CS 534 Artificial Intelligence
Assignment 2

Group 10

March 19, 2018

Group Member

Yixuan	Jiao	yjiao@wpi.edu
Yinkai	Ma	yma7@wpi.edu
Jiaming	Nie	jnie@wpi.edu
Pinyi	Xiao	pxiao@wpi.edu

1 Gibbs Sampling

1.1 Load The Data

The data of the Bayesian network is stored in the excel, so when loading the excel, this project could get the data we need in the next step. The format of the input is as follows:

gibbs price schools=good location=ugly -u 10000 -d 0

The price is the query node, and the schools and location are the evidence nodes. 10000 means the iteration, 0 means discard 0. In this step, this project could get the request of the Gibbs sampling.

1.2 Operation of Gibbs Sampling

In order to do the Gibbs sampling, we need to know the information from the excel and input. For each iteration, we need to update the state of every node and store the state after update. After the iteration, the probability could be computed by the stored state.

The sequence of updating is shown below: ['amenities', 'neighborhood', 'children', 'age', 'location', 'size', 'schools', 'price']

1.2.1 Set the evidence

For no update is made to evidence nodes, we need an approach to distinguish them from the rest nodes.We generate a list to store the information of evidence nodes.

1.2.2 Set the Initial State

Before the iteration, this project needs a start state to do the update. The start state of each node is randomly set with the same probability. For example, for the price, the probability of price=cheap, price=OK and price=expensive are 0.33333 when setting the start state.

1.2.3 Update of the Node

After setting the start state, we need to update the node one by one. For each node, if it is not an evidence node, we compute the conditional probabilities according to the Markov blanket. Then set this node according to the new probability. If it is the evidence node, we skip this node.

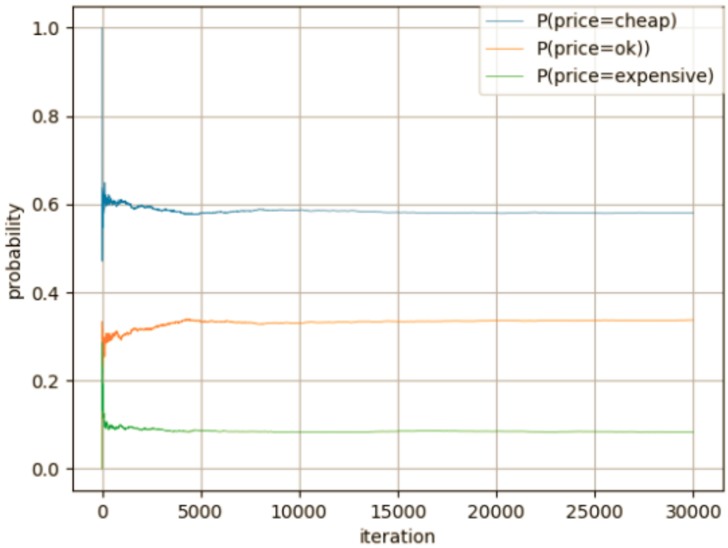
1.2.4 Iteration

After knowing the way updating the nodes, this project needs to update every node without updating evidence nodes. For each iteration, the order of updating node is the order of the sequence. After updating all the nodes in the sequence, if the discard number is 0, the number of the corresponding state of the query node plus 1, if not, discard number = discard number-1

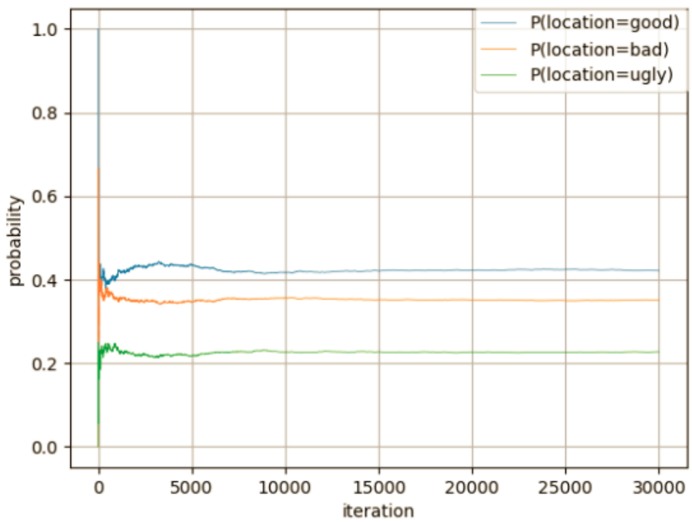
1.3 Result

After the iteration, this project calculates the probability of the query node by the number of the stored state divided by iteration time(this is calculated by total iteration time - discard time).

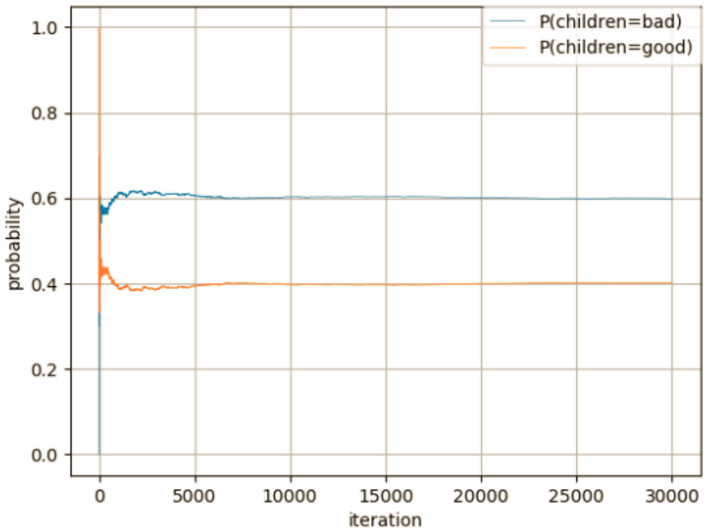
The input is: gibbs price schools=good location=ugly -u 30000 -d 0; The result plot is shown below: price1



The input is: gibbs location schools=good -u 30000 -d 0; The result plot is shown below: location1



The input is: gibbs children schools=good location=ugly -u 30000 -d 0; The result plot is shown below: children1.



1.4 Question Solution

1.4.1 Samples vs. estimated probability

For the 3 experiment, we get that for about 15000 iterations, the estimated probability to converge.

