# The Hierarchical Bayesian Model of Gold Production in Malaysia

#### Nora Muda\* and Norashikin Ramli

School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia; noramuda@ukm.edu.my, ikin2607@gmail.com

#### **Abstract**

Malaysia is a country which is rich in natural resources and one of it is gold. Gold has become an important national commodity. This study focused on gold production in Malaysia and also in 102 countries from 1995 to 2010. The Hierarchical Bayesian Model is selected to determine and explained the scenario of gold production in Malaysia. Through the analysis of hierarchical Bayesian model of the cumulative gold production in Malaysia and other countries for 15 years, the posterior mean of gold production in 102 countries are determined. Variance Partition Coefficient (VPC) is also calculated to quantify percentage of total variation in world gold production attributable to between countries variation. This study has found that the posterior mean for Malaysia and world gold production in hierarchical Bayesian are 3646 kg and 24 180 kg respectively. VPC reading obtained is 0.9226 which indicates that 93% of the total variations were contributed by variation between countries. This shows that the variation of gold production between countries is large and also the correlation of gold production within countries is high.

Keywords: Gold, Posterior Mean, Total Variation, Variance Partition Coefficient, World Gold Production

### 1. Introduction

The gold has become important these days as it gives substantial revenue to the country. It is reported that gold mining industry was the major contributor to the gross output of the mineral mining subsector in 2010. This industry contributed RM612.4 million or 40.1 percent to gross output¹. The gold mining in Malaysia is limited in the areas that is known as Central Belt. It stretches from highlands in Batu Melintang, Jeli, Kelantan up to Gunung Ledang in Johor. The majority of the gold production apparently come from the states of Pahang, mainly in the Penjom Gold Mine². Approximately 12 gold mines were operating in the country, all were located in the states of Kelantan, Pahang and Terengganu.

In 2010, the main gold producers in the world are China, Australia, USA, South Africa, Russia, Peru, Indonesia and Canada. China has become the largest gold producer in the world for the past three consecutive

years. Gold production in China rose for the eleventh year, even in other countries show a decline of gold production<sup>3</sup>. Gold production in Auatralia has also been encouraging for the last 30 years. Gold industry has become the third largest export commodity in Australia and thereby making the country the second largest gold producer in the world<sup>4</sup>.

The world gold production has already reached peak production in 2001. Lower production costs after the financial crisis in 2008 has increased the gold production in 2009 and 2010. The world gold production in 2008 and early 2009 showed a decline from the previous year. Most metal prices are fall and the world economic recession. This matter has caused mining companies face difficulties to continue operations. Therefore the rate of production had to be reduced<sup>5</sup>.

From previous study, it is found that the gold production is related to the gold price. Selvanathan and Selvanathan<sup>6</sup> have proven that gold price affected the gold production. While<sup>7</sup> shows that the gold price is

influenced by two macroeconomic factor which are oil price and inflation rate. However, Shafiee and Topal<sup>5</sup> proved that there is high correlation between gold price and oil price, while there is no significant relationship was found between gold price and inflation rate.

Malaysia are among the 30 largest countries which have gold. The country's gold reserves are said to be still retained by 106 000 tonnes based on studies by Anon8. In 2010, the production of commodities such as gold has risen by more than ten percent. Economic investment in the mining sector has seen tremendous growth over the 6.6% in 2010 compared to 3.3% in 2000<sup>2</sup>. These statistics clearly show minerals like gold has been beneficial to the country. Efforts to identify potential areas for mining activities are also underway so that the minerals can be utilized.

Based on the previous studies and the rapid activities of gold mining in Malaysia, therefore we able to see the variation of gold production in Malaysia and other countries. Therefore, the position of the Malaysia's gold production can be determined. Based on the hierarchical Bayesian approach, the level of the Malaysia's gold production and its position compared to other countries is known. It is able to help Malaysia in providing important information to achieve sustainable economic development.

# 2. Materials and Methods

The data for analysis are taken from Monthly Bulletin of Statistics Malaysia for the year of 1995 until 2010 and the book of World Mineral Production for the year 1995-2010. The data include the average gold production produced by 102 countries, including Malaysia.

