Chapter 14

Distribution Fitting: MLE and GMM

14.1 Data Generating Process and Model Estimation

Define a **data generating process (DGP)** as the real world process or mechanism by which some data are generated. This is the mechanism that we seek to mimic with our mathematical models. Sometimes, the model that is supposed to represent reality is called the data generating process. A common specification of a DGP is the following familiar system of potentially nonlinear equations,

$$\boldsymbol{F}\left(\boldsymbol{x}_t, \boldsymbol{z}_t; \boldsymbol{\theta}\right) = 0 \tag{14.1}$$

where $\boldsymbol{F}$ is a vector of equations, each of which is a function of an endogenous state vector $\boldsymbol{x}_t$, an exogenous state vector $\boldsymbol{z}_t$, and a parameter vector $\boldsymbol{\theta}$ at time-t. Note that both the endogenous state vector $\boldsymbol{x}_t$ and the exogenous state vector $\boldsymbol{z}_t$ can include variables from the current period and from previous periods. Another way to think of a data generating process is all the information necessary to simulate the time series of state variables.

As an example, the DGP or structural model for the Brock and Mirman (1972) stochastic growth model is the following.

$$(c_t)^{-1} - \beta E\left[r_{t+1}(c_{t+1})^{-1}\right] = 0 \tag{14.2}$$

$$c_t + k_{t+1} - w_t - r_t k_t = 0 \tag{14.3}$$

$$w_t - (1-\alpha)e^{z_t}\left(\frac{K_t}{L_t}\right)^{\alpha} = 0 \tag{14.4}$$

$$r_t - \alpha e^{z_t}\left(\frac{L_t}{K_t}\right)^{1-\alpha} = 0 \tag{14.5}$$

$$K_t - k_t = 0 \tag{14.6}$$

$$L_t - 1 = 0 \tag{14.7}$$

$$\begin{aligned} z_t &= \rho z_{t-1} + (1-\rho)\mu + \varepsilon_t \\ \text{where} &\quad \varepsilon_t \sim N(0,\sigma^2) \quad \text{or} \\ E&\left[z_{t+1} - \rho z_t - (1-\rho)\mu\right] = 0 \end{aligned} \tag{14.8}$$

Equations (14.2) through (14.8) represent the DGP for a growth economy. That is, if one draws a series of stochastic shocks $\{\varepsilon_t\}_{t=1}^n$, one can generate a time series for each of the variables in the model. Relating this DGP to the general formulation in (14.1), the system of nonlinear equations $\boldsymbol{F}$ is equations (14.2) to (14.8). The vector of endogenous state vector $\boldsymbol{x}_t$ is the singleton k_t, and the exogenous state vector $\boldsymbol{z}_t$ is z_t.[1] The parameter vector $\boldsymbol{\theta}$ of the DGP is $\{\beta, \alpha, \rho, \mu, \sigma\}$.

The question that the econometrician asks is the following. How should one choose the parameters of the model to make the model most closely reflect reality? With respect to the Brock and Mirman DGP in equations (14.2) through (14.8), the question is how to choose the parameters $\boldsymbol{\theta} = \{\beta, \alpha, \rho, \mu, \sigma\}$ so that the model reflects reality. With a more general system of equations represented by (14.1) $\boldsymbol{F}(\boldsymbol{x}_t, \boldsymbol{z}_t; \boldsymbol{\theta}) = 0$, the question is how to choose $\boldsymbol{\theta}$ to make the system of equations be as true as possible.

The two estimation methods we will detail here are maximum likelihood estimation (MLE) and generalized method of moments (GMM) estimation. These two methods both have advantages and both have disadvantages. Both methods leverage different assumptions

[1] We showed in Chapter 6 that the state in the Brock and Mirman (1972) model reduces to (k_t, z_t).

about the DGP to estimate the parameters. MLE requires the modeler to assign a specific functional form to the distribution of the shocks in the model. Once this is done, the modeler can use the distribution of the shocks and the data to create a likelihood function for the observed data. The MLE estimate of the parameter vector $\hat{\boldsymbol{\theta}}_{MLE}$ is the one that maximizes the likelihood function of the observed data. GMM remains completely agnostic as to the distribution of the shocks. The GMM estimate is simply the parameter vector $\hat{\boldsymbol{\theta}}_{GMM}$ that minimizes the errors in the characterizing or moment equations.