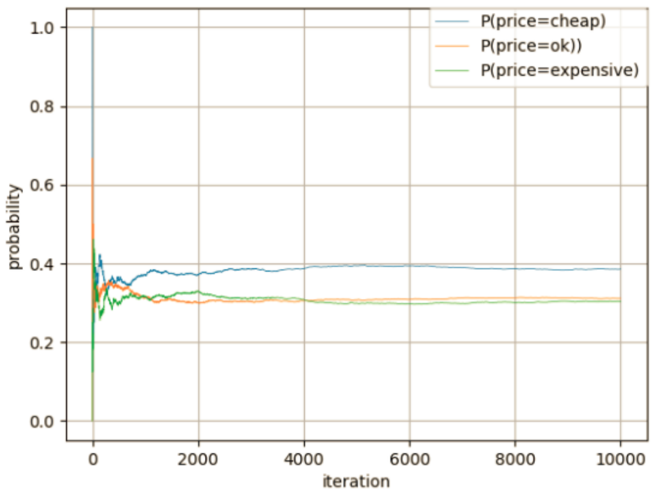
For the first plot, the probability converge at about 8000 iterations;

For the second plot, the probability converge at about 9000 iterations;

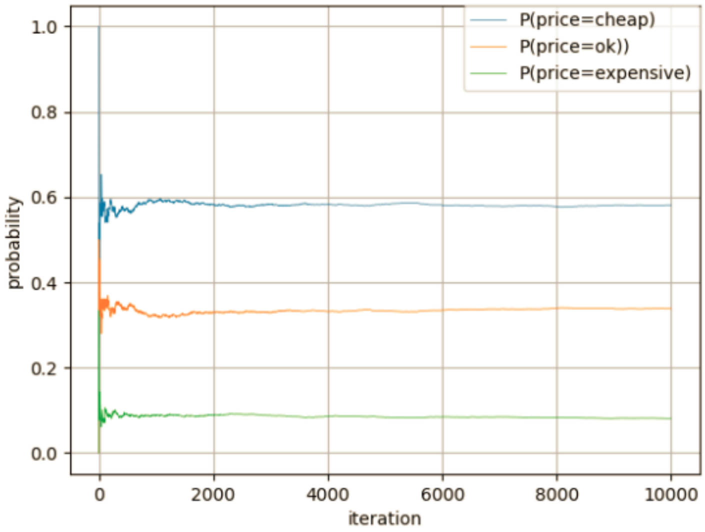
For the third plot, the probability converge at about 6000 iterations;

1.4.2 the parameter which influences the number of samples required

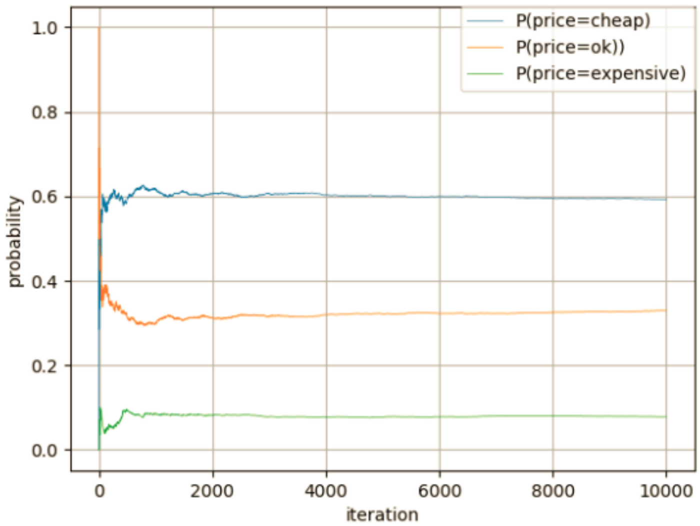
The number of the evidence node could influence. The input is: gibbs price schools=good -u 10000 -d 0; The result plot is shown below:



The input is: gibbs price schools=good location=ugly -u 10000 -d 0; The result plot is shown below:



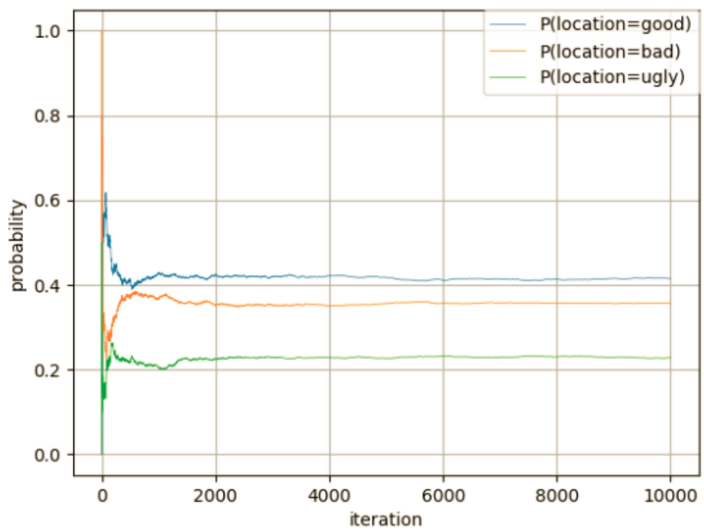
The input is: gibbs price schools=good location=ugly amenities=lots -u 30000 -d 0; The result plot is shown below:



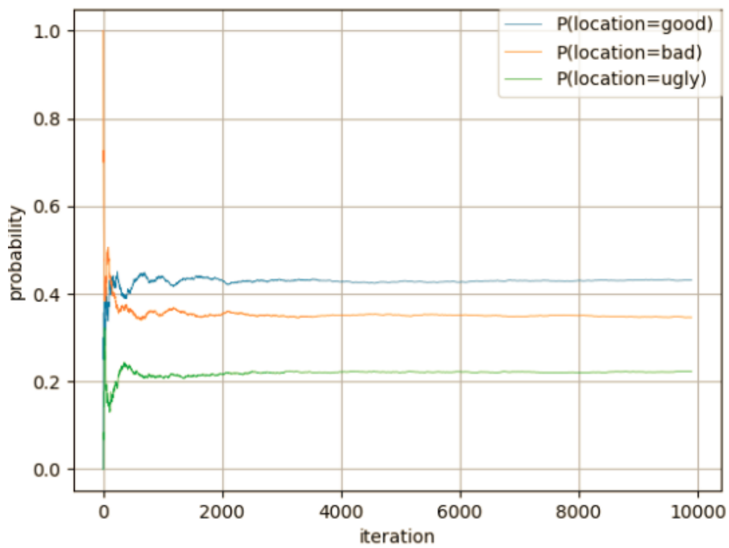
From the three plots above: For the first plot, the probability converge at about 9000 iterations; For the second plot, the probability converge at about 6000 iterations; For the third plot, the probability converge at about 4000 iterations; Thus, we conclude that the number of evidence nodes will have a positive influence on the speed of convergence.

