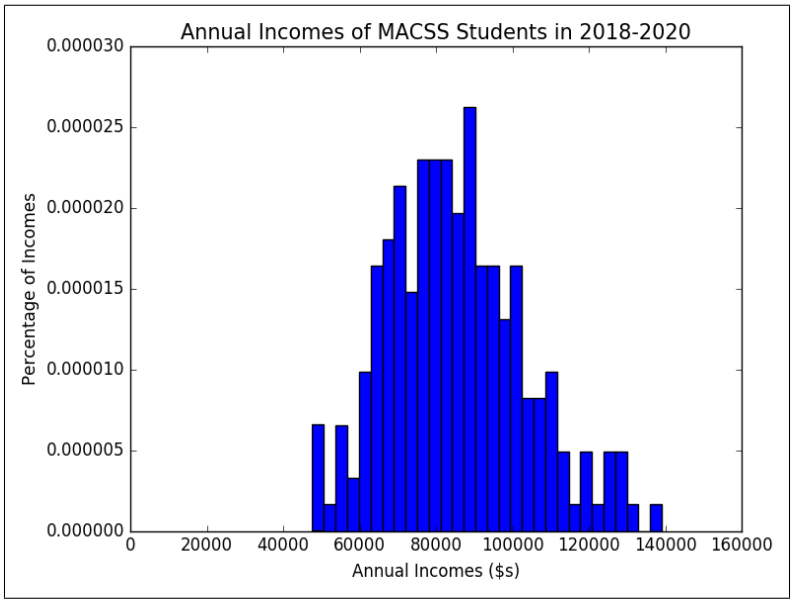
Problem Set #3

MACS 30100, Dr. Evans
Tong Ju

Problem 1 Lognormal Distribution of Income Data
The minimization processes in the 4 estimations in (b) to (e) all succeeded.

Part (a). A histogram of annual incomes of students who graduated in 2018, 2019, and 2020 from the University of Chicago M.A. Program in Computational Social Science.

Figure 1:

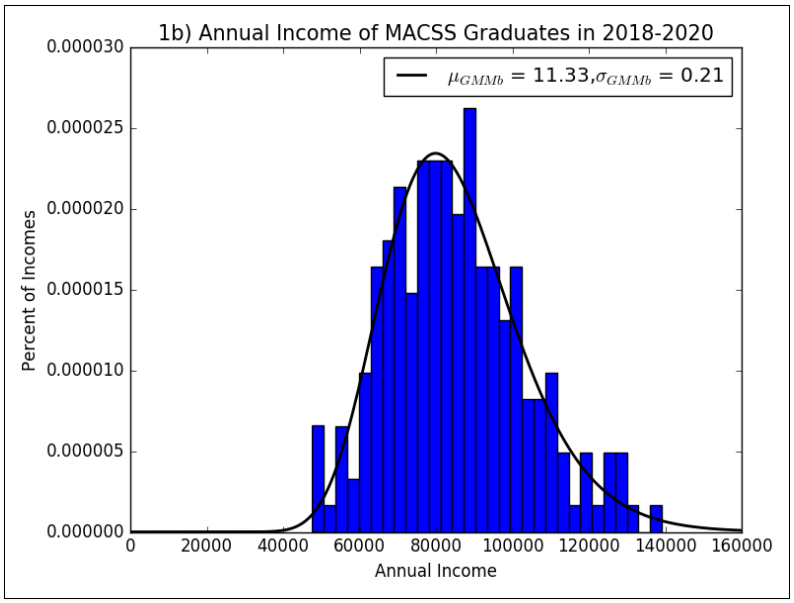


Part (b). One step GMM by using mean and variance as two moments

One step GMM is estimated with mean and variance as moments and identity weighting matrix. The initial guess of μ is 11 and σ is 0.1 .
The resulted criterion value is 3.12965575e-15. The estimated parameters are as follow: $\mu_{GMM1} = 11.3318808967; \sigma_{GMM1} = 0.208696648842$

moments	mean	var
data	85276.8236063	323731572.23
model	85276.8278958	323731564.303
data - model	-0.0042894944927	7.9263535738

