Analysis of the NLLS example in the scope of the OE theory.

In the context of the NLLS example remember that the statistical model is specified by $\{ \mathbb{E}(Y_n | \sigma(X_n)) = g(\theta, X_n)$, $\theta \in \Theta$, $Var(Y_n | \sigma(X_n)) = I_n \}$, where $g(\theta, X_n) : \Theta \times \mathbb{R}^{n \times k} \to \mathbb{R}^n$, e.g. (a) $g(\theta, X_n) = (\exp(X_{(i)} \theta))_{i=1,\dots,n}$ ($X_{(i)}$ is the i^{th} row of X_n), e.g. (b) $g(\theta, X_n) = X_n \theta$ (the usual LS case is recovered)

[As in the LS case, the simplifying $Var(Y_n | \sigma(X_n)) = I_n$ assumption is made for brevity]

Consider

$$c_n^*(\theta) : \mathbb{E}\left[\tfrac{1}{n} \overbrace{(Y_n - g(\theta, X_n))'(Y_n - g(\theta, X_n))}^{A_n} \right]$$

And notice that since $Y_n = g(\theta_0, X_n) + \varepsilon_n$ where $\mathbb{E}(\varepsilon_n | \sigma(X_n)) = 0_n$ and $Var(\varepsilon_n | \sigma(X_n)) = I_n$,

$$A_n = ((g(\theta_0, X_n) - g(\theta, X_n)) + \varepsilon_n)' \times$$

$$((g(\theta_0, X_n) - g(\theta, X_n)) + \varepsilon_n) =$$

$$\left((g(\theta_0, X_n) - g(\theta, X_n))' + \varepsilon_n' \right) \times$$

$$\left((g(\theta_0, X_n) - g(\theta, X_n)) + \varepsilon_n \right) =$$

$\varepsilon_n' u_n \overset{1\times 1}{=}$

$(\varepsilon_n' u_n)' = u_n' \varepsilon_n$

$$u_n' u_n + \varepsilon_n' u_n + u_n \varepsilon_n' + \varepsilon_n' \varepsilon_n =$$

$$\boxed{u_n' u_n + 2 u_n' \varepsilon_n + \varepsilon_n' \varepsilon_n \quad (A)}$$

where $u_n := \quad g(\theta_0, X_n) - g(\theta, X_n)$

Hence, taking expectations conditionally on $\sigma(X_n)$, and using (A),

$$C_n^*(\theta) = \frac{1}{n} (g(\theta_0, X_n) - g(\theta, X_n))' (g(\theta_0, X_n) -$$
$$g(\theta, X_n)) + 1,$$

by noticing that $\mathbb{E}(\mathbb{M}_n'\mathbb{M}_n / \sigma(X_n)) =$

$\mathbb{M}_n'\mathbb{M}_n$, $\mathbb{E}(\mathbb{M}_n'\varepsilon_n / \sigma(X_n)) = \mathbb{M}_n'\mathbb{E}(\varepsilon_n / \sigma(X_n))$

$= \mathbb{M}_n' 0_{n\times 1} = 0_{n\times 1}$, $\mathbb{E}(\varepsilon_n'\varepsilon_n / \sigma(X_n)) = \mathbb{E}(tr(\varepsilon_n'\varepsilon_n) / \sigma(X_n))$

$= \mathbb{E}(tr(\varepsilon_n\varepsilon_n') / \sigma(X_n)) = tr\,\mathbb{E}(\varepsilon_n\varepsilon_n' / \sigma(X_n))$

$= tr\,Var(\varepsilon_n / \sigma(X_n)) = tr\, I_n = n$

(since $\varepsilon_n'\varepsilon_n \overset{1\times 1}{=} tr(\varepsilon_n'\varepsilon_n) = tr(\varepsilon_n\varepsilon_n')$, while tr and $\mathbb{E}$ commute due to linearity).

