For office use only	Team Control Number	For office use only
T1 ________________	**93374**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**C**	F4 ________________

2018
MCM/ICM
Summary Sheet

Assessment of Renewable Energy Usage: A Combined Model of Multiple Linear Analysis and Grey Method

Summary

In order to accurately describe and assess the renewable energy usage profile of the four states: Arizona, California, New Mexico, and Texas, our team define the index Renewable Energy Ratio(RER) to be the ratio of cleaner, renewable energy to total energy consumption. We build a combined model of multiple linear analysis and grey method to characterize and predict the profile which is represented by RER. This model qualitatively characterizes the relationship between **RER** and a unique set of variables including energy price, energy production, GDP and population for each state. Those relationship reveals several important similarities and differences between states, and base on which policy suggestions can be proposed.

The similarity we found includes that RER of California and Texas is significantly related to GDP($p < 0.001$), which indicates that renewable energy usage is primarily promoted by commercial sector in these two states. Our team also find that different states have different renewable energy components as well. For example, Arizona's RER is significantly related to electricity production by nuclear power ($p < 0.001$), indicating that Arizona heavily relies on nuclear energy which is considered as a kind of clean energy by our team despite the raging debate.

The prediction of our model is quite interesting in that RER of each state either approaches 0 or 1. Only state that heavily relies on nuclear energy (Arizona) will be able to replace fossil fuel by renewable energy completely. As surprising as this result is, it is reasonable in the sense that apart from nuclear energy, the other renewable energy may not be able to satisfy the rapidly increasing energy demand. Our result, of course, cannot be the whole picture. Instead, it is only our team's insight regarding the provided data set. As history has told us, mankind's boundary will be pushed whenever boundary is reached. Our team believe that more efficient and cleaner renewable energy will be developed in the near future.

Governor Greg Abbott
Governor Jerry Brown
Governor Doug Ducey
Governor Susana Martinez

Dear Governors,

I write to you on behalf of the MCM team 93374. We are extremely concerned about the cleaner, renewable energy usage in Arizona, California, New Mexico, and Texas. Study shows that the ratio of the use of cleaner, renewable energy to total energy has much room for improvement. We hope to help you achieve the goal of increasing the ratio as much as possible by promoting cleaner, renewable energy use.

The energy profile in 2009 shows the ratio of cleaner, renewable energy to total energy consumed (**RER**) by each state. In Arizona, the ratio turns out to be 29.17%, which is the highest of all four states. In California, the ratio is 12.91%. In New Mexico, cleaner, renewable energy consists of 5.32% of all energy usage. In Texas, 3.16% of total energy is cleaner, renewable energy.

According to our model, it is predicted that the **RER** of the four states can barely maintain the present levels in 2025, and all states, except Arizona, which will achieve a 100

In order to accomplish the goals, four states should work together. First of all, it is beneficial to develop as much nuclear power plants as possible to generate the stable renewable energy and raise the ratio of cleaner, renewable energy. In the second place, safety measures in the nuclear power plants should be improved to guarantee safer usage of nuclear power. Moreover, more transmission lines can be constructed between Arizona and adjacent states such as New Mexico and California in order to transmit nuclear energy to other states that Arizona produced. Also, each state should make full use of their geological advantages and climate features to develop more suitable sources of energy. For example, more wind farms should be built in Arizona, and California should increase the use of hydroelectricity. Moreover, the government should grow wind energy and wood and waste energy in New Mexico. Also, more facilities should be built to utilize wind energy, wood and waste energy in Texas.

We appreciate all that you do for the states, and we look forward to seeing positive changes. We hope you can accept our suggestion. Sincerely,

Team # 93374

Contents

1 Introduction

1.1 Problem Statement

Along the U.S. border with Mexico, there are four states – California (CA), Arizona (AZ), New Mexico (NM), and Texas (TX) – that wish to form a realistic new energy compact focused on increased usage of cleaner, renewable energy sources. Our team has been asked by the four governors of these states to perform data analysis and modeling to inform their development of a set of goals for their interstate energy compact.

We rely on the attached data file **"ProblemCData.xlsx"** provided in the first worksheet (**"seseds"**) 50 years of data, from 1960 to 2009, in 605 variables on each of these four states' energy production and consumption, along with some demographic and economic information. The 605 variable names used in this dataset are defined in the second worksheet (**"msncodes"**).

We used multiple linear analysis to characterize how the energy profile of each of the four states has evolved from 1960 - 2009.

We used Grey method to predict the energy profile of each state, for 2025 and 2050 in the absence of any policy changes by each governor's office.

Based on our model, we listed several practical suggestions to the local government.

1.2 Background Knowledge

The given data file "ProblemCData.xlsx" comes from State Energy Data System (SEDS) Complete Dataset through 2009 (All 50 states), which is compiled by the U.S. Energy Information Administration (EIA). On EIA's website, we found technical notes and documentation, which can help us understand the details of and relationships between our given data.

The "msncodes" are names of 5 letters, ranging from ABICB to WYTCB. The first 2 letters of a MSN code indicate the energy type. The following 2 letters indicate the energy activity or energy-consuming sector. The last letter indicates the data type. For example, "**TE**" means total energy, "**RE**" means renewable energy, "**TC**" means total consumption, and "**B**" means the data is in British thermal units (billion Btu).

Particularly, the following background knowledge is essential to our problem-solving process.

Classification of cleaner, renewable energy Renewable energy are energy resources that are naturally replenishing but flow-limited. Cleaner energy are those which would produce less pollution. In this paper , cleaner, renewable sources of energy include

- **biomass** Organic non-fossil material of biological origin constituting a renewable energy source. Code: **BM**
- **fuel ethanol** Ethanol intended for fuel use. Code: **EM**

- **geothermal** Hot water or steam extracted from geothermal reservoirs in the Earth's crust. Code: **GE**
- **hydroelectric power** The use of flowing water to produce electric power. Code: **HY**
- **nuclear power** Electricity generated by the use of the thermal energy released from the fission of nuclear fuel in a reactor. Code: **NU**
- **solar energy** The radiant energy of the sun, which can be converted into other forms of energy, such as heat or electricity. Code: **SO**
- **wind energy** Kinetic energy present in wind motion that can be converted to mechanical energy for driving pumps, mills, and electric power generators. Code: **WY**
- **wood and waste energy** Wood, wood products, and waste used as fuel. Code: **WW**

2 Basic Energy Profile of Four States

2.1 Criterion and the "Best"

We create the energy profiles based on the **RER** of cleaner, renewable energy consumption to total energy consumption of the four states.

