APPENDIX TO "ON REGRESSION FOR SAMPLES WITH ALTERNATING PREDICTORS AND ITS APPLICATION TO PSYCHROMETRIC CHARTS"

Mosuk Chow, Bing Li, and Jackie Q. Xue

The Pennsylvania State University and Express Scripts China

This Appendix contains the proofs of the theorems in the paper. We will use numerals such as (A1), (A2) to label the equations generated in the Appendix. Numerals such as (1), (2) are reserved for the equations in the paper.

PROOF OF THEOREM 1. Let $g \in \mathscr{G}$. Then by the property of conditional expectation,

$$\begin{aligned} E[(Y - f(X))^2 | Z = 0] &\le E[(Y - g(X))^2 | Z = 0] \\ E[(X - f^{-1}(Y))^2 | Z = 1] &\le E[(X - g^{-1}(Y))^2 | Z = 1]. \end{aligned} \tag{1}$$

Hence $Q(f) \le Q(g)$.

Suppose f_1 is another function in $\mathscr{G}$ that minimizes $Q(g)$ over $\mathscr{G}$. Then at least one of the inequalities in (1) holds for f_1. Suppose, without loss of generality, the first inequality holds. Since f minimizes $E[(Y - h(X))^2 | Z = 0]$ over all $h \in L_2(P_{X|Z=0})$, we see that

$$E[(Y - f_1(X))^2 | Z = 0] = E[(Y - f(X))^2 | Z = 0].$$

Hence, by the uniqueness of conditional expectation, $f_1 = f$ a.s. $P_{X|Z=0}$. Since $P_{X|Z=0} \equiv P_X$, we have $f_1 = f$ almost surely. □

PROOF OF THEOREM 2. Let $\mathscr{F} = \{(\theta - \theta_0)^\intercal E_n q(U, \theta) : \|\theta - \theta_0\| = \delta\}$. An envelope of this class is

$$F(U) = \sup_{\|\theta - \theta_0\| = \delta} |(\theta - \theta_0)^\intercal q(U, \theta)| \le \|\delta\| \sup_{\|\theta - \theta_0\| = \delta} \|q(U, \theta)\|.$$

By assumption (iii), $E[F(U)] < \infty$. By assumption (iv), the function $\theta \mapsto (\theta - \theta_0)^\intercal q(u, \theta)$ is continuous. Hence, by van der Vaart (1998, Example 19.8), $\mathscr{F}$ is

Glivenko-Cantelli. That is,

$$\sup_{\|\theta-\theta_0\|=\delta} |(\theta-\theta_0)^{\mathsf{T}} E_n q(U,\theta) - (\theta-\theta_0)^{\mathsf{T}} Eq(U,\theta)| \to 0 \tag{2}$$

almost surely. By assumption (ii),

$$Eq(U,\theta) = Eq(U,\theta_0) + E[\partial q(U,\theta_0)/\partial\theta^{\mathsf{T}}](\theta-\theta_0) + o(\|\delta\|).$$

Noticing that $Eq(U,\theta_0)=0$, and multiplying both sides by $(\theta-\theta_0)^{\mathsf{T}}$ from the left, we see that

$$(\theta-\theta_0)^{\mathsf{T}} Eq(U,\theta) = (\theta-\theta_0)^{\mathsf{T}} E[\partial q(U,\theta_0)/\partial\theta^{\mathsf{T}}](\theta-\theta_0) + o(\|\delta\|^2).$$

Thus, for sufficiently small δ,

$$\inf_{\|\theta-\theta_0\|=\delta} (\theta-\theta_0)^{\mathsf{T}} Eq(U,\theta) > 0.$$

This, together with (2), implies that, there is an $\epsilon > 0$ such that

$$\liminf_{n\to\infty} \sup_{\|\theta-\theta_0\|=\delta} (\theta-\theta_0)^{\mathsf{T}} E_n q(U,\theta) > \epsilon$$

almost surely. The assertion of the theorem now follows from Theorem 12.1 of Heyde (1997). □

Proof of Theorem 4. By Taylor's mean value theorem, the law of large numbers, and the stochastic equicontinuity in Theorem 3, it can be shown that

$$\begin{aligned} Q_n(\hat\theta) &= Q_n(\theta) + E_n q^{\mathsf{T}}(U,\theta)(\hat\theta-\theta) + (\hat\theta-\theta)^{\mathsf{T}} J(\theta)(\hat\theta-\theta)/2 + o_P(n^{-1}), \\ Q_n(\tilde\theta) &= Q_n(\theta) + E_n q^{\mathsf{T}}(U,\theta)(\tilde\theta-\theta) + (\tilde\theta-\theta)^{\mathsf{T}} J(\theta)(\tilde\theta-\theta)/2 + o_P(n^{-1}). \end{aligned} \tag{3}$$