Then, if $g(\theta, X_n) = g(\theta_0, X_n) \Leftrightarrow \theta = \theta_0$ for almost every X_n (Ident) holds

$$\underset{\theta \in \Theta}{\arg\min} \left[\left(g(\theta_0, \mathbb{X}_n) - g(\theta, \mathbb{X}_n) \right)' \times \right.$$

$$\left. \left(g(\theta_0, \mathbb{X}_n) - g(\theta, \mathbb{X}_n) \right)' + L \right]$$

$$= \underset{\theta \in \Theta}{\arg\min} \; M_n' M_n = \{ \theta \in \Theta : g(\theta, \mathbb{X}_n) =$$

$$= g(\theta_0, \mathbb{X}_n) \} = \theta_0,$$ i.e. θ_0 is recoverable

as the unique minimizer of G_n^* due to (Ident) and the specification properties.

$\boxed{\text{Note:}}$ in e.g. (α) (Ident) holds if

$\exp(x_{(i)}\theta) = \exp(x_{(i)}\theta_0)$ for some $i = 1, \dots, n$

then $\theta = \theta_0$, equivalent to $x_{(i)}\theta = x_{(i)}\theta_0$

for some $i=1,\dots,n$ then $\theta=\Theta_0$, since e^x is 1-1, equivalent to $\text{rank}(X_n)=k$. The same condition holds for e.g. (b) as we already know. $\boxtimes$

C_n^* is latent, since it directly depends on $g(\Theta_0, X_n)$, yet it seems approximable
↳ latent

via its empirical version

$$C_n(\theta) := 1/n\,(Y_n - g(\theta, X_n))'(Y_n - g(\theta, X_n))$$

$$= 1/n\, Y_n'Y_n - 2/n\, Y_n' g(\theta, X_n) + 1/n\, g(\theta, X_n)' g(\theta, X_n)$$

$$= \frac{1}{n} U_n'U_n - \frac{2}{n} U_n'\varepsilon_n + \frac{1}{n}\varepsilon_n'\varepsilon_n$$ [latent form, useful in as. theory]

Hence the resulting OF, termed in this case Non Linear Least Squares Estimator (NLLS) is defined as:

$$C_n(\hat{\theta}_n) \leq \inf_{\theta \in \Theta} C_n(\theta) + \epsilon_n$$

(ϵ_n being the usual optimization error)

Or if $\epsilon_n = 0$,

$$\boxed{\hat{\theta}_n \in \underset{\theta \in \Theta}{\operatorname{argmin}}\ C_n(\theta) \qquad (B)}$$

Note: Since V_n does not depend on θ, (depends on θ_0), $\hat{\theta}_n$ can be defined as the minimizer of

$$\breve{C}_n(\theta) := 1/n \left[g(\theta, X_n)' g(\theta, X_n) - 2 V_n' g(\theta, X_n) \right]$$

Note: If $\Theta = \mathbb{R}^k$ and $g(\cdot, X_n)$ is two times differentiable then $\hat{\theta}_n$ satisfies:

$$\frac{\partial \breve{C}_n(\hat{\theta}_n)}{\partial \theta} = 0_{k\times L}, \text{ with } \frac{\partial \breve{C}_n(\theta)}{\partial \theta} =$$

$$= 1/n \; 2 \underbrace{\frac{\partial g(\theta, X_n)'}{\partial \theta}}_{k\times n} \underbrace{\left[g(\theta, X_n) - V_n \right]}_{n\times L} = 0_{k\times L} \qquad (\text{f.o.c})$$

and $\dfrac{\partial^2 \breve{c}_n(\theta_n)}{\partial\theta\partial\theta'}$ p.d. with

$$\frac{\partial^2 c_n(\theta)}{\partial\theta\partial\theta'} = \frac{2}{n}\underbrace{\frac{\partial g(\theta,X_n)'}{\partial\theta}}_{k\times n}\underbrace{\frac{\partial g(\theta,X_n)}{\partial\theta'}}_{n\times k}$$

$$\boxed{\begin{gathered}\text{S.O.C}\\ \frac{\partial^2 c_n}{\partial\theta\partial\theta'}(\theta_n)\ \text{p.d.}\end{gathered}}$$

$$+\frac{2}{n}\left(\underbrace{\frac{\partial^2 g(\theta,X_n)'}{\partial\theta\,\partial\theta_u}[g(\theta,X_n)-Y_n]}_{k\times 1}\right)_{u=1,\dots,k}$$