For one state, the **RER** =

$$\frac{RETCB + NUETB}{TETCB}$$

, where

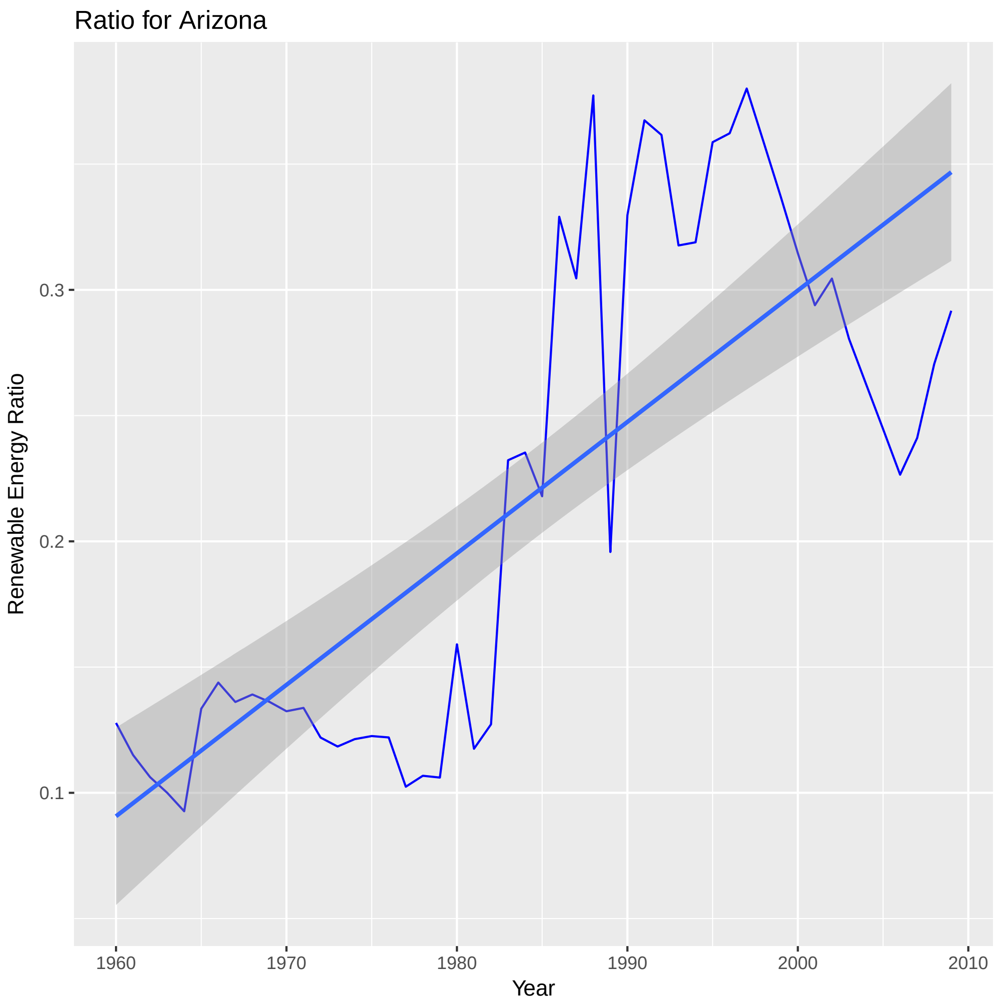
$$RETCB = EMTCB + EMTCB + GETCB + HYTCB + SOTCB + WWTCB + WYTCB$$

Thus, the criterion of deciding the "best" profile for the use of cleaner, renewable energy is which state has the highest **RER**.

Comparing the **RER** of four states, as of 2009, Arizona (29.17%) has the highest **RER**, much higher than California (12.91%), New Mexico (5.32%), and Texas (3.16%).

Thus, Arizona has the "best" profile for use of cleaner, renewable energy in 2009.

