arXiv:2603.03008v1 [econ.EM] 3 Mar 2026

聚焦加权平均最小二乘估计量

Shou-Yung Yin[1*]

[1]国立台北大学经济系，台湾新北市。

通讯作者。电子邮箱：syyin@mail.ntpu.edu.tw；

摘要

我们提出一种聚焦加权平均最小二乘（FWALS）估计量，以解决聚焦模型平均的计算负担。通过对辅助回归变量进行半正交化，加权问题从 2^{k_2}个子模型缩减为至多 k_2 个逐回归变量权重，从而得到一种易于处理的次优程序。在局部趋零条件下，我们推导了适用于平滑聚焦函数的 FWALS 极限分布，并提出一种用于数据驱动权重选择的插入式渐近均方误差准则。模拟结果表明，FWALS 的表现与聚焦信息准则（FIC）基准非常接近；当焦点函数设定为脉冲响应函数时，其表现稳定。基于先验的 WALS 在某些情形下具有竞争力，但其表现取决于信号机制和聚焦参数的设计。总体而言，FWALS 提供了一种实用且稳健的替代方案，并能显著降低计算成本。

关键词：模型　　平均，聚焦信息准则，正交 tra　　变换

1 引言

模型不确定性长期以来一直被视为计量经济学和应用统计学中的核心挑战。在实证研究中，研究者常常面临一个难题：如何从大量候选模型中进行选择，而每个模型都可能对数据作出合理描述。传统的模型选择方法，例如基于信息准则的方法，旨在选出单一的“最佳”模型。然而，已有大量研究反复指出，依赖单一选定模型可能会忽略其他合理设定中的相关信息，并导致推断产生偏差，尤其是在候选模型质量相近时。为此，大量文献围绕模型平均（MA）方法展开，这些方法整合多个模型中的信息，以获得更稳健的参数估计值和预测表现；例如，Steel (2020) 提供了全面的综述。

在模型平均（MA）文献中，一种常见做法是针对无约束模型的整个参数向量定义平均估计量，使其最小化一个可测的渐近风险函数；该风险是相对于候选模型的估计量来评估的。具有代表性的研究包括 Hansen (2007)、Hansen and Racine (2012)、Zhang and Liu (2023)、Zhang and Zhang (2023) 以及 Chen 等（2025）等。尽管已有研究表明，这些方法在整体模型拟合方面效果良好，但实证研究者往往主要关注某个特定的结构参数子集或这些参数的函数，而非与控制变量相关的讨厌参数。这一观察自然推动了聚焦模型平均方法的发展：该方法选择平均权重，以最小化对聚焦参数向量进行估计的渐近风险，而不是针对整个参数空间。

聚焦信息准则（FIC）最初由 Claeskens and Hjort (2003) 在似然框架下提出，为这一研究方向奠定了开创性基础。后续研究将 FIC 思想拓展至各种框架，推动了聚焦模型平均领域丰富文献的发展；例如，Hjort and Claeskens (2006)、Claeskens and Hjort (2008)、Zhang et al. (2012)、Liu (2015)、Lu (2015)、DiTraglia (2016)、Kitagawa and Muris (2016)、Lohmeyer et al. (2019) 以及 Yin et al. (2021)。近年来，Zhang and Liu (2024) 对渐近偏差和方差进行了统一表述，为适用于多种情形的通用框架提供了基础。该领域反复关注的一个主题，是如何准确估计渐近偏差和方差成分，因为它们既是模型选择的基础，也是确定平均权重的依据。尽管取得了重要

进展，但在实践中实施这些方法可能需要较高的计算成本，尤其是在辅助回归变量数量较多时，子模型数量呈指数增长会使权重的计算在数值上变得不稳定。

传统模型平均所带来的计算负担不容忽视。对于 k_2 个辅助回归变量，可能的子模型数量为 2^{k_2}；为了评估风险函数或计算平均权重，必须估计所有候选子模型。即使 k_2 适中，这一要求也会很快变得难以承受；此外，为了实施聚焦模型平均程序，还需要计算每个子模型的偏差和方差项，这进一步加剧了这一挑战。在实践中，研究人员可能会面临这样的情况：评估完整模型空间根本不可行，而这一局限性促使人们寻求计算效率更高的替代方案。

本文建立在 Magnus 等人（2010）和 DeLuca 等人（2018）提出的正交化思想之上；他们建议对辅助回归变量进行变换，以减少相关性，从而简化模型平均估计量的结构。通过采用正交化辅助回归变量，我们能够将聚焦模型平均估计量改写为一种仅依赖于与辅助回归变量相关的少量权重的形式，而不是依赖于定义在完整模型空间上的权重。这一变换大幅降低了计算成本。我们不再需要评估数量呈指数增长的子模型，而只需处理至多 k_2 个项；所得权重则可以解释为原始模型权重的部分和。尽管这一构造无法得到完整子模型平均所隐含的精确最优解，但它提供了一种具有计算吸引力的次优解，同时仍与聚焦模型平均的理念高度一致。因此，当传统方法因计算限制而无法实施时，所提出的方法有助于在实际情境中开展经验研究。

我们的方法在若干重要方面不同于近期也关注这一问题的两类研究。首先，Charkhi 等人（2016）在似然模型中考察了一种最小均方误差模型平均估计量，其平均权重由单变量方程推导得出，且仅受权重之和为一的约束，并未采用模型平均文献中通常施加的非负约束。相比之下，我们的方法要求所有权重均位于 $[0,1],$ ，从而确保其可解释为凸组合，并避免负权重可能导致的不稳定性。其次，Zhu等人（2023）提出了一种可扩展的频率学派模型平均方法，采用奇异值分解缩减模型空间，并证明了可扩展平均估计量与传统平均估计量的最小损失渐近等价。他们主要致力于实现统计效率和计算效率，而不依赖局部趋零

假设。我们的贡献有所不同：我们以聚焦参数框架为目标，重点关注结构参数及其变换，并证明基于正交化的加权方案可作为传统基于FIC的平均方法的一种实用且理论上连贯的次优替代方案。

此外，我们的方法不同于 Magnus 等人（2010）和 Luca 等人（2022, 2025）提出的基于先验的方法。基于先验的方法在正态位置框架中采用由选定先验所诱导的后验均值收缩，这意味着系数会获得收缩权重。相比之下，我们的方法通过最小化针对聚焦参数构造的代入式渐近均方误差（AMSE）来确定权重，因此，所选权重旨在优化聚焦目标估计量的偏差与方差权衡，而非优化通用风险准则。

我们的模拟研究为所提出的方法提供了进一步支持。在基准设计中，所提出的方法、Liu (2015) 的 FIC，以及 Charkhi 等人（2016）的最小均方误差模型平均估计量表现相当，说明这种变换不会损害风险表现。然而，在考察脉冲响应函数（IRF）的预测期时，出现了显著差异。我们的方法与 FIC 表现仍然十分接近，并且在不同预测期下风险表现稳定；相比之下，Charkhi 等人（2016）提出的方法，其表现会随所选 IRF 预测期而出现相对较大的波动。这些发现证实，所提出的方法表现稳定且计算效率高，为传统聚焦平均估计量提供了有力的替代方案。

综上所述，本文的贡献有两点。首先，我们通过引入一种基于正交化辅助回归变量的加权平均最小二乘估计量，扩展了聚焦模型平均框架，为聚焦参数估计提供了一种计算上易于实现的解决方案。其次，我们阐明了本文方法与文献中现有方法之间的关系，强调了使用基于渐近均方误差（AMSE）所得权重的优势，以及聚焦于结构参数的作用。

本文其余部分安排如下。第 2 节介绍模型设定，并对平均估计量进行形式化定义。第 3 节介绍所提出的次优平均估计量。第 4 节讨论其性质并研究其极限行为。第 5 节通过模拟结果展示所提方法在计算和统计方面的优势。第 6 节总结全文。

2 模型设定

本节介绍后续讨论所需的模型设定。我们遵循 Magnus 等人（2010）和 Liu (2015), 并考虑如下回归框架:

$$y_i = \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta} + \epsilon_i = \mathbf{x}_{1i}^{\mathsf{T}}\boldsymbol{\beta}_1 + \mathbf{x}_{2i}^{\mathsf{T}}\boldsymbol{\beta}_2 + \epsilon_i, \quad i = 1, ..., N, \tag{1}$$

其中， y_i 表示我们关注的目标变量，$\mathbf{x}_i = [\mathbf{x}_{1i}^{\mathsf{T}}\ \mathbf{x}_{2i}^{\mathsf{T}}]^{\mathsf{T}}$，以及 $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^{\mathsf{T}}\ \boldsymbol{\beta}_2^{\mathsf{T}}]^{\mathsf{T}}$。$\mathbf{x}_{1i}$ 和 $\mathbf{x}_{2i}$ 分别表示核心回归变量和辅助回归变量； ϵ_i为随机误差， N 表示样本量。$\mathbf{x}_{1i}$ 和 $\mathbf{x}_{2i}$的维度分别为 k_1 和 k_2 ，对应的斜率系数为 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$，并且我们进一步定义 $k = k_1 + k_2$ ，且 k 为有限值。在相关文献中，这种模型设定允许研究者在进行统计推断时，通过考虑辅助回归变量的不同组合，在所有可能的子模型中保留核心回归变量。

为便于解释所有子模型的估计结果，我们首先将方程(1)改写为矩阵表示法:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{2}$$

其中，$\mathbf{y}= [y_1\ \dots\ y_N]^{\mathsf{T}}$，$\mathbf{X} = [\mathbf{X}_1\mathbf{X}_2], \mathbf{X}_1 = [\mathbf{x}_{11}\ \dots\mathbf{x}_{1N}]^{\mathsf{T}}$，$\mathbf{X}_2 = [\mathbf{x}_{21}\ \dots\mathbf{x}_{2N}]^{\mathsf{T}}$以及 $\boldsymbol{\epsilon} = [\epsilon_1\ \dots\ \epsilon_N]^{\mathsf{T}}$。此外，我们定义一个 $k_{2m} \times k_2$选择矩阵 $\boldsymbol{\Pi}_m$，用于为子模型 m 选择辅助回归变量。例如，纳入的辅助回归变量可表示为$\mathbf{X}_{2m} =\mathbf{X}_2\boldsymbol{\Pi}_m^{\mathsf{T}}$。令

$$\mathbf{S}_m = \begin{bmatrix} \mathbf{I}_{k_1} & \mathbf{0}_{k_1 \times k_{2m}} \\ \mathbf{0}_{k_2 \times k_1} & \boldsymbol{\Pi}_m^{\mathsf{T}} \end{bmatrix}. \tag{3}$$

那么，我们可以定义子模型 m 的 $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathsf{T}}, \boldsymbol{\beta}_2^{\mathsf{T}})^{\mathsf{T}}$ 的最小二乘估计量:

$$\hat{\boldsymbol{\beta}}_m = \mathbf{S}_m \left(\mathbf{S}_m^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{S}_m\right)^{-1} \mathbf{S}_m^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{y}. \tag{4}$$

上述估计量的维度为 $k \times 1$。

遵循Liu (2015)，我们通过引入光滑实值函数 $\mu(\boldsymbol{\beta}_1)$来定义关注参数。相应地，对于任意子模型，我们可以得到该函数的估计值$\hat{}\mu_m = \mu(\boldsymbol{\beta}_{1m})$。基于这些估计值，Liu (2015)所提出的聚焦参数的平均估计量如下

$$\hat{\mu}(\mathbf{w}) = \sum_{m=1}^{M} w_m \hat{\mu}_m, \tag{5}$$

其中权重向量$\mathbf{w} = [w_1 \ \ldots \ w_M]^\mathsf{T}$满足如下条件：

$$\mathcal{H} = \left\{ \mathbf{w} \in [0,1]^M : \sum_{m=1}^{M} w_m = 1 \right\}, \tag{6}$$

其中， M 表示子模型总数。为了得到平均估计量，我们需要估计 M 个子模型。在上述情形中， $M = 2^{k_2}$。即使k_2 适中，这个数也会呈指数级增长，导致计算负担显著增加。下一节将介绍一种替代方法，它在保留平均程序优势的同时，大幅缩短计算时间；这一点已在我们的模拟研究中得到证实。