In eg.(b) $\dfrac{\partial g(\theta,X_n)'}{\partial\theta} = X_n'$, and

Subsequently $\dfrac{\partial g(\theta,X_n)'}{\partial\theta\,\partial\theta_u} = 0_{k\times n}$, while in e.g. ($\alpha$),

$$\frac{\partial g(\theta, X_n)'}{\partial \theta} = \left(\frac{\partial \exp(X_{(i)}\theta)}{\partial \theta} \right)'_{i=1,\dots,n}$$

$$= \left(\exp(X_{(i)}\theta) X'_{(i)} \right)_{i=1,\dots,n}$$

$$= \begin{pmatrix} \overbrace{\exp\left(\sum_{j=1}^{k} X_{1j}\theta_j\right)}^{\alpha_{1,1}} X_{1,1} & \dots & \overbrace{\exp\left(\sum_{j=1}^{k} X_{nj}\theta_j\right)}^{\alpha_{1,n}} X_{n,1} \\ \vdots & \ddots & \vdots \\ \overbrace{\exp\left(\sum_{j=1}^{k} X_{1j}\theta_j\right)}^{\alpha_{1,k}} X_{1,k} & \dots & \overbrace{\exp\left(\sum_{j=1}^{k} X_{nj}\theta_j\right)}^{\alpha_{k,n}} X_{n,k} \end{pmatrix}$$

and

$$\frac{\partial^2 g(\theta, X_n)'}{\partial \theta\, \partial \theta_\mu} = \dots =$$

$$\begin{pmatrix} \alpha_{1,1} X_{1,\mu} & \dots & \alpha_{1,n} X_{1,\mu} \\ \vdots & \ddots & \vdots \\ \alpha_{1,k} X_{1,\mu} & \dots & \alpha_{k,n} X_{1,\mu} \end{pmatrix}$$

which directly implies that the analytical form of the NLLS in e.g. (α) is difficult to derive, and numerical methods are needed (even when $\Theta = \mathbb{R}^k$)

Note: Regarding the issue of existence for the NLLS, our general theory says that for example, if ① Θ is compact and $g(\theta, X_n)$ is continuous in θ then $\hat{\theta}$ exists (this is true in both e.g. (α)-(b) for compact Θ), or ② if Θ is convex (e.g. $\Theta = \mathbb{R}^k$) and $g(\theta, X_n)$ is strictly convex, then $\hat{\theta}_n$ exists (as a matter of fact $\underset{\theta \in \Theta}{\operatorname{argmin}}\, G_n(\theta)$ is then unique).

For convex Θ, strict convexity holds if $\operatorname{rank}(X_n) = k$, in both e.g. (a)-(b). ☑

Limit Theory: Lets try to obtain the limit theory of $\hat{\theta}_n$ via the OE theory.

*** Consistency**

Notice that (using the derivations for G_n^*) $G_n(\theta) = \frac{1}{n} U_n' U_n - \frac{2}{n} U_n' \varepsilon_n + \frac{1}{n} \varepsilon_n' \varepsilon_n$

Similarly to the relevant analysis for the LS example, using the latent form of the criterion:

$$C_n(\theta) = \frac{1}{n} U_n' U_n - \frac{2}{n} U_n' \varepsilon_n + \frac{1}{n} \varepsilon_n' \varepsilon_n$$

(remember $U_n := g(\theta_0, X_n) - g(\theta, X_n)$)

and noting that the term $\frac{1}{n} \varepsilon_n' \varepsilon_n$ does not depend on θ, hence it does not affect optimization, it is obtained that (ignoring for simplicity issues regarding optimization errors) the NLLSE satisfies

$$\hat{\theta}_n \in \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{n} U_n' U_n - \frac{2}{n} U_n' \varepsilon_n$$

