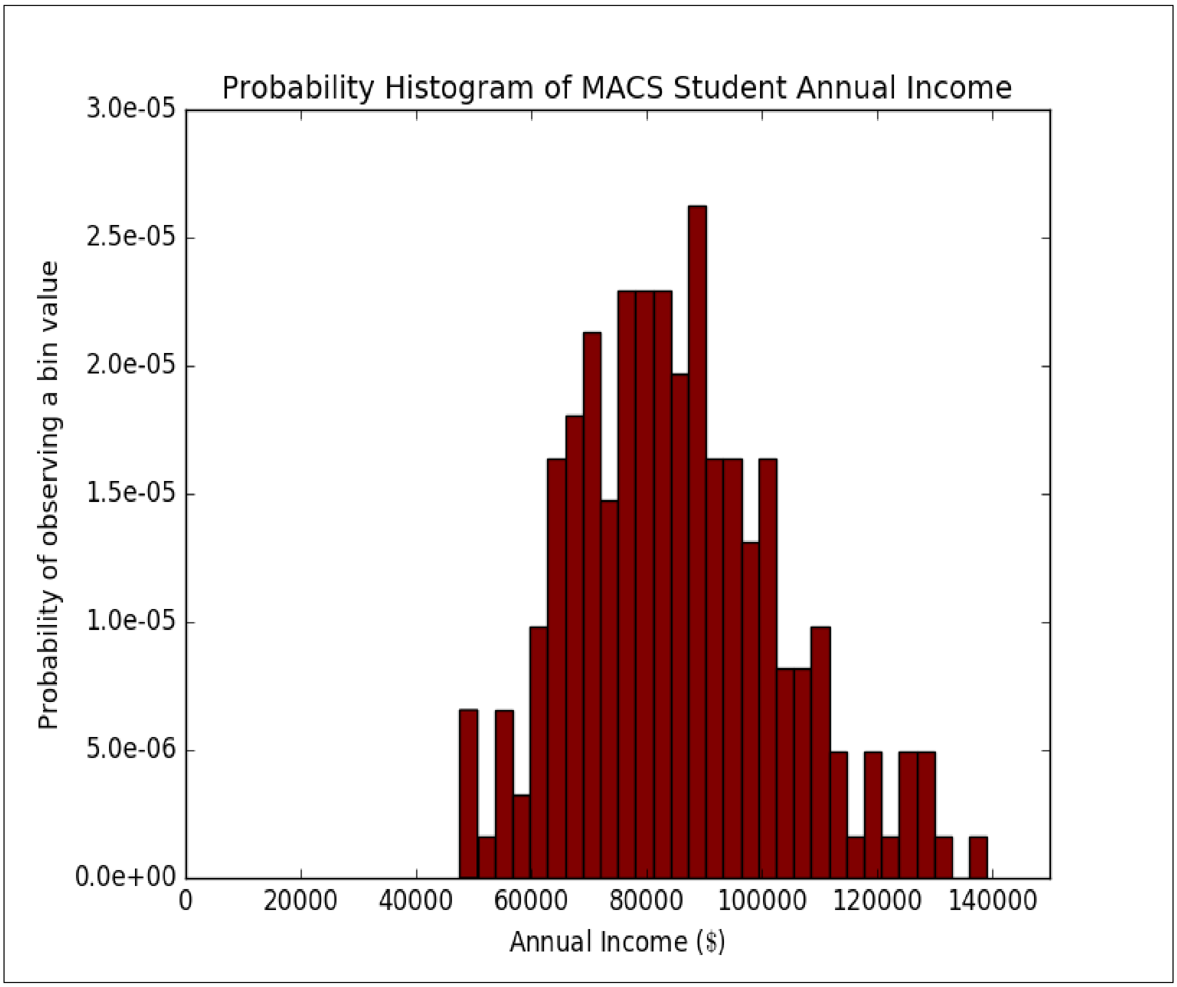
Problem Set #3

MACS 30100, Dr. Evans
Dongping Zhang

Problem 1: **Using Generalized Method of Moments (GMM) technique to estimate the parameters of a lognormal distribution that is supposedly resembleing the actual distribution of UChicago MACSS students' future incomes.**

