

## Problem: How Much Gas Should I Buy This Week?

Gas prices fluctuate significantly from week to week. Consumers would like to know whether to fill up the tank (gas price is likely to go up in the coming week) or buy a half tank (gas price is likely to go down in the coming week).

Consider the following cases:

- Consumer drives 100 miles per week
- Consumer drives 200 miles per week

Assume:

- gas tank holds 16 gallons and average mileage is 25 miles/gallon  $\Rightarrow$  400 miles/tank
- consumer buys gas once a week

Therefore, the consumer can drive for 2 weeks or 1 week on half a tank of gas for cases (1) and (2) respectively. Thus the choice each week is whether to buy a full or half tank of gas or no gas.

Use the weekly gas price data available by region, state and city at:

[http://www.eia.gov/dnav/pet/pet\\_pri\\_gnd\\_a\\_epmr\\_pte\\_dpgal\\_w.htm](http://www.eia.gov/dnav/pet/pet_pri_gnd_a_epmr_pte_dpgal_w.htm)

You can also use weekly crude price data [http://www.eia.gov/dnav/pet/pet\\_pri\\_spt\\_s1\\_w.htm](http://www.eia.gov/dnav/pet/pet_pri_spt_s1_w.htm) and any other publicly available data such as weather data, economic data, world events, etc.

- Develop a model that a consumer could use each week to determine how much gas – full tank or half tank – to purchase.
- Use the 2011 data to build/train your model and the 2012 data to test and validate your model.
- Is there an upper bound on “mileage driven” that changes the decision for buying weekly gasoline?
- Develop models for the following at least one large US city such as: Boston, Chicago, Denver, Dallas, Los Angeles, Houston, New York, or San Francisco.

In addition to the one page summary and complete project report prepare a short one page non-technical letter to your local paper describing how the average person might use your model.

## Paper of Problem B: How Much Gas Should I Buy This Week?

### Abstract

Gasoline is the bleed that surges incessantly within the muscular ground of city; gasoline is the feast that lures the appetite of drivers. “To fill or not fill?” That is the question flustering thousands of car owners. This paper will guide you to predict the gasoline prices of the coming week with the currently available data with respect to swift changes of oil prices. Do you hold any interest in what pattern of filling up the gas tank can lead to a lower cost in total?

By applying the Time series analysis method, this paper infers the price in the imminent week. Furthermore, we innovatively utilize the average prices of the continuous two weeks to predict the next two week’s average price; similarly, employ the four-week-long average prices to forecast the average price of four weeks later. By adopting the data obtained from 2011 and the comparison in different aspects, we can obtain the gas price prediction model :

$$\hat{G}_{t+1} = 0.0398 + 1.6002g_t - 0.7842g_{t-1} + 0.1207g_{t-2} + 0.4147g_t - 0.5107g_{t-1} + 0.1703g_{t-2} + \hat{\varepsilon}.$$

This predicted result of 2012 according to this model is fairly ideal. Based on the prediction model, We also establish the model for how to fill gasoline. With these models, we had calculated the lowest cost of filling up in 2012 when traveling 100 miles a week is 637.24 dollars with the help of MATLAB, while the lowest cost when traveling 200 miles a week is 1283.5 dollars. These two values are very close to the ideal value of cost on the basis of the historical figure, which are 635.24 dollars and 1253.5 dollars respectively. Also, we have come up with the scheme of gas fulfillment respectively. By analyzing the schemes of gas filling, we can discover that when you predict the future gasoline price going up, the best strategy is to fill the tank as soon as possible, in order to lower the gas fare. On the contrary, when the predicted price tends to decrease, it is wiser and more economic for people to postpone the filling, which encourages people to purchase a half tank of gasoline only if the tank is almost empty.

For other different pattern for every week’s “mileage driven”, we calculate the changing point of strategies-changed is 133.33 miles.

Eventually, we will apply the models -to the analysis of the New York City. The result of prediction is good enough to match the actual data approximately. However, the total gas cost of New York is a little higher than that of the average cost nationally, which might be related to the higher consumer price index in the city. Due to the limit of time, we are not able to investigate further the particular factors.

**Keywords:** gasoline price Time series analysis forecast lowest cost MATLAB



## Catalogue

Abstract -----	1
Restatement -----	2
1. Assumption-----	4
2. Definitions of Variables and Models-----	4
2.1 Models for the prediction of gasoline price in the subsequent week-----	4
2.2 The Model of oil price next two weeks and four weeks-----	5
2.3 Model for refuel decision-----	5
2.3.1 Decision Model for consumer who drives 100 miles per week-----	6
2.3.2 Decision Model for consumer who drives 200 miles per week-----	7
3. Train and Test Model by 2011 dataset-----	8
3.1 Determine the all the parameters in Equation ② from the 2011 dataset-----	8
3.2 Test the Forecast Model of gasoline price by the dataset of gasoline price in 2012-----	10
3.3 Calculating $\varepsilon$ -----	12
3.4 Test Decision Models of buying gasoline by dataset of 2012-----	14
3.4.1 100 miles per week-----	14
3.4.2 200 miles per week-----	14
3.4.3 Second Test for the Decision of buying gasoline-----	15
4. The upper bound will change the Decision of buying gasoline-----	15
5. An analysis of New York City-----	16
5.1 The main factor that will affect the gasoline price in New York City-----	16
5.2 Test Models with New York data-----	18
5.3 The analysis of result-----	19
6. Summery& Advantage and disadvantage-----	19
7. Report-----	20
8. Appendix-----	21
Appendix 1(main MATLAB programs) -----	21
Appendix 2(outcome and graph) -----	34

## Restatement

The world market is fluctuating swiftly now. As the most important limited energy, oil is much accounted of cars owners and dealer. We are required to make a gas-buying plan which relates to the price of gasoline, the volume of tank, the distance that consumer drives per week, the data from EIA and the influence of other events in order to help drivers to save money.

We should use the data of 2011 to build up two models that discuss two situations: 100miles/week or 200miles/week and use the data of 2012 to test the models to prove the model is applicable. In the model, consumer only has three choices to purchase gas each week, including no gas, half a tank and full tank. At the end, we should not only build two models but also write a simple but educational report that can attract consumer to follow this model.

### 1. Assumption

- Assume the consumer always buy gasoline according to the rule of minimum cost.
- Ignore the difference of the gasoline weight.
- Ignore the oil wear on the way to gas stations.
- Assume the tank is empty at the beginning of the following models.
- Apply the past data of crude oil price to predict the future price of gasoline.(The crude oil price can affect the gasoline price and we ignore the hysteresis effect on prices of crude oil towards prices of gasoline.)

### 2. Definitions of Variables and Models

$t$  stands for the sequence number of week in any time.( $t$  stands for the current week. ( $t-1$ ) stands for the last week. ( $t+1$ ) stands for the next week.

$c_t$ : Price of crude oil of the current week.

$g_t$ : Price of gasoline of the  $t^{\text{th}}$  week.

$P_t$ : The volume of oil of the  $t^{\text{th}}$  week.

$\hat{G}_{t+1}$ : Predicted price of gasoline of the  $(t+1)^{\text{th}}$  week.

$\alpha, \beta$ : The coefficient of the  $g_t$  and  $c_t$  in the model.

$d$ : The variable of decision of buying gasoline.( $d=1/2$  stands for buying a half tank gasoline)

#### 2.1 Model for the prediction of gasoline price in the subsequent week

Whether to buy half a tank oil or full tank oil depends on the short-term forecast about the gasoline prices. Time series analysis is a frequently-used method to expect the gasoline price trend. It can be expressed as:

$$\hat{G}_{t+1} = \alpha_1 g_t + \alpha_2 g_{t-1} + \alpha_3 g_{t-2} + \alpha_4 g_{t-3} + \dots + \alpha_{n+1} g_{t-n} + \hat{\varepsilon} \quad \text{---Equation ①}$$

$\hat{\varepsilon}$  is a parameter that reflects the influence towards the trend of gasoline price in relation to several aspects such as weather data, economic data, world events and so on.

Due to the prices of crude oil can influence the future prices of gasoline; we will adopt the past prices of crude oil into the model for gasoline price forecast.

$$\hat{G}_{t+1} = (\alpha_1 g_t + \alpha_2 g_{t-1} + \alpha_3 g_{t-2} + \alpha_4 g_{t-3} + \dots + \alpha_{n+1} g_{t-n}) + (\beta_1 g_t + \beta_2 g_{t-1} + \beta_3 g_{t-2} + \beta_4 g_{t-3} + \dots + \beta_{n+1} g_{t-n}) + \hat{\varepsilon} \quad \text{---Equation ②}$$

We will use the 2011 data set to calculate the all coefficients and the best delay periods  $n$ .

## 2.2 The Model of oil price next two weeks and four weeks

We mainly depend on the prediction of change of gasoline price in order to make decision that the consumer should buy half a tank or full tank gas. When consumer drives 100miles/week, he can drive whether 400miles most if he buys full tank gas or 200miles most if he buys half a tank gas. When consumer drives 200miles/week, full tank gas can be used two weeks most or half a tank can be used one week most. Thus, we should consider the gasoline price trend in four weeks in future.

