integral ploy line shows these two areas are high-risk areas using heroin.

Beside them, West Virginia has lowest-risk of using Heroin, cause the max heroin report number of each county in it will not exceed 50.

Considering the problem of identifying all possible locations where specific opioid may have started use, here we define two constraints when collecting target result:

1. The target locations will be selected when a kind of drug's report change from zero to a positive with time going by, and in year 2018 the special drug report number is positive.
2. We only consider 13 types of Synthetic opioid and Heroin.
3. We below choose low-rate happened drug report -- Propoxyphene to show.

State-Name	Possible Locations of using Propoxyphene
Ohio	018, 032, 050
Kentucky	None
Pennsylvania	023
Virginia	None
West Virginia	None

For a easier way to explain all possible danger drug event that can be happened, we here define several drug event concern table, given one event the special id.

Table 5.4. Two possible drug event happen inside one county

Concern-ID	Explaination
01	Special drug increase fastly inside the county
02	Specila drug spreads fastly from county to county

See Appendix for more data about any possible positions for each counties that have already used special medium in 2018.

According to the predictive result provided, the government should be worried about the following factors.

- The number of drug reports, predicted in 2018, which are caused by Fentanyl is largely more than other illegal medicine, making up to 72.1%, 80.3%, 46.3%, 61.1%,and 74.5% of total incidents in order of Ohio, Pen, Vir, West Vir, and Ken State.
- The drug incidents happened in PA and OH is more serious, totally 29307 and 25480 as predicted in 2018, and the number of incidents happened in West Virginia is only 550.

So the next step of DEA is keeping tight control over the supply of Fentanyl, and spending more energy on drug management in OH and PA state.

5.2 Modify with Social-economy Data

5.2.1 Correlation Test and Significant Difference

Using the U.S Census socio-economic data provided, we should find some related indicators which have a significant impact on total drug reports' variation trend. For each county, the data has recorded so many indexes from 2010 to 2016. According to calculate the correlation coefficient between the variation of drug reports' number and

socio-economic indicators, we could find the most remarkable indicators that has influence on drug reports' changing.

There are duplicate indicators that reflect the same social significance, such as MARITAL STATUS - Divorced and MARITAL STATUS - Males 15 years and over. To avoid considering the similar indicators, we select Top30 indicators with highest correlation coefficient and reliable significant difference as table 5.1.

Table 5.1 Partial Selected indicators only Top10

Rank	indicator	Rank	indicator
1	HOUSEHOLDS BY TYPE - Nonfamily households	6	PLACE OF BIRTH - Native - Born in United States - Different state
2	SCHOOL ENROLLMENT - Elementary school (grades 1-8)	7	HOUSEHOLDS BY TYPE - Households with one or more people under 18 years
3	HOUSEHOLDS BY TYPE - Nonfamily households - Householder living alone	8	SCHOOL ENROLLMENT - Elementary school (grades 1-8)
4	MARITAL STATUS - Never married	9	HOUSEHOLDS BY TYPE - Nonfamily households - Householder living alone
5	MARITAL STATUS - Divorced	10	MARITAL STATUS - Never married

Removing the duplicate indicators, we finally get a set contains five factors: marital status, educational attainment, veteran status, ancestry and households by type. Each factors has one typical indicator showing in table 5.2

Table 5.2. typical indicator

Factors	Typical Indicator
Households by Type	Non-family households
Marital Status	Divorced
Educational Attainment	9th to 12th grade, no diploma

Veteran Status	Civilian veterans
Ancestry	English

Based on the result, we propose hypotheses that explains why these indicators makes contributions to the growth of drug reports.

- The influence of social environment is an important factor for some drug addicts to take the road of drug abuse, which is related to the non-family households population and divorced population.
- Well-educated people are less likely to take drugs.
- The other indicators, such as civilian veterans population, naive speakers of English, also have a close contact with drug report number.

5.2.2 Partial Least Square Regression

To modify the original model, we use Partial Least Square Regression (PLS Regression) to analysis the number of drug reports changing with six index, they are original predictive result, Non-family households number, Divorced number, no diploma with 9th to 12th number, Civilian veterans number, and Ancestry is English.

The steps are below,

- Data standardization.
- Finding the correlation coefficient matrix.
- Select the independent variable group and dependent variable group, and take the first k components when the ratio of k component reach to 90%.
- Solving the regression equation.
- Reduce the normalized regression variables to original variable.

In order to realize the basic function of PLS, we here need to divide all variables into two part: Independent variables and Dependent variables. The target of PLS is to get the best weight of every independent variables t(i), and construct the result equation to predict result dependent variables u(i).

Basic PLS Model Equation:

$$Con = C(S(i), j)$$
$$factor\_result = w_1 * Factor_1(Y(p)) + w_2 * Factor_2(Y(p)) + w_3 * Factor_3(Y(p)) + w_4 * Factor_4(Y(p)) + w_5 * Factor_5(Y(p))$$
$$part1\_result = w_6 * SIS\_And\_Arima(Con, Y(p))$$
$$final\_result(Con, Y(p)) = factor\_result + part1\_result$$

Then the question is to ask: How to calculate the best weight? We may build equation set to get correlation of t(i) and u(i) the minimum, that's to solve the optimization question below:

$$\max\{Cov(t1, u1)\}$$
$$= \max\langle E_0 w_1, F_0 c_1\rangle$$
$$s.t\begin{cases} w_1^T w_1 = 1 \\ c_1^T c_1 = 1 \end{cases}$$

Use Lagrangian multiplier to obtain w1 and c1 satisfy:

$$E_0^T F_0 F_0^T E_0 w_1 = \theta_1^2 w_1$$

$$F_0^T E_0 E_0^T F_0 c_1 = \theta_1^2 c_1$$

Process normalization and residual equation, we finally get the suitable weight

$$F_0 = t_1 r_1^T + \ldots + t_A r_A^T + F_A = E_0\left[\sum_{j=1}^{A} w_j^* r_j^T\right] + F_A$$

Base on PLS Model establishing principle, and upon five related factors, Part-One result as independent set (Six independent variables), total drug reports number as dependent set (One dependent variable). We construct PLS Model for every county inside the state, use matlab program to decide weight result.

2	0.0317
3	0.0053
4	-0.0149
5	-0.0148
6	-0.0143
7	0.0629

Table of related weight value for sepcial independent variable:

Factors	Correlation coefficient

Households by Type - Non-family households	0.0317
Marital Status - Divorced	0.0053
Educational Attainment - 9th to 12th grade, no diploma	-0.0149
Veteran Status - Civilian veterans	-0.0148
Ancestry - English	-0.0143
SIS-ARIMA Model Prediction Result	0.0629

5.3 The Strategies of Limit Drug Abuse

Based on the results of the first two parts, we obtains a predictive model with the linear combination of three parts, such as history records, adjacent influence, and some social-economic indicators.

To limit the drug abuse, there are some tips from our model.