2.2 Arizona

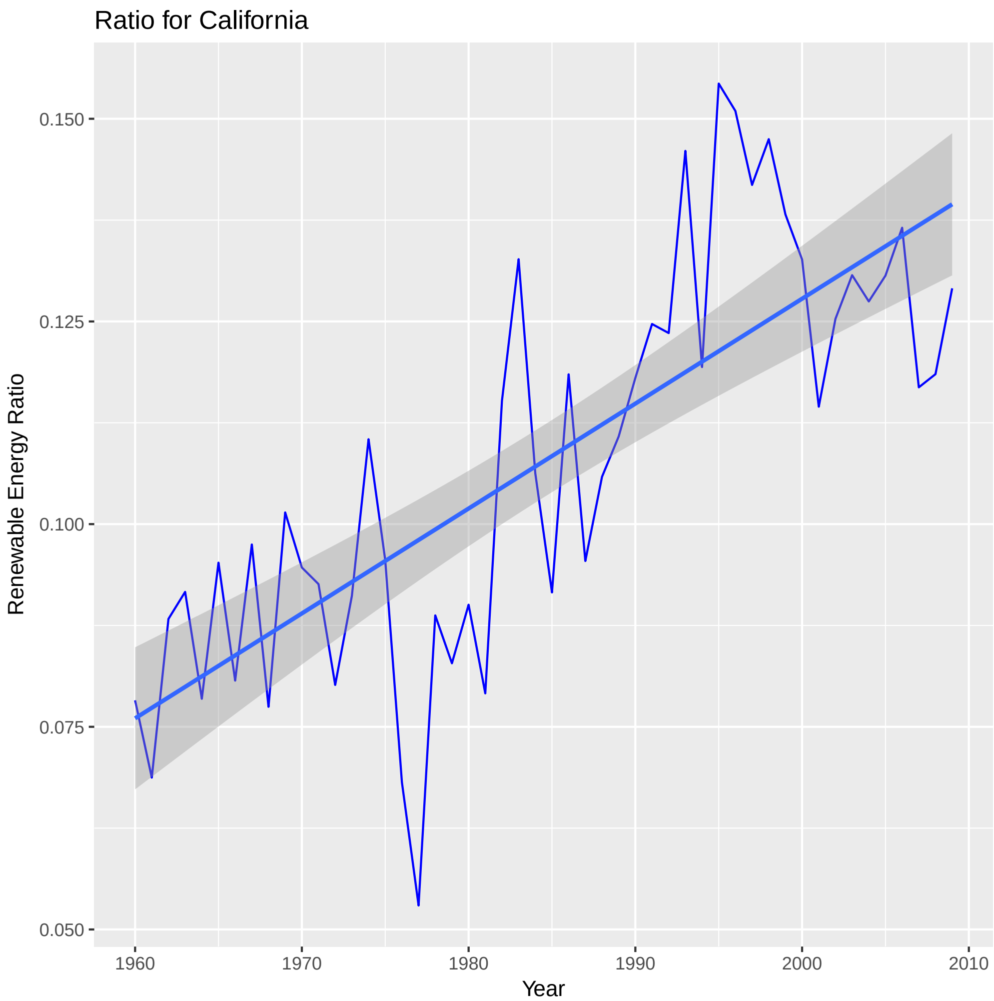


Arizona is known for its iconic vistas from the Grand Canyon in the north to the Saguaro deserts in the south. The state has few fossil fuel resources, but it does have abundant solar and geothermal energy potential. Along the more than 100-mile long, steep slope of the Mogollon Rim that marks the southern edge of the Colorado Plateau is some of Arizona's greatest wind potential. Also, abundant sunshine gives the entire state some of the nation's greatest solar power potential. Arizona is rich in minerals, and the state drew Spanish explorers seeking gold, silver, and copper as early as the 1600s.

Because Arizona's primary economic activities are not energy intensive, the state's per capita energy consumption is among the lowest in the nation. The transportation sector is Arizona's largest end-use energy consumer, followed by the residential sector. The majority of Arizona's residents live in a few urban areas, leaving the rest of the state lightly populated. Mild summers in the north and mild winters in the south make Arizona a popular vacation and retirement destination.

The **RER** shows an overall trend of increasing despite some fluctuations from 12.78% to 29.17% in Arizona. Despite the sudden decrease in the ratio from 2002 to 2003, the ratio started to increase again since 2006.

2.3 California

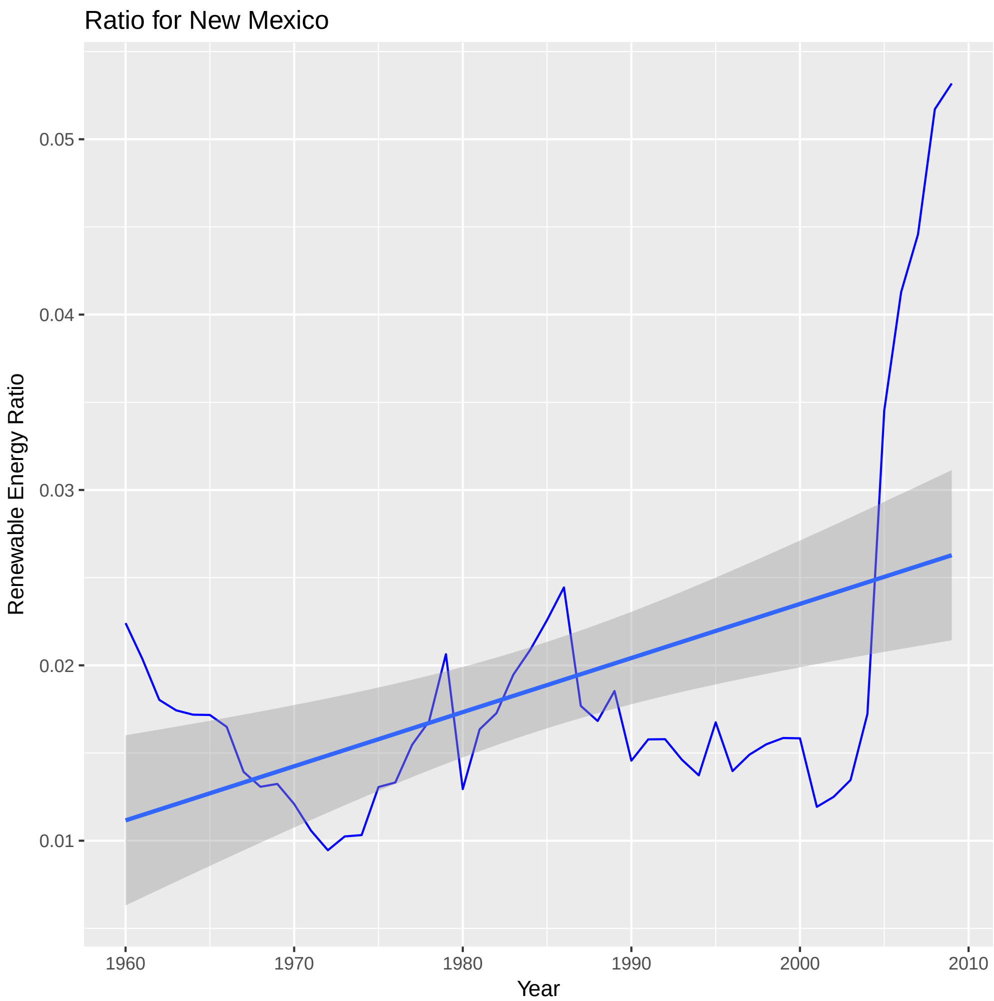


California is the most populated state in the nation, and, with the largest economy, its total energy demand is second only to Texas. California is also rich in energy resources. The state has an abundant supply of crude oil and is a top producer of conventional hydroelectric power. California also leads the nation in electricity generation from solar, geothermal, and biomass resources.

Stretching two-thirds of the way up the West Coast, California is the nation's third-largest state. Transportation dominates California's energy consumption profile. The state also accounts for one-fifth of the nation's jet fuel consumption. The industrial sector is the state's second-largest energy consumer. However, per capita energy use in California's residential sector is lower than that of every other state except Hawaii. In most of California's more densely populated areas, the climate is dry and relatively mild.

The **RER** shows an overall trend of increasing despite some fluctuations from 7.83% to 12.91% in California.

