arXiv:2304.07331v1 [econ.EM] 14 Apr 2023

广义自动最小二乘：从设定错误的异方差模型中提升效率

Bulat Gafarov†

2023年4月18日

摘要

众所周知，在存在异方差的情况下，普通最小二乘估计量并非有效。我提出一种广义自动最小二乘估计量（GALS），无需预检验，即可依据一个（可能）设定错误的模型对异方差进行部分修正。该估计量保证至少与普通最小二乘或加权最小二乘一样有效；如果设定错误的模型近似正确，还能比普通最小二乘实现一定的渐近效率提升。如果异方差模型正确，所提出的估计量可达到完全的渐近效率。其思路是将基于设定错误的异方差模型构造的普通最小二乘和加权最小二乘对应的矩条件，合并为一个联合广义矩估计问题。由此得到的最优广义矩估计量等价于采用估计权重矩阵的可行广义最小二乘法。我还提出一种适用于广义自动最小二乘的最优广义矩估计方差协方差估计量，以处理残差中任何残留的异方差。

JEL分类： C30。

关键词： 线性回归、异方差、有效估计。

1 引言

众所周知，在存在异方差的情况下，普通最小二乘估计量并非有效（参见Wooldridge (2019)第8.4节）；早期的研究见Cragg (1983)。我提出一种广义自动最小二乘估计量（GALS），它基于一个（可能）设定错误的模型对异方差进行部分校正。该估计量保证至少与普通最小二乘或加权最小二乘具有同等的渐近效率；若设定错误的模型近似正确，还能带来一定的效率提升。如果异方差模型设定正确，所提出的估计量可实现完全的渐近效率。其思路是将基于设定错误的异方差模型构造的普通最小二乘和加权最小二乘对应的矩条件，作为一个联合广义矩估计法问题来处理。由此得到的最优广义矩估计量等价于使用估计权重矩阵的广义最小二乘。我还提出了异方差稳健的广义矩估计方差—协方差公式，以处理残差中任何残留的异方差。

该方法分为三个步骤。首先，估计常规的普通最小二乘残差。然后，估计异方差的近似模型。第三，使用闭式公式计算广义自适应最小二乘估计量及其协方差矩阵。这三个步骤都不涉及任何优化，计算起来非常简单。

†加利福尼亚大学戴维斯分校农业与资源经济学系。电子邮件：bgafarov@ucdavis.edu。感谢 Colin Cameron、Takuya Ura 和 Kaspar Wüthrich 提供的宝贵意见。

在一篇密切相关的近期论文中，Romano 和 Wolf (2017年证明，如果采用White (1980年)标准误进行推断，那么使用估计得到的（可能设定错误的）异方差模型的加权最小二乘具有异方差稳健性。此外，他们建议先进行预检，以决定使用普通最小二乘还是加权最小二乘。Romano 和 Wolf (2017年)承认，对于某些设定错误的异方差模型，加权最小二乘的效率低于普通最小二乘；在这种情况下，他们提出的自适应最小二乘（自适应最小二乘）程序的效率也会低于普通最小二乘。我的建议则是不进行预检，而是使用广义自适应最小二乘估计量；其效率至少不低于普通最小二乘或加权最小二乘，甚至可能高于这两种估计量。

其他研究者也考察了在高阶矩上施加额外限制所带来的广义矩估计相对于普通最小二乘的效率提升（Im 和 Schmidt，2008年）。相比之下，除了标准的外生性条件以及存在近似非平凡异方差模型这一假设外，我不对回归残差作任何额外假设。（真实模型可能是同方差的，在这种情况下，广义渐近最小二乘与普通最小二乘渐近等价；加权最小二乘的矩在渐近意义下会被忽略。）

2 模型与方法

考虑一个线性回归模型

$$Y_i = X_i'\beta + \varepsilon_i. \tag{1}$$

照例，$E(\varepsilon_i|X_i) = 0$，回归变量向量 X_i 的维数为 p。假设 $E(\varepsilon_i^2|X_i) = \sigma_\star^2(X_i) = \exp(\sum_{k=1}^{\infty} \delta_k p_k(X_i))$，其中 $p_k(\cdot)$是逼近基函数，例如切比雪夫多项式。只要异方差函数平方可积，我们就可以选择三角多项式或其他定义在 X_i支撑集上的正交基函数。

