Applications of optimization

PAUL SCHRIMPF
OCTOBER 17, 2012
UNIVERSITY OF BRITISH COLUMBIA
ECONOMICS 526

The last two lectures have studied unconstrained and constrained optimization in some detail. Today, we will go over some examples of the use of the results from those two lectures.

1. Extremum estimators

Most estimators is econometrics are solutions to optimizations problems. For example, ordinary least squares solves

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^k} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2$$

where we have data on some outcome $y_i \in \mathbb{R}$ and some covariates, $x_i \in \mathbb{R}^k$.

Maximum likelihood estimators are the maximizer of the likelihood of y given x. That is, if we have some model that says the density of y given x is $f(y|x;\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^k$ are some unknown parameters, then the likelihood function is

$$\mathcal{L}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} f(y|x;\theta).$$

The maximum likelihood estimate of θ is

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}_n(\theta).$$

A third class of estimators that can be written as the solution to optimization problems are generalized method of moments (GMM) estimators. Suppose we have an economic model that perhaps does not fully specify the density of y given x, but does tell us that some moments of functions of y and x are zero. That is,

$$\mathrm{E}[m(y,x;\theta)] = 0$$

where $m(y,x;\theta) \in \mathbb{R}^m$. Then if $m \geq k$, we could estimate θ by solving these moment equations for θ. However, we do not know the true expectation $\mathrm{E}[m(y,x;\theta)]$, and we must replace it with the empirical expectation,

$$\mathbb{E}_n[m(y,x;\theta)] \equiv \frac{1}{n} \sum_{i=1}^{n} m(y_i, x_i;\theta).$$

Since $\mathbb{E}_n$ will differ from E there may not be any θ that exactly solves

$$\mathbb{E}_n[m(y,x;\theta)] = 0 \tag{1}$$

even if there is a solution with the true expectation. For this reason, we estimate θ by finding $\hat{\theta}$ that approximately solves (1). We want the best approximate solution, so one way to choose the best is to set

$$\hat{\theta} = \arg\min \mathbb{E}_n[m(y,x;\theta)]^T \mathbb{E}_n[m(y,x;\theta)].$$

More generally, we could say

$$\hat{\theta} = \arg\min \mathbb{E}_n[m(y,x;\theta)]^T W \mathbb{E}_n[m(y,x;\theta)],$$

where W is some positive definite weighting matrix.

Each of the three examples above have the common feature that they can be written as

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta)$$

where Q_n is some data dependent objective function. Estimators with this form are called extremum estimators. Given an extremum estimator, we would like to be know what $\hat{\theta}$ is estimating (i.e. what it converges to), and the asymptotic distribution of $\hat{\theta}$ so that we can perform inference (hypothesis tests, confidence intervals, etc). Suppose $Q_n(\theta)$ converges in probability to some fixed limit $Q_0(\theta)$. This means that as the sample size $n \to \infty$, $Q_n(\theta)$ gets closer and closer to $Q_0(\theta)$. The data is random, so $Q_n(\theta)$ is random as well, so the definition of convergence of limits that we have been using isn't appropriate. Instead, convergence in probability means that for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}(|Q_n(\theta) - Q_0(\theta)| > \epsilon) = 0.$$

I would guess that you have already learned about convergence in probability in you econometrics course, so we will not focus on it here.

Anyway, if

$$\theta_0 = \arg\min_{\theta\in\Theta} Q_0(\theta)$$

then with some additional assumptions we can show that $\hat{\theta}_n \xrightarrow{p} \theta_0$. In particular,

Theorem 1.1 (Consistency for extremum estimators)**.** *If (i) $Q(\theta)$ is uniquely minimized at the true parameter value θ_0, (ii) Θ is compact, (iii) $Q(\cdot)$ is continuous, and (iv) $\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0$, then $\hat{\theta}_n \xrightarrow{p} \theta_0$.*

Proof. See Newey and McFadden (1994). □

We will not go over the proof of this theorem, nor do you need to remember it for this course.