3 次优平均估计量

3.1 聚焦加权平均最小二乘估计量

本节提出一种基于正交化辅助回归变量的聚焦参数平均估计量。Magnus 等人（2010）和 De Luca 等人（2018）提出了正交化辅助回归变量的概念，该方法将新的辅助回归变量定义为

$$\mathbf{X}_2^* = \mathbf{X}_2 \hat{\mathbf{\Lambda}} \hat{\mathbf{P}}^{-1/2}, \tag{7}$$

其中，$\hat{\mathbf{\Lambda}} = \mathrm{Diag}\left(\mathrm{Diag}\left(\underline{\mathbf{X}_2^\mathsf{T} \mathbf{M}_{N}^{1}{}^{2}}\right)\right)^{-1/2}\mathbf{x}$，$\hat{\mathbf{P}} = \underline{\hat{\mathbf{\Lambda}} \mathbf{X}_2^\mathsf{T} \mathbf{M}_{N}^{1}{}^{2}} \mathbf{x} \hat{\mathbf{\Lambda}}$ 以及$\mathbf{M}_1 = \mathbf{I}_N - \mathbf{X}_1(\mathbf{X}_1^\mathsf{T}\mathbf{X}_1)^{-1}\mathbf{X}_1^\mathsf{T}$。给定这种半正交化，按照 Magnus 等人（2010）的做法，我们可以定义 $\boldsymbol{\beta}_1$ 的加权平均最小二乘估计量（WALS）：

$$\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}} = \sum_{m=1}^{M} w_m \hat{\boldsymbol{\beta}}_{1m}, \tag{8}$$

子模型估计量如下：

$$\hat{\boldsymbol{\beta}}_{1m} = \hat{\boldsymbol{\beta}}_{1\mathrm{narrow}} - \hat{\mathbf{\Xi}} \hat{\mathbf{\Lambda}} \hat{\mathbf{P}}^{-1/2} (\mathbf{\Pi}_m^\mathsf{T} \mathbf{\Pi}_m) \hat{\boldsymbol{\beta}}_2, \tag{9}$$

其中 $\hat{\mathbf{\Xi}} = (\mathbf{X}_1^\mathsf{T}\mathbf{X}_1)^{-1}\mathbf{X}_1^\mathsf{T}\mathbf{X}_2$, $\hat{\boldsymbol{\beta}}_{1\mathrm{narrow}} = (\mathbf{X}_1^\mathsf{T}\mathbf{X}_1)^{-1}\mathbf{X}_1^\mathsf{T}\mathbf{y}$ 且 $\hat{\boldsymbol{\beta}}_2 = \mathbf{x}_2^{*\mathsf{T}} \underline{\mathbf{M}_1 \mathbf{y} N_{\circ}}$

上述结果表明，方程（8）中定义的 $\boldsymbol{\beta}_1$ WALS 可以改写为：

$$\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}} = \hat{\boldsymbol{\beta}}_{1\mathrm{narrow}} - \hat{\mathbf{\Xi}} \hat{\mathbf{\Lambda}} \hat{\mathbf{P}}^{-1/2} \sum_{m=1}^{M} w_m (\mathbf{\Pi}_m^\mathsf{T} \mathbf{\Pi}_m) \hat{\boldsymbol{\beta}}_2$$

$$=\hat{\boldsymbol{\beta}}_{1\text{narrow}} - \hat{\boldsymbol{\Xi}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\hat{\boldsymbol{\beta}}_2 \tag{10}$$

其中 $\tilde{\mathbf{W}} = \sum_{m=1}^{M} w_m(\boldsymbol{\Pi}_m^{\mathsf{T}}\boldsymbol{\Pi}_m)$，且 $\tilde{\mathbf{W}}$ 是一个对角矩阵，这是正交化辅助回归变量的性质所致。因此，我们还可以进一步得到以下结果：

$$\tilde{\mathbf{W}}\hat{\boldsymbol{\beta}}_2 = \left[\tilde{w}_1\hat{\beta}_{21} \ \cdots \ \tilde{w}_{k_2}\hat{\beta}_{2k_2}\right]^{\mathsf{T}}, \tag{11}$$

其中，对于 $j = 1,\ldots,k_2$， $\tilde{w}_j$s 是$\mathbf{W}$的对角元素。与公式（6）中定义的 w_m 相比，这些对角元素所起的权重作用有所不同，因为 $\tilde{w}_j$ 是取决于选择矩阵 $\boldsymbol{\Pi}_m$ 的 w_ms 的部分和。根据对 w_m 所施加的条件，我们还可以对 $\tilde{w}_j$s 施加一个较弱的条件：

$$\tilde{\mathcal{H}} = \left\{\tilde{\mathbf{w}} \in [0,1]^{k_2}\right\}. \tag{12}$$

正如 Magnus 等人（2010）所讨论的，虽然对于 $\boldsymbol{\beta}_1$ ，加权平均最小二乘可以简化为仅涉及 k_2 $\tilde{w}_j$s 的估计量，但这种方法没有考虑结构权重 w_m，因此，与考虑所有可能子模型的方法相比，它不可能是最优解。然而，由于需要优化的权重数量为 k_2，这种方法可以减少计算时间，尤其是在采用模拟置信区间方法进行统计推断时。

现在，我们可以如下定义聚焦加权平均最小二乘：

$$\hat{\mu}(\tilde{\mathbf{w}}) = \mu(\hat{\boldsymbol{\beta}}_{1\text{WALS}}). \tag{13}$$

公式（5）和（13）的关键区别在于构造平均估计量的顺序。在聚焦加权平均最小二乘中，我们首先构造 $\boldsymbol{\beta}_1$的平均估计量，记为 $\boldsymbol{\beta}_{1\text{WALS}}$，然后直接将聚焦函数应用于 $\boldsymbol{\beta}_{1\text{WALS}}$，以获得 μ 的估计值。相比之下，Liu (2015) 的聚焦平均估计量会在每个子模型中计算聚焦参数，再对各模型的结果取平均。尽管这两种方法的步骤不同，但只要子模型和权重相同，其渐近行为仍然等价。根据德尔塔法，无论在平均之前还是之后应用聚焦函数，都不会改变极限分布。直观而言，这是因为聚焦函数是平滑的，因此围绕真实参数的线性近似意味着，在大样本下，先取平均还是先进行变换无关紧要。因此，可以将聚焦加权平均最小二乘的渐近均方误差视为评估模型并获得权重的另一种方式。然而，所提出方法构造的权重不同于 Liu (2015) 中的权重，因此其渐近性质也有所不同。下一节将给出正式结果。

注释1。 Liu (2015) 中平均估计量的具体构造方式，使其无法以类似方式降低计算负担，即便辅助回归变量经过公式（7）中的半正交化变换也是如此。回顾公式（5），对于给定的子模型 m，聚焦参数的估计量为$\mu_m = \mu(\boldsymbol{\beta}_{1m})$，其中 $\boldsymbol{\beta}_{1m}$ 在公式（9）中定义。对于任意两个不同的子模型 m 和 m'， $\boldsymbol{\beta}_{1m}$ 和 $\boldsymbol{\beta}_{1m'}$ 的结构形式共享相同的基准分量 $\boldsymbol{\beta}_{1\text{narrow}}$。因此，子模型估计量$\mu_m$ 和$\mu_{m'}$ 通过这一共享分量天然相关，尽管辅助回归变量本身彼此正交。由于这些子模型估计量相关，交叉乘积项不会消失。因此，权重优化问题虽可简化，但在 Liu (2015) 所采用的框架中，组合负担仍未解决。然而，值得注意的是，这一计算瓶颈严格来说仅与聚焦参数的评估有关。若目标改为 Zhang and Liu (2019) 所考虑的、始终包含核心回归变量的全局模型拟合，则正交变换可以发挥作用，从而彻底消除组合负担。[1]

4 聚焦WALS的渐近风险

本节将讨论所提出的聚焦加权平均最小二乘在局部趋零框架下的极限性质。在详述渐近框架之前，我们先陈述本文所作的假设。

假设1 （局部趋零条件）。 $\boldsymbol{\beta}_2 = \frac{\boldsymbol{\delta}}{\sqrt{N}}$。

假设 2 （样本矩收敛）。 $\frac{\mathbf{X}^\top \mathbf{X}}{N} \xrightarrow{p} \mathbf{Q} = \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^\top)$，并且$\mathbf{Q}$ 可利用$\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$这一事实分块为 $\begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix}$。

假设 3 （中心极限定理）。 $\frac{\mathbf{X}^\top \boldsymbol{\epsilon}}{\sqrt{N}} \xrightarrow{d} \mathbf{R} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$。

假设1是文献中的标准条件，有助于评估渐近偏差与方差之间的权衡。为方便起见，我们采用假设2和3作为高层次假设。此外，假定$\boldsymbol{\Omega}$ 是

[1]考察整体拟合时，子模型 m 的拟合值可根据 Frisch-Waugh-Lovell 定理分解为 $\hat{\mathrm{y}}_m = \mathrm{X}_1(\mathrm{X}_1^\top \mathrm{X}_1)^{-1}\mathrm{X}_1^\top \mathrm{y} + \sum_{j \in m} \mathrm{M}_1 \mathrm{x}_{2j}^* \hat{\beta}_{2j}^*$，其中 x_{2j}^* 是 X_2^* 的第 j列，而 $\hat{\ } \beta_{2j}^*$ 是 $\hat{\ } \boldsymbol{\beta}_2 = \mathrm{X}^*\mathrm{T}2\ \mathrm{M}1\mathrm{y}_N$ 的第 j个元素。由于正交化保证了对$j \neq l$而言 $\mathrm{x}_{2j}^{*\top}\ \mathrm{M}_1 \mathrm{x}_{2l}^* = 0$ ，风险准则中的所有交叉乘积项自然消失，从而将优化问题从 2^{k_2}种组合缩减为 k_2 种组合。

一般的，以涵盖时间序列数据中具有不同矩条件要求的异方差性和混合过程情形。此外，假设2还意味着公式（7）中的$\hat{\boldsymbol{\lambda}}$和$\hat{\mathbf{P}}$具有定义良好的极限，因此$\hat{\boldsymbol{\lambda}} \xrightarrow{p} \text{Diag}\left(\text{Diag}\left(\mathbf{Q}_{22}-\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\right)\right)^{-1/2} = \boldsymbol{\Lambda}$以及$\hat{\mathbf{P}} \xrightarrow{p} \boldsymbol{\Lambda}\left(\mathbf{Q}_{22}-\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\right)\boldsymbol{\Lambda} = \mathbf{P}$。

4.1 渐近均方误差

在假设 1-3 下，且$\mathbf{W}$为非随机矩阵时，我们有如下结果$N \to \infty$:

定理1. 在假设 1-3 下，且$\mathbf{W}$为非随机矩阵时:

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\beta}}_{1\text{WALS}}-\boldsymbol{\beta}_1) &\xrightarrow{d} \boldsymbol{\Xi}\mathbf{C}\left(\mathbf{I}-\tilde{\mathbf{W}}\right)\mathbf{C}^{-1}\boldsymbol{\delta}+\boldsymbol{\Psi}\mathbf{R} \\ &\equiv \mathbf{R}_{\boldsymbol{\beta}_1}(\tilde{\mathbf{W}}) \\ &\sim \mathrm{N}\left(\boldsymbol{\Xi}\mathbf{C}\left(\mathbf{I}-\tilde{\mathbf{W}}\right)\mathbf{C}^{-1}\boldsymbol{\delta}, \boldsymbol{\Psi}\boldsymbol{\Omega}\boldsymbol{\Psi}^{\mathsf{T}}\right),\end{aligned}$$

and

$$\sqrt{N}\hat{\boldsymbol{\beta}}_2 \xrightarrow{d} \mathbf{C}^{-1}\boldsymbol{\delta}+\mathbf{B}\mathbf{R},$$

