arXiv:2601.04101v1 [econ.EM] 7 Jan 2026

# 高维双向固定效应回归的岭估计

**Junnan He** [1] **和 Jean-Marc Robin**[1]

[1]*Economics Department, Sciences Po, Paris, e-mail:* junnan.he@sciencespo.fr; jeanmarc.robin@sciencespo.fr

**摘要：**我们研究了稀疏二部网络下高维双向固定效应回归模型的岭估计量。我们建立了集中不等式，表明当岭参数随网络规模的对数增长时，估计固定效应向量的偏差和方差-协方差矩阵收敛于仅取决于期望网络的确定性等价量。我们提供了模拟结果，并运用工人与企业匹配的工资行政数据进行了一项应用研究。

**关键词和短语：**岭回归、双向固定效应回归、大维度、二部网络、稀疏性。

## 1. 引言

本文研究如下双向固定效应模型：

$$y_{ijs} = \mu_i + \phi_j + u_{ijs}, i \in \{1, ..., n\}, j \in \{1, ..., p\}, s \in \{0, ..., d_{ij}\},$$

当观测到的匹配（$i, j$）具有稀疏二部图结构时。在我们的贯穿示例中，$y_{ijt}$ 是工人 $i$ 在企业 $j$ 受雇期间，第 $s$次出现该特定匹配时的年化收入对数，或者是对观测协变量进行初步回归后的残差（Abowd 等（1999））。在大多数应用中，工人面板的观测期仅覆盖有限的年数（5 年或 7 年）。因此，企业与工人的匹配网络是稀疏的，因为在观测期内，工人接触的企业数量较少（即大多数 $j$ 的 $d_{ij} = 0$ ）。每位工人的工资观测值数量（$d_{i\cdot} = \sum_j d_{ij}$）大致与观测年数相当；有时因失业而较少，有时则因年内工作流动而较多。类似的双向固定效应结构还会出现在各种情境中（见 Larremore 等（2014））。例如，若 $i$ 表示作者，$j$ 表示期刊（或领域），那么当作者在固定时间段内发表的论文数量较少时，可以将引用次数等论文层面结果建模为作者效应与发表渠道效应之和。同样，若 $i$ 表示传粉者物种，$j$ 表示植物物种，则可以将坐果率（或结籽率）等交互层面测度建模为传粉者“有效性”成分与植物“接受性”成分之和。为便于说明，下文将 $y$ 称为工资，将 $i$ 称为工人，将 $j$ 称为企业。

我们感兴趣的参数是固定效应分布。显然，在度数固定或有界的情况下，任何固定效应都无法被一致识别，但它们的分布又如何呢?

当网络规模增大时，我们能否对估计固定效应的分布极限得出一些结论？显然，这个问题的答案应取决于估计量。

如果网络是稀疏的，固定效应模型的OLS估计量会面临弱识别问题。Jochmans和Weidner (2019)表明，OLS估计量的方差取决于拉普拉斯矩阵第二小特征值的大小，而该特征值决定网络的连通性。Cheeger不等式（Chung(1997)对拉普拉斯矩阵的第二特征值给出了界，而Cheeger常数则衡量“瓶颈”的存在。Zhang和Rohe (2018)进一步表明，稀疏的非齐次图（边以节点依赖概率独立生成）会产生数量随节点数增加的多个瓶颈（“悬挂节点”）。拉普拉斯矩阵中非常小的特征值数量也会相应增加。

这一问题在社区检测文献中得到了充分研究，其目标是寻找彼此连接概率较高的节点子集。对于稀疏图，节点聚类通常会得到一个划分，其中一个大型聚类包含大部分数据，另有许多小型聚类。Chaudhuri等人 (2012)提出通过添加许多“弱连接”来降低图的稀疏程度，从而正则化图。具体而言，这相当于将邻接矩阵 $A$替换为 $A+\lambda \mathbf{1}_n \mathbf{1}_n^\top$，其中 $\lambda$较小。此后，人们又提出了其他几种正则化图的方法。Qin和Rohe (2013)提出的方案与我们尤为相关，因为它与岭回归估计量有关。他们只在度矩阵上加一个小数值，而不是加到整个邻接矩阵上。Dall'Amico等人 (2021)的近期研究表明，Qin和Rohe提出的特定正则化形式（以及Dall'Amico等人提出的等价形式）优于其他竞争算法。本文将以统计网络文献中发展起来的渐近理论为基础，进一步认识具有稀疏二部网络结构的双向固定效应模型中岭估计量的渐近性质。

底层网络的三个特征对于将其纳入本分析至关重要。首先，网络规模庞大（包含许多工人和企业），因此我们试图估计的固定效应数量也很多。其次，网络是稀疏的，因为在所有实际情形中，观测到的工人与企业匹配仅占所有可能匹配的一小部分。第三，渐近理论将假设工人 $n$ 和企业 $p$的数量很大，但 $p/n$ 基本固定且独立于 $n$。这意味着，工人节点和企业节点的度或多或少是固定的，且独立于网络规模。

高维设定下的岭回归是统计学文献中一个得到广泛研究的问题。相关文献关注如下回归模型 $y_i = x_i^\top \beta + u_i$，其中 $n$ 个观测值相互独立，回归变量的个数 $p$ 随样本量增长，且 $p/n \to \gamma$。假设回归变量向量是球对称的，即 $x_i = \Sigma^{1/2} z_i$ 其中 $z_i$ 具有独立元素。参数向量 $\beta$是均值为 0、方差为 $p^{-1}\alpha^2 I_p$的随机向量。于是，岭回归估计量为 $\widehat{\beta} = \left(X^\top X + \lambda I_p\right)^{-1} X^\top Y$，其中 $Y = (y_i)$，且 $X^\top = (x_1 ... x_n)$。在这些假设下，Dobriban和Wager (2018)表明，期望预测风险

几乎必然收敛到一个确定性极限，该极限在最优参数 $\lambda^* = n\gamma\alpha^{-2}\sigma^2$处取得最小值，其中 $\sigma^2$ 是残差方差。[1] 这一设定与我们的设定大不相同。在我们的设定中， $\beta$ 的元素不会趋于零。与此同时，线性形式 $x_i^\top\beta$ 不会因稀疏性而发散: $x_i$ 中的非零元素个数随 $n$ 增长，且 $p$。

我们需要为底层网络建立一个模型。我们假设采用度修正随机块模型（Karrerand Newman（2011）；Larremore 等人（2014），该模型已成为社区检测文献中的标准模型。具体而言，我们首先假设工人和企业的集合可以划分为 $K$ 组（社区）。随后， $d_{ij}$ 之间的连接数 $i$ 和 $j$ 从二项分布（$T$， $p_{ij}$）中独立抽取，其中 $p_{ij}$ 是发生连接的概率。如果 $i$ 和 $j$ 属于同一社区，该概率就更高。[2]

对于任意形成的工人与企业连接网络，我们随后生成一个具有同方差残差的双向固定效应模型。令 $\beta = (\mu^\top, \phi^\top)^\top$ 表示$\widehat{}$参数向量。我们考虑岭估计量 $\beta$，其中 $X = (W, F)$，这里 $W$ 是工人虚拟变量的矩阵， $F$ 是企业虚拟变量的矩阵，并使用两个不同的正则化参数，分别用于 $\mu$ 和 $\phi$。我们的主要理论贡献是推导出岭估计量一阶和二阶误差的确定性等价物。在我们的随机图模型下，如果岭惩罚项$\widehat{}$的阶为 $\ln(n+p)$，那么高维偏差向量 $\mathbb{E}[\beta - \beta]$ 及其方差矩阵在算子矩阵范数意义下接近确定性极限。这些确定性等价物根据期望网络计算得出，在该网络中，每个节点具有期望度，且邻接矩阵为期望邻接矩阵。这有助于解释为何固定效应的总体分布可以得到准确近似，尽管无法一致估计每个固定效应，因为每个工人和企业的观测值数量并不会随网络规模增加而增加。利用随机矩阵理论进一步研究这些确定性等价物超出了本文的范围。

我们开展了一系列蒙特卡洛实验，结果表明，与普通最小二乘法相比，岭正则化显著改善了固定效应分布的恢复效果。此外，$\ln(n+p)$ 缩放是一个充分条件；我们的模拟表明，使用预测误差（交叉验证）来调整岭惩罚项，往往会得到较小的岭参数。遵循实证文献的做法，我们将结果变量的方差分解为工人效应方差、企业效应方差、工人效应与企业效应之间的协方差项，以及残差方差。当网络非常稀疏时，同方差偏差校正会失效，[3] 而岭估计量则能得出更准确的方差分解——

[1]其他参考文献参见 Dicker（2013）；El Karoui（2018）；Richards 等（2020）；Wu 和 Xu（2020）；Bigot 等（2024）；Patil 等（2025）。[2]另请参见 Nimczik（2017）；Abowd 等（2019）；Bonhomme 等（2019）；Lentz 等（2023）。这些论文通常允许工作流动服从马尔可夫过程，但 $d_{ij}$ 与 $d_{ij'}$ 之间存在少量依赖不应影响网络的总体随机结构。[3]参见 Andrews 等（2008）、Gaure（2014）、Kline 等（2020）以及 Azkarate-Askasua 和 Zerecero（2022）。

。我们还报告了对实际雇主—雇员匹配数据（法国 DADS 面板）的估计结果。基于最大连通分量的普通最小二乘法估计所得到的对数工资方差分解显示，固定效应协方差的贡献为负。我们发现，同方差偏差校正的影响很小。然而，岭估计量得出的方差分解更为合理。工人效应在对数工资总方差中所占的比例更为合理，且协方差项为正而非负。我们还使用最大且最强的连通分量进行估计，这使得我们能够采用 Kline 等人 (2020)提出的留一法（LOO）偏差校正方法。普通最小二乘法和同方差偏差校正的效果有所改善，但幅度很小。岭估计量的结果变化不大，而留一法偏差校正的结果接近岭估计量。

本文的结构如下。在下一节中，我们将介绍岭估计量，并将其与工人和企业之间连接关系的底层图联系起来。第 3 节为该网络构建随机块模型。第 4 节（网络）和第 5 节（岭估计量）将发展渐近理论。证明见附录。

## 2. 双向固定效应模型

双向固定效应模型可以写成矩阵形式：

$$Y = W\mu + F\phi + U = X\beta + U, \quad X = (W, F), \quad \beta = (\mu^\top, \phi^\top)^\top.$$

向量 $Y = (y_{ijs}) \in \mathbb{R}^N$ （其中 $N = \sum_{i,\ j} d_{ij}$）是连续结果的向量。矩阵 $W \in \{0,1\}^{N\times n}$ 将结果观测值与工人标识符 $i \in \{1, ..., n\}$ 相关联，矩阵 $F \in \{0,1\}^{N\times p}$ 将观测值与企业标识符 $j \in \{1, ..., p\}$相关联。我们假设 $U$是均值为 0、方差为 $\sigma^2 I_n$ 的独立同分布变量。

### *2.1.* 普通最小二乘法与二部图解释

OLS估计量 $\widehat{\beta} = (\widehat{\mu}^\top,\ \widehat{\phi}^\top)^\top$ 满足正规方程 $X^\top X\widehat{\beta} = X^\top Y$，其中

$$X^\top X = \begin{pmatrix} W^\top W & W^\top F \\ F^\top W & F^\top F \end{pmatrix} := \begin{pmatrix} D_w & B \\ B^\top & D_f \end{pmatrix},$$

and

$$\det\left(X^\top X\right) = \det\left(D_w\right)\det\left(D_f - B^\top D_w^{-1} B\right) = \det\left(D_f\right)\det\left(D_w - BD_f^{-1}B^\top\right).$$