2.4 New Mexico

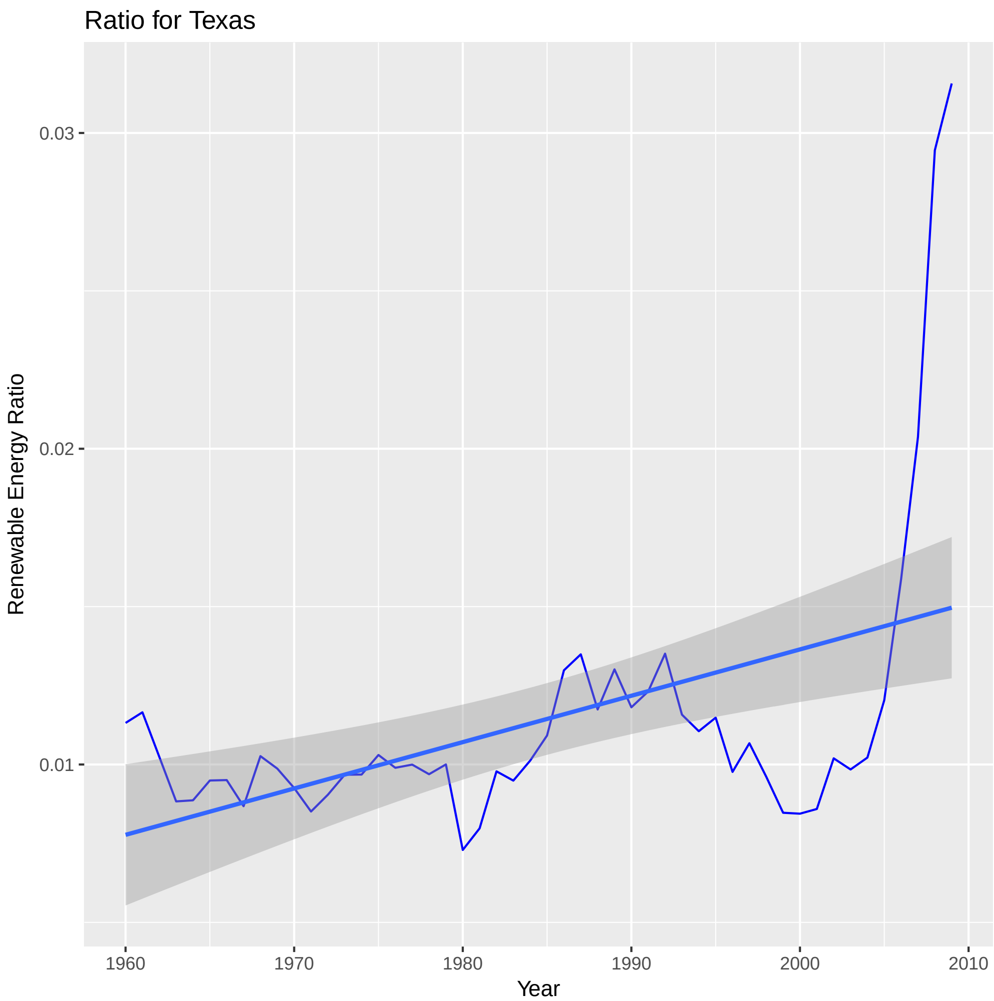


New Mexico is home to the forested peaks and valleys of the southern Rocky Mountains, high plateaus of the Great Plains, and many spectacular desert canyons and mesas. New Mexico also contains a wealth of fossil fuel, mineral, and renewable energy resources. The climate varies widely by location and elevation, from the deserts in the south, where summer temperatures above 100 degrees Fahrenheit are common, to snowy peaks in the north, where winter temperatures have fallen to 50 degrees below zero. Although New Mexico is the fifth-largest state by area, it is the sixth-least densely populated.

New Mexico is the seventh-largest net supplier of energy to the nation, primarily because of its petroleum, natural gas, natural gas liquids, and coal production. Among the state's end-use sectors, the industrial sector is the largest consumer of energy, followed by the transportation sector. Despite the state's climate extremes, energy consumption per capita by the residential sector is among the lowest in the nation.

The **RER** shows an overall trend of increasing despite some fluctuations from 2.24% to 5.32%. The **RER** experienced a rapid growth from 2004 to 2005.

2.5 Texas



Texas leads the nation in energy production, primarily from crude oil and natural gas, providing more than one-fifth of U.S. domestically-produced energy. Second only to Alaska in total land area, Texas stretches about 800 miles at its widest points both east to west and north to south. Crude oil and natural gas fields are present across the entire state, and coal is found in bands that cut across the eastern Texas coastal plain and in other coal-producing areas in the north-central and southwestern parts of the state. Texas also has abundant renewable energy resources and has rapidly developed its wind energy, becoming first in the nation in wind generated electricity. With a significant number of sunny days across vast distances, Texas is among the leading states in solar energy potential as well. Geothermal resources suitable for power generation are present in East Texas. Additionally, uranium deposits are found in South Texas. Overall, Texas is a large state with a wealth of energy resources.

Among the states, Texas has the second-largest population and the second-largest economy after California. The state leads the nation in total energy consumption, accounting for more than one-eighth of the U.S. total. The transportation sector accounts for the second-largest share of energy consumption. The Texas climate varies significantly from east to west. The result is a climate that ranges from humid and subtropical along the coast. Temperatures average in the 90s during the summer in the most densely populated parts of Texas, and energy use for cooling is high. Although the residential sector accounts for just one-eighth of state energy consumption, Texas leads the nation in residential energy use. The **RER** shows an overall trend of increase despite some fluctuations from 1.13% to 3.16%. The **RER** showed a steep growth from 2004 to 2005.

3 Model Construction and Result

3.1 Basic Assumptions

On the basis of our defined energy profile criteria, further assumptions are made in order to construct an accurate model for each state. It is assumed that the renewable energy ratio is only possibly related to fossil fuel price, fossil fuel production, renewable energy production, electricity production from nuclear power, real GDP and total population. The description and unit of these variables are listed in table 1. It is assumed that all variable has fixed effects through time. Since consumption is statistically related to the price of the energy (adjusted R-square > 0.7), consumption variables are not considered in our model. It turns out that our model is still significant without those variables ($p < 10^{-10}$).

Table 1: Variable Description and Unit

Variable	Description
CLTCD	Coal average price, all sectors
NGTCD	Natural gas average price, all sectors (including supplemental gaseous fuels)
PATCD	All petroleum products average price, all sectors
CLPRB	Coal production
NGMPB	Natural gas marketed production
PAPRB	Crude oil production (including lease condensate)
REPRB	Renewable energy production
EMFDB	Biomass inputs (feedstock) for the production of fuel ethanol
GETCB	Geothermal energy total consumption
HYTCB	Hydroelectricity total production
SOTCB	Photovoltaic and solar thermal energy total consumption
WYTCB	Electricity produced from wind energy
WWTCB	Wood and waste total consumption
NUETB	Electricity produced from nuclear power
GDPRX	Real gross domestic product
TPOPP	Resident population including Armed Forces

Due to the small sample size of our data set, data should be carefully chosen so as to honestly reflect the trend. We find the data to be unstable or missing in the early years especially for renewable energy production. For example, Arizona only started its electricity production from wind energy in 2009, leaving the data for previous years to be 0. Moreover, it's explained in the official document that many unrecorded data in the early years is assumed to be zero, which can be different from the reality. Taking these factors into consideration, we only use data starting from 1980 for our forecasting model, and data that has the value of 0 is discarded.