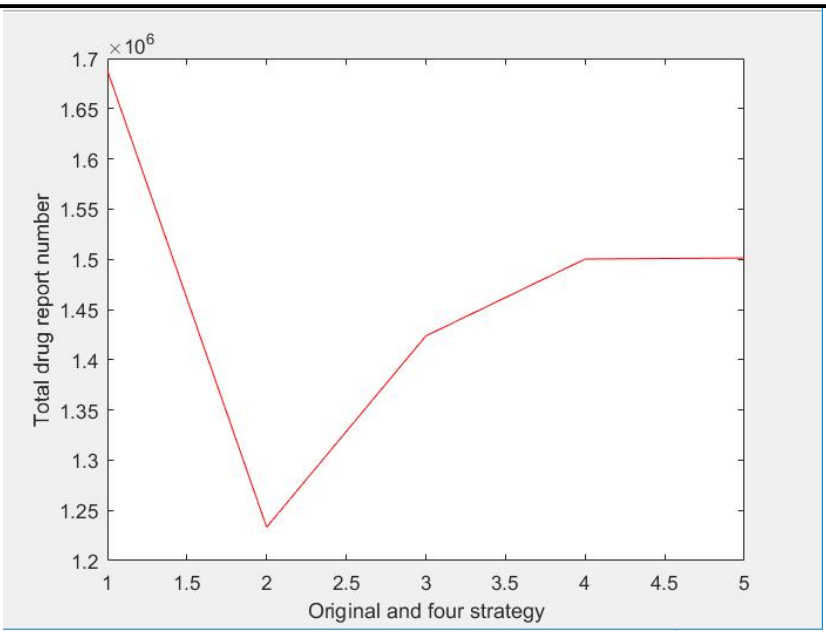
- Limit the spreading of opiods between different counties to prevent abuse, decreasing the infectious rate of each nodes.
- Increase the punishment of drug reports, and this policy can reduce the number of drug users.
- Expanding school education for opiods medicine's harms to avoid drug abuse by young person.
- Taking care of marginalized groups in society, such as divorce population, retire soldiers, and non-family households, and giving them more community care.

According to the following limit constraints, we define four strategy to limit the drug abuse, as followed:

- Strategy 1 -- Infectious Rate Beta reduce to 80% of itself.
- Strategy 2 -- Recovery Rate Gamma imporve to 120 % of itself.
- Strategy 3 -- 9th to 12th grade, no diploma people number reduce to 80%.
- Strategy 4 -- Divorce, Civilian veterans and no-family people number reduce to 95%.

Table 5.4. Strategy result table

Case	Total incident number
Original	1687961
Strategy 1	1233284
Strategy 2	1423706
Strategy 3	1500229
Strategy 4	1501341



Result Analysis:

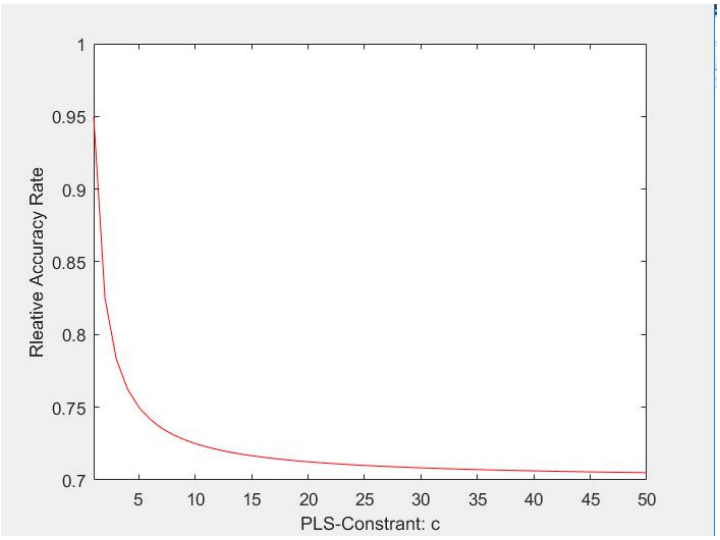
- We are aware of that all these four strategies can reduce the total drug report number, cause it do help to society improved, making people healthier in mind and body, so result drug number can be less that original data.
- Inside these four strategy, we find reduce the infectious rate is the best way to stop drug event imporve, it means that drug abuse usually develops by transmission.
- By using Strategy 1 and 2, county inside can have less drug report number with SIS model, and Strategy 3 and 4 improve society conditions, as they reduce divorced and no-family people, making happy peopel more, as well as improving people's culture level, so that they can be aware of opiods' danger.

6 Model Analysis

6.1 Sensitivity Analysis

Considering SIS-ARIMA and Modified Model with PLS, it contains several constant parameters, such as: PLS usually produce the constant number *c*, and in SIS model, we usually have infectious rate *beta* and recovery rate *gamma.* Here we change the PLS-Constant and SIS-beta-gamma, by personally change their values in range, to test our model sensitivity.

6.1.1 Impact of PLS-Constant *c* for prediction result



For above chart, we change PLS-Constrant in range [1, 50], and calculate model's relative accuracy rate (judged by compared predicted data with origin data), and find the error rate is reduced in a lower way, that means PLS-Constrant can slightly influence our modified model result, when it get larger, of course we select the lowest constrant for this question, to suit the special situation.

6.1.2 Impact of SIS-beta-gamma for predicition result

Here we define several beta-gamma pair for testing its influence to model sensitivity, the following relation table is shown below.

Table 6.1.Accuracy for different beta-gamma pairs

Beta-Gamma pairs	Accuracy Rate
Beta = [0.3,0.2,0.1] Gamma=[0.1,0.2,0.3]	91.03%
Beta = [0.5,0.3,0.1] Gamma=[0.1,0.3,0.5]	87.65%
Beta = [0.5,0.4,0.3] Gamma=[0.3,0.4,0.5]	83.55%

6.2 Strengths and Weaknesses

6.2.1 Strengths

- SIS-ARIMA-PLS Model uses NFLIS DataBase as the background data support, the result of it have strong reliability, can be applied in real life.
- Our Model can use NFLIS Excel datas as input, so it can carry on predicting in the long run, accepting changing NFLIS input data and predict.
- We consider several society factors that may have influence to drug abuse events, according to Corrleation and T-test, select Top-5 relative factors to make our model greater and diversity.
- All math equations and model usage are based on expert web site, and we combine them to make the prediction better.

6.2.2 Weaknesses

- Our model accuracy rate can not totally reach 100%, and we lack to analyse other factors that can have influence in drug abuse.
- Testing range and values are mostly selected by person, lacking strong foundation.

7 Conclusions

As our term set out to come up with a strategy on what would be the most efficient way to describe the variation trend of drug reports provided by

NFLIS reports for each counties, the first aspect that we took into major consideration was the adjacent influence, which we have used SIS model to analysis. Then we also considered the history records' influence. Based on the two aspects, we constructed a predictive model using ARIMA method, and predicted any possible positions that might occurs accidents.

At the second part, we add some indicators that have an impact on the variation trend from U.S.Census socio-enconomic data. We select five representational factors by comparing their correlation coefficients and significant difference, and then combined these factors and original model by partial least square method.

Finally, based on the relevant factors of the model, we raise some strategies to limit the drug abuse, and evaluate those strategies whether are effective.

References

[1] APC/EEB/MOL 514 Tutorial 4: Seasonal Epidemic Models. MECHANICAL ENGINEERING-University of California, Santa Barbara.

[2] Gareth Janacek. Time series analysis forecasting and control. Journal of Time Series Analysis, 31(4); 303-303, 2010.