$B = W^\top F = (d_{ij})$ 的每个元素表示每个匹配的工资观测值数量（$i,\ j$）。它是连接工人与企业的二部图的加权邻接矩阵。我们关注 $B$ 稀疏的情形：对于大多数潜在连接，$d_{ij} = 0$ 。我们还拥有

$$D_w = W^\top W = \mathrm{diag}(d_{i\cdot}), \quad D_f = F^\top F = \mathrm{diag}(d_{\cdot j}).$$

这些是工人和企业的度矩阵： $d_{i\cdot} = \sum_j d_{ij}$ 表示工人 $i$ 的观测值数量，$d_{\cdot j} = \sum_i d_{ij}$ 表示企业 $j$ 的观测值数量。我们假设工人度和企业度均为正，从而确保 det（$D_w$） $\neq 0$ 以及 det（$D_f$） $\neq 0$。

假设没有工人或企业的度为零，则矩阵

$$\widetilde{A}_f = B^\top D_w^{-1} B = \left( \sum_i \frac{d_{ij} d_{ij'}}{d_{i\cdot}} \right)_{j,j'}.$$

是一个加权邻接矩阵，表示通过共同雇员连接企业的无向图，也称为原始二部图的加权单模投影（参见Zhou 等人，(2007年)）。如果对于大多数企业对（$j$， $j'$），在观测窗口内没有工人在两家企业都被观测到，那么该矩阵就是稀疏的。实际上，对于通常的观测期长度（5 年或 7 年），情况通常如此。矩阵 $\widetilde{L}_f$：$= D_f - B^\top D_w^{-1} B = F^\top M_W F$，其中$M_W = I_N - W(W^\top W)^{-1} W^\top$，是相应的拉普拉斯矩阵。

类似的解释也适用于矩阵

$$\widetilde{A}_w = B D_f^{-1} B^\top = \left( \sum_j \frac{d_{ij} d_{i'j}}{d_{\cdot j}} \right)_{i,i'}.$$

它是一个邻接矩阵，表示通过共同雇主连接工人的无向加权图。矩阵 $\widetilde{L}_w$：$= D_w - B D_f^{-1} B^\top = W^\top M_f W$是工人拉普拉斯矩阵。

下文将使用这些邻接矩阵的以下归一化对称版本：

$$A_w = \left( \sum_j \frac{d_{ij} d_{i'j}}{\sqrt{d_{i\cdot} d_{i'\cdot} d_{\cdot j}}} \right)_{i,i'} = EE^\top, \quad A_f = \left( \sum_i \frac{d_{ij} d_{ij'}}{\sqrt{d_{\cdot j} d_{\cdot j'} d_{i\cdot}}} \right)_{j,j'} = E^\top E,$$

with

$$E = D_w^{-1/2} B D_f^{-1/2} = \left( \frac{d_{ij}}{\sqrt{d_{i\cdot} d_{\cdot j}}} \right)_{i,j}.$$

我们还定义拉普拉斯矩阵的归一化版本： $L_w = I_n - A_w$ 和$L_f = I_p - A_f$。归一化邻接矩阵和拉普拉斯矩阵满足以下性质。

**引理 2.1**。 $A_f$ （记作 $\alpha_p \leq \ldots \leq \alpha_1$）的特征值以及 $L_f$ （$\lambda_1 = 1 - \alpha_1 \leq \ldots \leq \lambda_p = 1 - \alpha_p$）的特征值均属于 $[0, 1]$。此外， $\lambda_1 = 0$ 和 $\alpha_1 = 1$。

证明。 见附录。除非另有说明，所有证明均在附录中给出。

□

引理 2.1将拉普拉斯矩阵的特征值限制在 0 和 1 之间。它还指出，无论图的连通性如何，0 始终是一个

拉普拉斯矩阵的特征值。这是因为 $W$ 和 $F$ 的列和为 1。实际应用中，这意味着双向固定效应回归要求对固定效应 $\mu$ 和 $\phi$进行归一化。下一个引理将特征值 0 的重数与图的连通分量数量联系起来。

**引理 2.2**。 *(i)* 拉普拉斯矩阵的不同形式具有完全相同的特征值。*(ii)* 非连通分量的数量等于拉普拉斯矩阵的特征值 *0* 或邻接矩阵的特征值 *1* 的重数。

证明。 例如参见 vonLuxburg (2007年)。 □

要计算 OLS估计量，每个连通分量都需要进行一次参数归一化。为了估计双向固定效应模型，必须首先找到连通分量，例如使用 Tarjan算法。估计通常在最大的连通分量上进行。

### *2.2* 岭估计量

在实践中，最大的连通分量可能是弱连通的。也就是说，拉普拉斯矩阵的第二小特征值可能接近于零。这是一个弱识别问题。以下岭估计量是一个自然的解决方案（Hastie(2020)）：

$$\widehat{\beta} = \begin{pmatrix} \widehat{\mu} \\ \widehat{\phi} \end{pmatrix} = \begin{pmatrix} D_w + \lambda_w I_n & B \\ B^\top & D_f + \lambda_f I_p \end{pmatrix}^{-1} \begin{pmatrix} W^\top Y \\ F^\top Y \end{pmatrix}$$

其中 $\lambda_w$，$\lambda_f$ 为非负参数。记 $D_{w,\lambda} = D_w + \lambda_w I_n$ 和$D_{f,\lambda} = D_f + \lambda_f I_p$。当且仅当以下条件成立时，该估计量是可行的:

$$\det \begin{pmatrix} D_{w,\lambda} & B \\ B^\top & D_{f,\lambda} \end{pmatrix} = \det\left(D_{w,\lambda}\right) \det\left(\widetilde{L}_{f,\lambda}\right) = \det\left(D_{f,\lambda}\right) \det\left(\widetilde{L}_{w,\lambda}\right) \neq 0,$$

其中 $\widetilde{L}_{w,\lambda} := D_{w,\lambda} - BD_{f,\lambda}^{-1} B^\top$ 和 $\widetilde{L}_{f,\lambda} := D_{f,\lambda} - B^\top D_{w,\lambda}^{-1} B$ 为正则化拉普拉斯矩阵。在这种情况下，分块求逆得到

$$\begin{pmatrix} \widehat{\mu} \\ \widehat{\phi} \end{pmatrix} = \begin{pmatrix} \widetilde{L}_{w,\lambda}^{-1}\left(W^\top Y - BD_{f,\lambda}^{-1} F^\top Y\right) \\ \widetilde{L}_{f,\lambda}^{-1}\left(F^\top Y - B^\top D_{w,\lambda}^{-1} W^\top Y\right) \end{pmatrix}.$$

令 $L_{w,\lambda} = I_n - D_{w,\lambda}^{-1/2} BD_{f,\lambda}^{-1} B^\top D_{w,\lambda}^{-1/2}$和 $L_{f,\lambda} = I_p - D_{f,\lambda}^{-1/2} B^\top D_{w,\lambda}^{-1} BD_{f,\lambda}^{-1/2}$表示正则化拉普拉斯矩阵的归一化版本。那么，$\widetilde{L}_{w,\lambda} = D_{w,\lambda}^{1/2} L_{w,\lambda} D_{w,\lambda}^{1/2}$，和 $\widetilde{L}_{f,\lambda} = D_{f,\lambda}^{1/2} L_{f,\lambda} D_{f,\lambda}^{1/2}$。

以下引理表明，岭估计量始终是可行的（即使不对固定效应参数进行归一化），因为它证明了正则化拉普拉斯矩阵的特征值均与零保持有界距离。令 $E_\lambda = D_{w,\lambda}^{-1/2} BD_{f,\lambda}^{-1/2}$，正则化企业拉普拉斯矩阵可写为 $L_{f,\lambda} = I_p -$

$E_\lambda^\top E_\lambda$，正则化工人拉普拉斯矩阵可写为 $L_{w,\ \lambda} = I_n - E_\lambda E_\lambda^\top$。下文中，矩阵范数均指谱矩阵范数（对于任意矩阵 $A$，$\|A\| = \sqrt{\text{eigmax}(A^\top A)}$）。

**引理 2.3**。如果相应的岭参数为正（$\lambda_f > 0$），则企业拉普拉斯矩阵的最小特征值严格为正；具体而言：$\text{eigmin}(L_{f,\ \lambda}) \geq \lambda_f \max_j d_{\cdot j} + \lambda_f > 0$。由此可得 $\|E_\lambda\| = \sqrt{\text{eigmax}(E_\lambda^\top E_\lambda)} \leq \sqrt{\max_j d_{\cdot j} \max_j d_{\cdot j} + \lambda_f} < 1$。$\lambda_w$ 的情形类似。

### *2.3.* 样本内偏差与方差

假设 $U$ 与 $X = (W,\ F)$ 独立，均值为 0，且 $\mathbb{V}U = \sigma^2 I_N$。那么，岭估计量存在偏差，其偏差为

$$\mathbb{E}\left(\widehat{\mu} - \mu \mid X, \beta\right) = \widetilde{L}_{w,\lambda}^{-1}\left(-\lambda_w \mu + \lambda_f B D_{f,\lambda}^{-1}\phi\right),$$
$$\mathbb{E}\left(\widehat{\phi} - \phi \mid X, \beta\right) = \widetilde{L}_{f,\lambda}^{-1}\left(\lambda_w B^\top D_{w,\lambda}^{-1}\mu - \lambda_f \phi\right).$$

$\widehat{\ }(\widehat{\mu},\ \phi)$ 的方差为

$$\mathbb{V}\left(\widehat{\mu} \mid X, \beta\right) = \sigma^2 \widetilde{L}_{w,\lambda}^{-1} - \sigma^2 \widetilde{L}_{w,\lambda}^{-1}\left[\lambda_w I_n + \lambda_f B D_{f,\lambda}^{-2} B^\top\right]\widetilde{L}_{w,\lambda}^{-1},$$
$$\mathbb{V}\left(\widehat{\phi} \mid X, \beta\right) = \sigma^2 \widetilde{L}_{f,\lambda}^{-1} - \sigma^2 \widetilde{L}_{f,\lambda}^{-1}\left[\lambda_w B^\top D_{w,\lambda}^{-2} B + \lambda_f I_p\right]\widetilde{L}_{f,\lambda}^{-1},$$
$$\text{Cov}\left(\widehat{\mu}, \widehat{\phi} \mid X, \beta\right) = -\sigma^2 \widetilde{L}_{w,\lambda}^{-1} B D_{f,\lambda}^{-1} + \sigma^2 \widetilde{L}_{w,\lambda}^{-1}\left(\lambda_w D_{w,\lambda}^{-1} B + \lambda_f B D_{f,\lambda}^{-1}\right)\widetilde{L}_{f,\lambda}^{-1}.$$

## 3. 随机块模型

为了研究这些估计量的统计性质，我们需要对工人与企业匹配构成的图建立一个模型。我们将二部图建模为度修正随机块模型（Dasgupta 等人 (2004)、Karrer 和 Newman(2011)、Larremore 等人 (2014)、Razaee 等人 (2019)）。

**节点类型** 我们首先将任意数量的 $n$工人和 $p$企业划分为$K$组。令 $k \in \{1, ..., K\}$ 为工人类型的索引，$\ell \in \{1, ..., K\}$ 为企业类型的索引。为简化起见，我们假设组数相同。工人最初以概率 $\pi_w(k)$独立抽取类型 $k_i = k$，企业则以概率$\pi_f(\ell)$抽取类型 $\ell_j = \ell$。令 $n_k$ 表示类型为$k$的工人数。其次，我们引入参数 $\theta_j$ 来控制企业 $j$的期望度，并要求其在企业组内的总和为 1：$\sum_{j'} \theta_{j'} \delta_{\ell_{j'}, \ell} = 1$。因此，参数 $\theta_j$ 可解释为：在已从组 $\ell$ 中抽取任意一家企业的条件下，抽取该组企业 $j$ 的概率。对于工人，我们假设在给定类型的情况下均匀抽样（即概率为 $1/n_{k_i}$）。我们还可以引入另一个固定效应 $\theta$，以控制给定类型下的工人抽样。通过对一种节点类型进行度修正而不对另一种进行度修正，我们涵盖了标准随机块模型及其度修正扩展。