Equation ② can also be rewritten as

$$\hat{G}_{t+1} = (\alpha_1 g_t + \beta_1 g_t) + (\alpha_2 g_{t-1} + \beta_2 g_{t-1}) + (\alpha_3 g_{t-2} + \beta_3 g_{t-2}) + \dots + (\alpha_{n+1} g_{t-n} + \beta_{n+1} g_{t-n}) + \hat{\varepsilon} \quad \text{---Equation ③}$$

If we define  $y_t = \alpha_1 g_t + \beta_1 g_t$ ,  $y_{t-1} = \alpha_2 g_{t-1} + \beta_2 g_{t-1}$ ,  $y_{t-2} = \alpha_3 g_{t-2} + \beta_3 g_{t-2} \dots$ , and so on.

Equation ③ can change to

$$\hat{G}_{t+1} = y_t + y_{t-1} + y_{t-2} + \dots + y_{t-n} + \hat{\varepsilon} \quad \text{---Equation ④}$$

We use  $y_{(t-1,t)}$  denote the average price from week  $(t-1)$  to week  $(t)$ , which is

$$y_{(t-1,t)} = \frac{y_{t-1} + y_t}{2}$$

Accordingly, the average price from week  $(t-3)$  to week  $(t)$  is

$$y_{(t-3,t)} = \frac{y_{t-3} + y_{t-2} + y_{t-1} + y_t}{4}$$

Apply Time series analysis, we can get the average price from week  $(t+1)$  to week  $(t+2)$  by Equation ④,

$$\hat{G}_{(t+1,t+2)} = y_{(t-1,t)} + y_{(t-3,t-2)} + y_{(t-5,t-4)}, \quad \text{---Equation ⑤}$$

As well, the average price from week  $(t+1)$  to week  $(t+4)$  is

$$\hat{G}_{(t+1,t+4)} = y_{(t-3,t)} + y_{(t-7,t-4)} + y_{(t-11,t-8)}. \quad \text{---Equation ⑥}$$

### 2.3 Model for refuel decision

By comparing the present gasoline price with the future price, we can decide whether to fill half or full tank.

The process for decision can be shown through the following flow chart.

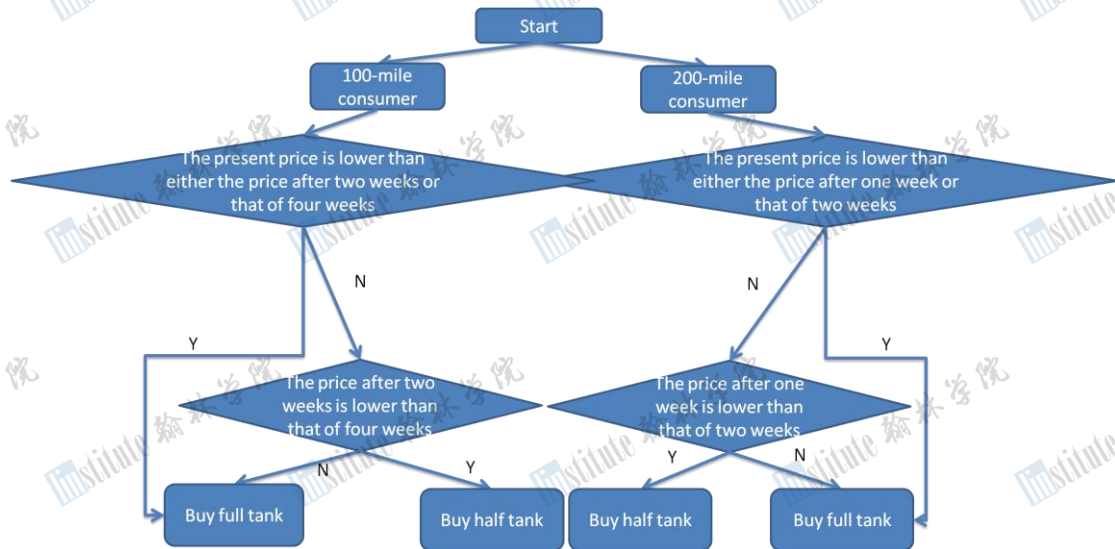


Chart 1

For the consumer, the best decision is to get gasoline with the lowest prices. Because a tank of gasoline can run 2 or 4 week, so we should choose a time point that the price is lowest by comparison of the gas prices at present, 2 weeks and 4 weeks later separately. The refuel decision also depends on how many free spaces in the tank because we can only choose half or full tank each time. If the free spaces are less than  $1/2$ , we can refuel nothing even if we think the price is the lowest at that time.

#### 2.3.1 Decision Model for consumer who drives 100 miles per week.

We assume the oil tank is empty at the beginning time  $(t=0)$ . There are four cases for a consumer to choose a best refuel time when the tank is empty.

- i.  $g_t > \hat{G}_{t+4}$  and  $g_t > \hat{G}_{t+2}$ , which means the present gasoline price is higher than that either two weeks or four weeks later. It is economic to fill half tank under such condition.



- ii.  $g_t < \hat{G}_{t+4}$  and  $g_t < \hat{G}_{t+2}$ , which means the present gasoline price is lower than that either two weeks or four weeks later. It is economic to fill full tank under such condition.
- iii.  $\hat{G}_{t+4} > g_t > \hat{G}_{t+2}$ , which means the present gasoline price is higher than that two weeks later but lower than that four weeks later. It is economic to fill half tank under such condition.
- iv.  $\hat{G}_{t+4} < g_t < \hat{G}_{t+2}$ , which means the present gasoline price is higher than that four weeks later but lower than that two weeks later. It is economic to fill full tank under such condition.

If other time, we should consider both the gasoline price and the oil volume in the tank to pick up a best refuel time. In summary, the decision model for running 100 miles a week is

$$d_i = \begin{cases} 0 & \frac{1}{2} < \sum_1^{t-i} d_i - \frac{1}{4}(t-1) < 1 \text{ or } (\frac{1}{4} < \sum_1^{t-i} d_i - \frac{1}{4}(t-1) < \frac{1}{2} \& g_t > \max(\hat{G}_{t+2}, \hat{G}_{t+4})) \\ 1/2 & (\frac{1}{4} < \sum_1^{t-i} d_i - \frac{1}{4}(t-1) < \frac{1}{2} \& (g_t < \hat{G}_{t+2} \& \hat{G}_{t+4})) \text{ or } (0 < \sum_1^{t-i} d_i - \frac{1}{4}(t-1) < \frac{1}{4} \& \hat{G}_{t+2} < g_t < \hat{G}_{t+4}) \\ 1 & \sum_1^{t-i} d_i - \frac{1}{4}(t-1) = 0 \& ((g_t > \hat{G}_{t+2} \& \hat{G}_{t+4}) \text{ or } \hat{G}_{t+4} < g_t < \hat{G}_{t+2}) \end{cases}$$

----Equation ⑦

$d_i$  is the decision variable,  $d_i = 1$  means we fill full tank,  $d_i = 1/2$  means we fill half tank.  $\sum_1^{t-i} d_i - \frac{1}{4}(t-1)$  represents the residual gasoline volume in the tank.

The method of prices comparison was analyzed in the beginning part of 2.3.1.

### 2.3.2 Decision Model for consumer who drives 200 miles per week.

Because even full tank can run only two weeks, the consumer must refuel during every two weeks. There are two cases to decide whether to buy half or full tank when the tank is empty. This situation is much simpler than that of 100 miles a week. The process for decision can also be shown through the following flow chart.

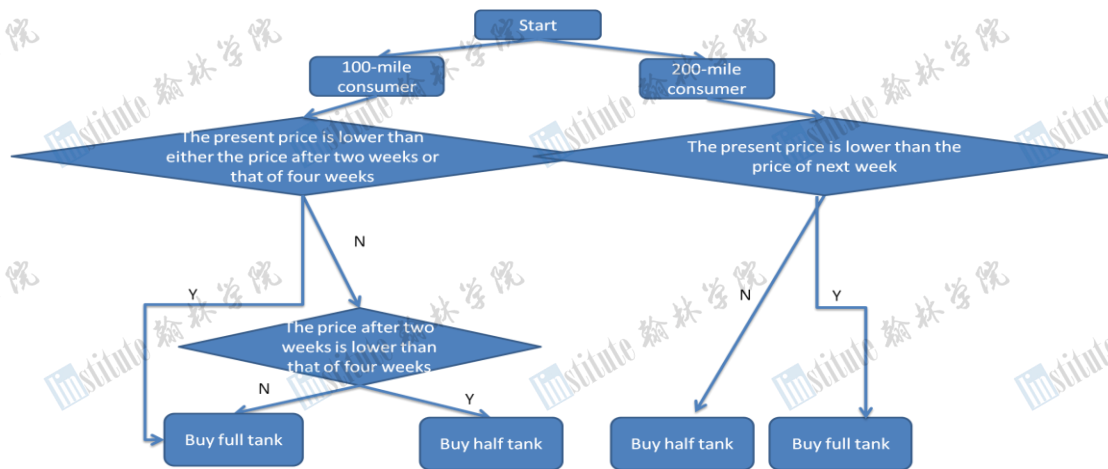


Chart 2

The two cases for deciding buy half or full tank are:

- i.  $g_t > \hat{G}_{t+1}$ , which means the present gasoline price is higher than the next

week. We will buy half tank because we can buy the cheaper gasoline in the next week.

