系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
32(1) (2012, 1), 1–14

线性测量误差模型的平均估计*

王 海 鹰

(中国科学院数学与系统科学研究院, 北京 100190;
美国密苏里大学哥伦比亚分校统计学院, 密苏里 65211)

邹 国 华

(中国科学院数学与系统科学研究院, 北京 100190)

摘要 频率模型平均估计近年来受到了较大的关注, 但对有测量误差的观测数据尚未见到任何研究. 文章主要考虑了线性测量误差模型的平均估计问题, 导出了模型平均估计的渐近分布, 基于 Hjort 和 Claeskens (2003) 的思想构造了一个覆盖真实参数的概率趋于预定水平的置信区间, 并证明了该置信区间与基于全模型正态逼近所构造的置信区间的渐近等价性. 模拟结果表明当协变量存在测量误差时, 模型平均估计能明显增加点估计的效率.

关键词 模型选择, 模型平均, 测量误差, 渐近分布.

MR(2000) 主题分类号 62F12, 62J05

1 引 言

模型选择在统计学的研究中有较长的历史, 自 20 世纪 60 年代以来一直是统计学界研究的热点问题. 对于一组观测数据, 我们可以用许多种模型来描述其产生的机制, 究竟哪个模型能更好地反映其真实的产生过程显然是一个十分重要的问题. 为此, 统计学家们提出了许多模型选择的方法和准则, 例如逐步回归, AIC[1], Mallows' C_p [2], 交叉核实 [3], BIC [4], 广义交叉核实 [5], RIC [6] 等等. 根据这些方法或准则, 可以从众多待选模型中挑选出适当的模型, 其后的统计推断都基于这个选定的模型. 然而, 这一过程忽略了模型选择阶段所带来的不确定性, 其直接后果是低估实际的变异, 这样对基于所选模型而得到的估计或预测的精确度难以令人信服, 所报告的置信区间也过于乐观. 文 [7] 首先考虑了统计推断中由于忽略模型选择过程的不确定性而带来的种种问题并利用贝叶斯模型平均方法加以解决, 尽管模型平均估计这个概念其实提出得更早, 如见文 [8]. 文 [7] 通过油价预测和挑战者号航天飞机事故两个例子验证了其方法. 之后, 贝叶斯模型平均方法得到了极大关注, 已有大量文献进

* 国家自然科学基金 (70625004, 11021161) 和中国科学院“百人计划”资助课题.
通讯作者: 邹国华, Email: ghzou@amss.ac.cn.
收稿日期: 2011-03-01.

行讨论. 例如, 文 [9] 讨论了此前提出的一些贝叶斯模型平均方法并针对各种方法举出例子加以说明, 他们还提供了一份可用于贝叶斯模型平均的软件列表; 文 [10] 提出了一种基于列正交设计阵的正态线性回归问题的近似方法, 以解决高维数据时贝叶斯模型平均的计算问题; 文 [11] 总结了贝叶斯模型平均的发展并指出了一些很好的方法; 文 [12] 则提出了一种稳健的贝叶斯模型平均方法.

尽管贝叶斯模型平均近些年发展很快, 但它有明显的缺陷. 首先是先验分布如何确定的问题, 因为不同模型的先验概率对分析结果的影响很大, 如何确定足够好的先验概率是贝叶斯模型平均研究的一个重点, 也是一个难点; 其次是贝叶斯模型平均经常需要对感兴趣的参数假定一些有冲突的先验分布, 这使得贝叶斯模型平均方法难以被直观地接受. 注意到这些困难, 统计学家们也从频率的观点研究了模型平均估计方法. 例如, 文 [13] 为解决模型选择过程带来的不确定性, 提出用 AIC 或 BIC 的指数形式作为权重, 将基于不同子模型得到的估计进行加权. 这种加权方式与贝叶斯模型平均的先验分布方法不同, 是一种完全基于似然的估计形式. 文 [14] 继承了这一思想并更加详尽地阐述了这一方法的应用: 如何量化模型选择阶段的不确定性并将其整合到精度的估计中, 在分析中如何确定协变量的相对重要性以及如何构建模型的 K-L 最优置信集等. 相对于贝叶斯模型平均, 频率模型平均估计的研究要少很多, 但是近些年来这个领域受到了越来越多的关注, 取得了许多重要进展, 可参见文献 [15–21] 等.

频率模型平均估计的一个核心问题是基于不同模型之估计的权重的选取. 最常见的权重选取方式是上面提到的 AIC 和 BIC 指数形式的权重 (即光滑 AIC 权重和光滑 BIC 权重). 文 [16] 提出了一种自适应混合回归 (ARM) 的加权方法; 文 [17] 提出了一种筛选自适应混合回归 (ARMS) 方法来选择待加权的子模型及其权重; 在模型误差为正态分布时, 文 [19] 提出并分析了模型平均估计的性质, 给出了平方损失下模型平均估计风险的无偏估计, 而且基于他们的框架, 在未知何模型为真实模型的前提下, 模型平均估计要优于任何一个基于单个模型的估计. 最近, 文 [20] 提出了一种通过极小化 Mallows 准则来选择权重的方法, 并证明了所得到的模型平均估计 (MMA 估计) 的渐近最优性, 在权重的最优选择方面迈出了实质性的一步. 然而, 他的方法要求回归元有一个具体的排序, 且其优良性是建立在权重属于离散集合的基础之上的, 这是两个较强的条件. 文 [22] 推广了文 [20] 的工作, 在不要求回归元有具体排序的框架下, 对一般的连续权重集合证明了 MMA 估计的渐近最优性. 关于权重选取的其它结果, 可参见文 [21, 23–25] 等.

频率模型平均估计的另一个核心问题是确定估计量的分布. 文 [15] 在局部误设定的框架下, 对一般的参数模型导出了模型平均估计的渐近分布, 这是频率模型平均领域的一个重要进展. 此后对不同的模型, 类似的结果相继被获得, 如文 [18] 将文 [15] 的结果推广到 Cox 风险回归模型; 文 [26] 又将这一结果推广到一般的半参数广义部分线性模型; 文 [27] 则考虑了广义可加部分线性模型. 关于频率模型平均估计的一个综合介绍可参见文 [28]. 然而, 对于观测数据带有测量误差的情形目前还没有见到任何讨论, 本文的目的是将文 [15] 的思想推广到样本独立但不同分布的协变量有测量误差的线性模型.

本文的结构安排如下: 第 2 节是主体部分, 将推导感兴趣参数的模型平均估计的渐近分布, 构造一个覆盖真实参数的概率趋于预定水平的置信区间, 并证明该置信区间与基于全模

型正态逼近所构造的置信区间其实是渐近等价的；第 3 节是模拟结果；第 4 节包含了总结和若干讨论．引理和定理的证明见附录 A.