**网络** 网络的块结构由亲和矩阵$C$决定，其中 $C(k,\ \ell) \geq 0$。对于每个工人与企业配对（$i,\ j$），我们独立地在 $i$ 和 $j$ 之间抽取$d_{ij}$ 条边，边出现的概率服从伯努利分布，概率为 $p_{ij} = \frac{1}{n_{k_i}} \theta_j C(k_i,\ \ell_j)$。类型为 $k$ 的工人与类型为 $\ell$ 的企业之间的连接期望数量为

$$\mathbb{E}\left(\sum_{i,j} d_{ij}\delta_{k_i k}\delta_{\ell_j \ell}\right) = \sum_{i,j} p_{ij}\delta_{k_i k}\delta_{\ell_j \ell} = C(k,\ell),$$

即 $\sum_k \delta_{k_i k} = n_{k_i}$ 和 $\sum_j \delta_{\ell_j \ell}\theta_j = 1$。预期企业度为

$$\mathbb{E}\left(d_{\cdot j}\right) = \mathbb{E}\left(\sum_i d_{ij}\right) = \sum_i p_{ij} = \sum_i \frac{1}{n_{k_i}}\theta_j C(k_i,\ell_j) = \theta_j C(\cdot,\ell_j),$$

其中，$C(\cdot,\ \ell) = \sum_k C(k,\ \ell)$，预期工人度为

$$\mathbb{E}\left(d_{i\cdot}\right) = \mathbb{E}\left(\sum_j d_{ij}\right) = \sum_j p_{ij} = \frac{1}{n^0_{k_i}} \sum_\ell \sum_j \theta_j \delta_{\ell_j \ell} C(k_i,\ell) = \frac{1}{n_{k_i}} C(k_i,\cdot),$$

其中，$C(k,\ \cdot) = \sum_\ell C(k,\ \ell)$。

例如，在模拟中我们使用

$$C = c\frac{p}{K}\left[I_K + \delta(J_K - I_K)\right],$$

其中，$c$ 是一个常数，$J_K = \mathbf{1}_K \mathbf{1}_K^\top$ 和 $\delta$ 是调节参数：$\delta = 0$ 表示完全隔离，$\delta = 1$ 表示无差异。在这一设定下，每对之间的连接数（$k,\ \ell$）与企业数量成正比。如果工人均匀地分配到各组中，则 $n_k \simeq n/K$，且平均工人度与 $p/n = \gamma$ 成正比。因此，当工人和企业数量以相同速度增长时，它保持不变，不会随 $p$ 和 $n$ 增长。

实际生成的图通常包含一个主要的连通分量，其中的观测值数量比其他任何分量都多出几个数量级。出于识别原因，通常将其用作估计样本。此时，$n$ 和 $p$表示估计样本中的工人和企业数量，$n_k$ 表示样本中类型为 $k$ 的工人数。

## 4. 网络渐近理论

我们为包含众多企业和工人的大型网络建立渐近理论。工人和企业的类型均已给定，参数 $\theta_j$ 和 $C$也已给定。边权重 $d_{ij} =$服从 Bernoulli($p_{ij}$) 分布，其中 $p_{ij} = \frac{1}{n_{k_i}} \theta_j C(k_i,\ \ell_j)$ 是唯一的随机变量。以下分析不假设选取最大连通分量。

### *4.1.* 确定性等价网络

遵循Qin和Rohe （2013），我们用花体字体表示关于 $B = (d_{ij})$ 的期望。因此记作

$$\mathfrak{B} := \mathbb{E}\, B = (p_{ij}) = \left[\tfrac{1}{n_{k_i}}\theta_j C(k_i, \ell_j)\right] = \operatorname{diag}\left(\tfrac{1}{n_{k_i}}\right) Z_w C Z_f^\top \operatorname{diag}(\theta_j), \qquad (1)$$

其中，$Z_w = (\delta_{k_i k}) \in \{0,1\}^{n\times K}$和$Z_f = (\delta_{\ell_j \ell}) \in \{0,1\}^{p\times K}$是选择矩阵，包含有关工人和企业社群成员身份的信息。此外，

$$\mathfrak{B}^\top\mathfrak{B} = \left[\theta_j \sum\nolimits_k \frac{C(k,\ell_j)C(k,\ell_{j'})}{n_k}\theta_{j'}\right] = \operatorname{diag}(\theta_j)\, Z_f C^\top \operatorname{diag}\left(\tfrac{1}{n_k}\right) C Z_f^\top \operatorname{diag}(\theta_j).$$

使用归一化 $\sum_j \theta_j \delta_{\ell_j \ell} = 1$ 和记号 $C(k, \cdot) = \sum_\ell C(k, \ell)$，我们有

$$\mathfrak{D}_{w,\lambda} = \operatorname{diag}\left(\tfrac{1}{n_{k_i}} \sum\nolimits_j \theta_j C(k_i, \ell_j) + \lambda_w\right) = \operatorname{diag}\left(\tfrac{1}{n_{k_i}} C(k_i, \cdot) + \lambda_w\right).$$

并使用记号 $C(\cdot, \ell) = \sum_k C(k, \ell)$ 以及以下事实

$$\sum_i \tfrac{1}{n_{k_i}} C(k_i, \ell_j) = \sum_k \sum_i \frac{\delta_{k_i k}}{n_{k_i}} C(k, \ell_j) = \sum_k C(k, \ell_j),$$

我们有

$$\mathfrak{D}_{f,\lambda} = \operatorname{diag}\left(\theta_j \sum\nolimits_i \tfrac{1}{n_{k_i}} C(k_i, \ell_j) + \lambda_f\right) = \operatorname{diag}(\theta_j C(\cdot, \ell_j) + \lambda_f).$$

令 $\mathfrak{E}_\lambda = \mathfrak{D}_w^{-1/2}\, \mathfrak{B}\mathfrak{D}_f^{-1/2}$，$\mathfrak{E}_\lambda$ 的通用元素为

$$(\mathfrak{E}_\lambda)_{ij} = \frac{\theta_j C(k_i,\ell_j)/n_{k_i}}{\sqrt{C(k_i,\cdot)/n_{k_i}+\lambda_w}\sqrt{\theta_j C(\cdot,\ell_j)+\lambda_f}}$$
$$= \tfrac{1}{n_{k_i}} \sqrt{\frac{C(k_i,\cdot)}{\frac{1}{n_{k_i}}C(k_i,\cdot)+\lambda_w}} \frac{C(k_i,\ell_j)}{\sqrt{C(k_i,\cdot)C(\cdot,\ell_j)}} \theta_j \sqrt{\frac{C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f}} = \omega_{\lambda,k_i} \widetilde{C}(k_i,\ell_j)\phi_{\lambda,j},$$

表示 $\widetilde{C}(k,\ell) = \frac{C(k,\ell)}{\sqrt{C(k,\cdot)C(\cdot,\ell)}}$，$\omega_{\lambda,k} = \frac{1}{n_k\sqrt{\frac{1}{n_k}C(k,\cdot)+\lambda_w}}$ 以及 $\phi_{\lambda,j} = \theta_j\sqrt{\frac{C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f}}$。正则化降低了人均就业连接较少（$\frac{1}{n_k}C(k,\cdot)$ 较小）的工人社群以及每家企业员工人数较少（$\theta_j C(\cdot,\ell_j)$ 较小）的企业组所受到的噪声影响。

然后，定义正则化的邻接矩阵和拉普拉斯矩阵

$$\mathcal{A}_{f,\lambda} := \mathfrak{E}_\lambda^\top \mathfrak{E}_\lambda = \left[\phi_{\lambda,j}\left(\sum_k n_k \widetilde{C}(k,\ell_j)\omega_{\lambda,k}^2 \widetilde{C}(k,\ell_{j'})\right)\phi_{\lambda,j'}\right]_{j,j'}, \quad \mathfrak{L}_{f,\lambda} = I_p - \mathcal{A}_{f,\lambda},$$

$$\mathcal{A}_{w,\lambda} := \mathfrak{E}_\lambda \mathfrak{E}_\lambda^\top = \left[\omega_{\lambda,i}\left(\sum_j \widetilde{C}(k_i,\ell_j)\phi_{\lambda,j}^2 \widetilde{C}(k_{i'},\ell_j)\right)\omega_{\lambda,i'}\right]_{i,i'}, \quad \mathfrak{L}_{w,\lambda} = I_p - \mathcal{A}_{w,\lambda}.$$

下面的引理表明，矩阵 $\mathfrak{E}_\lambda$ 包含随机图的社区结构，并且模拟拉普拉斯矩阵 $\mathfrak{L}_{f,\lambda} = I_p - \mathfrak{E}_\lambda^\top \mathfrak{E}_\lambda$ 的特征值与 0 保持一定距离。

引理 *4.1*。*1)* $\mathfrak{E}_\lambda$、$\mathcal{A}_{f,\lambda}$ 和 $\mathcal{A}_{w,\lambda}$ 的秩为 $K$。*2)* 如果 $\lambda_f > 0$，则 $eigmin(\mathfrak{L}_{f,\lambda}) \geq \lambda_f \min_j \theta_j C(\cdot, \ell_j) + \lambda_f > 0$，并且 $\|\mathfrak{E}_\lambda\| \leq \sqrt{\max_j \theta_j C(\cdot, \ell_j) \max_j \theta_j C(\cdot, \ell_j) + \lambda_f} < 1$。*3)* 如果 $\lambda_w > 0$，则 $eigmin(\mathfrak{L}_{w,\lambda}) \geq \frac{\lambda_w \min_k \frac{1}{n_k} C(k, \cdot)}{+\lambda_w} > 0$，并且 $\|\mathfrak{E}_\lambda\| \leq \sqrt{\max_k \frac{1}{n_k} C(k, \cdot) \max_k \frac{1}{n_k} C(k, \cdot) + \lambda_w}} < 1$。

### *4.2.* 拉普拉斯矩阵的渐近界

用于推导渐近界的主要工具是随机矩阵的 Bernstein 不等式的一个版本（参见 Chung 和 Radcliffe（2011）以及 Tropp（2015），定理 6.1.1）。设 $S_1, \ldots, S_n$ 是相互独立、均值为零且维数相同的随机矩阵，$d_1 \times d_2$ 并假设每个矩阵都有一致的界：

$$\mathbb{E}\, S_k = 0 \quad \text{and} \quad \|S_k\| \leq L, \quad \forall k = 1, \ldots, n.$$

引入总和 $Z = \sum_{k=1}^n S_k$，并令 $v(Z)$ 表示该总和的矩阵方差统计量：

$$v(Z) = \max\left\{\left\|\mathbb{E}\left(ZZ^\top\right)\right\|, \left\|\mathbb{E}\left(Z^\top Z\right)\right\|\right\} = \max\left\{\left\|\sum_{k=1}^n \mathbb{E}\left(S_k S_k^\top\right)\right\|, \left\|\sum_{k=1}^n \mathbb{E}\left(S_k^\top S_k\right)\right\|\right\}.$$

那么，

$$\Pr\{\|Z\| \geq t\} \leq (d_1 + d_2) \exp\left(-\frac{1}{2} t^2 / \left(v(Z) + \frac{1}{3} L t\right)\right), \quad \forall t \geq 0.$$

此外，

$$\mathbb{E}\, Z \leq \sqrt{2 v(Z) \ln(d_1 + d_2)} + \frac{1}{3} L \ln(d_1 + d_2).$$

