Problem Set #3

MACS 40200, Dr. Evans
Due Wednesday, Feb. 5 at 1:30pm

1. **Matching the U.S. income distribution by GMM (5 points).** In this problem set, you will use the comma-delimited data file `hh_inc_synth.txt` in the `PS3/data` folder, which contains the 121,085 observations (synthetic) on household U.S. income. Table 1 displays histogram counts and percentages moments listed in along with the midpoints of each bin. The first column in the data file gives the percent of the population in each income bin (the third column of Table 1). The second column in the data file has the midpoint of each income bin. So the midpoint of the first income bin of all household incomes less than \$5,000 is \$2,500.

   (a) (0.5 point) Use the `numpy.histogram()` function to create the moments in Table 1 from the synthetic household income data in comma-delimited text file `hh_inc_synth.txt` by inputing the appropriate list of bin edges for the `bins` argument.

   (b) (1 points) Plot the histogram of the data `hh_inc_synth.txt` using the bins described in the first column of Table 1, which you used as an input to part (a), and the height being the percent of observations in that bin and not the count frequency (use the `weights` option rather than the `density` option in `matplotlib.pyplot.hist`). List the dollar amounts on the x-axis as thousands of dollars. That is, divide them by 1,000 to put them in units of thousands of dollars (\$000s). Even though the top bin is listed as \$250,000 and up in Table 1, the synthetic data are top-coded at \$350,000, so set to last bin edge to \$350,000. (It doesn't look very good graphing it between 0 and ∞.) Because the 41st bar is 10 times bigger than the first 40 bars, divide it's percentage by 10 just for plotting purposes. And because the 42nd bar is 20 times bigger than the first 40 bars, divide it's percentage by 20 just for plotting purposes. You can do this by dividing the weights for observations in the last two bins by 10 and 20, respectively. In summary, your histogram should have 42 bars. The first 40 bars for the lowest income bins should be the same width. However, the last two bars should be different widths from each other and from the rest of the bars. It should look like Figure 1. [Hint: look at the `matplotlib.pyplot.hist` command option of `bins` and submit a list of bin edges for the `bins` option.]

   (c) (1 points) Using GMM, fit the lognormal $LN(x; \mu, \sigma)$ distribution defined in the MLE notebook to the distribution of household income data using the moments from the data file. Make sure to try various initial guesses. (HINT: $\mu_0 = \ln(avg.inc.)$ might be good.) For your weighting matrix $\boldsymbol{W}$, use a 42×42 diagonal matrix in which the diagonal elements are the moments from the data file. This will put the most weight on the

moments with the largest percent of the population. Report your estimated values for $\hat{\mu}$ and $\hat{\sigma}$, as well as the value of the minimized criterion function $\boldsymbol{e}(\boldsymbol{x}|\hat{\boldsymbol{\theta}})^T \boldsymbol{W} \boldsymbol{e}(\boldsymbol{x}|\hat{\boldsymbol{\theta}})$. Plot the histogram from part (a) overlayed with a line representing the implied histogram from your estimated lognormal (LN) distribution. Each point on the line is the midpoint of the bin and the implied height of the bin. Do not forget to divide the values for your last two moments by 10 and 20, respectively, so that they match up with the histogram.

(d) (1 points) Using GMM, fit the gamma $GA(x; \alpha, \beta)$ distribution defined in the MLE notebook to the distribution of household income data using the moments from the data file. Use $\alpha_0 = 3$ and $\beta_0 = 20,000$ as your initial guess.[1] Report your estimated values for $\hat{\alpha}$ and $\hat{\beta}$, as well as the value of the minimized criterion function $\boldsymbol{e}(\boldsymbol{x}, \hat{\boldsymbol{\theta}})^T \boldsymbol{W} \boldsymbol{e}(\boldsymbol{x}, \hat{\boldsymbol{\theta}})$. Use the same weighting matrix as in part (b). Plot the histogram from part (a) overlayed with a line representing the implied histogram from your estimated gamma (GA) distribution. Do not forget to divide the values for your last two moments by 10 and 20, respectively, so that they match up with the histogram.

(e) (0.5 point)Plot the histogram from part (a) overlayed with the line representing the implied histogram from your estimated lognormal (LN) distribution from part (b) and the line representing the implied histogram from your estimated gamma (GA) distribution from part (c). What is the most precise way to tell which distribution fits the data the best? Which estimated distribution—LN or GA—fits the data best?