[3] U.S. list of counties in Ohio. https://en.wikipedia.org/wiki/List_of_counties_in_Ohio

[4] U.S. list of counties in Pennsylvani. https://en.wikipedia.org/wiki/List_of_counties_in_Pennsylvani

[5] U.S. list of counties in Virginia. https://en.wikipedia.org/wiki/List_of_cities_and_counties_in_Virginia

[6] U.S. list of counties in West Virginia. https://en.wikipedia.org/wiki/List_of_counties_in_West_Virginia

[7] U.S. list of counties in Kentucky. https://en.wikipedia.org/wiki/List_of_counties_in_Kentucky

Appendix

Part I
Ohio-Part predict 14 kinds drug report in 2018

County-ID \ Drug-Nam	Fentanyl	Pethidine	Propoxyphene	Dextropropoxy	Methadone	Pentazocine	Buprenorphine	Nalbuphine	Tramadol	Remifentanil	Butorphanol	Mitragynine	Levorphanol	Heroin
1	8	0	0	0	2	0	12	0	1	0	0	0	0	17
3	23	0	0	0	0	0	3	0	3	0	0	0	0	8
5	113	0	0	0	4	0	30	0	3	0	0	0	0	31
7	375	0	0	0	3	0	25	0	5	0	0	0	0	78
9	12	0	0	0	1	0	52	0	2	0	0	0	0	7
11	1	0	0	0	0	0	1	0	0	0	0	0	0	3
13	62	0	0	0	2	0	49	0	0	0	0	0	0	7
15	10	0	0	0	1	0	17	0	1	0	0	0	0	26
17	527	0	0	0	5	0	54	0	13	0	0	0	0	346
19	0	0	0	0	0	0	9	0	0	0	0	0	0	9
21	36	0	0	0	0	0	0	0	2	0	0	0	0	7
23	271	0	0	0	0	0	31	0	5	0	0	0	0	10
25	274	0	0	0	1	0	55	0	7	0	0	0	0	402
27	36	0	0	0	1	0	54	0	3	0	0	0	0	15
29	192	0	0	0	0	0	68	0	17	0	0	0	0	38
31	0	0	0	0	0	0	9	0	0	0	0	0	0	14
33	10	0	0	0	1	0	142	0	1	0	0	0	0	11
35	3763	0	2	0	25	0	135	0	217	0	0	81	0	30
37	35	0	0	0	0	0	1	0	1	0	0	0	0	5
39	37	0	0	0	1	0	9	0	7	0	0	0	0	0
41	68	0	0	0	1	0	26	0	3	0	0	0	0	93
43	40	0	0	0	2	0	10	0	4	0	0	0	0	43
45	38	0	0	0	1	0	76	0	8	0	0	0	0	108
47	82	0	0	0	0	0	2	0	7	0	0	0	0	14
49	1383	0	0	0	8	0	207	0	91	0	0	0	0	8
51	22	0	0	0	0	0	1	0	1	0	0	0	0	23
53	79	0	0	0	4	0	35	0	1	0	0	0	0	6
55	0	0	0	0	0	0	1	0	9	0	0	0	0	141
57	144	0	0	0	1	0	13	0	2	0	0	0	0	47
59	45	0	0	0	0	0	51	0	1	0	0	0	0	10
61	4980	0	0	0	18	0	161	0	110	0	0	0	0	5
63	104	0	1	0	1	0	16	0	7	0	0	0	0	7
65	1	0	0	0	1	0	5	0	0	0	0	0	0	14
67	0	0	0	0	0	0	1	0	0	0	0	0	0	4
69	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Kentucky-Part predict 14 kinds drug report in 2018