Since $E_n q(U,\hat\theta)=0$, by taking Taylor expansion of $E_n q(U,\hat\theta)$ about θ we have

$$0 = E_n q(U,\theta) + J(\theta)(\hat\theta-\theta) + o_P(n^{-1/2}). \tag{4}$$

Substitute this into the first equation in (3) to obtain

$$Q_n(\hat\theta) = Q_n(\theta) - (1/2)[E_n q(U,\theta)]^{\mathsf{T}} J(\theta)[E_n q(U,\theta)] + o_P(n^{-1}).$$

Because $\tilde\theta$ minimizes $Q_n(\theta)$ subject to $h(\theta)=0$ there is a Lagrangian multiplier $\tilde\tau \in \mathbb{R}^r$ such that

$$E_n q(U,\tilde\theta) - H(\tilde\theta)\tilde\tau = 0. \tag{5}$$

Expanding $E_n q(U, \tilde{\theta})$ about θ, we see that

$$E_n q(U,\theta) + J(\theta)(\tilde{\theta} - \theta) - H(\tilde{\theta})\tilde{\tau} = o_P(n^{-1/2}).$$

This implies

$$H^{\mathsf{T}}(\tilde{\theta})J^{-1}(\theta)H(\tilde{\theta})\tilde{\tau} = H^{\mathsf{T}}(\tilde{\theta})J^{-1}(\theta)E_n q(U,\theta) + H^{\mathsf{T}}(\tilde{\theta})(\tilde{\theta} - \theta) + o_P(n^{-1/2}).$$

However, because $h(\theta) = h(\tilde{\theta}) = 0$ we have

$$0 = 0 + H^{\mathsf{T}}(\tilde{\theta})(\tilde{\theta} - \theta) + o_P(n^{-1/2}) \;\Rightarrow\; H^{\mathsf{T}}(\tilde{\theta})(\tilde{\theta} - \theta) = o_P(n^{-1/2}).$$

Hence,

$$H^{\mathsf{T}}(\tilde{\theta})J^{-1}(\theta)H(\tilde{\theta})\tilde{\tau} = H^{\mathsf{T}}(\tilde{\theta})J^{-1}(\theta)E_n q(U,\theta) + o_P(n^{-1/2}).$$

Because $H(\tilde{\theta}) = H(\theta) + o_P(n^{-1/2})$, the above implies

$$\tilde{\tau} = [H^{\mathsf{T}}(\theta)J^{-1}(\theta)H(\theta)]^{-1}H^{\mathsf{T}}(\theta)J^{-1}(\theta)E_n q(U,\theta) + o_P(n^{-1/2}). \tag{6}$$

This implies that $\tilde{\tau} = O_P(n^{-1/2})$.

Since $\tilde{\tau} = O_P(n^{-1/2})$, we can replace $H(\tilde{\theta})$ in (5) by $H(\theta)$ without incurring error greater than $o_P(n^{-1/2})$. That is,

$$E_n q(U,\theta) + J(\tilde{\theta} - \theta) - H(\theta)\tilde{\tau} = o_P(n^{-1/2}).$$

Substitute (6) into the above equation, and then solve for $\tilde{\theta} - \theta$, to obtain

$$\tilde{\theta} - \theta = [J^{-1}(\theta)\Pi(\theta) - J^{-1}(\theta)]E_n q(U,\theta) + o_P(n^{-1/2}),$$

where $\Pi(\theta) = H(\theta)[H^{\mathsf{T}}(\theta)J^{-1}(\theta)H(\theta)]^{-1}H^{\mathsf{T}}(\theta)J^{-1}(\theta)$. Note that $\Pi(\theta)$ is a projection. Substituting the above equation into the right hand side of the second equation in (3), we find

$$Q_n(\tilde{\theta}) = Q_n(\theta) + (1/2)[E_n q(U,\theta)]^{\mathsf{T}}[J^{-1}(\theta)\Pi(\theta) - J^{-1}(\theta)][E_n q(U,\theta)] + o_P(n^{-1}).$$

Subtract (4) from the above equation to obtain

$$Q_n(\tilde{\theta}) - Q_n(\hat{\theta}) = (1/2)[E_n q(U,\theta)]^{\mathsf{T}}J^{-1}(\theta)\Pi(\theta)[E_n q(U,\theta)] + o_P(n^{-1}).$$

The desired result follows because $\sqrt{n}E_n q(U,\theta) \xrightarrow{\mathscr{D}} N(0, I(\theta))$. □

Proof of Theorem 5. The proof is essentially the same as that of Theorem 4 except that here we have m independent i.i.d. samples, instead of one i.i.d. sample, and the ratios of sample sizes must be taken into account when taking the limit $\lim_{n\to\infty}$. Thus instead of using

$$\sqrt{n}\,\partial Q_n(\theta)/\partial\theta \xrightarrow{\mathscr{D}} N(0, J(\theta)), \quad \partial^2 Q_n/\partial\theta\partial\theta^{\mathsf{T}} \xrightarrow{P} I(\theta),$$

as in the proof of Theorem 4, here we use

$$\sqrt{n}\,\partial Q^{(k)}_{n_k}(\theta^{(k)})/\partial\theta^{(k)} \xrightarrow{\mathscr{D}} N(0, \alpha_k J^{(k)}(\theta), \quad \partial^2 Q^{(k)}_{n_k}(\theta^{(k)})/\partial\theta^{(k)}\partial\theta^{(k)\mathsf{T}} \xrightarrow{P} \alpha_k I^{(k)}(\theta^{(k)}).$$