2　线性测量误差模型的平均估计

2.1　模型假设及估计方法

考虑如下的线性回归模型

$$\boldsymbol{Y}=\boldsymbol{X}_1\beta+\boldsymbol{X}_2\gamma+\boldsymbol{\varepsilon}, \tag{1}$$

其中响应变量 $\boldsymbol{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$ 是 $n\times 1$ 维向量， $\boldsymbol{X}_1=(X_1^{(1)},\cdots,X_n^{(1)})^{\mathrm{T}}$ 和 $\boldsymbol{X}_2=(X_1^{(2)},\cdots,X_n^{(2)})^{\mathrm{T}}$ 分别是 $n\times p$ 维和 $n\times q$ 维的非随机协变量样本矩阵，前者包括了基于理论或专业背景可以确定是相关的协变量，后者包含了不能确定的协变量， β 是 p 维向量，而 γ 是 q 维向量，模型随机误差 $\boldsymbol{\varepsilon}=(\varepsilon_1,\cdots,\varepsilon_n)^{\mathrm{T}}$ 是 $n\times 1$ 维向量且其均值为 0 、协方差矩阵为 $\sigma^2 I_n$. 为方便，我们记 $X_i\triangleq((X_i^{(1)})^{\mathrm{T}},(X_i^{(2)})^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{X}\triangleq(\boldsymbol{X}_1|\boldsymbol{X}_2)$, $\theta\triangleq(\beta^{\mathrm{T}},\gamma^{\mathrm{T}})^{\mathrm{T}}$.

在实际中，协变量 X_i 的精确值常常难以获得，也就是说我们观测到的并不是 X_i, 而是 $W_i=X_i+U_i$, 其中 U_i 是期望为 0 协方差矩阵为

$$\varSigma_u=\begin{pmatrix}\varSigma_{u11} & \varSigma_{u12}\\ \varSigma_{u21} & \varSigma_{u22}\end{pmatrix}$$

且与 ε_i 独立的随机变量. 对于这样的观测数据，如果忽略测量误差而直接采用最小二乘估计

$$\widehat{\theta}_{naive}=(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{Y},$$

其中 $\boldsymbol{W}\triangleq(W_1,\cdots,W_n)^{\mathrm{T}}$, 则得到的估计值在线性可加误差下往往要比真实参数小，更多的结果可参见文 [29]. 为此可以通过极小化下面的修正最小二乘来排除测量误差带来的影响 (如见文 [30])

$$\left\|\boldsymbol{Y}-\boldsymbol{W}\theta\right\|^2-n\theta^{\mathrm{T}}\varSigma_u\theta,$$

其中 $\|\cdot\|$ 为欧几里得范数；类似地，我们定义一个矩阵 $Z=(z_{ij})$ 的欧几里得范数为 $\|Z\|=\big(\sum\limits_{i,j}z_{ij}^2\big)^{\frac{1}{2}}$. 极小化上式得到的参数估计是

$$\widehat{\theta}=\left\{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}-n\varSigma_u\right\}^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{Y},$$

即

$$\begin{pmatrix}\widehat{\beta}\\ \widehat{\gamma}\end{pmatrix}=\begin{pmatrix}\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1-n\varSigma_{u11} & \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2-n\varSigma_{u12}\\ \boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_1-n\varSigma_{u21} & \boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_2-n\varSigma_{u22}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{Y}\\ \boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{Y}\end{pmatrix}. \tag{2}$$

利用分块矩阵的求逆公式，通过直接的运算可以得到

$$\begin{aligned}\widehat{\beta}&=\left(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1-n\varSigma_{u11}\right)^{-1}\left[\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{Y}-(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2-n\varSigma_{u12})\widehat{\gamma}\right],\\ \widehat{\gamma}&=A_n^{-1}\left\{\boldsymbol{W}_2^{\mathrm{T}}-(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_1-n\varSigma_{u21})(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1-n\varSigma_{u11})^{-1}\boldsymbol{W}_1^{\mathrm{T}}\right\}\boldsymbol{Y},\end{aligned} \tag{3}$$

其中 $A_n=\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_2-n\varSigma_{u22}-\left(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_1-n\varSigma_{u21}\right)\left(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1-n\varSigma_{u11}\right)^{-1}\left(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2-n\varSigma_{u12}\right)$.

下面考虑基于由 $\{1, 2, \cdots, q\}$ 的子集 S 所代表的子模型的估计问题. 我们利用 $q \times |S|$ 的选择矩阵 V_S 来选择我们想要包含在子模型中的 γ 的分量，$|S|$ 为子模型中包含的 γ 分量的个数. 于是 $\gamma_S = V_S^{\mathrm{T}}\gamma$, 也就是说 V_S^{T} 每行的 q 个元素中，只有与我们想要保留在子模型中的协变量相对应的元素为 1, 其余均为 0. 相应地，记 $\boldsymbol{X}_S \triangleq (\boldsymbol{X}_1|\boldsymbol{X}_{2S}) = (\boldsymbol{X}_1|\boldsymbol{X}_2 V_S)$ 为子模型 S 假设下的协变量矩阵. 由此，假设的子模型可表示为

$$\boldsymbol{Y} = \boldsymbol{X}_1\beta + \boldsymbol{X}_{2S}\gamma_S + \boldsymbol{\varepsilon}.$$

通过与全模型下的估计方法类似的推导，我们可以得到基于子模型的参数估计如下：

$$\begin{aligned}
\widehat{\beta}_S &= \left(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1 - n\Sigma_{u11}\right)^{-1}\left[\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{Y} - (\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2 - n\Sigma_{u12})V_S\widehat{\gamma_S}\right], \\
\widehat{\gamma}_S &= \left(V_S^{\mathrm{T}}A_nV_S\right)^{-1}V_S^{\mathrm{T}}\left\{\boldsymbol{W}_2^{\mathrm{T}} - (\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_1 - n\Sigma_{u21})(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1 - n\Sigma_{u11})^{-1}\boldsymbol{W}_1^{\mathrm{T}}\right\}Y.
\end{aligned} \tag{4}$$

由此不难得到回归系数在全模型下的估计与基于子模型的估计之间的关系，下面以引理的形式将其给出.

引理 2.1 在全模型下回归系数的估计与基于子模型 S 回归系数的估计之间有如下关系成立

$$\begin{pmatrix}\widehat{\beta}_S \\ \widehat{\gamma}_S\end{pmatrix} = \begin{pmatrix} I_p & C_{nS} \\ 0_{|S|\times p} & (V_S^{\mathrm{T}}A_nV_S)^{-1}V_S^{\mathrm{T}}A_n\end{pmatrix}\begin{pmatrix}\widehat{\beta} \\ \widehat{\gamma}\end{pmatrix} \triangleq G_{nS}\begin{pmatrix}\widehat{\beta} \\ \widehat{\gamma}\end{pmatrix},$$

其中 $C_{nS} = (\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1 - n\Sigma_{u11})^{-1}\left(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2 - n\Sigma_{u12}\right)\left(I - A_n^{-\frac{1}{2}}H_{nS}A_n^{\frac{1}{2}}\right)$, $H_{nS} = A_n^{\frac{1}{2}}V_S(V_S^{\mathrm{T}} \cdot A_nV_S)^{-1}V_S^{\mathrm{T}}A_n^{\frac{1}{2}}$ 为一投影矩阵.

