COWLES COMMISSION DISCUSSION PAPER: STATISTICS: 329

June 17, 1949

ASYMPTOTIC PROPERTIES OF LIMITED INFORMATION ESTIMATES IN LINEAR AND NON-LINEAR SYSTEMS

by

Herman Rubin

I. Linear Case. Suppose we are presented with a complete linear system of stochastic equations

(1) $$\beta_{qy}\, y_t' + \Gamma_{qz}\, z_t' = q_t',$$

where y_t is jointly dependent at "time" t, z_t is predetermined at "time" t, and q_t are independently and normally distributed with mean 0 and constant non-singular covariance matrix Σ. Suppose that $q = (u\ r)$. Then (1) may be written as

(2) $$\beta_{uy}\, y_t' + \Gamma_{uz}\, z_t' = u_t',$$
$$\beta_{ry}\, y_t' + \Gamma_{rz}\, z_t' = r_t'.$$

Then the maximum likelihood estimates of $\alpha_{ux} = (\beta_{uy}\ \Gamma_{uz})$ is that matrix $\hat{A}_{ux}$ which minimizes

(3) $$V = \frac{|A_{ux}\, M_{xx}\, A_{ux}'|}{|A_{ux}\, W_{xx}\, A_{ux}'|}$$

where

(4) $$W_{xx} = M_{xx} - M_{xz}\, M_{zz}^{-1}\, M_{zx},$$

or the moment matrix[1] of the residuals of the coordinates of X_t about their regression on the coordinates of z_t.

II. Non-Linear Case. Suppose q_t are serially independent random variables (not necessarily in a Euclidean space) and suppose there exist measurable vector

1. By M_{ab} we mean $\frac{1}{T} \sum_{t=1}^{T} a_t'\, b_t$.

functions $x_t = f_t\,(q_t, q_{t-1}, \text{---})$, $u_t = g_t\,(q_t, q_{t-1}, \text{---})$ such that

(5) $\quad \alpha_{ux}\, x_t' = u_t'.$

Suppose further that the vector $z_t = h_t\,(q_{t-1}, q_{t-2}, \text{---})$. Then why not estimate α_{ux} by the same procedure as before?

For example, suppose

$$
\begin{aligned}
& y_{1t} + \alpha\, v_{1t}\, y_{2t} + \beta y_{3t-1} + \gamma = q_{1t}\, y_{1,t-1}, \\
(6) \qquad & y_{3t} + \delta\, y_{1t}\, y_{3t} + \varepsilon\, v_{2t} + \zeta = q_{2t}\, v_{1t}\, v_{2t}, \\
& f(y_{1t}, y_{2t}, y_{3t}, y_{2,t-1}, v_{1t}, q_{3t}, q_{4t}, q_{5t}) = 0,
\end{aligned}
$$

and v_1 and v_2 are exogenous. Then we may take

$x_t = (y_{1t}, v_{1t}, v_{1t}\, y_{2t}, y_{3,t-1}, 1, y_{3t}, y_{1t}\, y_{3t}, v_{2t})$, $u_t = (q_{1t}\, y_{1,t-1}, q_{2t}\, v_{1t}\, v_{2t}$ and $z_t = (v_{1t}, y_{3,t-1}, v_{2t}, 1, v_{1t}\, v_{2t}, v_{1t}^2)$. Then $\alpha_{ux} = \begin{bmatrix} 1 & \alpha & \beta & \gamma & 0 & 0 & 0 \\ 0 & 0 & 0 & \zeta & 1 & \delta & \varepsilon \end{bmatrix}.$

The important properties of maximum likelihood estimates which cause us to use them are consistency, efficiency and asymptotic normality. We in general cannot get efficiency without a much better knowledge of the process then that contained in α_{ux}, but we may still hope for the other properties. <u>We find that certain "reasonable" requirements, which are also the same in the case of maximum likelihood estimates, suffice here.</u> Let us examine these conditions.