假设我们选择一个灵活的异方差模型，其形式为 $\sigma^2(X_i)$（例如，$\sigma^2(X_i) = \exp(\sum_{k=1}^{K} \delta_k p_k(X_i))$）。众所周知，WLS 程序对应的矩条件为

$$E(\varepsilon_i \frac{X_i}{\sigma^2(X_i)}) = 0. \tag{2}$$

如果函数形式正确，即 $\sigma^2(X_i) == \sigma_\star^2(X_i)$，那么根据高斯–马尔可夫定理，基于（2）的 WLS 对 β 的估计在渐近意义下是有效率的。当然，OLS 估计量也对应于满足以下矩条件的广义矩估计量

$$E(\varepsilon_i X_i) = 0. \tag{3}$$

根据我们对异方差建模的准确程度，选择（2）可能会提高效率，也可能不会，相较于（3）。然而，如果我们基于（2）和（3）联合进行广义矩估计，并采用最优广义矩估计权重，那么可以保证其效率至少不低于仅使用矩（3）（OLS 估计量）或矩（2）（效率较低的 WLS 估计量）的估计量。因此，如果我们的异方差模型 $\sigma^2(X_i)$ 近似正确，根据最优广义矩估计量的性质，联合广义矩估计将比 OLS 更有效率。

2.1 估计量的描述

考虑一个基于 (2) 和 (3) 的联合GMM估计量。结果表明，它本身就是一个广义最小二乘估计量，能够自动适应残差的异方差（因此得名）。它由以下优化问题给出

$$\hat{\beta}^{GALS} = \underset{b}{\operatorname{argmin}} \frac{1}{n} \begin{pmatrix} \varepsilon' X' \\ \varepsilon' D X' \end{pmatrix} W \begin{pmatrix} X\varepsilon \\ XD\varepsilon \end{pmatrix}, \tag{4}$$

其中， ε 是残差向量， X 是维度为 $(k \times n)$ 的回归变量矩阵， W 是最优权重矩阵（将在下一小节中明确给出）， $D = diag(\sigma^2{}^{-1}(X_1) \cdots, \sigma^2{}^{-1}(X_n))$。事实上，经过化简后，广义最小二乘形式可明确写为，

$$\hat{\beta}^{GALS} = \underset{b}{\operatorname{argmin}} \frac{1}{n} \varepsilon'(X'W_{11}X + X'W_{12}XD + DX'W_{21}X + DX'W_{22}XD)\varepsilon \tag{5}$$

$$= \underset{b}{\operatorname{argmin}} (Y - X'b)'A_n(Y - X'b) = (XA_nX')^{-1}(XA_nY) \tag{6}$$

其中 $A_n = \frac{1}{n}(X'W_{11}X + X'W_{12}XD + DX'W_{21}X + DX'W_{22}XD)$ 是一个对称 $n \times n$ 矩阵。

定理 1。 假设数据是独立同分布的，矩阵 $G = (EX_iX_i', E\frac{1}{\sigma^2(X_i)}X_iX_i')'$ 满秩，矩条件 (2)-(3) 具有非退化的协方差矩阵 Ω。那么 $\sqrt{n}(\hat{\beta}^{GALS} - \beta) \xrightarrow{d} N(0, (G'WG)^{-1})$。此外，$\hat{\beta}^{GALS}$ 的效率弱高于 $\hat{\beta}^{OLS}$ 和 $\hat{\beta}^{WLS}$（即相应渐近协方差矩阵之差为正定）。