以下我们采用文 [15] 的局部误设定框架而假设 $\gamma = \frac{\delta}{\sqrt{n}}$, 其中 δ 为 q 维未知的常数向量，并定义 $\theta_0 = (\beta^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$, $\theta_S = (\beta^{\mathrm{T}}, \gamma_S^{\mathrm{T}})^{\mathrm{T}}$.

在这一前提下，可以得到关于回归系数估计的渐近性质，下面以引理的形式给出，它是后面定理证明的基础.

引理 2.2 假设存在 $t > 0$ 满足 $\sup\limits_i \boldsymbol{E}|\varepsilon_i|^{2+t} < \infty$ 和 $\sup\limits_i \boldsymbol{E}\|U_i\|^{4+2t} < \infty$, 且 $\sup\limits_i \|X_i\| < \infty$, $\frac{1}{n}\sum\limits_{i=1}^{n} X_iX_i^{\mathrm{T}} \to B \triangleq \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{pmatrix}$, 其中 B 为非随机正定矩阵. 进一步假定 U_i 的分布对称、有相同的前四阶矩，则估计量 (4)有渐近正态性，即当 $n \to \infty$ 时有下式成立

$$\sqrt{n}\begin{pmatrix}\widehat{\beta}_S - \beta \\ \widehat{\gamma}_S\end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} C_S\delta \\ (V_S^{\mathrm{T}}AV_S)^{-1}V_S^{\mathrm{T}}A\delta\end{pmatrix}, G_SB^{-1}FB^{-1}G_S^{\mathrm{T}}\right), \tag{5}$$

特别地

$$\sqrt{n}\begin{pmatrix}\widehat{\beta} - \beta \\ \widehat{\gamma}\end{pmatrix} \triangleq \begin{pmatrix} M_{n1} \\ M_{n2}\end{pmatrix} \xrightarrow{d} \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \delta\end{pmatrix}, B^{-1}FB^{-1}\right), \tag{6}$$

其中 $C_S = B_{11}^{-1}B_{12}\big(I_q - A^{-\frac{1}{2}}H_SA^{\frac{1}{2}}\big)$, $H_S = A^{\frac{1}{2}}V_S(V_S^{\mathrm{T}}AV_S)^{-1}V_S^{\mathrm{T}}A^{\frac{1}{2}}$, $A = B_{22} - B_{21}B_{11}^{-1}B_{12}$, $G_S = \begin{pmatrix} I_p & C_S \\ 0_{|S|\times p} & (V_S^{\mathrm{T}}AV_S)^{-1}V_S^{\mathrm{T}}A\end{pmatrix}$, $F = \sigma^2B + \sigma^2\Sigma_u + B\theta_0^{\mathrm{T}}\Sigma_u\theta_0 + \boldsymbol{E}(UU^{\mathrm{T}}\theta_0 - \Sigma_u\theta_0)^{\otimes 2}$, 而 $Z^{\otimes 2}$ 代表 ZZ^{T}.

证 见附录 A.

对于实际的数据，Σ_u 经常是未知的. 若我们对协变量 X_i 有重复观测值：$W_{ij}=X_i+U_{ij}$, $j=1,2,\cdots,J_i$, $i=1,2,\cdots,n$, 那么可以用下式估计 Σ_u, 此估计是无偏和相合的 (见文 [31–32]):

$$\widehat{\Sigma}_u=\frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{J_i}(W_{ij}-\overline{W}_i)(W_{ij}-\overline{W}_i)^{\mathrm{T}}}{\sum\limits_{i=1}^{n}(J_i-1)},$$

其中 $\overline{W}_i=\sum\limits_{j=1}^{J_i}\frac{W_{ij}}{J_i}$. 在这种情况下, 如果我们使用 $\widehat{\Sigma}_u$ 来代替未知的 Σ_u, 则前述的理论结果依然成立. 注意到 $\overline{U}_i=\sum\limits_{j=1}^{J_i}\frac{U_{ij}}{J_i}$ 的协方差矩阵比单独 U_{ij} 的要小, 所以我们应使用 $\overline{W}_i=X_i+\overline{U}_i$ 作为回归协变量，这样可以减少测量误差的影响并且能够提高计算的稳定性.

2.2　感兴趣参数的模型平均估计及其性质

假设我们感兴趣的参数是回归系数的函数 $\mu_{\text{true}}=\mu(\theta)=\mu(\beta,\gamma)$, 并记在子模型 S 下参数 μ 的估计为 $\widehat{\mu}_S=\mu(\widehat{\beta}_S,\widehat{\gamma}_S)$. 可以证明 $\widehat{\mu}_S$ 有下面的渐近性质

定理 2.1　设 μ 在 θ_0 处可导，且引理 2.2 的条件成立，则基于子模型 S 的估计 $\widehat{\mu}_S$ 具有渐近正态性，即

$$\sqrt{n}(\widehat{\mu}_S-\mu_{\text{true}})\xrightarrow{d}\Lambda_S=\mu_\beta^{\mathrm{T}}\left[M_1+B_{11}^{-1}B_{12}(M_2-\delta)\right]+\omega^{\mathrm{T}}\left(\delta-A^{-\frac{1}{2}}H_SA^{\frac{1}{2}}M_2\right),\tag{7}$$

其中 $\mu_\beta=\frac{\partial\mu(\theta_0)}{\partial\beta}$, $\mu_\gamma=\frac{\partial\mu(\theta_0)}{\partial\gamma}$, $\omega=B_{21}B_{11}^{-1}\mu_\beta-\mu_\gamma$.

证　见附录 A.

本文考虑如下形式的模型平均估计

$$\widehat{\mu}=\sum_S c(S|\widehat{\delta})\widehat{\mu}_S\tag{8}$$

其中 $\widehat{\delta}=\sqrt{n}\widehat{\gamma}=M_{n2}$, $c(S|\cdot)$ 为权函数. 对于这种形式的估计量，我们实际上假设了其权重只通过参数 γ 在全模型下的估计而依赖于数据. 有关这种形式估计量的更多论述可参见文 [15, 33]. 下面的定理给出了 $\widehat{\mu}$ 的渐近性质：

定理 2.2　设 μ 在 θ_0 处可导，权函数 $c(S|d)$ 关于 d 连续，且对任意的 d, $\sum\limits_S c(S|d)=1$. 若引理 2.2 的条件成立，则有

$$\begin{aligned}
\sqrt{n}(\widehat{\mu}-\mu_{\text{true}})\xrightarrow{d}\Lambda&=\mu_\beta^{\mathrm{T}}\Big[M_1+B_{11}^{-1}B_{12}(M_2-\delta)\Big]+\omega^{\mathrm{T}}\Big(\delta-Q(M_2)M_2\Big),\\
\boldsymbol{E}\Lambda&=\omega^{\mathrm{T}}\Big(\delta-\boldsymbol{E}[Q(M_2)M_2]\Big),\\
\mathrm{Var}(\Lambda)&=\mu_\beta^{\mathrm{T}}(I,B_{11}^{-1}B_{12})P(I,B_{11}^{-1}B_{12})^{\mathrm{T}}\mu_\beta+\omega^{\mathrm{T}}\mathrm{Var}\big[Q(M_2)M_2\big]\omega\\
&\quad-2\mu_\beta^{\mathrm{T}}(I,B_{11}^{-1}B_{12})\mathrm{Cov}\big[(M_1^{\mathrm{T}},M_2^{\mathrm{T}})^{\mathrm{T}},Q(M_2)M_2\big]\omega,
\end{aligned}$$