我们首先证明以下集中不等式，它将 Qin和Rohe (2013) 的定理 4.1 推广到二部图的正则化拉普拉斯矩阵。岭系数 $\lambda_w$ 和 $\lambda_f$ 使我们能够选择 $M_w \vee M_f := \max(M_w, M_f)$ 足够大，从而以任意指定的概率使楔形 $\|E_\lambda - \mathfrak{E}_\lambda\|$ 小于任意选定的值。

**定理 4.2**。令 $M_w = (\delta_w + \lambda_w)^{-1}$ 和 $M_f = \left(\delta_f + \lambda_f\right)^{-1}$，其中 $\delta_w = \min_k \frac{1}{n_k} C(k, \cdot)$ 和 $\delta_f = \min_j \theta_j C(\cdot, \ell_j)$ 分别为工人和企业的最小期望度。对于任意 $\epsilon > 0$，若 $M = M_f \vee M_w \leq \left(3 \ln \frac{n+p}{\epsilon}\right)^{-1}$，则以至少 $1 - \frac{3+4\gamma}{1+\gamma}\epsilon$ 的概率（其中 $\gamma = p/n$），

$$\|E_\lambda - \mathfrak{E}_\lambda\| \leq 4t, \quad \|L_{w,\lambda} - \mathfrak{L}_{w,\lambda}\| \leq 8t, \quad \|L_{f,\lambda} - \mathfrak{L}_{f,\lambda}\| \leq 8t,$$

其中 $t = \sqrt{3M \ln \frac{n+p}{\epsilon}} \leq 1$。

备注 1。例如，设 $\epsilon = (n+p)^{-\nu}$ 满足 $\nu \in (0,1)$。那么只需满足 $\lambda_w \wedge \lambda_f := \min(\lambda_w, \lambda_f) \geq 3(1+\nu)\ln(n+p)$，定理的条件便可成立，即使最小期望工人度 $\delta_w = \min_k \frac{1}{n_k} C(k,\cdot)$ 和最小期望企业度 $\delta_f = \min_j \theta_j C(\cdot, \ell_j)$ 要小得多。[4]特别地，选择 $\lambda_w \wedge \lambda_f \propto (\ln(n+p))^{1+\nu'}$ 并满足 $\nu' > 0$，即可推出拉普拉斯矩阵在谱范数意义下依概率收敛到其期望值（当 $t \to 0$ 时，如 $n+p \to \infty$）。

下一个定理表明，上述性质可传递到逆矩阵。

**定理 4.3**。对于任意 $\epsilon > 0$，在与定理 *4.2*相同的条件下，以至少 $1 - \frac{3+5\gamma}{1+\gamma}\epsilon$的概率，

$$\left\| L_{f,\lambda}^{-1} - \mathfrak{L}_{f,\lambda}^{-1} \right\| \leq 16t \left( \frac{\overline{\delta}_f + \lambda_f}{\lambda_f} \right)^2,$$

并且以至少 $1 - 4\epsilon$ 的概率，

$$\left\| L_{w,\lambda}^{-1} - \mathfrak{L}_{w,\lambda}^{-1} \right\| \leq 16t \left( \frac{\overline{\delta}_w + \lambda_w}{\lambda_w} \right)^2,$$

其中，$\overline{\delta}_w =$ 最大值$_k \frac{1}{n_k} C(k,\cdot)$ 和 $\overline{\delta}_f =$ 最大值$_j\, \theta_j C(\cdot, \ell_j)$ 分别表示期望工人度和企业度的最大值。

**注释 2**。我们利用以下不等式得到这些界：

$$\left\| L_{f,\lambda}^{-1} - \mathfrak{L}_{f,\lambda}^{-1} \right\| = \left\| L_{f,\lambda}^{-1} \left( L_{f,\lambda} - \mathfrak{L}_{f,\lambda} \right) \mathfrak{L}_{f,\lambda}^{-1} \right\| \leq \left\| L_{f,\lambda}^{-1} \right\| \left\| \mathfrak{L}_{f,\lambda}^{-1} \right\| \left\| L_{f,\lambda} - \mathfrak{L}_{f,\lambda} \right\|.$$

然后，我们使用引理 4.1 对 $\text{eigmin}(\mathfrak{L}_{f,\lambda}) \geq \frac{\lambda_f}{\delta_f + \lambda_f}$ 进行界定，因此对 $\text{eigmax}(\mathfrak{L}_{f\lambda,}^{-1}) \geq \frac{\delta_f + \lambda_f}{\lambda_f}$ 进行界定；经验拉普拉斯矩阵同理。若不进行正则化，拉普拉斯矩阵的最小特征值等于零。使用摩尔－彭若斯逆可以很容易地避免这一问题。然而，即使图是连通的，第二小特征值也不一定远离零。拉普拉斯矩阵的第二小特征值较小，反映出网络中存在瓶颈；随着稀疏性的增加，瓶颈会变得更加频繁（Zhang和Rohe (2018)）。

**注释 3.** 在许多实证设置中，度不会随节点数增加。在经济学工资应用中，工人度和企业度会随观测期数增加，而观测期数是一个固定且相当小的数字。因此，节点数增加时，取逆操作不会使集中界显著恶化。

[4]Chung和Radcliffe (2011)以及 Chaudhuri等人 (2012)假设最小期望度为 $\delta \geq \nu \ln n$，其中 $n$ 是图中的节点数。正如 Qin和Rohe (2013)所强调的，正则化消除了提高节点度的必要性。

本节最后，我们将证明一个关于非归一化拉普拉斯矩阵逆矩阵的类似集中不等式，

$$\widetilde{L}_{f,\lambda}^{-1} = D_{f,\lambda}^{-1/2} L_{f,\lambda}^{-1} D_{f,\lambda}^{-1/2} = \left(D_{f,\lambda} - B^\top D_{w,\lambda}^{-1} B\right)^{-1},$$
$$\widetilde{L}_{w,\lambda}^{-1} = D_{w,\lambda}^{-1/2} L_{w,\lambda}^{-1} D_{w,\lambda}^{-1/2} = \left(D_{w,\lambda} - B D_{f,\lambda}^{-1} B^\top\right)^{-1}.$$

**定理 4.4** (正则化非归一化拉普拉斯矩阵的逆矩阵的集中性)。 对于任意 $\epsilon > 0$, 在与定理 *4.2*相同的条件下，以至少 $1 - \frac{3+9\gamma}{1+\gamma}\epsilon$的概率，

$$\left\|\widetilde{L}_{f,\lambda}^{-1} - \widetilde{\mathfrak{L}}_{f,\lambda}^{-1}\right\| = \left\|D_{f,\lambda}^{-1/2} L_{f,\lambda}^{-1} D_{f,\lambda}^{-1/2} - \mathfrak{D}_{f,\lambda}^{-1/2} \mathfrak{L}_{f,\lambda}^{-1} \mathfrak{D}_{f,\lambda}^{-1/2}\right\| \leq \left(5 + 16\frac{\overline{\delta}_f + \lambda_f}{\underline{\delta}_f + \lambda_f}\right) \frac{\overline{\delta}_f + \lambda_f}{\lambda_f^2} t,$$

并且，以至少 $1 - \frac{8+4\gamma}{1+\gamma}\epsilon$ 的概率,——

$$\left\|\widetilde{L}_{w,\lambda}^{-1} - \widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\| = \left\|D_{w,\lambda}^{-1/2} L_{w,\lambda}^{-1} D_{w,\lambda}^{-1/2} - \mathfrak{D}_{w,\lambda}^{-1/2} \mathfrak{L}_{w,\lambda}^{-1} \mathfrak{D}_{w,\lambda}^{-1/2}\right\| \leq \left(5 + 16\frac{\overline{\delta}_w + \lambda_w}{\underline{\delta}_w + \lambda_w}\right) \frac{\overline{\delta}_w + \lambda_w}{\lambda_w^2} t.$$

**注 4**。 如果期望节点度不随节点数增加，而正则化参数随网络规模增大，那么界的阶为 $t/\lambda_f$ 和 $t/\lambda_w$ ，因此对于非归一化拉普拉斯矩阵的逆矩阵，界更紧。这是合理的，因为非归一化拉普拉斯矩阵等于归一化拉普拉斯矩阵乘以正则化度矩阵，而后者的最小特征值有下界，即正则化参数的倒数。

## 5. 岭回归

本节将说明，上一节的集中不等式保证岭固定效应在偏差和方差方面都收敛到定义良好的极限。

将岭估计量改写为

$$\widehat{\mu} - \mu = \widetilde{L}_{w,\lambda}^{-1}\left[-\lambda_w \mu + \lambda_f B D_{f,\lambda}^{-1}\phi + \left(W^\top - B D_{f,\lambda}^{-1} F^\top\right) U\right],$$
$$\widehat{\phi} - \phi = \widetilde{L}_{f,\lambda}^{-1}\left[\lambda_w B^\top D_{w,\lambda}^{-1}\mu - \lambda_f \phi + \left(F^\top - B^\top D_{w,\lambda}^{-1} W^\top\right) U\right],$$

使用 $Y = W\mu + F\phi + U$，其中 $U$具有均值为 0、方差为 $\sigma^2$ 的独立元素。

我们假设参数（$\mu$， $\phi$）是随机的，满足

$$\mu = Z_w \mu^* + U_w, \quad \phi = Z_f \phi^* + U_f,$$

其中， $Z_w$和 $Z_f$是表示工人和企业社群的 $n \times K$和 $p \times K$矩阵， $\mu^*$ 和 $\phi^*$ 是组固定效应的 $K$向量， $U_w$ 和 $U_f$ 是均值为零、方差分别为 $\sigma_w^2$ 和 $\sigma_f^2$的独立分量向量。进一步假设 $U_w$ 和 $U_f$ 与 $U$相互独立。

### *5.1.* 偏差

给定社区结构 $Z=(Z_w,\ Z_f)$ 和 $\beta^*=(\mu^*,\ \phi^*)$，以及网络结构 $X=(W,\ F)$，$\widehat{\mu}$ 和 $\phi$ 上的偏差为

$$b_{\mu,\lambda}:=\mathbb{E}\left(\widehat{\mu}-\mu\mid X,Z,\beta^*\right)=\widetilde{L}_{w,\lambda}^{-1}\left(-\lambda_w Z_w\mu^*+\lambda_f BD_{f,\lambda}^{-1}Z_f\phi^*\right),$$
$$b_{\phi,\lambda}:=\mathbb{E}\left(\widehat{\phi}-\phi\mid X,Z,\beta^*\right)=\widetilde{L}_{f,\lambda}^{-1}\left(-\lambda_f Z_f\phi^*+\lambda_w B^\top D_{w,\lambda}^{-1}Z_w\mu^*\right).$$

将 $\mu$ 和 $\phi$ 上的确定性偏差极限定义为使用期望邻接矩阵 $\mathfrak{B}$ 得到的值，

$$\mathfrak{b}_{\mu,\lambda}:=\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\left(-\lambda_w Z_w\mu^*+\lambda_f\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}Z_f\phi^*\right)$$
$$b_{\phi,\lambda}:=\widetilde{L}_{f,\lambda}^{-1}\left(-\lambda_f Z_f\phi^*+\lambda_w B^\top D_{w,\lambda}^{-1}Z_w\mu^*\right).$$

下面的定理证明，只要 $\lambda_w,\lambda_f$ 的增长速度快于 $\ln(n+p)$，那么 $\mathfrak{b}_{\mu,\lambda}$ 和 $\mathfrak{b}_{\phi,\lambda}$ 分别是偏差 $b_{\mu,\lambda}$ 和 $b_{\phi,\lambda}$ 在均方误差意义下的渐近良好预测量。

**定理 5.1.** 对于任意 $\epsilon>0$，在与定理 *4.2*相同的条件下，以至少 $1-\frac{11+7\gamma}{1+\gamma}\epsilon$的概率，

$$\|b_{\mu,\lambda}-\mathfrak{b}_{\mu,\lambda}\|\le\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t\left(\lambda_w\sqrt{n}\,\|\mu^*\|+\lambda_f\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}\sqrt{p}\,\|\phi^*\|\right)$$
$$+2\lambda_f\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}\left(\sqrt{\chi_f}+2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right)t\sqrt{p}\,\|\phi^*\|,$$