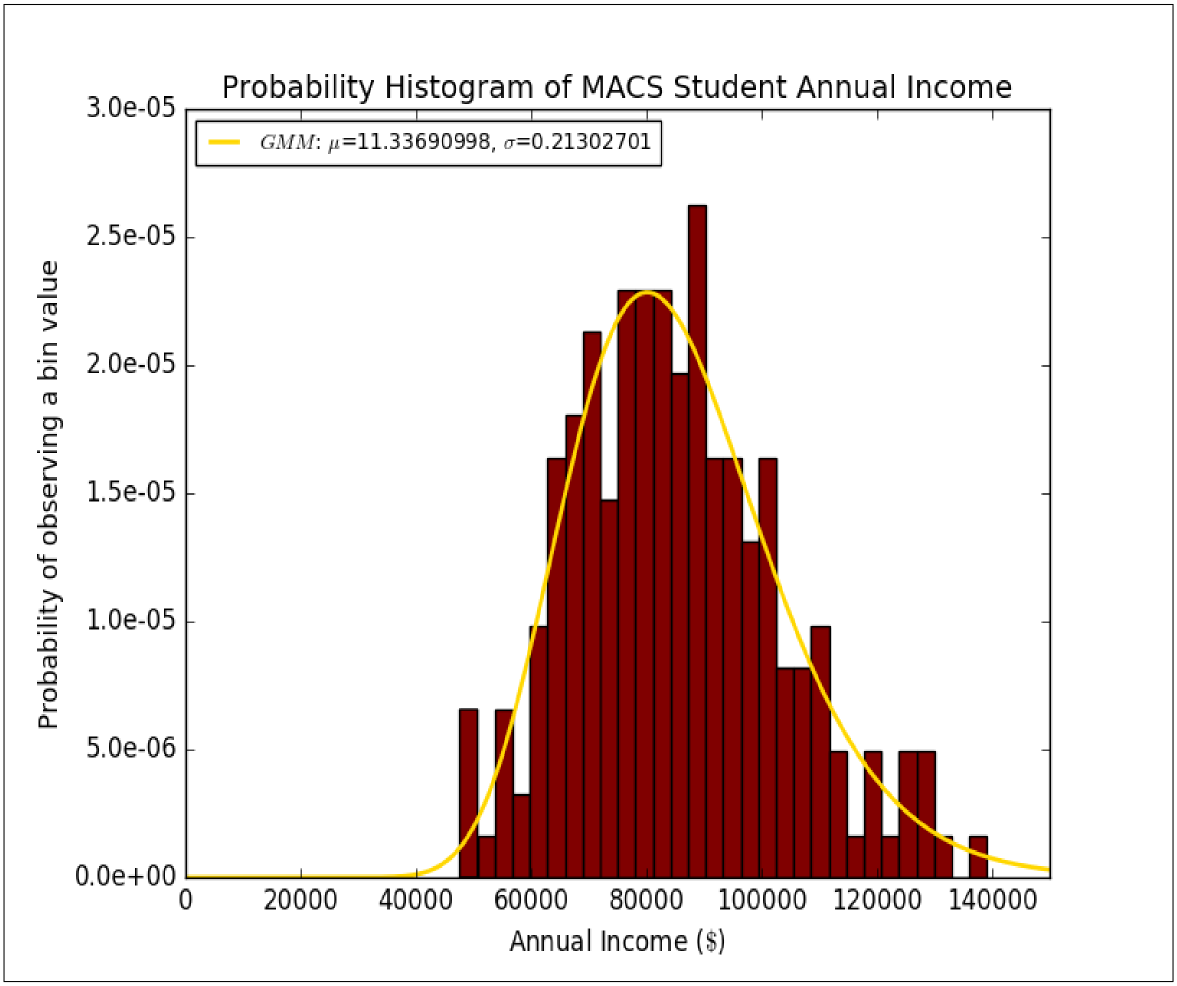
Part (a). Plot a histogram of percentages using *income.txt* data with 30 bins and *normed = True*.

Figure 1: Probability Histogram of MACSS Students' Annual Incomes



Part (b). Estimate the parameters of the lognormal distribution by GMM, using average income and standard deviation of income as two moments. Plot the estimated lognormal PDF against the histogram from part (a). Report the value of your GMM criterion function at the estimated parameter values. Report and compare your two data moments against your two model moments at the estimated parameter values.