- ii.  $g_t < \hat{G}_{t+1}$ , which means the present gasoline price is lower than the next week. To buy full tank is economic under such situation.

But we should consider both gasoline prices and free tank volume to decide our refueling plan. The Model is

$$d_t = \begin{cases} 0 & \frac{1}{2} < \sum_1^{t-i} d_t - \frac{1}{2}(t-1) < 1 \\ 1/2 & 0 < \sum_1^{t-i} d_t - \frac{1}{2}(t-1) < \frac{1}{2} \text{ or } \sum_1^{t-i} d_t - \frac{1}{4}(t-1) = 0 \text{ \& } g_t > \hat{G}_{t+1} \\ 1 & \sum_1^{t-i} d_t - \frac{1}{2}(t-1) = 0 \text{ \& } g_t < \hat{G}_{t+1} \end{cases}$$

---Equation ⑧

### 3. Train and Test Model by the 2011 dataset

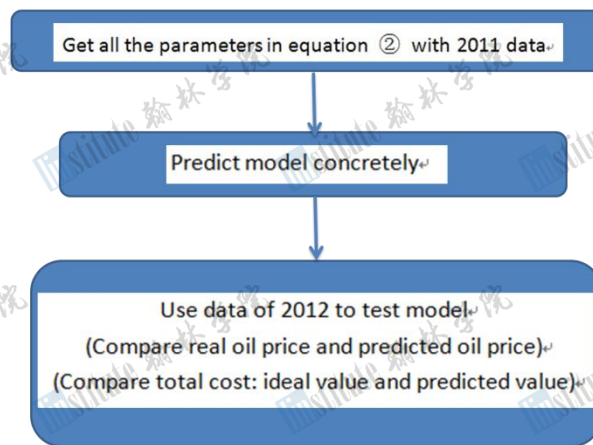
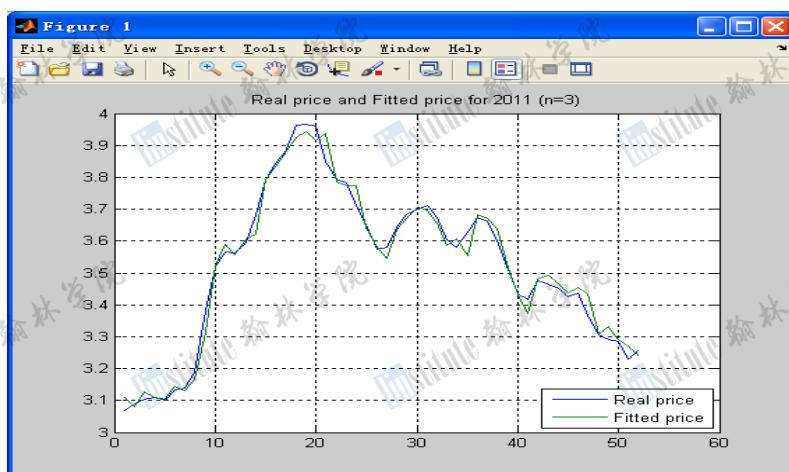


Chart 3

#### 3.1 Determine all the parameters in Equation ② from the 2011 dataset.

Using the weekly gas data from the website <http://www.eia.gov> and the weekly crude price data from <http://www.eia.gov>, we can determine the best delay periods  $n$  and calculate all the parameters in Equation ②. For there are two crude oil price dataset (Weekly Cushing OK WTI Spot Price FOB and Weekly Europe Brent Spot Price FOB), we use the average value as the crude oil price without loss of generality. We tried  $n=3, 4$  and  $5$  respectively with 2011 dataset and received comparison graph of predicted value and actual value, including corresponding coefficient.

(A)  $n=3$ (the hysteretic period is 3)



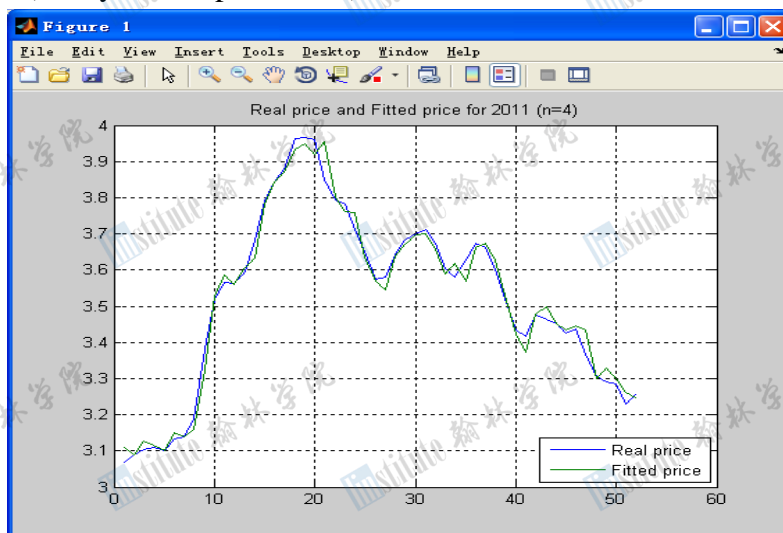
Graph 1 The fitted price and real price of gasoline in 2011(n=3)

Constant	$\alpha_3$	$\alpha_2$	$\alpha_1$	$\beta_3$	$\beta_2$	$\beta_1$
0.0398	0.1207	-0.7842	1.6002	0.1713	-0.5107	0.4147

Table 1

We find that the nearby effect coefficient of the price of crude oil and gasoline. This result is same as our anticipation.

(B)  $n=4$ (the hysteretic period is 4)

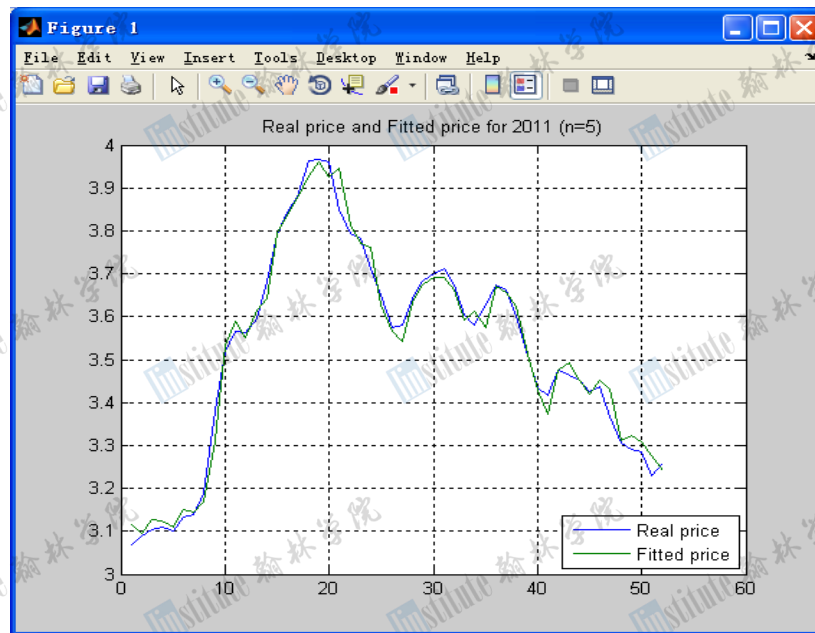


Graph 2 The fitted price and real price of gasoline in 2011 (n=4)

Constant	$\alpha_4$	$\alpha_3$	$\alpha_2$	$\alpha_1$	$\beta_4$	$\beta_3$	$\beta_2$	$\beta_1$
0.0537	0.0280	-0.0925	-0.5570	1.5358	0.1646	0.0048	-0.4766	0.4093

Table 2

(C)  $n=5$ (the hysteretic period is 5)



Graph 3 The fitted price and real price of gasoline in 2011 (n=5)

constant	$\alpha_5$	$\alpha_4$	$\alpha_3$	$\alpha_2$	$\alpha_1$	$\beta_5$	$\beta_4$	$\beta_3$	$\beta_2$	$\beta_1$
0.0849	-0.1209	0.1782	-0.1431	-0.5065	1.5005	0.0499	0.0976	0.0076	-0.4597	0.4027

Table 3

Via comparing the three figures above, we can easily found that the predictive validity of  $n=3$ (the hysteretic period is 3) is slightly better than that of  $n=4$ (the hysteretic period is 4) and  $n=5$ (the hysteretic period is 5) so we choose the model of  $n=3$  to be the prediction model of gasoline price.