From your econometrics course, you probably know that for OLS $\hat{\beta}$ is asymptotically, meaning that the distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ converges to a normal distribution as n gets large. We denote this as

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V).$$

We can develop an analogous result for extremum estimators. Suppose Q_n is twice continuously differentiable in Θ and $\hat{\theta}$ is in the interior of Θ. $\hat{\theta}$ must satisfy the first order condition because it is in the interior and Q_n is differentiable.

$$0 = DQ_{n\hat{\theta}}$$

We want to end up with something like $\sqrt{n}(\hat{\theta} - \theta_0)$, so let's take a mean value expansion of DQ_n around θ_0, then

$$\begin{aligned} 0 =& DQ_{n\hat{\theta}} \\ =& DQ_{n\theta_0} + D^2 Q_{n\bar{\theta}}(\hat{\theta} - \theta_0) \end{aligned}$$

If $D^2 Q_n \bar{\theta}$ is invertible then,

$$(\hat{\theta} - \theta_0) = -(D^2 Q_{n\bar{\theta}})^{-1} DQ_{n\theta_0}$$

Generally, $D^2 Q_n \xrightarrow{p} D^2 Q$ and $DQ_n \xrightarrow{p} DQ$, but this is not what we want. We want to get a distribution for $\hat{\theta} - \theta_0$, not a point limit. Therefore, we should multiple by some a_n to make $a_n(\hat{\theta} - \theta_0)$ converge to a non-degenerate distribution. The right choice of a_n depends somewhat on the details of the problem, but the vast majority of the time the right choice is $a_n = \sqrt{n}$. This is the right choice because, you can often show that

$$\sqrt{n} DQ_{n\theta_0} \xrightarrow{d} N(0, V). \tag{2}$$

This fact follows from a central limit theorem. For example, it is often the case that

$$DQ_{n\theta_0} = \frac{1}{n} \sum_{i=1}^{n} D_\theta f(y_i, x_i, \theta_0).$$

Then, if observations are independent and identically distributed, $\mathrm{E}[D_\theta f(y_i, x_i, \theta)] = 0$, and $\mathrm{E}[f(y, x, \theta_0)^2]$ is finite then the classical central limit theorem says that (2) holds[1]. Regardless of why, if we just assume (2), then

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) =& D^2 Q_{n\bar{\theta}}(\sqrt{n} DQ_{n\theta_0}) \\ & \xrightarrow{d} N\left(0, (D^2 Q_{n\theta_0})^{-1} V (D^2 Q_{n\theta_0})^{-1}\right). \end{aligned}$$

1.1. **With constraints.** Sometimes we have an extremum estimator with constraints. Suppose we have a constrained extremum estimate of the form

$$\hat{\theta} = \arg\min_{\theta \in \Theta} Q_n(\theta) \text{ s.t. } h_n(\theta) = 0.$$

As above, we will assume that $Q_n \xrightarrow{p} Q$ and $h_n \xrightarrow{p} h$. We will let θ_0 be the solution to

$$\theta_0 = \arg\min_{\theta \in \Theta} Q(\theta) \text{ s.t. } h(\theta) = 0.$$

[1] There are other types of central limit theorems with different assumptions. For example, independence can be relaxed and the finite variance condition strengthened, and the conclusion still holds.

We will assume that θ_0 is a unique strict minimizer in the interior of Θ. As above, $\hat{\theta}$ solve the first order condition:

$$0 = \begin{pmatrix} DQ_{n\hat{\theta}} - \hat{\mu}^T Dh_{n\hat{\theta}} \\ h_n(\hat{\theta}) \end{pmatrix}$$

If we expand around θ_0 and μ_0 we get

$$0 = \begin{pmatrix} DQ_{n\theta_0} - \mu_0^T Dh_{n\theta_0} \\ h_n(\theta_0) \end{pmatrix} + \begin{pmatrix} D^2Q_{n\bar{\theta}} - \sum_{j=1}^m \bar{\mu}_j D^2 h_{jn\bar{\theta}} & Dh_{n\bar{\theta}}^T \\ Dh_{n\bar{\theta}} & 0 \end{pmatrix} \begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\mu} - \mu_0 \end{pmatrix}$$

Using the same sort of reasoning as in the previous section, we could show that

$$\sqrt{n} \begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\mu} - \mu_0 \end{pmatrix} \xrightarrow{d} N(0, V)$$

for some V.

2. A MODEL OF POLLUTION

Let's analyze a simple model of pollution. Suppose there is a representative firm that takes one input, ℓ, and produces two outputs, a consumption good c, and pollution z. The firm allocates its input to two activities: production and pollution reduction. If it devotes α portion of ℓ to production, then it produces $c = f_c(\alpha\ell)$ and $z = f_z(\alpha\ell) - g((1-\alpha)\ell)$ We will assume that $f_c' > 0$, $f_z' > 0$ and $g' > 0$.