其中 $P=B^{-1}FB^{-1}$, $Q(M_2)=A^{-\frac{1}{2}}\big(\sum_S c(S|M_2)H_S\big)A^{\frac{1}{2}}$. 如果协变量 X 没有测量误差, 那么 $M_1+B_{11}^{-1}B_{12}(M_2-\delta)$ 和 M_2 是相互独立的，此时随机变量 Λ 的方差可简化为 $\sigma^2\mu_\beta^{\mathrm{T}}B_{11}^{-1}\mu_\beta$ $+\omega^{\mathrm{T}}\mathrm{Var}\big[Q(M_2)M_2\big]\omega$.

证　见附录 A.

2.2 置信区间

由于模型平均估计量 $\widehat{\mu}$ 的极限分布一般不是正态的，本节我们利用文 [15] 提出的方法，在有测量误差的情况下构造一个覆盖概率趋于预期水平的置信区间，并进一步证明这个置信区间与基于全模型正态逼近所构造的置信区间是渐近等价的.

若假设引理 2.2 的各条件得到满足，且 ω 和 $\kappa=\sqrt{(\mu_\beta^{\mathrm{T}},\mu_\gamma^{\mathrm{T}})P(\mu_\beta^{\mathrm{T}},\mu_\gamma^{\mathrm{T}})^{\mathrm{T}}}$ 分别有相合的估计 $\widehat{\omega}$ 和 $\widehat{\kappa}$, z 为标准正态分布的分位数，则可构造置信区间如下

$$\begin{aligned}\mathrm{low}_n&=\widehat{\mu}-\frac{\widehat{\omega}^{\mathrm{T}}[M_{n2}-Q_n(M_{n2})M_{n2}]}{\sqrt{n}}-\frac{z\widehat{\kappa}}{\sqrt{n}},\\ \mathrm{up}_n&=\widehat{\mu}-\frac{\widehat{\omega}^{\mathrm{T}}[M_{n2}-Q_n(M_{n2})M_{n2}]}{\sqrt{n}}+\frac{z\widehat{\kappa}}{\sqrt{n}},\end{aligned}\tag{9}$$

其中 $Q_n(M_{n2})=A_n^{-\frac{1}{2}}\big(\sum_S c(S|M_{n2})H_{nS}\big)A_n^{\frac{1}{2}}$. 不难证明此置信区间覆盖真实参数的概率随着 n 的增大趋近于预期的水平 (证明见附录 A), 即

$$\Pr\big\{\mu_{\mathrm{true}}\in(\mathrm{low}_n,\mathrm{up}_n)\big\}\to 2\Phi(z)-1,$$

其中 Φ 为标准正态分布函数.

现在我们研究 (9) 所给出的置信区间的本质. 首先注意式 (7) 的一个特殊情况是

$$\begin{aligned}\sqrt{n}(\widehat{\mu}_{\mathrm{full}}-\mu_{\mathrm{true}})&\xrightarrow{d}\mu_\beta^{\mathrm{T}}[M_1+B_{11}^{-1}B_{12}(M_2-\delta)]+\omega^{\mathrm{T}}(\delta-M_2)\\ &=\mu_\beta^{\mathrm{T}}M_1+\mu_\gamma^{\mathrm{T}}(M_2-\delta),\end{aligned}$$

其中上式右端的分布为正态 $N(0,\kappa^2)$. 基于此正态分布，可以构造 μ 的另一个置信区间

$$\mathrm{low}_{\mathrm{full}}=\widehat{\mu}_{\mathrm{full}}-\frac{z\widehat{\kappa}}{\sqrt{n}},\quad \mathrm{up}_{\mathrm{full}}=\widehat{\mu}_{\mathrm{full}}+\frac{z\widehat{\kappa}}{\sqrt{n}}.\tag{10}$$

下面我们考察式 (9) 与 (10) 中两个置信区间的联系. 可以证明， $\mathrm{low}_n=\mathrm{low}_{\mathrm{full}}+o_P(\frac{1}{\sqrt{n}})$ 以及 $\mathrm{up}_n=\mathrm{up}_{\mathrm{full}}+o_P(\frac{1}{\sqrt{n}})$, 这意味着这两个置信区间在大样本的情况下是等价的. 为证明这一结论，我们只需证

$$\widehat{\mu}_{\mathrm{full}}=\widehat{\mu}-\frac{\omega^{\mathrm{T}}[M_{n2}-Q_n(M_{n2})M_{n2}]}{\sqrt{n}}+o_P\Big(\frac{1}{\sqrt{n}}\Big).\tag{11}$$

一方面，由式 (8) 中 $\widehat{\mu}$ 的定义有

$$\begin{aligned}\widehat{\mu}=&\mu(\beta,0)+\sum_S c(S|\widehat{\delta})\begin{pmatrix}\mu_\beta\\ V_S^{\mathrm{T}}\mu_\gamma\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\widehat{\beta}_S-\beta\\ \widehat{\gamma}_S\end{pmatrix}+o_P\Big(\frac{1}{\sqrt{n}}\Big)\\ =&\mu(\beta,0)+\mu_\beta^{\mathrm{T}}(\widehat{\beta}-\beta)+\sum_S c(S|\widehat{\delta})\Big(\mu_\beta^{\mathrm{T}}C_{ns}\widehat{\gamma}+\mu_\gamma^{\mathrm{T}}A_n^{-\frac{1}{2}}H_{nS}A_n^{\frac{1}{2}}\widehat{\gamma}\Big)+o_P\Big(\frac{1}{\sqrt{n}}\Big)\\ =&\mu(\beta,0)+\mu_\beta^{\mathrm{T}}(\widehat{\beta}-\beta)\\ &+\sum_S c(S|\widehat{\delta})\Big((\omega^{\mathrm{T}}+\mu_\gamma^{\mathrm{T}})(I_q-A_n^{-\frac{1}{2}}H_{nS}A_n^{\frac{1}{2}})+\mu_\gamma^{\mathrm{T}}A_n^{-\frac{1}{2}}H_{nS}A_n^{\frac{1}{2}}\Big)\frac{M_{n2}}{\sqrt{n}}+o_P\Big(\frac{1}{\sqrt{n}}\Big)\\ =&\mu(\beta,0)+\mu_\beta^{\mathrm{T}}(\widehat{\beta}-\beta)+\mu_\gamma^{\mathrm{T}}\widehat{\gamma}+\frac{\omega^{\mathrm{T}}[M_{n2}-Q_n(M_{n2})M_{n2}]}{\sqrt{n}}+o_P\Big(\frac{1}{\sqrt{n}}\Big).\end{aligned}\tag{12}$$