# 2.1 Hierarchical Bayesian Analysis

Bayesian methods use a different approach compare to the classic method. In Bayesian, parameter is considered as a random variable and the probability is defined as the 'degree of belief'. Bayesian terms come from the use of Bayes's theorem, which was discovered by Reverend Thomas Bayes in the 18th century9. Bayes theorem can be explained as:

$$p(\theta|\mathbf{y}) = \frac{p(\theta, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{p(\mathbf{y})}$$

or can also be written in the form of proportions:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta) \pi(\theta)$$

which  $\pi(\theta)$  is a prior distribution,  $p(y|\theta)$  is a likelihood and  $p(\theta|\mathbf{y})$  is a posterior distribution. Prior distribution refers to beliefs about the parameters before the data are examined. Further, reliance on  $\theta$  will be updated to incorporate the information from the prior distribution

Hierarchical Bayesian ideas have become increasingly important in recent years. Hierarchical model which is also known as multi-level model is used when the information is available at different levels of observation units<sup>10</sup>.

and data through calculation of the posterior distribution.

Hierarchical model associated with parameters variations between groups when there is a model for these parameters. Occurrence of hierarchy is exist when the model parameters are located 'on' the model of data.

Based on the foregoing, all the bayesian model is hierarchy; when prior for located 'on' the model of 11. In general, the Bayesian hierarchical has shape

$$y_{j} \mid \theta_{j} \sim p(y_{j} \mid \theta_{j}) \text{ (model for data in grou } j = 1,...,J)$$

$$\theta_{j}|v\sim p\left(\theta_{j}|v\right)$$
 (model between group or prior for parameter  $\theta_{j}$  )

$$v \sim p(v)$$
 (prior for hyperparameter  $v$ )

by writing the hierarchy from the bottom up. Next, the all posterior density of parameters,  $\mathbf{\theta} = (\theta_1, ..., \theta_T, \nu)'$  and the marginal posterior density for

interested will be calculated.

To begin a hierarchical Bayesian analysis, the distribution of gold production in 102 countries for the years 1995-2010 should be identified in advance with the help of Easy Fit software.

For the prior distribution, the fuzzy prior or noninformative priors is used, since no information is known about the priors, ie by taking the large variance. Therefore, the normal and inverse gamma are used as a prior distribution. In overall, the Bayesian hierarchical model is described as follows:

$$y_{ij} | \theta_j, \sigma_e^2 \sim \text{Normal } (\theta_j, \sigma_e^2)$$
  
 $\theta_j | \mu, \sigma_j^2 \sim \text{Normal } (\phi, \sigma_j^2)$ 

 $\phi \sim N_{ormal} (0,100000)$  $\sigma_i^2 \sim \text{Inverse Gamma } (0.01, 0.01)$  $\sigma_e^2 \sim \text{Inverse Gamma (0.01,0.01)}$ for i = 1, 2, ..., 1632 and j = 1, 2, ..., 102. Variance components  $\sigma_i^2$  (variance of the random effects)

measuring the variability between countries while  $\sigma_a^2$ measure variability of within-country.

Then the parameters of the model are collected in the vector. Hierarchical  $\mathbf{\omega} = (\theta_1, \dots, \theta_{102}, \varphi, \sigma_0^2, \sigma_1^2)$ 

structure of the model allows for density pior factored as follows:

$$\begin{split} \mathbf{p}\left(\boldsymbol{\omega}\right) &= \mathbf{p}\left(\boldsymbol{\theta}_{1},...,\boldsymbol{\theta}_{102},\boldsymbol{\varphi},\boldsymbol{\sigma}_{e}^{2},\boldsymbol{\sigma}_{j}^{2}\right) \\ &= \mathbf{p}\left(\boldsymbol{\theta}_{1},...,\boldsymbol{\theta}_{102},\boldsymbol{\varphi},\boldsymbol{\sigma}_{j}^{2}\right) \mathbf{p}\left(\boldsymbol{\varphi}\right) \mathbf{p}\left(\boldsymbol{\sigma}_{e}^{2}\right) \mathbf{p}\left(\boldsymbol{\sigma}_{j}^{2}\right) \\ &= \prod_{j=1}^{102} \mathbf{p}\left(\boldsymbol{\theta}_{j}\left|\boldsymbol{\varphi},\boldsymbol{\sigma}_{j}^{2}\right) \mathbf{p}\left(\boldsymbol{\varphi}\right) \mathbf{p}\left(\boldsymbol{\sigma}_{e}^{2}\right) \mathbf{p}\left(\boldsymbol{\sigma}_{j}^{2}\right) \end{split}$$

