M1 INTERMEDIATE ECONOMETRICS

SEMIPARAMETRIC ESTIMATION

Koen Jochmans

November 13, 2025

1. Nonlinear regression

Suppose that

$$\mathbb{E}(Y|X = x) = m(x; \theta_0)$$

where m is a known function. Let $e = Y - m(X; \theta_0)$. To use maximum likelihood we would need to specify the distribution of $e|X = x$ up to a finite set of parameters. Suppose that we are primarily interested in the parameter θ_0. This suffices to evaluate average marginal effects, for example. In that case we may want to devise an estimator of θ_0 that only relies on the fact that

$$\mathbb{E}(e|X = x) = \mathbb{E}(Y - m(X; \theta_0)|X = x) = 0 \qquad \text{(a.s.)}, \tag{1.1}$$

that is, on the correct specification of the conditional mean function alone. An example where we did this was the linear regression model,

$$m(x; \theta) = x'\theta,$$

where we known that an ordinary least-squares approach gives a consistent and asymptotically-normal estimator of θ_0 without strong restrictions on $e|X = x$.

1.1. Nonlinear least squares

A natural extension of the ordinary least-squares idea is nonlinear least squares. The population problem is

$$\min_{\theta} \mathbb{E}\left((Y - m(X;\theta))^2\right).$$

and because

$$\begin{aligned}
\mathbb{E}\left(\{Y - m(X;\theta)\}^2\right) &= \mathbb{E}\left(\{(Y - m(X;\theta_0)) + (m(X;\theta_0) - m(X;\theta))\}^2\right) \\
&= \mathbb{E}(e^2) + \mathbb{E}\left(\{m(X;\theta_0) - m(X;\theta)\}^2\right)
\end{aligned}$$

we know that this problem is minimized at any θ for which

$$\mathbb{E}\left(\{m(X;\theta_0) - m(X;\theta)\}^2\right) = 0,$$

which obviously includes $\theta = \theta_0$. Furthermore, θ_0 is the unique minimizer if and only if

$$\mathbb{P}(m(X,\theta) \neq m(X,\theta_0)) > 0$$

for all $\theta \neq \theta_0$. In the linear model this is the no-multicolinearity condition because, there,

$$\mathbb{E}\left((m(X,\theta_0) - m(X,\theta))^2\right) = (\theta - \theta_0)'\mathbb{E}(XX')(\theta - \theta_0)$$

and by linear independence $\alpha'\,\mathbb{E}(XX')\,\alpha > 0$ for any $\alpha \neq 0$ when $\mathbb{E}(XX')$ is positive definite.

Under differentiability of the function m the first-order condition for the least-squares problem is

$$\mathbb{E}\left(\frac{\partial m(X;\theta)}{\partial\theta}(Y - m(X;\theta))\right) = 0.$$

Again, in the linear model $\partial m(x;\theta)/\partial\theta = x$, and so the moment condition becomes $\mathbb{E}(X(Y - X'\theta)) = 0$, which is are least-squares normal equation from before.

The nonlinear least-squares estimator is then

$$\arg\min_{\theta} \sum_{i=1}^{n} (Y_i - m(X_i;\theta))^2.$$

Using the same type of arguments as for maximum likelihood it should be easy to see that (under regularity conditions), the estimator, say $\hat{\theta}$, satisfies

$$\sqrt{n}(\hat{\theta} - \theta_0) \underset{d}{\rightarrow} N(V_\theta)$$

for

$$V_\theta = \mathbb{E}\left(\tfrac{\partial m(X;\theta_0)}{\partial\theta}\tfrac{\partial m(X;\theta_0)}{\partial\theta'}\right)^{-1} \mathbb{E}\left(\tfrac{\partial m(X;\theta_0)}{\partial\theta}\tfrac{\partial m(X;\theta_0)}{\partial\theta'}e^2\right) \mathbb{E}\left(\tfrac{\partial m(X;\theta_0)}{\partial\theta}\tfrac{\partial m(X;\theta_0)}{\partial\theta'}\right)^{-1}$$

as $n \to \infty$. An estimator of the asymptotic variance follows in the usual way.

1.2. Generalized method of moments

Nonlinear least squares is not the only estimator that can be constructed under (1.1), and it usually will not be the efficient choice either. In fact, by the law of iterated expectations, (1.1) implies that

$$\mathbb{E}(\varphi(X;\theta_0)(Y - m(X;\theta_0)) = 0$$

for any (vector) function φ. For any choice of such function this leads to the method-of-moment estimator

$$\arg\operatorname{solve}_\theta \sum_{i=1}^n \varphi(X_i;\theta)\,(Y_i - m(X_i;\theta)) = 0$$

when the problem is just identified, or to the generalized method-of-moment estimator

$$\arg\min_\theta \left(\sum_{i=1}^n \varphi(X_i;\theta)\,(Y_i - m(X_i;\theta))\right)' W \left(\sum_{i=1}^n \varphi(X_i;\theta)\,(Y_i - m(X_i;\theta))\right),$$

for a chosen full-rank weight matrix W, when the problem is overidentified. We know that the optimal choice for W for a given choice of φ is the inverse of the variance of the moment conditions. However, in the conditional-moment problem (1.1) there are also many possible choices for φ. We discuss this next.