另一方面，由 Taylor 展开得

$$\widehat{\mu}_{\rm full}=\mu(\beta,0)+\mu_\beta^{\rm T}(\widehat{\beta}-\beta)+\mu_\gamma^{\rm T}\widehat{\gamma}+o_P\Big(\frac{1}{\sqrt{n}}\Big). \tag{13}$$

比较式 (12) 与 (13) 知式 (11) 成立．若进一步假设 μ 是回归系数 β 和 γ 的线性函数，那么 $\widehat{\mu}$ 与 $\widehat{\mu}_{\rm full}$ 在 θ_0 处 Taylor 展开的余项都为零．注意到 κ^2 和 ω 与全模型有关而与特定的子模型无关，所以在实际估计时 $\widehat{\kappa}$ 与 $\widehat{\omega}$ 应为基于全模型的估计．这样若感兴趣的参数为回归系数的线性组合，那么式 (9) 与 (10) 中的两个置信区间将完全相同．

3 模拟研究

考虑从下面的线性模型产生数据，

$$Y=X_1\beta_1+X_2\beta_2+X_3\beta_3+X_4\gamma_1+X_5\gamma_2+X_6\gamma_3+X_7\gamma_4+\varepsilon,$$

其中 $X_1,\cdots,X_7$ 是联合分布为多元正态分布的协变量，且它们均值均为 0 、方差均为 1, 对它们之间的相关性我们考虑两种情况：其一为它们之间相互独立，其二是它们之间的相关系数为 0.5; 模型误差 ε 服从标准正态分布．由于有测量误差，我们用 $W=X+U$ 代替 X 作估计，其中 U 服从均值为 0 、协方差矩阵为 $\sigma_u^2I_7$ 的多元正态分布．关于 σ_u 我们也考虑两种情况，分别为 $\sigma_u^{(1)}=0.1$ 和 $\sigma_u^{(2)}=0.2$. $\theta=(\beta^{\rm T},\gamma^{\rm T})^{\rm T}=\Big(\beta^{\rm T},\frac{\delta^{\rm T}}{\sqrt{n}}\Big)^{\rm T}$, 且 $\beta=(\beta_1,\beta_2,\beta_3)^{\rm T}=(3,1.5,2)^{\rm T}$, 而对于 $\delta=(\delta_1,\delta_2,\delta_3,\delta_4)^{\rm T}$, 我们考虑三种情况，分别为 $\delta^{(1)}=(0,0,0,0)^{\rm T}$, $\delta^{(2)}=(4,0,4,0)^{\rm T}$, 和 $\delta^{(3)}=(4,4,4,4)^{\rm T}$. 假设感兴趣的参数是回归系数的线性函数，即 $\mu_{\rm true}=\ell^{\rm T}\theta$, 其中 ℓ 为常数向量．考虑的两个参数 μ_1 和 μ_2 分别对应于 $\ell_1=(1,1,-2,0,0,0,0)^{\rm T}$ 和 $\ell_2=(1,1,1,1,1,1,1)^{\rm T}$. 对于模型平均估计，我们使用光滑 AIC 权重和光滑 BIC 权重，其表达式分别为

$$\frac{\exp\big(-\frac{1}{2}{\rm AIC}_{ns}\big)}{\sum\limits_S\exp\big(-\frac{1}{2}{\rm AIC}_{ns}\big)} \qquad 和 \qquad \frac{\exp\big(-\frac{1}{2}{\rm BIC}_{ns}\big)}{\sum\limits_S\exp\big(-\frac{1}{2}{\rm BIC}_{ns}\big)}.$$

针对各种不同情况，分别抽取 1000 个独立样本，然后计算均方误差 $MSE=\frac{1}{T}\sum\limits_{t=1}^{T}\big(\widehat{\mu}^{(t)}-\mu\big)^2$, 其中 $\widehat{\mu}^{(t)}$ 为基于第 t 个样本的估计值．结果见表 1 和表 2.

从表 1 和表 2 不难看出在大多数情况下，模型平均估计在精度上 (MSE) 相对于模型选择 (AIC 或 BIC) 有较大的优势，尤其是当协变量的测量误差较大时，模型平均估计的优势更加明显．当 δ 取为 $\delta^{(1)}$ 或 $\delta^{(2)}$ 时，基于全模型的估计往往是最不精确的，这说明当有些回归系数的数值较小时，进行模型选择或者模型平均是十分必要的，特别是当协变量有较大测量误差的时候，进行模型平均很有必要．自然，如果 $\delta=\delta^{(3)}$, 则基于全模型的估计一般是最好的，因为此时全模型是真实模型．

表 1 协变量相互独立时各估计方法的 MSE

		$\delta^{(1)}$		$\delta^{(2)}$		$\delta^{(3)}$	
		$\sigma_u^{(1)}$	$\sigma_u^{(2)}$	$\sigma_u^{(1)}$	$\sigma_u^{(2)}$	$\sigma_u^{(1)}$	$\sigma_u^{(2)}$
n=70							
μ_1	光滑 AIC	0.1044	0.7146	0.1080	0.7563	0.1117	0.7789
	AIC	0.1058	0.7242	0.1082	0.7600	0.1102	0.7778
	光滑 BIC	0.1027	**0.7038**	0.1085	**0.7539**	0.1142	0.7814
	BIC	**0.1022**	0.7101	**0.1079**	0.7581	0.1127	0.7852
	全模型	0.1095	0.7301	0.1094	0.7603	**0.1098**	**0.7771**
μ_2	光滑 AIC	0.0843	0.9763	**0.1181**	1.1023	0.1538	**1.2183**
	AIC	0.1003	1.0067	0.1195	1.1184	0.1378	1.2225
	光滑 BIC	**0.0684**	**0.9282**	0.1229	**1.0844**	0.1857	1.2234
	BIC	0.0748	0.9698	0.1239	1.1037	0.1768	1.2490
	全模型	0.1303	1.0560	0.1306	1.1379	**0.1314**	1.2198
n=120							
μ_1	光滑 AIC	0.0632	0.3337	**0.0637**	0.3421	0.0643	**0.3450**
	AIC	0.0634	0.3356	0.0638	0.3432	0.0640	0.3451
	光滑 BIC	**0.0630**	**0.3315**	0.0638	**0.3406**	0.0649	0.3457
	BIC	0.0632	0.3329	0.0638	0.3416	0.0648	0.3457
	全模型	0.0641	0.3367	0.0641	0.3443	**0.0640**	0.3451
μ_2	光滑 AIC	0.0452	0.4481	**0.0637**	0.4947	0.0817	0.5316
	AIC	0.0513	0.4661	0.0648	0.5030	0.0738	0.5362
	光滑 BIC	**0.0356**	**0.4237**	0.0673	**0.4895**	0.1037	0.5416
	BIC	0.0377	0.4437	0.0669	0.5050	0.0990	0.5483
	全模型	0.0708	0.4796	0.0713	0.5049	**0.0715**	**0.5245**