Figure 2: Estimated lognorm PDF against the income histogram



The estimated parameters of the lognormal distribution by GMM are:

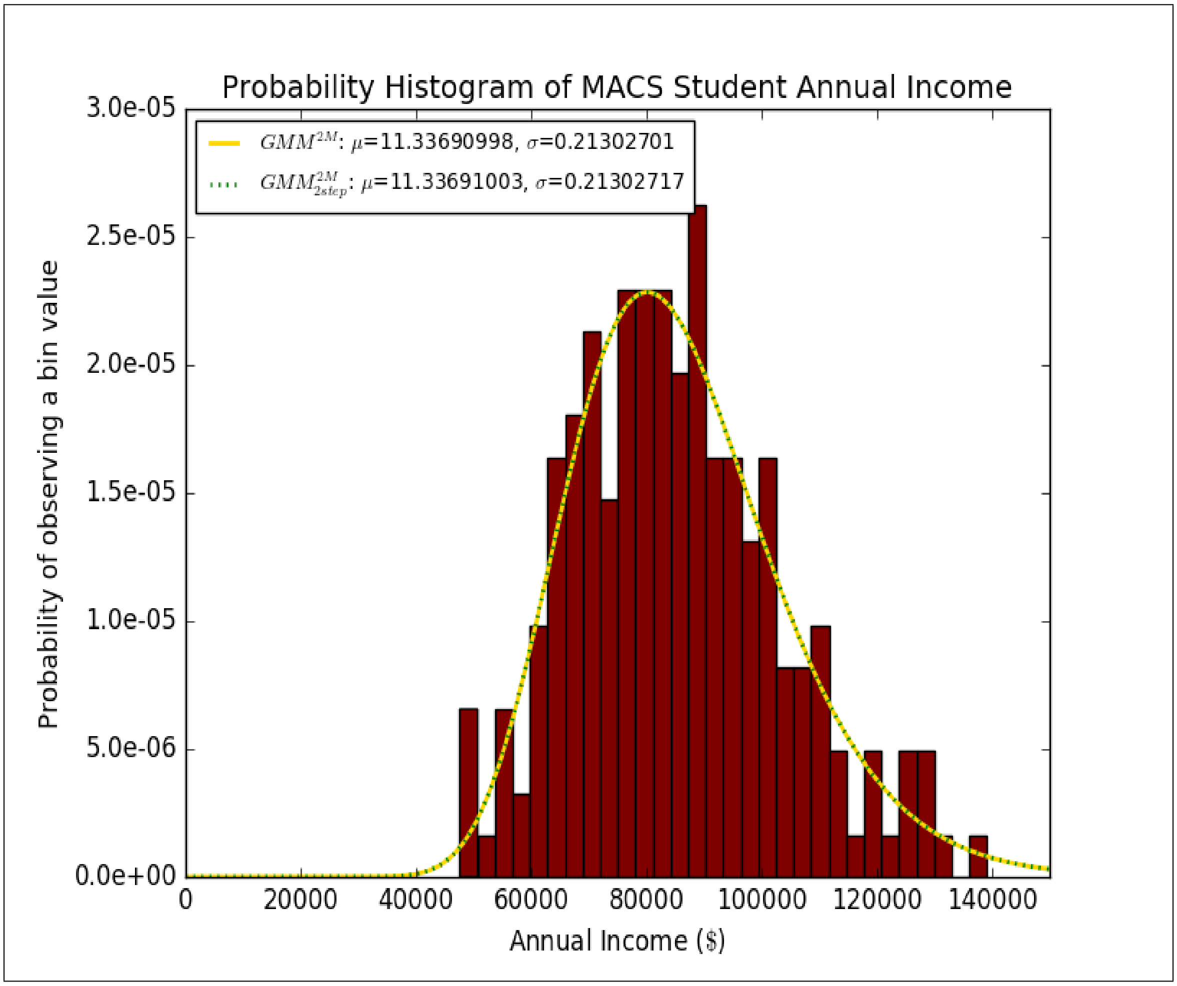
- $\hat{\mu}_{2M}^{GMM}$ = **11.3369099824**
- $\hat{\sigma}_{2M}^{GMM}$ = **0.213027011191**
- Value of GMM criterion function at using GMM parameters obtained is: $\mathbf{3.93983088 \times 10^{-13}}$
- Data Moments vs. Model Moments: from the table below, we are able to observe that the data moments and model moments are so similar and that

only the decimal points vary. Thus we can claim that our estimation procedure is a success.