Lets consider first the asymptotic behaviour of the term $\frac{1}{n} d_n' \varepsilon_n$;

we have that:

$$d_n = g(\theta_0, X_n) - g(\hat{\theta}, X_n).$$

Assume for simplicity that

$$g(\theta, X_n) = \begin{pmatrix} G(\theta, X_{(1)}) \\ G(\theta, X_{(2)}) \\ \vdots \\ G(\theta, X_{(n)}) \end{pmatrix}$$

where $G : \Theta \times \mathbb{R}^k \to \mathbb{R}$

(In e.g. (a) $G(\theta, X_{(i)}) := \exp(X_{(i)}\theta)$ and in e.g. (b) $G(\theta, X_{(i)}) := X_{(i)}\theta$)

And Notice that

this implies that:

$$\frac{1}{n} M_n' \varepsilon_n = \frac{1}{n} \sum_{i=1}^{n} G(\theta, X_{(i)}) \varepsilon_{(i)}$$

($\varepsilon_{(i)}$ is similarly the i^{th} "row" of ε_n, i.e., its i^{th} component)

We have that $\mathbb{E}\left[G(\theta, X_{(i)}) \varepsilon_{(i)}\right]$

$$\overset{L.I.E.}{=} \mathbb{E}\left[\mathbb{E}\left(G(\theta, X_{(i)}) \varepsilon_{(i)} / \mathcal{G}(X_n)\right)\right] =$$

$$= \mathbb{E}\left[G(\theta, X_{(i)}) \mathbb{E}\left(\varepsilon_{(i)} / \mathcal{G}(X_n)\right)\right]$$

$$= \mathbb{E}\left[G(\theta, X_{(i)}) 0\right] = 0.$$

Thereby if some LLN is applicable to $\frac{1}{n}\sum_{i=1}^{n} G(\theta, \mathbf{X}_{(i)})\varepsilon_{(i)}$, then this would converge to zero (in time series contexts this should occur as long as, e.g. $(Y_n, \mathbf{X}_n)$ is a stationary ergodic process [e.g. iid]).

Hence lets assume:

$$\boxed{\alpha.\quad 1/n \sum_{i=1}^{n} G(\theta, X_{(i)})\,\varepsilon_{(i)} \xrightarrow{P} 0 \quad \forall \theta \in \Theta.}$$

e.g. (α) $1/n \sum_{i=1}^{n} \exp(\mathbf{X}_{(i)}\theta)\,\varepsilon_{(i)} \xrightarrow{P} 0 \quad \forall \theta \in \Theta$

e.g. (β) $1/n \sum_{i=1}^{n} X_{(i)}\theta\,\varepsilon_i = \left(\frac{1}{n}\sum_{i=1}^{n} X_{(i)}\varepsilon_{(i)}\right)\theta \xrightarrow{P} 0 \quad \forall \theta \in \Theta \overset{\text{CMT.}}{\Leftarrow} \frac{1}{n}\sum_{i=1}^{n} X_{(i)}\varepsilon_{(i)} \xrightarrow{P} 0$

Furthermore, lets try to discern the asymptotic behavior of

$$\frac{1}{n} M_n' M_n = \frac{1}{n}\left(\begin{pmatrix} G(\theta_0, X_{(1)}) \\ \vdots \\ G(\theta_0, X_{(n)}) \end{pmatrix} - \begin{pmatrix} G(\theta, X_{(1)}) \\ \vdots \\ G(\theta, X_{(n)}) \end{pmatrix}\right)' \times$$

$$\left(\begin{pmatrix} G(\theta_0, X_{(1)}) \\ \vdots \\ G(\theta_0, X_{(n)}) \end{pmatrix} - \begin{pmatrix} G(\theta, X_{(1)}) \\ \vdots \\ G(\theta, X_{(n)}) \end{pmatrix}\right) =$$

$$\frac{1}{n}\sum_{i=1}^{n} \left(G(\theta_0, X_{(i)}) - G(\theta, X_{(i)}) \right)^2$$

if $\mathbb{E}\left[G(\theta_0, X_{(1)}) - G(\theta, X_{(1)})\right]^2 < +\infty \ \forall \theta \in \Theta$

and an LLN is valid, then it could be plausible that: $\frac{1}{n}\sum_{i=1}^{n}\left(G(\theta_0, X_{(i)}) - G(\theta, X_{(i)})\right)^2 \xrightarrow{P}$