In the case of maximum likelihood estimates, $\mathcal{E}(z_t'\, u_t) = 0$ for all t. We need an asymptotic condition similar to this.

Condition I. $\operatorname*{plim}_{t \to \infty} M_{uz}\, M_{zz}^{-1}\, M_{zu} = 0$, $\mathcal{E}(u_t'\, z_t) = 0$.[2]

I.e., for large enought T, M_{zu} is close to 0 "most of the time." This condition stated mathematically is: For any $\varepsilon > 0$, any any $\delta > 0$, for $T > T(\varepsilon, \delta)$, the probability that[3] $\operatorname{tr}(M_{zu}\, M_{uz}) < \delta^2$ is greater than $1 - \varepsilon$.

Let us write $\operatorname{tr} HH' = \|H\|^2$. This notation will be convenient in what follows:

---

2. A weaker condition than this can be used.

3. Any other definition of norm could be used.

For maximum likelihood estimates, $\operatorname{plim}_{t\to\infty} M_{uu} = \Sigma$. We need merely that the u's are not "too wild," i.e.,

Condition II: For any $\varepsilon > 0$, there is an N_1 such that $\| M_{uu} \| \leq N_1$ with probability greater than $1 - \varepsilon$, and an N_2 such that for sufficiently large T, $\| M_{uu}^{-1} \| \leq N_2$ with probability greater than $1 - \varepsilon$.

We also need that the z's do not get too big, i.e.,

Condition III. For every $\varepsilon > 0$, there is an N_3 such that $\| M_{zz} \| < N_3$ with probability greater than $1 - \varepsilon$.

Let

(7) $\mu_{xz} = \frac{1}{T} \sum_{t=1}^{T} (x_t' \mid t-1)\, z_t.$

Then

(8) $\alpha_{ux} \mu_{xz} = \frac{1}{T} \sum_{t=1}^{T} (u_t' \mid t-1)\, z_t = 0.$

If M_{zz} approaches a singular matrix, the estimates need not be consistent. Let Q be an orthogonal matrix such that $Q M_{zz} Q'$ is diogonal and the roots are arranged in descending order. Let Q_ε be the matrix composed of those rows of Q for which the characteristic roots are greater than ε. We also need something corresponding to i~~n~~dentifiability, but slightly stronger (also insurance that z's don't get too small)

Condition IV: For every $\delta > 0$, there is an $\varepsilon > 0$, such that for T sufficiently large, $\alpha_{ux} \mu_{xz} Q_\varepsilon' = 0$ has a unique solution with probability greater than $1 - \delta$.

We also need

Condition V: $\operatorname{plim}_{t\to\infty} M_{xz} - \mu_{xz} = 0$

Condition VI: For every $\varepsilon > 0$, there is a bounded closed set $\mathcal{S}$ in μ_{xz} - space such that μ_{xz} is in $\mathcal{S}$ with probability greater than $1 - \varepsilon$ for T sufficiently large, and if H is in $\mathcal{S}$, then for every neighborhood $\mathcal{N}$ of α_{ux}, if K is sufficiently near H and AK = 0, then A is in $\mathcal{N}$.

This last condition states that μ_{xz} does not "misbehave" too violently and that it stays in a region in which we may "continuously" solve for A as a function of μ_{xz}.

Under these conditions, $\hat{A}_{ux}$ is a consistent estimate of α_{ux}.

If, in addition, the elements of $\sqrt{T}\, M_{uz}$ are asymptotically normally distributed, so are the elements of $\hat{A}_{ux}$, provided the elements of α_{ux} are functions of the elements of μ_{xz} with continuous first partial derivatives. If the covariance of $\sqrt{T}\, m_{u_i z_k}$ and $\sqrt{T}\, m_{u_j z_\ell}$ is approximately $m_{u_i u_j}\, m_{z_k z_\ell}$, then the asymptotic covariance matrix of the parametric system for α_{ux} is twice the inverse of the second partial derivative matrix of log V with respect to those parameters.