Nonlinear Regression Asymptotics [1]

A. Ronald Gallant
University of North Carolina

July 1992
Revised September 1992
Corrections November 1992

[1] Handout for Econ 275.

ABSTRACT

This is terse summary of the asymptotic theory of least mean distance estimation and testing for nonlinear statistical models together with illustrative application to least squares and maximum likelihood estimation. It is less general than Gallant (1987, Chapter 3) because the true parameter may not drift, estimates may not depend on a preliminary estimate of a nuisance parameter, and the model must be correctly specified.

1 Estimation Theory

1.1 Setup

Structural model: $q(y_t, x_t, \lambda^o) = e_t$ (must be computable)

Reduced form: $y_t = Y(e_t, x_t, \lambda^o)$ (need only exist)

Distance function: $s(y, x, \lambda)$ (small when x & λ fit y well)

e_t independent, each with distribution $P(e)$

λ^o in Λ which is a closed & bounded subset of $\Re^p$

$s_n(\lambda) = (1/n) \sum_{t=1}^{n} s(y_t, x_t, \lambda)$ (sample objective function)

$\hat{\lambda}_n = \text{argmin}_\Lambda s_n(\lambda)$ (the estimator)

$s_n^o(\lambda) = (1/n) \sum_{t=1}^{n} \int s[Y(e, x_t, \lambda^o), x_t, \lambda]\, dP(e)$ (mean of sample objective function)

$\lambda^o = \text{argmin}_\Lambda s_n^o(\lambda)$ (implies correct specification)

1.2 Limits

Throughout, the expression $\lim_{n\to\infty} U_n = c$ a.s. is to be interpreted as follows. U_n is a random variable depending on $\{(e_t, x_t)\}_{t=1}^{\infty}$ and c is a constant. The sequence $\{x_t\}_{t=1}^{\infty}$ is held fixed and the probability is zero that a sequence $\{e_t\}_{t=1}^{\infty}$ occurs such that $\lim_{n\to\infty} U_n \neq c$. If x_t is random then this means that probabilities are being computed with respect to the conditional distribution of $\{e_t\}_{t=1}^{\infty}$ given $\{x_t\}_{t=1}^{\infty}$. The a.s. qualifier, meaning almost surely, is often omitted below.

1.3 Strong law of large numbers

If the errors $\{e_t\}$ are independently and identically distributed, as assumed above, then sample averages converge to population averages:

$$\lim_{n\to\infty} \left|\frac{1}{n}\sum_{t=1}^{n} g(e_t) - \int g(e)\, dP(e)\right| = 0$$

for any $g(e)$ such that $\int_{\mathcal{E}} |g(e)|\, dP(e) < \infty$.

1.4 Stability condition on x_t

Chaotic data, data obtained by replication, data obtained by sampling a distribution, etc. have the property that there is a distribution $\mu(x)$, sometimes called the design measure, such that

$$\lim_{n\to\infty} \left|\frac{1}{n}\sum_{t=1}^{n} g(x_t) - \int_{\mathcal{X}} g(x)\, d\mu(x)\right| = 0$$

for continuous $g(x)$ such that $\int_{\mathcal{X}} |g(x)|\, d\mu(x) < \infty$.

1.5 Uniform strong law of large numbers

For sequences $\{(e_t, x_t)\}_{t=1}^{\infty}$ as above, which are called Cesaro sum generators, sample averages converge uniformly to population averages:

$$\lim_{n\to\infty} \max_{\lambda\in\Lambda} \left|\frac{1}{n}\sum_{t=1}^{n} g(e_t, x_t, \lambda) - \int_{\mathcal{X}}\int_{\mathcal{E}} g(e, x, \lambda)\, dP(e) d\mu(x)\right| = 0$$

for continuous $g(e, x, \lambda)$ such that $\int\int \max_{\lambda\in\Lambda} |g(e, x, \lambda)|\, dP(e) d\mu(x) < \infty$. Furthermore,

$$\lim_{n\to\infty} \max_{\lambda\in\Lambda} \left|\frac{1}{n}\sum_{t=1}^{n} \int_{\mathcal{E}} g(e, x_t, \lambda)\, dP(e) - \int_{\mathcal{X}}\int_{\mathcal{E}} g(e, x, \lambda)\, dP(e) d\mu(x)\right| = 0.$$

1.6 Relevance of USLLN

The uniform strong law of large numbers implies that if $\lim_{n\to\infty} \hat{\lambda}_n = \lambda^o$ then $\lim_{n\to\infty}$ $(1/n)\sum_{t=1}^{n} g(e_t, x_t, \hat{\lambda}_n) = \int_{\mathcal{X}}\int_{\mathcal{E}} g(e, x, \lambda^o)\, dP(e) d\mu(x)$. This can be used to show that sample averages that depend upon an estimate, such as a variance estimate computed from residuals, converge to their population counterpart.