其中 $\boldsymbol{\Xi} = p \lim_{N\to\infty} \boldsymbol{\Xi} = p \lim_{N\to\infty} \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12} = \tilde{\mathbf{Q}}_{11}^{-1}\tilde{\mathbf{Q}}_{12}$, $\boldsymbol{\Lambda} = p \lim_{N\to\infty} \hat{\boldsymbol{\Lambda}}^{-1/2} = \text{diag}\left(\tilde{\mathbf{Q}}_{22}-\tilde{\mathbf{Q}}_{21}\tilde{\mathbf{Q}}_{11}^{-1}\tilde{\mathbf{Q}}_{12}\right)^{-1/2}$, $\boldsymbol{\Psi} = p \lim_{N\to\infty} \hat{\mathbf{P}}, \mathbf{B} = \left[-\mathbf{C}\mathbf{C}^{\mathsf{T}}\right]$ 以及 $\boldsymbol{\Psi} = \left[\mathbf{Q}_{11}^{-1}+\boldsymbol{\Xi}\mathbf{C}\,\mathbf{W}\mathbf{C}\mathbf{C}\,\mathbf{W}\mathbf{C}^{\mathsf{T}}\right]$。

上述定理的第一个结果给出了在局部趋零框架下采用加权平均最小二乘时$\boldsymbol{\beta}_{1\text{WALS}}$的渐近偏差和方差。容易看出，当$\mathbf{W} = \mathbf{I}_{k_2}$时，渐近偏差为零，加权平均最小二乘便成为考虑所有辅助回归变量的估计量。若令$\mathbf{W} = \mathbf{0}$，渐近偏差则变为 $\boldsymbol{\Xi}\mathbf{C}\mathbf{C}^{-1}\boldsymbol{\delta} = \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\boldsymbol{\delta}$，而渐近方差为$\mathbf{Q}_{11}^{-1}\boldsymbol{\Omega}_{11}\mathbf{Q}_{11}^{-1}$。这些结果表明，可以通过选择$\mathbf{W}$来权衡偏差与方差。此外，渐近方差继承了加权平均最小二乘的一项优良性质：无需计算任意两个子模型之间的关系，只需关注$\mathbf{X}_1$与$\mathbf{X}_2^*$之间的协方差。另一方面，第二个结果表明，$\sqrt{N}\hat{\boldsymbol{\beta}}_2$是$\mathbf{C}^{-1}\boldsymbol{\delta}$的无偏估计量，这对于恢复$\boldsymbol{\beta}_{1\text{WALS}}$的渐近偏差信息起着重要作用。

—

基于前面的结果，我们可以通过纳入焦点函数对其进行扩展。令$\mathbf{D}_{\boldsymbol{\beta}_1} = \frac{\partial \mu}{\partial \boldsymbol{\beta}_1}$，并假设偏导数在 $\boldsymbol{\beta}_1$的所有实数取值上连续。然后应用德尔塔法，可得以下定理。

定理 2. 在假设 1-3 下，假设 $N \to \infty$，可得

$$\begin{aligned}\sqrt{N}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}}) - \mu(\boldsymbol{\beta}_1)) &\xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}} \boldsymbol{\Xi}\mathbf{C}\left(\mathbf{I} - \tilde{\mathbf{W}}\right)\mathbf{C}^{-1}\boldsymbol{\delta} + \mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}} \boldsymbol{\Psi}\mathbf{R} \\ &\equiv R_{\mu}(\tilde{\mathbf{W}}).\end{aligned}$$

ˆ 该定理表明，聚焦加权平均最小二乘 $\mu(\boldsymbol{\beta}_{1\mathrm{WALS}})$的渐近均方误差为

$$\begin{aligned}\mathrm{AMSE}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}})) =& \mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}} \boldsymbol{\Xi}\boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\beta}_1} + \tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}\mathbf{C}\mathbf{V}\tilde{\mathbf{w}} - 2\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\beta}_1} \\ &+ \mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\boldsymbol{\beta}_1} + \tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{V}\tilde{\mathbf{w}} + 2\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{B}\boldsymbol{\Omega}\mathbf{H}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\boldsymbol{\beta}_1},\end{aligned} \tag{14}$$

其中$\mathbf{V} = \mathrm{Diag}(\mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}}\boldsymbol{\Xi}\mathbf{C}) = \mathrm{Diag}(\mathbf{C}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\beta}_1})$，且$\mathbf{H} = [\mathbf{I}\ \mathbf{0}]^{\mathsf{T}}$；$\mathbf{B}$ 已在定理1中定义。最优$\mathbf{W}$，$\mathbf{W}^o$，可通过在˜$\mathbf{w} \in \mathcal{H}$ 上最小化渐近均方误差得到，其中 $\boldsymbol{\Xi}$、$\mathbf{D}_{\boldsymbol{\beta}_1}$、 $\boldsymbol{\Lambda}$、$\mathbf{P}$ 和 $\boldsymbol{\delta}$ 为固定值，即定义为

$$\tilde{\mathbf{w}}^o = \arg\min_{\tilde{\mathbf{w}} \in \tilde{\mathcal{H}}} \mathrm{AMSE}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}})), \tag{15}$$

˜且$\mathbf{W}^o = \mathrm{Diag}(\tilde{\mathbf{w}}^o)$。

4.2 插件式聚焦加权平均最小二乘

我们已讨论聚焦加权平均最小二乘的渐近均方误差；然而，W 的最优解˜不可行，因为它依赖于未知参数，包括公式（14）中的 $\mathrm{D}_{\boldsymbol{\beta}_1}$、$\mathrm{Q}_{11}$, B、C、V、 Ξ和 $\boldsymbol{\delta}$。显然，Q_{11}, C、 Ξ 可分别通过 ˆQ_{11}, ˆC $=$ ˆ$\boldsymbol{\Lambda}$ˆ$\mathrm{P}^{-1/2}$ 和 ˆΞ 一致估计。对于 $\mathrm{D}_{\boldsymbol{\beta}_1}$ 和 V，我们可以使用完整模型估计量 ˆ $\boldsymbol{\beta}_1 =$ ˆ$\boldsymbol{\beta}_{1\mathrm{narrow}} -$ ˆΞ ˆ$\boldsymbol{\Lambda}$ˆ$\mathrm{P}^{-1/2}$ˆ$\boldsymbol{\beta}_2$；当 W $=$I 时，它是公式（10）的一个特殊˜ ˆ ˆ 情形，并将 $\mathrm{D}_{\hat{\boldsymbol{\beta}}_1}$ 定义为在 $\boldsymbol{\beta}_1$处取值的 $\mathrm{D}_{\boldsymbol{\beta}_1}$。由于 $\boldsymbol{\beta}_1$是一致的，$\mathrm{D}_{\hat{\boldsymbol{\beta}}_1}$ 和 ˆV $=\mathrm{Diag}(\mathrm{D}_{\hat{\boldsymbol{\beta}}_1}^{\mathsf{T}}$ ˆΞˆC) $=$ $\mathrm{Diag}($ ˆCᵀˆΞᵀ$\mathrm{D}_{\hat{\boldsymbol{\beta}}_1})$ 也一致。

然而，根据定理1中的第二个结果，我们只能在局部趋近于零框架下获得$\mathbf{C}^{-1}\boldsymbol{\delta}$的无偏估计。因此，我们可以进一步定义ˆ$\boldsymbol{\delta} = \sqrt{N}$ˆ$\mathbf{C}$ˆ$\boldsymbol{\beta}_2$，由此可得:

$$\hat{\boldsymbol{\delta}} \xrightarrow{d} \boldsymbol{\delta} + \mathbf{C}\mathbf{B}\mathbf{R} \equiv \mathbf{R}_{\boldsymbol{\delta}} \sim \mathrm{N}(\boldsymbol{\delta}, \mathbf{C}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}}), \tag{16}$$

$$\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^{\mathsf{T}} \xrightarrow{d} \boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}} + \boldsymbol{\delta}\mathbf{C}\mathbf{B}\mathbf{R} + \mathbf{R}^{\mathsf{T}}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}}\boldsymbol{\delta}^{\mathsf{T}} + \mathbf{C}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}} \equiv \boldsymbol{\Sigma}_{\boldsymbol{\delta}}. \tag{17}$$

方程（16）和（17）给出了估计值 $\boldsymbol{\delta}$ 的性质，该估计值用于计算渐近偏差。上述结果所蕴含的第一个性质是， $\boldsymbol{\delta}$ 是 $\boldsymbol{\delta}$ 的无偏估计量；第二个性质是，我们可以得到

如果我们要计算$\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^{\mathsf{T}}$，它就不是 $\boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}$ 的无偏估计量。因此，应将其修改为：

$$\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^{\mathsf{T}} - \hat{\mathbf{C}}\hat{\mathbf{B}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{B}}^{\mathsf{T}}\hat{\mathbf{C}}^{\mathsf{T}} \xrightarrow{d} \boldsymbol{\Sigma}_{\boldsymbol{\delta}} - \mathbf{C}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}}, \tag{18}$$

并且 $\mathbb{E}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}} - \mathbf{C}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}}) = \boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}$。因此，我们可以直接将$\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^{\mathsf{T}}$ _ ˆ**C**ˆ**B** ˆ**Ω**ˆ**B**^{T} ˆ**C**^{T}代入公式（14），从而得到聚焦加权平均最小二乘的渐近均方误差估计量如下：

$$\begin{aligned}
\widehat{\mathrm{AMSE}}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}})) =& \mathbf{D}_{\hat{\boldsymbol{\beta}}_1}^{\mathsf{T}}\hat{\boldsymbol{\Xi}}\left(\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^{\mathsf{T}} - \hat{\mathbf{C}}\hat{\mathbf{B}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{B}}^{\mathsf{T}}\hat{\mathbf{C}}^{\mathsf{T}}\right)\hat{\boldsymbol{\Xi}}^{\mathsf{T}}\mathbf{D}_{\hat{\boldsymbol{\beta}}_1} \\
&+ \tilde{\mathbf{w}}^{\mathsf{T}}\hat{\mathbf{V}}\hat{\mathbf{C}}^{-1}\left(\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^{\mathsf{T}} - \hat{\mathbf{C}}\hat{\mathbf{B}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{B}}^{\mathsf{T}}\hat{\mathbf{C}}^{\mathsf{T}}\right)\hat{\mathbf{C}}\hat{\mathbf{V}}\tilde{\mathbf{w}} \\
&- 2\tilde{\mathbf{w}}^{\mathsf{T}}\hat{\mathbf{V}}\hat{\mathbf{C}}^{-1}\left(\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^{\mathsf{T}} - \hat{\mathbf{C}}\hat{\mathbf{B}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{B}}^{\mathsf{T}}\hat{\mathbf{C}}^{\mathsf{T}}\right)\hat{\boldsymbol{\Xi}}^{\mathsf{T}}\mathbf{D}_{\hat{\boldsymbol{\beta}}_1} + \mathbf{D}_{\hat{\boldsymbol{\beta}}_1}^{\mathsf{T}}\hat{\mathbf{Q}}_{11}^{-1}\mathbf{D}_{\hat{\boldsymbol{\beta}}_1} \\
&+ \tilde{\mathbf{w}}^{\mathsf{T}}\hat{\mathbf{V}}\hat{\mathbf{B}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{B}}^{\mathsf{T}}\hat{\mathbf{V}}\tilde{\mathbf{w}} + 2\tilde{\mathbf{w}}^{\mathsf{T}}\hat{\mathbf{V}}\hat{\mathbf{B}}\hat{\boldsymbol{\Omega}}\mathbf{H}\hat{\mathbf{Q}}_{11}^{-1}\mathbf{D}_{\hat{\boldsymbol{\beta}}_1} \\
\xrightarrow{d}& \mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}}\boldsymbol{\Xi}\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}} - \mathbf{C}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}}\right)\boldsymbol{\Xi}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\beta}_1} \\
&+ \tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{-1}\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}} - \mathbf{C}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}}\right)\mathbf{C}\mathbf{V}\tilde{\mathbf{w}} \\
&- 2\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{-1}\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}} - \mathbf{C}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{C}^{\mathsf{T}}\right)\boldsymbol{\Xi}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\beta}_1} + \mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\boldsymbol{\beta}_1} \\
&+ \tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathsf{T}}\mathbf{V}\tilde{\mathbf{w}} + 2\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{B}\boldsymbol{\Omega}\mathbf{H}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\boldsymbol{\beta}_1} \\
\equiv& \mathrm{AMSE}^{*}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}})).
\end{aligned} \tag{19}$$

此外，上述结果还表明，$\mathbb{E}\left(\mathrm{AMSE}^{*}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}}))\right) = \hat{}\mathrm{AMSE}(\mu(\boldsymbol{\beta}_{1\mathrm{WALS}})$，因此，公式（19）中定义的渐近均方误差估计量是公式（14）中定义的不可行渐近均方误差的渐近无偏估计量。