Model is fitted by performed the Markov Chain Monte Carlo (MCMC) which implemented in WinBUGS software. A total of 20 000 simulations were done with the first fuel (burn-in) is 10 000. The MCMC goal is to achieve convergence. Therefore, the sample must be generated to reach the equilibrium distribution. The equilibrium distribution is known as the stationary distribution or the target distribution for the MCMC algorithm. The MCMC convergence refers to situations in which the algorithm has reached equilibrium and generate value from the desired target distribution. To determine whether convergence has been achieved or not, some diagnostic tests should be carried out.

#### 2.2 Diagnostic Tests

The density of the Markov chain can be examined visually with some plot, namely trace plot ( $\varphi$ ), autocorrelation plot  $(\sigma_{\varrho}^{2})$  and quantile plot,  $(\sigma_{i}^{2})$ . The Markov Chain

error can also be seen, to reflect the posterior density is estimated accurately based on the MC error which is smaller than the standard error.

In the hierarchical model or a multi-level model, residual variation in the response variable is separated into components according to different levels12. Therefore, to measure the percentage of the total variation, partition coefficients of variance (Variance Partition Coefficient, VPC) should be obtained. The VPC equation is shown as:

$$VPC = \frac{\sigma_j^2}{\sigma_e^2 + \sigma_j^2}$$

which  $\sigma_a^2$  is the variance of level 1, the variance of the

likelihood and  $\sigma_j^2$  is the variance of level 2, the variance of the random effects. Then the VPC posterior samples are obtained for point and interval estimation.

## 3. Results and Discussions

Figure 1 shows of gold production in Malaysia for the years 1995 to 2010. Based on the graph, it appears that the total gold production was 64.43 tons. The gold production in Malaysia has already reached its peak around the year 2003, where the cumulative gold production for that year was 37.91 tons or 34 387 kg of 58% of total production.

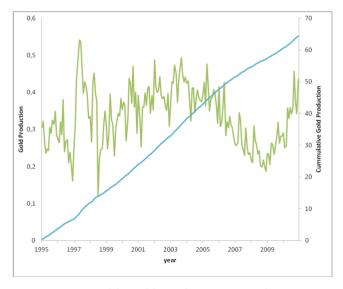


Figure 1. Monthly gold production in Malaysia 1995-2010.

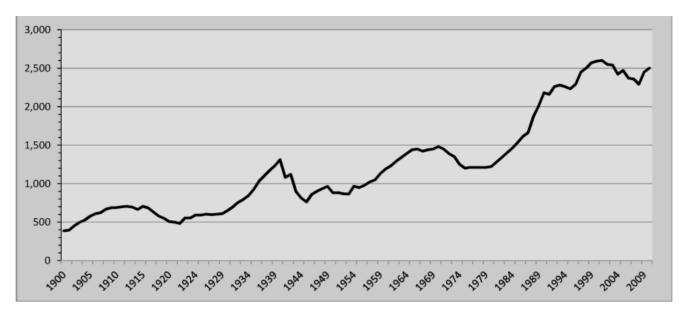


Figure 2. World Gold Production 1900-2010.

Meanwhile, the world gold production for a period of 110 years, ie from 1900 to 2010 is shown in Figure 2. During this period, the gold production is seen rising. In 1900, a total of 386 tonnes of gold were produced. Of these only 15% of gold production in 2010 which is 2,500 tons. The year 2001 recorded the highest gold production of 2600 tons. Overall, the world gold production from the year 1900-2010 has reached production below 141 000 tonnes.

Based on the findings from Easyfit software in order to determine the distribution of gold production in 102 countries, it has found that Normal distribution has given the best fitted distribution among the other distributions. To prove and validate the results, the Kolmogorov-Smirnov test is done to validate that the gold production for 102 countries for the year 1995-2010 are follow Normal distribution with the hypothesis:

H0: Data follows Normal distribution HA: Data is not follow Normal distribution

It has found that  $\rm H_0$  are failed to reject at all significant level in most countries. Therefore there are strong evidence that gold production in 102 countries are follow Normal distribution.