County-ID	Fentanyl	Pethidine	Propoxyph	Dextropro	Methadone	Pentazoci	Buprenorp	Nalbuphir	Tramadol	Remifenta	Butorphar	Mitragyni	Levorphar	Heroin
1	0	0	0	0	0	0	12	0	0	0	0	0	0	0
3	0	0	0	0	0	0	3	0	0	0	0	0	0	0
5	5	0	0	0	0	0	1	0	0	0	0	0	0	4
7	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	5	0	2	0	0	0	0	0
11	3	0	0	0	0	0	2	0	0	0	0	0	0	2
13	0	0	0	0	1	0	30	0	2	0	0	0	0	0
15	97	0	0	0	0	0	6	0	0	0	0	0	0	34
17	8	0	0	0	0	0	6	0	0	0	0	0	0	7
19	135	0	0	0	0	0	17	0	4	0	0	0	0	2
21	96	0	0	0	0	0	10	0	0	0	0	0	0	2
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	1	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	1
29	83	0	0	0	0	0	0	0	0	0	0	0	0	84
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	1	0	0	0	0	0	0	0
35	0	0	0	0	2	0	2	0	0	0	0	0	0	0
37	228	0	0	0	3	0	15	0	5	0	0	0	0	42
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	29	0	0	0	1	0	10	0	0	0	0	0	0	0
43	47	0	0	0	0	0	11	0	0	0	0	0	0	4
45	0	0	0	0	0	0	5	0	0	0	0	0	0	1
47	0	0	0	0	0	0	3	0	0	0	0	0	0	0
49	16	0	0	0	1	0	4	0	1	0	0	0	0	31
51	0	0	0	0	0	0	9	0	1	0	0	0	0	0
53	0	0	0	0	0	0	4	0	0	0	0	0	0	0
55	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57	0	0	0	0	0	0	4	0	0	0	0	0	0	0

```
%% Combining the county neighbor relationship between the five states in the United States
%% Build the SIS model to establish a spatial regional opioid drug propagation model

%% Step 1. Read in Five States' County neighbor relationship matrix
[id_1, ohi] = xlsread('Ohio.xlsx');
[id_2, ken] = xlsread('Kentucky.xlsx');
[id_3, ten] = xlsread('Pennsylvania.xlsx');
[~, vir] = xlsread('Virginia.xlsx');
[~, wes] = xlsread('West Virginia.xlsx');

id_2(:,2) = [];
id_4 = vir(:,1); id_4(1,:) = []; id_4(1,:) = [];
id_5 = wes(:,1); id_5(1,:) = []; id_5(1,:) = [];
```

```
%% Perform useless words substracting
ohi(:,1) = []; ohi(1,:) = [];
ken(:,1) = []; ken(1,:) = [];
ten(:,1) = []; ten(1,:) = [];
vir(:,1) = []; vir(1,:) = []; vir(1,:) = [];
wes(:,1) = []; wes(1,:) = []; wes(1,:) = [];

%% New every five states basic adjency matrix -- Base on the county
size
ohi_size = size(id_1, 1); ken_size = size(id_2, 1);
ten_size = size(id_3, 1); vir_size = size(id_4, 1);
wes_size = size(id_5, 1);

map_2 = zeros(1000, 1);
for i=1:ken_size
   map_2(id_2(i)) = i;
end

map_4 = zeros(1000, 1);
for i=1:vir_size
   map_4(str2num(char(id_4(i)))) = i;
end

ohi_adj = zeros(ohi_size, ohi_size);
ken_adj = zeros(ken_size, ken_size);
ten_adj = zeros(ten_size, ten_size);
vir_adj = zeros(vir_size, vir_size);
wes_adj = zeros(wes_size, wes_size);

for i=1:ohi_size
    row_index = round(id_1(i)/2);
    str = char(ohi(i));
    str_array = strsplit(str, ',');
    for j=1:size(str_array, 2)
        nei_char = char(str_array(j));
        neiID = str2num(nei_char);
        col_index = round(neiID/2);
        ohi_adj(row_index, col_index) = 1;
    end
end

row_index = 0;
for i=1:ken_size
    row_index = row_index + 1;
```

```
    str = char(ken(i));
    str_array = strsplit(str, ',');
    for j=1:size(str_array, 2)
        nei_char = char(str_array(j));
        neiID = str2num(nei_char);
        col_index = map_2(neiID);
        ken_adj(row_index, col_index) = 1;
    end
end

for i=1:ten_size
    temp = char(id_3(i));
    row_index = round(temp/2);
    str = char(ten(i));
    str_array = strsplit(str, ', ');
    for j=1:size(str_array, 2)
        nei_char = char(str_array(j));
        neiID = str2num(nei_char);
        col_index = round(neiID/2);
        ten_adj(row_index, col_index) = 1;
    end
end

row_index = 0;
for i=1:vir_size
    row_index = row_index + 1;
    str = char(vir(i));
    str_array = strsplit(str, ', ');
    for j=1:size(str_array, 2)
        nei_char = char(str_array(j));
        neiID = str2num(nei_char);
        col_index = map_4(neiID);
        vir_adj(row_index, col_index) = 1;
    end
end

for i=1:wes_size
    temp = char(id_5(i));
    temp = str2num(temp);
    row_index = round(temp/2);
    str = char(wes(i));
    str_array = strsplit(str, ', ');
    for j=1:size(str_array, 2)
        nei_char = char(str_array(j));
```

```
        neiID = str2num(nei_char);
        col_index = round(neiID/2);
        wes_adj(row_index, col_index) = 1;
    end
end

%% Step 2. Read in SyntheticOpioids and Heroin excel records, first
consider its exploration between counties
BETA = [0.1, 0.2, 0.3];
GAMMA = [0.3, 0.2, 0.1];
[~, syn] = xlsread('SyntheticOpioids.xls');
[~, her] = xlsread('Heroin.xls');

%% Treat a county as a node, containing beta as opioid transforming
rate
%% And gamma as opioid reports decreasing rate

%% Define Synthetic Opioids' transfering process between Counties'
reports, beta and gamma matrix
ohi_syn = zeros(8, ohi_size, 4);
ken_syn = zeros(8, ken_size, 4);
ten_syn = zeros(8, ten_size, 4);
vir_syn = zeros(8, vir_size, 4);
wes_syn = zeros(8, wes_size, 4);

%% Define Heroin's transfering process between Counties' reports, beta
and gamma matrix
ohi_her = zeros(8, ohi_size, 3);
ken_her = zeros(8, ken_size, 3);
ten_her = zeros(8, ten_size, 3);
vir_her = zeros(8, vir_size, 3);
wes_her = zeros(8, wes_size, 3);

%% Fill beta and gamma matrix using drug reports range threshold
syn(1,:) = [];
her(1,:) = [];
row_size = size(syn, 1);
row_size_1 = size(her, 1);
syn_drugs =
{'Fentanyl';'Pethidine';'Propoxyphene';'Dextropropoxyphene';'Meth
adone';'Pentazocine';'Buprenorphine';'Nalbuphine';'Tramadol';'Rem
ifentanil';'Butorphanol';'Mitragynine';'Levorphanol'};

for i=1:row_size
```

```
year = str2num(char(syn(i, 1))) - 2010 + 1;
state_str = char(syn(i, 2));
county_id = str2num(char(syn(i, 5)));
syn_kind = char(syn(i, 7));
report_num = str2num(char(syn(i, 8)));
beta = 0;
gamma = 0;
syn_id = 0;
index = round(county_id/2);

for tt=1:13
   if isequal(char(syn_drugs(tt)), syn_kind)
      syn_id = tt;