证明。 由采用最优权重的广义矩估计的效率性质可得 W （参见 Hall(2004) 中的定理 3.4）。 □

当然，在实际应用中， A_n 必须替换为相合估计量，下一小节将对此进行讨论。

2.2 矩阵 A_n 的计算

通常，最优广义矩估计权重矩阵 W 等于矩条件协方差矩阵 Ω 的逆矩阵，其形式如下:

$$\Omega = \begin{pmatrix} E\varepsilon_i^2 X_i X_i' & E\frac{1}{\sigma^2(X_i)}\varepsilon_i^2 X_i X_i' \\ E\frac{1}{\sigma^2(X_i)}\varepsilon_i^2 X_i X_i' & E\frac{1}{\sigma^4(X_i)}\varepsilon_i^2 X_i X_i' \end{pmatrix}. \tag{7}$$

假设矩阵 $\Omega_{11} = E\varepsilon_i^2 X_i X_i'$ 和 $(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})$ 均可逆。根据Lu 和 Shiou (2002中的定理 2.1，我们可以得到其逆矩阵的显式分块表示:

$$W = \Omega^{-1} = \begin{pmatrix} \Omega_{11}^{-1} + \Omega_{11}^{-1}\Omega_{12}(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1}\Omega_{12}\Omega_{11}^{-1} & -\Omega_{11}^{-1}\Omega_{12}(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1} \\ -(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1}\Omega_{12}\Omega_{11}^{-1} & (\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1} \end{pmatrix} \tag{8}$$

因此，矩阵 A_n 的显式形式为

$$A_n = \frac{1}{n}\big(X'(\Omega_{11}^{-1} + \Omega_{11}^{-1}\Omega_{12}(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1}\Omega_{12}\Omega_{11}^{-1})X \tag{9}$$

$$- X'(\Omega_{11}^{-1}\Omega_{12}(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1})XD \tag{10}$$

$$- DX'(\Omega_{11}^{-1}\Omega_{12}(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1})X \tag{11}$$

$$+ DX'(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{12})^{-1}XD\big) \tag{12}$$

该矩阵的可行版本使用以下分块构造

$$\hat{\Omega} = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n \hat{\varepsilon}_i^2 X_i X_i' & \sum_{i=1}^n \frac{1}{\hat{\sigma}^2(X_i)}\hat{\varepsilon}_i^2 X_i X_i' \\ \sum_{i=1}^n \frac{1}{\hat{\sigma}^2(X_i)}\hat{\varepsilon}_i^2 X_i X_i' & \sum_{i=1}^n \frac{1}{\hat{\sigma}^4(X_i)}\hat{\varepsilon}_i^2 X_i X_i' \end{pmatrix} \tag{13}$$

而不是 Ω 和 $\hat{}$ $D = diag(\frac{1}{\hat{\sigma}^2(X_1)} \cdots, \frac{1}{\hat{\sigma}^2(X_n)})$，而不是 D。这里 $\hat{}\sigma^2(X_i) = \exp(\sum_{k=1}^K \hat{}\delta_k p_k(X_i))$，其中 $\hat{}$ δ_k 由对 $\log \hat{}\varepsilon_i^2$ 关于 $(p_1(X_i) \ldots, p_k(X_i))$ 的普通最小二乘回归得到。残差 $\hat{}\varepsilon_i$ 来自对 Y_i 关于 X_i 的普通最小二乘回归。在确保相合性 δ_k 以及 Ω 的 $\hat{}$ $\hat{}$ 的正则矩条件下，可以在定理 1 中用 A_n 替换 A_n。

参考文献

Cragg, J. G.（1983）：“未知形式异方差情况下的更有效估计”，《计量经济学》：计量经济学会期刊，751–763。Hall, A. R.（2004）：广义矩估计法，牛津大学出版社。Im, K. S. 和 P. Schmidt（2008）：“当高阶矩不依赖于回归变量时，在非正态性情况下使用残差增广最小二乘法进行更有效估计”，《计量经济学杂志》，144，219–233。Lu, T.-T. 和 S.-H. Shiou（2002）：“2×2分块矩阵的逆”，《计算机与应用数学》，43，119–129。Romano, J. P. 和 M. Wolf（2017）：“重振加权最小二乘法”，《计量经济学杂志》，197，1–19。White, H.（1980年）：“异方差稳健协方差矩阵估计量及异方差直接检验”，《计量经济学》：计量经济学会期刊，817–838。Wooldridge, J. M.（2019）：《计量经济学导论：现代方法》，圣智学习。