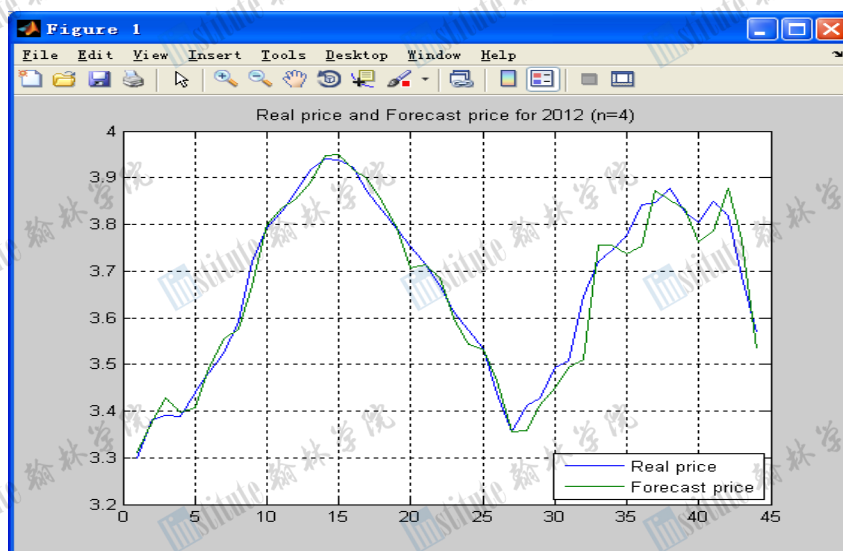
$$\hat{G}_{t+1} = 0.0398 + 1.6002g_t - 0.7842g_{t-1} + 0.1207g_{t-2} + 0.4147g_t - 0.5107g_{t-1} + 0.1703g_{t-2} + \hat{\varepsilon}$$

----Equation ⑨

### 3.2 Test the Forecast Model of gasoline price by the dataset of gasoline price in 2012

Next, we apply models in terms of different hysteretic periods( $n=3,4,5$  respectively), which are shown in Equation ②, to forecast the gasoline price which can be acquired currently in 2012 and get the graph of the forecast price and real price of gasoline:



Graph 4 The real price and forecast price in 2012 ( $n=3$ )Graph 5 The real price and forecast price in 2012 ( $n=4$ )Graph 6 The real price and forecast price in 2012 ( $n=5$ )

Conserving the error of observation, predictive validity is best when  $n$  is 3, but the differences are not obvious when  $n=4$  and  $n=5$ . However, a serious problem should be drawn to concerns: consumers determines how to fill the tank by using the trend of oil price. If the trend prediction is wrong (like predicting oil price will rise when it actually falls), consumers will lose. We use MATLAB software to calculate the amount of error time when we use the model of Equation ⑨ to predict the price of gasoline in 2012. The graph below shows the result.

The value of $n$	$n=3$	$n=4$	$n=5$
The amount of error time	5	7	8

Table 4

It's not difficult to find the prediction effect is the best when  $n$  is 3. Therefore, we determined to use Equation ⑨ as the prediction model of oil price in 2012.

$$\hat{G}_{t+1} = 0.0398 + 1.6002g_t - 0.7842g_{t-1} + 0.1207g_{t-2} + 0.4147g_t - 0.5107g_{t-1} + 0.1703g_{t-2} + \hat{\varepsilon}$$

### 3.3 Calculating $\hat{\varepsilon}$

Since political occurrences, economic events and climatic changes can affect gasoline price, it is undeniable that a  $\hat{\varepsilon}$  exists between predicted prices and real prices. We can use Equation ② to predict gasoline prices in 2011 and then compare them with real data. Through the difference between predicted data and real data, we can estimate the value of  $\hat{\varepsilon}$ . The estimating process can be shown through the following flow chart

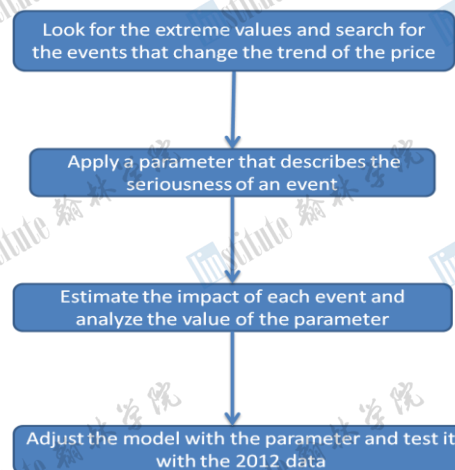


Chart 4

We divide the international events into three types: extra serious event, major event and ordinary event according to the criteria of influence on gas prices. Then we evaluate the value: extra serious event is 3a, major event is 2a, and ordinary event is a. With inference to the comparison of the forecast price and real price in 2011, we find that large deviation of data exists at three time points: May 16, 2011, Aug 08, 2011 and

Oct 10, 2011. After searching, we find that some important international events happened nearly at the three time points. We believe that these events which occurred by chance affect the international prices of gasoline so the predicted prices deviate from the actual prices. The table of events and the calculation of the value of  $a$  are below:

Events in 2011 and The Corresponding Coefficient

Date	Reality	Prediction	Error (Magnification is 1000)	Coefficient	Event
May 16, 2011	3.96	3.909322	50.7	2a	Bin Laden was dead.
Aug 08, 2011	3.674	3.61854	55.45994	2a	Financial aid for Greek
Oct 10, 2011	3.417	3.388994	28.00626	a	Hurricane

Table 4

Therefore, by generalizing several sets of particular data and events, we can estimate the value of  $a$ :

$$a=26.84$$

----Equation ⑩

The calculating process is shown as the following graph.

Since now we have obtained the approximate value of  $a$ , we can evaluate the future prices according to currently known gasoline prices and crude oil prices. To improve our model, we can look for factors resulting in some major turning point in the graph of gasoline prices. On the ground that the most influential factors on prices in 2012 are respectively graded, the difference between fact and prediction can be calculated.

### 3.4 Test Decision Models of buying gasoline by the dataset of 2012

First, we use Equation ⑨ to calculate the gasoline price of next week and use Equation ⑤ and Equation ⑥ to calculate the gasoline price trend of next two to four weeks. On the basis above, we calculate the total cost, and thus receive schemes of buying gasoline of 100miles per week according to Equation ⑦ and Equation ⑧. Using the same method, we can easily obtain the pattern when driving 200 miles per week. The result is presented below.

We collect the important events which will affect the gasoline price in 2012 as well. Therefore, we calculate and adjust the predicted price of gasoline by Equation ⑩. We calculate the scheme of buying gasoline again. The result is below:

#### 3.4.1 100 miles per week

$T_{2012} = 637.2400$  (If the consumer drives 100 miles per week, the total cost in 2012 is 637.24 USD). The scheme calculated by software is below:

The seque	1	3	5	7	9	1	1	1	1	2	2	2	2	2	2	3	3	3	3	4	4
						1	3	7	9	1	3	5	7	8	9	1	3	5	7	1	2

Sequence number of week																				
The volume of buying gasoline/gallon	16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Table 5

**3.4.2 200 miles per week**

T2012 = 1283.5 (If the consumer drives 200 miles per week, the total cost in 2012 is 1283.5 USD). The scheme calculated by software is below:

Columns 1 through 13

The sequence number of week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
The volume of buying gasoline/gallon	16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Columns 14 through 26

The sequence number of week	22	23	24	25	26	27	28	29	30	31	32	33	34	35	37	39	40	41	42	44
The volume of	8	8	8	8	8	8	16	8	8	8	8	8	8	8	16	8	8	8	16	8



buyin g gasoli ne/ gallon																			
---------------------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table 6

According to the result of calculating the buying-gasoline scheme from the model, we can know: when the gasoline price goes up, we should fill up the tank first and fill up again immediately after using half of gasoline. It is economical to always keep the tank full and also to fill the tank in advance in order to spend least on gasoline fee. However, when gasoline price goes down, we have to use up gasoline first and then fill up the tank. In another words, we need to delay the time of filling the tank in order to pay for the lowest price. In retrospect to our model, it is very easy to discover that the situation is consistent with life experience. However, there is a difference. The result is based on the calculation from the model, while experience is just a kind of intuition.

### 3.4.3 Second Test for the Decision of buying gasoline

Since the data in 2012 is historical data now, we use artificial calculation to get the optimal value of buying gasoline. The minimum fee of driving 100 miles per week is 635.7440 USD. The result of calculating the model is 637.44 USD. The minimum fee of driving 200 miles per week is 1253.5 USD. The result of calculating the model is 1283.5 USD. The values we calculate is close to the result of the model we build. It means our model prediction effect is good. (we mention the decision people made every week and the gas price in the future is unknown. We can only predict. It's normal to have deviation. The buying-gasoline fee which is based on predicted calculation must be higher than the minimum buying-gasoline fee which is calculated when all the gas price data are known.)

We use MATLAB again to calculate the total buying-gasoline fee when  $n=4$  and  $n=5$ . When  $n=4$ , the total fee of driving 100 miles per week is 639.4560 USD and the total fee of driving 200 miles per week is 1285 USD. When  $n=5$ , the total fee of driving 100 miles per week is 639.5840 USD and the total fee of driving 200 miles per week is 1285.9 USD. The total fee are all higher the fee when  $n=3$ . It means it is best for us to take the average prediction model of 3 phases.