其中 $\chi_f=$最大 $\left(1,\frac{\delta_w}{\delta_f+\lambda_f}\right)$。并且以至少 $1-\frac{6+12\gamma}{1+\gamma}\epsilon$ 的概率，

$$\|b_{\phi,\lambda}-\mathfrak{b}_{\phi,\lambda}\|\le\left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}t\left(\lambda_f\sqrt{p}\,\|\phi^*\|+\lambda_w\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_w+\lambda_w}}\sqrt{n}\,\|\mu^*\|\right)$$
$$+2\lambda_w\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}\left(\sqrt{\chi_w}+2\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\lambda_w}}\right)t\sqrt{n}\,\|\mu^*\|,$$

其中 $\chi_w=$最大 $\left(1,\frac{\delta_f}{\delta_w+\lambda_w}\right)$。

**备注 5。** 如果 $p/n\to\gamma$ 并且岭参数以相同的速率增长，且快于 $\ln(n+p)$，例如 $(\ln(n+p))^{1+\nu'}$ 且 $\nu'>0$；参见备注 1。因此，偏差界的阶为 $\sqrt{n}t$。于是，$b_{\mu,\lambda}$ 和$\mathfrak{b}_{\mu}$之间的均方根误差$_{\lambda}$ 有界于若干倍的 $t$，如定理 4.2所示，因此依概率收敛于 0。

### *5.2.* 方差

$\widehat{\mu}-\mu$ 的条件方差为

$$\begin{aligned}V_{w,\lambda} &= \mathbb{V}\left(\widehat{\mu}-\mu \mid X, Z, \beta^*\right)\\ &= \widetilde{L}_{w,\lambda}^{-1}\,\mathbb{V}\left(-\lambda_w U_w+\lambda_f B D_{f,\lambda}^{-1}U_f+W^\top U-BD_{f,\lambda}^{-1}F^\top U \mid X, Z, \beta^*\right)\widetilde{L}_{w,\lambda}^{-1}\\ &= \widetilde{L}_{w,\lambda}^{-1}\left[\left(\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right)I_n+\left(\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right)BD_{f,\lambda}^{-2}B^\top+\sigma^2\widetilde{L}_{w,\lambda}\right]\widetilde{L}_{w,\lambda}^{-1}\\ &= \sigma^2\widetilde{L}_{w,\lambda}^{-1}+\widetilde{L}_{w,\lambda}^{-1}\left[\left(\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right)I_n+\left(\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right)BD_{f,\lambda}^{-2}B^\top\right]\widetilde{L}_{w,\lambda}^{-1}.\end{aligned}$$

将方差矩阵的确定性等价物定义为

$$\mathfrak{V}_{w,\lambda}=\sigma^2\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}+\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\left[\left(\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right)I_n+\left(\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right)\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-2}\mathfrak{B}^\top\right]\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}.$$

这些方差的企业端对应量为

$$\begin{aligned}V_{f,\lambda}&=\sigma^2\widetilde{L}_{f,\lambda}^{-1}+\widetilde{L}_{f,\lambda}^{-1}\left[\left(\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right)I_p+\left(\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right)B^\top D_{w,\lambda}^{-2}B\right]\widetilde{L}_{f,\lambda}^{-1},\\ \mathfrak{V}_{f,\lambda}&=\sigma^2\widetilde{\mathfrak{L}}_{f,\lambda}^{-1}+\widetilde{\mathfrak{L}}_{f,\lambda}^{-1}\left[\left(\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right)I_p+\left(\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right)\mathfrak{B}^\top\mathfrak{D}_{w,\lambda}^{-2}\mathfrak{B}\right]\widetilde{\mathfrak{L}}_{f,\lambda}^{-1}.\end{aligned}$$

下一个定理表明，只要岭参数以相同的速率增长，且增长速度快于 $\ln(n+p)$，$\mathfrak{V}_{w,\,\lambda}$ 和 $\mathfrak{V}_{f,\,\lambda}$ 就渐近等价于方差 $V_{w,\,\lambda}$ 和 $V_{f,\,\lambda}$。

**定理 5.2**。 对于任意 $\epsilon>0$，在满足定理 *4.2*相同条件时，以至少 $1-\frac{13+6\gamma}{1+\gamma}\epsilon$ 的概率，

$$\begin{aligned}\|V_{w,\lambda}-\mathfrak{V}_{w,\lambda}\| &\le \sigma^2\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t\\ &+\left|\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right|\left(\frac{2}{\lambda_w}+\frac{1}{\underline{\delta}_w+\lambda_w}\right)\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\left(\overline{\delta}_w+\lambda_w\right)^2}{\lambda_w^3}t\\ &+\left|\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right|\left(4\frac{1}{\lambda_w}\sqrt{\frac{1}{\lambda_f}}+\frac{1}{\underline{\delta}_w+\lambda_w}\sqrt{\frac{1}{\underline{\delta}_f+\lambda_f}}\right)\frac{\left(\overline{\delta}_w+\lambda_w\right)^{5/2}}{\lambda_w^3}\\ &\quad\times\left[\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}+2\left(\sqrt{\chi_w}+2\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\lambda_w}}\right)\right]t.\end{aligned}$$

同理，以至少 $1-\frac{5+13\gamma}{1+\gamma}\epsilon$ 的概率，

$$\begin{aligned}\|V_{f,\lambda}-\mathfrak{V}_{f,\lambda}\| \leq \sigma^2\left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}t \\ +\left|\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right|\left(\frac{2}{\lambda_f}+\frac{1}{\underline{\delta}_f+\lambda_f}\right)\left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\frac{\left(\overline{\delta}_f+\lambda_f\right)^2}{\lambda_f^3}t \\ +\left|\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right|\left(4\frac{1}{\lambda_f}\sqrt{\frac{1}{\lambda_w}}+\frac{1}{\underline{\delta}_f+\lambda_f}\sqrt{\frac{1}{\underline{\delta}_w+\lambda_w}}\right)\frac{\left(\overline{\delta}_f+\lambda_f\right)^{5/2}}{\lambda_f^3} \\ \times\left[\left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_w+\lambda_w}}+2\left(\sqrt{\chi_w}+2\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\lambda_w}}\right)\right]t.\end{aligned}$$

**注 6**。如果 $p/n\to\gamma$，并且岭参数以相同的速率增长，且其增长速度快于 $\ln(n+p)$，即 $(\ln(n+p))^{1+\nu'}$ 且 $\nu'>0$，则不等式右侧的第一项的阶为 $t/\lambda_w$ 或 $t/\lambda_f$，最后两项的阶为 $t$。

### *5.3.* 预测

假设我们从相同的数据生成过程中抽取另一个网络和结果变量，即 $\sim\sim\sim\sim\sim\sim\sim X=(W,\ F)$ 和 $Y=W\mu+F\phi+U$。预测 $SSE$ 为

$$\begin{aligned}SSE&=\mathbb{E}\left[\left(\widetilde{W}(\mu-\widehat{\mu})+\widetilde{F}(\phi-\widehat{\phi})+\widetilde{U}\right)^\top\left(\widetilde{W}(\mu-\widehat{\mu})+\widetilde{F}(\phi-\widehat{\phi})+\widetilde{U}\right)\mid X,Z,\beta^*\right]\\&=\sigma^2+\mathbb{E}\left[(\mu-\widehat{\mu})^\top\mathfrak{D}_w(\mu-\widehat{\mu})+2(\mu-\widehat{\mu})^\top\mathfrak{B}(\phi-\widehat{\phi})+(\phi-\widehat{\phi})^\top\mathfrak{D}_f(\phi-\widehat{\phi})\mid X,Z,\beta^*\right],\end{aligned}$$

因为 $\widetilde{W}^\top\widetilde{W}$、$\widetilde{F}^\top\widetilde{F}$ 和 $\widetilde{W}^\top\widetilde{F}$ 的期望与之前相同。

第一个非平凡项为

$$\begin{aligned}\mathbb{E}\left[(\mu-\widehat{\mu})^\top\mathfrak{D}_w(\mu-\widehat{\mu})\mid X,Z,\beta^*\right]&=\mathbb{E}\left(\widehat{\mu}-\mu\mid X,Z,\beta^*\right)^\top\mathfrak{D}_w\,\mathbb{E}\left(\widehat{\mu}-\mu\mid X,Z,\beta^*\right)\\&\quad+\operatorname{tr}\left[\mathfrak{D}_w\,\mathbb{V}\left(\widehat{\mu}-\mu|X,Z,\beta^*\right)\right]=b_{\mu,\lambda}^\top\mathfrak{D}_w b_{\mu,\lambda}+\operatorname{tr}\left[\mathfrak{D}_w V_{w,\lambda}\right].\end{aligned}$$

其余各项可类似地得到，从而有

$$SSE=b_{\mu,\lambda}^\top\mathfrak{D}_w b_{\mu,\lambda}+\operatorname{tr}\left[\mathfrak{D}_w V_{w,\lambda}\right]+2b_{\mu,\lambda}^\top\mathfrak{B}b_{\phi,\lambda}+\operatorname{tr}\left[\mathfrak{B}C_\lambda\right]+b_{\phi,\lambda}^\top\mathfrak{D}_f b_{\phi,\lambda}+\operatorname{tr}\left[\mathfrak{D}_f V_{f,\lambda}\right],$$

其中 $C_\lambda=\mathbb{E}\left[(\widehat{\mu}-\mu-b_{\mu,\lambda})\left(\widehat{\phi}-\phi-b_{\phi,\lambda}\right)^\top\mid X,Z,\beta^*\right]$ 表示协方差矩阵。

将偏差和方差替换为它们的确定性等价物，即可得到一个确定性等价物。

考虑第一项:

$$\left|b_{\mu,\lambda}^\top\mathfrak{D}_w b_{\mu,\lambda}-\mathfrak{b}_{\mu,\lambda}^\top\mathfrak{D}_w\mathfrak{b}_{\mu,\lambda}\right|\leq\|\mathfrak{D}_w\|\left(\|b_{\mu,\lambda}\|+\|\mathfrak{b}_{\mu,\lambda}\|\right)\|b_{\mu,\lambda}-\mathfrak{b}_{\mu,\lambda}\|,$$

¯其中， $\|\mathfrak{D}_w\| = \delta_w$ 是最大工人期望度数。容易证明，在前两个定理的条件下，如果岭参数以相同的速率趋于无穷大，且该速率快于 $\ln(n+p)$，那么此量的阶为 $nt$。

考虑第二项：

$$|\mathrm{tr}\left[\mathfrak{D}_w V_{w,\lambda}\right] - \mathrm{tr}\left[\mathfrak{D}_w \mathfrak{V}_{w,\lambda}\right]| = |\mathrm{tr}\left[\mathfrak{D}_w \left(V_{w,\lambda} - \mathfrak{V}_{w,\lambda}\right)\right]| \leq \delta_w n \left\|V_{w,\lambda} - \mathfrak{V}_{w,\lambda}\right\|,$$

其阶数与第一项相同。

所有项均依此类推。这表明，均方误差（$SSE$ 除以观测值数量 $N$）接近 $n$ 乘以平均工人度，并将收敛于一个确定性等价极限。研究这一极限并推导高维数据集的最优岭参数不在本文讨论范围内。