遵循 Lu (2015) 的方法，我们可以应用 argmax连续映射定理，得到给定数据驱动的 ˆ**w** 后聚焦加权平均最小二乘的极限分布。

ˆ定理3。 在假设 1-3 下，$\boldsymbol{\beta}_{1\mathrm{WALS}}$ 基于 ˆ**w** 和 ˆ**w** $=\widehat{\quad}$ ˆ arg min$_{\tilde{\mathbf{w}}\in\tilde{\mathcal{H}}}$ AMSE($\mu(\boldsymbol{\beta}_{1\mathrm{WALS}})$) 计算，当 $N \to \infty$ 时，我们有

$$\begin{aligned}
&\hat{\mathbf{w}} \xrightarrow{d} \tilde{\mathbf{w}}^{*} = \arg\min_{\tilde{\mathbf{w}}\in\tilde{\mathcal{H}}} \mathrm{AMSE}^{*}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}})), \\
&\sqrt{N}(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}} - \boldsymbol{\beta}_1) \xrightarrow{d} R_{\mu}(\tilde{\mathbf{W}}^{*}), \quad \tilde{\mathbf{W}}^{*} = \mathrm{Diag}(\tilde{\mathbf{w}}^{*}).
\end{aligned}$$

注释2。我们的计算简化依赖于将式（7）中的半正交化与加权表示相结合，从而将权重选择问题从 2^{k_2} 维模型空间简化为 k_2 个标量权重。这一思路与 Zhu 等人（2023）的可扩展平均框架有关，但权重的构造方式不同。Zhu 等人（2023）考虑 Mallows 模型平均（Hansen 2007）

和 Jackknife 模型平均（Hansen 和 Racine 2012）准则，而我们的权重则通过最小化一个插入式渐近均方误差得到；该误差是针对局部趋零条件下的聚焦参数，在假设1中明确设计的。在 MMA/JMA 文献中，通常也会将关注范围限制在一部分候选模型上，相应的渐近最优性则是在该受限集合条件下建立的。我们的框架也可以施加类似限制；在这种情况下，无偏渐近均方误差的刻画及由此得到的权重，应相对于所选候选集来理解。然而，除非对$\mathbf{X}_2$ 施加额外结构（例如近似因子结构），或采用筛选步骤（例如确定性独立筛选）来论证缩减候选模型的合理性，否则在受限集合下计算的渐近均方误差通常无法与在完整集合下计算的结果直接比较。我们将此类结构化限制的正式分析留待未来研究。

4.3 与贝叶斯方法的比较

尽管本文提出的方法与 Magnus 等人（2010）采用的正交化理念相同，但我们以根本不同的方式确定子模型权重。为说明主要区别，令

$$t_j = \frac{\hat{\beta}_{2j}}{\sigma_j}, \tag{20}$$

其中，$\hat{}\ \beta_{2j}$ 是$\hat{}\ \boldsymbol{\beta}_2 = \mathbf{x}_2^{*\intercal}\ \mathbf{M}_1\mathbf{y}\ N$ 中的第 j个估计值。假设σ_j 已知。在 Magnus 等人（2010）和 Luca 等人（2022）所采用的分布假设下，或在 Luca 等人（2025）讨论的渐近框架下，t_j 可视为均值为 η_j 、方差为 1 的正态分布。根据这一结果，并选取合适的先验密度 π（η_j），我们可以应用 Tweedie公式（例如 Pericchi 和 S mith，1992）求得后验均值 m（t_j）$= \mathbb{E}$（$\eta|t_j$），其表达式为:

$$m(t_j) = t_j + \frac{p'(t_j)}{p(t_j)}, \tag{21}$$

其中 $p(t_j) = \int_{-\infty}^{\infty} \phi(t_j - \eta)\pi(\eta)d\eta$。因此，我们可以将先验为 $\pi(\eta)$ 的 WALS 估计量定义为:

$$\begin{aligned}
\hat{\beta}_{2j,WALS-prior} =& m(t_j)\sigma_j \\
=& \left(t_j + \frac{p'(t_j)}{p(t_j)}\right)\sigma_j \\
=& \hat{\beta}_{2j}\left(1 + \frac{p'(t_j)}{t_j p(t_j)}\right)
\end{aligned}$$

$$= \hat{\beta}_{2j}\omega_j. \tag{22}$$

这一表述与 Magnus 等人（2010）提出的标准 WALS 估计量一致。采用这种表述方式，我们强调贝叶斯收缩的作用恰好相当于一个标量权重 ω_j，并将其应用于 β_{2j}。术语 ω_j 的作用与公式（11）中定义的权重类似。不过，ω_j 是通过后验均值获得的，而后验均值的作用是界定 $\beta_{2j,\ WALS-prior}$ 的极大极小遗憾，这一点与 Magnus 等人（2010）、Luca 等人（2022）和 Luca 等人（2025）的建议一致。相比之下，我们的方法通过直接最小化渐近均方误差来求得权重，而渐近均方误差计算的是偏差平方与方差的精确总和。更重要的是，我们的方法可以针对聚焦参数来设计渐近均方误差。

为了展示贝叶斯后验均值所隐含的权重与我们采用渐近均方误差的方法之间的行为差异，我们考虑一个简单的数据生成过程 $y_i = x_i\beta + \epsilon_i$。我们假设局部趋零框架 $\beta = \frac{\delta}{\sqrt{N}}$，$\mathbb{E}(x_i^2) = 1$，且 $\mathbb{E}(\epsilon_i^2) = \sigma^2$。考虑平均估计量

$$\hat{\beta}(w) = w\hat{\beta}_{OLS} + (1-w)\cdot 0, \qquad w \in [0,1], \tag{23}$$

其中 β_{OLS} 是无偏的，但具有估计方差；而无意义估计量等于零，因此没有估计方差。假设我们事先知道 δ 的真实值。理论上，基于渐近均方误差得到的最优权重形式为 $\frac{(\delta/\sigma)^2}{(\delta/\sigma)^2+1}$，而贝叶斯后验均值所隐含的权重为 $\left(1 + \frac{p'}{\delta/\sigma}\frac{(\delta/\sigma)}{p(\delta/\sigma)}\right)^{-1}$。

在图 1 中，我们绘制了理论最优渐近均方误差权重的变化情况（不包括当 $\delta = 0$ 时的点），并将其与 Luca 等人（2025）讨论的四种先验所隐含的权重进行了比较。[2] 如图所示，当信噪比（δ/σ）趋近于零时，理论最优渐近均方误差权重严格收敛于零。相比之下，当 $\delta/\sigma \to 0$时，贝叶斯后验均值所隐含的权重仍为正，因此收缩程度较为温和、更加保守。此外，随着 δ/σ 增大，我们的渐近均方误差方法会迅速收敛至权重 1，完全依赖预言估计量。尽管重尾先验（柯西分布、威布尔分布、帕累托分布）最终也会收敛至 1，但标准拉普拉斯先验会保留永久收缩惩罚，从而产生无法完全消除的恒定偏差。尽管这个数值例子假定已知真实参数，而这在实践中不可行，但它阐明了这两个框架在机制上的显著差异。我们将在模拟部分进一步考察这些差异如何转化为有限样本表现。

[2]四种先验分别是拉普拉斯分布、威布尔分布、帕累托分布和柯西分布。这些先验的完整定义及数值积分的详细内容见附录 B。

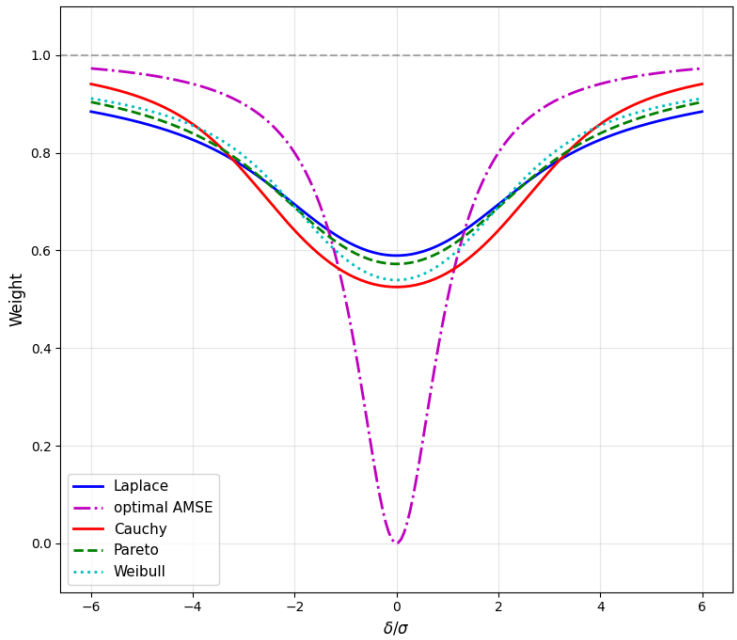

图1 权重：理论渐近均方误差与贝叶斯先验的比较

注3。 Luca 等人（2025）在先验满足适当条件的情况下，推导了 WALS 型估计量的渐近分布。具体而言，他们证明，在局部趋零设定下，所得渐近分布可能呈现显著的非正态性（例如偏度和超额峰度），且其极限分布取决于先验通过后验均值所隐含的收缩规则产生的影响。我们的方法则不同于这一贝叶斯视角。在我们的框架中，权重通过最小化专为聚焦参数构造的插入式渐近均方误差准则来选择（定理3），而非由预先指定的先验生成。因此，定理3中的极限分布由渐近均方误差最优加权机制决定，而贝叶斯加权平均最小二乘极限则随先验而异，这种差异通过先验诱导的收缩函数体现出来。因此，即使在相同的局部趋零假设下，这两种方法仍由不同的权重选择映射所支配（先验诱导的收缩与渐近均方误差最小化）。关于使用先验实施加权平均最小二乘的实践方法，参见 Luca 和 Magnus（2025b,a）；相应的R实现可通过`WALS`软件包获取。

5 模拟研究

本节通过蒙特卡洛模拟评估所提出的聚焦平均法的表现。我们考察两个关键方面。对于两种数据生成过程，我们通过比较不同方法所得聚焦参数的均方误差（MSE）与不可行子模型的均方误差，来评估风险。我们用记号 N 表示基本设计的样本量，用 T 表示脉冲响应函数设计的样本量。

5.1 基本设计

数据生成过程通常基于 Liu (2015) 中的设计，以便于进行比较。我们考虑以下模型设定：

$$y_i = \mathbf{x}_{1i}^{\mathsf{T}}\boldsymbol{\beta}_1 + \mathbf{x}_{2i}^{\mathsf{T}}\boldsymbol{\beta}_2 + \epsilon_i, \quad i = 1, ..., N, \tag{24}$$

其中，$\mathbf{x}_i = [\mathbf{x}_{1i}^{\mathsf{T}}\ \mathbf{x}_{2i}^{\mathsf{T}}]$ 服从均值为零、协方差矩阵为 $\boldsymbol{\Sigma}_{\mathbf{x}}$ 的多元正态分布，其中对角元素为 1，非对角元素为 τ。ϵ_i 服从标准正态分布。核心回归变量 $\mathbf{x}_{1i}$和辅助回归变量 $\mathbf{x}_{2i}$的维度分别为 k_1 和 k_2。斜率系数设定如下：

$$\boldsymbol{\beta}_1 = \frac{c_x}{a}[1\ \dots\ 1]^{\mathsf{T}}, \tag{25}$$

$$\boldsymbol{\beta}_2 = c_x\left[1\quad \frac{k_2-1}{k_2}\ \dots\ \frac{1}{k_2}\right]^{\mathsf{T}}, \tag{26}$$

其中，a 控制核心回归变量和辅助回归变量的相对重要性。较大的a意味着核心回归变量的系数较小，使其相对不那么重要。参数 c_x 决定总解释力，并通过对 c_x的不同选择控制 R^2 值。

在我们的第一个实验设计中，我们考虑在给定 $a = 12$和 $k_1 = 3$的情况下，$\tau = \{0.3, 0.5, 0.7\}$、$k_2 = \{2, 4, 7\}$、$R^2 = \{0.1, 0.2, ..., 0.9\}$、$N = \{100, 200\}$的所有组合。该设计旨在考察小样本性能如何受到回归变量之间的相关性以及模型复杂度变化的影响。本例中所考虑的聚焦参数定义为

$$\begin{aligned}\mu =& \mu(\boldsymbol{\beta}_1)\\ =& \beta_{11} + \beta_{12} + \beta_{13}.\end{aligned} \tag{27}$$