Figure 2: 1-step GMM by mean and variance

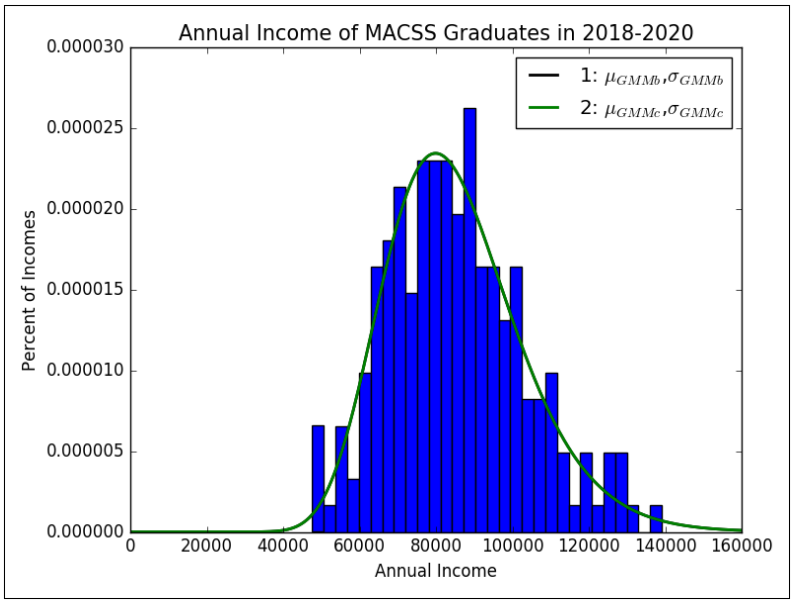


Part (c). Two step GMM with mean and variance

The estimated weighting matrix caculated from b) is used to the 2-step GMM modeling. The initial guess of μ and σ is from (b).
The resulted criterion value is 6.98194849. The estimated parameters are as follow: $\mu_{GMM1} = 11.3318796375; \sigma_{GMM1} = 0.208696655465$

moments	mean	var
data	85276.8236063	323731572.23
model	85276.7206263	323730770.861
data - model	0.102979999676	801.368354678

Figure 3: 2-step GMM by mean and variance

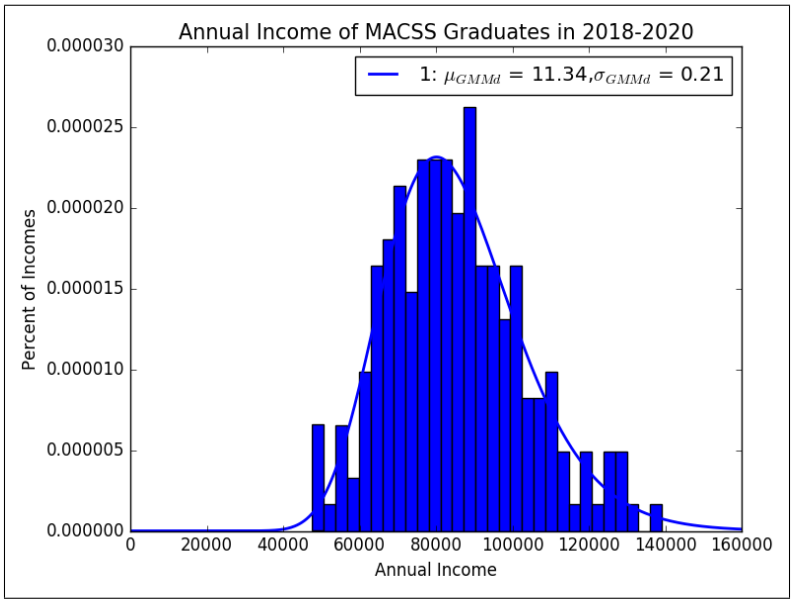


Part (d). One step GMM with three proportion moments

One step GMM is estimated with percent of individuals who earn less than $75,000, between $75,000 and $100,000, and more than $100,000 as moments and identity weighting matrix. The initial guess of μ is 11 and σ is 0.2.
The resulted criterion value is 1.45549442e-12

moments	less than 75,000	75,000-100,000	more than 100,000
data	0.3	0.5	0.2
model	0.2999999	0.4999996	0.200000057
data - model	5.51066977e-08	2.5350653798e-09	-5.7641763084e-08 height