1.4.3 Does discarding initial samples speed the convergence of probabilities

Yes, it could help speed the convergence of probabilities The input is: gibbs location schools=good -u 10000 -d 0; The result plot is shown below:



The input is: gibbs location schools=good -u 10000 -d 100; The result plot is shown below:



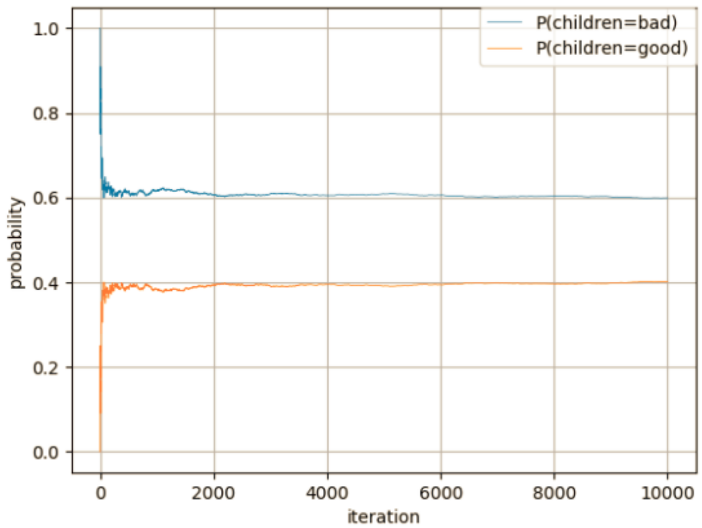
The input is: gibbs children schools=good location=ugly -u 10000 -d 0; The result plot is shown below:

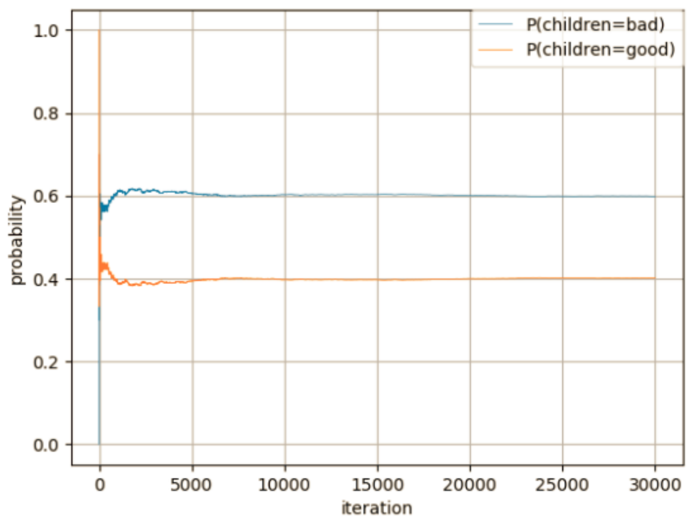
The input is: gibbs children schools=good location=ugly -u 30000 -d 0; The result plot is shown below: children1.

1.5 Question Solution

1.5.1 samples vs. estimated probability

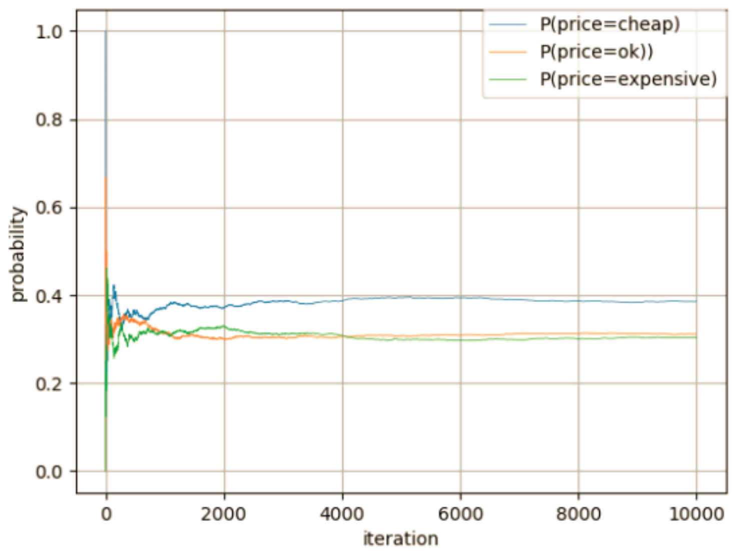
For the 3 experiment, we get that for about 15000 iterations, the estimated probability to converge. For the first plot, the probability converge at about 8000 iterations; For the second plot, the probability converge at about 9000 iterations; For the third plot, the probability converge at about 6000 iterations;


P(children=bad)
P(children=good)
1.0
0.8
0.6
0.4
0.2
0.0
probability
0
2000
4000
6000
8000
10000
iteration

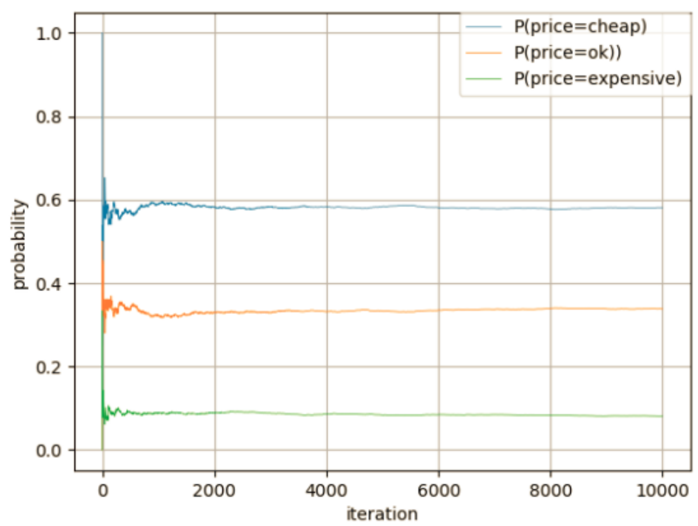

P(children=bad)
P(children=good)
1.0
0.8
0.6
0.4
0.2
0.0
probability
0
5000
10000
15000
20000
25000
30000
iteration

1.5.2 the parameter which influences the number of samples required

The number of the evidence node could influence. The input is: gibbs price schools=good -u 10000 -d 0; The result plot is shown below:

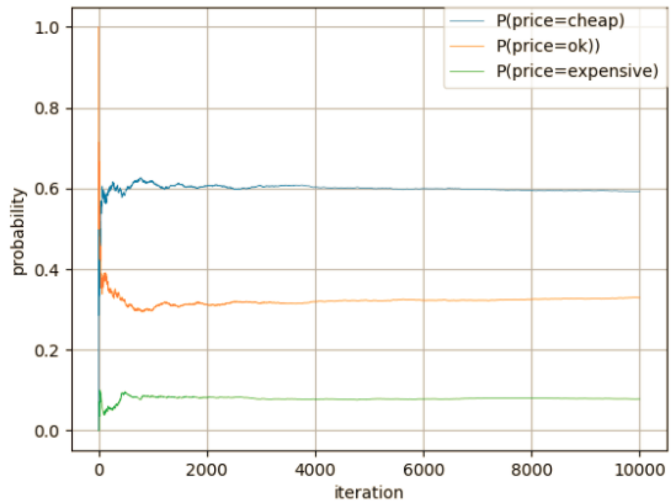


The input is: gibbs price schools=good location=ugly -u 10000 -d 0; The result plot is shown below:



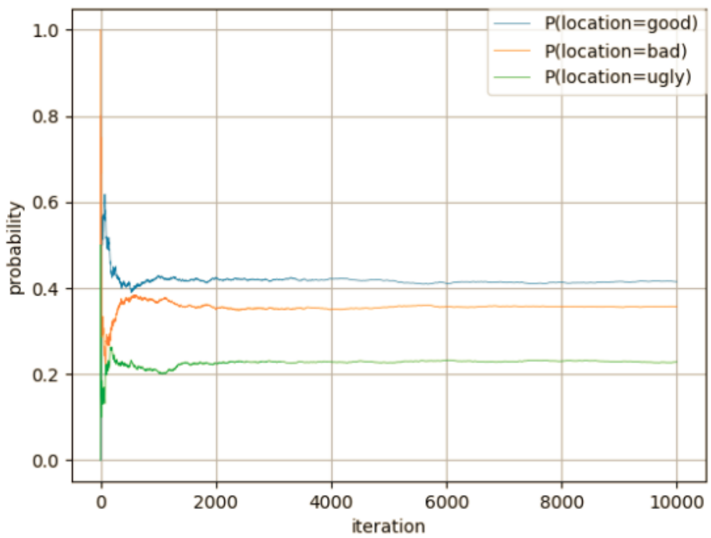
The input is: gibbs price schools=good location=ugly amenities=lots -u 30000 -d 0; The result plot is shown below:

From the three plots above: For the first plot, the probability converge at about 9000 iterations; For the second plot, the probability converge at about 6000 iterations; For the third plot, the probability converge at about 4000 iterations; Thus, we conclude that the number of evidence nodes will have a positive influence on the speed of convergence.

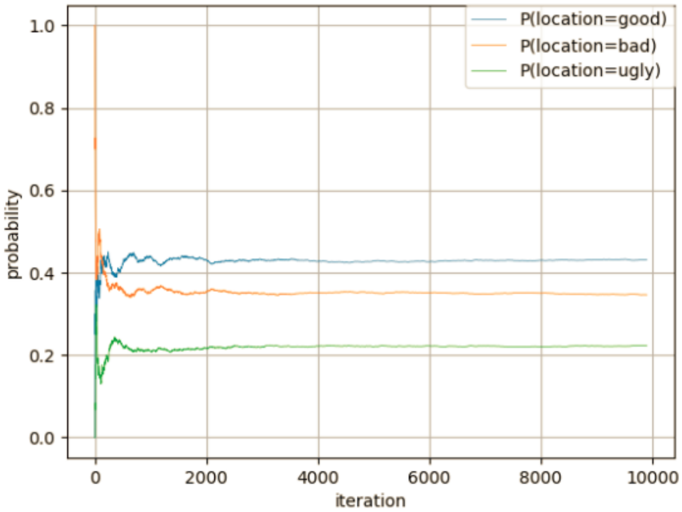


1.5.3 Does discarding initial samples speed the convergence of probabilities

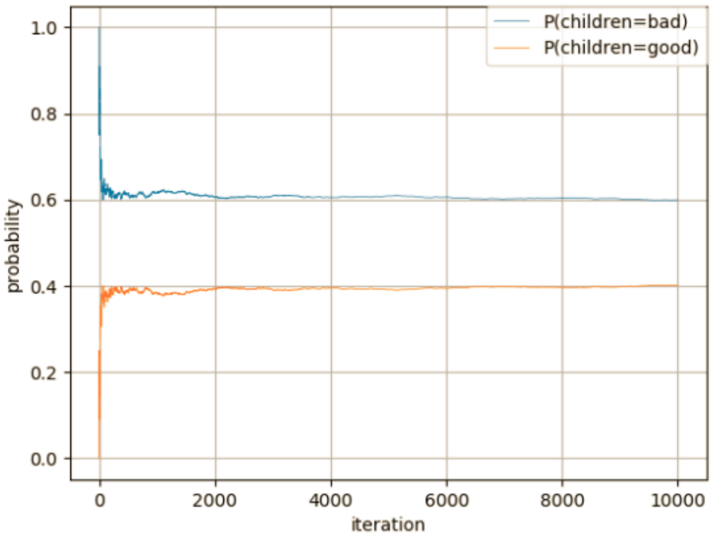
Yes, it could help speed the convergence of probabilities The input is: gibbs location schools=good -u 10000 -d 0; The result plot is shown below:



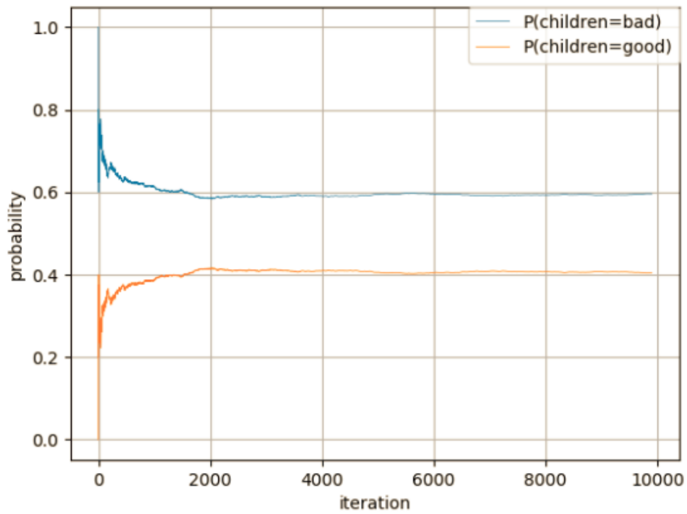
The input is: gibbs location schools=good -u 10000 -d 100; The result plot is shown below:



The input is: gibbs children schools=good location=ugly -u 10000 -d 0; The result plot is shown below:



The input is: gibbs children schools=good location=ugly -u 10000 -d 100; The result plot is shown below:



For the first couple, 0 discard converge at about 6000 iterations, while 100 discard converge at about 4000 iterations;

For the second couple, 0 discard converge at about 7000 iterations, while 100 discard converge at about 5000 iterations.

