Problem Set 2: Estimation frameworks

Mitch Downey
Econometrics I

January 27, 2024

1 Problems

1. *Extraordinary Least Squares.* Consider the class of linear models: $Y = X\beta + \varepsilon$ where Y is an $n \times 1$ vector of outcomes, X is an $n \times k$ matrix of characteristics, β is a $k \times 1$ parameter vector, and ε is an $n \times 1$ vector of innovations or errors. Throughout this problem, A' denotes the transpose of some matrix A.

   (a) *Least squares.* Derive the $\hat{\beta}$ that minimizes the sum of squared errors: $\varepsilon'\varepsilon$.

   (b) *MLE.* Assume each error ε_i is mean zero iid normally distributed: $\varepsilon_i \sim N(0, \sigma^2)$. Find the maximum likelihood estimate of β.

   (c) *Bayesian.* For this question, let $k = 1$. Assume each error ε_i is mean zero iid normally distributed: $\varepsilon_i \sim N(0, \sigma^2)$, where we assume that σ^2 is known (for simplicity). Your prior is that $\beta \sim N(\theta, \tau^2)$. Find the mean of the posterior distribution of β.

   (d) *Bayesian.* Assume each error ε_i is mean zero iid normally distributed: $\varepsilon_i \sim N(0, \sigma^2)$, where we assume that σ^2 is known (for simplicity). You don't want your prior to unduly influence your results, so you assume a uniform prior $\beta \sim U[-a, a]$ and take the limit as $a \to \infty$. This means that your prior considers any value between $-a$ and a to be equally plausible, and you are taking the limit as this includes all possible values. Find the limit of the mode of the posterior distribution of β.

   (e) *GMM.* Assume that the errors are uncorrelated with the regressors (equivalently, all regressors are exogenous): $E(X'\varepsilon) = 0$. Derive the GMM estimate of β.

2. *Actually using MLE.*

   (a) You are interested in generating a teacher-level estimate of the propensity to inflate grades.[1] All students are given an identical test. Based on the test design, the number of points a student receives is known to be distributed according to a Poisson

[1] This question was inspired by Diamond and Persson (2016), "The long-term consequences of teacher discretion in grading of high-stakes tests", *NBER Working Paper 22207*.

distribution with parameter λ. Assume λ is known because of special test-design-magic. If a student gets a score of 20 or more, s/he will pass the exam and graduate. Grading is done centrally (i.e., not by the students' teacher), but upon seeing the grades, the teacher can appeal to have the exam re-reviewed by another grader. This re-review process is known to be generous: In expectation, it increases the score received on the exam, although it varies by how much. Assume that the number of additional points the student receives, conditional on a re-review, is drawn from $\{0, 1\}$ with equal probabilities of each (i.e., a discrete version of a uniform distribution). Assume that *i*) no teacher would bother appealing for a re-review of an exam with fewer than 19 points (since the student would never pass) or more than 20 points (since the student has already passed), and *ii*) a teacher realizes that if she appeals every exam with 19 points, s/he will get in trouble for gaming the system, and so s/he randomizes and only sends p share of those exams out for a re-review. You are interested in estimating p for each teacher, but you only observe the final grades (post re-review) and you don't observe how many or which exams were sent out for re-review.

i. Write the likelihood function for the problem.
ii. Write the score.
iii. Determine the maximum likelihood estimate $\hat{p}$.
iv. Derive the Fisher Information Matrix for $\hat{p}$ (note it's a 1×1 "matrix" because there's only one parameter).
v. At what value of λ is the Fisher Information Matrix maximized? Why? Interpret what this means.