      break;
   end
end

if report_num < 10
    beta = BETA(1);
    gamma = GAMMA(1);
elseif report_num < 30
    beta = BETA(2);
    gamma = GAMMA(2);
else
    beta = BETA(3);
    gamma = GAMMA(3);
end

switch state_str
    case 'OH'
        ohi_syn(year, index, 1) = report_num;
        ohi_syn(year, index, 2) = beta;
        ohi_syn(year, index, 3) = gamma;
        ohi_syn(year, index, 4) = syn_id;
    case 'KY'
        ken_syn(year, map_2(county_id), 1) = report_num;
        ken_syn(year, map_2(county_id), 2) = beta;
        ken_syn(year, map_2(county_id), 3) = gamma;
        ken_syn(year, map_2(county_id), 4) = syn_id;
    case 'PA'
        ten_syn(year, index, 1) = report_num;
        ten_syn(year, index, 2) = beta;
        ten_syn(year, index, 3) = gamma;
        ten_syn(year, index, 4) = syn_id;
```

```
        case 'VA'
            vir_syn(year, map_4(county_id), 1) = report_num;
            vir_syn(year, map_4(county_id), 2) = beta;
            vir_syn(year, map_4(county_id), 3) = gamma;
            vir_syn(year, map_4(county_id), 4) = syn_id;
        case 'WV'
            wes_syn(year, index, 1) = report_num;
            wes_syn(year, index, 2) = beta;
            wes_syn(year, index, 3) = gamma;
            wes_syn(year, index, 4) = syn_id;
    end
end

for i=1:row_size_1
    year = str2num(char(her(i, 1))) - 2010 + 1;
    state_str = char(her(i, 2));
    county_id = str2num(char(her(i, 5)));
    report_num = str2num(char(her(i, 8)));
    beta = 0;
    gamma = 0;
    index = round(county_id/2);

    if report_num < 10
        beta = BETA(1);
        gamma = GAMMA(1);
    elseif report_num < 30
        beta = BETA(2);
        gamma = GAMMA(2);
    else
        beta = BETA(3);
        gamma = GAMMA(3);
    end

    switch state_str
        case 'OH'
            ohi_her(year, index, 1) = report_num;
            ohi_her(year, index, 2) = beta;
            ohi_her(year, index, 3) = gamma;
        case 'KY'
            ken_her(year, map_2(county_id), 1) = report_num;
            ken_her(year, map_2(county_id), 2) = beta;
            ken_her(year, map_2(county_id), 3) = gamma;
        case 'PA'
            ten_her(year, index, 1) = report_num;
```

```
            ten_her(year, index, 2) = beta;
            ten_her(year, index, 3) = gamma;
        case 'VA'
            if map_4(county_id) ~= 0
                vir_her(year, map_4(county_id), 1) = report_num;
                vir_her(year, map_4(county_id), 2) = beta;
                vir_her(year, map_4(county_id), 3) = gamma;
            end
        case 'WV'
            wes_her(year, index, 1) = report_num;
            wes_her(year, index, 2) = beta;
            wes_her(year, index, 3) = gamma;
    end
end

%% Step 3. Use Odinary Difference equation to forcast growing and
transfering model -- SIS
ohi_syn_result = zeros(13, 8, ohi_size);
ken_syn_result = zeros(13, 8, ken_size);
ten_syn_result = zeros(13, 8, ten_size);
vir_syn_result = zeros(13, 8, vir_size);
wes_syn_result = zeros(13, 8, wes_size);

ohi_her_result = zeros(8, ohi_size);
ken_her_result = zeros(8, ken_size);
ten_her_result = zeros(8, ten_size);
vir_her_result = zeros(8, vir_size);
wes_her_result = zeros(8, wes_size);

%% For Synethic Opioids classification using SIS model forecasting
for i=1:8
    for j=1:ohi_size
        result_report_num = 0;
        src_syn_id = ohi_syn(i, j, 4);
        if src_syn_id == 0
            continue;
        end
        neigh_sum = 0;
        for k=1:ohi_size
            if src_syn_id == ohi_syn(i, k, 4)
                result_report_num = result_report_num + ohi_adj(j, k) *
ohi_syn(i, k, 1);
            end
            if ohi_adj(j, k) == 1
```

```
          neigh_sum = neigh_sum + 1;
        end
      end
      ohi_syn_result(src_syn_id, i, j) = round((round(ohi_syn(i, j,
2) * result_report_num - ohi_syn(i, j, 3) * ohi_syn(i, j, 1))) /
neigh_sum);
      if ohi_syn_result(src_syn_id, i, j) < 0
         ohi_syn_result(src_syn_id, i, j) = 0;
      end
   end
end

for i=1:8
   for j=1:ken_size
      result_report_num = 0;
      src_syn_id = ken_syn(i, j, 4);
      if src_syn_id == 0
         continue;
      end
      neigh_sum = 0;
      for k=1:ken_size
         if src_syn_id == ken_syn(i, k, 4)
            result_report_num = result_report_num + ken_adj(j, k) *
ken_syn(i, k, 1);
         end
         if ken_adj(j, k) == 1
            neigh_sum = neigh_sum + 1;
         end
      end
      ken_syn_result(src_syn_id, i, j) = round((round(ken_syn(i, j,
2) * result_report_num - ken_syn(i, j, 3) * ken_syn(i, j, 1))) /
neigh_sum);
      if ken_syn_result(src_syn_id, i, j) < 0
         ken_syn_result(src_syn_id, i, j) = 0;
      end
   end
end

for i=1:8
   for j=1:ten_size
      result_report_num = 0;
      src_syn_id = ten_syn(i, j, 4);
      if src_syn_id == 0
         continue;
```

```
      end
      neigh_sum = 0;
      for k=1:ten_size
         if src_syn_id == ten_syn(i, k, 4)
            result_report_num = result_report_num + ten_adj(j, k) *
ten_syn(i, k, 1);
         end
         if ten_adj(j, k) == 1
           neigh_sum = neigh_sum + 1;
         end
      end
      ten_syn_result(src_syn_id, i, j) = round((round(ten_syn(i, j,
2) * result_report_num - ten_syn(i, j, 3) * ten_syn(i, j, 1))) /
neigh_sum);
      if ten_syn_result(src_syn_id, i, j) < 0
         ten_syn_result(src_syn_id, i, j) = 0;
      end
   end
end

for i=1:8
   for j=1:vir_size
      result_report_num = 0;
      src_syn_id = vir_syn(i, j, 4);
      if src_syn_id == 0
         continue;
      end
      neigh_sum = 0;
      for k=1:vir_size
         if src_syn_id == vir_syn(i, k, 4)
            result_report_num = result_report_num + vir_adj(j, k) *