## 4. The upper bound will change the Decision of buying gasoline.

Assume the consumer has a mileage driven of  $x_1$  miles per week. Then, we can use  $\frac{200}{x_1}$  to indicate the period of consumption, for half of a tank can supply 200-mile driving. Here are two situations:

$$\textcircled{1} \frac{200}{x_1} < 1.5$$

$$\textcircled{2} \frac{200}{x_1} > 1.5$$

In situation①, the consumer is more likely to apply the decision of 200-mile consumer's; otherwise, it is wiser to adopt the decision of 100-mile consumer's.

Therefore,  $x_1$  is a critical value that changes the decision if

$$\frac{200}{x_1} = 1.5$$

$$x_1 = 133.3.$$

Thus, the mileage driven of 133.3 miles per week changes the buying decision.

Then, we consider the full-tank buyers likewise. The 100-mile consumer buys half a tank once in four weeks; the 200-mile consumer buys half a tank once in two weeks. The midpoint of buying period is 3 weeks.

Assume the consumer has a mileage driven of  $x_2$  miles per week. Then, we can use  $\frac{400}{x_2}$  to illustrate the buying period, since a full tank contains 400 gallons. There are still two situations:

$$\textcircled{3} \frac{400}{x_2} < 3$$

$$\textcircled{4} \frac{400}{x_2} > 3$$

In situation③, the consumer needs the decision of 200-mile consumer's to prevent the gasoline from running out; in the latter situation, it is wiser to tend to the decision of 100-mile consumer's. Therefore,  $x_2$  is a critical value that changes the decision if

$$\frac{400}{x_2} = 3$$

$$x_2 = 133.3$$

We can find that  $x_2 = x_1 = 133.3$ .

To wrap up, there exists an upper bound on “mileage driven”, that 133.3 miles per week is the value to switch the decision for buying weekly gasoline. The following picture simplifies the process.

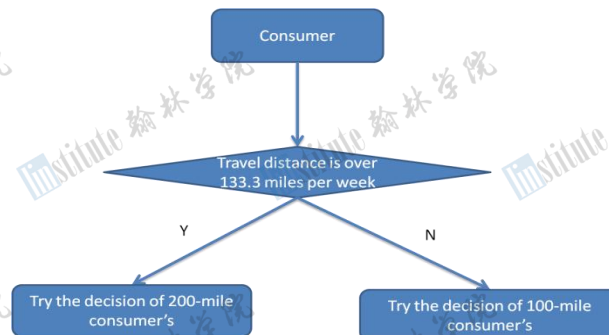


Chart 4

## 5. An analysis of New York City

### 5.1 The main factors that will affect the gasoline price in New York City

Based on the models above, we decide to estimate the price of gasoline according to the data collected and real circumstances in several cities. Specifically, we choose New York City as a representative one.

New York City stands in the North East in the United States, with the largest population throughout the country as 8.2 million. The total area of New York City is around 1300 km<sup>2</sup>, with the land area as 785.6 km<sup>2</sup> (303.3 mi<sup>2</sup>). One of the largest trading centers in the world, New York City has a high level of resident's consumption. As a result, the level of the price of gasoline in New York City is higher than the average regular oil price of the United States. The price level of gasoline and its fluctuation are the main factors of buying decision.

Another reasonable factor we expect is the distribution of gas stations. According to the latest report, there are approximately 1670 gas stations in the city area (However, after the impact of hurricane Sandy, about 90 gas stations have been temporarily out of use because of the devastation of Sandy, and there is still around 1580 stations remaining). From the information above, we can calculate the density of gas stations that

$$D(\text{gasoline station}) = \frac{\text{the amount of gas stations}}{\text{total land area}} = \frac{1670 \text{ stations}}{303.3 \text{ mi}^2} = 5.506 \text{ stations per mi}^2$$

This is a respectively high value compared with several other cities the United States. It also indicates that the average distance between gas stations is relatively small. The fact that we can neglect the distance for the cars to get to the station highlights the role of the fluctuation of the price of gasoline in New York City.

Also, there are approximately 1.8 million residents of New York City hold the driving license. Because the exact amount of cars in New York City is hard to determine, we choose to analyze the distribution of possible consumers. Thus, we can directly estimate the density of consumers in New York City in a similar way as that of gas stations:

$$D(\text{gasoline consumers}) = \frac{\text{the amount of consumers}}{\text{total land area}} = \frac{1.8 \text{ million consumers}}{303.3 \text{ mi}^2} = 5817 \text{ consumers per mi}^2$$

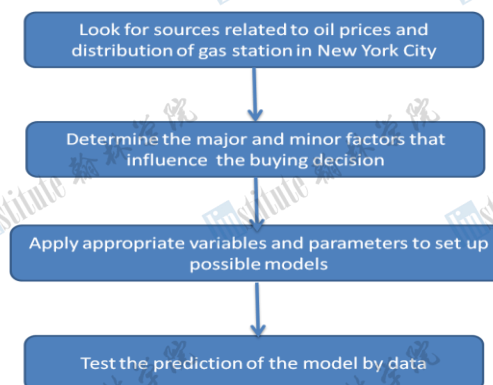


Chart 5



In addition, we expect that the fluctuation of the price of crude oil plays a critical role of the buying decision. The media in New York City is well developed, so it is convenient for citizens to look for the data of the instant price of crude oil, then to estimate the price of gasoline for the coming week if the result of our model conforms to the assumption. We will include all of these considerations in our modification of the model, which we will discuss in the next few steps.

For the analysis of New York City, we apply two different models to estimate the price and help consumers make the decision.

### 5.2 Test Models with New York data

Among the cities in US, we pick up New York as an typical example. The gas price data is downloaded from the website ([www.eia.gov](http://www.eia.gov)) and is used in the model described in Section 2 and 3.

The gas price curves between the observed data and prediction data are compared in next Figure.

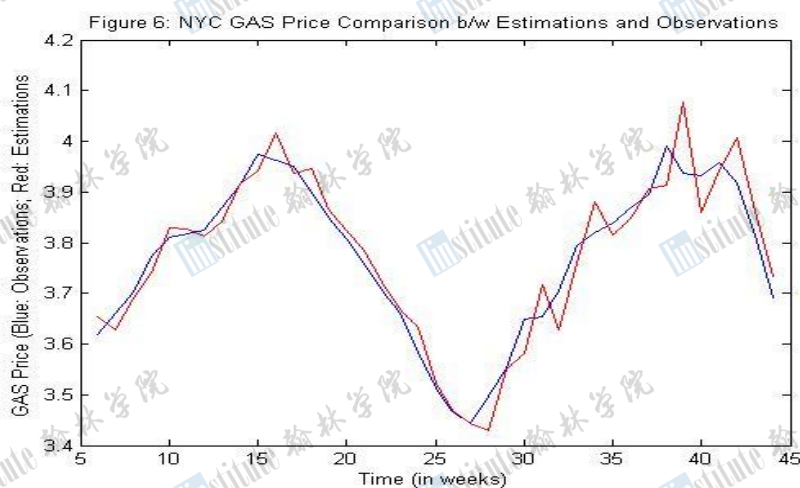


Figure 6

The gas price between the observed data and predicted data of New York is very similar to Figure 3 in US case.

Since there is little difference between the National case and New York case, the purchase strategy is same. Following the same procedure, we can compare the gas cost between the historical result and predicted result.

For the case of 100 miles per week, the total cost of observed data from Feb to Oct of 2012 in New York is 636.26USD, while the total cost of predicted data in the same period is 638.78USD, which is very close. It proves that our prediction model is good. For the case of 200 miles per week, the total cost of observed data from Feb to Oct of 2012 in New York is 1271.2USD, while the total cost of predicted data in the same period is 1277.6USD, which is very close. It proves that our prediction model is good also.

### 5.3 The analysis of result

By comparing, though density of gas stations and density of consumers of New York is a little higher than other places but it can't lower the total buying-gas fee. In



another words, density of gas stations and density of consumers are not the actual factors of affecting buying-gas fee.

On the other hand, we find the gas fee in New York is a bit higher than the average fee in US. We can only analyze preliminary it is because of the higher goods price in New York. We need to add price factor into prediction model. We can't improve deeper because of the limited time. The average CPI table of New York City and USA is below:

Consumer Price Index (1982-84=100 unless otherwise noted)(not seasonally adjusted)

Area (Links provide news releases)	Annual average 2011	Sep 2012	Oct 2012
<u>U.S. City Average</u>	224.939	231.407	231.317
<u>New York-Northern New Jersey</u>	247.718	254.554	254.277

Datas are from: Bureau of Labor Statistics website([http://www.bls.gov/xg\\_shells/ro2xg01.htm](http://www.bls.gov/xg_shells/ro2xg01.htm))

## 6. Summery& Advantage and disadvantage

To reach the solution, we make graphs of crude oil and gasoline respectively and find the similarity between them. Since the conditions are limited that consumers can only drive 100miles per week or 200miles per week, we separate the problem into two parts according to the limitation. we use Time series analysis Method to predict the gasoline price of a future period by the data of several periods in the past. Then we take the influence of international events, economic events and weather changes and so on into consideration by adding a parameter. We give each factor a weight consequently and find the rules of the solution of 100miles per week and 200miles per week. Then we discuss the upper bound and clarify the definition of upper bound to solve the problem.