The function $\hat{\lambda} = \mathrm{argmin}_{\Lambda} s_n(\lambda)$ is continuous with respect to uniform convergence so uniform convergence of $s_n(\lambda)$ to a function $s^*(\lambda)$ that has a unique minimum implies $\hat{\lambda}_n$ converges to $\mathrm{argmin}_{\Lambda} s^*(\lambda)$. This fact is proved below. Usually $s^*(\lambda)$ is easy to compute and the result is easy to apply. Note that $\lambda^o = \mathrm{argmin}_{\Lambda} s_n^o(\lambda)$ together with uniform convergence of $s_n^o(\lambda)$ to $s^*(\lambda)$ implies that $\lambda^o = \mathrm{argmin}_{\Lambda} s^*(\lambda)$.

1.7 Continuity of argmin

If Λ is closed and bounded, $\{s_n(\lambda)\}_{n=1}^{\infty}$ is a sequence of continuous functions, and $s^*(\lambda)$ is continuous with a unique minimum λ^o on Λ then

$$\lim_{n\to\infty} \max_{\lambda\in\Lambda} |s_n(\lambda) - s^*(\lambda)| = 0$$

implies

$$\lim_{n\to\infty} \arg\min_{\lambda\in\Lambda} s_n(\lambda) = \arg\min_{\lambda\in\Lambda} s^*(\lambda).$$

Proof. Let $\lambda^o = \text{argmin}_{\lambda\in\Lambda} s^*(\lambda)$ and let $\hat{\lambda}_n = \text{argmin}_{\lambda\in\Lambda} s_n(\lambda)$. If Λ is closed and bounded then every subsequence $\{\hat{\lambda}_{n_m}\}$ of $\{\hat{\lambda}_n\}$ has a convergent subsubsequence $\{\hat{\lambda}_{n_{m_j}}\}$ with limit point $\lim_{j\to\infty} \hat{\lambda}_{n_{m_j}} = \lambda^{\#}$. Now $s_{n_{m_j}}(\hat{\lambda}_{n_{m_j}}) < s_{n_{m_j}}(\lambda^o)$ and uniform convergence imply $s^*(\lambda^{\#}) \leq s^*(\lambda^o)$. Uniqueness of λ^o implies $\lambda^{\#} = \lambda^o$. Thus, every limit point of $\{\hat{\lambda}_n\}$ is λ^o.

1.8 First order conditions

$$(\partial/\partial\lambda) s_n(\hat{\lambda}_n) = 0$$

1.9 Taylor's expansion of first order conditions

$$[(\partial^2/\partial\lambda\partial\lambda') s_n(\bar{\lambda}_n)]\sqrt{n}(\hat{\lambda}_n - \lambda^o) = -\sqrt{n}(\partial/\partial\lambda) s_n(\lambda^o)$$

$\bar{\lambda}_n$ is on the line segment joining λ^o to $\hat{\lambda}$ hence $\lim_{n\to\infty} \bar{\lambda}_n = \lambda^o$.

1.10 Asymptotics of right hand side

$$\sqrt{n}(\partial/\partial\lambda) s_n(\lambda^o) = (1/\sqrt{n}) \sum_{t=1}^{n} (\partial/\partial\lambda) s[Y(e_t, x_t, \lambda^o), x, \lambda]|_{\lambda=\lambda^o}$$

Mean: $\mathcal{E}[\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o)] = 0$ because $\lambda^o = \text{argmin}_\Lambda s_n^o(\lambda)$

Variance: $\mathcal{I}_n = Var[\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o)]$

$$= (1/n)\sum_{t=1}^n \int_\mathcal{E} (\partial/\partial\lambda)s[Y(e,x_t,\lambda^o),x_t,\lambda](\partial/\partial\lambda')s[Y(e,x_t,\lambda^o),x_t,\lambda]\,dP(e)|_{\lambda=\hat{\lambda}_n}$$

