Semiparametric Models and Estimators

Whitney Newey

October 2007

Semiparametric Models

Data: $Z_1, Z_2, ...$ i.i.d.

Model: $\mathcal{F}$ a set of pdfs.

Correct specification: pdf f_0 of Z_i in $\mathcal{F}$.

Parametric model: $\mathcal{F} = \{f(z|\theta) : \theta \in \Theta\}$, Θ is finite dimensional.

Semiparametric model: $\mathcal{F}$ is not finite dimensional but has finite dimensional components.

Ex: Probit, $Z = (Y, X)$, $Y \in \{0, 1\}$, $\Pr(Y = 1|X) = \Phi(X'\beta_0)$

$$\mathcal{F} = \{\Phi(x'\beta)^y[1 - \Phi(x'\beta)]^{1-y}h(x) : \beta \in B, h \text{ is pdf of } X\}$$

β is parametric component, nonparametric component is $h(X)$.

Ex: Linear model $Z = (Y, X)$, $E[Y|X] = X'\beta_0$; $\mathcal{F} = \{f(z) : E_f[Y|X] = X'\beta\}$ parametric component is β, everything else nonparametric. β depends on f.

Probit pdf is an explicit function of a parametric component β and a nonparametric component $h(x)$. Linear model parameters imposes constraints on pdf.

Binary Choice with Unknown Disturbance Distribution:

$Z = (Y, X)$; Let $v(x, \beta)$ be a known function,

$$Y = 1(Y^* > 0), Y^* = v(X, \beta_0) - \varepsilon, \varepsilon \text{ independent of } X,$$

This equation implies that for the CDF $G(t)$ of ε

$$\Pr(Y = 1|X) = G(v(X, \beta_0)),$$

Here the parameter is β and everything else is nonparametric. The model can be written as an explicit function of the parametric component β and two nonparametric components.

$$\mathcal{F} = \qquad\qquad\qquad\qquad .$$

The $v(x, \beta)$ notation allows location and scale normalization, e.g. $v(x, \beta) = x_1 + x_2'\beta$, $x = (x_1, x_2')'$, x_1 scalar. Generalization to nonmonotonic $G(\cdot)$.

Censored Regression with Unknown Disturbance Distribution: $Z = (Y, X)$,

$$Y = \max\{0, Y^*\}, Y^* = X'\beta_0 + \varepsilon,\ med(\varepsilon|X) = 0;$$

Parameter β, everything else, including distribution of ε, is nonparametric.

$$\mathcal{F} = \qquad\qquad\qquad\qquad\qquad\qquad .$$

Binary choice and censored regression are limited dependent variable models. Semiparametric models are important here because misspecifying the distribution of the disturbances leads to inconsistency of MLE.

Partially Linear Regression: $Z = (Y, X, W)$,

$$E[Y|X, W] = X'\beta_0 + g_0(W).$$

Parameter β, everything else nonparametric, including additive component of regression. Can help with curse of dimensionality, with covariates X entering parametrically. In Hausman and Newey (1995) W is log income and log price, and X includes about 20 time and location dummies. X may be variable of interest and $g_0(Z)$ some covariates, e.g. sample selection.

$$\mathcal{F} = \qquad\qquad\qquad\qquad\qquad\qquad .$$

Index Regression: $Z = (Y, X)$, $v(x, \beta)$ a known function,

$$E[Y|X] = \tau(v(X, \beta_0)),$$

where the function $\tau(\cdot)$ is unknown. Binary choice model has $E[Y|X] = \Pr(Y = 1|X) = \tau(v(X, \beta_0))$, with $\tau(\cdot) = G(\cdot)$. If allow conditional distribution of ε given X to depend (only) on $v(X, \beta_0)$, then binary choice model becomes index model.

$$\mathcal{F} = \qquad\qquad\qquad\qquad .$$

Semiparametric Estimators

Estimators of β_0. Two kinds; do and do not require nonparametric estimation. Really model specific, but beyond scope to say why. One general kind of estimator:

$$\hat{\beta} = \arg\min_{\beta \in B} \sum_{i=1}^{n} q(Z_i, \beta)/n, \ \beta_0 = \arg\min_{\beta \in B} E[q(Z_i, \beta)],$$

B set of parameter values. Extremum estimator. Clever choices of $q(Z, \beta)$ in some semiparametric models.