1.3. Generalized least squares

It turns out that the optimal choice for φ is

$$\varphi(x;\theta) = -\frac{\partial m(x;\theta)}{\partial\theta}\,\frac{1}{\sigma^2(x)},$$

where $\sigma^2(x) = \mathbb{E}(e^2|X=x)$. Note that this returns a just-identified system. The estimating equation becomes

$$\sum_{i=1}^n \frac{\partial m(X_i;\theta)}{\partial\theta}\,\frac{(Y_i - m(X_i;\theta))}{\sigma^2(X_i)} = 0.$$

Furthermore (up to the factor -2), this is the first-order condition of the minimization problem

$$\min_{\theta} \sum_{i=1}^{n} \frac{(Y_i - m(X_i;\theta))^2}{\sigma^2(X_i)},$$

which is generalized nonlinear least squares.

If the function $\sigma^2(x)$ is known the asymptotic distribution of the feasible estimator follows from the arguments above. In this case the asymptotic variance becomes

$$\mathbb{E}\left(\frac{\partial m(X;\theta_0)}{\partial \theta}\frac{1}{\sigma^2(X)}\frac{\partial m(X;\theta_0)}{\partial \theta'}\right)^{-1}.$$

Note how this no longer has a sandwich structure. Under homoskedasticity it simplifies to

$$\sigma^2\,\mathbb{E}\left(\frac{\partial m(X;\theta_0)}{\partial \theta}\frac{\partial m(X;\theta_0)}{\partial \theta'}\right)^{-1}.$$

This is the direct nonlinear counterpart to the optimality result for ordinary least squares; indeed, when $m(x;\theta) = x'\theta$ we have $\partial m(x;\theta)/\partial\theta = x$ and so the variance reduces to

$$\sigma^2\,\mathbb{E}(XX')^{-1}.$$

1.4. Example: Exponential regression

For count data popular conditional distributions are the poisson distribution and the negative-binomial distribution. The conditional mean is usually modelled as

$$m(x;\theta) = e^{x'\theta}.$$

This type of conditional mean is also popular for more general non-negative outcomes, such as quantities, and appears in the estimation of trade flows and production functions, for example. As such variables are continuous, there, the poisson and negative binomial are not suitable candidates to construct a likelihood.

The nonlinear least-squares estimator solves

$$\sum_{i=1}^{n} X_i e^{X_i'\theta}(Y_i - e^{X_i'\theta}) = 0.$$

The optimal moment condition depends on the conditional variance. The poisson distribution, for example, has the property that its mean is equal to its variance, i.e., $\sigma^2(x) = e^{x'\theta}$. The maximum-likelihood score for the poisson problem is thus

$$\sum_{i=1}^{n} X_i(Y_i - e^{X_i'\theta}) = 0,$$

and corresponds to the optimal moment equation as long as both the mean and variance are equal to $e^{x'\theta}$. In this sense, the estimating equation has a certain robustness property similar to the maximum-likelihood estimator of the regression slopes in the classical linear regression model. On the other hand, for nonlinear least squares to be optimal in this context we would need that $\sigma^2(x)$ does not depend on x.

2. General moment problems

2.1. Example

Consider the following simplified version of a CAPM. The problem is to choose an (infinite) consumption stream $\{c_i\}$ in order to maximize expected discounted utility stream,

$$\mathbb{E}\left(\sum_{h=0}^{\infty} \alpha_0^h\, u(c_{i+h}, \beta_0)\,\middle|\, I_i\right),$$

where I_i is the information set at time period i. Here, u is a chosen utility function parametrized by β_0, α_0 is the discount rate, and we aim to learn $\theta_0 = (\alpha_0, \beta_0)$. The risk free return is r (taken to be constant over time here). To estimate θ_0 we need to obtain estimating equations. One way of doing so is by exploiting the fact that, at the optimum, we must have that the Euler equation

$$\alpha_0^i \frac{\partial u(c_i, \beta_0)}{\partial c} = \alpha_0^{i+1} \frac{\partial u(c_{i+1}, \beta_0)}{\partial c} r$$

holds. The intuition of this equation is the following. At the optimum one must be indifferent between consuming a small amount in period i and guarding the amount for one additional period before consuming it then. These are the left- and right-hand sides of the above equation, respectively. Hence,

$$\mathbb{E}\left(r\,\alpha_0 \frac{\partial u(c_{i+1},\beta_0)/\partial c}{\partial u(c_i,\beta_0)/\partial c} - 1 \middle| I_i \right) = 0$$

holds true.

2.2. General case

The example illustrates that an economic model will usually imply certain moment conditions that can be used to estimate the underlying parameters. The generic formulation is that

$$\mathbb{E}(g(V; \theta_0)) = 0,$$

where the (vector-valued) random variable V can be comprised of outcomes, regressors, and instruments, i.e,, $V = (Y, X, Z)$.

Let

$$G = \mathbb{E}\left(\frac{g(V; \theta_0)}{\partial \theta'} \right)$$

be the Jacobian matrix. Then the GMM estimator based on our moment conditions, and using weight matrix W, will (again under suitable regularity conditions) behave like a normal random variable with asymptotic variance

$$(GWG')^{-1}(GW\Omega WG')(GWG')^{-1}$$

for $\Omega = \mathbb{E}\left(g(V;\theta_0)g(V;\theta_0)'\right)$. The optimal choice for W is still equal to (a consistent estimator of) Ω^{-1}, which yields asymptotic variance $(G\Omega^{-1}G')^{-1}$ and will usually lead to a two-step implementation of the efficient estimator.