(f) (1 point) Repeat your estimation of the GA distribution from part (c), but use the two-step estimator for the optimal weighting matrix $\hat{\boldsymbol{W}}_{twostep}$. Do your estimates for α and β change much? How can you compare the goodness of fit of this estimated distribution versus the goodness of fit of the estimated distribution in part (c)?

[1] These initial guesses come from the property of the gamma (GA) distribution that $E(x) = \alpha\beta$ and $Var(x) = \alpha\beta^2$.

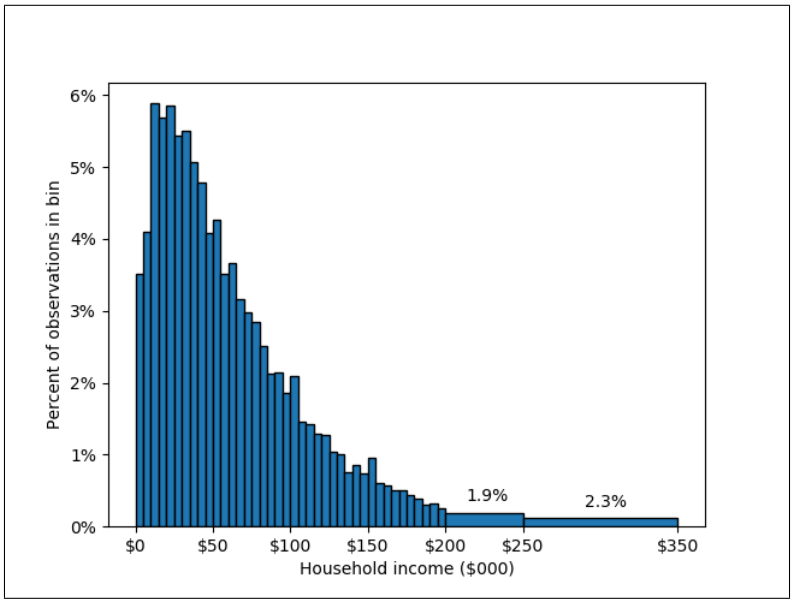
Table 1: Distribution of Household Money Income by Selected Income Class, 2011

Income class	# households (000s)	households %
All households	121,084	100.0
Less than $5,000	4,261	3.5
$5,000 to $9,999	4,972	4.1
$10,000 to $14,999	7,127	5.9
$15,000 to $19,999	6,882	5.7
$20,000 to $24,999	7,095	5.9
$25,000 to $29,999	6,591	5.4
$30,000 to $34,999	6,667	5.5
$35,000 to $39,999	6,136	5.1
$40,000 to $44,999	5,795	4.8
$45,000 to $49,999	4,945	4.1
$50,000 to $54,999	5,170	4.3
$55,000 to $59,999	4,250	3.5
$60,000 to $64,999	4,432	3.7
$65,000 to $69,999	3,836	3.2
$70,000 to $74,999	3,606	3.0
$75,000 to $79,999	3,452	2.9
$80,000 to $84,999	3,036	2.5
$85,000 to $89,999	2,566	2.1
$90,000 to $94,999	2,594	2.1
$95,000 to $99,999	2,251	1.9
$100,000 to $104,999	2,527	2.1
$105,000 to $109,999	1,771	1.5
$110,000 to $114,999	1,723	1.4
$115,000 to $119,999	1,569	1.3
$120,000 to $124,999	1,540	1.3
$125,000 to $129,999	1,258	1.0
$130,000 to $134,999	1,211	1.0
$135,000 to $139,999	918	0.8
$140,000 to $144,999	1,031	0.9
$145,000 to $149,999	893	0.7
$150,000 to $154,999	1,166	1.0
$155,000 to $159,999	740	0.6
$160,000 to $164,999	697	0.6
$165,000 to $169,999	610	0.5
$170,000 to $174,999	617	0.5
$175,000 to $179,999	530	0.4
$180,000 to $184,999	460	0.4
$185,000 to $189,999	363	0.3
$190,000 to $194,999	380	0.3
$195,000 to $199,999	312	0.3
$200,000 to $249,999	2,297	1.9
$250,000 and over	2,808	2.3
Mean income*	$67,270	
Median income*	$50,090	

Source: 2011 CPS household income count data Current Population Survey (2012, Table HINC-01)

* Mean and median are from synthesized data `hh_inc_synth` and are both slightly lower than the true survey statistics.