表 2 协变量间相关系数为 0.5 时各估计方法的 MSE

		$\delta^{(1)}$		$\delta^{(2)}$		$\delta^{(3)}$	
		$\sigma_u^{(1)}$	$\sigma_u^{(2)}$	$\sigma_u^{(1)}$	$\sigma_u^{(2)}$	$\sigma_u^{(1)}$	$\sigma_u^{(2)}$
n=70							
μ_1	光滑 AIC	0.2121	2.8273	**0.2215**	2.9838	0.2302	3.0402
	AIC	0.2151	2.8458	0.2222	3.0080	0.2288	3.0560
	光滑 BIC	**0.2090**	**2.7667**	0.2236	**2.9521**	0.2371	**3.0318**
	BIC	0.2106	2.7668	0.2249	2.9741	0.2421	3.0575
	全模型	0.2220	2.9227	0.2218	3.0325	**0.2226**	3.0524
μ_2	光滑 AIC	0.0299	0.2244	**0.0326**	0.2344	0.0347	0.2433
	AIC	0.0320	0.2270	0.0334	0.2352	0.0347	0.2445
	光滑 BIC	**0.0286**	**0.2207**	0.0334	**0.2332**	0.0368	**0.2429**
	BIC	0.0292	0.2245	0.0346	0.2358	0.0376	0.2443
	全模型	0.0326	0.2297	0.0326	0.2381	**0.0328**	0.2451

(续表 2)

		$\delta^{(1)}$		$\delta^{(2)}$		$\delta^{(3)}$	
		$\sigma_u^{(1)}$	$\sigma_u^{(2)}$	$\sigma_u^{(1)}$	$\sigma_u^{(2)}$	$\sigma_u^{(1)}$	$\sigma_u^{(2)}$
n=120							
μ_1	光滑 AIC	0.1272	0.9458	**0.1286**	0.9659	0.1308	0.9619
	AIC	0.1276	0.9510	0.1287	0.9740	0.1305	0.9615
	光滑 BIC	**0.1268**	**0.9362**	0.1292	**0.9599**	0.1331	0.9636
	BIC	0.1271	0.9429	0.1300	0.9688	0.1341	0.9668
	全模型	0.1294	0.9571	0.1295	0.9718	**0.1292**	**0.9614**
μ_2	光滑 AIC	0.0169	0.0980	0.0187	0.1015	0.0198	0.1031
	AIC	0.0171	0.0994	0.0187	0.1021	0.0194	0.1034
	光滑 BIC	**0.0159**	**0.0966**	0.0195	**0.1010**	0.0215	0.1033
	BIC	0.0160	0.0980	0.0198	0.1023	0.0217	0.1040
	全模型	0.0186	0.0999	**0.0187**	0.1022	**0.0187**	**0.1030**

4 总 结

本文在线性模型的协变量不能精确观测的情况下导出了模型平均估计的渐近分布，并证明了基于文 [15] 的方法得到的置信区间事实上渐近等同于基于全模型估计的置信区间. 对于参数模型情形，两种置信区间的渐近等价性在文 [15] 中已通过模拟研究观察到，文 [34] 则给出了一个理论证明概要. 这样的等价性说明如果按文 [15] 的方法作区间估计，则模型平均估计中权重的选择是不重要的. 然而数值模拟的结论告诉我们模型平均的点估计一般要优于全模型，因此如何利用模型平均估计构造感兴趣参数的更好的置信区间仍是一个有待解决的问题. 另一方面，本文假定了模型的误差方差是相同的. 在实际问题中，模型误差为异方差的情况也很常见. 尽管我们的方法对于异方差情形也适用，但因异方差时普通最小二乘估计没有最优性质，所以如何利用加权最小二乘来解决模型平均中的异方差问题还需进一步研究. 此外，在其它更复杂的测量误差模型下研究模型平均方法也是有趣的问题.

参 考 文 献

[1] Akaike H. Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory, B. N. Petrov and F. Cźaki eds, Akademiai Kaidó, Budapest, 1973.

[2] Mallows C L. Some comments on C_p. *Technometrics*, 1973, **15**: 661–675.

[3] Stone M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B*, 1974, **36**: 111–147.
[4] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 1978, **6**: 461–464.
[5] Craven P, Wahba G. Smoothing noisy data with spline functions. *Numerische Mathematik*, 1979, **31**: 377–403.
[6] Foster D P, George E I. The risk inflation criterion for multiple regression. *The Annals of Statistics*, 1994, **22**: 1947–1975.
[7] Draper D. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B*, 1995, **57**: 45–97.
[8] Barnard G A. New methods of quality control. *Journal of the Royal Statistical Society Series A*, 1963, **126**: 255–258.
[9] Hoeting J A, Madigan D, Raftery A E, Volinsky C T. Bayesian model averaging: A tutorial. *Statistical Science*, 1999, **14**: 382–417.
[10] Clyde M. Bayesian model averaging and model search strategies (with discussion). *Bayesian Statistics*, 1999, **6**: 157–185.
[11] Clyde M, George E I. Model uncertainty. *Statistical Science*, 2004, **19**: 81–94.
[12] Doppelhofer G, Weeks M. Robust model averaging. Tech. Rep., CESifo, 2008.
[13] Buckland S T, Burnham K P, Augustin N H. Model selection: An integral part of inference. *Biometrics*, 1997, **53**: 603–618.
[14] Burnham K P, Andreson D R. Model Selection and Multimodel Inference: A Practical Infromation-Theoretic Approach, New York: Springer, 2006.
[15] Hjort N L, Claeskens G. Frequestist model average estimators. *Journal of the American Statistical Association*, 2003, **98**: 879–899.
[16] Yang Y. Adaptive regression by mixing. *Journal of the American Statistical Association*, 2001, **96**: 574–586.
[17] Yuan Z, Yang Y. Combining linear regression models: When and how? *Journal of the American Statistical Association*, 2005, **100**: 1202–1214.
[18] Hjort N L, Claeskens G. Focused information criteria and model averaging for the cox hazard regression model. *Journal of the American Statistical Association*, 2006, **101**: 1449–1464.
[19] Leung G, Barron A R. Infromation theory and mixing least-squares regressions. *IEEE Transactions on Information Theory*, 2006, **52**: 3396–3410.
[20] Hansen B E. Least squares model averaging. *Econometrica*, 1997, **75**: 1175–1189.
[21] Hansen B E. Least squares forecast averaging. *Journal of Econometrics*, 2008, **146**: 342–350.
[22] Wan A T K, Zhang X, Zou G. Least squares model averaging by Mallows criterion. *Journal of Econometrics*, 2010, **156**: 277–283.
[23] Hansen B E. Averaging estimators for regressions with a possible structural break. *Econometric Theory*, 2009, **25**: 1498–1514.
[24] Hansen B E. Averaging estimators for autoregressions with a near unit root. *Journal of Econometrics*, 2010, **158**: 142–155.
[25] Zou G, Wan A T K, Wu X, Chen T. Estimation of regression coefficients of interest when other regression coefficient are of no interest: The case of non-normal errors. *Statistics and Probability Letters*, 2007, **77**: 803–810.
[26] Claeskens G, Carroll R J. An asymptotic theory for model selection inference in general semiparametric problems. *Biometrika*, 2007, **94**: 249–265.
[27] Zhang X, Liang H. Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics*, **39**: 174–200.
[28] Wang H, Zhang X, Zou G. Frequentist model averaging estimation: A review. *Journal of Systems Science & Complexity*, 2009, **22**(4): 732–748.
[29] Fuller W A. Measurement Error Models. New York: Wiley, 1997.