We draw to our conclusion that discarding initial samples do speed the convergence of probabilities.

2 Kalman Filter

2.1 Model Hypothesis

The transition model is constructed using the GDP and the gdp growth rate from 1970 to 2017, data source is from world bank website [1]. Unit is trillion dollar.

There are 2 sensor models, in the first model our group use the Export and Import data from 1970 to 2017 [2], the unit is value in million dollars, to measure the GDP only, which is use the 2 sensors to measure the GDP.

The second model is use 3 sensors, the 2 sensors are still import and export, the additional sensor is the employment rate [3], to measure the GDP growth rate. The data source is from Office of Management and Budget.

2.1.1 Kalman Filter Network

2.2 Methodology

2.2.1 Kalman Filter Graph Model and Equations

The graph model of the Kalman filter is on the following:

The parameters of the Kalman filter and their equations are on the following [4].

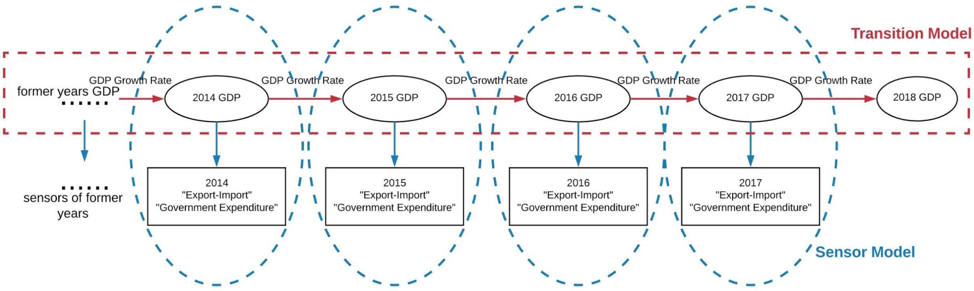


Figure 1: Kalman Filter Network

For Kalman filter, its parameters are listed in the table 1:

Table 1: Kalman Filter Parameters Notation

Parameter	Notation
x_t	The state vector containing the terms of interest for the system (e.g., position, velocity, heading) at a time
F_t	The state transition matrix which applies the effect of each system state at time t. (e.g., the position and velocity at time t-1 both affect the position at time t)
w_t	the vector which contains the noise for each parameter in the state vector x_t. A zero mean multivariate Gaussian distribution with covariance matrix Q_t is used to describe the process noise vector.
z_t	The measurements vector.
H_t	the transformation matrix which can interpret the state vector x_t into measurement domain of z_t.
v_t	the measurement noise vector for each observed value in the measurement vector z_t. The model that to establish the measurement noise is also drawn from a zero mean multivariate normal distribution with covariance matrix R_t.
R_t	The covariance matrix for the observation noise.
Q_t	The covariance matrix for the process noise.

In the Kalman filter, the state matrix can be described as 2.2.1, which is transition model.

$$x_t = F_t x_{t-1} + w_t \tag{1}$$

At the time t an observation z_t of the true state x_t is made according to 2.2.1, which represents the observation model.

$$z_t = H_t x_t + v_t \tag{2}$$

The state of the filter is represented by 2 variables:

$\widehat{x}_{t\|t}$	a posterior state estimate given the observations up to t
$\widehat{P}_{t\|t}$	a posterior error covariance estimate given the observations up to t

The predict process of the Kalman filter can be described using equation 2.2.1 and 2.2.1.

$$\widehat{t}_{t|t-1} = F_t \widehat{t}_{t-1|t-1} \tag{3}$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t \tag{4}$$

The update process of the Kalman filter is on the following, $\widetilde{y}_t$ is the residual covariance, K_t is the Kalman gain matrix:

$$\widetilde{y}_t = z_t - H_t \widehat{x}_{t|t-1} \tag{5}$$

$$S_t = R_t + H_t P_{t|t-1} H_t^T \tag{6}$$

$$K_t = P_{t|t-1} H_t^T S_t^{-1} \tag{7}$$

$$\widehat{x}_{t|t} = \widehat{x}_{t|t-1} + K_t \widetilde{y}_t \tag{8}$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1} (I - K_t H_T)^T + K_t \tag{9}$$

$$\widetilde{y}_{t|t} = z_t - H_t \widehat{x}_{t|t} \tag{10}$$

2.2.2 Transition Model

In the transition model, a state is defined by the GDP and GDP growth rate. For a state x_k, $x_k = \begin{bmatrix} x_g \\ x_{gr} \end{bmatrix}$. x_g represents the state of the GDP and x_{gr} represents the state of the GDP growth rate. The transition matrix F_t is calculated based on the matrix Q_t.

In the transition model, our group used the linear fitting to estimate the relationship between the GDP and GDP growth rate on the condition that the GDP growth rate is set to constant.

The estimated linear relationship is on the following equation 11:

$$y_{GDP} = -0.687 \times x_{growthrate} + 11.3324 \tag{11}$$

According to the linear relationship, the transition matrix F is $F = \begin{bmatrix} 1 & -0.687 \\ 0 & 1 \end{bmatrix}$

In the transition model, assume the the Pearson's coefficient ρ between the GDP growth rate and the GDP is 0, which is that the variation between GDP x_g and GDP growth rate x_{gr} does not have any dependency. Then the process noise matrix Q is

$$Q = \begin{bmatrix} var(x_g) & 0 \\ 0 & var(x_{gr}) \end{bmatrix}$$

Then the matrix w_k is $w_k = \mathcal{N}(0, Q)$.

2.2.3 Sensor Model

Export and import data will be 2 sensors to measure the GDP data and the employment rate will be the sensor to measure the GDP growth rate as the extra credit part.