3.2 Multiple Linear Analysis and Step-wise Regression

To characterize the energy profile through 1960-2009, multiple linear analysis is employed. The best-fitted multiple regression equation is derived for each of the four states. The renewable energy ratio is used as dependent variable and explanatory variable is defined to be those variables included in table 3.1. Then step-wise regression method is used to select variables so that we can minimize the AIC (Akaike information criterion) index and derive the regression equation in which every variable is significantly related to our dependent variable,the renewable energy ratio ($p < 0.05$). The multiple linear regression of our data has the following general form of equation

$$Y_{RER} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{1}$$

where Y_{RER} denotes the renewable energy ratio through time. β_0 is the constant intercept different by state. β_1 through β_n are explanatory variables and ϵ is the residual. The variables and corresponding coefficients are listed in table 2. All computation is done by R 3.4.2 .

Table 2: Coefficients and Standard Errors in Multiple Linear Analysis Model

	Arizona	California	New Mexico	Texas
PATCD	0.008164(1.418e-03)***		-6.253e-04(1.670e-04)**	
PAPRB	-0.00004718(6.981e-06)***			
EMFDB	-4.786e-06(2.242e-06)*			
HYTCB	1.053e-06(8.087e-08)***	1.105e-07(8.397e-09)***	1.077e-06(4.520e-07)*	1.048e-07(2.607e-08)***
WYTCB	3.027e-04(5.889e-05)***	1.202e-05(8.116e-07)***	1.495e-06(1.823e-07)***	1.140e-07(7.552e-09)***
TPOPP	-1.071e-04(7.875e-06)***			
NUETB	1.006e-06(4.787e-08)***			
CLTCD	6.619e-03(6.619e-03)*	6.834e-03(1.580e-03)***	4.476e-03(1.430e-03)**	
GDPRX		-1.098e-07(1.094e-08)***		-7.441e-09(1.532e-09)***
NGMPB			-3.597e-09(9.831e-10)**	
WWTCB			2.049e-06(1.540e-07)***	8.481e-08(1.040e-08)***
SOTCB				-2.308e-06(8.471e-07)*

*We denote statistical significance at the following levels: * p<0.05, ** p<0.01, *** p<0.001. Standard errors are shown in parentheses.*

3.3 Grey Method and Other Nonlinear Model for RER Forecasting

Grey prediction model is a time series model that is based on grey system theory. Grey system theory is usually applied to uncertain system where information is incomplete which is likely the condition of forecasting energy usage. Grey model is applied to a broad range of fields like economy, agriculture and biological system. In our model, grey prediction model is used to forecast the price of non-renewable energy, production of energy sources, real GDP and population. GM(1,1) (i.e. First order grey model with one variable) specifically is used in our **RER** forecasting model. The process of model construction is described below. Assume that we have a set of time series data for a certain variable listed in table2, X

$$X = \{X(1), X(2), X(3), \dots, X(n)\} \tag{2}$$

The cumulative sum of the original series is then calculated

$$X' = \{X'(1), X'(2), X'(3), \ldots, X'(n)\} \tag{3}$$

where

$$X'(t) = \sum_{i=1}^{t} X(i) \tag{4}$$

A first-order dynamic system is then constructed,

$$\frac{dX'}{dt} + \eta X' = \alpha \tag{5}$$

where α and η are constants determined by the least square estimate

$$\hat{a} = \begin{pmatrix} \hat{\eta} \\ \hat{\alpha} \end{pmatrix} = \left(\mathbf{B}^T \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{Y} \tag{6}$$

where

$$\mathbf{B} = \begin{bmatrix} -\frac{1}{2}(X'(1) + X'(2)) & 1 \\ -\frac{1}{2}(X'(2) + X'(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(X'(n-1) + X'(n) & 1 \end{bmatrix} \tag{7}$$

$$\mathbf{Y} = \begin{bmatrix} X(2) \\ X(3) \\ \vdots \\ X(n) \end{bmatrix} \tag{8}$$

After α and η are estimated, differential equation(5) is solved with the initial condition $X'(1) = X(1)$ by integrating both sides

$$\frac{1}{\frac{\alpha}{\eta} - X} dx = \eta dt \tag{9}$$

and prediction formula is derived for the variable

$$\hat{X}'(t+1) = \left(X(1) - \frac{\alpha}{\eta}\right) e^{-\eta t} + \frac{\alpha}{\eta} \tag{10}$$