## 6. 应用

### *6.1.* 模拟结果

我们从初始数量为 $n_0$ 名工人和 $p_0$ 家企业开始。我们对模型进行模拟，

$$C = c\frac{p_0}{K}\left[I_K + \delta(J_K - I_K)\right],$$

其中 $c$ 为常数， $J_K = \mathbf{1}_K \mathbf{1}_K^\top$ 和 $\delta$ 为调节参数： $\delta = 0$ 表示完全隔离，$\delta = 1$ 表示无差异。初始节点数的设定使得 $n_0 = 3p_0$ ，社区数量为 $K = 5$。社区强度由 $\delta = 0.1$控制。我们从尺度参数为 $\alpha = 2$ 、最小值为 $\theta_{min} = 1$的帕累托分布中独立抽取 $\theta_j$： $\Pr\{\theta_j > x\} = (\theta_{min}/x)^\alpha$。然后，对每个组 $\ell$分别处理，将属于组 $\ell$ 的所有企业 $j$ 的 $\theta_j$除以 $\sum_{j'} \theta_{j'} \delta_{\ell_{j'},\ell}$进行归一化。我们选取最大连通分量，且 $n$ 和 $p$ 分别表示保留下来的工人和企业数量。

最后，我们考虑以下工资参数取值：

$$\mu_i = k_i - 1 + \sqrt{2}\mathcal{N}(0,1), \quad \phi_j = 0.4(\ell_j - 1) + \mathcal{N}(0,1).$$

这意味着工资存在某种正向排序（工资固定效应呈正相关）。残差方差为 $\sigma = 2$。

表 1 展示了模拟网络的主要特征，第一次使用 $c = 1$，第二次使用 $c = 2$。当 $c = 1$时，该网络非常稀疏，包含许多非连通分量（移除没有连接的节点后）。随着 $c$ 增加，网络变得更加稠密。因此，为了生成一个规模大致相同、接近 10000 个观测值的连通图，我们对 $p_0$ 使用了比 $c = 1$ 更大的初始企业节点值 $c = 2$ 。 $N$

我们在表 2 中展示了对数工资方差分解的相应模拟结果。第一行（“真实值”）展示了真实方差分解

TABLE 1
*Characteristics of the simulated networks*

| | $c$ | 1 | 2 |
|---|---|---|---|
| 初始企业数 | $p_0$ | 30000 | 4500 |
| 连通分量数 | | 11757 | 714 |
| 企业数 | $p$ | 3059 | 3081 |
| 工人数 | $n$ | 7249 | 6928 |
| 连接数 | $N = d_{\cdot\cdot}$ | 10341 | 11267 |
| 稀疏性 | $\frac{N}{pn}$ | 0.05% | 0.05% |
| 工人平均度数 | $\frac{N}{n}$ | 1.4265 | 1.6263 |
| 企业平均度数 | $\frac{N}{p}$ | 3.3805 | 3.6569 |

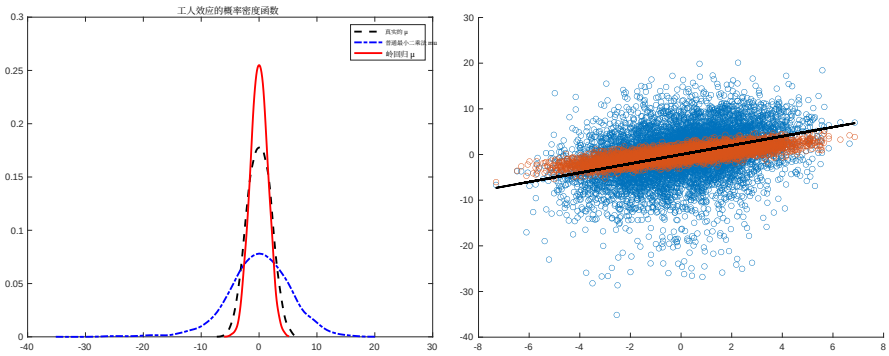
TABLE 2
*Simulated regression results*

| | 工资方差份额 | | | | 固定效应 | 样本外 | $\frac{\lambda_w}{N/n}$ | $\frac{\lambda_f}{N/p}$ |
|---|---|---|---|---|---|---|---|---|
| | 工人 | firm | 2*协方差 | 残差 | 相关性 | MSE | | |
| | | | $c=1$ | | | | | |
| true | 0.4259 | 0.1378 | 0.1212 | 0.3113 | 0.2500 | | | |
| OLS | 3.1724 | 2.8607 | -5.0343 | 0.0012 | -0.8356 | 7.4778 | | |
| 去偏普通最小二乘法 | -0.0453 | -0.2110 | 0.8959 | 0.3605 | | | | |
| 岭回归 | 0.2552 | 0.1028 | 0.1101 | 0.5319 | 0.3399 | 2.1289 | 0.58 | 0.83 |
| | | | $c=2$ | | | | | |
| true | 0.4363 | 0.1463 | 0.1062 | 0.3292 | 0.2101 | | | |
| OLS | 0.9798 | 0.5792 | -0.5949 | 0.0358 | -0.3948 | 3.6589 | | |
| 去偏普通最小二乘法 | 0.4089 | 0.1178 | 0.1526 | 0.3208 | | | | |
| 岭回归 | 0.2858 | 0.1026 | 0.1115 | 0.5002 | 0.3255 | 2.1167 | 0.48 | 0.53 |

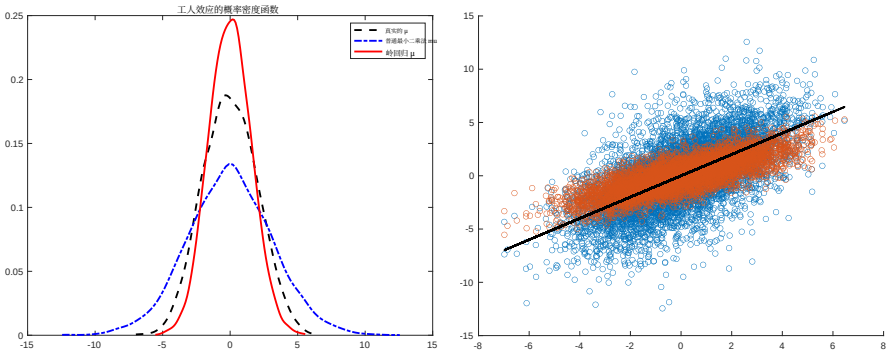
（以及固定效应之间的相关性）。固定效应参数经过选择，使分解结果与常见的实证结果相似，即工人贡献较大、残差方差较大，而企业和排序贡献较小。接着，我们展示使用固定效应的普通最小二乘法估计值以及去偏方差分量（Andrews 等人 （2008））得到的方差分解。[5] 标记为“岭回归”的一行展示了使用岭估计量得到的方差分解。我们展示了 $\widehat{\mathbb{V}}(W\hat{\mu})$、$\widehat{\mathbb{V}}(F\phi)$、$2\,\widehat{\text{Cov}}(W\hat{\mu}, F\phi)$的占比。表 2的第六列展示了样本外均方误差。最后，我们报告通过交叉验证得到的岭正则化参数。进行交叉验证时，我们按如下方式模拟测试样本：使用相同的节点和社区，但为每一对工人—企业节点抽取一个新的 $d_{ij}$。随后，我们抽取一个新的残差 $u_{ijs}$。同样，我们保留最大连通分量用于预测。

普通最小二乘法估计始终存在较大的偏差，但只要图不太稀疏，偏差校正就非常有效（即适用于 $c=2$，但不适用于 $c=1$）。与普通最小二乘法相比，使用岭估计量进行的方差分解受稀疏性的影响要小得多。岭估计量往往会低估 $\widehat{\mathbb{V}}(W\hat{\mu})$、$\widehat{\mathbb{V}}(F\phi)$ 和 $2\,\widehat{\text{Cov}}(W\hat{\mu}, F\phi)$ 以下各项的份额；并高估“残差份额”。我们将岭正则化参数报告为平均

[5] 我们不考虑其他偏差校正方法（Gaure (2014)、Kline 等人 (2020)、Azkarate-Askasua 和 Zerecero (2022)），因为模型是同方差的。




(a) $c = 1$: $\mu_i$ 的概率密度函数　(b) $c = 1$: $\mu_i$ 的散点图




(c) $c = 2$: $\mu_i$ 的概率密度函数　(d) $c = 2$: $\mu_i$ 的散点图

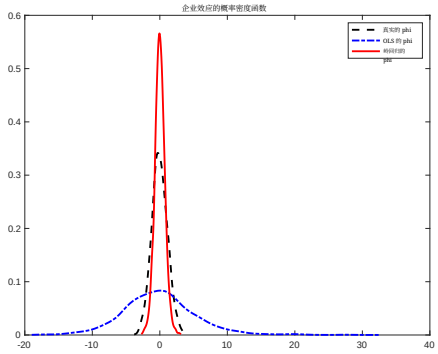
图1：工人固定效应分布：真实值、普通最小二乘法和岭回归

工人度和企业度的比例。

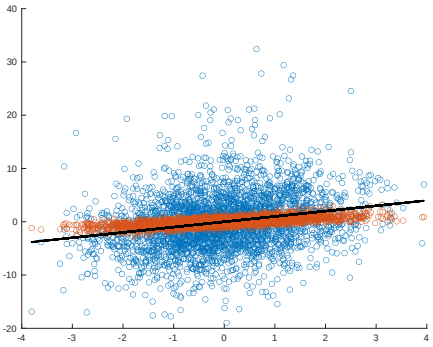
接下来，我们转向固定效应分布。图1 和2 展示了真实固定效应和估计固定效应的密度及其散点图。散点图展示了估计值与真实值的关系。黑线是表示完美估计的45度线。我们看到，普通最小二乘法的估计值比真实分布分散得多，而岭估计量的分布仅略微集中一些。散点图显示，岭估计量的偏差存在一定规律：真实固定效应较大的值对应更强的负偏差，反之亦然。

### *6.2* 真实数据应用

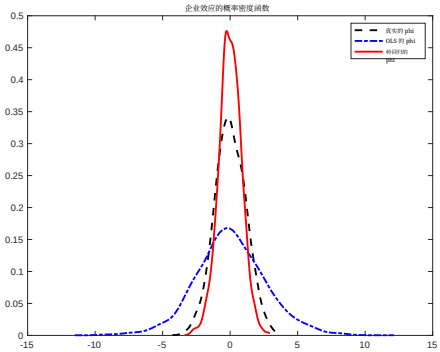
随后，我们使用法国雇主—雇员匹配数据（DADS面板数据，1995—2001年）进行一项应用，数据与Abowd 等人(1999)使用的数据类似。我们仅保留从事非农、有薪、私营部门全职工作的55岁以下雇员。我们保留所有就业时段，包括记录在



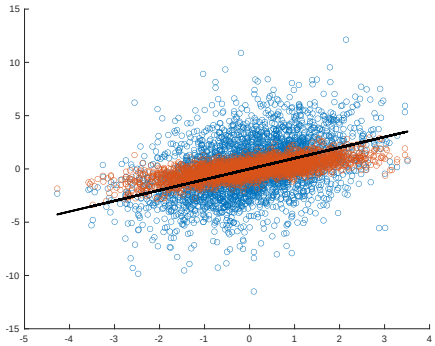

(a) $c = 1$: $\phi_i$ 的概率密度函数



(b) $c = 1$: $\phi_i$ 的散点图




(c) $c = 2$: $\phi_i$ 的概率密度函数



(d) $c = 2$: $\phi_i$ 的散点图

图 2：企业固定效应分布：真实值、普通最小二乘法和岭回归

TABLE 3
*Actual data results*

| | 工资方差份额 | | | |
|---|---|---|---|---|
| | 工人 | firm | 2*协方差 | 残差 |
| 最大连通分量 | | | | |
| OLS | 0.9584 | 0.2332 | -0.3163 | 0.1247 |
| 去偏普通最小二乘法 | 0.8841 | 0.1912 | -0.2347 | 0.1594 |
| 岭回归 | 0.6005 | 0.0846 | 0.1067 | 0.2083 |
| 最大强连通分量 | | | | |
| OLS | 0.8626 | 0.1008 | -0.0978 | 0.1324 |
| 去偏普通最小二乘法 | 0.8153 | 0.0845 | -0.067 | 0.1655 |
| 岭回归 | 0.5611 | 0.1257 | 0.1001 | 0.2131 |
| KSS | 0.7496 | 0.0655 | 0.0088 | 0.1762 |