$\mathbb{E}\left[\left(G(\theta_0, X_{(1)}) - G(\theta, X_{(1)})\right)^2\right] \quad \forall \theta \in \Theta$

[Again in time series settings such an LLN could be valid when (Y_n, X_n) is stationary ergodic, e.g. iid]

Hence we assume:

$$\boxed{(b) \quad \tfrac{1}{n} U_n' U_n \xrightarrow{p} \mathbb{E}\left[G(\theta_0, X_{(1)}) - G(\theta, X_{(1)})\right]^2, \forall \theta \in \Theta}$$

---

e.g. (a) $\tfrac{1}{n} U_n' U_n = \tfrac{1}{n} \sum_{i=1}^{n} \left(\exp(X_{(i)} \theta_0) - \exp(X_{(i)} \theta)\right)^2$

$\xrightarrow{p} \mathbb{E}\left[\exp(X_{(1)} \theta_0) - \exp(X_{(1)} \theta)\right]^2$

e.g. (b) $\tfrac{1}{n} U_n' U_n = \tfrac{1}{n} \sum_{i=1}^{n} \left(X_{(i)} \theta_0 - X_{(i)} \theta\right)^2$

$= \frac{1}{n} (\theta_0 - \theta)' X_n' X_n (\theta_0 - \theta) \xrightarrow{p} (\theta_0 - \theta)' \mathbb{E}(X_{(1)} X_{(1)}') \times (\theta_0 - \theta)$

for which it suffices that $\tfrac{1}{n} X_n' X_n \xrightarrow{p} \mathbb{E}(X_{(1)} X_{(1)}')$

Under (a), (b) and due to the

CMT we obtain that:

$$\frac{1}{n} M_n' M_n - \frac{2}{n} M_n' \varepsilon_n \xrightarrow{p} \mathbb{E}\left[\left(G(\theta_0, X_{(1)}) - G(\theta, X_{(1)})\right)^2\right]$$

$\forall \theta \in \Theta$.

Notice also that if

(c) for every $\theta \in \Theta$, and θ^*, θ^{**} in some neighborhood of θ, say B_θ, $\forall x \in \mathbb{R}^k$, $\exists\, k_x > 0$:

$$|G(\theta^*, x) - G(\theta^{**}, x)| \leq k_x \|\theta^* - \theta^{**}\|$$

with $\frac{1}{n}\sum_{i=1}^{n} k_{X_{(i)}} |\varepsilon_{(i)}| = O_p(1)$ and

$$\boxed{1/n \sum_{i=1}^{n} K_{X_{(i)}} = O_P(1)},$$

then using arguments analogous to the LS case the previous can be strengthened to

$$1/n\, M_n' M_n - \frac{2}{n} M_n' \varepsilon_n \xrightarrow{cp} \mathbb{E}\left[G(\theta_0, X_{(1)}) - G(\theta, X_{(1)})\right]^2 \quad (A)$$

e.g. (a) $\qquad K_X = X \max_{\theta^* \in B_\theta} \exp(X\theta)$

and in e.g. (b) $\quad K_X = X$

if furthermore:

(d) $\mathbb{E}\left[\left(G(\theta_0, x_{(1)}) - G(\theta, x_{(1)})\right)^2\right] = 0$
iff $\theta = \theta_0$ (this is the asymptotic identification condition)

Then we have that, since

$$0 \leq \mathbb{E}\left[\left(G(\theta_0, x_{(1)}) - G(\theta, x_{(1)})\right)^2 \quad \forall \theta \in \Theta,\right.$$

$$\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmin}} \; \mathbb{E}\left(G(\theta_0, x_{(1)}) - G(\theta, x_{(1)})\right)^2.$$

---

e.g. (a), if $\operatorname{rank}(X_n) = k$ then
$\left(\exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)\right)^2 \geq 0$ with equality
iff $\theta = \theta_0$, due to that $x \to e^x$ is 1-1.