Limit: $\lim_{n\to\infty}\mathcal{I}_n = \mathcal{I}$

$$\mathcal{I} = \int_\mathcal{X}\int_\mathcal{E}(\partial/\partial\lambda)s[Y(e,x,\lambda^o),x,\lambda](\partial/\partial\lambda')s[Y(e,x,\lambda^o),x,\lambda]\,dP(e)d\mu(x)|_{\lambda=\lambda^o}$$

Estimator: $\hat{\mathcal{I}}_n = (1/n)\sum_{t=1}^n(\partial/\partial\lambda)s(y_t,x_t,\hat{\lambda}_n)(\partial/\partial\lambda')s(y_t,x_t,\hat{\lambda}_n)$

Limit: $\lim_{n\to\infty}\hat{\mathcal{I}}_n = \mathcal{I}$

Central limit theorem: $\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) \stackrel{\mathcal{L}}{\to} N_p(0,\mathcal{I})$

1.11 Asymptotics of left hand side

$$\bar{\mathcal{J}}_n = (\partial^2/\partial\lambda\partial\lambda')s_n(\bar{\lambda}_n)$$

Limit: $\lim_{n\to\infty}\bar{\mathcal{J}}_n = \mathcal{J}$

$$\mathcal{J} = \int_\mathcal{X}\int_\mathcal{E}(\partial^2/\partial\lambda\partial\lambda')s[Y(e,x,\lambda^o),x,\lambda]\,dP(e)d\mu(x)|_{\lambda=\lambda^o}$$

Estimator: $\hat{\mathcal{J}}_n = (1/n)\sum_{t=1}^n(\partial^2/\partial\lambda\partial\lambda')s(y_t,x_t,\hat{\lambda}_n)$

Limit: $\mathcal{J} = \lim_{n\to\infty}\hat{\mathcal{J}}_n$

1.12 Slutsky's theorem

$\bar{\mathcal{J}}_n\sqrt{n}(\hat{\lambda}_n - \lambda^o) = -\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o)$

$-\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) \stackrel{\mathcal{L}}{\to} N_p(0,\mathcal{I})$

$\lim_{n\to\infty}\bar{\mathcal{J}}_n = \mathcal{J}$

$\Rightarrow$

$\sqrt{n}(\hat{\lambda}_n - \lambda^o) \stackrel{\mathcal{L}}{\to} N_p(0,V)$

$V = (\mathcal{J})^{-1}\mathcal{I}(\mathcal{J}')^{-1}$

2 Hypothesis Testing Theory

2.1 Summary of Estimation Theory

Structural model: $q(y_t, x_t, \lambda^o) = e_t$	(must be computable)
Reduced form: $y_t = Y(e_t, x_t, \lambda^o)$	(need only exist)
Distance function: $s(y, x, \lambda)$	(small when x & λ fit y well)

Hypothesis: $H : h(\lambda^o) = 0$ against $A : h(\lambda^o) \neq 0$

e_t independent, each with distribution P(e)

(e_t, x_t) a Cesaro sum generator

λ^o in Λ which is a closed and bounded subset of $\Re^p$

$s_n(\lambda) = (1/n)\sum_{t=1}^{n} s(y_t, x_t, \lambda)$	
$\hat{\lambda}_n = \text{argmin}_{\Lambda} s_n(\lambda)$	(unconstrained estimator)
$\tilde{\lambda}_n = \text{argmin}_{h(\lambda)=0} s_n(\lambda)$	(constrained estimator)

$s_n^o(\lambda) = (1/n)\sum_{t=1}^{n} \int s[Y(e, x_t, \lambda^o), x_t, \lambda]\, dP(e)$	
$\lambda^o = \text{argmin}_{\Lambda} s_n^o(\lambda)$	(implies correct specification)
$\lambda^o = \text{argmin}_{h(\lambda)=0} s_n^o(\lambda)$	(implies null hypothesis true)

$\lim_{n\to\infty} \hat{\lambda}_n = \lambda^o$	(proved in Subsections 1.6 & 1.7)
$\lim_{n\to\infty} \tilde{\lambda}_n = \lambda^o$	(the proof is the same)

$\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) = (1/\sqrt{n})\sum_{t=1}^{n}(\partial/\partial\lambda)s[Y(e_t,x_t,\lambda^o),x,\lambda]|_{\lambda=\lambda^o}$