There is also a representative consumer with an endowment of L units of ℓ and preferences over consumption and pollution represented by a utility function $u : \mathbb{R}^2 \to \mathbb{R}$. We will assume that $\frac{\partial u}{\partial c} > 0$ and $\frac{\partial u}{\partial z} < 0$.

2.1. **Optimal allocation.**

First let's solve the social planner's problem and find the optimal allocation in this economy. The social planner maximizes utility subject to the production function.

$$\begin{aligned} \max_{c,\alpha,\ell,z} u(c,z) \text{ s.t. } \ell &= L \\ c &= f_c(\alpha\ell) \\ z &= f_z(\alpha\ell) - g((1-\alpha)\ell) \\ 0 \leq \alpha &\leq 1 \end{aligned}$$

Now we could just substitute in the constraints and have an unconstrained maximization over α, but the multipliers are going to be useful later, so we will work with them. The Lagrangian is

$$\begin{aligned} L(c, a, \ell, z, \mu) =& u(c,z) - \mu_\ell(\ell - L) - \mu_c\left(c - f_c(\alpha\ell)\right) - \\ & - \mu_z\left(f_z(\alpha\ell) - g((1-\alpha)\ell)\right) - \mu_{\alpha_1}(\alpha - 1) + \mu_{\alpha_0}\alpha \end{aligned}$$

The first order conditions are

$$
\begin{aligned}
[c]: && \frac{\partial u}{\partial c} =& \mu_c && (3)\\
[z]: && \frac{\partial u}{\partial z} =& \mu_z && (4)\\
[\ell]: && -\mu_\ell + \mu_c f_c'(\alpha\ell)\alpha + \mu_z \left(f_z'(\alpha\ell)\alpha - g'((1-\alpha)\ell)(1-\alpha)\right) =& 0 && (5)\\
[\alpha]: && \mu_c f_c'(\alpha\ell)\ell + \mu_z \left(f_z'(\alpha\ell)\ell + g'((1-\alpha)\ell)\ell\right) - \mu_{\alpha_1} + \mu_{\alpha_0} =& 0 && (6)
\end{aligned}
$$

As well as the first three constraints. We will assume the last constraint does not bind. If we combine (3), (4), and (6) we get

$$
\frac{\partial u}{\partial c} f_c'(\alpha\ell) + \frac{\partial u}{\partial z}\left(f_z'(\alpha\ell) + g'((1-\alpha)\ell)\right) = 0 \tag{7}
$$

This characterizes the optimal choice of α^*. Under reasonable assumptions, it will be true that $\ell^* = L$. Then, once we know α^* we can find c^* and z^* from the constraints. (7) says that the marginal benefit of devoting more resource to producing c to be equal to the marginal disutility of devoting fewer resources to preventing pollution.

2.2. **Competitive equilibrium.** Let the price of c be p and ℓ be w. Pollution, z, is not traded. The firm's problem is

$$
\max_{c,\alpha,\ell,z} p f_c(\alpha\ell) - w\ell \text{ s.t. } 0 \leq \alpha \leq 1
$$

Let the multipliers be λ_0 and λ_1. The first order conditions are:

$$
\begin{aligned}
[\ell]: && p f_c'(\alpha\ell)\alpha - w =& 0\\
[\alpha]: && p f_c'(\alpha\ell)\ell + \lambda_0 - \lambda_1 =& 0\\
[\lambda_1]: && \lambda_1(\alpha - 1) =& 0\\
[\lambda_0]: && \lambda_0 \alpha =& 0
\end{aligned}
$$

Since $f_c' > 0$ and α only enters the objective function, the maximum must have $\alpha = 1$ and $\lambda_0 = 0$.