Figure 4: One step GMM with three proportion moments

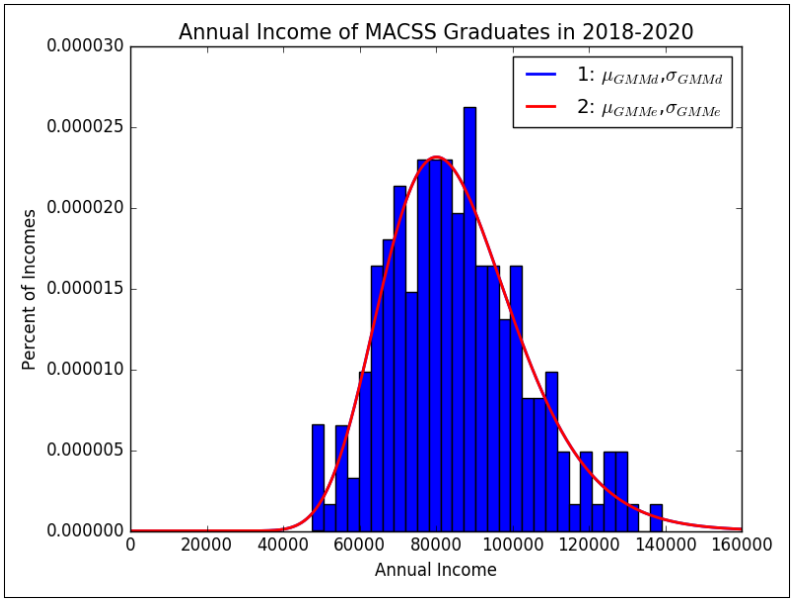


Part (e). Two step GMM with three proportion moments

Two step GMM is estimated with percent of individuals who earn less than $75,000, between $75,000 and $100,000, and more than $100,000 as moments and weighting matrix derived from (d). The initial guess of μ and σ is from (d).
The resulted criterion value is 2.6828065. The estimated parameters are as follow: $\mu_{GMM2_3} = 11.3356815244 \sigma_{GMM2_3} = 0.210598561303$

moments	less than 75,000	75,000-100,000	more than 100,000
data	0.3	0.5	0.2
model	0.300000000	0.49999998	0.200000012
data - model	-5.650838130e-10	1.3310548885e-08	-1.274546507e-08 height

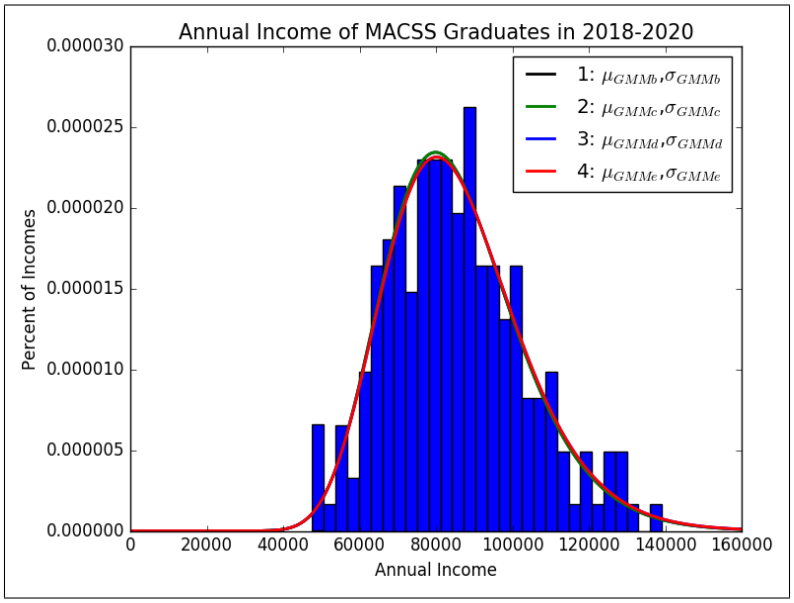
Figure 5: Two step GMM with three proportion moments



Part (f). Estimation comparison

According to the above data and the plot below , we can find that the resulted modelings are very similar to each other. However, based on the difference of the modeled and real data, I think e) would be the best estimation. First, it is obvious b) and c) estimations are visually indistinguishable in the plot. Though, comparing the difference of data moment and model moment of these two estimations, we can still find that the two-step estimation in (c) has been a little bit improved than (b). The same happens to the estimation form d) and e). The difference between predicted moments and real moments are smaller in e) than d). Second, looking at the histogram plot of our "real" data, we will find the real data does not exactly follow the normal distribution (data itself is right-skewed). For the estimation in b) and c) is only focusing on the mean and variance, which cannot really catch the features of our data. However, d) and e) consider the asymmetry of disribution of real data, which, I think, could predict the real data much better. Thirdly, in the plot below, we can also find visually the e)'s plot (red) shifted a little bit right than the c)'s plot. Sine e) is much better than d). So I deem the e) estimation is the best one.

Figure 6: comparison of GMM methods from b) to e)



Problem 2
Part (a). Linear regression and GMM

The parameters of the model are estimated by GMM by solving the minimization problem of the GMM criterion function. 200 sick data are considered as the moments of data, and simple difference between 'sick' and the estimated sick from linear model is used as the error function. The initial guesses for β_0, β_1, β_2, β_3 are 0, 0, 0, 1.
The resulted criterion value is 0.00182128984298. The estimated parameters are as follow:

$$\beta_{0,GMM} = 0.2516448071$$
$$\beta_{1,GMM} = 0.0129334339724$$
$$\beta_{2,GMM} = 0.400501320985$$
$$\beta_{3,GMM} = -0.00999168730293$$