In this chapter, we will use an example of fitting a distribution $f(x_t;\boldsymbol{\theta})$ to data using different information about the data x_t. The most intuitive progression is to start with MLE and then do GMM.

14.2 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) was formalized in the early twentieth century by Edgeworth (1908a,b) and Fisher (1925). MLE requires the strong assumption of a functional form for the distribution of the random variables in the data generating process. However, if one believes that the distributional assumption is close to the truth, then MLE produces very nice properties.[2]

As a general formulation let $\boldsymbol{x}$ be a vector of variables of interest. MLE requires an assumption about the joint probability density function (pdf) of the vector of variables. Let $f(\boldsymbol{x},\boldsymbol{\theta})$ be the assumed pdf of $\boldsymbol{x}$ as characterized by parameter vector $\boldsymbol{\theta}$. Suppose you have one observation of your vector of variables $\boldsymbol{x}_i$. That is, you have one data point. The PDF at that data point $f(\boldsymbol{x}_i,\boldsymbol{\theta})$ tells you the probability density of that observation given the particular distribution characterized by $\boldsymbol{\theta}$. Another interpretation of $f(\boldsymbol{x}_i,\boldsymbol{\theta})$ is the probability of observing data $\boldsymbol{x}_i$ given distribution $f(\cdot,\boldsymbol{\theta})$.

This last interpretation gives us the intuition for MLE. Suppose we had n independent observations of the data $\{\boldsymbol{x}_i\}_{i=1}^n$. The total probability or likelihood of observing those n observations $\{\boldsymbol{x}_i\}_{i=1}^n$ given the distribution $f(\cdot,\boldsymbol{\theta})$ is the product of the PDF values at each

[2] Fuhrer et al. (1995) find evidence that MLE estimators dominate GMM estimators in a certain class of DSGE models.

observation.

$$\mathcal{L} = \prod_{i=1}^{n} f(\boldsymbol{x}_i, \boldsymbol{\theta}) \tag{14.9}$$

Equation (14.9) is called the **likelihood function** because it can be interpreted as the probability of observing data $\{\boldsymbol{x}_i\}_{i=1}^n$ given the probability distribution $f(\cdot, \boldsymbol{\theta})$. MLE chooses $\boldsymbol{\theta}$ to maximize the likelihood function or the probability of observing the n data points $\{\boldsymbol{x}_i\}_{i=1}^n$.

Because the maximization of a product can be computationally difficult—especially if you have a large number of observations—a more tractable approach that is nearly always followed in MLE is to maximize the logarithm of the likelihood function.

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}} \ln \mathcal{L} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \ln\Big(f(\boldsymbol{x}_i, \boldsymbol{\theta})\Big) \tag{14.10}$$

The **log likelihood function** $\ln\mathcal{L}$ turns the product of n terms in (14.9) into the sum of n terms in (14.10). Using this sum as a criterion function to be maximized is much better behaved computationally than the product of n terms. Furthermore, one can sometimes get analytical solutions for $\hat{\boldsymbol{\theta}}_{MLE}$ for some pdf's f.