[30] Liang H, Härdle W, Carroll R J. Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics*, 1999, **27**: 1519–1535.

[31] Carroll R J, Ruppert D, Stefanski L A, Crainiceanu C M. Measurement Error in Nonlinear Models: A Modern Perspective. New York: Chapman and Hall, 2006.

[32] Liang H, Li R. Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association*, 2009, **104**: 234–248.

[33] Liang H, Zou G, Wan A T K, Zhang X. Optimal weight choice for frequentist model average estimators. *Journal of the American Statistical Association*, 2011, **106**: 1053–1066.

[34] Kabaila P, Leeb H. On the large-sample minimal coverage probability of confidence intervals after model selection. *Journal of the American Statistical Association*, 2006, **101**: 619–629.

[35] Serfling R J. Approximation Theorems of Mathematical Statistics. New York: John Wiley and Sons, 1980.

FREQUENTIST MODEL AVERAGE ESTIMATION FOR LINEAR ERRORS-IN-VARIABLES MODELS

WANG Haiying

(*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190;
Department of Statistics, University of Missouri, Columbia, Missouri 65211, *USA*)

ZOU Guohua

(*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190)

Abstract Frequentist model average estimation receives much attention in recent years. However, no investigation has been conducted for the data with measurement errors. In this paper, we consider frequentist model average estimation for linear errors-in-variables models. The asymptotic distribution of the model average estimator is derived, and a confidence interval having a coverage probability that tends toward the nominal level in large samples is constructed. Further, the confidence interval constructed based on the model average estimator is shown to be asymptotically the same as that obtained under the full model. A simulation study shows that the finite sample performance of the model average estimator is better than that of the model selection approach or of the full model approach.

Key words Asymptotic distribution, measurement errors, model averaging, model selection.

附录 A 引理和定理的证明

引理 2.2 的证明 首先注意若式 (6) 成立，则结合引理 2.1 的结论并利用连续映射定理，立即可知式 (5) 成立，因此下面我们只证明式 (6).

注意到

$$\sum_{i=1}^{n} W_iW_i^{\rm T} = \sum_{i=1}^{n} X_iX_i^{\rm T} + \sum_{i=1}^{n} X_iU_i^{\rm T} + \sum_{i=1}^{n} U_iX_i^{\rm T} + \sum_{i=1}^{n} U_iU_i^{\rm T}.$$

又因为

$$0 \le \frac{\sum\limits_{i=1}^{n} \boldsymbol{E}\|X_iU_i^{\rm T}\|^2}{n^2} \le \frac{\boldsymbol{E}\|U_j\|^2 \sup_j \|X_j\|^2}{n} \to 0,$$

所以由 Chebyshew 弱大数定律 (例如见文 [35] 定理 1.8.C) 知

$$\frac{1}{n}\sum_{i=1}^{n} W_iW_i^{\rm T} \xrightarrow{p} B + \Sigma_u,$$

于是根据 $\widehat{\theta}$ 的定义 (见式 (2)) 可以得到

$$\begin{aligned}\sqrt{n}(\widehat{\theta}-\theta) &= \Big(\frac{1}{n}\sum_{i=1}^{n} W_iW_i^{\rm T} - \Sigma_u\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(W_iY_i - W_iW_i^{\rm T}\theta + \Sigma_u\theta\big)\\ &= \Big(\frac{1}{n}\sum_{i=1}^{n} W_iW_i^{\rm T} - \Sigma_u\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(X_i\varepsilon_i - X_iU_i^{\rm T}\theta + U_i\varepsilon_i - (U_iU_i^{\rm T}\theta - \Sigma_u\theta)\big)\\ &= \Big(\frac{1}{n}\sum_{i=1}^{n} W_iW_i^{\rm T} - \Sigma_u\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_i + o_P(1),\end{aligned}$$

其中 $\eta_i = X_i\varepsilon_i - X_iU_i^{\rm T}\theta_0 + U_i\varepsilon_i - (U_iU_i^{\rm T}\theta_0 - \Sigma_u\theta_0)$.

$\sum\limits_{i=1}^{n}\eta_i$ 为独立和的形式，故可利用多维中心极限定理 (例如见文 [35]1.9.2 定理 B) 得到其渐近分布．下面我们来验证定理成立的条件．首先

$$\begin{aligned}&\boldsymbol{E}\eta_i = 0,\\ &{\rm Var}(\eta_i) = \sigma^2X_iX_i^{\rm T} + \sigma^2\Sigma_u + X_iX_i^{\rm T}\theta_0^{\rm T}\Sigma_u\theta_0 + \boldsymbol{E}(UU^{\rm T}\theta_0 - \Sigma_u\theta_0)^{\otimes 2},\end{aligned}$$

所以

$$\frac{1}{n}{\rm Var}\left(\sum_{i=1}^{n}\eta_i\right) \to \sigma^2B + \sigma^2\Sigma_u + B\theta_0^{\rm T}\Sigma_u\theta_0 + \boldsymbol{E}(UU^{\rm T}\theta_0 - \Sigma_u\theta_0)^{\otimes 2},$$

又对任意的 ε, 有

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{E}_{\eta_i}\Big[\|\eta_i\|^2 I_{\{\|\eta_i\|>\sqrt{n}\varepsilon\}}\Big]\\ \le&\frac{1}{n}\frac{1}{(\sqrt{n}\varepsilon)^t}\sum_{i=1}^{n}\boldsymbol{E}_{\eta_i}\Big[\|\eta_i\|^{2+t} I_{\{\|\eta_i\|>\sqrt{n}\varepsilon\}}\Big] \qquad (14)\\ \le&\frac{1}{n}\frac{1}{(\sqrt{n}\varepsilon)^t}\sum_{i=1}^{n}\boldsymbol{E}_{\eta_i}\|\eta_i\|^{2+t},\end{aligned}$$

而利用 C_r 不等式可知

$$
\begin{aligned}
\boldsymbol{E}_{\eta_i}\|\eta_i\|^{2+t} =&\boldsymbol{E}_{\varepsilon_i}\boldsymbol{E}_{U_i}\|X_i\varepsilon_i - X_iU_i^{\mathrm{T}}\theta_0 + U_i\varepsilon_i - U_iU_i^{\mathrm{T}}\theta_0 + \Sigma_u\theta_0\|^{2+t}\\
\leq& C_r\big[\boldsymbol{E}\|X_i\varepsilon_i\|^{2+t} + \boldsymbol{E}\|X_iU_i^{\mathrm{T}}\theta_0\|^{2+t} + \boldsymbol{E}\|U_i\varepsilon_i\|^{2+t} + \boldsymbol{E}\|U_iU_i^{\mathrm{T}}\theta_0\|^{2+t} + \|\Sigma_u\theta_0\|^{2+t}\big].
\end{aligned}
$$

根据所给的条件，上式一致有界，所以式 (14) 收敛到 0. 因此多维中心极限定理的条件得到满足. 再利用 Slutsky 定理即可知式 (6) 成立.