Mean: $\mathcal{E}[\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o)] = 0$ because $\lambda^o = \text{argmin}_\Lambda s_n^o(\lambda)$

Variance: $\mathcal{I}_n = Var[\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o)]$

Limit: $\lim_{n\to\infty}\mathcal{I}_n = \lim_{n\to\infty}\hat{\mathcal{I}}_n = \lim_{n\to\infty}\tilde{\mathcal{I}}_n = \mathcal{I}$

$\hat{\mathcal{I}}_n = \lim(1/n)\sum_{t=1}^{n}(\partial/\partial\lambda)s(y_t,x_t,\hat{\lambda}_n)(\partial/\partial\lambda')s(y_t,x_t,\hat{\lambda}_n)$

$\tilde{\mathcal{I}}_n = \lim(1/n)\sum_{t=1}^{n}(\partial/\partial\lambda)s(y_t,x_t,\tilde{\lambda}_n)(\partial/\partial\lambda')s(y_t,x_t,\tilde{\lambda}_n)$

Central limit theorem: $\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{I})$

$\mathcal{J}_n = (\partial^2/\partial\lambda\partial\lambda')s_n^o(\lambda^o)$

Limit: $\lim_{n\to\infty}\mathcal{J}_n = \lim_{n\to\infty}\hat{\mathcal{J}}_n = \lim_{n\to\infty}\tilde{\mathcal{J}}_n = \mathcal{J}$

$\hat{\mathcal{J}}_n = (1/n)\sum_{t=1}^{n}(\partial^2/\partial\lambda\partial\lambda')s(y_t,x_t,\hat{\lambda}_n)$

$\tilde{\mathcal{J}}_n = (1/n)\sum_{t=1}^{n}(\partial^2/\partial\lambda\partial\lambda')s(y_t,x_t,\tilde{\lambda}_n)$

$\sqrt{n}(\hat{\lambda}_n - \lambda^o) \xrightarrow{\mathcal{L}} N_p(0,V)$

$V = (\mathcal{J})^{-1}\mathcal{I}(\mathcal{J}')^{-1}$

2.2 Order

Let $\{a_n\}_{t=1}^{\infty}$ be a sequence of numbers. Writing $a_n = o(n^\alpha)$ means $\lim_{n\to\infty} a_n/n^\alpha = 0$. Writing $a_n = O(n^\alpha)$ means there is a bound B and an N such that $|a_n/n^\alpha| < B$ for all $n > N$. The exponent α may be positive, negative, or zero.

Let $\{X_n\}_{t=1}^{\infty}$ be a sequence of random variables. Writing $X_n = o_s(n^\alpha)$ means $P(\lim_{n\to\infty} X_n/n^\alpha = 0) = 1$. Writing $X_n = O_s(n^\alpha)$ means there is a bound B and an N such that $P(|X_n/n^\alpha| < B \text{ for all } n > N) = 1$. The exponent α may be positive, negative, or zero.

Let $\{X_n\}_{t=1}^{\infty}$ be a sequence of random variables. Writing $X_n = o_p(n^\alpha)$ means given $\epsilon > 0$ and $\delta > 0$ there is an N such that $P(|X_n/n^\alpha| > \epsilon) < \delta$ for all $n > N$. Writing $X_n = O_p(n^\alpha)$ means given $\delta > 0$ there is a bound B and an N such that $P(|X_n/n^\alpha| > B) < \delta$ for all $n > N$. The exponent α may be positive, negative, or zero.

If X_n converges in distribution then $X_n = O_p(1)$.

The obvious algebra holds: $o(n^\alpha)o(n^\beta) = o(n^{\alpha+\beta})$, $o(n^\alpha)O(n^\beta) = o(n^{\alpha+\beta})$, $o(n^\alpha)o_p(n^\beta) = o_p(n^{\alpha+\beta})$, $o_s(n^\alpha)o_p(n^\beta) = o_p(n^{\alpha+\beta})$, $o(n^\alpha)o_s(n^\beta) = o_s(n^{\alpha+\beta})$, $o_s(n^\alpha)O_p(n^\beta) = o_p(n^{\alpha+\beta})$, etc.