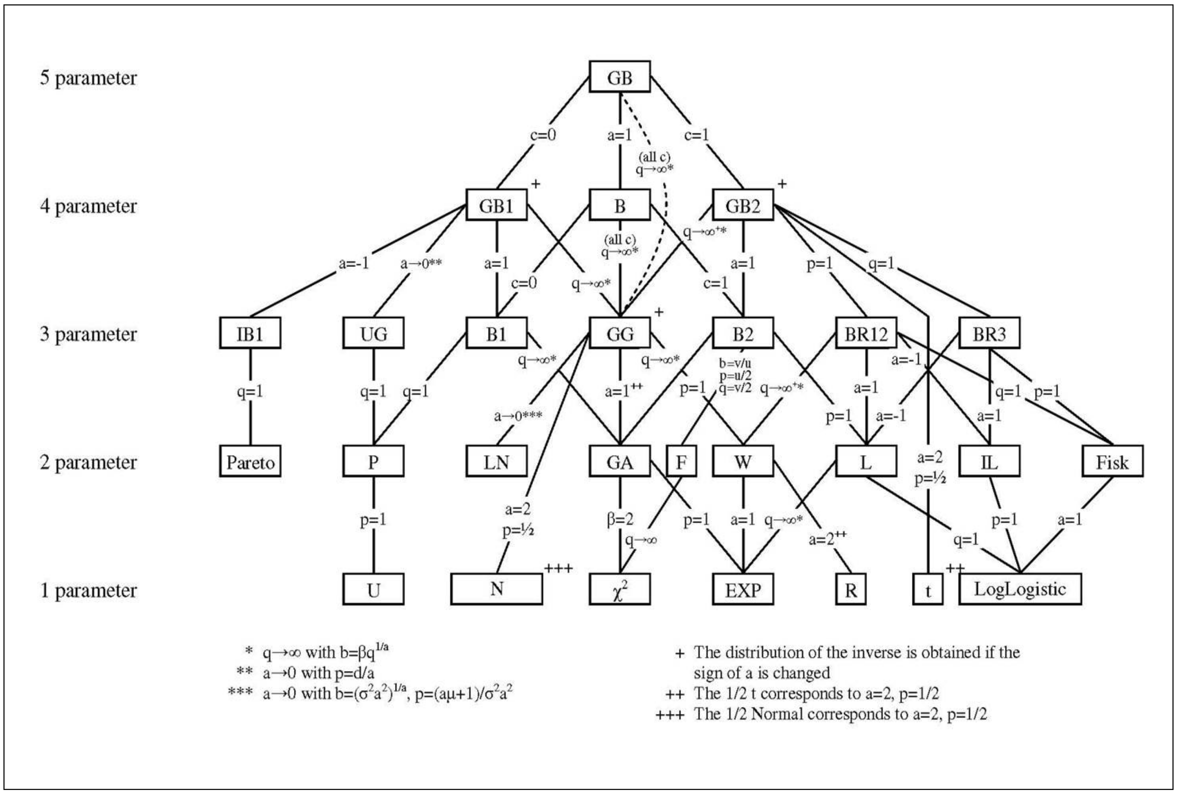
14.2.1 Generalized Beta Distributions

For exercises in this chapter, you will need to know the functional forms of four continuous univariate probability density functions (pdf's), each of which are part of the generalized beta family of distributions (See Figure 14.1). The lognormal distribution (LN) is the distribution of the exponential of a normally distributed variable with mean μ and standard deviation σ. If the variable x_i is lognormally distributed $x_i \sim LN(\mu, \sigma)$, then the log of x_i is normally distributed $\ln(x_i) \sim N(\mu, \sigma)$. The pdf of the lognormal distribution is the following.

$$\text{(LN):}\quad f(x;\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}, \quad x \in (0,\infty),\ \mu \in (-\infty,\infty),\ \sigma > 0 \tag{14.11}$$

Note that the lognormal distribution has a support that is strictly positive. This is one reason why it is commonly used to approximate income distributions. A household's total income is rarely negative. The lognormal distribution also has a lot of the nice properties of the normal distribution.

Figure 14.1: Generalized beta family of distributions [taken from McDonald and Xu (1995, Fig. 2)]



Another two-parameter distribution with strictly positive support is the gamma (GA) distribution. The pdf of the gamma distribution is the following.

$$\begin{aligned} \text{(GA)}: \quad f(x;\alpha,\beta) &= \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{\beta}}, \quad x\in[0,\infty),\ \alpha,\beta>0 \\ \text{where} \quad \Gamma(z) &\equiv \int_0^\infty t^{z-1}e^{-t}dt \end{aligned} \tag{14.12}$$

The gamma function $\Gamma(\cdot)$ within the gamma (GA) distribution is a common function that has a preprogrammed function in most programming languages.