Table 1 shows the estimated mean  $E(\theta_j | y_{ij})$  and standard deviation of gold production in all countries. Based on these, the highest mean was recorded by South Africa with a mean production was 359 500 kg, followed by

United States and Australia. While the lowest mean was recorded by Taiwan with a mean production of only about 20.6 kg. The mean of gold production in Malaysia is 3646 kg, after taking into account the variations between each country's contribution. Malaysia was ranked at 46th out of 102 countries in gold production. From the original data, the average gold production in Malaysia is 3641 kg. Due to the informations from other countries which are allowed in hierarchical Bayesian model, the average gold production of Malaysia has increased to 3646 kg. Meanwhile, the estimated mean of world total gold production,  $\hat{\mu}$  was 24 180 kg per annum with refer to Table 2. Table 3. shows a summary of the posterior's parameters

for  $\varphi$ ,  $\sigma_e^2$ ,  $\sigma_j^2$  and . From the table, the error of Monte Carlo (MC) for the three parameters are smaller than the posterior standard deviation. Thus, there is strong evidence that the mean posterior was estimated with high precision. For the components of variance, the variability

between countries,  $\sigma_j^2$  is large compared to the variability within-country,  $\sigma_e^2$ . For a more accurate explanation, the variability among countries is twelve times that of the variability in countries when the each mean posterior are

# 3.1 Diagnostic Tests

compared.

Figure 3 are trace plots for the generated values. From the figure, there is no pattern or irregular shape seen in

Table 1. Mean and Standard Deviation of  $\boldsymbol{\theta}_j$ 

Country	Mean	Std Dev	Country	Mean	Std Dev
Armenia 1024		4611	Zimbabwe	16470	4592
Azerbaijan	1187	12700	Kanada	135500	4611
Bulgaria	2681	4573	Costa Rica	529.3	4590
Finland	2112	4567	Cuba	867.4	10580
Perancis	3559	6171	Dominican Republic	1510	6451
Georgia	1927	4525	Guatemala	5483	5492
Greenland	1775	8194	Honduras	2290	4746
Itali	618	6974	Jamaica	214.8	9162
Poland	643.3	4579	Mexico	32800	4608
Romania	2087	4579	Nicaragua	3304	4596
Rusia	154600	4567	Panama	1054	6456
Serbia	602.7	6057	USA	290400	4596
Slovakia	243.3	4619	Argentina	29610	4611
Spanyol	4196	5484	Bolivia	10390	4606
Sweden	5701	4563	Brazil	49040	4603
Turki	5325	4545	Chile	43420	4582
United Kingdom	137	9148	Colombia	29910	4600
Yugoslavia	2333	7461	Ecuador	4232	4613
Algeria	603.8	6156	French Guiana	2786	4583
Botswana	1509	5762	Guyana	10790	4576
Burkina Faso	3769	4739	Peru	143400	4589
Burundi	2103	4590	Suriname	10500	6905
Cameroon	721.5	4596	Uruguay	2288	5837
Chad	164	8147	Venezuela	10080	4566
Congo	3585	4720	Burma	126.1	6090
Mesir	4403	17720	China	216600	4560
Equatorial Guinea	137	6944	India	2637	4591
Eritrea	176.4	5573	Indonesia	113500	4628
Ethiopia	3294	4771	Iran	577.6	4602
Gabon	197.5	5813	Jepun	8388	4563
Ghana	71150	4595	Kazakhstan	16460	4573
Guinea	14470	4586	Korea(Rep Of)	362.7	4599
Ivory Coast	2908	4573	Korea, Dem. P.R.	1790	6505
Kenya	975.6	4570	Kyrgyzstan	16230	4577
Liberia	629.7	5070	Laos	5063	6456
Madagascar	424.4	6507	Malaysia	3646	4595
Mali	34840	4615	Mongolia	12550	4591
Mauritania	3698	6894	Myanmar	120.7	6937
Morocco	946	4579	Oman	365	4567
Mozambique	169.2	6509	Filipina	34820	4573
Namibia	2369	4585	Arab Saudi	5660	4587
Niger	1740	5500	Taiwan	20.6	5788
Nigeria	1170	6909	Tajikistan	2157	4588
Senegal	1609	5817	Thailand	3824	5810
Sierra Leone	71.86	6911	Uzbekistan	79510	4598
Afrika Selatan	359500	4584	Vietnam	2404	4757
Sudan	4252	4618	Australia	267700	4575
		+			
Tanzania	26950	4642	Fiji Novy Zooland	2994	4589
Togo	9774	7460	New Zealand	10590	4594
Uganda	8079	4612	Papua New Guinea	62850	4599
Zambia	772.1	4730	Solomon Islands	631.8	5293