2.3 Distribution of the Constrained Estimator

2.3.1 Lagrangian

$$\mathcal{L}(\lambda, \theta) = s_n(\lambda) + \theta' h(\lambda)$$

2.3.2 First order conditions

$$0 = (\partial/\partial\lambda') s_n(\tilde{\lambda}) + \tilde{\theta}'(\partial/\partial\lambda') h(\tilde{\lambda})$$

$$0 = h(\tilde{\lambda})$$

2.3.3 Taylor's expansions

Note that for any $\bar{\lambda}_n$ on the line segment joining $\tilde{\lambda}_n$ to λ^o we have $(\partial^2/\partial\lambda\partial\lambda') s_n(\bar{\lambda}_n) = (\partial^2/\partial\lambda\partial\lambda') s_n(\lambda^o) + o_s(1)$, $(\partial/\partial\lambda') h(\bar{\lambda}) = (\partial/\partial\lambda') h(\lambda^o) + o_s(1)$. Write $H = (\partial/\partial\lambda') h(\lambda^o)$, $\mathcal{J}$ as above, $\bar{H} = (\partial/\partial\lambda') h(\bar{\lambda}_n)$, $\bar{\mathcal{J}} = (\partial^2/\partial\lambda\partial\lambda') s_n(\bar{\lambda}_n)$, $\tilde{H} = (\partial/\partial\lambda') h(\tilde{\lambda}_n)$, $\tilde{\mathcal{J}} = (\partial^2/\partial\lambda\partial\lambda') s_n(\tilde{\lambda}_n)$, etc.

By Taylor's theorem

$$0 = \bar{H}\sqrt{n}(\tilde{\lambda} - \lambda^o)$$

$$\sqrt{n}(\partial/\partial\lambda) s_n(\tilde{\lambda}_n) = \sqrt{n}(\partial/\partial\lambda) s_n(\lambda^o) + \bar{\mathcal{J}}\sqrt{n}(\tilde{\lambda}_n - \lambda^o).$$

Thus

$$\begin{aligned}
[\bar{H}\bar{\mathcal{J}}^{-1}\tilde{H}']^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\sqrt{n}(\partial/\partial\lambda) s_n(\lambda^o) &= [\bar{H}\bar{\mathcal{J}}^{-1}\tilde{H}']^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\sqrt{n}(\partial/\partial\lambda) s_n(\tilde{\lambda}_n) \\
&\quad -[\bar{H}\bar{\mathcal{J}}^{-1}\tilde{H}']^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\bar{\mathcal{J}}\sqrt{n}(\tilde{\lambda}_n - \lambda^o) \\
&= [\bar{H}\bar{\mathcal{J}}^{-1}\tilde{H}']^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\tilde{H}'\sqrt{n}\tilde{\theta} - 0 \\
&= \sqrt{n}\tilde{\theta}.
\end{aligned}$$

Since $[\bar{H}\bar{\mathcal{J}}^{-1}\tilde{H}']^{-1}$, $\bar{H}\bar{\mathcal{J}}^{-1}$, and $\sqrt{n}(\partial/\partial\lambda) s_n(\lambda^o)$ are each $O_p(1)$ we have that

$$\sqrt{n}\tilde{\theta} = O_p(1)$$

2.3.4 Key equations

$$\begin{aligned}
&H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}\sqrt{n}(\partial/\partial\lambda) s_n(\lambda^o) \\
&\quad = [H'(H\mathcal{J}^{-1}H')^{-1}\bar{H}\bar{\mathcal{J}}^{-1} + o_s(1)]\sqrt{n}(\partial/\partial\lambda) s_n(\lambda^o)
\end{aligned}$$

$$
\begin{aligned}
&= H'(H\mathcal{J}^{-1}H')^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) + o_p(1) \\
&= H'(H\mathcal{J}^{-1}H')^{-1}\bar{H}\bar{\mathcal{J}}^{-1}[\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) - \bar{\mathcal{J}}\sqrt{n}(\tilde{\lambda}_n - \lambda^o)] + o_p(1) \\
&= H'(H\mathcal{J}^{-1}H')^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) - H'(H\mathcal{J}^{-1}H')^{-1}\bar{H}\sqrt{n}(\tilde{\lambda}_n - \lambda^o) + o_p(1) \\
&= H'(H\mathcal{J}^{-1}H')^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) - 0 + o_p(1) \\
&= -H'(H\mathcal{J}^{-1}H')^{-1}\bar{H}\bar{\mathcal{J}}^{-1}\tilde{H}'\sqrt{n}\theta' + o_p(1) \\
&= -[\tilde{H}' + o_s(1)][I + o_s(1)]\sqrt{n}\theta' + o_p(1) \\
&= -\tilde{H}'\sqrt{n}\tilde{\theta} + o_p(1) + o_p(1) \\
&= \sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_p(1).
\end{aligned}
$$