The lognormal (LN) and gamma (GA) distributions are both two-parameter distributions and are both special cases of the three-parameter generalized gamma (GG) distribution. The pdf of the generalized gamma distribution is the following.

$$\begin{aligned}\text{(GG):}\quad f(x;\alpha,\beta,m) = &\frac{m}{\beta^\alpha\Gamma\left(\frac{\alpha}{m}\right)}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^m},\quad x\in[0,\infty),\ \alpha,\beta,m>0 \\ &\text{where}\quad \Gamma(z)\equiv\int_0^\infty t^{z-1}e^{-t}dt\end{aligned} \tag{14.13}$$

The relationship between the generalized gamma (GG) distribution in (14.13) and the gamma (GA) distribution in (14.13) is straightforward. The GA distribution equals the GG distribution at $m=1$.

$$GA(\alpha,\beta) = GG(\alpha,\beta,m=1) \tag{14.14}$$

The relationship between the generalized gamma (GG) distribution in (14.13) and the lognormal (LN) distribution in (14.11) is less straightforward. The LN distribution equals the GG distribution as α goes to zero, $\beta=(\alpha\sigma)^{\frac{2}{\alpha}}$, and $m=\frac{\alpha\mu+1}{\alpha^2\sigma^2}$.[3]

$$LN(\mu,\sigma) = \lim_{\alpha\rightarrow 0} GG\left(\alpha,\beta=(\alpha\sigma)^{\frac{2}{\alpha}},m=\frac{\alpha\mu+1}{\alpha^2\sigma^2}\right) \tag{14.15}$$

The last distribution we describe is the generalized beta 2 (GB2) distribution. Like the GG, GA, and LN distributions, it also has a strictly positive support. The pdf of the generalized beta 2 distribution is the following.

$$\begin{aligned}\text{(GB2):}\quad f(x;a,b,p,q) = &\frac{ax^{ap-1}}{b^{ap}B(p,q)\left(1+\left(\frac{x}{b}\right)^a\right)^{p+q}},\quad x\in[0,\infty),\ a,b,p,q>0 \\ &\text{where}\quad B(v,w)\equiv\int_0^1 t^{v-1}(1-t)^{w-1}dt\end{aligned} \tag{14.16}$$

The beta function $B(\cdot,\cdot)$ within the GB2 distribution is a common function that has a preprogrammed function in most programming languages. The three-parameter generalized gamma (GG) distribution in (14.13) is a nested case of the four-parameter generalized beta 2 (GB2) distribution in (14.16) as q goes to ∞ and for $a=m$, $b=q^{1/m}\beta$, and $p=\frac{\alpha}{m}$.[4]

$$GG(\alpha,\beta,m) = \lim_{q\rightarrow\infty} GB2\left(a=m,b=q^{1/m}\beta,p=\frac{\alpha}{m},q\right) \tag{14.17}$$

[3]See McDonald et al. (2013) for derivation.
[4]See McDonald (1984, p. 662) for a derivation.

Figure 14.1 shows the all the relationships between the various pdf's in the generalized beta family of distributions.

14.2.2 Exercises

For exercises 14.1 through 14.5, use the monthly health expenditure data in the tab-delimited data file `clms.txt`. Each observation in this file x_i is total monthly expenditures on health care in dollars for individual i. The data file contains 10,619 observations.

Exercise 14.1. Calculate and report the mean, median, maximum, minimum, and standard deviation of monthly health expenditures for these data. Plot two histograms of the data in which the y-axis gives the percent of observations in the particular bin of health expenditures and the x-axis gives the value of monthly health expenditures.[5] In the first histogram, use 1,000 bins to plot the frequency of all the data. In the second histogram, use 100 bins to plot the frequency of only monthly health expenditures less-than-or-equal-to \$800 ($x_i \leq 800$). Adjust the frequencies of this second histogram to account for the observations that you have not displayed ($x_i > 800$). Comparing the two histograms, why might you prefer the second one?

Exercise 14.2. Using MLE, fit the gamma $GA(x; \alpha, \beta)$ distribution from (14.12) to the individual observation data. Use $\beta_0 = Var(x)/E(x)$ and $\alpha_0 = E(x)/\beta_0$ as your initial guess.[6] Report your estimated values for $\hat{\alpha}$ and $\hat{\beta}$, as well as the value of the maximized log likelihood function $\ln \mathcal{L}$. Plot the second histogram from exercise 14.1 overlayed with a line representing the implied histogram from your estimated gamma (GA) distribution.