Hence $\mathbb{E}\left[\left(\exp(X_{(1)}(\theta)) - \exp(X_{(1)}\theta_0)\right)^2\right] = 0$ iff $\theta = \theta_0$.

eg. (B) $\mathbb{E}\left(G(\theta_0, X_{(1)}) - G(\theta, X_{(1)})\right)^2 =$

$\overset{(B)}{=} (\theta - \theta_0)' \mathbb{E}(X_{(1)} X_{(1)}')(\theta - \theta_0)$ and

as long as $\mathbb{E}(X_{(1)} X_{(1)}')$ has rank k the function is zero iff $\theta = \theta_0$. Notice that if $\text{rank}(X_n) = k$ then it can be proven using the theory of quadratic forms that $\mathbb{E}(X_{(1)} X_{(1)}')$ has also rank k.

THEOREM 1:

Thereby, using our general theory we have proven that under a, b, c, d,

$$\theta_n \xrightarrow{P} \theta_0,$$

i.e., the NLLSE is weakly consistent.

* Rate and Limiting Distribution

We assume for simplicity that θ_0 lies in the interior of Θ (e.g. trivial when $\Theta = \mathbb{R}^k$]

e) θ_0 lies in the interior of Θ

Assume also for simplicity that $G(\theta, x)$ is sufficiently smooth:

$\boxed{\text{f) for all } x \in \mathbb{R}^k,\ G(\theta, x) \text{ is twice continuously differentiable w.r.t. } \theta, \text{ for all } \theta \text{ in a neighborhood of } B_{\theta_0}.}$

Notice that this implies that for any such θ, $\frac{\partial \|m_n\|^2}{\partial \theta} =$

$$= \frac{\partial m_n' m_n}{\partial \theta} = \frac{\partial}{\partial \theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left(G(\theta_0, x_{(i)}) - G(\theta_0, x_{(i)}\right)^2\right]$$

$$= -\frac{2}{n}\sum_{i=1}^{n}\left[G(\theta_0, x_{(i)}) - G(\theta, x_{(i)})\right]\frac{\partial G}{\partial \theta}(\theta, x_{(i)}) \quad (*)$$

$$\frac{\partial^2 M_n' M_n}{\partial\theta\,\partial\theta'} =$$

$$= \frac{2}{n}\sum_{i=1}^{n} \frac{\partial G(\theta, X_{(i)})}{\partial\theta}\,\frac{\partial G(\theta, X_{(i)})}{\partial\theta'} \qquad (**)$$

$$- \frac{1}{n}\sum_{i=1}^{n}\left(G(\theta_0, X_{(i)}) - G(\theta, X_{(i)})\right)\frac{\partial^2 G(\theta, X_{(i)})}{\partial\theta\,\partial\theta'}$$

$$\frac{\partial M_n' \varepsilon_n}{\partial\theta} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial G(\theta, X_{(i)})}{\partial\theta}\,\varepsilon(i) \qquad (***),$$

and $$\frac{\partial^2 M_n' \varepsilon_n}{\partial\theta\,\partial\theta'} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 G(\theta, X_{(i)})}{\partial\theta\,\partial\theta'}\,\varepsilon_i \qquad (****)$$

.

In the context of e.g. (α):

$$(*) = -\frac{2}{n}\sum_{i=1}^{n}\left[\exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)\right]\exp(x_{(i)}\theta)x_{(i)}'$$

$$(**) = \frac{2}{n}\sum_{i=1}^{n}\exp(2x_{(i)}\theta)\,x_{(i)}x_{(i)}'$$

$$-\frac{1}{n}\sum_{i=1}^{n}\left[\exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)\right]\left[\exp(x_{(i)}\theta)x_{(i)}x_{(i)}'\right]$$

$$(***) = -\frac{1}{n}\sum_{i=1}^{n}\exp(x_{(i)}\theta)\,x_{(i)}'\,\varepsilon_i$$

$$(****) = -\frac{1}{n}\sum_{i=1}^{n}\exp(x_{(i)}\theta)\,x_{(i)}x_{(i)}'\,\varepsilon_i$$

In the context of e.g. (b)

$$(*) = -\frac{2}{n}\sum_{i=1}^{n} x_{(i)}(\theta_0-\theta)x_{(i)}'$$

$$(**) = \frac{2}{n}\sum_{i=1}^{n} x_{(i)}x_{(i)}'$$

$$(***) = -\frac{1}{n}\sum_{i=1}^{n} x_{(i)}'\varepsilon_i$$

$$(****) = 0_{k\times k}$$

Notice also that in the general case:

$$\frac{\partial M_n' M_n}{\partial \theta}\bigg|_{\theta=\theta_0} = 0_{k\times k}.$$

Given the previous, and following our general theory, we assume:

f) $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial g(\theta_0, X_{(i)})}{\partial \theta} \varepsilon_i \rightsquigarrow Z_{\theta_0} \sim N(0_{k\times 1}, V_{\theta_0})$

for V_{θ_0} a p.d. matrix.