(b) You are interested in the parameters governing Calvo pricing. Calvo pricing asserts that firms cannot necessarily adjust their prices in response to changes in costs. Instead, in each period, with probability ϕ, a firm is randomly selected by the "Calvo fairy" to be allowed to change its prices. Note that by "random selection," I mean that being selected by the fairy is independent of other firm characteristics: It is a constant probability for all firms, regardless of prices or cost shocks. Assume that *i*) there is no aggregate inflation, only idiosyncratic firm-level cost shocks in which each firm, each period, draws a cost change from $N(0, \sigma^2)$, and *ii*) when selected, a firm sets its price equal to its marginal cost.[2] You observe only prices, not costs.

i. You are interested in estimating ϕ and σ^2. Why are those parameters interesting, from an economic perspective?
ii. Assume that at time t, the Calvo fairy was feeling generous, and every firm was allowed to set price equal to marginal cost. At $t + 1$, the fairy's behavior returned to normal and only a random ϕ share of firms were allowed to. You observe price changes at $t + 1$. Write the likelihood function for the problem.
iii. Assume that the Calvo fairy wasn't generous at time t, and that it has always been only a randomly selected ϕ share of firms who could change prices. Write the likelihood function for the problem.

[2] Note that these two assumptions rule out a lot of the important economic issues around inflation expectations that macroeconomists are interested in with these models.

iv. Return to the likelihood function from ii (the one with the generous fairy, not iii with the mean one). Write the score.

v. Determine the maximum likelihood estimates $\hat{\phi}$ and $\hat{\sigma}^2$.

vi. Derive the Fisher Information Matrix and check that it is positive semi-definite.

vii. In our model, all of the costs are passed through to prices. Assume instead that only a share $\gamma \in [0, 1]$ is passed on (with γ being the same for all firms). Can you estimate ϕ, σ^2, and γ? Why or why not? In your answer, use a word that rhymes with mimentification.

3. *Misspecified MLE.* Consider a random sample $X_1, X_2, ..., X_n$ such that each X_i is iid distributed according to an exponential distribution: $X_i \sim exp(\lambda) \Leftrightarrow f(x) = \lambda e^{-\lambda x}$ if and only if $x > 0$ (and zero otherwise). Note that this implies that $E(x) = 1/\lambda$ and $Var(x) = 1/\lambda^2$. An analyst incorrectly assumes that the data comes from an iid normal distribution with mean μ and variance σ^2.

(a) Derive the true MLE for λ assuming that you know the data are exponentially distributed. Show that this estimator is biased but consistent.

(b) Let $\hat{\mu}$ denote the analyst's MLE estimator for the mean. Show that this is a consistent estimator for $E(x)$ despite the misspecification.

(c) Let $\hat{\sigma}^2$ denote the analyst's MLE estimator for the variance. Show that this is a consistent estimator for $Var(x)$ despite the misspecification.

(d) A property of the normal distribution is that the probability of an observation falling below the mean is .5: $F(\mu) = 1/2$. Propose an estimator for the probability that an observation falls below the mean, and show that it is consistent. Discuss how the analyst might use the estimates from your estimator to asess her assumptions about the distribution from which the data is drawn.

(e) Simulations. For each $n \in \{50, 100, 250, 1000\}$, simulate n observations from the exponential distribution with $\lambda = 1$. Let $F_{exp}(x|\lambda)$ denote the CDF of the exponential distribution and $F_{norm}(x|\mu, \sigma^2)$ denote the CDF of the normal distribution. Consider four estimators – each a choice of $\hat{F}(\tilde{x})$ – for $Pr(x < \tilde{x})$ for fixed $\tilde{x}$: $F_{exp}(\tilde{x}|\hat{\lambda}_{MLE} = 1/\bar{x})$, $F_{exp}(\tilde{x}|\hat{\lambda}_2 = \frac{1}{2\bar{x}} + \frac{1}{2sd(x)})$,[3] $F_{norm}(\tilde{x}|\hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE})$, and the share of the sample observed below $\tilde{x}$ (where $\hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE}$ are the MLE estimators for the mean and variance, respectively, of a normal distribution). For each n and each $\tilde{x} \in \{1/2, 1, 2, 3\}$, conduct 500 iterations under each set of conditions, and compare these three estimators by plotting the empirical average bias (the sample mean of $\hat{F}(\tilde{x}) - F_{exp}(\tilde{x}|\lambda = 1)$), the empirical variance (across the 500 iterations), and the MSE ($bias^2 + var$), all separately by n. Interpret these results. What do you learn?