Exercise 14.3. Using MLE, fit the generalized gamma $GG(x; \alpha, \beta, m)$ distribution from (14.13) to the individual observation data. Use your estimates for α and β from exercise 14.2, as well as $m = 1$, as your initial guess. Report your estimated values for $\hat{\alpha}$, $\hat{\beta}$, and $\hat{m}$, as well as the value of the maximized log likelihood function $\ln \mathcal{L}$. Plot the second

[5] As a reminder, a histogram is a bar chart, in which each of the bars represents the percent of observations in a particular x-axis bin (monthly health expenditures, in this case). As such, the bars should be touching each other because each edge of each bar represents the cutoff level of each income category bin. I prefer to color the histogram bars red.

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

histogram from exercise 14.1 overlayed with a line representing the implied histogram from your estimated generalized gamma (GG) distribution.

Exercise 14.4. Using MLE, fit the generalized beta 2 $GB2(x;a,b,p,q)$ distribution from (14.16) to the individual observation data. Use your estimates for α, β, and m from exercise 14.3, as well as $q = 10,000$, as your initial guess. Report your estimated values for $\hat{a}$, $\hat{b}$, $\hat{p}$, and $\hat{q}$, as well as the value of the maximized log likelihood function $\ln\mathcal{L}$. Plot the second histogram from exercise 14.1 overlayed with a line representing the implied histogram from your estimated generalized beta 2 (GB2) distribution.

Exercise 14.5. Plot the second histogram from exercise 14.1 overlayed with the line representing the implied histogram from your estimated gamma (GA) distribution from exercise 14.2, the line representing the implied histogram from your estimated generalized gamma (GG) distribution from exercise 14.3, and the line representing the implied histogram from your estimated generalized gamma (GB2) distribution from exercise 14.4. What is the most precise way to tell which distribution fits the data the best?

14.3 Generalized Method of Moments (GMM) Estimation

GMM was first formalized by Hansen (1982). A strength of GMM estimation is that the econometrician can remain completely agnostic as to the distribution of the random variables in the DGP. For identification, the econometrician simply needs at least as many moment conditions from the data as he has parameters to estimate.

A **moment** of the data is broadly defined as any statistic that summarizes the data to some degree. A data moment could be as narrow as an individual observation from the data or as broad as the sample average. GMM estimates the parameters of a model or data generating process to make the model moments as close as possible to the corresponding data moments.[7]

[7] See Davidson and MacKinnon (2004, ch. 9) for a more detailed treatment of GMM. The estimation methods of linear least squares, nonlinear least squares, generalized least squares, and instrumental variables estimation are all specific cases of the more general GMM estimation method.

Let $\bar{\boldsymbol{m}}$ be an $N \times 1$ vector of moments from the real world. If the DGP of the model is $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\theta})$, then define $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\theta})$ as a vector of moments from the model that correspond to the real-world moment vector $\bar{\boldsymbol{m}}$. Note that GMM requires both real world data and deterministic moments from the model in order to estimate $\hat{\boldsymbol{\theta}}$. There is also a stochastic way to generate moments from the model, which we discuss in Chapter 12. The GMM approach of estimating the parameter vector $\hat{\boldsymbol{\theta}}$ is to choose $\boldsymbol{\theta}$ to minimize some distance measure of the data moments $\bar{\boldsymbol{m}}$ from the model moments $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\theta})$.

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}} \; \boldsymbol{e}(\boldsymbol{x}, \boldsymbol{\theta})^T \, \boldsymbol{W} \, \boldsymbol{e}(\boldsymbol{x}, \boldsymbol{\theta}) \quad \text{where} \quad \boldsymbol{e}(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \left(\frac{\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\theta}) - \bar{\boldsymbol{m}}}{\bar{\boldsymbol{m}}}\right) \tag{14.18}$$

The $N \times 1$ vector $\boldsymbol{e}(\boldsymbol{x}, \boldsymbol{\theta})$ is a vector of percent deviations of each model moment from its corresponding data moment. It is important to put the moment difference vector in percent deviations so that the problem is scaled properly and does not suffer from ill conditioning. We call the quadratic form expression $\boldsymbol{e}(\boldsymbol{x}, \boldsymbol{\theta})^T \, \boldsymbol{W} \, \boldsymbol{e}(\boldsymbol{x}, \boldsymbol{\theta})$ the **criterion function** because it is a strictly positive scalar that is the object of the minimization in (14.18). The $N \times N$ weighting matrix $\boldsymbol{W}$ in the criterion function allows the econometrician to control how each moment is weighted in the minimization problem. For example, an $N \times N$ identity matrix for $\boldsymbol{W}$ would give each moment equal weighting, and the criterion function would be a simply sum of squared percent deviations (errors). Other weighting strategies can be dictated by the nature of the problem or model.