Sensor Model for GDP only In the sensor model for GDP purely, the GDP is measured by the Export and Import, construct the linear polyfit model for the GDP using export and import,

the equation is on the following equation 12 and 13:

$$y_{GDP} = 0.4943 \times x_1 + 5.1524 (x_1 : Export) \tag{12}$$

$$y_{GDP} = 0.3435 \times x_2 + 5.1855 (x_2 : Import) \tag{13}$$

Then the observe matrix H is $H = \begin{bmatrix} 0.4943 & 0 \\ 0.3435 & 0 \end{bmatrix}$

Assume the export and import data does not have any dependency, then the noise matrix R is:

$$R = \begin{bmatrix} var(Export) & 0 \\ 0 & var(Import) \end{bmatrix}$$

Then the matrix v_k is $v_k = \mathcal{N}(0, R)$

Sensor Model for GDP and GDP growth Rate In the sensor model which already has 2 sensors for the GDP, add 1 sensor for the GDP growth rate.

The sensor for the GDP growth rate y adopts the employment rate x,the linear fitting result is illustrated in equation 14:

$$y = -0.074 \times x + 7.7465 \tag{14}$$

Then the observe matrix H is H is $H = \begin{bmatrix} 0.4943 & 0 \\ 0.3435 & 0 \\ 0 & -0.074 \end{bmatrix}$

Assume the export data, import data and the employment rate does not have any dependency, then the noise matrix R is:

$$R = \begin{bmatrix} var(Export) & 0 & 0 \\ 0 & var(Import) & 0 \\ 0 & 0 & var(employRate) \end{bmatrix}$$

Then the matrix v_k is $v_k = \mathcal{N}(0, R)$

2.3 Results

2.3.1 Kalman Estimate Result: Sensor for GDP only (2 Sensors)

Probability Distribution The probability distribution for GDP is illustrated in 2.

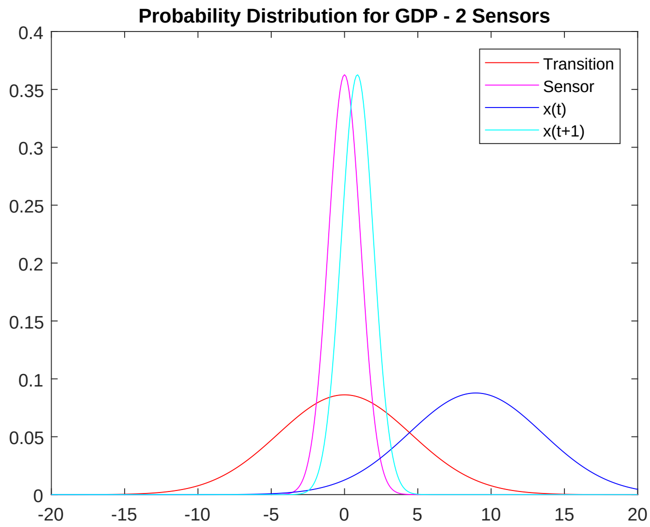


Figure 2: 2-Sensor Model GDP Probability Distribution

The probability distribution for GDP growth rate is illustrated in 3.

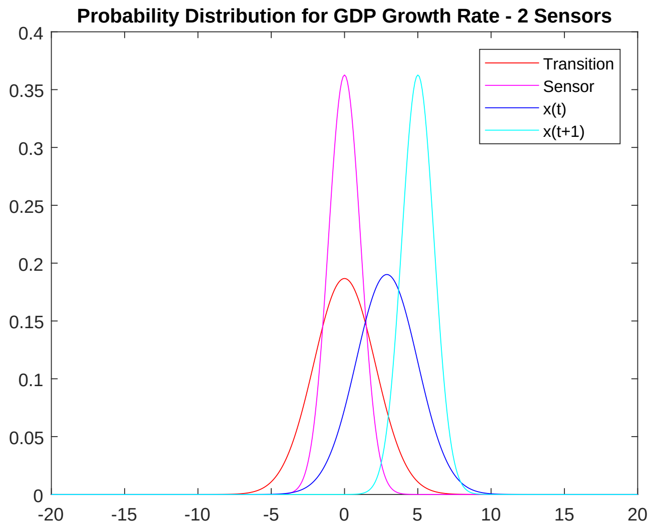


Figure 3: 2-Sensor Model GDP Growth Rate Probability Distribution

The Estimate Result Under Transition Model Only

GDP Estimate The estimate result of GDP under the transition model only is illustrated in figure 4.

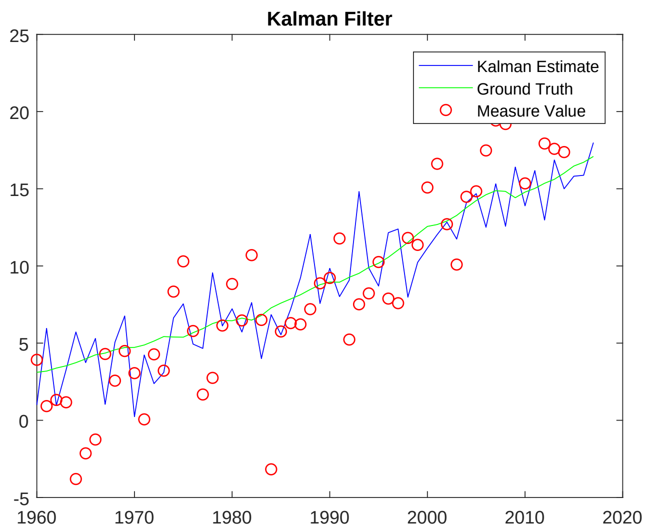


Figure 4: Kalman Estimate for GDP (Transition Model Only)

GDP Growth Rate Estimate The estimate result of GDP growth rate under the transition model only is illustrated in figure 5.

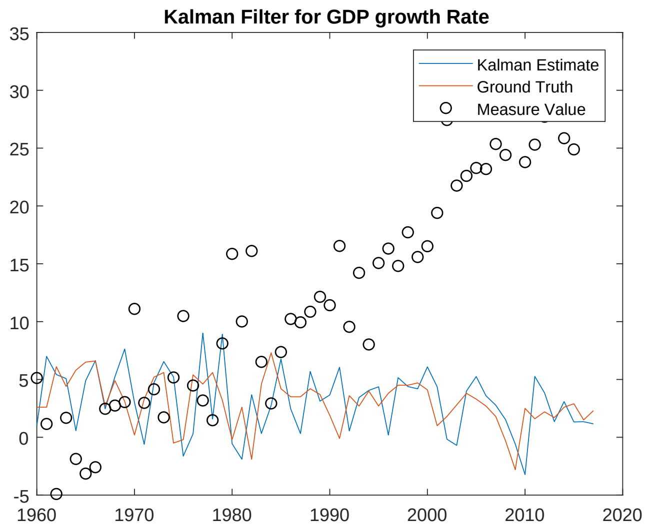


Figure 5: Kalman Estimate for GDP growth rate (Transition Model Only)

Estimate Result for Transition and Sensor Model

GDP Estimate The estimate result of GDP under the complete model is illustrated in figure 6.