vir_syn(i, k, 1);
         end
         if vir_adj(j, k) == 1
           neigh_sum = neigh_sum + 1;
         end
      end
      vir_syn_result(src_syn_id, i, j) = round((round(vir_syn(i, j,
2) * result_report_num - vir_syn(i, j, 3) * vir_syn(i, j, 1))) /
neigh_sum);
      if vir_syn_result(src_syn_id, i, j) < 0
         vir_syn_result(src_syn_id, i, j) = 0;
      end
   end
```

```
end

for i=1:8
   for j=1:wes_size
      result_report_num = 0;
      src_syn_id = wes_syn(i, j, 4);
      if src_syn_id == 0
         continue;
      end
      neigh_sum = 0;
      for k=1:wes_size
         if src_syn_id == wes_syn(i, k, 4)
            result_report_num = result_report_num + wes_adj(j, k) *
wes_syn(i, k, 1);
         end
         if wes_adj(j, k) == 1
           neigh_sum = neigh_sum + 1;
         end
      end
      wes_syn_result(src_syn_id, i, j) = round((round(wes_syn(i, j,
2) * result_report_num - wes_syn(i, j, 3) * wes_syn(i, j, 1))) /
neigh_sum);
      if wes_syn_result(src_syn_id, i, j) < 0
         wes_syn_result(src_syn_id, i, j) = 0;
      end
   end
end

%% For Heroin classification using SIS model forecasting
for i=1:8
   for j=1:ohi_size
      result_report_num = 0;
      neigh_sum = 0;
      for k=1:ohi_size
         result_report_num = result_report_num + ohi_adj(j, k) *
ohi_her(i, k, 1);
         if ohi_adj(j, k) == 1
           neigh_sum = neigh_sum + 1;
         end
      end
      ohi_her_result(i, j) = round((round(ohi_her(i, j, 2) *
result_report_num - ohi_her(i, j, 3) * ohi_her(i, j, 1))) / neigh_sum);
      if ohi_her_result(i, j) < 0
         ohi_her_result(i, j) = 0;
```

```
    end
  end
end

for i=1:8
  for j=1:ken_size
    result_report_num = 0;
    neigh_sum = 0;
    for k=1:ken_size
      result_report_num = result_report_num + ken_adj(j, k) *
ken_her(i, k, 1);
      if ken_adj(j, k) == 1
        neigh_sum = neigh_sum + 1;
      end
    end
    ken_her_result(i, j) = round((round(ken_her(i, j, 2) *
result_report_num - ken_her(i, j, 3) * ken_her(i, j, 1))) / neigh_sum);
    if ken_her_result(i, j) < 0
      ken_her_result(i, j) = 0;
    end
  end
end

for i=1:8
  for j=1:ten_size
    result_report_num = 0;
    neigh_sum = 0;
    for k=1:ten_size
      result_report_num = result_report_num + ten_adj(j, k) *
ten_her(i, k, 1);
      if ten_adj(j, k) == 1
        neigh_sum = neigh_sum + 1;
      end
    end
    ten_her_result(i, j) = round((round(ten_her(i, j, 2) *
result_report_num - ten_her(i, j, 3) * ten_her(i, j, 1))) / neigh_sum);
    if ten_her_result(i, j) < 0
      ten_her_result(i, j) = 0;
    end
  end
end

for i=1:8
  for j=1:vir_size
```

```
      result_report_num = 0;
      neigh_sum = 0;
      for k=1:vir_size
         result_report_num = result_report_num + vir_adj(j, k) *
vir_her(i, k, 1);
         if vir_adj(j, k) == 1
            neigh_sum = neigh_sum + 1;
         end
      end
      vir_her_result(i, j) = round((round(vir_her(i, j, 2) *
result_report_num - vir_her(i, j, 3) * vir_her(i, j, 1))) / neigh_sum);
      if vir_her_result(i, j) < 0
          vir_her_result(i, j) = 0;
      end
   end
end

for i=1:8
   for j=1:wes_size
      result_report_num = 0;
      neigh_sum = 0;
      for k=1:wes_size
         result_report_num = result_report_num + wes_adj(j, k) *
wes_her(i, k, 1);
         if wes_adj(j, k) == 1
            neigh_sum = neigh_sum + 1;
         end
      end
      wes_her_result(i, j) = round((round(wes_her(i, j, 2) *
result_report_num - wes_her(i, j, 3) * wes_her(i, j, 1))) / neigh_sum);
      if wes_her_result(i, j) < 0
          wes_her_result(i, j) = 0;
      end
   end
end

%% Step 4. Output SIS Model's county opioids drug reports excel tables
for i=1:ohi_size
   temp = ohi_syn_result(:,:,i);
   tt = ohi_her_result(:,i);
   temp = [temp;tt'];
   xlswrite('SIS_Ohi.xlsx', temp, i);
end
```

```
for i=1:ken_size
   temp = ken_syn_result(:,:,i);
   tt = ken_her_result(:,i);
   temp = [temp;tt'];
   xlswrite('SIS_Ken.xlsx', temp, i);
end

for i=1:ten_size
   temp = ten_syn_result(:,:,i);
   tt = ten_her_result(:,i);
   temp = [temp;tt'];
   xlswrite('SIS_Pen.xlsx', temp, i);
end

for i=1:vir_size
   temp = vir_syn_result(:,:,i);
   tt = vir_her_result(:,i);
   temp = [temp;tt'];
   xlswrite('SIS_Vir.xlsx', temp, i);
end

for i=1:wes_size
   temp = wes_syn_result(:,:,i);
   tt = wes_her_result(:,i);
   temp = [temp;tt'];
   xlswrite('SIS_Wes.xlsx', temp, i);
end

%% Create ARIMA Model to analyse county's drug report growing
trends with time sequences
[id_1, ohi] = xlsread('Ohio.xlsx');
[id_2, ken] = xlsread('Kentucky.xlsx');
[id_3, ten] = xlsread('Pennsylvania.xlsx');
[~, vir] = xlsread('Virginia.xlsx');
[~, wes] = xlsread('West Virginia.xlsx');

id_2(:,2) = [];
id_4 = vir(:,1); id_4(1,:) = []; id_4(1,:) = [];
id_5 = wes(:,1); id_5(1,:) = []; id_5(1,:) = [];


%% Perform useless words substracting
ohi(:,1) = []; ohi(1,:) = [];
ken(:,1) = []; ken(1,:) = [];
```

```
ten(:,1) = []; ten(1,:) = [];
vir(:,1) = []; vir(1,:) = []; vir(1,:) = [];
wes(:,1) = []; wes(1,:) = []; wes(1,:) = [];

%% New every five states basic adjency matrix -- Base on the
county size
ohi_size = size(id_1, 1); ken_size = size(id_2, 1);
ten_size = size(id_3, 1); vir_size = size(id_4, 1);
wes_size = size(id_5, 1);

%% Step 1. Collect Synethic Opidiods and Heroin's reports
number in counties
[~, syn] = xlsread('SyntheticOpioids.xls');
[~, her] = xlsread('Heroin.xls');

ohi_syn = zeros(13, 8, ohi_size);
ken_syn = zeros(13, 8, ken_size);
ten_syn = zeros(13, 8, ten_size);
vir_syn = zeros(13, 8, vir_size);
wes_syn = zeros(13, 8, wes_size);

ohi_her = zeros(8, ohi_size);
ken_her = zeros(8, ken_size);
ten_her = zeros(8, ten_size);
vir_her = zeros(8, vir_size);