Note: This takes care of the limiting behavior of $r_n \frac{\partial G_n}{\partial \theta}(\theta_0)$ in our general theory. r_n is now the classical rate $\sqrt{n}$.

(f) can be validated in the context of time series, when the CLT in the relevant part of the notes is applicable. For example, when (Y_n, X_n) are iid, then

(f) holds as long as

$$\mathbb{E}\left[\left\|\frac{\partial G(\theta_0, X_{(1)})}{\partial\theta}\varepsilon_{(1)}\right\|^2\right] < +\infty$$

But $\mathbb{E}\left[\left\|\frac{\partial G(\theta_0, X_{(1)})}{\partial\theta}\varepsilon_{(1)}\right\|^2\right] =$

$$= \mathbb{E}\left[\left\|\frac{\partial G(\theta_0, X_{(1)})}{\partial\theta}\right\|^2 \varepsilon_{(1)}^2\right] \overset{L.I.E.}{=}$$

$$\mathbb{E}\left[\mathbb{E}\left[\left\|\frac{\partial G(\theta_0, X_{(1)})}{\partial\theta}\right\|^2 \varepsilon_{(1)}^2 \,\middle|\, \mathcal{G}(X_n)\right]\right]$$

$$= \mathbb{E}\left[\left\|\frac{\partial G(\theta_0, X_{(1)})}{\partial\theta}\right\|^2 \mathbb{E}\left(\varepsilon_{(1)}^2 \,\middle|\, \mathcal{G}(X_n)\right)\right] =$$

$$= \mathbb{E}\left[\left\|\frac{\partial G(\theta_0, X_{(1)})}{\partial\theta}\right\|^2\right].$$ Hence, then

it suffices that $\mathbb{E}\left(\frac{\partial G(\theta_0, X_{(1)})}{\partial\theta_j}\right)^2 < +\infty, \forall j = 1, \dots, k$

For example, in eq. (a), $\frac{\partial G(\theta_0, x_{(i)})}{\partial \theta_j} = \exp(x_{(i)}\theta_0)\, x_{i,j}$, and thereby it suffices that $\mathbb{E}[\exp(2x_{(i)}\theta_0)\, x_{i,j}^2] < +\infty \ \forall j=1,\dots,k$.

In eq. (b), $\frac{\partial G(\theta_0, x_{(i)})}{\partial \theta_j} = x_{i,j}$, and thereby it suffices that $\mathbb{E}(x_{i,j}^2) < +\infty \ \forall j=1,\dots,k$.

Finally we have to take care the local (near θ_0) behaviour of the Hessian of c_n: it can be proven that under the following technical conditions:

(g) $\forall \theta \to \theta_0$, $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial G(\theta, x_{(i)})}{\partial \theta} \frac{\partial G(\theta, x_{(i)})}{\partial \theta'} \xrightarrow{P} \mathcal{U}_{\theta_0}$

where $M_{\theta_0} := \mathbb{E}\left[\frac{\partial G(\theta_0, x_{(i)})}{\partial \theta} \frac{\partial G(\theta_0, x_{(i)})}{\partial \theta'} \right]$,

$$\frac{1}{n} \sum_{i=1}^{n} \sup_{\theta \in B_{\theta_0}} \left\| \frac{\partial G(\theta, x_{(i)})}{\partial \theta} \right\|^2 = O_p(1),$$

$$1/n \sum_{i=1}^{n} \sup_{\theta \in B_{\theta_0}} \left\| \frac{\partial^2 G(\theta, x_{(i)})}{\partial \theta \partial \theta'} \right\| = O_p(1)$$

that $\frac{1}{n} \frac{\partial^2 M_n' M_n}{\partial \theta \partial \theta'} - \frac{2}{n} \frac{\partial^2 M_n' \varepsilon_n}{\partial \theta \partial \theta'} \xrightarrow{p} M_0 \quad \forall \theta \to \theta_0.$

Note: The above can be shown to be valid in contexts of stationarity and ergodicity (e.g. iid), given the smoothness assumption in (f), as long as particular moments related to the regressors exist.