[3] The mean of the exponential distribution is $1/\lambda$ and the variance is $1/\lambda^2$. Thus, unlike the normal distribution (where the centered second moment is independent of the first moment) with the exponential distribution, the second moment is a function of the first moment (other distributions are like this; the Poisson is an example). Thus, empirically, both the first and second (centered) moment are informative about the same parameter, unlike in the case of the normal distribution. So one estimator of λ is $1/\bar{x}$ and another is $\sqrt{1/Var(x)} = 1/sd(x)$. The estimator $\hat{\lambda}_2$ above is the simple mean of these two potential estimators.

4. *Actually using GMM.* Hansen and Singleton (1982) analyze an asset pricing model in order to show that GMM can be used to estimate rational expectations models that aren't linear (unlike other methods).[4] Here, we analyze a simplified version of their model. Suppose a representative agent maximizes expected discounted lifetime utility subject to a budget constraint. Assume a constant relative risk aversion utility function $U(c) = (c^\gamma - 1)/\gamma$. The agent's problem is to solve the following problem:

$$\max_{c_t, I_t} \; E_t\left[\sum_{\tau=0}^{\infty} \beta^\tau U(c_{t+\tau})\right]$$
$$s.t. \;\; c_t + I_t \leq r_t I_{t-1} + w_t \;\forall t$$

where $E_t(\cdot)$ denotes expectations given information available at time t, c_t is consumption, I_t is investment, r_t is the return on the single asset, and w_t is the wage. We assume all of these are observable to both the agent and the econometrician, for each t. However, we assume that the return to assets is drawn iid from some distribution. We assume that households are price takers, and so the return at time $t+1$ is independent of all quantities at time t (including c_t, w_t, r_t). We are interested in estimating β and γ.

(a) Take the first order condition with respect to c_t. This generates the Euler equation, which can be written as saying that the marginal utility of consumption today is equal to the discounted expected marginal utility of consumption tomorrow (given expectations about the returns to investment today). Write this Euler equation.

(b) Moment conditions are written as $E\big(g(\theta, x)\big) = 0$, where θ is a parameter vector. Rewrite this Euler equation as a moment condition. You have one moment condition and two parameters. You cannot solve this model; it is under-identified.

(c) As noted above, we assume r_{t+1} is independent of c_t, w_t, r_t. Thus, it is uncorrelated with marginal utility in period t. Write this as a moment condition. Now you have two moment conditions and two parameters. You can solve this model; it is just-identified.

(d) Let $g_b(\theta, x)$ be the moment condition you solved for in (b) above. Let $g_c(\theta, x)$ be the moment condition you solved for in (c) above. The gradient can be written as:

$$\begin{pmatrix} \frac{\partial g_b(\theta,x)}{\partial \beta} & \frac{\partial g_b(\theta,x)}{\partial \gamma} \\ \frac{\partial g_c(\theta,x)}{\partial \beta} & \frac{\partial g_c(\theta,x)}{\partial \gamma} \end{pmatrix}$$

Solve for the gradient.

(e) Download data from this dropbox link: `https://www.dropbox.com/scl/fi/7sh1te900ken36sydge23/gmmdata.csv?rlkey=k0iqjqcvp16x1tz4jyj54ut76&dl=0`. Use the gmm package in R to solve for β, γ (see `https://cran.r-project.org/web/packages/gmm/vignettes/gmm_with_R.pdf` for helpful documentation).

[4] Hansen, Lars Peter, and Kenneth J. Singleton. "Generalized instrumental variables estimation of nonlinear rational expectations models." *Econometrica* (1982): 1269-1286.