5.2 脉冲响应函数设计

该数据生成过程旨在考察平均估计量在脉冲响应函数（IRF）不同预测期下的表现。我们考虑的样本量为 $T = 100$。因变量 y_t 与 k_2 个辅助回归变量联合生成，其中 $k = k_1 + k_2$ 和 $k_1 = 3$ 固定为 y_t的自回归滞后项个数。据此，我们令 $k \in \{5, 7, 10\}$，从而 $k_2 \in \{2, 4, 7\}$。

因变量由具有 $k_1 = 3$ 阶滞后的自回归结构生成：

$$\begin{aligned} y_t &= \mathbf{x}_{1t}^{\mathsf{T}}\boldsymbol{\beta}_1 + \mathbf{x}_{2t}^{\mathsf{T}}\boldsymbol{\beta}_2 + u_t, \\ &= \sum_{j=1}^{k_1} \beta_{1j} y_{t-j} + \mathbf{x}_{2t}^{\mathsf{T}}\boldsymbol{\beta}_2 + u_t. \end{aligned} \tag{28}$$

自回归系数为

$$\beta_{11} = 0.5, \qquad \beta_{1j} = \frac{d}{\sqrt{T}(j-1)}, \ \ j = 2, 3. \tag{29}$$

辅助回归变量遵循平稳自回归过程

$$\mathbf{x}_{2t} = 0.2\,\mathbf{x}_{2,t-1} + \mathbf{e}_{\mathbf{x},t}, \tag{30}$$

其中 $\mathbf{e}_{\mathbf{x},t} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_x)$，其中 $\boldsymbol{\Sigma}_x$ 是一个 $k_2 \times k_2$ 协方差矩阵，其对角元素为 1，非对角元素为 τ。创新项 $\mathbf{e}_{\mathbf{x},t}$ 与结构冲击 u_t 相互独立，其中 $u_t \sim \mathrm{N}(0,1)$。

对于辅助系数，我们令 $s = \lfloor k_2/2 \rfloor$ 并采用近稀疏局部设定：

$$\boldsymbol{\beta}_2 = \frac{c_y}{\sqrt{T}}\boldsymbol{\theta}_y, \qquad \boldsymbol{\theta}_y = \Big[\underbrace{1 \ \cdots \ 1}_{s \text{ entries}} \ \underbrace{0.05 \ \cdots \ 0.05}_{k_2 - s \text{ entries}}\Big]^{\mathsf{T}}, \tag{31}$$

而辅助回归变量的贡献由缩放常数 c_y 控制。

初始值（y_{-2}，y_{-1}，y_0）设为零，且$\mathbf{x}_{2,\,0} = \mathbf{0}$。模拟$T + 100$ 个观测值后，舍弃前 100 个，保留最后的 T 个作为有效样本。我们考虑 $c_y = \{0.1, .., 4\}$ 的所有组合，其中包含 10 个网格，分别取值为 $d = 1$ 和 $\tau = 0.2$。对于每次模拟，我们评估平均估计量在不同 IRF期数 $h = \{1, 3, 5, 7\}$ 下的表现。

在此设计中考虑的聚焦参数，是在不同期数（h）下计算的脉冲响应。更具体地说，其定义为

$$\begin{aligned}\mu_h =&\mu_h(\boldsymbol{\beta}_1)\\ =&\frac{\partial y_{t+h}}{\partial u_t}\\ =&e_1^{\mathsf{T}} A_\beta^h e_1,\end{aligned} \tag{32}$$

其中， e_1 是第一个标准基向量，定义为第一个分量为 1、其余分量为 0 的向量，$A_\beta^h = (A_\beta)^h$，而 A_β 表示伴随矩阵

$$A_\beta = \begin{bmatrix}\beta_{11} & \beta_{12} & \beta_{13}\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}. \tag{33}$$

聚焦参数的相应导数为

$$\frac{\partial \mu_h}{\partial \beta_j} = \sum_{i=0}^{h} e_1^{\mathsf{T}} A_\beta^i e_1 e_j^{\mathsf{T}} A_\beta^{h-1-i} e_1, \tag{34}$$

其中， e_j 表示第 j个标准基向量。详情见 Lohmeyer et al. (2019)。

5.3 替代方法

除所提出的聚焦加权平均最小二乘（记为FWALS）外，我们还进一步给出了使用 Liu (2015) 提出的聚焦平均估计量（记为FIC）、平滑赤池信息准则（SAIC）、平滑贝叶斯信息准则（SBIC）、Charkhi 等人 (2016) 的最小化均方误差方法（记为 mMSE），以及分别采用拉普拉斯、柯西分布、帕累托分布和威布尔分布先验的加权平均最小二乘（WALS − Lap,WALS − Cau, WALS − Par, WALS −Wei）所得的结果。[3][4][5] 我们对所有情形均以 1,000 次重复试验进行模拟。

---

[3]候选模型 m 的赤池信息准则值按公式 $AIC_m = N \ln\left(\sum_{i=1}^{N} \hat{\epsilon}_{mi}^2/N\right) + 2(k_1 + k_{2m})$ 计算，其中 k_1 和 k_{2m} 分别表示残差和所选辅助回归变量的数量。SAIC 的权重可定义为 $\hat{w}_m = \exp(-AIC_m/2)/\sum_{m=1}^{M}\exp(-AIC_m/2)$。[4]候选模型 m 的贝叶斯信息准则值按公式 $BIC_m = N \ln\left(\sum_{i=1}^{N} \hat{\epsilon}_{mi}^2/N\right) + \ln(N)(k_1 + k_{2m})$ 计算，其中 $\hat{\epsilon}_{mi}$ 和 k_{2m} 分别表示残差和所选辅助回归变量的数量。SBIC 的权重可定义为 $\hat{w}_m = \exp(-BIC_m/2)/\sum_{m=1}^{M}\exp(-BIC_m/2)$。[5]先验的详情见附录 B。

5.4 小样本性质

我们在图 2 至图 7 中总结了第一个实验的模拟结果。每幅图展示了辅助回归变量个数取三个不同数值时的结果；并且，对于给定的辅助回归变量个数，我们给出了所有方法中风险相对于 R^2 的动态变化。

我们首先讨论 $N = 100$时的情形。在图2中，就非 WALS 型方法而言，我们观察到，在大多数情况下，当回归变量之间的相关性相对较弱时，无论辅助回归变量的数量如何，FWALS、FIC和mMSE 的表现都非常相似。这表明所提出方法中的变换不会对聚焦参数的渐近均方误差产生负面影响。与SAIC 和SBIC相比，FWALS、FIC 和 mMSE 通常表现更好，除非 R^2 较小且 k_2 较大。此外，当真实模型较为复杂、包含许多弱预测变量时，SBIC往往表现较差（风险较高），这很可能是因为它更偏好较简单的模型，而非关注聚焦参数的准确性。另一方面，SAIC表现更好，很可能是因为它对相对复杂的模型给予了更多重视。

至于 WALS 型方法，无论先验如何选择，我们都观察到加权平均最小二乘 –Lap、加权平均最小二乘 – Cau、加权平均最小二乘 –Par和加权平均最小二乘 –Wei 呈现出相似的模式。当 R^2 较大时，带先验的加权平均最小二乘表现更差。这主要是因为该设计并未用于最小化聚焦参数的风险。此外，加权平均最小二乘 –Lap 在 R^2 较大时显示出永久收缩（权重小于1）的潜在缺陷，这可能使聚焦参数产生偏差。总体而言，我们观察到，增大 k_2 进一步体现了使用聚焦型方法的优势。

接下来，我们考察不同 τ 值下的结果，如图3和图4所示。可以观察到类似的模式：FWALS、FIC 和 mMSE 通常表现更好，尤其是在 $\tau = 0.7$、 $k_2 = 7$ 和 R^2 较高时。此外，我们观察到，随着τ 增加，当 R^2 相对较小时，FWALS 略优于FIC 。这可以通过变换在分离经变换回归变量的贡献方面的优势来解释。然而，随着 R^2 增加，这一优势会逐渐减弱，因为较高的 R^2 意味着所有辅助回归变量对聚焦参数的影响都更大。在这种情况下，通过FWALS 施加更多权重限制反而成为一种劣势，而考虑所有组合则会取得更好的结果。这种权衡凸显了该方法的一个关键方面。总体而言，模拟结果表明，在上述所有情形中，FWALS 与FIC 之间的差异都很小。此外，当 τ 较大时，使用预设先验的劣势仍会随着R^2 增加而变得明显。然而，我们观察到，当 R^2 较小且 τ 较大时，FWALS、FIC和mMSE的表现略差。这可以看作是我们

必须付出的代价，因为我们需要估计偏差，而这可能会扭曲渐近均方误差（AMSE）的估计，从而导致不令人满意的结果。在考虑情形 $N = 200$时，也可以发现类似的模式，如图5至图7所示。

图8至图10报告了在不同的 h 预测期及若干个 k_2 取值下，脉冲响应函数的风险表现。在所有预测期内，FWALS 和FIC 的风险动态都非常相似，这证实了所提出的变换在大幅节省计算成本的同时，并未牺牲相较于传统聚焦平均估计量的风险表现。当 $k_2 = 2$ 时，如图8所示，在所考察的预测期范围内，几乎所有方法的表现都相当。随着 k_2 增大，情况有所变化。当 k_2 增大、相对较弱的辅助回归变量数量增加时，我们再次发现，当 c_y 较大时，FWALS 和FIC 的表现不逊于基于先验的 WALS 方法，甚至更优（图9—图10）。当辅助回归变量的信号变弱（ c_y较小）时，FWALS和FIC 的偏差估计可能不够精确，正如第一个设计中所观察到的；不过，总体而言，这两种方法的表现仍然相似。相比之下，mMSE在不同预测期内的表现不稳定。这可能反映了基于单变量方程允许较宽权重空间的缺点，尤其是在聚焦函数为非线性时。最后，SBIC 和SAIC的表现模式总体上与FWALS、FIC及基于先验的 WALS 方法相似，不过就风险而言，SBIC 的波动似乎略大。

总的来说，这些模拟结果表明，FWALS 在不同预测期和 k_2 取值下，都是 FIC的有竞争力的替代方法，兼具计算效率和稳定的统计表现，其表现也比 mMSE更加稳定。基于先验的 WALS方法在某些情况下也具有竞争力，尤其是在辅助回归变量的信号较弱时。