定理 2.1 的证明　由 Taylor 展开知

$$
\begin{aligned}
\mu_{\mathrm{true}} &= \mu\left(\beta, \frac{\delta}{\sqrt{n}}\right) = \mu(\beta, 0) + \mu_\gamma^{\mathrm{T}}\frac{\delta}{\sqrt{n}} + o\Big(\frac{1}{\sqrt{n}}\Big),\\
\widehat{\mu}_S &= \mu(\widehat{\beta}_S, \widehat{\gamma}_S) = \mu(\beta, 0) + \begin{pmatrix}\mu_\beta\\ V_S^{\mathrm{T}}\mu_\gamma\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\widehat{\beta}_S - \beta\\ \widehat{\gamma}_S\end{pmatrix} + o_P\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned}
$$

因此，

$$
\begin{aligned}
\widehat{\mu}_S - \mu_{\mathrm{true}} &= \begin{pmatrix}\mu_\beta\\ V_S^{\mathrm{T}}\mu_\gamma\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\widehat{\beta}_S - \beta\\ \widehat{\gamma}_S\end{pmatrix} - \mu_\gamma^{\mathrm{T}}\frac{\delta}{\sqrt{n}} + o_P\Big(\frac{1}{\sqrt{n}}\Big)\\
&= \begin{pmatrix}\mu_\beta\\ V_S^{\mathrm{T}}\mu_\gamma\end{pmatrix}^{\mathrm{T}}G_{nS}\begin{pmatrix}\widehat{\beta} - \beta\\ \widehat{\gamma}\end{pmatrix} - \mu_\gamma^{\mathrm{T}}\frac{\delta}{\sqrt{n}} + o_P\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned}
$$

注意到 $M_{n1} = \sqrt{n}(\widehat{\beta} - \beta)$, $M_{n2} = \sqrt{n}\widehat{\gamma}$, 并将 G_{nS} 的表达式代入可以得到

$$
\sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true}}) = \mu_\beta^{\mathrm{T}}[M_{n1} + B_{11}^{-1}B_{12}(M_{n2} - \delta)] + \omega^{\mathrm{T}}\Big(\delta - A_n^{-\frac{1}{2}}H_{nS}A_n^{\frac{1}{2}}M_{n2}\Big) + o_P(1),
$$

故由连续映射定理以及 Slutsky 定理可知上式依分布收敛到

$$
\Lambda_S = \mu_\beta^{\mathrm{T}}[M_1 + B_{11}^{-1}B_{12}(M_2 - \delta)] + \omega^{\mathrm{T}}\Big(\delta - A^{-\frac{1}{2}}H_SA^{\frac{1}{2}}M_2\Big).
$$

定理 2.2 的证明　根据 $\widehat{\mu}$ 的定义有

$$
\sqrt{n}(\widehat{\mu} - \mu_{\mathrm{true}}) = \sum_S c(S|\widehat{\delta})\sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true}}).
$$

由定理 2.1 的证明过程可知，上式右端中 $\sqrt{n}(\widehat{\mu}_S - \mu_{\mathrm{true}})$ 可以写成 M_{n1} 和 M_{n2} 的线性函数的形式，而 $c(S|\cdot)$ 又是连续函数，所以上式左端实际上是 M_{n1} 和 M_{n2} 的连续函数，故由连续映射定理以及 Slutsky 定理知上式依分布收敛到 $\sum\limits_S c(S|M_2)\Lambda_S$, 经过化简即得结论. 若协变量没有测量误差，为证明 $M_1 + B_{11}^{-1}B_{12}(M_2 - \delta)$ 和 M_2 的独立性，我们只需证明二者的协方差矩阵为 0. 为此首先注意若协变量没有测量误差，则式 (6) 的渐近协方差矩阵中的 F 将简化为 σ^2B, 于是

$$\begin{aligned}
&\operatorname{Var}\begin{pmatrix} M_1 + B_{11}^{-1}B_{12}(M_2-\delta) \\ M_2 \end{pmatrix} \\
=&\operatorname{Var}\left(\begin{pmatrix} I_p & B_{11}^{-1}B_{12} \\ 0 & I_q \end{pmatrix}\begin{pmatrix} M_1 \\ M_2 \end{pmatrix}\right) \\
=&\sigma^2\begin{pmatrix} I_p & B_{11}^{-1}B_{12} \\ 0 & I_q \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1}\begin{pmatrix} I_p & 0 \\ B_{21}B_{11}^{-1} & I_q \end{pmatrix} \\
=&\sigma^2\begin{pmatrix} B_{11}^{-1} & 0 \\ 0 & (B_{22}-B_{21}B_{11}^{-1}B_{12})^{-1} \end{pmatrix}.
\end{aligned}$$

置信区间 (9) 覆盖率的证明 由式 (9) 知, $\Pr\big\{\mu_{\rm true}\in(\mathrm{low}_n,\mathrm{up}_n)\big\}=\Pr\{-z\le T_n\le z\}$, 其中

$$T_n=\frac{\sqrt{n}(\widehat{\mu}-\mu_{\rm true})-\widehat{\omega}^{\rm T}[M_{n2}-Q_n(M_{n2})M_{n2}]}{\widehat{\kappa}}.$$

由定理 2.2 的证明知, $\sqrt{n}(\widehat{\mu}-\mu_{\rm true})$ 是 M_{n1} 和 M_{n2} 的连续函数, 所以由连续映射定理以及 Slutsky 定理知

$$\big(\sqrt{n}(\widehat{\mu}-\mu_{\rm true}),M_{n2}\big)\xrightarrow{d}\big(\Lambda_0+\omega^{\rm T}[\delta-Q(M_2)M_2],M_2\big),$$

其中 $\Lambda_0=\mu_\beta^{\rm T}[M_1+B_{11}^{-1}B_{12}(M_2-\delta)]$. 因此

$$T_n\xrightarrow{d}\frac{\Lambda_0+\omega^{\rm T}(\delta-M_2)}{\kappa}=\frac{\mu_\beta^{\rm T}M_1+\mu_\gamma^{\rm T}(M_2-\delta)}{\kappa}.$$

上式右端服从标准正态分布, 所以此置信区间覆盖真实参数的概率趋近于所指定的覆盖水平.