According to comparison from many different aspects, we confirm that the model expressed by Equation ⑨ is the best. On the basis of historical data and the decision model of buying gasoline(Equation ⑦ and Equation ⑧), we calculate that the actual least cost of buying gasoline is 635.7440 USD if the consumer drives 100 miles per week (the result of our model is 637.24 USD) and the actual least cost of buying gasoline is 1253.5 USD(the result of our model is 1283.5 USD) if the consumer drives 100 miles per week. The result we predicted is similar to the actual result so the predictive validity of our model is finer.

Disadvantages:

1. The events which we predicted are difficult to quantize accurately. The turning point is difficult for us to predict accurately as well.
2. We only choose two kinds of train of thought to develop models so we cannot evaluate other methods that we did not discuss in this paper. Other models which are built up by other train of thought are possible to be the optimal solution.

We only considerate 100miles/week to 200miles/week so we cannot predict if the distance consumer drives is less than 100miles/week or more than 200miles/week. From these, our model has limitation.

3. On the basis of analysis of New York City, we find that the commodity price may be one of the factors that will affect the total cost of buying gasoline, but we cannot build a better model that fully discuss this factor because of the time limit.

## 7. Report

Many people face this question when they stop at a gas station: To fill the full tank every time we reach the station, or to purchase only half of the tank? Is there a margin which people may act differently on two sides of it? Will the price of gasoline and crude oil from the past affect our decisions? To how much degree is the market of gasoline affected by political events, economic events and natural calamities? Besides these, a bunch of questions can be added on the list.

In order to make the right decisions, we should predict how the gasoline price is likely to change in the future more concisely.

Firstly, we want to remind you that the awareness of the price change in crude oil from the past may present you with an index of the gasoline price in the future. According to our calculation, we find the changing pattern of crude oil resembles that of gasoline. However, considering the period of refinement and the efficiency of the transportation, we find that the graph of crude oil price is approximate a week earlier than that of gasoline. Therefore, we strongly recommend drivers to notice the change in crude oil and make their gas-filling choices by predicting the future prices.

Secondly, besides the effect of the crude oil prices, the influence of unpredictable events cannot be neglected. Real life can never be explained only by the scientific statistics data, we should also pay attention to those influential events all over the world, including both human activities and natural destructions. For example, on Oct.21<sup>st</sup>, 2012, hurricane sandy swept through the eastern of America. Many people in the country are likely to benefit from the lower gasoline price, which is caused by Hurricane Sandy's depressing of regional fuel demand. Therefore, we can apply the factor of events to the prediction of gasoline price. On the part of political reasons, OPEC can be an important factor that may change the situation of gasoline price because the members of OPEC are the main oil export countries which control a large quantity of supply of crude oil all over the world. If you want to know more about how the price of gasoline may change, you'd better read some news about main crude oil export countries, for example, countries in the Middle East.

After predicting the future gasoline price, now you are able to decide whether to fill full or half of the tank. Here is a simple principle for your decision. If you predict that the gasoline price may rise in the near future, you should fill the tank firstly and fill the tank after using a half tank. That is, in order to pay less money, the best way is to keep the tank full by buying gasoline as earlier as possible. If you predict the price will lower, you can keep the tank half. You should go to buy gasoline after you deplenish the gasoline in order to prolong the time that you buy gasoline.

Hope you and your family enjoy your cars in a more economical way!

## 8. Appendix

### Appendix1 ( main MATLAB Programs)

Computer program 1:

```
% Use the data of today, yesterday and the day before yesterday to predict the data
of tomorrow.
```

```
clear
```

```
% Read the data
```

```
A = xlsread('PET_PRI_GND_A_EPMR_PTE_DPGAL_W.xls','Data 1');
```

```
Afit = A(1055 : end);
```

```
A2011 = A(1058 : 1109);
```

```
A2012 = A(1109 : end);           % Start to collect value from the last week
```

```
B1 = xlsread('PET_PRI_SPT_S1_W.xls','Data 1');
```

```
B1fit = B1(1303 : end,:) / 42;
```

```
B1fit_aver = (B1fit(:,1) + B1fit(:,2)) / 2;
```

```
% Use the data of 2011 to do fitting and get the corresponding coefficient
```

```
X = [ones(size(B1fit,1) - 2,1),Afit(1 : end - 3),Afit(2 : end - 2),Afit(3 : end - 1),...
```

```
    B1fit_aver(1 : end - 2),B1fit_aver(2 : end - 1),B1fit_aver(3 : end)];
```

```
X1 = X(4 : 52,:);
```

```
Y1 = Afit(7 : 55);
```

```
[bpre1,bint1,r1,rint1,stat1] = regress(Y1,X1);
```

```
bpre1
```

```
% Calculate the fitted price of 2011
```

```
nihe2011 = X(1 : 52,:) * bpre1;
```

```
% Calculate the predicted price of 2012
```

```
Apr1 = X(53 : end,:) * bpre1;
```

```
% Make the graph of the fitted price and predicted price of 2011 to compare
intuitively
```

```
plot([1 : 52],[A2011,nihe2011])
```

```
grid on
```

```
legend('Real price','Fitted price','Location','SouthEast')
```

```
title('Real price and Fitted price for 2011 (n=3)');
```

```
pause()
```

```
% Make the graph of the actual price and predicted price of 2012 to compare
intuitively.
```

```
plot([1:length(Afit(56 : end))],[Afit(56 : end),Apr1])
```

```
grid on
```

```
legend('Real price','Forecast price','Location','SouthEast')
```

```
title('Real price and Forecast price for 2012 (n=3)');
```



```

% Calculate the amount of error number
d_A2012 = diff(A2012);
d_Apre1 = Apre1 - A2012(1 : end - 1);
d2 = d_A2012 .* d_Apre1;
sum(d2<0)

% The decision of buying gasoline if the consumer drives 100 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1(i) >= A2012(i)
        if fuel == 0
            intime_100(k,1) = i;
            intime_100(k,2) = 16;
            fuel = fuel + 16;
            k = k + 1;
        elseif fuel <= 8
            intime_100(k,1) = i;
            intime_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    else
        if fuel == 0
            intime_100(k,1) = i;
            intime_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
        end
        fuel = fuel - 4;
        if fuel < 0 || fuel > 16
            [i,fuel]
        end
    end
end

% The total cost of buying gasoline if the consumer drives 100 miles per week
disp('100 kilometers per week');
T2012 = intime_100'
Total100_2012 = 0;
for i = 1 : length(T2012(1,:))
    Total100_2012 = Total100_2012 + T2012(2,i) * A2012(T2012(1,i) + 1);
end
Total100_2012

```



```

% The decision of buying gasoline if the consumer drives 200 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1(i) >= A2012(i)
        if fuel == 0
            intime_200(k,1) = i;
            intime_200(k,2) = 16;
            fuel = fuel + 16;
            k = k + 1;
        elseif fuel <= 8
            intime_200(k,1) = i;
            intime_200(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    else
        if fuel == 0
            intime_200(k,1) = i;
            intime_200(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    end
    if fuel > 16
        [i,fuel]
    end
    fuel = fuel - 8;
    if fuel < 0
        [i,fuel]
    end
end
% The total cost of buying gasoline if the consumer drives 200 miles per week.
disp('200 kilometers per week');
T2012 = intime_200'
Total200_2012 = 0;
for i = 1 : length(T2012(1,:))
    Total200_2012 = Total200_2012 + T2012(2,i) * A2012(T2012(1,i) + 1);
end
Total200_2012