Ex: Censored regression quantiles, Powell (1984, 1986).

$med(Y^*|X) = X'\beta_0$. $\max\{0, y\}$ is a monotonic transformation, so $med(Y|X) = \max\{0, X'\beta_0\}$,

$$\beta_0 = \arg\min_\beta E[|Y_i - \max\{0, X_i'\beta\}|]$$

Sample analog is

$$\hat{\beta} = \arg\min_\beta \sum_{i=1}^{n} |Y_i - \max\{0, X_i'\beta\}|.$$

Censored least absolute deviations estimator of Powell (1984). Only requires

$$med(Y^*|X) = X'\beta_0.$$

Allows for heteroskedasticity.

Consistency and Asymptotic Normality of Minimization Estimators

Consistency: *If i) $E[q(Z,\beta)]$ has a unique minimum at β_0, ii) $\beta_0 \in B$ and B is compact; iii) $q(Z,\beta)$ is continuous at β with probability one; iv) $E[\sup_{\beta\in B}|q(Z_i,\beta)|] < \infty$; then $\hat{\beta} = \arg\min_{\beta\in B}\sum_{i=1}^{n} q(Z_i,\beta) \xrightarrow{p} \beta_0$.*

Well known. Allows for $q(Z,\beta)$ to be discontinuous.

Asymptotic Normality: (Van der Vaart, 1995). *If $\hat{\beta} \xrightarrow{p} \beta_0$, β_0 is in the interior of B, and i) $E[q(Z_i,\beta)]$ is twice differentiable at β_0 with nonsingular Hessian H; ii) there is $d(z)$ such that $E[d(Z)^2]$ exists and for all $\beta, \tilde{\beta} \in B$, $|q(Z,\tilde{\beta}) - q(Z,\beta)| \le d(Z)\|\tilde{\beta} - \beta\|$; iii) with probability one $q(Z,\beta)$ is differentiable at β_0 with derivative $m(Z)$, then for $\Sigma = E[m(Z)m(Z)']$,*

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1}).$$

Tests and confidence intervals: Either bootstrap or estimate $H^{-1}\Sigma H^{-1}$

Estimating H : Plug into formula for H, or $\hat{H}$ is a finite difference approximation to second derivative of $\sum q(Z_i, \hat{\beta})/n$ evaluated at $\hat{\beta}$ where differences not too small.

Estimating Σ :

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} m(Z_i, \hat{\beta}) m(Z_i, \hat{\beta})'.$$

Alternative approach to estimating parameters of semiparametric models is to find $g(z, \beta)$ with

$$E[g(Z_i, \beta_0)] = 0,$$

then do GMM. Using objective function is better. Why?

Ex: Censored Least Absolute Deviations

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} |Y_i - \max\{0, X_i'\beta\}|.$$

What is corresponding moment condition?

$$g(z, \beta) =$$

When does $\hat{\beta}$ solve $\sum_{i=1}^{n} g(Z_i, \beta) = 0$.

In some examples we first find $g(Z, \beta)$ and then "integrate" to get $q(Z, \beta)$.

Estimators with Nonparametric Components

Some models require use of nonparametric estimators.

Include the partially linear and index regressions.

Generalize extremum estimator: There is a function h such that

$$\beta_0 = \arg\min_{\beta,h} E[q(Z_i, \beta, h)]$$

Does not work to solve

$$\hat{\beta} = \arg\min_{\beta,h} \sum_{i=1}^{n} q(Z_i, \beta, h)$$

In general $\hat{\beta}$ not consistent.

General approach is "profile" method or "concentrating out" h.