**Table 2.** Summary of the posterior  $\mu$ 

Node	Maan	Median	Std Dev	MC error	Percentile	
	Mean				2.5%	97.5%
$\mu$	24180	24180	577.7	4.173	23040	25310

Table 3. Summary of posterior for parameter in the model

Node	Mean	Madian	Std Dev	MC	Percentile	
		Median		error	2.5%	
$\varphi$	557	552	991	7.227	-1375	2508
$\sigma_e^{2}$	3.39e <sup>8</sup>	3.39e <sup>8</sup>	1.39e <sup>7</sup>	111800	3.13e <sup>8</sup>	3.67e <sup>8</sup>
$\sigma_{j}^{2}$	4.37e <sup>9</sup>	4.31e <sup>9</sup>	6.28e <sup>8</sup>	4.22e <sup>6</sup>	3.30e <sup>9</sup>	5.76e <sup>9</sup>

Table 4. Summary of posterior VPC

Node	Mean	Median	Std Dev	MC error	Percentile	
	Mean			MC error	2.5%	97.5%
VPC	0.9268	0.9271	0.009955	6.736×10 <sup>-5</sup>	0.9061	0.945

 $\varphi$  ,  $\sigma_e^2$  and  $\sigma_j^2$  . Therefore, convergence has been achieved. Figure 4. is a plot of autocorrelation. Based on this figure, the autocorrelation of the three parameters is low. This shows that convergence has been obtained on the low number of iterations. Quantile plot is shown in Figure 5. The plot shows that the selected quantile was balanced and this shows that the algorithms have been converging.

The trace plots, autocorrelation and quantile for param-

eter  $\theta_j$  also shows the convergence was achieved in all countries. Besides the plots, the errors of Markov Chain (MC) also clearly shows that the posterior density is estimated accurately based on the MC error which is smaller than the standard error. Based on the evidence obtained, it is obviously that Markov chain has converges to the target distribution, ie in this case the posterior distribution.

From the Figure 6. and Table 4., it found that the mean posterior is 0.9268. It means almost 93% of the total variation in world gold production were contributed by variations between countries. According to Ntzoufras<sup>13</sup>, the VPC value which is close to '1' means the correlation of gold production in the country is high.

This shows that the random effects has improved the model and further demonstrate the importance of using a hierarchical model.

In the Bayesian hierarchical model analysis, information on the posterior distribution according to the gold production in 102 countries have been obtained. In Malaysia, the mean of gold production is 3646 kg, while the mean of total world gold production was 24 180 kg. Based on the analysis of variations in gold production, it is found that 93% of the total variation was contributed by the variations between countries. Correlation was found to be high in the country of Malaysia. This shows that there has been improvement in the model and the selection of using hierarchical Bayesian model in this analysis is accurate.

## 4. Conclusion

In Hierarchical Bayesian analysis, we concentrated on looking at the posterior distribution. Based on the Kolmogorov-Smirnov test, we have found that the gold production in 102 countries are follow Normal distribution. Since we don't have any information on the prior