```

Computer program 2:

```
% Use the datas of today, yesterday, the day before yesterday and three days ago to
predict the data of tomorrow.
```

```
clear
```

```
% Read data
```

```
A = xlsread('PET_PRI_GND_A_EPMR_PTE_DPGAL_W.xls','Data 1');
```

```
Afit = A(1055 : end);
```

```
A2011 = A(1058 : 1109);
```

```
A2012 = A(1109 : end); % Start to deferencing value from the last week of
2011
```

```
B1 = xlsread('PET_PRI_SPT_S1_W.xls','Data 1');
```

```
B1fit = B1(1303 : end,:) / 42;
```

```
B1fit_aver = (B1fit(:,1) + B1fit(:,2)) / 2;
```

```
% Use the data of 2011 to do fitting and get corresponding coefficient
```

```
X = [ones(size(B1fit,1) - 3,1), ...
```

```
    Afit(1 : end - 4),Afit(2 : end - 3),Afit(3 : end - 2),Afit(4 : end - 1), ...
```

```
    B1fit_aver(1 : end - 3),B1fit_aver(2 : end - 2),B1fit_aver(3 : end -
1),B1fit_aver(4 : end)];
```

```
X1 = X(4 : 51,:);
```

```
Y1 = Afit(8 : 55);
```

```
[bpre1,bint1,r1,rint1,stat1] = regress(Y1,X1);
```

```
bpre1
```

```
% Calculate the fitted price of 2011
```

```
nihe2011 = X(1 : 52,:) * bpre1;
```

```
% Calculate the predicted price of 2012
```

```
Apr1 = X(52 : end,:) * bpre1;
```

```
% Make the graph of the fitted price and predicted price of 2011 to compare
intuitively
```

```
plot([1 : 52],[A2011,nihe2011])
```

```
grid on
```

```
legend('Real price','Fitted price','Location','SouthEast')
```

```
title('Real price and Fitted price for 2011 (n=4)');
```

```
pause()
```

```
% Make the graph of the actual price and predicted price of 2012 to compare
intuitively
```

```
plot([1:length(Afit(56 : end))],[Afit(56 : end),Apr1])
```

```
grid on
```

```
legend('Real price','Forecast price','Location','SouthEast')
```

```
title('Real price and Forecast price for 2012 (n=4)');
```

```

% The decision of buying gasoline if the consumer drives 100 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1(i) >= A2012(i)
        if fuel == 0
            intime_100(k,1) = i;
            intime_100(k,2) = 16;
            fuel = fuel + 16;
            k = k + 1;
        elseif fuel <= 8
            intime_100(k,1) = i;
            intime_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    else
        if fuel == 0
            intime_100(k,1) = i;
            intime_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    end
    fuel = fuel - 4;
    if fuel < 0 || fuel > 16
        [i,fuel]
    end
end

% The total cost of buying gasoline if the consumer drives 100 miles per week
disp('100 kilometers per week');
T2012 = intime_100'
Total100_2012 = 0;
for i = 1 : length(T2012(1,:))
    Total100_2012 = Total100_2012 + T2012(2,i) * A2012(T2012(1,i) + 1);
end
Total100_2012

% The decision of buying gasoline if the consumer drives 200 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1(i) >= A2012(i)
        if fuel == 0

```

```

        intime_200(k,1) = i;
        intime_200(k,2) = 16;
        fuel = fuel + 16;
        k = k + 1;
    elseif fuel <= 8
        intime_200(k,1) = i;
        intime_200(k,2) = 8;
        fuel = fuel + 8;
        k = k + 1;
    end
else
    if fuel == 0
        intime_200(k,1) = i;
        intime_200(k,2) = 8;
        fuel = fuel + 8;
        k = k + 1;
    end
end
if fuel > 16
    [i,fuel]
end
fuel = fuel - 8;
if fuel < 0
    [i,fuel]
end
end
% The total cost of buying gasoline if the consumer drives 200 miles per week.
disp('200 kilometers per week');
T2012 = intime_200'
Total200_2012 = 0;
for i = 1 : length(T2012(1,:))
    Total200_2012 = Total200_2012 + T2012(2,i) * A2012(T2012(1,i) + 1);
end
Total200_2012

% Calculate the amount of error number
d_A2012 = diff(A2012);
d_Apre1 = Apre1 - A2012(1 : end - 1);
d2 = d_A2012 .* d_Apre1;
sum(d2 < 0)

```



Computer program 3:

```
% Use the data of today, yesterday, the day before yesterday and three days ago and
four days ago to predict the data of tomorrow
```

```
clear
```

```
% Read data
```

```
A = xlsread('PET_PRI_GND_A_EPMR_PTE_DPGAL_W.xls','Data 1');
```

```
Afit = A(1055 : end);
```

```
A2011 = A(1058 : 1109);
```

```
A2012 = A(1109 : end); % Start dereferencing the data
```

```
B1 = xlsread('PET_PRI_SPT_S1_W.xls','Data 1');
```

```
B1fit = B1(1303 : end,:) / 42;
```

```
B1fit_aver = (B1fit(:,1) + B1fit(:,2)) / 2;
```

```
% Use the data of 2011 to do fitting and get the corresponding coefficient
```

```
X = [ones(size(B1fit,1) - 4,1), ...
```

```
    Afit(1 : end - 5),Afit(2 : end - 4),Afit(3 : end - 3),Afit(4 : end - 2),Afit(5 : end -
    1), ...
```

```
    B1fit_aver(1 : end - 4),B1fit_aver(2 : end - 3),B1fit_aver(3 : end -
    2),B1fit_aver(4 : end - 1),B1fit_aver(5 : end)];
```

```
X1 = X(4 : 50,:);
```

```
Y1 = Afit(9 : 55);
```

```
[bpre1,bint1,r1,rint1,stat1] = regress(Y1,X1);
```

```
bpre1
```

```
% Calculate the fitted price of 2011
```

```
nihe2011 = X(1 : 52,:) * bpre1;
```

```
% Calculate the predicted price of 2012
```

```
Apr1 = X(51 : end,:) * bpre1;
```

```
% Make the graph of the fitted price and predicted price of 2011 to compare
intuitively
```

```
plot([1 : 52],[A2011,nihe2011])
```

```
grid on
```

```
legend('Real price','Fitted price','Location','SouthEast')
```

```
title('Real price and Fitted price for 2011 (n=5)');
```

```
pause()
```

```
% Make the graph of the actual price and predicted price of 2012 to compare
intuitively.
```

```
plot([1:length(Afit(56 : end))],[Afit(56 : end),Apr1])
```

```
grid on
```

```
legend('Real price','Forecast price','Location','SouthEast')
```

```
title('Real price and Forecast price for 2012 (n=5)');
```

```

% The decision of buying gasoline if the consumer drives 100 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1(i) >= A2012(i)
        if fuel == 0
            intime_100(k,1) = i;
            intime_100(k,2) = 16;
            fuel = fuel + 16;
            k = k + 1;
        elseif fuel <= 8
            intime_100(k,1) = i;
            intime_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    else
        if fuel == 0
            intime_100(k,1) = i;
            intime_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    end
    fuel = fuel - 4;
    if fuel < 0 || fuel > 16
        [i,fuel]
    end
end
end
% The total cost of buying gasoline if the consumer drives 100 miles per week
disp('100 kilometers per week');
T2012 = intime_100'
Total100_2012 = 0;
for i = 1 : length(T2012(1,:))
    Total100_2012 = Total100_2012 + T2012(2,i) * A2012(T2012(1,i) + 1);
end
Total100_2012

% The decision of buying gasoline if the consumer drives 200 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1(i) >= A2012(i)

```

```

    if fuel == 0
        intime_200(k,1) = i;
        intime_200(k,2) = 16;
        fuel = fuel + 16;
        k = k + 1;
    elseif fuel <= 8
        intime_200(k,1) = i;
        intime_200(k,2) = 8;
        fuel = fuel + 8;
        k = k + 1;
    end
else
    if fuel == 0
        intime_200(k,1) = i;
        intime_200(k,2) = 8;
        fuel = fuel + 8;
        k = k + 1;
    end
end
if fuel > 16
    [i,fuel]
end
fuel = fuel - 8;
if fuel < 0
    [i,fuel]
end
end
% The total cost of buying gasoline if the consumer drives 200 miles per week.
disp('200 kilometers per week');
T2012 = intime_200'
Total200_2012 = 0;
for i = 1 : length(T2012(1,:))
    Total200_2012 = Total200_2012 + T2012(2,i) * A2012(T2012(1,i) + 1);
end
Total200_2012

% Calculate the amount of error number
d_A2012 = diff(A2012);
d_Apre1 = Apre1 - A2012(1 : end - 1);
d2 = d_A2012 .* d_Apre1;
sum(d2<0)

```

Computer program 4:

```

% The decision program of buying gasoline in New York City
clear
A = xlsread('PET_PRI_GND_A_EPMR_PTE_DPGAL_W.xls','Data 1');
Afit = A(1055 : end);
A2011 = A(1058 : 1109);
A2012 = A(1109 : end);
Acity = xlsread('PET_PRI_GND_A_EPMR_PTE_DPGAL_W.xls','Data 4');
Acity2009_2012 = Acity(449:end,:);
Acityfit = Acity2009_2012(102 : end,:);
Acity2011 = Acity2009_2012(105 : 156,:);
Acity2012 = Acity2009_2012(156 : end,:);
Anewyorkfit = Acityfit(:,8);
Anewyork2011 = Acity2011(:,8);
Anewyork2012 = Acity2012(:,8);

B1 = xlsread('PET_PRI_SPT_S1_W.xls','Data 1');
B1fit = B1(1303 : end,:) / 42;
B1fit_aver = (B1fit(:,1) + B1fit(:,2)) / 2;

Xnewyork = [ones(size(B1fit,1) - 2,1),...
    Anewyorkfit(1 : end - 3),Anewyorkfit(2 : end - 2),Anewyorkfit(3 : end - 1),...
    B1fit_aver(1 : end - 2),B1fit_aver(2 : end - 1),B1fit_aver(3 : end)];

X1newyork = Xnewyork(4 : 52,:);
Y1newyork = Anewyorkfit(7 : 55);
[bpre1newyork,bint1,r1,rint1,stat1] = regress(Y1newyork,X1newyork);
bpre1newyork
% Calculate the fitted price of 2011 in New York City
nihenewyork2011 = Xnewyork(1 : 52,:) * bpre1newyork;
% Calculate the predicted price of 2012 in New York City
Apre1newyork = Xnewyork(53 : end,:) * bpre1newyork;

% Make the graph of the fitted price and predicted price of 2011 to compare
intuitively in New York City
plot([1 : 52],[Anewyork2011,nihenewyork2011])
grid on
legend('Real price','Fitted price','Location','SouthEast')
title('New York Real price and Fitted price for 2011 (n=3)');
pause();

% Make the graph of the actual price and predicted price of 2012 to compare
intuitively in New York City
plot([1:length(Anewyorkfit(56 : end))],[Anewyorkfit(56 : end),Apre1newyork])
grid on