一年中的一部分时间。不过，我们剔除了工人-工资观测值少于 50 个的企业所对应的匹配，以及记录期少于 5 个的工人。工资指标为日收入，并剔除了低于第 0.1 百分位数和高于第 99.9 百分位数的工资。我们将对数工资对工作年限、年龄及年龄的平方进行回归，并使用残差进行分析。

选出最大连通分量（共 2663 个）后，我们得到一个包含 $n = 94491$ 名工人、 $p = 3408$ 家企业和 $N = 450125$个工资观测值的网络。平均工人度为 $d_w = 4.7637$ , 平均企业度为 $d_f = 132.08$。每个匹配的平均工资记录数为 3.5903。该网络非常稀疏，共有 127457 个匹配，占 $np$ 潜在匹配的 0.03958%。

对数工资残差的方差为 0.2045，可按表 3所示进行分解。使用工人效应和企业效应的 OLS估计量得到的方差分解表明，固定效应协方差对工资的贡献为较大的负值。插件方差贡献的偏差（假设同方差性）估计较小。

接下来，我们展示通过岭回归得到的方差分解。[6]我们使用 1995—2000 年的数据估计岭模型，并将最后一年（2001 年）留作交叉验证。通过最小化样本外误差，我们将岭参数 $\lambda_w$ 和 $\lambda_f$ 分别调至工人度和企业度的 0.055 倍和 0.060 倍。[7]现在，我们估计出工人效应与企业效应之间存在正协方差，且工人效应和回归残差的贡献更为合理。

接下来，由于我们担心残差异方差性和自相关，因此希望将岭回归与Kline等人（2020）提出的留一法偏差校正方法进行比较。留一法要求企业在移除一个观测值后仍保持连通。

[6]我们展示工人效应、企业效应、工人效应与企业效应协方差的两倍，以及残差贡献对总方差的贡献；残差贡献是残差方差与残差和固定效应之间协方差之和。与普通最小二乘法估计量不同，岭估计量不会产生正交残差。[7]之前的 $\log(n + p)$ 下界是我们集中性论证的一个充分条件；我们并未声称交叉验证会产生具有这种缩放方式的惩罚项。相反，我们将所选惩罚项表示为平均度的比例，并通过模拟评估其表现。

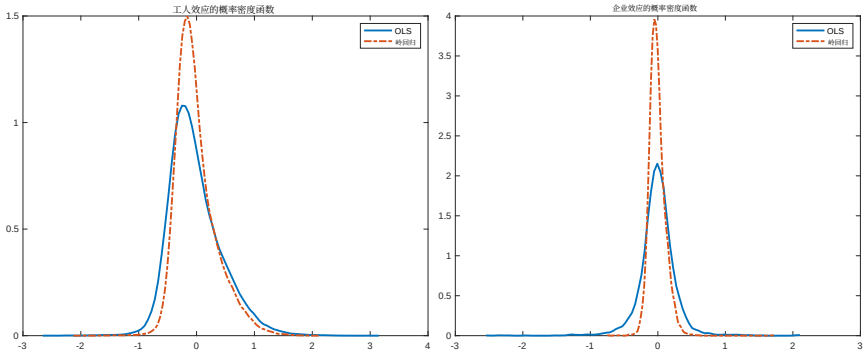



图 3：估计固定效应的分布（普通最小二乘法和岭回归）

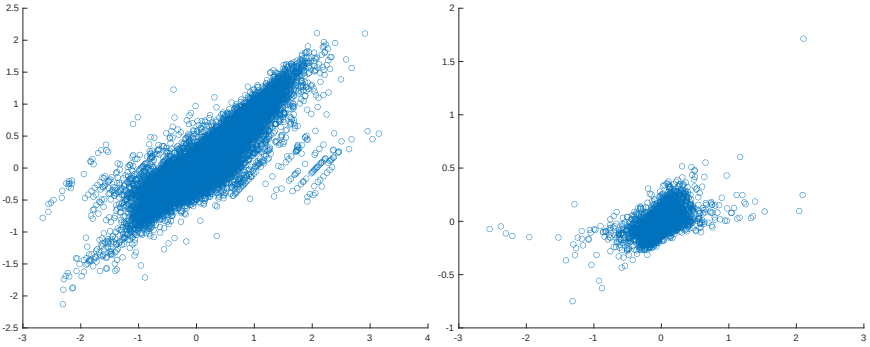


图 4：岭回归与普通最小二乘法估计的固定效应对比（左图：工人；右图：企业）

我们采用最严格的版本，即每次移除一个匹配；当匹配内的残差存在异方差或自回归时，该方法渐近地消除二阶偏差。普通最小二乘法和同方差偏差校正的结果仍非常接近，且协方差项为负，但幅度较小，这表明留一法样本的稀疏程度较低。然而，岭回归似乎几乎不受样本选择的影响。留一法校正得到的协方差贡献非常接近于零，因此更接近岭回归。这些估计结果也与Babet等人（2022）以及Azkarate-Askasua和Zerecero（2022）的结果非常接近。他们使用的是2002年之后完整的DADS 数据（5200万条观测值），而非2001年之前的面板数据，并使用小时工资而非日工资。Babet等人采用留一法校正，Azkarate-Zerecero则提出了一种相关的参数自助法校正。

接下来，我们在图 3中展示工人和企业固定效应的分布。岭估计量比OLS估计量集中得多。图4 展示了每个岭估计值与对应 OLS估计值之间的关系。企业固定效应的估计值比工人固定效应的估计值更容易出现分歧

。

## 7. 结论

本文研究了岭惩罚，作为普通最小二乘法的一种替代方法，用于估计双向固定效应回归模型中的固定效应，例如匹配的雇主—雇员数据的对数工资方程。我们构建了一个度修正随机块模型来生成网络。结果变量在第二阶段给定网络后生成。我们推导了固定效应估计量的渐近性质，并得到了该估计量的偏差和方差的确定性等价物。未来，对这些确定性等价物的进一步研究将有助于刻画最优岭参数。

## 证明

### *A.1.* 引理 *2.1*

对于任意 $x = (x_j) \in \mathbb{R}^p$， $x^\top L_f x$ 的表达式为

$$\sum_j x_j^2 - \sum_{j,j'} \frac{x_j x_{j'}}{\sqrt{d_{\cdot j} d_{\cdot j'}}} \sum_i \frac{d_{ij} d_{ij'}}{d_{i\cdot}} = \frac{1}{2} \sum_i \frac{1}{d_{i\cdot}} \sum_{j,j'} \left( \frac{x_j}{\sqrt{d_{\cdot j}}} - \frac{x_{j'}}{\sqrt{d_{\cdot j'}}} \right)^2 d_{ij} d_{ij'} \geq 0.$$

因此， $L_f$ 具有非负特征值，且 $\lambda_1 = 0$ 作为 $L_f x = \mathbf{0}_p$ 适用于 $x = \left(\sqrt{d_{\cdot j}}\right)_j$。然后，

$$x^\top L_f x \geq 0 \Rightarrow 1 \geq \frac{x^\top A_f x}{x^\top x}.$$

该瑞利商给出了 $\alpha_1$ 的上界 1，而 $\alpha_1 = 1$ 确实是 $A_f$ 的最大特征值，其对应的特征向量为 $v_1 = \left(\sqrt{d_{\cdot j}}\right)$。最后， $A_f = E^\top E$ 显然是半正定的。其特征值均为非负数。

### *A.2.* 引理 *2.3* 的证明

对于所有 $x \in \mathbb{R}^p$,

$$x^\top L_{f,\lambda} x = \frac{1}{2} \sum_i \frac{1}{d_{i\cdot} + \lambda_w} \sum_{j,j'} \left( \frac{x_j}{\sqrt{d_{\cdot j} + \lambda_f}} - \frac{x_{j'}}{\sqrt{d_{\cdot j'} + \lambda_f}} \right)^2 d_{ij} d_{ij'}$$

$$+ \sum_j \frac{x_j^2}{d_{\cdot j} + \lambda_f} \left( \lambda_f + \lambda_w \sum_i \frac{d_{ij}}{d_{i\cdot} + \lambda_w} \right) \geq \lambda_f \sum_j \frac{x_j^2}{d_{\cdot j} + \lambda_f} \geq \frac{\lambda_f x^\top x}{\max_j d_{\cdot j} + \lambda_f}.$$

为得到第二个等式，写作

$$\sum_i \frac{1}{d_{i\cdot}+\lambda_w}\sum_{j,j'}\left(\frac{x_j}{\sqrt{d_{\cdot j}+\lambda_f}}-\frac{x_{j'}}{\sqrt{d_{\cdot j'}+\lambda_f}}\right)^2 d_{ij}d_{ij'}$$
$$=2\sum_i \frac{d_{i\cdot}}{d_{i\cdot}+\lambda_w}\sum_j \frac{x_j^2}{d_{\cdot j}+\lambda_f}d_{ij}-2\sum_i \frac{1}{d_{i\cdot}+\lambda_w}\sum_{j,j'}\frac{x_j}{\sqrt{d_{\cdot j}+\lambda_f}}\frac{x_{j'}}{\sqrt{d_{\cdot j'}+\lambda_f}}d_{ij}d_{ij'}.$$

那么，$\sum_i \frac{d_{i\cdot}}{d_{i\cdot}+\lambda_w}\sum_j \frac{x_j^2}{d_{\cdot j}+\lambda_f}d_{ij}$ 可以写为

$$\sum_j \frac{x_j^2}{d_{\cdot j}+\lambda_f}\sum_i \frac{d_{i\cdot}}{d_{i\cdot}+\lambda_w}d_{ij}=\sum_j x_j^2-\sum_j \frac{x_j^2}{d_{\cdot j}+\lambda_f}\left(\lambda_f+\lambda_w\sum_i \frac{d_{ij}}{d_{i\cdot}+\lambda_w}\right).$$

### *A.3.* 引理*4.1* 的证明

对于所有 $x\in\mathbb{R}^p$，内积 $x^\top \mathfrak{L}_{f,\lambda}x$ 为

$$\sum_j x_j^2-\sum_{j,j'}\frac{x_j}{\sqrt{\theta_j C(\cdot,\ell_j)+\lambda_f}}\frac{x_{j'}}{\sqrt{\theta_{j'}C(\cdot,\ell_{j'})+\lambda_f}}\sum_k \frac{1}{n_k}\frac{\theta_j C(k,\ell_j)\theta_{j'}C(k,\ell_{j'})}{C(k,\cdot)/n_k+\lambda_w}$$
$$=\frac{1}{2}\sum_k \frac{1/n_k}{C(k,\cdot)/n_k+\lambda_w}\sum_{j,j'}\left(\frac{x_j}{\sqrt{\theta_j C(\cdot,\ell_j)+\lambda_f}}-\frac{x_{j'}}{\sqrt{\theta_{j'}C(\cdot,\ell_{j'})+\lambda_f}}\right)^2\theta_j C(k,\ell_j)\theta_{j'}C(k,\ell_{j'})$$
$$+\sum_j \frac{x_j^2}{\theta_j C(\cdot,\ell_j)+\lambda_f}\left(\lambda_f+\sum_k \frac{\lambda_w\theta_j C(k,\ell_j)}{C(k,\cdot)/n_k+\lambda_w}\right)\geq \lambda_f\sum_j \frac{x_j^2}{\theta_j C(\cdot,\ell_j)+\lambda_f}\geq x^\top x\frac{\lambda_f}{\max_j \theta_j C(\cdot,\ell_j)+\lambda_f},$$

as

$$\frac{1}{2}\sum_k \frac{1/n_k}{C(k,\cdot)/n_k+\lambda_w}\sum_{j,j'}\left(\frac{x_j}{\sqrt{\theta_j C(\cdot,\ell_j)+\lambda_f}}-\frac{x_{j'}}{\sqrt{\theta_{j'}C(\cdot,\ell_{j'})+\lambda_f}}\right)^2\theta_j C(k,\ell_j)\theta_{j'}C(k,\ell_{j'})$$
$$=\sum_k \frac{\frac{1}{n_k}C(k,\cdot)}{\frac{1}{n_k}C(k,\cdot)+\lambda_w}\sum_j \frac{x_j^2\theta_j C(k,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f}-\sum_{j,j'}\frac{x_j}{\sqrt{\theta_j C(\cdot,\ell_j)+\lambda_f}}\frac{x_{j'}}{\sqrt{\theta_{j'}C(\cdot,\ell_{j'})+\lambda_f}}\sum_k \frac{\frac{1}{n_k}\theta_j C(k,\ell_j)\theta_{j'}C(k,\ell_{j'})}{\frac{1}{n_k}C(k,\cdot)+\lambda_w}$$