2.3.5 Main result

Joining the first line of Subsection 2.3.4 to the last we have

$$H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) = \sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_p(1).$$

2.4 Continuity theorem

If $X_n \xrightarrow{\mathcal{L}} X$ and $g(x)$ is continuous then $g(X_n) \xrightarrow{\mathcal{L}} g(X)$. For example, if $X_n \xrightarrow{\mathcal{L}} X$ and $A_n \to A$ where X is $N(0,\Sigma)$, A is symmetric, and $A\Sigma$ is idempotent then $X_n'A_nX_n$ converges in distribution to a chi square with rank(A) degrees freedom.

2.5 Wald test

By Taylor's theorem

$$\sqrt{n}[h(\hat{\lambda}_n) - h(\lambda^o)] = \bar{H}\sqrt{n}(\hat{\lambda}_n - \lambda^o).$$

Because $h(\lambda^o) = 0$

$$\sqrt{n}h(\hat{\lambda}_n) = \bar{H}\sqrt{n}(\hat{\lambda}_n - \lambda^o).$$

Because $\sqrt{n}(\hat{\lambda}_n - \lambda^o) \xrightarrow{\mathcal{L}} N_p(0,V)$ and $\bar{H} = H + o_s(1)$

$$\sqrt{n}h(\hat{\lambda}_n) \xrightarrow{\mathcal{L}} N_q(0, HVH').$$

Thus

$$W = \hat{h}'(HVH')^{-1}\hat{h}$$

converges in distribution to a chi square with q degrees freedom by Subsection 2.4.

2.6 Likelihood ratio test

By Taylor's theorem

$$\begin{aligned}(\partial/\partial\lambda')s_n(\tilde{\lambda}_n) &= (\partial/\partial\lambda')s_n(\hat{\lambda}_n) + \bar{\mathcal{J}}(\tilde{\lambda}_n - \hat{\lambda}_n) \\ &= 0 + \bar{\mathcal{J}}(\tilde{\lambda}_n - \hat{\lambda}_n)\end{aligned}$$

$$\begin{aligned}L &= 2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)] \\ &= 2n(\partial/\partial\lambda)s_n(\hat{\lambda}_n)(\tilde{\lambda}_n - \hat{\lambda}_n) + \sqrt{n}(\tilde{\lambda}_n - \hat{\lambda}_n)'\bar{\bar{\mathcal{J}}}\sqrt{n}(\tilde{\lambda}_n - \hat{\lambda}_n) \\ &= 0 + \sqrt{n}(\partial/\partial\lambda')s_n(\tilde{\lambda}_n)\bar{\mathcal{J}}^{-1}\bar{\bar{\mathcal{J}}}\bar{\mathcal{J}}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) \\ &= \sqrt{n}(\partial/\partial\lambda')s_n(\tilde{\lambda}_n)[\mathcal{J}^{-1} + o_s(1)]\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) \\ &= \sqrt{n}(\partial/\partial\lambda')s_n(\tilde{\lambda}_n)\mathcal{J}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_p(1)\end{aligned}$$

Substituting

$$\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) + o_p(1)$$

from Subsection 2.3.5 and noting that this equation implies $\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = o_p(1)$ we have

$$L = \sqrt{n}(\partial/\partial\lambda')s_n(\lambda^o)\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) + o_p(1)$$

Recall $\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) \xrightarrow{\mathcal{L}} N_p(0, \mathcal{I})$. If $H'VH = H'\mathcal{J}^{-1}H$ then $\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}\mathcal{I}$ is idempotent and

$$L = 2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)]$$

converges in distribution to a chi square with q degrees freedom by Subsection 2.4.