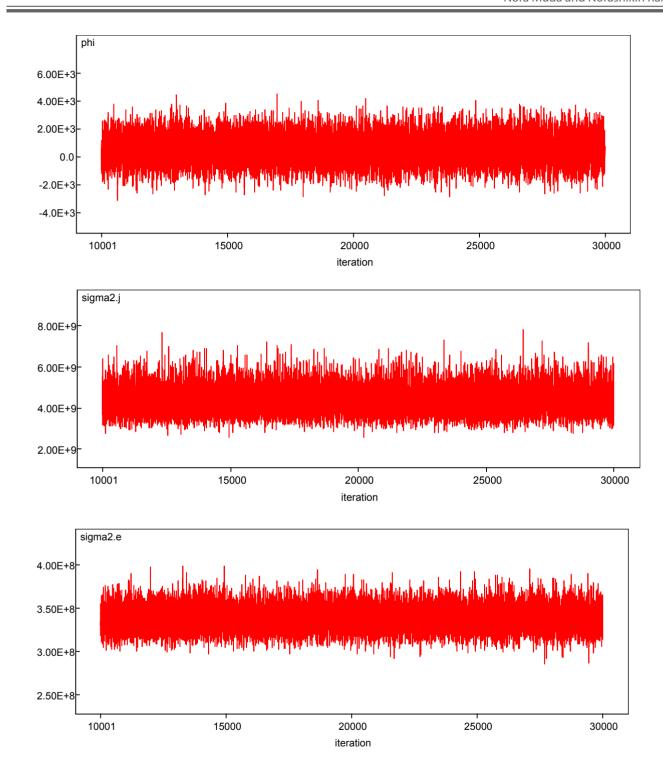
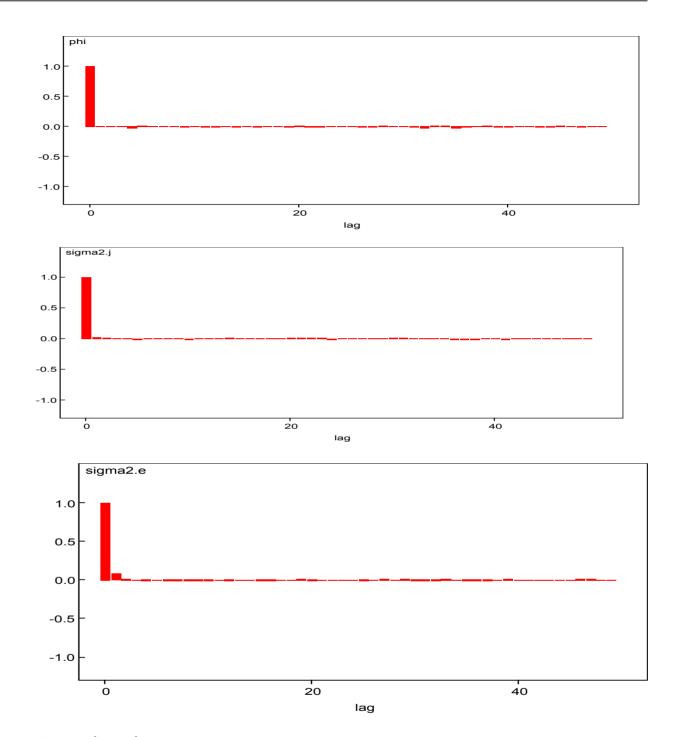


Figure 3. Trace Plot.

distribution, therefore we use an uninformative prior in the analysis and using large variance. For the prior distribution, mean of prior is follow Normal distribution and variance following inverse-Gamma distribution. The results showed that the posterior distribution of all parameters are estimated and obtained by using 20,000 samples. The Monte Carlo error for all parameters are smaller than the standard deviation of their posteriors.



**Figure 4.** Autocorrelation plot.

This shows that the mean posterior is estimated by high precision. To prove MCMC convergence has been achieved, a number of diagnostic tests were performed. The diagnostic tests are based on trace plots, autocorrelation and quantile plot. Based on the plots, it found that the Markov chain has converges to the target distribution; the posterior distribution.

From this analysis, the estimated mean of gold production in Malaysia for each year is around 3646 kg, while the mean of total world gold production is 24,180 kg. In addition, the partition coefficients of variance (VPC) is also calculated in the Hierarchical Bayesian model. The VPC which obtained can help researchers in identifying the percentage of variation within and between the amount of variation.