Figure 1: Histogram of household income: $N = 121,085$



2. **Estimating the Brock and Mirman (1972) model by GMM (5 points).** You can observe time series data in an economy for the following variables: (c_t, k_t, w_t, r_t). Data on (c_t, k_t, w_t, r_t) can be loaded from the file `MacroSeries.txt` in the `PS3/data` folder. This file is a comma separated text file with no labels. The variables are ordered as (c_t, k_t, w_t, r_t). These data have 100 periods, which are quarterly (25 years). Suppose you think that the data are generated by a process similar to the Brock and Mirman (1972). A simplified set of characterizing equations of the Brock and Mirman model are the following.

$$(c_t)^{-1} - \beta E\left[r_{t+1}(c_{t+1})^{-1}\right] = 0 \tag{1}$$

$$c_t + k_{t+1} - w_t - r_t k_t = 0 \tag{2}$$

$$w_t - (1-\alpha)e^{z_t}(k_t)^{\alpha} = 0 \tag{3}$$

$$r_t - \alpha e^{z_t}(k_t)^{\alpha-1} = 0 \tag{4}$$

$$z_t = \rho z_{t-1} + (1-\rho)\mu + \varepsilon_t \quad \text{where} \quad E\left[\varepsilon_t\right] = 0 \tag{5}$$

The variable c_t is aggregate consumption in period t, k_{t+1} is total household savings and investment in period t for which they receive a return in the next period (this model assumes full depreciation of capital). The wage per unit of labor in period t is w_t and the interest rate or rate of return on investment is r_t. Total factor productivity is z_t, which follows an AR(1) process given in (5). The rest of the symbols in the equations are parameters that must be estimated or must be otherwise given $(\alpha, \beta, \rho, \mu, \sigma)$. The constraints on these parameters are the following.

$$\alpha, \beta \in (0,1), \quad \mu, \sigma > 0, \quad \rho \in (-1,1)$$

Assume that the first observation in the data file variables is $t = 1$. Let k_1 be the first observation in the data file for the variable k_t.

(a) Estimate α, ρ, and μ by GMM using the unconditional moment conditions that $E[\varepsilon_t] = 0$ and $E[\beta r_{t+1} c_t / c_{t+1} - 1] = 0$. Assume $\beta = 0.99$. Use the identity matrix $I(4)$ as your estimator of the optimal weighting matrix. Use the following four moment conditions to estimate the four parameters.

$$E\Big[z_{t+1} - \rho z_t - (1-\rho)\mu\Big] = 0 \tag{6}$$

$$E\Big[\Big(z_{t+1} - \rho z_t - (1-\rho)\mu\Big) z_t\Big] = 0 \tag{7}$$

$$E\left[\beta\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} \frac{c_t}{c_{t+1}} - 1\right] = 0 \tag{8}$$

$$E\left[\left(\beta\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} \frac{c_t}{c_{t+1}} - 1\right) w_t\right] = 0 \tag{9}$$

The estimation inside each iteration of the minimizer of the GMM objective function is the following.

- Given a guess for (α, ρ, μ) and data (c_t, k_t, w_t, r_t), use (4) to back out an implied series for z_t.
- Given z_t, parameters (α, ρ, μ) and data (c_t, k_t, w_t, r_t), calculate four empirical analogues of the moment conditions (6), (7), (8), and (9).
- Update guesses for parameters (α, ρ, μ) until minimum criterion value is found.

Report your estimated parameter values $(\hat{\alpha}, \hat{\rho}, \hat{\mu})$ and the value of your minimized criterion function.

(b) (Optional) Compute the two-step GMM estimator of (α, ρ, μ) and use the finite difference Jacobian method for the estimator of the variance-covariance of the two-step GMM point estimates $(\hat{\alpha}_{GMM}, \hat{\rho}_{GMM}, \hat{\mu}_{GMM})$.

References

Brock, William A. and Leonard J. Mirman, "Optimal economic growth and uncertainty: The discounted case," *Journal of Economic Theory*, June 1972, *4* (3), 479–513.

Current Population Survey, "2012 Annual Social and Economic (ASEC) Supplement," Technical Report, Bureau of the Census and Bureau of Labor Statistics 2012.