```



```

legend('Real price','Forecast price','Location','SouthEast')
title('NewYork Real price and Forecast price for 2012 (n=3)');

% Calculate the amount of error number in New York City
d_Anewyork2012 = diff(Anewyork2012);
d_Apre1newyork = Apre1newyork - Anewyork2012(1 : end - 1);
d2newyork = d_Anewyork2012 .* d_Apre1newyork;
sum(d2newyork<0)

% The decision of buying gasoline if the consumer drives 100 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1newyork(i) >= Anewyork2012(i)
        if fuel == 0
            intimenewyork_100(k,1) = i;
            intimenewyork_100(k,2) = 16;
            fuel = fuel + 16;
            k = k + 1;
        elseif fuel <= 8
            intimenewyork_100(k,1) = i;
            intimenewyork_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    else
        if fuel == 0
            intimenewyork_100(k,1) = i;
            intimenewyork_100(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    end
    fuel = fuel - 4;
    if fuel < 0 || fuel > 16
        [i,fuel]
    end
end
% The total cost of buying gasoline if the consumer drives 100 miles per week
disp('100 kilometers per week at NewYok');
Tnewyork2012 = intimenewyork_100'
Totalnewyork100_2012 = 0;
for i = 1 : length(Tnewyork2012(1,:))
    Totalnewyork100_2012 = Totalnewyork100_2012 + Tnewyork2012(2,i) *

```

```

Anewyork2012(Tnewyork2012(1,i) + 1);
end
Totalnewyork100_2012

% The decision of buying gasoline if the consumer drives 200 miles per week
fuel = 0;
k = 1;
for i = 1 : 44
    if Apre1newyork(i) >= Anewyork2012(i)
        if fuel == 0
            intimenewyork_200(k,1) = i;
            intimenewyork_200(k,2) = 16;
            fuel = fuel + 16;
            k = k + 1;
        elseif fuel <= 8
            intimenewyork_200(k,1) = i;
            intimenewyork_200(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    else
        if fuel == 0
            intimenewyork_200(k,1) = i;
            intimenewyork_200(k,2) = 8;
            fuel = fuel + 8;
            k = k + 1;
        end
    end
    if fuel > 16
        [i,fuel]
    end
    fuel = fuel - 8;
    if fuel < 0
        [i,fuel]
    end
end
end
% The total cost of buying gasoline if the consumer drives 200 miles per week.
disp('200 kilometers per week at NewYok');
Tnewyork2012 = intimenewyork_200'
Totalnewyork200_2012 = 0;
for i = 1 : length(Tnewyork2012(1,:))
    Totalnewyork200_2012 = Totalnewyork200_2012 + Tnewyork2012(2,i) *
    Anewyork2012(Tnewyork2012(1,i) + 1);
end

```

2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4

2 3 4 5 6 7 8 9 0 1 2 3 4 5 7 9 0 1 2 4  
8 8 8 8 8 8 1 8 8 8 8 8 8 1 8 8 8 1 8  
6 6



bpre1 =  
0.0537 0.0280 -0.0925 -0.5570 1.5358 0.1646 0.0048 -0.4766 0.4093

ans = 7

100 kilometers per week

T2012 = 639.4560

1 3 5 7 9 1 1 1 1 2 2 2 2 2 3 3 3 4 4  
1 3 5 9 1 3 5 7 8 9 2 3 7 1 2  
1 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 1 8  
6 6

200 kilometers per week

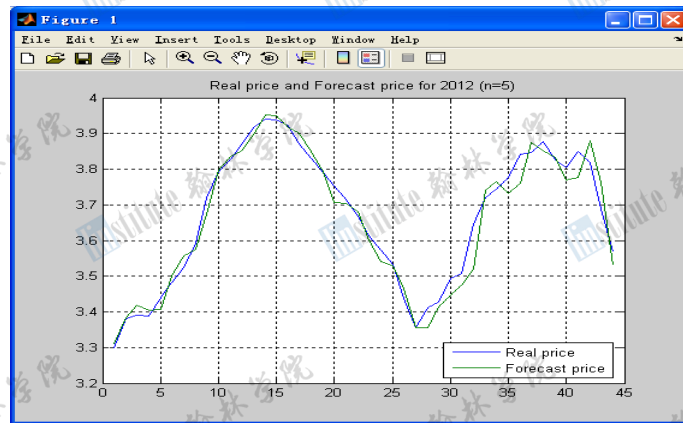
T2012 = 1285.7

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 19 20 21  
16 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8

Columns 14 through 26



2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4
2	3	4	5	6	7	8	9	0	2	3	4	6	7	8	0	1	2	4
8	8	8	8	8	8	1	8	8	1	8	8	8	1	8	8	8	1	8
						6			6				6				6	



bpref1 =

0.08	-0.12	0.17	-0.14	-0.50	1.50	0.04	0.09	0.00	-0.45	0.40
49	09	82	31	65	05	99	76	76	97	27

ans = 8

100 kilometers per week

T2012 = 639.5840

1	3	5	7	9	11	13	15	19	21	23	25	27	29	32	33	37	41	42
16	8	8	8	8	8	8	8	8	8	8	8	8	16	8	8	16	8	8

200 kilometers per week

T2012 = 1285.9

Columns 1 through 13

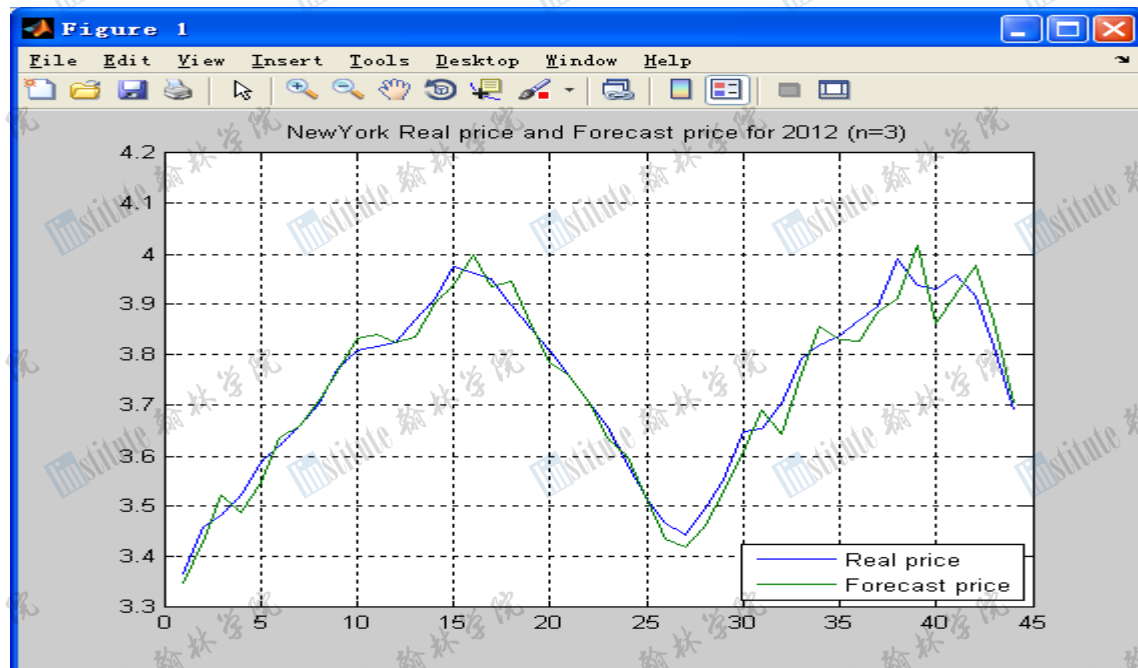
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	20	21
16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Columns 14 through 26

2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	
2	3	4	5	6	7	8	9	0	2	3	4	6	7	8	0	1	2	4	
8	8	8	8	8	8	8	1	8	1	8	8	8	1	8	8	8	1	8	

6 6 6 6

## Appendix 3: New York City Model



bpre1newyork =

-0.0096	0.1454	-0.8698	1.6856	0.0185	-0.2589	0.3014
---------	--------	---------	--------	--------	---------	--------

ans = 6

100 kilometers per week at NewYok

Tnewyork2012 = 649.6560

Columns 1 through 12

1	3	5	7	9	1	1	1	1	2	2	2	2	2	2	3	3	3	3	4
1	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
6																			

200 kilometers per week at NewYok

Tnewyork2012 = 1309.6

Columns 1 through 12

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	19	20	21
16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Columns 13 through 24

2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4
2	3	4	5	6	7	8	9	0	1	3	4	5	7	8	9	1	2	4	
8	8	8	8	8	8	1	8	8	8	1	8	8	1	8	8	8	1	6	8