2.7 Lagrange multiplier test

Using

$$\sqrt{n}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) + o_p(1)$$

from Subsection 2.3.5 we have that

$$\begin{aligned}R &= n(\partial/\partial\lambda')s_n(\tilde{\lambda}_n)\tilde{\mathcal{J}}^{-1}\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{\mathcal{J}}^{-1}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) \\ &= n(\partial/\partial\lambda')s_n(\tilde{\lambda}_n)\mathcal{J}^{-1}H'(HVH')^{-1}H\mathcal{J}^{-1}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_p(1) \\ &= n(\partial/\partial\lambda')s_n(\lambda_n^o)\mathcal{J}^{-1}H'(HVH')^{-1}H\mathcal{J}^{-1}(\partial/\partial\lambda)s_n(\lambda_n^o) + o_p(1).\end{aligned}$$

Recall that $\sqrt{n}(\partial/\partial\lambda)s_n(\lambda^o) \xrightarrow{\mathcal{L}} N_p(0,\mathcal{I})$. Because $\mathcal{J}^{-1}H'(HVH')^{-1}H\mathcal{J}^{-1}\mathcal{I}$ is idempotent

$$R = n(\partial/\partial\lambda')s_n(\tilde{\lambda}_n)\tilde{\mathcal{J}}^{-1}\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{\mathcal{J}}^{-1}(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)$$

converges in distribution to a chi square with q degrees freedom by Subsection 2.4.

3 Applications

3.1 Least squares

$y_t = f(x_t, \theta^o) + e_t$

$\mathcal{E}(e_t) = 0$

$\text{Var}(e_t) = \sigma^2$

$\lambda = \theta$

Structural model: $e_t = q(y_t, x_t, \theta^o) = y_t - f(x_t, \theta^o)$

Reduced form: $y_t = Y(e_t, x_t, \theta^o) = f(x_t, \theta^o) + e_t$

Distance function: $s(y, x, \theta) = [y - f(x, \theta)]^2$

$(\partial/\partial\theta) s(y, x, \theta) = -2[y - f(x, \theta)](\partial/\partial\theta) f(x, \theta)$

$(\partial^2/\partial\theta\partial\theta') s(y, x, \theta) = 2[(\partial/\partial\theta) f(x, \theta)][(\partial/\partial\theta) f(x, \theta)]' - 2[y - f(x, \theta)](\partial^2/\partial\theta\partial\theta') f(x, \theta)$

$s_n(\theta) = (1/n) \sum_{t=1}^n [y_t - f(x_t, \theta)]^2$

$s_n^o(\theta) = \sigma^2 + (1/n) \sum_{t=1}^n [f(x_t, \theta^o) - f(x_t, \theta)]^2$

$s^*(\theta) = \sigma^2 + \int_{\mathcal{X}} [f(x_t, \theta^o) - f(x_t, \theta)]^2 \, d\mu(x)$

$\mu\{x : f(x, \theta) \neq f(x, \theta^o)\} \Rightarrow \theta^o = \operatorname{argmin} s_n^o(\theta) \Rightarrow \theta^o = \operatorname{argmin} s^*(\theta)$

$\mathcal{I} = \int_{\mathcal{X}} 4\sigma^2 [(\partial/\partial\theta) f(x, \theta)][(\partial/\partial\theta) f(x, \theta)]' \, d\mu(x)|_{\theta=\theta^o}$

$\mathcal{I} = \lim_{n\to\infty} (1/n) 4\sigma^2 \sum_{t=1}^n [(\partial/\partial\theta) f(x_t, \theta)][(\partial/\partial\theta) f(x_t, \theta)]'|_{\theta=\hat{\theta}_n}$

$\mathcal{I} = \lim_{n\to\infty} (1/n) \sum_{t=1}^n 4[y_t - f(x_t, \hat{\theta}_n)]^2 [(\partial/\partial\theta) f(x_t, \hat{\theta}_n)][(\partial/\partial\theta) f(x_t, \hat{\theta}_n)]'$

$\sigma^2 = \lim_{n\to\infty} \ (1/n) \sum_{t=1}^n [y_t - f(x_t, \hat{\theta}_n)]^2$

$\mathcal{J} = \int_{\mathcal{X}} 2[(\partial/\partial\theta) f(x, \theta)][(\partial/\partial\theta) f(x, \theta)]' \, d\mu(x)|_{\theta=\theta^o}$

$\mathcal{J} = \lim_{n\to\infty} (1/n) \sum_{t=1}^n \ 2[(\partial/\partial\theta) f(x, \hat{\theta}_n)][(\partial/\partial\theta) f(x, \hat{\theta}_n)]'$

$V = \sigma^2 \Big\{ \int_{\mathcal{X}} [(\partial/\partial\theta) f(x, \theta)][(\partial/\partial\theta) f(x, \theta)]' \, d\mu(x)|_{\theta=\theta^o} \Big\}$