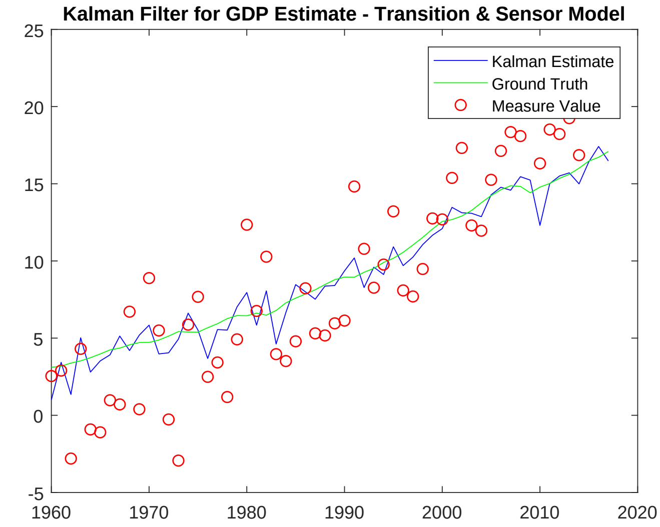


Figure 6: Kalman Estimate for GDP (Complete Model)

GDP Growth Estimate The estimate result of GDP under the complete model is illustrated in figure 7.

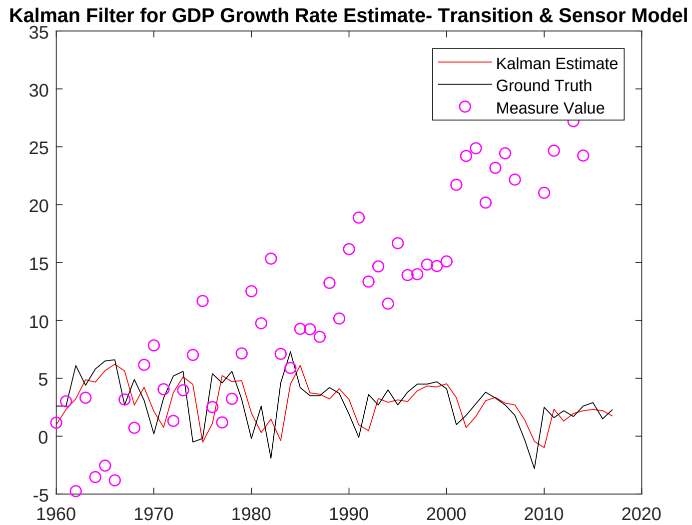


Figure 7: Kalman Estimate for GDP Growth Rate (Complete Model)

2.3.2 Extra Part: Additional Sensor for GDP growth rate

Probability Distribution The probability distribution for GDP is illustrated in 8.

The probability distribution for GDP growth rate is illustrated in 9.

Estimate Result for Transition Model

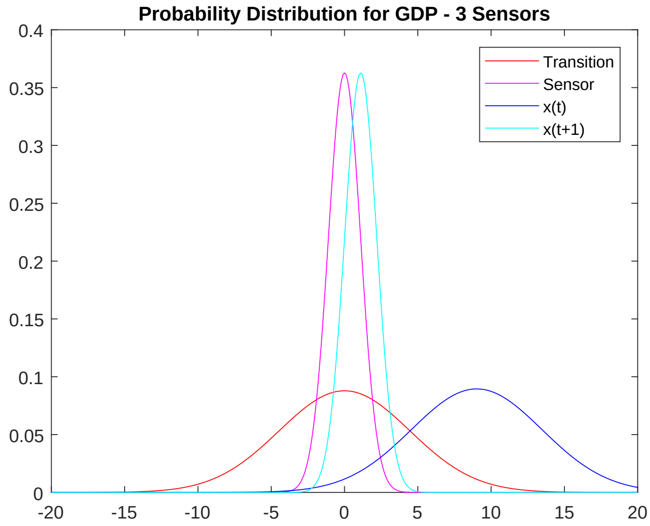


Figure 8: 3-Sensor Model GDP Probability Distribution

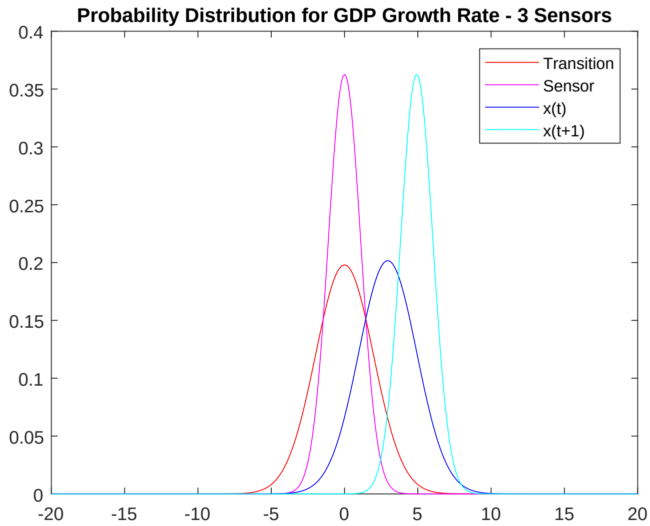


Figure 9: 3-Sensor Model GDP Growth Rate Probability Distribution

GDP Estimate The estimate result of GDP under the transition model is illustrated in figure 10.