The rest of the proof is omitted. □

Proof of Theorem 6. It is equivalent to show that

$$\mathrm{AV}^{-1}(\beta^*_\theta) \ge \mathrm{AV}^{-1}(\beta_\theta). \tag{7}$$

From the definition (21) and notation in (22) we see that

$$I(\beta_\theta) = E[\beta_\theta(X,Y)v_\theta(X,Y)c(Z)\beta^{\mathsf{T}}_\theta(X,Y)], \; J(\beta_\theta) = E[\gamma_\theta(X,Y)c(Z)\beta^{\mathsf{T}}_\theta(X,Y)].$$

Let

$$\begin{aligned}\eta_\theta(X,Y,Z) &= J(\beta_\theta)I^{-1}(\beta_\theta)\beta_\theta(X,Y)c(Z)\delta_\theta(X,Y)\\ &\quad - J(\beta^*_\theta)I^{-1}(\beta^*_\theta)\beta^*_\theta(X,Y)c(Z)\delta_\theta(X,Y) \equiv A_\theta(X,Y,Z) - B_\theta(X,Y,Z).\end{aligned}$$

Note that

$$E_\theta[\eta^{\mathsf{T}}_\theta(X,Y,Z)v_\theta(X,Y)c(Z)\eta_\theta(X,Y,Z)] \ge 0, \tag{8}$$

where ≥ 0 means being positive semidefinite. The left hand side can be decomposed into four terms:

$$\begin{aligned}&E_\theta A_\theta(X,Y,Z)A^{\mathsf{T}}_\theta(X,Y,Z) - E_\theta A_\theta(X,Y,Z)B^{\mathsf{T}}_\theta(X,Y,Z)\\ &- E_\theta B_\theta(X,Y,Z)A^{\mathsf{T}}_\theta(X,Y,Z) + E_\theta B_\theta(X,Y,Z)B^{\mathsf{T}}_\theta(X,Y,Z).\end{aligned} \tag{9}$$

The first term is

$$\begin{aligned}&E_\theta A_\theta(X,Y,Z)A^{\mathsf{T}}_\theta(X,Y,Z)\\ &\quad = J(\beta_\theta)I^{-1}(\beta_\theta)E_\theta[\beta_\theta(X,Y)c(Z)v_\theta(X,Y)\beta^{\mathsf{T}}_\theta(X,Y)]I^{-1}(\beta_\theta)J(\beta_\theta)\\ &\quad = J(\beta_\theta)I^{-1}(\beta_\theta)J^{\mathsf{T}}(\beta_\theta)\end{aligned} \tag{10}$$

Notice that

$$J(\beta_\theta^*) = E_\theta[\gamma_\theta(X,Y)c(Z)v_\theta^{-1}(X,Y)\gamma_\theta^\mathsf{T}(X,Y)] = I(\beta_\theta^*).$$

From this it is easy to deduce the last three terms in (9) as

$$\begin{aligned} E_\theta[A_\theta(X,Y,Z)B_\theta^\mathsf{T}(X,Y,Z)] &= J(\beta_\theta)I^{-1}(\beta_\theta)J^\mathsf{T}(\beta_\theta) \\ E_\theta B_\theta(X,Y,Z)A_\theta^\mathsf{T}(X,Y,Z) &= J(\beta_\theta)I^{-1}(\beta_\theta)J^\mathsf{T}(\beta_\theta) \\ E_\theta[B_\theta(X,Y,Z)B_\theta^\mathsf{T}(X,Y,Z)] &= I(\beta_\theta^*). \end{aligned} \tag{11}$$

Substituting (10) and (11) into (9), we find

$$I(\beta_\theta^*) \geq J(\beta_\theta)I^{-1}(\beta_\theta)J^\mathsf{T}(\beta_\theta).$$

The inequality (7) now follows from $I(\beta_\theta^*) = J(\beta_\theta^*)I^{-1}(\beta_\theta^*)J^\mathsf{T}(\beta_\theta^*)$, which is a consequence of (8). □

References

Heyde, C. C. (1997). *Quasi-Likelihood and its Application: a General Approach to Optimal Parameter Estimation.* Springer.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

Mosuk Chow, Department of Statistics, The Pennsylvania State University
E-mail: mchow@stat.psu.edu
Bing Li, Department of Statistics, The Pennsylvania State University
E-mail: bing@stat.psu.edu

Jackie Q. Xue, Express Scripts China
E-mail: qxue_2001@yahoo.com