	μ	σ
data moments	85276.823606258113	17992.542128046523
model moments	85276.79520501554	17992.532555365196

Part (c). Perform a two-step GMM estimator by using estimates from part (b) with two moments to generate an estimator for the variance-covariance matrix, which then use to get the two-step estimator for the optimal weighting matrix.

Figure 3: GMM^{2M} vs. GMM^{2M}_{2Step}



The estimated parameters of the lognormal distribution by GMM are:

- $\hat{\mu}^{GMM}_{2M2S}$ = **11.3369099824**
- $\hat{\sigma}^{GMM}_{2M2S}$ = **0.213027011191**

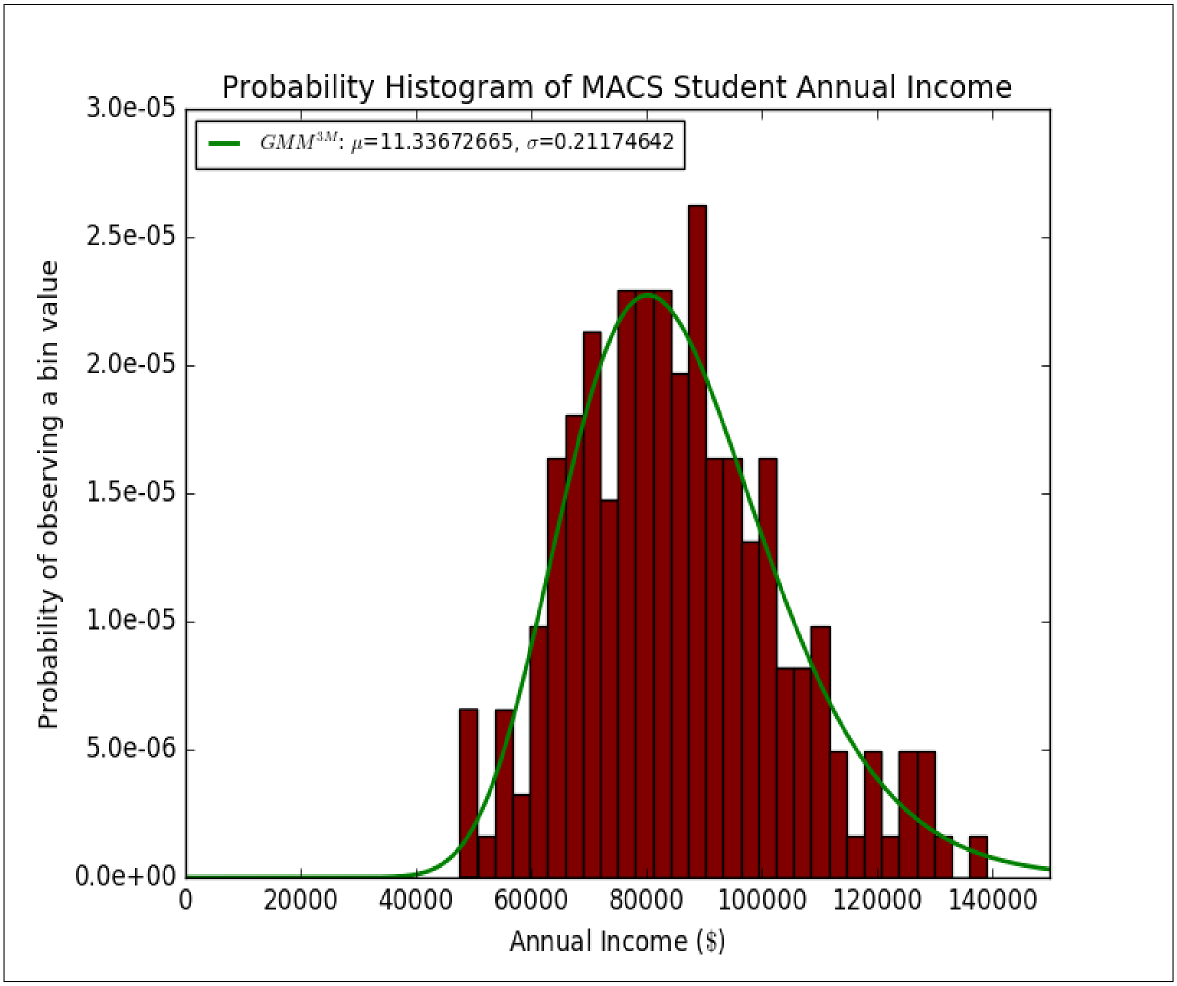
- Value of GMM criterion function at using GMM parameters obtained is: **0.05659968**

- Data Moments vs. Model Moments: from the table below, we are able to observe that the data moments and model moments are so similar and that only the decimal points vary. Thus we can claim that our estimation procedure is a success.