Find

$$h(\beta) = \arg\min_h E[q(Z_i, \beta, h)]$$

Form nonparametric estimator $\hat{h}(\beta)$ of $h(\beta)$ (e.g. locally linear or series regression).

Get

$$\hat{\beta} = \arg\min_\beta \sum_i q(Z_i, \beta, \hat{h}(\beta)).$$

Ex: Partially linear model (Robinson, 1988).

Know $E[Y|X,W]$ minimizes $E[(Y - H(X,W))^2]$ over H, so that

$$(\beta_0, h_0(\cdot)) = \arg\min_{\beta, h(\cdot)} E[\{Y_i - X_i'\beta - h(W)\}^2].$$

We can use $q(Z,\beta,h) = \{Y - X'\beta - h(W)\}^2$

What would we get if we tried to solve

$$\tilde{\beta} = \arg\min_{\beta,h} \sum_i q(Z_i,\beta,h) = \arg\min_{\beta,h} \sum_i \{Y_i - X_i'\beta - h(W_i)\}^2?$$

Instead we concentrate out h.

$$h(\beta) = \arg\min_h E[\{Y_i - X_i'\beta - h(W_i)\}^2] = ?$$

Get

$$\hat{h}(\beta, W) = ?$$

Plug this back into sample minimization problem

$$\begin{aligned}\hat{\beta} &= \arg\min_\beta \sum_i \{Y_i - X_i'\beta - \hat{h}(\beta, W_i)\}^2 \\ &= \arg\min_\beta\end{aligned}$$

For series estimator same as

$$\hat{\beta} = \arg\min_{\beta,\gamma} \sum_i \{Y_i - X_i'\beta - p^K(W_i)'\gamma\}^2.$$

Also, $\hat{H}(x, w) = x'\hat{\beta} + \hat{h}(w)$ is often of interest. Hausman and Newey (1995) w is ln(price) and ln(income) and Y is ln(demand).

Index regression, as in Ichimura (1993). By $E[Y|X] = \tau_0(v(X,\beta_0))$,

$$(\beta_0, \tau_0(\cdot)) = \arg\min_{\beta,\tau(\cdot)} E[\{Y_i - \tau(v(X_i,\beta))\}^2].$$

Why?

Concentrating out the τ, $\tau(X,\beta) =$?.

Estimator is

$$\begin{aligned} \hat{\beta} &= \arg\min_{\beta} \sum_{i=1}^{n} \{Y_i - \hat{\tau}(X_i,\beta)\}^2 \\ &= \arg\min_{\beta} . \end{aligned}$$

Asymptotic Theory: Difficult because of presence of nonparametric estimator.

Need $\hat{h}(\beta)$ to converge faster than $n^{-1/4}$ and other conditions.

One helpful result is that estimation of $h(\beta)$ does not affect asymptotic distribution.

Straightforward to estimate asymptotic variance.

$$\hat{H} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial q(Z_i, \beta, \hat{h}(\beta))}{\partial \beta}\bigg|_{\beta=\hat{\beta}}, \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} q(Z_i, \hat{\beta}, \hat{h}(\hat{\beta}))q(Z_i, \hat{\beta}, \hat{h}(\hat{\beta}))'.$$

$$\hat{V} = \hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1}$$

Under appropriate regularity conditions

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

Could also work with moment conditions, where

$$E[g(z, \beta_0, h)] = 0.$$

For exactly identified case get $\hat{h}(\beta)$ and solve

$$\sum_{i=1}^{n} g(Z_i, \beta, \hat{h}(\beta)) = 0.$$

In this case estimation of h does affect asymptotic variance.

Formulae available for kernel and series estimates.

Simplest for series: Just treat as if it were parametric and calculate two-step adjustment.

Ex: Average derivative. $h_0(X) = E[Y|X]$, $\beta_0 = E[\partial h(X)/\partial X]$;

$$g(Z, \beta, h) = \beta - \partial h(X)/\partial \beta.$$

Estimator is

$$\hat{\beta} = \frac{1}{n} \sum \frac{\partial \hat{h}(X_i)}{\partial X}.$$