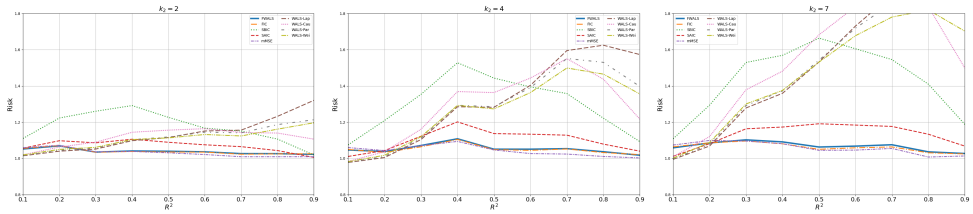

图2 不同 k_2s 下的风险（$N = 100$，$\tau = 0.3$）

5.5 计算时间比较

为突出所提方法的计算优势，我们记录了获取权重和估计值所需的实际计算时间。作为

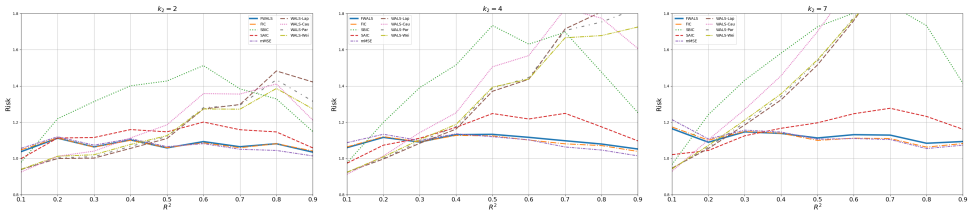

图 3 不同 k_2 下的风险（$N = 100,\ \tau = 0.5$）

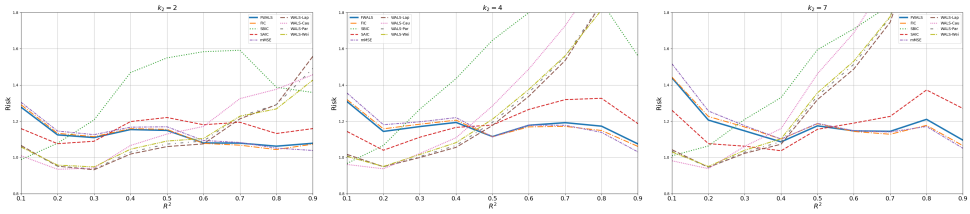

图 4 不同 k_2 下的风险（$N = 100,\ \tau = 0.7$）

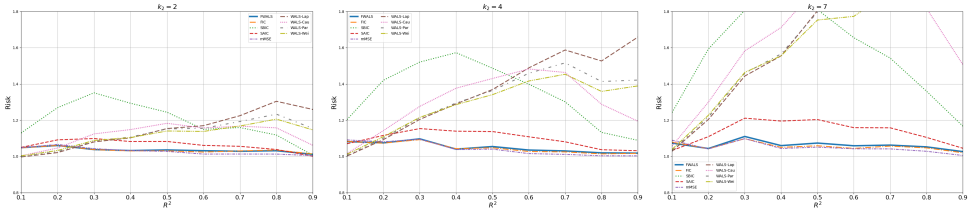

图 5 不同 k_2 下的风险（$N = 200,\ \tau = 0.3$）

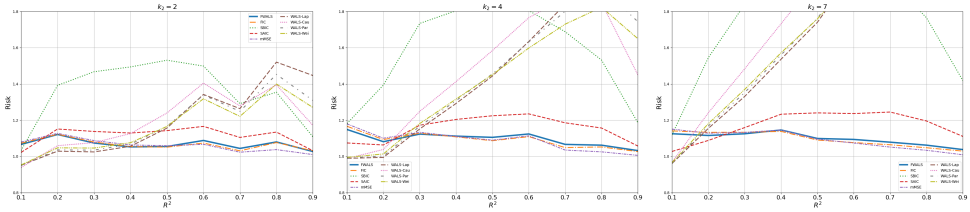

图 6 不同 k_2 下的风险（$N = 200,\ \tau = 0.5$）

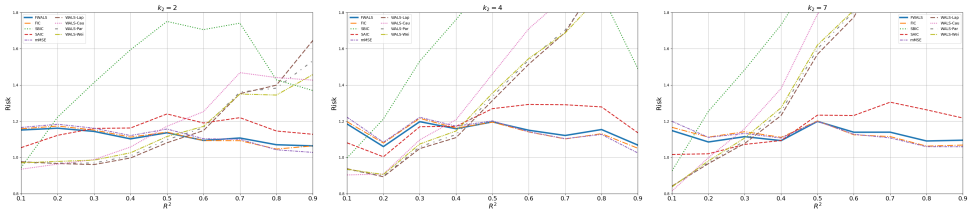

图 7 不同 k_2 下的风险（$N = 200,\ \tau = 0.7$）

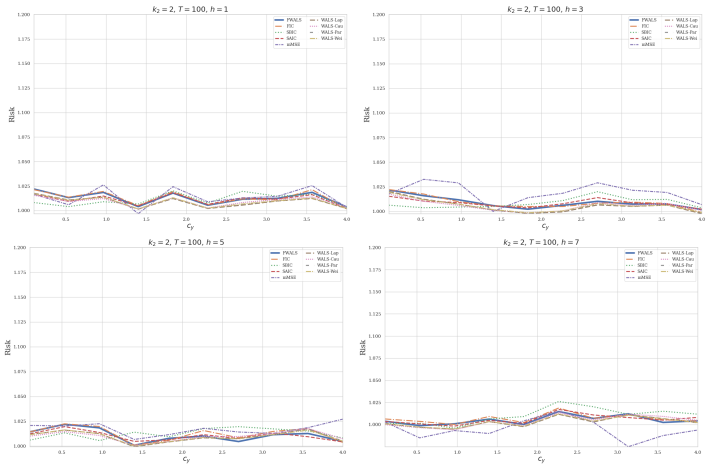

图 8 不同 hs ($T = 100$,$k_2 = 2$) 下的风险

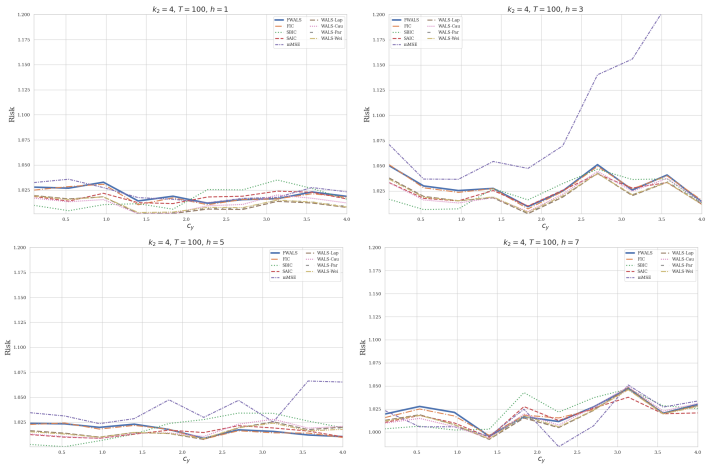

图 9 不同 hs ($T = 100$,$k_2 = 4$) 下的风险

正如理论部分所指出的，传统的频率学派模型平均技术需要估计并组合所有可能的 2^{k_2} 子模型。相比之下，我们提出的方法依赖于半正交变换，可将加权问题有效地从 2^{k_2} 子模型缩减至 k_2 正交化项。当辅助回归变量数量较少时（例如 $k_2 \leq 4$），计算负担可能并不明显；但随着 k_2增加，这种差异会变得十分显著。

为了说明这一点，我们使用 N =100的模拟数据集开展实验。我们将所提出的方法与FIC 和mMSE 方法进行比较，涵盖

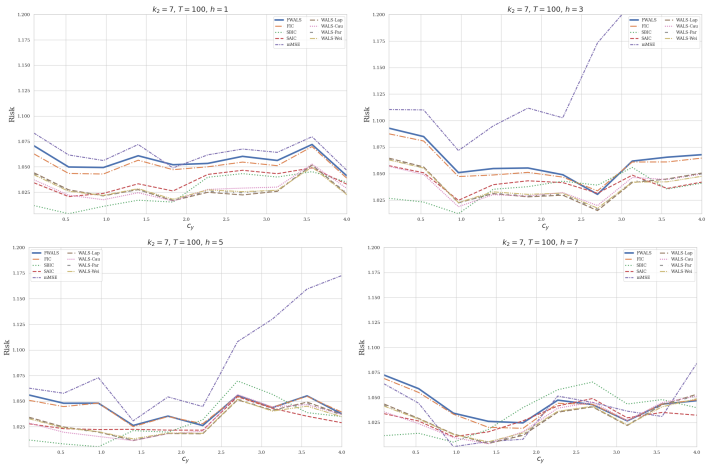

图10 不同 hs 下的风险（$T = 100$，$k_2 = 7$）

中等较大的辅助回归变量维度 $k_2 \in \{8, 9, 10, 11\}$。所有计算均在一台运行Pop! OS 22.04（基于Linux）的台式计算机上执行，该计算机配备Intel® Core™ i 9-10980XE CPU。所报告的计算时间按第一个实验中指定的所有设计上100次重复试验的平均值计算。

	$N = 100$		
k_2	WFIC	FIC	mMSE
8	0.012	0.851	0.130
9	0.015	10.330	0.992
10	0.016	137.098	9.804
11	0.017	1407.540	97.055

表1 不同方法的计算时间（秒）

结果列于表 1。如我们所见，当 $k_2 = 8$时，FIC 耗时 0.851 秒，mMSE 耗时 0.130 秒，而我们提出的方法仅需 0.012 秒。当 k_2 达到 11 时，FIC 需要 1,400 多秒（超过 23 分钟）来计算权重。相比之下，我们提出的方法的计算时间呈线性增长，且仍然极低，仅为 0.017 秒。这些证据表明，当 k_2 为中等或较大值时，所提出的方法具有良好的可扩展性和实用性；在这种情况下，对 2^{k_2} 个候选模型进行枚举在计算上

难以承受。有限样本表现与 k_2 较小时的情况相似，结果见附录 C（图 11–13）。

6 总结

本文通过引入基于正交化辅助回归变量的聚焦加权平均最小二乘（FWALS）估计量，为聚焦模型平均构建了一个计算高效的框架。该方法解决了传统聚焦模型平均的一项关键局限，即随着辅助回归变量数量增加，子模型数量呈指数级增长。通过按照 Magnus 等人（2010）和 De Luca 等人（2018）的方法对辅助回归变量进行变换，我们降低了加权问题的维度，从而在不牺牲统计有效性的前提下，提供了一种易于处理的替代方案。

在我们的模拟研究中，FWALS、FIC 和基于单变量方程的最小均方误差平均方法在基准设计中的表现相似，证实正交化步骤不会损害效率。在脉冲响应函数设定下，FWALS 与 FIC 仍然高度一致，并在各预测期内展现出稳定的风险表现；相比之下，基于单变量方程的方法可能在较长预测期内表现出更高的风险，这是由负权重所引发的不稳定性导致的。我们还将 FWALS 与基于先验的 WALS 方法（拉普拉斯分布、柯西分布、帕累托分布和威布尔先验）进行了比较。这些基于先验的程序在某些配置下具有竞争力，尤其是在辅助信号较弱时，但它们并非为最小化聚焦参数的风险而构造。相比之下，FWALS 通过最小化针对聚焦参数设计的插入式渐近均方误差准则来选择逐回归变量权重，从而为聚焦推断提供了一种有理论依据的方法。总体而言，结果表明，FWALS 是传统聚焦平均的一种具有竞争力且计算上颇具吸引力的替代方法，尤其是在无法对全部 2^{k_2} 候选模型进行穷举时。

参考文献

Charkhi, A.、G. Claeskens 和 B.E. Hansen。2016。似然模型中的最小均方误差模型平均。《中华统计学》。https://doi.org/10.5705/ ss.202014.0067 。Chen, Y.T.、C.A. Liu 和 J.H. Su。2025。预测组合的 Bregman模型平均。《计量经济学杂志》251：106076。https://doi.org/10.1016/j.jeconom. 2025.106076 。Claeskens, G. 和 N.L. Hjort。2003。聚焦信息准则。《美国统计学会会刊 98》(464)：900–916。Claeskens, G. 和 N.L. Hjort。2008。模型选择与模型平均。《剑桥图书》。De Luca, G.、J.R. Magnus 和 F. Peracchi。2018。广义线性模型的加权平均最小二乘估计。《计量经济学杂志 204》(1)：1–17。DiTraglia, F.J.。2016。有意使用无效工具变量：广义矩估计的聚焦矩选择与平均。《计量经济学杂志 195》(2)：187–208。Hansen, B.E.。2007。最小二乘模型平均。《计量经济学 75》(4)：1175–1189。Hansen, B.E. 和 J.S. Racine。2012。刀切模型平均。《计量经济学杂志 167》(1)：38–46。Hjort, N.L. 和 G. Claeskens。2006。Cox风险回归模型的聚焦信息准则与模型平均。《美国统计学会会刊 101》(476)：1449–1464。Kitagawa, T. 和 C. Muris。2016。处理效应半参数估计中的模型平均。《计量经济学杂志 193》(1)：271–289。Liu, C.A.。2015。最小二乘平均估计量的分布理论。《计量经济学杂志 186》(1)：142–159。https://doi.org/10.1016/j.jeconom.2014.07.002 。Lohmeyer, J.、F. Palm、H. Reuvers 和 J.P. Urbain。2019。局部错设向量自回归模型的聚焦信息准则。《计量经济学评论 38》(7)：763–792。