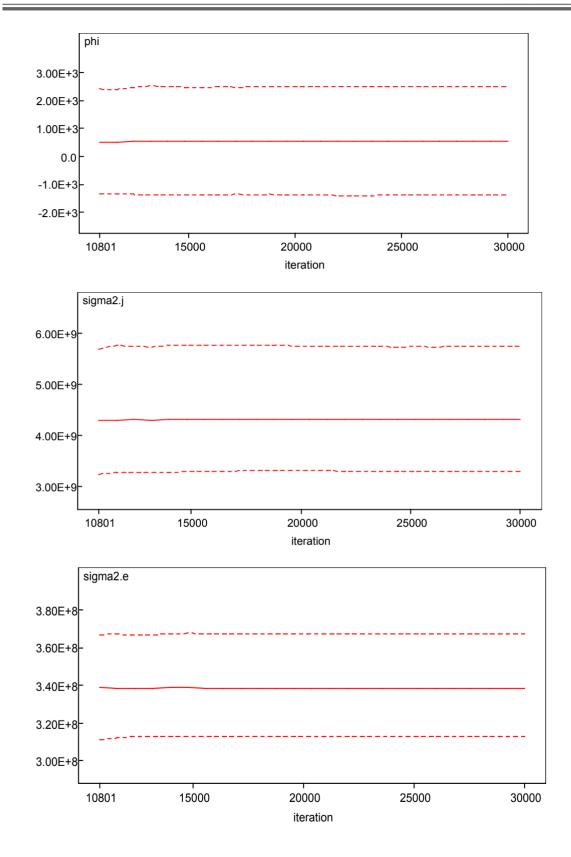
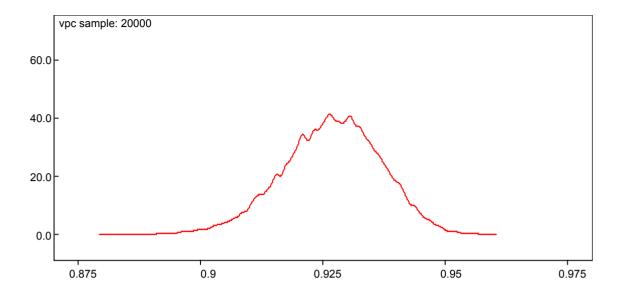


Figure 5. Quantile plot.



**Figure 6.** Posterior distribution of VPC.

The mean posterior which obtained from the VPC samples is 0.9268. It means 93% of the total variations are contributed by the variations between countries. It also showed that the correlation in the country is high as its value is close to one. The use of hierarchical bayesian models in the study of world gold production is very sensible and right as of the study are clearly shows that the random effects have been able to improve the model and therefore prove the effectiveness of this model.

# 5. Acknowledgement

This research was supported by the Research Grant DPP-2014-008.

## 6. References

- Department of Statistics Malaysia. Banci Ekonomi 2011. Perlombongan dan Pengkuarian, Putrajaya: Jabatan Perangkaan Malaysia; 2012.
- Tse P. The mineral industry of Malaysia 2010. Minerals Yearbook, Virginia, US: Geological Department; 2012
- Chaize T. World Production of gold. 2011 http://news. goldseek.com/Dani/1309290922.php, 2011.

- Huleatt MB, Jaques AL. Australian Gold Exploration 1976-2003. Resources Policy 2005; 30: 29-37.
- Shafiee S, Topal E. An Overview of Global Gold Market and Gold Price Forecasting. Resources Policy 2010; 35: 178-89.
- Selvanathan S, Selvanathan EA. The effect of the Price of Gold on its Production: a Time Series Analysis. Resources Policy 1999; 25: 265-75.
- Tully E, Lucy BM. A power GARCH examination of the gold market. Research in International Business and Finance 2007; 21(2): 316-25.
- Anon. Bumi Kelantan kaya sumber mineral. Berita PKINK 2012; 510.
- SAS Institute Inc. SAS/STAT 9.2 User's Guide, Cary: SAS Institute Inc; 2008.
- 10. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis, Chapman & Hall: Boca Raton; 2004.
- 11. Jackman S. Bayesian Analysis for the Social Sciences, John Wiley & Sons, Inc.: New Jersey; 2008.
- 12. Goldstein H, Browne W, Rasbash J. Partitioning Variation in Multilevel Models. Understanding Statistics 2002; 1(4): 22331.
- 13. Ntzoufras I. Bayesian Modeling Using WinBUGS. Wiley Series in Computational Statistics: Hoboken, USA; 2009.