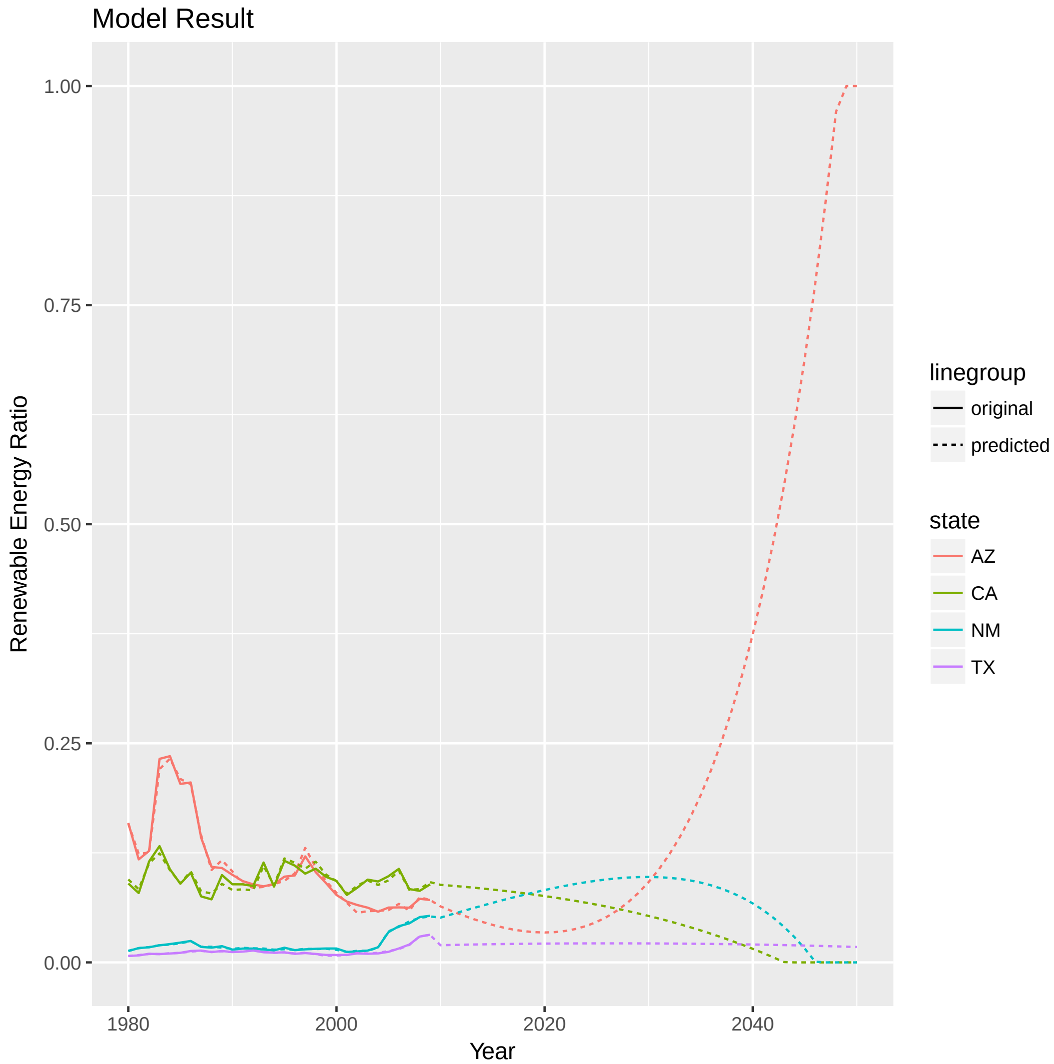
where $\hat{X}'$ denotes the predicted value, and the predicted value for X, $\hat{X}$ can be obtained by calculating the difference between two consecutive value of $\hat{X}'$ Predicted time series data is finally substituted back into equation(1) where we calculated our predicted **RER** data.

It's observed that the production data of certain kind of renewable energy like wind energy increases rapidly. In this situation, grey model yield exponential growth rate. In this case, the prediction yield results that do not align with the real situation. This makes reasonable sense since production using wind energy, for example, has upper limit due to the restriction of geography and climate, etc. As a result, we establish a linear relationship between time and the logarithm of the production data

$$t = aln(x) + b \tag{11}$$

where t denotes time and x denotes the production of a certain kind of renewable energy.

With the combined model, we get the following result:



4 Model Interpretation

4.1 Similarities and Differences Between States

- In Arizona, the **RER** has positive correlations with petroleum price, hydroelectricity consumption, wind energy consumption, and nuclear power production; has a negative correlation with crude oil production.
- In California, the **RER** has a positive correlation with hydroelectricity production; has a negative correlation with Gross Domestic Product.
- In New Mexico, the **RER** has positive correlations with coal price, wind energy consumption and wood & waste energy production.
- In Texas, the **RER** has positive correlations with hydroelectricity consumption, wind energy consumption, and wood & waste energy production; has a negative correlation with Gross Domestic Product.

From the above findings, we can see that in all of the four states, high prices of traditional energy (as apposed to cleaner, renewable energy) and cleaner, renewable energy consumption would cause the **RER** to increase. The former is easy to understand, as people generally prefer energy with low prices. The latter can be seen as a part of our conclusion: if we promote the use of cleaner, renewable energy by using more of each kind of them, we would of course increase the total cleaner, renewable energy consumption, thus achieving the increase of the **RER**.

In some states such as California and Texas, high GDP would cause the **RER** to decrease. This means that the development of economy is faster than the popularization of cleaner, renewable energy use.

Different states have different components of cleaner, renewable energy.

In **Arizona**, fossil fuels consist of a large part of total energy consumption. Nuclear electric power is a major source of renewable energy.

In **California**, natural gas is the most widely used energy source, and non-hydroelectric renewable technologies contribute to about 1/4 of the state's net generation.

As for **New Mexico** and **Texas**, fossil fuel is still a major source of energy consumption. Biomass and wind energy are the main contributions of renewable energy.

Therefore, to increase the **RER**, they should adopt customized policies to best utilize their cleaner, renewable energy resources.

4.2 Compact Goal Statements

In 2025, the ratio of renewable energy of the four states should reach to the same level.

In 2050, the ratio of renewable energy of the four states should keep rising with same speed in order to reach the compact goal.

4.3 Policy Implications

a According to the energy profile of Arizona, nuclear energy consists of a large part of total energy that generates electricity for the state. Arizona has the nation's largest nuclear power plant (Palo Verde Nuclear Generating Station). As a result, Arizona should increase safety measures that can be applied when accidents happen to the nuclear power plants. For example, portable equipment should be available that can respond flexibly if the safety equipment is damaged. It is also important to improve the robustness of the reactor containment vessel to prevent leaks of radioactive material. Moreover, more transmission lines can be constructed between Arizona and adjacent states such as New Mexico and California. In this way, the ratio of traditional energy out of total energy consumed to generate electricity is reduced in New Mexico and California. Although the ratio to generate electricity is high in Texas, the high cost of transmitting electricity through transmission lines has to be considered.

b According to the positivity of coefficient of multiple linear analysis model, the price of fossil fuels is positively related to the **RER**. Therefore, the tax of fossil fuels should be raised by the government in order to encourage the usage of renewable energy.

c More wind farms should be built in Arizona in order to increase the use of wind energy since the coefficient of electricity produced from wind energy is much higher than other energy in the multiple linear analysis model. Hydroelectricity should also be encouraged according to the coefficient of hydroelectricity total production. As for California, the government should increase hydroelectricity use because the hydroelectricity production is positively related to the **RER** according to the model. In New Mexico, the government should grow wind energy and wood and waste energy use in generating electricity according to the coefficient in the multiple linear analysis model. Also, average coal price should be raised because the coefficient of average coal price in all sectors is negatively related to the **RER**. In Texas, more facilities should be built to utilize wind energy, wood and waste energy, and hydroelectricity energy.