3.2 Maximum likelihood

$y_t = f(x_t, \theta^o) + \sigma^o e_t$

$\mathcal{E}(e_t) = 0$

$\mathrm{Var}(e_t) = 1$

$\lambda = (\theta', \sigma^2)'$

Structural model: $e_t = q(y_t, x_t, \lambda^o) = [y_t - f(x_t, \theta^o)]/\sigma^o$

Reduced form: $y_t = Y(e_t, x_t, \theta^o) = f(x_t, \theta^o) + \sigma^o e_t$

Distance function: $s(y, x, \lambda) = (1/2)\{\log \sigma^2 + \sigma^{-2}[y - f(x, \theta)]^2\}$

$(\partial/\partial\theta)s(y, x, \lambda) = -\sigma^{-2}[y - f(x, \theta)](\partial/\partial\theta)f(x, \theta)$

$(\partial^2/\partial\theta\partial\theta')s(y, x, \lambda) = \sigma^{-2}[(\partial/\partial\theta)f(x, \theta)][(\partial/\partial\theta)f(x, \theta)]' - \sigma^{-2}[y - f(x, \theta)](\partial^2/\partial\theta\partial\theta')f(x, \theta)$

$(\partial/\partial\sigma^2)s(y, x, \lambda) = (1/2)\{\sigma^{-2} - \sigma^{-4}[y - f(x, \theta)]^2\}$

$(\partial^2/\partial\sigma^2\partial\sigma^2)s(y, x, \lambda) = (1/2)\{-\sigma^{-4} + \sigma^{-6}[y - f(x, \theta)]^2\}$

$(\partial^2/\partial\sigma^2\partial\theta)s(y, x, \lambda) = \sigma^{-4}[y - f(x, \theta)](\partial/\partial\theta)f(x, \theta)$

$s_n(\lambda) = (1/n)\sum_{t=1}^{n}(1/2)\{\log \sigma^2 + \sigma^{-2}[y - f(x, \theta)]^2\}$

$s_n^o(\lambda) = (1/2)\{\log \sigma^2 + (\sigma^o/\sigma)^2 + (1/n)\sum_{t=1}^{n} \sigma^{-2}[f(x_t, \theta^o) - f(x_t, \theta)]^2\}$

$s^*(\lambda) = (1/2)\{\log \sigma^2 + (\sigma^o/\sigma)^2 + \int_{\mathcal{X}} \sigma^{-2}[f(x_t, \theta^o) - f(x_t, \theta)]^2 \, d\mu(x)\}$

$\mu\{x : f(x, \theta) \neq f(x, \theta^o)\} \Rightarrow (\theta^o, \sigma^o) = \operatorname{argmin} s_n^o(\lambda) \Rightarrow (\theta^o, \sigma^o) = \operatorname{argmin} s^*(\lambda)$

$$\mathcal{I} = \begin{pmatrix} (\sigma^o)^{-2}Q & (1/2)(\sigma^o)^{-3}\mathcal{E}(e^3)q \\ (1/2)(\sigma^o)^{-3}\mathcal{E}(e^3)q' & (1/4)(\sigma^o)^{-4}\mathrm{Var}(e^2) \end{pmatrix}$$

$$\mathcal{J} = \begin{pmatrix} (\sigma^o)^{-2}Q & 0 \\ 0' & (1/2)(\sigma^o)^{-4} \end{pmatrix}$$

$$q = \int_{\mathcal{X}} (\partial/\partial\theta) f(x,\theta)]\, d\mu(x)|_{\theta=\theta^o}$$

$$Q = \int_{\mathcal{X}} [(\partial/\partial\theta) f(x,\theta)][(\partial/\partial\theta) f(x,\theta)]'\, d\mu(x)|_{\theta=\theta^o}$$

$$V = \begin{pmatrix} (\sigma^o)^2 Q^{-1} & (\sigma^o)^3 \mathcal{E}(e^3) Q^{-1} q \\ (\sigma^o)^3 \mathcal{E}(e^3) q' Q^{-1} & (\sigma^o)^4 \mathrm{Var}(e^2) \end{pmatrix}$$

4 References

Burguete, Jose F., A. Ronald Gallant, and Geraldo Souza (1982), "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews* 1, 151–190.

Gallant, A. Ronald (1987), *Nonlinear Statistical Models,* New York: Wiley.