wes_her = zeros(8, wes_size);

%% Read in map_2 and map_4
ken_size = size(id_2, 1);
vir_size = size(id_4, 1);

map_2 = zeros(1000, 1);
for i=1:ken_size
   map_2(id_2(i)) = i;
end

map_4 = zeros(1000, 1);
for i=1:vir_size
   map_4(str2num(char(id_4(i)))) = i;
end

%% Clear useless datas and define different syn-drugs id
syn(1,:) = [];
her(1,:) = [];
```

```
row_size = size(syn, 1);
row_size_1 = size(her, 1);
syn_drugs =
{'Fentanyl';'Pethidine';'Propoxyphene';'Dextropropoxyphen
e';'Methadone';'Pentazocine';'Buprenorphine';'Nalbuphine';
'Tramadol';'Remifentanil';'Butorphanol';'Mitragynine';'Le
vorphanol'};

for i=1:row_size
   year = str2num(char(syn(i, 1))) - 2010 + 1;
   state_str = char(syn(i, 2));
   county_id = str2num(char(syn(i, 5)));
   drug_kind = char(syn(i, 7));
   drug_id = 0;
   report_num = str2num(char(syn(i, 8)));
   index = round(county_id/2);

   for j=1:13
      if isequal(char(syn_drugs(j)), drug_kind)
         drug_id = j;
         break;
      end
   end

   switch state_str
      case 'OH'
         ohi_syn(drug_id, year, index) = report_num;
      case 'KY'
         ken_syn(drug_id, year, map_2(county_id)) =
report_num;
      case 'PA'
         ten_syn(drug_id, year, index) = report_num;
      case 'VA'
         vir_syn(drug_id, year, map_4(county_id)) =
report_num;
      case 'WV'
         wes_syn(drug_id, year, index) = report_num;
   end
end

for i=1:row_size_1
   year = str2num(char(syn(i, 1))) - 2010 + 1;
   state_str = char(syn(i, 2));
   county_id = str2num(char(syn(i, 5)));
```

```
    report_num = str2num(char(syn(i, 8)));
    index = round(county_id/2);

    switch state_str
        case 'OH'
            ohi_her(year, index) = report_num;
        case 'KY'
            ken_her(year, map_2(county_id)) = report_num;
        case 'PA'
            ten_her(year, index) = report_num;
        case 'VA'
            vir_her(year, map_4(county_id)) = report_num;
        case 'WV'
            wes_her(year, index) = report_num;
    end
end

%% Step 2. For five states, we are here to get sums of all
ARIMA aic to decide best q and p
%% And build ARIMA Model
ohi_arima_predict = zeros(14, 8, ohi_size);
ken_arima_predict = zeros(14, 8, ken_size);
ten_arima_predict = zeros(14, 8, ten_size);
vir_arima_predict = zeros(14, 8, vir_size);
wes_arima_predict = zeros(14, 8, wes_size);

p = 1;
q = 1;
for k=1:ohi_size
    for t=1:14
        sequence = zeros(8, 1);
        if t < 14
            for v=1:8
                sequence(v) = ohi_syn(t, v, k);
            end
        else
            for v=1:8
                sequence(v) = ohi_her(v, k);
            end
        end
        z = iddata(sequence);
        model = armax(z, [p q]);
        sequence = [sequence;0];
        yp = predict(model, sequence, 1);
```

```
      for temp=1:9
         if yp(temp) < 0
            yp(temp) = 0;
         end
      end
      for temp=2:9
         ohi_arima_predict(t, temp-1, k) = round(yp(temp));
      end
   end
end

for k=1:ken_size
   for t=1:14
      sequence = zeros(8, 1);
      if t < 14
        for v=1:8
           sequence(v) = ken_syn(t, v, k);
        end
      else
        for v=1:8
           sequence(v) = ken_her(v, k);
        end
      end
      z = iddata(sequence);
      model = armax(z, [p q]);
      sequence = [sequence;0];
      yp = predict(model, sequence, 1);
      for temp=1:9
         if yp(temp) < 0
            yp(temp) = 0;
         end
      end
      for temp=2:9
         ken_arima_predict(t, temp-1, k) = round(yp(temp));
      end
   end
end

for k=1:ten_size
   for t=1:14
      sequence = zeros(8, 1);
      if t < 14
        for v=1:8
           sequence(v) = ten_syn(t, v, k);
```

```
      end
    else
      for v=1:8
        sequence(v) = ten_her(v, k);
      end
    end
    z = iddata(sequence);
    model = armax(z, [p q]);
    sequence = [sequence;0];
    yp = predict(model, sequence, 1);
    for temp=1:9
      if yp(temp) < 0
        yp(temp) = 0;
      end
    end
    for temp=2:9
      ten_arima_predict(t, temp-1, k) = round(yp(temp));
    end
  end
end

for k=1:vir_size
  for t=1:14
    sequence = zeros(8, 1);
    if t < 14
      for v=1:8
        sequence(v) = vir_syn(t, v, k);
      end
    else
      for v=1:8
        sequence(v) = vir_her(v, k);
      end
    end
    z = iddata(sequence);
    model = armax(z, [p q]);
    sequence = [sequence;0];
    yp = predict(model, sequence, 1);
    for temp=1:9
      if yp(temp) < 0
        yp(temp) = 0;
      end
    end
    for temp=2:9
      vir_arima_predict(t, temp-1, k) = round(yp(temp));
```

```
      end
   end
end

for k=1:wes_size
   for t=1:14
      sequence = zeros(8, 1);
      if t < 14
        for v=1:8
          sequence(v) = wes_syn(t, v, k);
        end
      else
        for v=1:8
          sequence(v) = wes_her(v, k);
        end
      end
      z = iddata(sequence);
      model = armax(z, [p q]);
      sequence = [sequence;0];
      yp = predict(model, sequence, 1);
      for temp=1:9
         if yp(temp) < 0
            yp(temp) = 0;
         end
      end
      for temp=2:9
         wes_arima_predict(t, temp-1, k) = round(yp(temp));
      end
   end
end

%% Step 3. Ouput ARIMA Model predict sequence to excel
for i=1:ohi_size
   temp = ohi_arima_predict(:,:,i);
   xlswrite('ARIMA_Ohi.xlsx', temp, i);
end

for i=1:ken_size
   temp = ken_arima_predict(:,:,i);
   xlswrite('ARIMA_Ken.xlsx', temp, i);
end

for i=1:ten_size
   temp = ten_arima_predict(:,:,i);
```

```
   xlswrite('ARIMA_Pen.xlsx', temp, i);
end

for i=1:vir_size
   temp = vir_arima_predict(:,:,i);
   xlswrite('ARIMA_Vir.xlsx', temp, i);
end

for i=1:wes_size
   temp = wes_arima_predict(:,:,i);
   xlswrite('ARIMA_Wes.xlsx', temp, i);
end
```