5 Conclusions

In this paper, an important index named Renewable Energy Ratio(**RER**) is defined and justified. This index is used to describe the renewable energy profile of four states in United States. Our combined model of multiple linear analysis and grey method qualitatively characterizes the relationship between **RER** and a unique set of variables including energy price, energy production, GDP and population for each state. Several important similarities and differences between states are identified and the change of this index after 2009 is predicted. Based on the result of our model, policy suggestions and compact goal statement are proposed. The result shows that **RER** of all four states either approaches 0 or 1, which might indicates that apart from nuclear energy, the other renewable energy may not be able to satisfy the increasing energy demand. According to our prediction, only state that heavily relies on nuclear energy (Arizona) will be able to replace fossil fuel by renewable energy completely.

6 Discussions

We utilize the given data file as much as we can, and come to our conclusion. We realize that the data file itself has shortcomings. For example, some data begin from 1960, and some others begin from 1970. Also, we admit that our analysis and prediction may be different from the reality in the years to come. Policies regarding cleaner, renewable energy use change and can have great influences over future planning and development. Thus, as stated in the problem prompt, our results are only true in the absence of any policy changes by each governor's office. Also, new energy may be developed, which may also influence our analysis and predictions.

7 References

[1] State Energy Data System (SEDS) Complete Dataset through 2009 (All 50 states), U.S. Energy Information Administration, https://catalog.data.gov/dataset/state-energy-data-system-seds-complete-dataset-through-2009#sec-dates

[2] Technical notes & documentation - complete 2015, U.S. Energy Information Administration, https://www.eia.gov/state/seds/seds-technical-notes-complete.php?sid=US#Consumption

[3] Appendix A. Mnemonic Series Names (MSN), U.S. Energy Information Administration, https://www.eia.gov/state/seds/sep_use/notes/use_a.pdf

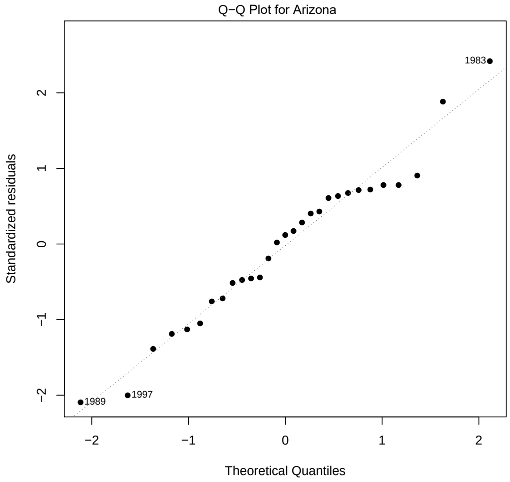
[4] Efforts to Improve Safety of Nuclear Power Plants, Masayoshi Matsuura, et al., http://large.stanford.edu/courses/2017/ph241/kim-d1/docs/matsuura.pdf

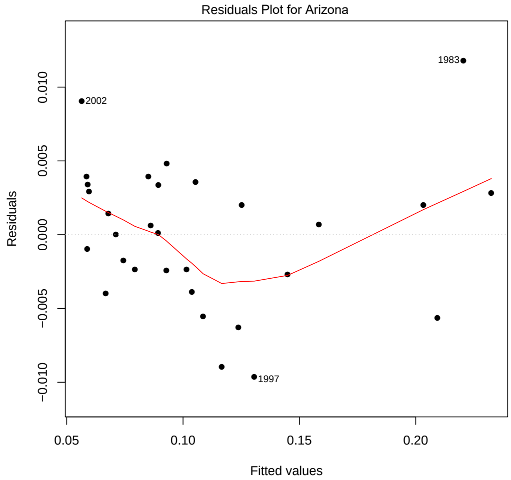
[5] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

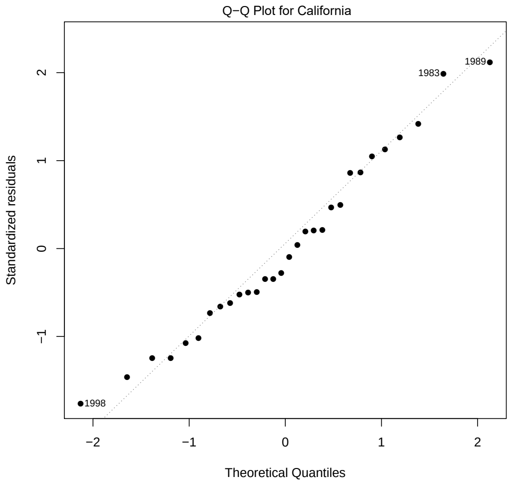
Appendices

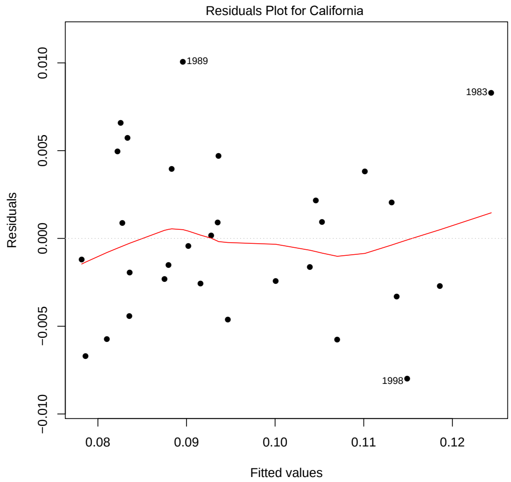
Appendix A Model Validation

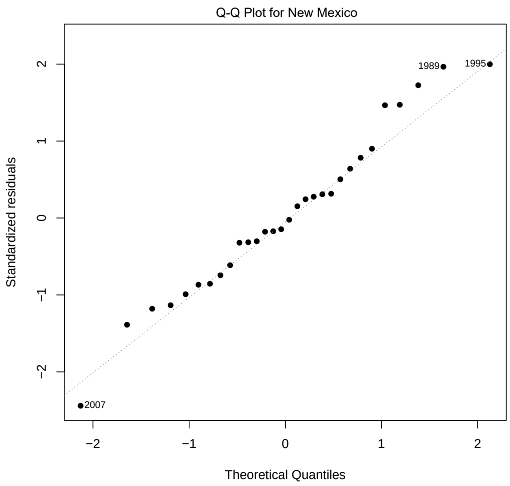
Four assumptions including linearity of residuals, independence of residuals, normal distribution of residuals are made when we use multiple linear analysis. We use QQ plot and residual plot to examine whether these conditions are met.