Lu, X. 2015. 用于估计处理效应的协变量选择准则。《商业与经济统计期刊》*33*(4): 506–522. https://doi.org/10. 1080/07350015.2014.982755 . Luca, G.D. and J.R. Magnus. 2025a. 加权平均最小二乘：超越经典线性回归模型。《*Stata*期刊》*25*(4): 772–811 . Luca, G.D. and J.R. Magnus. 2025b. 加权平均最小二乘：改进与扩展。《*Stata*期刊》*25*(3): 587–626 . Luca, G.D., J.R. Magnus, and F. Peracchi. 2022. 贝叶斯后验均值的抽样性质及其在 WALS 估计中的应用。《计量经济学杂志》*230*(2): 299–317. https://doi.org/10.1016/j.jeconom.2021.04.008 . Luca, G.D., J.R. Magnus, and F. Peracchi. 2025. 正态位置模型的贝叶斯估计：一种非标准方法。《牛津经济学与统计学公报》*87*(5): 913–923. https://doi.org/10.1111/obes.12672 . Magnus, J.R., O. Powell, and P. Pr¨ufer. 2010. 两种模型平均技术的比较及其在增长实证研究中的应用。《计量经济学杂志》*154*(2): 139–153. https://doi.org/10.1016/j.jeconom.2009.07.004 . Pericchi, L. and A. Smith. 1992. 正态位置参数的精确与近似后验矩。《英国皇家统计学会杂志*B*辑：统计方法学》*54*(3): 793–804 . Steel, M.F. 2020. 模型平均及其在经济学中的应用。《经济文献杂志》*58*(3): 644–719 . Yin, S.y., C.a. Liu, and C.C. Lin. 2021. 具有多因子误差结构的大型面板数据的聚焦信息准则与模型平均。《商业与经济统计期刊》*39*(1): 54–68 . Zhang, X. and C.A. Liu. 2019. 线性回归模型中模型平均后的推断。《计量经济理论》*35*(4): 816–841 . Zhang, X. and C.A. Liu. 2023. 通过 k 折交叉验证进行模型平均预测。《计量经济学杂志》*235*(1): 280–301. https://doi.org/10.1016/j.jeconom.2022.04.007 .

Zhang, X. and C.A. Liu. 2024. 聚焦信息准则与插件平均法的统一方法。《中华统计学》34: 771–792 .

Zhang, X.、A.T. Wan 和 S.Z. Zhou。2012。非零阈值 Tobit 模型中的聚焦信息准则、模型选择与模型平均。《商业与经济统计期刊》*30*(1)：132–142。

Zhang, X. 和 X. Zhang。2023年。基于前向验证的最优模型平均。《计量经济学杂志》*237*(2)：105295。https://doi.org/10.1016/j.jeconom.2022. 03.010。

Zhu, R.、H. Wang、X. Zhang 和 H. Liang。2023年。一种可扩展频率学派模型平均方法。《商业与经济统计期刊》*41*(4)：1228–1237。https://doi.org/10.1080/07350015.2022.2116442。

附录A

为简化记号，在整个证明过程中我们令 $\boldsymbol{\Xi} = \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}$ 和 $\mathbf{C} = \boldsymbol{\Lambda}\mathbf{P}^{-1/2}$ 。

Proof of Theorem 1 我们首先讨论$\hat{} \boldsymbol{\beta}_2$的极限行为。回顾一下，$\hat{} \boldsymbol{\beta}_2 = \mathbf{X}_2^* \intercal \mathbf{M}_1 \mathbf{y} \, N$ ，我们可以得到如下表示:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_2 =& \frac{\mathbf{X}_2^{*\mathsf{T}}\mathbf{M}_1(\mathbf{X}_2^*\boldsymbol{\beta}_2^* + \boldsymbol{\epsilon})}{N} \\ =& \boldsymbol{\beta}_2^* + \frac{1}{N}\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\mathbf{X}_2^{\mathsf{T}}\left(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^{\mathsf{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathsf{T}}\right)\boldsymbol{\epsilon} \\ =& \boldsymbol{\beta}_2^* + \frac{1}{N}\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\mathbf{X}_2^{\mathsf{T}}\boldsymbol{\epsilon} - \frac{1}{N}\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\mathbf{X}_2^{\mathsf{T}}\mathbf{X}_1(\mathbf{X}_1^{\mathsf{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathsf{T}}\boldsymbol{\epsilon}.\end{aligned}$$

因此，可以证明

$$\begin{aligned}\sqrt{N}\hat{\boldsymbol{\beta}}_2 =& \sqrt{N}\boldsymbol{\beta}_2^* + \hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\frac{1}{\sqrt{N}}\mathbf{X}_2^{\mathsf{T}}\boldsymbol{\epsilon} - \hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\frac{1}{N}\mathbf{X}_2^{\mathsf{T}}\mathbf{X}_1\left(\frac{1}{N}\mathbf{X}_1^{\mathsf{T}}\mathbf{X}_1\right)^{-1}\frac{1}{\sqrt{N}}\mathbf{X}_1^{\mathsf{T}}\boldsymbol{\epsilon} \\ =& \hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\sqrt{N}\boldsymbol{\beta}_2 + \begin{bmatrix}\mathbf{0} & \hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\end{bmatrix}\frac{1}{\sqrt{N}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\epsilon} + \begin{bmatrix}-\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\frac{1}{N}\mathbf{X}_2^{\mathsf{T}}\mathbf{X}_1\left(\frac{1}{N}\mathbf{X}_1^{\mathsf{T}}\mathbf{X}_1\right)^{-1} & \mathbf{0}\end{bmatrix}\frac{1}{\sqrt{N}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\epsilon} \\ =& \hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\delta} + \begin{bmatrix}\mathbf{0} & \hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\end{bmatrix}\frac{1}{\sqrt{N}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\epsilon} + \begin{bmatrix}-\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{Q}}_{21}\hat{\mathbf{Q}}_{11}^{-1} & \mathbf{0}\end{bmatrix}\frac{1}{\sqrt{N}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\epsilon} \\ =& \hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\delta} + \begin{bmatrix}-\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{Q}}_{21}\hat{\mathbf{Q}}_{11}^{-1} & \hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\end{bmatrix}\frac{1}{\sqrt{N}}\mathbf{X}^{\mathsf{T}}\boldsymbol{\epsilon}.\end{aligned}$$

第二个等式由公式（7）得出，第三个等式则来自假设1。利用假设2和3，我们可以得到以下结果:

$$\sqrt{N}\hat{\boldsymbol{\beta}}_2 - \hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\delta} \xrightarrow{d} \begin{bmatrix}-\mathbf{C}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}} & \mathbf{C}^{\mathsf{T}}\end{bmatrix}\mathbf{R}.$$

至于 $\boldsymbol{\beta}_{1\text{WALS}}$，根据公式（10）和 $\boldsymbol{\beta}_2$ 的结果，我们可以得到

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{1\text{WALS}} =& \boldsymbol{\beta}_1 + \left(\frac{\mathbf{X}_1^\mathsf{T}\mathbf{X}_1}{N}\right)^{-1}\frac{\mathbf{X}_1^\mathsf{T}\mathbf{X}_2}{N}\boldsymbol{\beta}_2 + \left(\frac{\mathbf{X}_1^\mathsf{T}\mathbf{X}_1}{N}\right)^{-1}\frac{\mathbf{X}_1^\mathsf{T}\boldsymbol{\epsilon}}{N} \\
& - \left(\frac{\mathbf{X}_1^\mathsf{T}\mathbf{X}_1}{N}\right)^{-1}\frac{\mathbf{X}_1^\mathsf{T}\mathbf{X}_2}{N}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\left(\hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\beta}_2 + \left[-\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{Q}}_{21}\hat{\mathbf{Q}}_{11}^{-1} \quad \hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\right]\frac{1}{N}\mathbf{X}^\mathsf{T}\boldsymbol{\epsilon}\right) \\
=& \boldsymbol{\beta}_1 + \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\boldsymbol{\beta}_2 + \hat{\mathbf{Q}}_{11}^{-1}\frac{\mathbf{X}_1^\mathsf{T}\boldsymbol{\epsilon}}{N} \\
& - \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\left(\hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\beta}_2 + \left[-\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{Q}}_{21}\hat{\mathbf{Q}}_{11}^{-1} \quad \hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\right]\frac{1}{N}\mathbf{X}^\mathsf{T}\boldsymbol{\epsilon}\right) \\
=& \boldsymbol{\beta}_1 + \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\left(\mathbf{I} - \hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\right)\boldsymbol{\beta}_2 \\
& + \left[\hat{\mathbf{Q}}_{11}^{-1} + \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{Q}}_{21}\hat{\mathbf{Q}}_{11}^{-1} \quad - \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\right]\frac{1}{N}\mathbf{X}^\mathsf{T}\boldsymbol{\epsilon}.
\end{aligned}$$

因此，

$$\begin{aligned}
\sqrt{N}(\hat{\boldsymbol{\beta}}_{1\text{WALS}} - \boldsymbol{\beta}_1) =& \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\left(\mathbf{I} - \hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\hat{\mathbf{P}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\right)\boldsymbol{\delta} \\
& + \left[\hat{\mathbf{Q}}_{11}^{-1} + \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{Q}}_{21}\hat{\mathbf{Q}}_{11}^{-1} \quad - \hat{\mathbf{Q}}_{11}^{-1}\hat{\mathbf{Q}}_{12}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^{-1/2}\tilde{\mathbf{W}}\hat{\mathbf{P}}^{-1/2}\hat{\boldsymbol{\Lambda}}\right]\frac{1}{\sqrt{N}}\mathbf{X}^\mathsf{T}\boldsymbol{\epsilon} \\
\xrightarrow{d} & \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\left(\mathbf{I} - \boldsymbol{\Lambda}\mathbf{P}^{-1/2}\tilde{\mathbf{W}}\mathbf{P}^{1/2^\mathsf{T}}\boldsymbol{\Lambda}^{-1}\right)\boldsymbol{\delta} \\
& + \left[\mathbf{Q}_{11}^{-1} + \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\boldsymbol{\Lambda}\mathbf{P}^{-1/2}\tilde{\mathbf{W}}\mathbf{P}^{-1/2^\mathsf{T}}\boldsymbol{\Lambda}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} \quad - \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\boldsymbol{\Lambda}\mathbf{P}^{-1/2}\tilde{\mathbf{W}}\mathbf{P}^{-1/2^\mathsf{T}}\boldsymbol{\Lambda}\right]\mathbf{R} \\
=& \boldsymbol{\Xi}\mathbf{C}\left(\mathbf{I} - \tilde{\mathbf{W}}\right)\mathbf{C}^{-1}\boldsymbol{\delta} + \left[\mathbf{Q}_{11}^{-1} + \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\boldsymbol{\Xi}^\mathsf{T} \quad - \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\right]\mathbf{R}.
\end{aligned}$$

□

Proof of Theorem 2 and Equation 14 在此证明中，我们推导聚焦参数的极限行为。由于聚焦参数是 $\boldsymbol{\beta}_1$ 的函数，并令 $\frac{\partial\mu}{\partial\boldsymbol{\beta}_1} = \mathbf{D}^{\boldsymbol{\beta}_1}$，因此很容易采用德尔塔法得到

$$\begin{aligned}
\sqrt{N}\left(\mu(\hat{\boldsymbol{\beta}}_{1\text{WALS}}) - \mu(\boldsymbol{\beta}_1)\right) \xrightarrow{d} & \mathbf{D}_{\boldsymbol{\beta}_1}^\mathsf{T}\boldsymbol{\Xi}\mathbf{C}\left(\mathbf{I} - \tilde{\mathbf{W}}\right)\mathbf{C}^{-1}\boldsymbol{\delta} \\
& + \mathbf{D}_{\boldsymbol{\beta}_1}^\mathsf{T}\left[\mathbf{Q}_{11}^{-1} + \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\boldsymbol{\Xi}^\mathsf{T} \quad - \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\right]\mathbf{R} \equiv R_\mu.
\end{aligned}$$

根据上述结果，可以看出 R_μ 的均值满足

$$\mathbb{E}(R_\mu) = \mathbf{D}_{\boldsymbol{\beta}_1}^\mathsf{T}\boldsymbol{\Xi}\mathbf{C}\left(\mathbf{I} - \tilde{\mathbf{W}}\right)\mathbf{C}^{-1}\boldsymbol{\delta}.$$