The consumer's problem is

$$
\begin{aligned}
\max_{c,\ell} u(c,z) \text{ s.t. } & pc \leq w\ell\\
& \ell \leq L
\end{aligned}
$$

The Lagrangian is

$$
L(c,\ell,\psi) = u(c,z) - \psi_c(pc - w\ell) - \psi_\ell(\ell - L)
$$

The first order conditions are

$$
\begin{aligned}
[c] && \frac{\partial u}{\partial c} - p\psi_c =& 0\\
[\ell] && w\psi_c - \psi_\ell =& 0\\
[\psi_c] && \psi_c(pc - w\ell) =& 0\\
[\psi_\ell] && \psi_\ell(\ell - L) =& 0.
\end{aligned}
$$

Since u is increasing in c and if p and w are positive, then the first constraint must bind, so $\psi_c > 0$. From the envelope theorem, ψ_c is the marginal value of increase $w\ell$. If this is positive, then the consumer will want ℓ as large as possible, so $\ell = L$. Then from the budget constraint, $\frac{p}{w} = \frac{L}{c}$. Also, given the firm's problem has $\alpha = 1$, it must be that $c = f_c(L)$. Also, $z = f_z(L) - g(0)$. Given our assumption that $f_z' > 0$ and $g' > 0$, z is greater in this competitive equilibrium than in the optimal allocation.

2.3. **Competitive equilibrium with taxes.** Suppose the production of pollution is taxed at rate τ. Also, assume that the government gives all tax revenue to the consumer as a lump sum transfer. In this case, the firm's problem becomes

$$\max_{c,\alpha,\ell,z} p f_c(\alpha\ell) - w\ell - \tau\left(f_z(\alpha\ell) - g((1-\alpha)\ell)\right) \text{ s.t. } 0 \leq \alpha \leq 1$$

Let the multipliers be λ_0 and λ_1. The first order conditions are:

$$\begin{aligned}
[\ell]: &\quad p f_c'(\alpha\ell)\alpha - w - \tau\left(f_z'(\alpha\ell)\alpha - g((1-\alpha)\ell)(1-\alpha)\right) = 0 \\
[\alpha]: &\quad p f_c'(\alpha\ell)\ell - \tau\left(f_z'(\alpha\ell)\ell + g'((1-\alpha)\ell)\ell\right) + \lambda_0 - \lambda_1 = 0 \\
[\lambda_1]: &\quad \lambda_1(\alpha - 1) = 0 \\
[\lambda_0]: &\quad \lambda_0\alpha = 0
\end{aligned}$$

Notice that $[\ell]$ is nearly the same as the social planner's first order condition for ℓ (5), except the former has p, w, and τ whereas the later has μ_c, μ_ℓ, and μ_z. Suppose the government sets the tax rate to

$$\tau = -\mu_z = -\frac{\partial u}{\partial z}(c^*, z^*)$$

. If $p = \mu_c$ and $w = \mu_\ell$, then α^* and c^* solve the firm's first order condition with taxes. To show this is an equilibrium, we just need to verify that the consumer's first order conditions hold. The consumer's problem is now

$$\max_{c,\ell} u(c,z) \text{ s.t. } pc \leq w\ell + T$$
$$\ell \leq L$$

The first order conditions are

$$[c] \qquad \frac{\partial u}{\partial c} - p\psi_c = 0 \tag{8}$$
$$[\ell] \qquad w\psi_c - \psi_\ell = 0 \tag{9}$$
$$[\psi_c] \qquad \psi_c(pc - w\ell - T) = 0 \tag{10}$$
$$[\psi_\ell] \qquad \psi_\ell(\ell - L) = 0. \tag{11}$$

From (3), $\frac{\partial u}{\partial c} = \mu_c = p$, so it must be that $\psi_c = 1$. Then from (9), $\psi_\ell = 1$ as well. Finally, integrating the firm's first order condition for ℓ with respect to ℓ, and using the fact that $\tau z = T$, we see that the consumer's budget constraint must hold.

Thus, by appropriately setting the pollution tax, we could achieve the optimal allocation. This sort of result–that certain taxes can achieve first (or sometimes second) best outcomes—appears quite often in public finance. Also, this approach—finding the social planner's solution, and showing that setting prices and taxes equal to certain multipliers give a competitive equilibrium with taxes—is the standard way of showing it.

This has gotten long enough, but a useful exercise might be to consider a cap and trade program as well. Suppose the consumers are given tradeable permits that allow up to Z units of pollution. You can show that if $Z = z^*$, then the price of the permits will end up being μ_z, and $p = \mu_c$ and $w = \mu_\ell$[2] as above, so this is another way of decentralizing the efficient allocation.

[2] Actually only relative prices are determined, and this is one possible set of prices. Nonetheless all possible prices led to the same allocation.