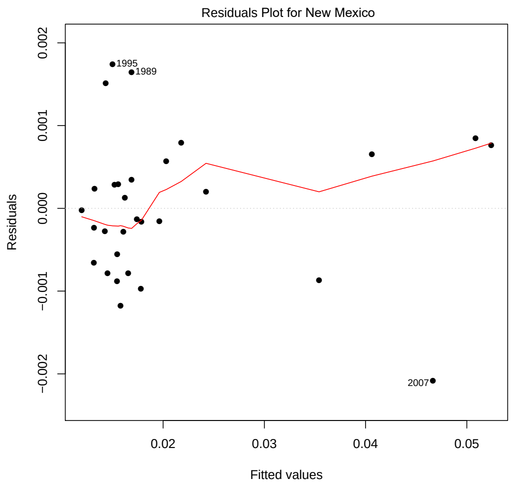

Q−Q Plot for Arizona
Standardized residuals
Theoretical Quantiles
1983
1989
1997

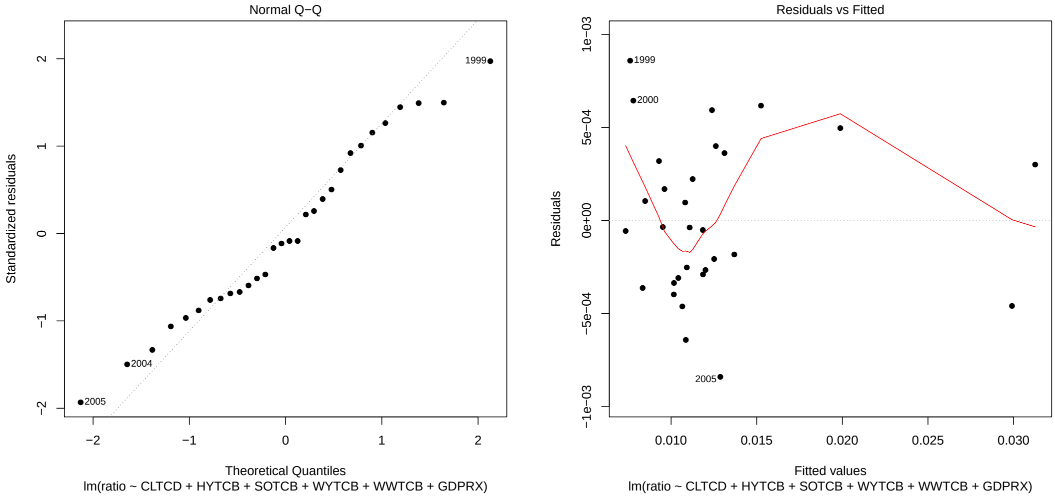

Residuals Plot for Arizona
Residuals
Fitted values
1983
2002
1997


Q−Q Plot for California
Standardized residuals
Theoretical Quantiles
1983
1989
1998


Residuals Plot for California
Residuals
Fitted values
1989
1983
1998


Q-Q Plot for New Mexico
Standardized residuals
Theoretical Quantiles
1989
1995
2007


Residuals Plot for New Mexico
Residuals
Fitted values
1995
1989
2007



Appendix B Program Code

Here is a part of a statistical calculation program we used in our model.

```
cpi = read.csv("CPI.txt", header = F)
rownames(cpi) = 1960:2009
baseyrindex = cpi["2005",]
alldata = list(AZ = azdata, CA = cadata, NM = nmdata, TX = txdata)
ratiodata = list(AZ = azratio, CA = caratio, NM = nmratio, TX = txratio)
adjprice = function(p){
    (p/baseyrindex)*cpi[21:50,]
}
tmp = sapply(states,function(sn){           #sn = state names
    modeldata = sapply(1980:2009, function(x){
        alldata[[sn]][alldata[[sn]][,"Year"]==x,"Data"]
    })
    modeldata = t(modeldata)
    rownames(modeldata) = 1980:2009
    colnames(modeldata) = levels(as.factor(alldata[[sn]][,"MSN"]))
    modeldata = data.frame(modeldata)
    prices = c("CLTCD","NGTCD","PATCD")
    tmp = sapply(prices,function(x){
        modeldata[[x]] <<- adjprice(modeldata[[x]])
    })
    modeldata = data.frame(modeldata,ratio = ratiodata[[sn]][21:50])
    lm.sol = lm(ratio ~ CLTCD + NGTCD + PATCD + CLPRB + NGMPB + PAPRB
+ EMFDB + GETCB + HYTCB + SOTCB + WYTCB + WWTCB + GDPRX + TPOPP + NUETB, data = modeldata)
    print(summary(lm.sol))
    lm.step = step(lm.sol)
})

grey_model = function(ts){  #tm = time series data
    tsp = ts[ts != 0] #time series positive
    if(length(tsp) == 0){
        result = rep(0,times = 41)
        return(list(result = result, C = -1, q = -1, P = -1))
```

```
    }
    if(length(tsp)==1){
        result = rep(tsp, times = 41)
        return(list(result = result, C = -1, q = -1, P = -1))
    }
    if(length(tsp) < 10){
        data = cbind(1:length(tsp), tsp)
        colnames(data) = c("x","y")
        data = data.frame(data)
        m = lm(y~x,data = data)
        result = sapply((length(tsp) + 1):(length(tsp) + 41),function(x){
            sum(c(1,x)*coef(m))
        })
        return(list(result = result, C = -1, q = -1, P = -1))
    }else{
        cs = cumsum(tsp)
        tmp = -0.5*(cs[1:(length(cs)-1)]+cs[2:length(cs)])
        B = cbind(tmp,rep(1, times = length(tmp)))
        Y = as.matrix(tsp[2:length(tsp)])
        coefgrey = solve(t(B) %*% B) %*% t(B) %*% Y      #least square estimate sequence
        a = coefgrey[1]
        u = coefgrey[2]
        cumresult = sapply(0:(length(cs) + 41),function(t){
            (tsp[1] - u/a) * exp(-a*t) + u/a
        })
        result1 = c(cumresult[1],diff(cumresult))[(length(cs) + 1):(length(cs) + 41)]
        result2 = c(cumresult[1],diff(cumresult))[1:(length(cs) + 41)]
        residual = tsp - result2[1:length(tsp)]
        qbar = mean(residual/tsp)  #relative error
        rebar = mean(residual)
        resd = sd(residual)
        tspsd = sd(tsp)
        C = resd/tspsd
        P = sum(abs(residual - rebar) < 0.6745 * tspsd)/length(residual)
        return(list(result = result1, allresult = result2, C = C, q = qbar, P = P))

    }

}
```