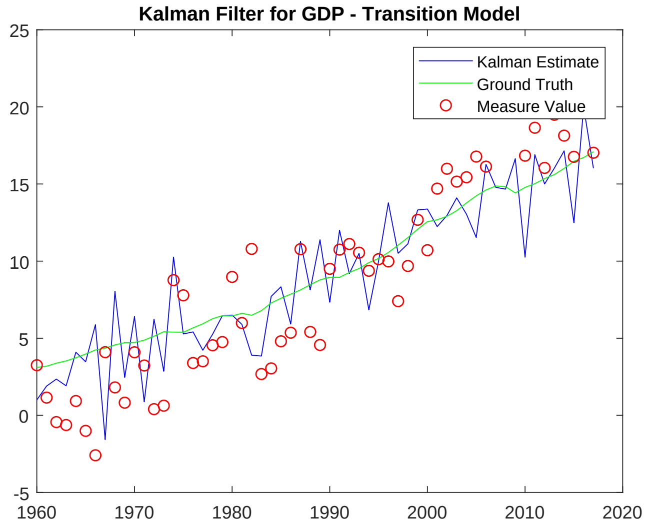


Figure 10: Kalman Estimate for GDP (Transition Model Only)

GDP Growth Rate Estimate The estimate result of GDP growth rate under the transition model is illustrated in figure 11.

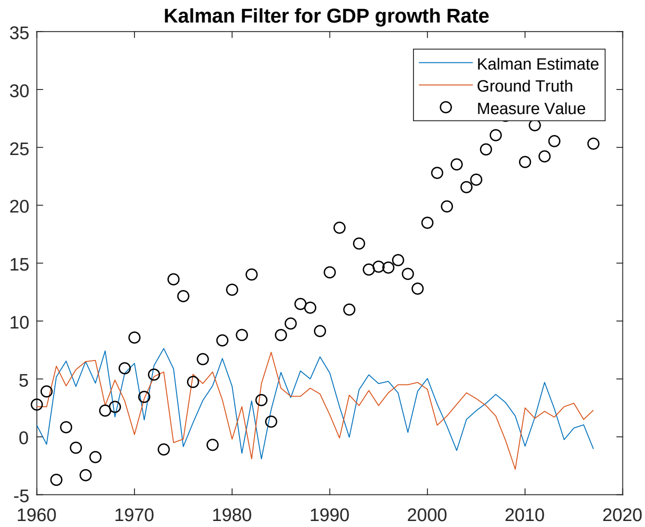


Figure 11: Kalman Estimate for GDP Growth Rate (Transition Model Only)

Estimate Result for The Complete Model

GDP Estimate The estimate result of GDP under the complete model is illustrated in figure 12.

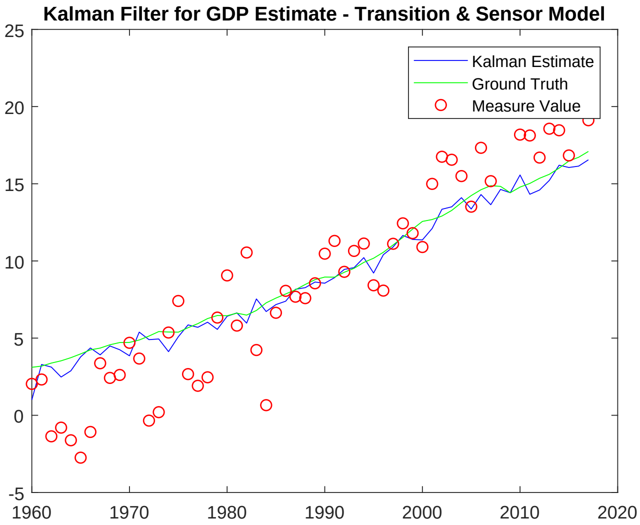


Figure 12: Kalman Estimate for GDP Growth Rate (Complete Model)

GDP Growth Rate Estimate The estimate result of GDP growth rate under the complete model is illustrated in figure 13.

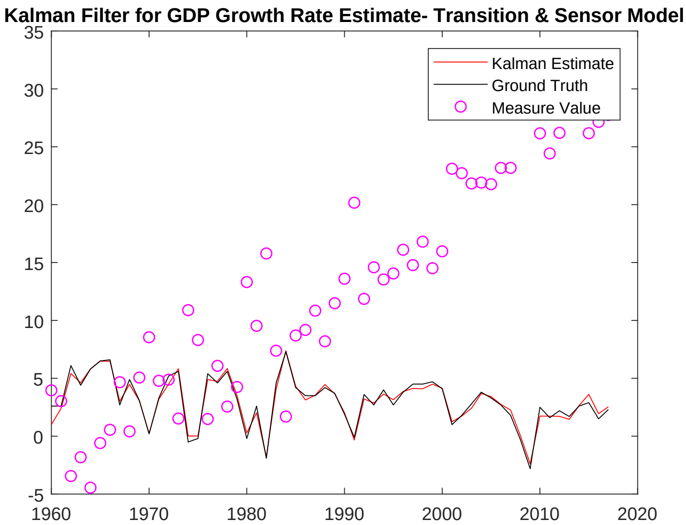


Figure 13: Kalman Estimate for GDP Growth Rate (Complete Model)

2.4 Discussion

In the result part, the Kalman estimate result for the sensor model with 2 sensors and 3 sensors are given.

In this part, the Pearson's coffeicient and the variance of the data to compare the fitting result for the 2 sensors and 3 sensors.

The Pearson's coefficient table. The higher value means that the better fitting result.

Table 2: Variance of Different Data

	Category	Transition Model	Complete Model
GDP	**GDP Ground Variance**	18.6672	18.6672
	2 Sensors Fitting Variance	22.5705	19.8861
	3 Sensors Fitting Variance	23.5823	18.6355
GDP Growth	**GDP growth Rate Variance**	4.3208	4.3208
	2 Sensors Fitting Variance	9.0279	3.5090
	3 Sensors Fitting Variance	7.1595	4.1567

Table 3: Pearson's Coefficient ρ

	Category	Transition Model	Whole Model
GDP	2 Sensor Model	0.8712	0.9745
	3 Sensor Model	0.8938	0.9932
GDP Growth Rate	2 Sensor Model	0.2433	0.5353
	3 Sensor Model	0.2480	0.9694

From table 2 and 3, the 3 sensor model illustrated a better fitting result, the variance deviation is smaller and the Pearson's coefficieint is higher.

The sensor model for the GDP growth rate improve the accuracy of the estimate of the GDP growth rate efficiently.

References

[1] W. B. Website, "United states gdp and gdp growth rate," https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.

[2] F. Trade, "Guide to foreign trade statistics," https://www.census.gov/foreign-trade/guide/sec2.html#bop.

[3] "United states economic historical data," https://www.whitehouse.gov/omb/historical-tables/.

[4] R. Faragher, "Understanding the basis of the kalman filter via a simple and intuitive derivation, ieee signal processing magazine," https://www.cl.cam.ac.uk/~rmf25/papers/Understanding%20the%20Basis%20of%20the%20Kalman%20Filter.pdf, Cambridge University, England, Cambridge.