根据假设 3，$\mathbb{E}(\mathbf{R}) = \mathbf{0}$成立。通过适当的分块，可以立即推导出方差，其中 $\boldsymbol{\Omega} = \begin{bmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{bmatrix}$。

$$\begin{aligned}
\text{Var}(R_\mu) =& \mathbf{D}_{\boldsymbol{\beta}_1}^\mathsf{T}\left[\mathbf{Q}_{11}^{-1} + \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\boldsymbol{\Xi}^\mathsf{T} \quad - \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\right]\boldsymbol{\Omega}\left[\mathbf{Q}_{11}^{-1} + \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\boldsymbol{\Xi}^\mathsf{T} \quad - \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\right]^\mathsf{T}\mathbf{D}_{\boldsymbol{\beta}_1} \\
=& \mathbf{D}_{\boldsymbol{\beta}_1}^\mathsf{T}\left[\mathbf{Q}_{11}^{-1} + \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\boldsymbol{\Xi}^\mathsf{T} \quad - \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\right]\begin{bmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{bmatrix}\left[\mathbf{Q}_{11}^{-1} + \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\boldsymbol{\Xi}^\mathsf{T} \quad - \boldsymbol{\Xi}\mathbf{C}\tilde{\mathbf{W}}\mathbf{C}^\mathsf{T}\right]^\mathsf{T}\mathbf{D}_{\boldsymbol{\beta}_1}
\end{aligned}$$

$=\mathbf{D}_{\beta_1}^{\mathsf{T}}(\mathbf{Q}_{11}^{-1}+\Xi\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}})\Omega_{11}(\mathbf{Q}_{11}^{-1}+\Xi\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}})\mathbf{D}\beta_1-\mathbf{D}_{\beta_1}^{\mathsf{T}}(\mathbf{Q}_{11}^{-1}+\Xi\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}})\Omega_{12}$
$\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}}\mathbf{D}\beta_1-\mathbf{D}_{\beta_1}^{\mathsf{T}}\Xi\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}^{\mathsf{T}}\Omega_{12}(\mathbf{Q}_{11}^{-1}+\Xi\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}})\mathbf{D}\beta_1+\mathbf{D}_{\beta_1}^{\mathsf{T}}\Xi\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}^{\mathsf{T}}\Omega_{22}\mathbf{C}\ \tilde{}\mathbf{W}\mathbf{C}$
$^{\mathsf{T}}\Xi^{\mathsf{T}}\mathbf{D}\beta_1=\mathbf{D}_{\beta_1}^{\mathsf{T}}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\beta_1}+\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}}\Omega_{11}\Xi\mathbf{C}\mathbf{V}\ \tilde{}\mathbf{w}\ -\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}}\Omega_{12}\mathbf{C}\mathbf{V}\ \tilde{}\mathbf{w}\ -\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{\mathsf{T}}\Omega_{21}\Xi$
$\mathbf{C}\mathbf{V}\ \tilde{}\mathbf{w}+\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{\mathsf{T}}\Omega_{22}\mathbf{C}\mathbf{V}\ \tilde{}\mathbf{w}\ +2\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}}\Omega_{11}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\beta_1}-2\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{\mathsf{T}}\Omega_{21}\mathbf{Q}_{11}^{-1}\mathbf{D}\beta_1=\mathbf{D}_{\beta_1}^{\mathsf{T}}\mathbf{Q}_{11}^{-1}$
$\mathbf{D}_{\beta_1}+\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{B}\Omega\mathbf{B}^{\mathsf{T}}\mathbf{V}\ \tilde{}\mathbf{w}\ +2\ \tilde{}\mathbf{w}^{\mathsf{T}}\mathbf{V}\mathbf{B}\Omega\mathbf{H}\mathbf{Q}_{11}^{-1}\mathbf{D}\beta_1$，其中 $\tilde{}\mathbf{w}\ =(\tilde{w}_1,\ldots,\tilde{w}_{k_2})^{\mathsf{T}}$，$\mathbf{V}\ =$
$\mathbf{Diag}(\mathbf{D}_{\beta_1}^{\mathsf{T}}\Xi\mathbf{C})\ =\mathbf{Diag}(\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}}\mathbf{D}\beta_1)$，$\mathbf{B}\ =\ \left[-\mathbf{C}^{\mathsf{T}}\Xi^{\mathsf{T}}\ \mathbf{C}^{\mathsf{T}}\right]$，$\mathbf{H}=[\mathbf{I}\quad\mathbf{0}]^{\mathsf{T}}$。

结合 R_μ 的均值和方差，我们可以通过计算R_μ 的均值平方与方差之和，得到渐近均方误差:

$$\begin{aligned}\mathrm{AMSE}(\mu(\hat{\boldsymbol{\beta}}_{1\mathrm{WALS}})) =&\mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}}\boldsymbol{\Xi}\boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\beta}_1}+\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}\mathbf{C}\mathbf{V}\tilde{\mathbf{w}}-2\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{C}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\beta}_1}\\&+\mathbf{D}_{\boldsymbol{\beta}_1}^{\mathsf{T}}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\boldsymbol{\beta}_1}+\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{B}\Omega\mathbf{B}^{\mathsf{T}}\mathbf{V}\tilde{\mathbf{w}}+2\tilde{\mathbf{w}}^{\mathsf{T}}\mathbf{V}\mathbf{B}\Omega\mathbf{H}\mathbf{Q}_{11}^{-1}\mathbf{D}_{\boldsymbol{\beta}_1}.\end{aligned}$$

□

附录 B

本附录详细介绍正文所讨论的四种先验对应的密度函数 $\pi(\eta)$，以及由此得到的收缩权重 $\omega(t)$的计算过程。按照标准 WALS框架，拉普拉斯先验定义为 $\pi(\eta)\ =\ \frac{c}{2}\exp($ $-c|\eta|)$，其中 $c\ =\ \ln(2)$。由于它可以得到闭式解，其对应的收缩权重可解析地表示为 $\omega(t)\ =\ 1\ -\ \frac{c}{t}h(t)$，其中 $h(t)\ =\exp(-ct)\Phi(t-c)-\exp(ct)\Phi(-t-c)\exp(-ct)\Phi(t-c)+$ $\exp(ct)\Phi(-t-c)$，而 $\Phi(\cdot)$是标准正态累积分布函数。

对于重尾先验，密度函数及参数选择规定如下。柯西先验无需额外参数，其密度为$\pi(\eta)=\ \pi^{-1}(1+\eta^2)$。对于帕累托先验，我们采用文献中的首选取值，将形状参数设为 $a\ =\ \cancel{0.0862}$、尺度参数设为 $c\ =\ 0.0676$，其密度为 $\pi(\eta)\ =\ {}^{c}(1-a)\ 2a\ (1+c|\eta|)$ $^{-1/a}$。同样，为确保威布尔先验具有足够的尾部厚度与平滑性，我们在其密度$\pi(\eta)$ $=\ \frac{bc}{2}|\eta|^{b-1}\exp(-c|\eta|^{b})$中设定 $b=0.8876$ 和 $c=\ln(2)$。

与拉普拉斯先验不同，将这些重尾先验与正态似然 $\phi(t-\eta)$ 结合，并不能得到闭式后验均值。因此，我们使用高斯求积计算后验均值 $m(t)=\int_{-\infty}^{\infty}\eta\phi(t-\eta)\pi(\eta)$ $d\eta/\int_{-\infty}^{\infty}\phi(t-\eta)\pi(\eta)d\eta$。我们将积分区间截断至 $[-20,20]$，该区间涵盖了绝大部分概率质量，并确保计算稳定性。由此得到的

权重随后计算为 $\omega(t) = m(t)/t$。最后，为避免原点 ($t \to 0$) 附近出现数值除零错误，我们在实现中对 t 施加严格为正的下界（例如，10^{-10}）。

附录 C

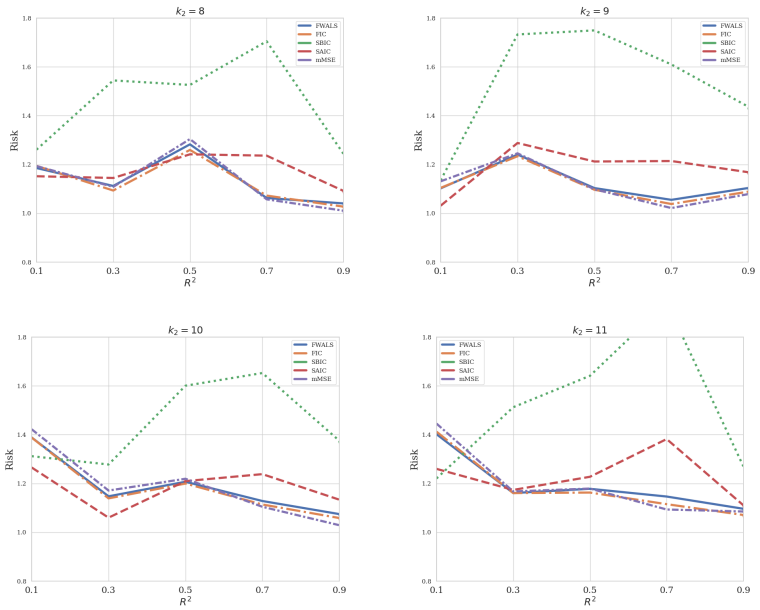

图 11 不同 k_2s（$N = 100$，$\tau = 0.3$）的风险

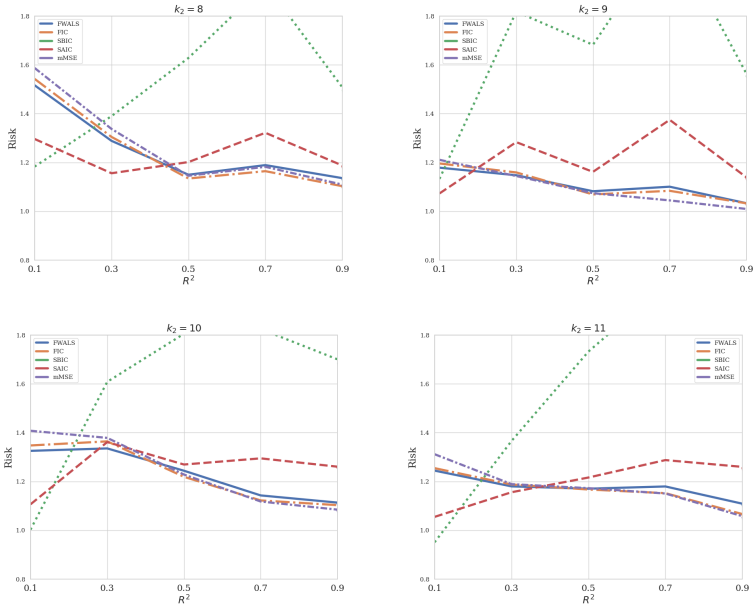

图12 不同 k_2的风险（$N = 100$、$\tau = 0.5$）

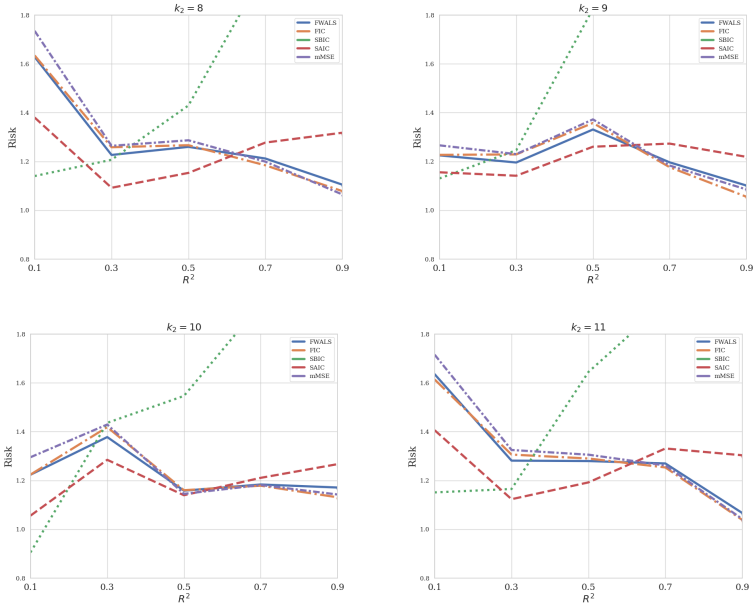

图13 不同 k_2的风险（$N = 100$、$\tau = 0.7$）