	μ	σ
data moments	85276.823606258113	17992.542128046523
model moments	85276.79890729053	17992.545166067979

Part (d). Estimate the lognormal PDF to fit the data by GMM using different moments. Use percent of individuals who earn less than \$75,000, percent of individuals who earn between \$75,000 and \$100,000, and percent of individuals who earn more than \$100,000 as your three moments. Use the identity matrix as your estimator for the optimal weighting matrix.

Figure 4: GMM using 3 percentile moments



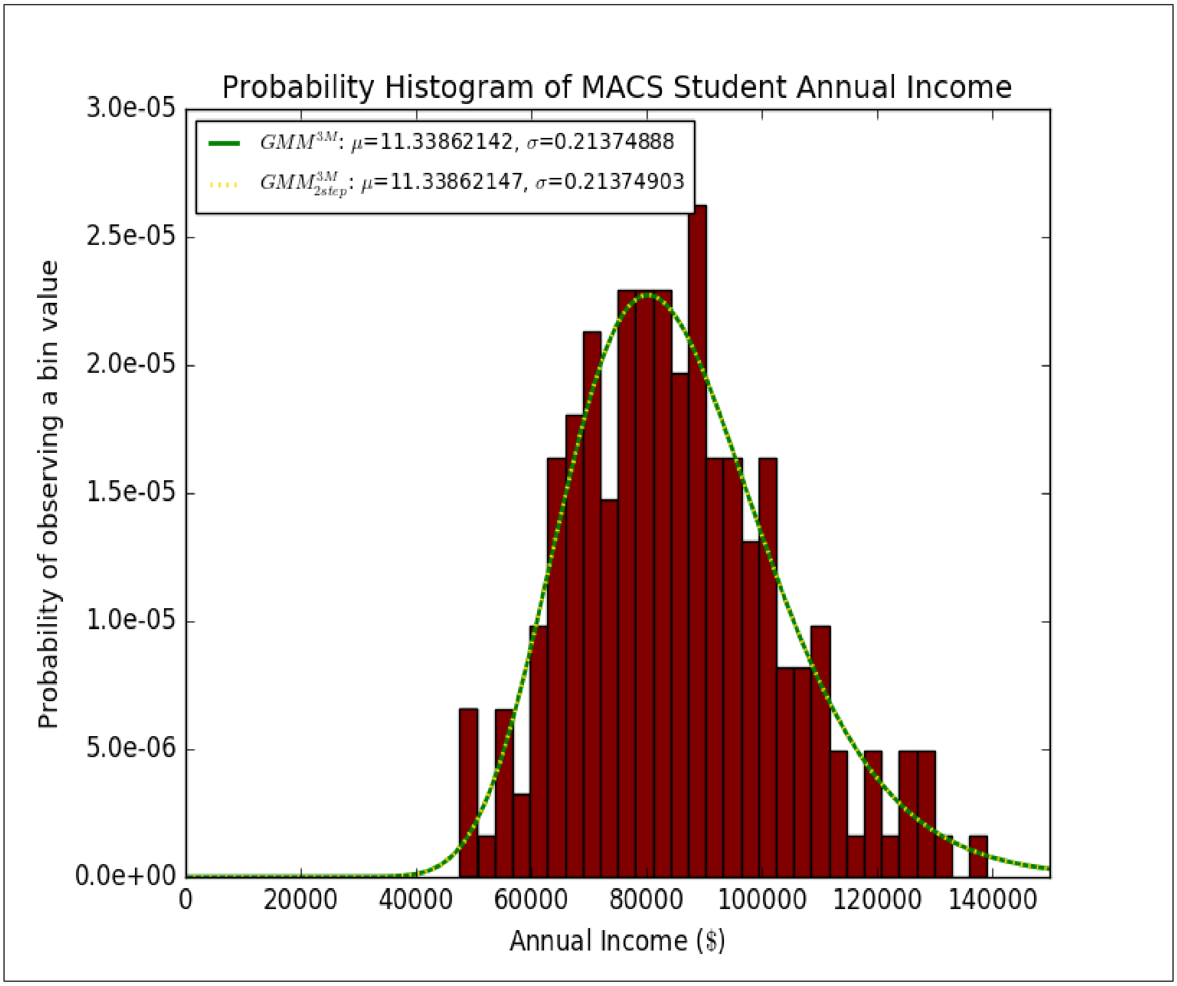
The estimated parameters of the lognormal distribution by GMM are:

- $\hat{\mu}_{3M}^{GMM}$ = **11.3386214217**
- $\hat{\sigma}_{3M}^{GMM}$ = **0.213748881097**
- Value of GMM criterion function at using GMM parameters obtained is: **1.87974871**
- Data Moments vs. Model Moments: from the table below, we are able to observe that the data moments and model moments are so similar and that only the decimal points vary. Thus we can claim that our estimation procedure is a success.

	$\Pr(x < \$75,000)$	$\Pr(\$75,000 \leq x < \$100,000)$	$\Pr(x \geq 100,000)$
data moments	0.3	0.5	0.2
model moments	0.2979081959482765	0.4946877876432526	0.19897086052259172

Part (e). Perform the two-step GMM estimator by using the estimates from part (d) with three moments to generate an estimator for the variance covariance matrix, which you then use to get the two-step estimator for the optimal weighting matrix.

Figure 5: GMM_{2Step} using 3 percentile moments



The estimated parameters of the lognormal distribution by GMM are:

- $\hat{\mu}^{GMM}_{3M2S}$ = **11.3386214746**
- $\hat{\sigma}^{GMM}_{3M2S}$ = **0.213749032771**
- Value of GMM criterion function at using GMM parameters obtained is: **199.99999993**

- Data Moments vs. Model Moments: from the table below, we are able to observe that the data moments and model moments are so similar and that only the decimal points vary. Thus we can claim that our estimation procedure is a success.

	$\Pr(x < \$75,000)$	$\Pr(\$75,000 \leq x < \$100,000)$	$\Pr(x \geq 100,000)$
data moments	0.3	0.5	0.2
model moments	0.29790824055180337	0.4946875066281751	0.19897105231742282

Part (f). Which of the four estimations from parts (b), (c), (d), and (e) fits the data best? Justify your answer.

In problem 1, we use GMM to estimate two parameters, μ and σ, for the lognormal distribution. Both part (b) and part (c) are exactly identified because we have exactly as many moments as parameters to be estimated while both parts (d) and parts (e) are over-identified because we have more moments than parameters to be estimated. In GMM estimation, it is supposedly better to have the case of over-identified model, R > K, because not all moments are orthogonal. That is, some moments convey roughly the same information about the data and, therefore, do not separately identify any extra parameters. So a good GMM model often is over-identified R > K. Thus, GMM estimations from part (d) and part (e) should be better than GMM estimations from part (b) and part (c).

In addition, because the GMM estimators from part (e) are obtained through a two-step procedure using the inverse of the variance-covariance matrix as the weighting matrix. Thus, it down-weights moments that have a high variance, and weight more heavily the moments that are generated more precisely. Thus, estimations obtained in part (e) is supposedly to be better than part (d).

In conclusion, **the parameter estimations from part (e) fit the data the best**, which is also proved by the values obtained from the criterion function using corresponding GMM estimators reported above.

Problem 2: Using Generalized Method of Moments (GMM) technique to estimate the coefficients of a linear regression.

The proposed sickness model is:

$$sick_i = \beta_0 + \beta_1 age_i + \beta_2 children_i + \beta_3 temp\_winter_i + \varepsilon_i \tag{1}$$

Part (a). Estimate the parameters of the model $(\beta_0, \beta_1, \beta_2, \beta_3)$ by GMM by solving the minimization problem of the GMM criterion function. Use the identity matrix as the estimator for the optimal weighting matrix.

- $\hat{\beta}_{0GMM}$ = **0.251645437572**
- $\hat{\beta}_{1GMM}$ = **0.0129335422436**
- $\hat{\beta}_{2GMM}$ = **0.400500401025**
- $\hat{\beta}_{3GMM}$ = **-0.00999176584737**
- Value of GMM criterion function at using GMM parameters obtained is: **0.00182128980860659 73**