Part II

```
%% Use PLS methods to decide the weight of selected five factors
and Part1 values

%% Step 1. Read in five factors sequence and Part1 result
acs_10 = xlsread("ACS_10_5YR_DP02_with_ann.xlsx");
acs_11 = xlsread("ACS_11_5YR_DP02_with_ann.xlsx");
acs_12 = xlsread("ACS_12_5YR_DP02_with_ann.xlsx");
acs_13 = xlsread("ACS_13_5YR_DP02_with_ann.xlsx");
acs_14 = xlsread("ACS_14_5YR_DP02_with_ann.xlsx");
acs_15 = xlsread("ACS_15_5YR_DP02_with_ann.xlsx");
acs_16 = xlsread("ACS_16_5YR_DP02_with_ann.xlsx");

[~,~,nflis] = xlsread('nflis.xlsx', 'Data');
nflis(1,:) = [];

[~, meta] = xlsread('ACS_10_5YR_DP02_metadata.xlsx');
meta(1,:) = []; meta(1,:) = []; meta(1,:) = [];

%% Get origin county datas
[id_1, ohi] = xlsread('Ohio.xlsx');
[id_2, ken] = xlsread('Kentucky.xlsx');
[id_3, ten] = xlsread('Pennsylvania.xlsx');
[~, vir] = xlsread('Virginia.xlsx');
[~, wes] = xlsread('West Virginia.xlsx');

id_2(:,2) = [];
id_4 = vir(:,1); id_4(1,:) = []; id_4(1,:) = [];
id_5 = wes(:,1); id_5(1,:) = []; id_5(1,:) = [];

%%Base on the county size
```

```
ohi_size = size(id_1, 1); ken_size = size(id_2, 1);
ten_size = size(id_3, 1); vir_size = size(id_4, 1);
wes_size = size(id_5, 1);

map_2 = zeros(1000, 1);
for i=1:ken_size
   map_2(id_2(i)) = i;
end

map_4 = zeros(1000, 1);
for i=1:vir_size
   map_4(str2num(char(id_4(i)))) = i;
end

%% Record five states county's total drug report
ohi_store = zeros(ohi_size, 7);
ken_store = zeros(ken_size, 7);
ten_store = zeros(ten_size, 7);
vir_store = zeros(vir_size, 7);
wes_store = zeros(wes_size, 7);

for i=1:size(nflis, 1)
    year = cell2mat(nflis(i, 1)) - 2010 + 1;
    state_str = char(nflis(i, 2));
    county_id = str2num(cell2mat(nflis(i, 5)));
    report_num = cell2mat(nflis(i, 9));

    switch state_str
        case 'VA'
            if map_4(county_id) > 0
                vir_store(map_4(county_id), year) =
report_num;
            end
        case 'OH'
            ohi_store(round(county_id/2), year) = report_num;
        case 'PA'
            ten_store(round(county_id/2), year) = report_num;
        case 'KY'
            ken_store(map_2(county_id), year) = report_num;
        case 'WV'
            wes_store(round(county_id/2), year) = report_num;
    end
end
```

```
%% Get origin all kinds drug reports of every county's
2010-2016 col-matrix
origin_result = zeros(7, 5);

for i=1:ken_size
    for j=1:7
      origin_result(j, 1) = origin_result(j, 1) + ken_store(i,
j);
    end
end

for i=1:ohi_size
    for j=1:7
      origin_result(j, 2) = origin_result(j, 2) + ohi_store(i,
j);
    end
end

for i=1:ten_size
    for j=1:7
      origin_result(j, 3) = origin_result(j, 3) + ten_store(i,
j);
    end
end

for i=1:vir_size
    for j=1:7
      origin_result(j, 4) = origin_result(j, 4) + vir_store(i,
j);
    end
end

for i=1:wes_size
    for j=1:7
      origin_result(j, 5) = origin_result(j, 5) + wes_store(i,
j);
    end
end

%% Get five factors' report sequences
target_factors_id = [10,20,60,69,128];
target_factors_val = zeros(7, 5, 5);

for i=1:5
```

```
   col = (target_factors_id(i)-1) * 4 + 3;
   sum = 0;
   for countyIndex=1:463
       statics = [acs_10(countyIndex,
col);acs_11(countyIndex, col);acs_12(countyIndex,
col);acs_13(countyIndex, col);acs_14(countyIndex,
col);acs_15(countyIndex, col);acs_16(countyIndex, col)];
       if countyIndex <= ken_size
          target_factors_val(:, 1 ,i) = target_factors_val(:,
1 ,i) + statics;
       elseif countyIndex <= ken_size + ohi_size
           target_factors_val(:, 2 ,i) = target_factors_val(:,
2 ,i) + statics;
       elseif countyIndex <= ken_size + ohi_size + ten_size
           target_factors_val(:, 3 ,i) = target_factors_val(:,
3 ,i) + statics;
       elseif countyIndex <= ken_size + ohi_size + ten_size
+ vir_size
           target_factors_val(:, 4 ,i) = target_factors_val(:,
4 ,i) + statics;
       else
           target_factors_val(:, 5 ,i) = target_factors_val(:,
5 ,i) + statics;
       end
   end
end


%% Read Part1 results
sis_arima_result = zeros(7, 5);
sis_arima_result(1,:) = origin_result(1,:);

for i=1:ken_size
    result_ken = xlsread('Final_Ken.xlsx', i);
    for j=1:14
       for k=1:6
          if ~isnan(result_ken(j,k))
             sis_arima_result(k+1, 1) = sis_arima_result(k+1,
1) + result_ken(j, k);
          end
       end
    end
end
```

```
for i=1:ohi_size
    result_ohi = xlsread('Final_Ohi.xlsx', i);
    for j=1:14
       for k=1:6
          sis_arima_result(k+1, 2) = sis_arima_result(k+1, 2)
+ result_ohi(j, k);
       end
    end
end

for i=1:ten_size
    result_ten = xlsread('Final_Pen.xlsx', i);
    for j=1:14
       for k=1:6
          if ~isnan(result_ten(j, k))
            sis_arima_result(k+1, 3) = sis_arima_result(k+1,
3) + result_ten(j, k);
          end
       end
    end
end

for i=1:vir_size
    result_vir = xlsread('Final_Vir.xlsx', i);
    for j=1:14
       for k=1:6
          if ~isnan(result_vir(j, k))
            sis_arima_result(k+1, 4) = sis_arima_result(k+1,
4) + result_vir(j, k);
          end
       end
    end
end

for i=1:wes_size
    result_wes = xlsread('Final_Wes.xlsx', i);
    for j=1:14
       for k=1:6
          if ~isnan(result_wes(j, k))
             sis_arima_result(k+1, 5) = sis_arima_result(k+1,
5) + result_wes(j, k);
          end
       end
    end
```

```
end

%% Step 2. Combine the five factors sequence and arima-sis
result sequence and origin sequence
result = zeros(7, 7);

for i=1:5
    for j=1:5
       result(:,j) = result(:,j) + target_factors_val(:,i,j);
    end

    result(:,6) = result(:,6) + sis_arima_result(:, i);
    result(:,7) = result(:,7) + origin_result(:, i);
end

%% Begin PLS
mu = mean(result);
sig = std(result);

zs = zscore(result);

a = zs(:, 1:6);
b = zs(:, 7);

ncomp = 2;
[xl, yl, xs, ys, beta, pctvar, mse, stats] = plsregress(a,
b, ncomp);
constr = cumsum(pctvar, 2);

n = size(a, 2);
m = size(b, 2);

res(1, :) = mu(n+1:end) + mu(1:n) ./ sig(1:n) *
beta(2:end, :) .* sig(n+1:end);
res(2:n+1, :) = (1 ./ sig(1:n))' * sig(n+1:end) .*
beta(2:end, :);

res = [];
%% Step 3.Use new model to predict sequence result
c = res(1);
w = res(2:end);

%% Read every county five factors datas
target_factors_county_val = zeros(7, 463, 5);
```

```
for i=1:5
   col = (target_factors_id(i)-1) * 4 + 3;
   sum = 0;
   for countyIndex=1:463
       statics = [acs_10(countyIndex,
col);acs_11(countyIndex, col);acs_12(countyIndex,
col);acs_13(countyIndex, col);acs_14(countyIndex,
col);acs_15(countyIndex, col);acs_16(countyIndex, col)];
       target_factors_county_val(:, countyIndex ,i) =
statics;
   end
end

%% Read every county ARIMA-SIS model result datas
origin_result = zeros(7, 463);

for i=1:ken_size
    for j=1:7
      origin_result(j, i) = origin_result(j, i) + ken_store(i,
j);
    end
end

for i=1:ohi_size
    for j=1:7
      origin_result(j, i+ken_size) = origin_result(j,
i+ken_size) + ohi_store(i, j);
    end
end

for i=1:ten_size
    for j=1:7
      origin_result(j, i+ken_size+ohi_size) =
origin_result(j, i+ken_size+ohi_size) + ten_store(i, j);
    end
end

for i=1:vir_size
    for j=1:7
      origin_result(j, i+ken_size+ohi_size+ten_size) =
origin_result(j, i+ken_size+ohi_size+ten_size) +
vir_store(i, j);
    end
```

```
end

for i=1:wes_size
    for j=1:7
      origin_result(j,
i+ken_size+ohi_size+ten_size+vir_size) = origin_result(j,
i+ken_size+ohi_size+ten_size+vir_size) + wes_store(i, j);
    end
end

sis_arima_result = zeros(7, 463);
sis_arima_result(1,:) = origin_result(1,:);

for i=1:ken_size
    result_ken = xlsread('Final_Ken.xlsx', i);
    for j=1:14
      for k=1:6
        if ~isnan(result_ken(j,k))
          sis_arima_result(k+1, i) = sis_arima_result(k+1,
i) + result_ken(j, k);
        end
      end
    end
end

for i=1:ohi_size
    result_ohi = xlsread('Final_Ohi.xlsx', i);
    for j=1:14
      for k=1:6
        sis_arima_result(k+1, i+ken_size) =
sis_arima_result(k+1, i+ken_size) + result_ohi(j, k);
      end
    end
end

for i=1:ten_size
    result_ten = xlsread('Final_Pen.xlsx', i);
    for j=1:14
      for k=1:6
        if ~isnan(result_ten(j, k))
          sis_arima_result(k+1, i+ken_size+ohi_size) =
sis_arima_result(k+1, i+ken_size+ohi_size) + result_ten(j,
k);
        end
```

```
      end
    end
end

for i=1:vir_size
    result_vir = xlsread('Final_Vir.xlsx', i);
    for j=1:14
      for k=1:6
        if ~isnan(result_vir(j, k))
          sis_arima_result(k+1,
i+ken_size+ohi_size+ten_size) = sis_arima_result(k+1,
i+ken_size+ohi_size+ten_size) + result_vir(j, k);
        end
      end
    end
end

for i=1:wes_size
    result_wes = xlsread('Final_Wes.xlsx', i);
    for j=1:14
      for k=1:6
        if ~isnan(result_wes(j, k))
          sis_arima_result(k+1,
i+ken_size+ohi_size+vir_size+ten_size) =
sis_arima_result(k+1,
i+ken_size+ohi_size+vir_size+ten_size) + result_wes(j, k);
        end
      end
    end
end

%% Define result model with pls result
part2_result = zeros(7, 463);
for i=1:463
   for j=1:7
     temp = round(w(1) * target_factors_county_val(j, i, 1)
+ w(2) * target_factors_county_val(j, i, 2)+ w(3) *
target_factors_county_val(j, i, 3) + w(4) *
target_factors_county_val(j, i, 4)+ w(5) *
target_factors_county_val(j, i, 5) + w(6) *
sis_arima_result(j, i));
     if temp > 0
       part2_result(j, i) = temp;
     else
```

```
            part2_result(j, i) = 0;
        end
    end
end
```