Another issue in GMM estimation is identification. Suppose the parameter vector $\boldsymbol{\theta}$ has M elements, or rather, M parameters to be estimated. In order to estimate $\boldsymbol{\theta}$ by GMM, you must have at least as many moments as parameters to estimate $N \geq M$. If you have exactly as many moments as parameters to be estimated $N = M$, the model is said to be **exactly identified**. If you have more moments than parameters to be estimated $N > M$, the model is said to be **overidentified**. If you have fewer moments than parameters to be estimated $N < M$, the model is said to be **underidentified**. There are good reasons to overidentify $N > M$ the model in GMM estimation. The main reason is that 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 overidentified $N > M$.

One last point about GMM regards moment selection and verification of results. The real world has an infinite supply of potential moments that describe some part of the data. Choosing moments to estimate parameters by GMM requires understanding of the model, intuition about its connections to the real world, and artistry. A good GMM estimation will include moments that have some relation to or story about their connection to particular parameters of the model to be estimated. In addition, a good verification of a GMM estimation is to take some moment from the data that was not used in the estimation and see how well the corresponding moment from the estimated model matches that **outside moment**.

14.3.1 Exercises

For exercises 14.6 through 14.9, you will use the tab-delimited data file `usincmoms.txt`, which contains the 42 moments listed in Table 14.1 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 14.1). 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.

Exercise 14.6. Plot the histogram implied by the moments in the tab-delimited text file `usincmoms.txt`.[8] The centers of each bin are in the second column of the data file `usincmoms.txt`. Divide them by 1,000 to put them in units of thousands of dollars ($000s). The cutoffs are given in Table 14.1. Even though the top bin is all incomes of $250,000 and up, only graph the histogram up to the maximum income of $350,000. (It doesn't look very good graphing it between 0 and ∞.) In summary, your histogram should have 42 bars. The first 40 bars for the lowest income bins should be the same size. However, the last two bars should be different sizes from each other and from the rest of the bars. Because the 41st bar is 10 times bigger (fatter) than the first 40 bars, divide its height by 10. Because the 42nd bar is 20 times bigger (fatter) than the first 40 bars, divide its height by 20. This

[8] As a reminder, a histogram is a bar chart, in which each of the bars represents the percent of observations in a particular x-axis bin (income, in this case). As such, the bars should be touching each other because each edge of each bar represents the cutoff level of each income category bin.

Table 14.1: 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	$69,677	
Median income	$50,054	

Source: 2011 Current Population Survey household income count data Current Population Survey (2012, Table HINC-01) is analogous to dividing the last two bars into 10 and 20 bars, respectively, and spreading frequency of each evenly among its divisions.

Exercise 14.7. Using GMM, fit the lognormal $LN(x;\mu,\sigma)$ distribution from (14.11) to the distribution of household income data using the moments from the data file. Use $\mu_0 = \ln(69677)$ and $\sigma_0 = 0.2\mu_0$ as your initial guess. Report your estimated values for $\hat{\mu}$ and $\hat{\sigma}$, as well as the value of the minimized criterion function $\boldsymbol{e}(\boldsymbol{x},\hat{\boldsymbol{\theta}})^T\,\boldsymbol{W}\,\boldsymbol{e}(\boldsymbol{x},\hat{\boldsymbol{\theta}})$. 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. Plot the histogram from exercise 14.6 overlayed with a line representing the implied histogram from your estimated lognormal (LN) 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.

Exercise 14.8. Using GMM, fit the gamma $GA(x;\alpha,\beta)$ distribution from (14.12) 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.[9] 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 exercise 14.7. Plot the histogram from exercise 14.6 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.

Exercise 14.9. Plot the histogram from exercise 14.6 overlayed with the line representing the implied histogram from your estimated lognormal (LN) distribution from exercise 14.7 and the line representing the implied histogram from your estimated gamma (GA) distribution from exercise 14.8. What is the most precise way to tell which distribution fits the data the best?

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