We will not delve more into such considerations. As far as the consequential convergence is concerned, this takes care of the asymptotic behaviour of the Hessian

$\frac{\partial^2 G_n(\theta)}{\partial\theta\partial\theta'}$ $\forall \theta \to \theta_0$ [for the modified pattern $G_n(\theta) = \frac{1}{n} M_n' M_n - \frac{2}{n} M_n' \varepsilon_n$]. In the context of e.g. (α), $M_0 = 2\mathbb{E}[\exp(2X_{(1)}\theta_0) X_{(1)} X_{(1)}']$ while in e.g. (B), $M_0 = 2\,\mathbb{E}[X_{(1)} X_{(1)}']$

The final assumption takes care of the issue of invertibility of the limiting Hessian:

(h) U_0 is p.d.

Note: Given that $\exp(2 x_{(1)} \theta) > 0$ for any value of $x_{(1)}$, in both eq. (α), (b) this would follow as long as $\operatorname{rank} X_n = k$.

Hence using at last our general theory, it is obtained:

THEOREM 2

Under (α)-(h)

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \simeq -U_0^{-1} Z_{\theta_0} \sim N\left(0_{k\times 1}, U_0^{-1} V_{\theta_0} U_0^{-1}\right)$$

$\boxed{\text{Note}}$: If $\sup_n \mathbb{E}\left[\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\partial G(\theta_0, X_{(i)})}{\partial \theta} \varepsilon_{(i)} \right\|^{2+\epsilon}\right] < +\infty$

for some $\epsilon > 0$, then the concept of uniform integrability can be invoked and be further proven that $V_{\theta_0} = U_{\theta_0}$ and thereby,

$$\boxed{\sqrt{n}\,(\theta_n - \theta_0) \sim N\left(0_{k\times 1}, U_{\theta_0}^{-1}\right).}$$

Then, and since U_{θ_0} is latent (both as a function of θ_0, as well as due to its dependence of θ_0), and due to (g), the CLT implies that:

$$U_n^{-1} := \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial G(\theta_n, X_{(i)})}{\partial \theta} \frac{\partial G(\theta_n, X_{(i)})}{\partial \theta'}\right]^{-1} \xrightarrow{p} U_{\theta_0}^{-1}$$

And thereby U_n^{-1} is a consistent estimator of the latent asymptotic variance of NLLSE.

In eq. (a) $U_n^{-1} = \left[\frac{1}{n}\sum_{i=1}^{n} \exp(2x_{(i)}'\theta_n)\, x_{(i)} x_{(i)}'\right]^{-1}$

and in eq. (b) $U_n^{-1} = \left[1/n \sum_{i=1}^{n} x_{(i)} x_{(i)}'\right]^{-1}$.

Due to (h), U_n^{-1} is, at least asymptotically well defined. $\boxtimes$

Note: The previous can be used for the construction of a Wald-test for the hypotheses structure
$H_0: \theta_0 = \theta^*$
$H_1: \theta_0 \neq \theta^*$

Using the statistic

$$n\,(\hat{\theta}_n-\theta_0)'\,\mathcal{U}_n^{-1}\,(\hat{\theta}_n-\theta_0)$$

$$= n^2(\hat{\theta}_n-\theta_0)'\left(\sum_{i=1}^{n}\frac{\partial g(\hat{\theta}_n, X_{(i)})}{\partial\theta}\frac{\partial g(\hat{\theta}_n, X_{(i)})}{\partial\theta'}\right)^{-1}(\hat{\theta}_n-\theta_0).$$

The limiting properties of the procedure fall into our general theory under the totality of the previous assumptions.