由于 $\sum_j \theta_j C(k,\ell_j)=\sum_\ell \sum_j \theta_j\delta_{\ell_j\ell}C(k,\ell)=C(k,\cdot)$，并且

$$\sum_k \frac{\frac{1}{n_k}C(k,\cdot)}{\frac{1}{n_k}C(k,\cdot)+\lambda_w}\sum_j \frac{x_j^2\theta_j C(k,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f}=\sum_j x_j^2-\sum_j \frac{x_j^2}{\theta_j C(\cdot,\ell_j)+\lambda_f}\left(\lambda_f+\sum_k \frac{\lambda_w\theta_j C(k,\ell_j)}{\frac{1}{n_k}C(k,\cdot)+\lambda_w}\right).$$

### *A.4.* 定理*4.2*的证明

按照 Chung 和 Radcliffe (2011)（定理 2）以及 Qin和Rohe (2013)，首先写出

$$\begin{aligned}\|E_\lambda - \mathfrak{E}_\lambda\| &= \left\|D_{w,\lambda}^{-1/2} B D_{f,\lambda}^{-1/2} - \mathfrak{D}_{w,\lambda}^{-1/2}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \\ &= \left\|\mathfrak{D}_{w,\lambda}^{-1/2}(B-\mathfrak{B})\mathfrak{D}_{f,\lambda}^{-1/2} + D_{w,\lambda}^{-1/2} B D_{f,\lambda}^{-1/2} - \mathfrak{D}_{w,\lambda}^{-1/2} B \mathfrak{D}_{f,\lambda}^{-1/2}\right\| \\ &\le \left\|\mathfrak{D}_{w,\lambda}^{-1/2}(B-\mathfrak{B})\mathfrak{D}_{f,\lambda}^{-1/2}\right\| + \left\|D_{w,\lambda}^{-1/2} B D_{f,\lambda}^{-1/2} - \mathfrak{D}_{w,\lambda}^{-1/2} B \mathfrak{D}_{f,\lambda}^{-1/2}\right\|.\end{aligned}$$

接下来，我们分别对每一项进行界定。

*Bound for* $\left\|\mathfrak{D}_{w,\lambda}^{-1/2}(B-\mathfrak{B})\mathfrak{D}_{f,\lambda}^{-1/2}\right\|$

令 $X := \mathfrak{D}_{w\,\lambda}^{-1/2}(B-\mathfrak{B})\mathfrak{D}_{f\,\lambda}^{-1/2}$。将矩阵 $X$ 写成如下各项之和 $X = \sum_{i,j} X_{ij}$，其中

$$X_{ij} = \mathfrak{D}_{w,\lambda}^{-1/2}(d_{ij}-p_{ij})\Delta_{ij}\mathfrak{D}_{f,\lambda}^{-1/2},$$

其中，$\Delta_{ij}$ 是一个 $n \times p$ 矩阵，除位置 $(i,j)$ 上的元素为 1 外，其余元素均为 0。这些矩阵 $X_{ij}$ 相互独立、均值为 0 且一致有界。具体而言，我们有

$$X_{ij}^\top X_{ij} = (d_{ij}-p_{ij})^2 \Delta_{jj}\left(\frac{1}{(C(k_i,\cdot)/n_{k_i}+\lambda_w)(\theta_j C(\cdot,\ell_j)+\lambda_f)}\right)$$

$(\Delta_{ij}^\top \Delta_{ij} = \Delta_{jj})$ 和 $\|X_{ij}\| = \sqrt{\text{eigmax}\,(X_{ij}^\top X_{ij})} \le \sqrt{M_w M_f} \le \max(M_f,\ M_w) :=$

$M$，当 $(d_{ij}-p_{ij})^2 \in \{p_{ij}^2, (1-p_{ij})^2\} \le 1$ 时，并记 $M_w = \left(\min_k \frac{1}{n_k} C(k,\cdot) + \lambda_w\right)^{-1}$

以及 $M_f = (\min_j \theta_j C(\cdot,\ell_j) + \lambda_f)^{-1}$。此外，令

$$v(X) = \max\left\{\left\|\mathbb{E}(X^\top X)\right\|, \left\|\mathbb{E}(XX^\top)\right\|\right\} = \max\left\{\left\|\sum_{ij}\mathbb{E}(X_{ij}^\top X_{ij})\right\|, \left\|\sum_{ij}\mathbb{E}(X_{ij}X_{ij}^\top)\right\|\right\}.$$

考虑 $\sum_{ij}\mathbb{E}(X_{ij}^\top X_{ij}) = \text{diag}\left(\sum_i \frac{p_{ij}}{C(k_i,\cdot)(1-p_{ij})(1/n_{k_i}+\lambda_w)(\theta_j C(\cdot,\ell_j)+\lambda_f)}\right)_j$。我们有

$$\begin{aligned}\left\|\sum_{ij}\mathbb{E}(X_{ij}^\top X_{ij})\right\| &\le \max_j \sum_k \frac{\theta_j C(k,\ell_j)}{(C(k,\cdot)/n_k+\lambda_w)(\theta_j C(\cdot,\ell_j)+\lambda_f)} \\ &\le \max_j \frac{\theta_j C(\cdot,\ell_j)}{(\min_k C(k,\cdot)/n_k+\lambda_w)(\theta_j C(\cdot,\ell_j)+\lambda_f)} \le M_w.\end{aligned}$$

类似地，$\left\|\sum_{ij}\mathbb{E}(X_{ij}X_{ij}^\top)\right\| \le M_f$. 接下来，我们可以应用矩阵 Bernstein 不等式。对于所有 $t \ge 0$,

$$\Pr\{\|X\| \ge t\} \le (n+p)\exp\left(\frac{-t^2/2}{v(X)+Mt/3}\right) \le (n+p)\exp\left(\frac{-t^2/2}{M+Mt/3}\right).$$

我们要寻找一个有界的 $t$，例如 $0 \le t \le 1$，使得 $\Pr\{\|X\| \ge t\} \le \epsilon$。只需满足

$$(n+p)\exp\left(\frac{-t^2/2}{M+Mt/3}\right) \le (n+p)\exp\left(\frac{-t^2}{3M}\right) = \epsilon,$$

这对 $t = \sqrt{3M\ln\frac{n+p}{\epsilon}}$ 成立。此外，为使 $t \le 1$ 成立，我们还需要 $M \le \frac{1}{3\ln\frac{n+p}{\epsilon}}$。

*Bound for* $\left\|D_{w,\lambda}^{-1/2} B D_{f,\lambda}^{-1/2} - \mathfrak{D}_{w,\lambda}^{-1/2} B \mathfrak{D}_{f,\lambda}^{-1/2}\right\|$

由 $E_\lambda = D_{w,\lambda}^{-1/2} B D_{f,\lambda}^{-1/2}$，且由于 $\|E_\lambda\| < 1$ 根据引理 2.3，我们有

$$\begin{aligned}\left\|E_\lambda - \mathfrak{D}_{w,\lambda}^{-1/2} B \mathfrak{D}_{f,\lambda}^{-1/2}\right\| &= \left\|E_\lambda\left(I_p - D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right) + \left(I_n - \mathfrak{D}_{w,\lambda}^{-1/2} D_{w,\lambda}^{1/2}\right) E_\lambda D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \\ &\le \left\|I_p - D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| + \left\|I_n - \mathfrak{D}_{w,\lambda}^{-1/2} D_{w,\lambda}^{1/2}\right\|\end{aligned}$$

首先考虑项 $\left\|D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| = \max_j \sqrt{\frac{d_{\cdot j}+\lambda_f}{\theta_j C(\cdot,\ell_j)+\lambda_f}}$，我们有

$$\sqrt{\frac{d_{\cdot j}+\lambda_f}{\theta_j C(\cdot,\ell_j)+\lambda_f}} \ge \sqrt{1+t} \Leftrightarrow \frac{d_{\cdot j} - \theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f} \ge t.$$

令 $X_j = \frac{d_{\cdot j}-\theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f} = \sum_i \frac{d_{ij}-p_{ij}}{\theta_j C(\cdot,\ell_j)+\lambda_f} := \sum_i X_{ij}$。可对每个 $X_j$ 应用伯恩斯坦不等式。首先，$|X_{ij}| \le M_f$ 和

$$v_j = \sum_i \mathbb{E}\left(X_{ij}^2\right) = \left(\frac{1}{\theta_j C(\cdot,\ell_j)+\lambda_f}\right)^2 \sum_i p_{ij}(1-p_{ij}) \le M_f^2 \sum_i p_{ij} = M_f^2 \theta_j C(\cdot,\ell_j) \le M_f.$$

Hence,

$$\Pr\left\{\frac{d_{\cdot j}-\theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f} \ge t\right\} \le \exp\left(\frac{-t^2/2}{v_j + M_f t/3}\right) \le \exp\left(\frac{-t^2/2}{M_f(1+t/3)}\right) \le \exp\left(\frac{-t^2}{3M}\right)$$

若 $0 \le t \le 1$。与之前相同的 $t$，最终得到 $\Pr\left\{\frac{d_{\cdot j}-\theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f} \ge t\right\} \le$

同样的推理表明，$\Pr\{$最终得到

$$\Pr\left\{\left|\frac{d_{\cdot j}-\theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f}\right| \ge t\right\} \le \frac{2\epsilon}{n+p}.$$

我们还可以根据各度的独立性 $d_{ij}$ 推导出

$$\begin{aligned}\Pr\left\{\left\|D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \ge \sqrt{1+t}\right\} &= \Pr\left\{\max_j \frac{d_{\cdot j}+\lambda_f}{\theta_j C(\cdot,\ell_j)+\lambda_f} \ge 1+t\right\} \\ &\le \sum_j \Pr\left\{\frac{d_{\cdot j}-\theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f} \ge t\right\} = p\frac{\epsilon}{n+p}.\end{aligned}$$

接下来，对于 $a \in [0,1]$，我们可以写出

$$\begin{aligned}\Pr\left\{\left\|I_p - D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \geq a\right\} &= \Pr\left\{\max_j \left|\sqrt{\frac{d_{\cdot j}+\lambda_f}{\theta_j C(\cdot,\ell_j)+\lambda_f}} - 1\right| \geq a\right\} \\ &= \Pr\left\{\exists j, \frac{d_{\cdot j}-\theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f} \geq a^2+2a \text{ or } \leq a^2-2a\right\} \\ &\leq \sum_j \Pr\left\{\left|\frac{d_{\cdot j}-\theta_j C(\cdot,\ell_j)}{\theta_j C(\cdot,\ell_j)+\lambda_f}\right| \geq 2a-a^2\right\} \leq p\frac{2\epsilon}{n+p}\end{aligned}$$

如果我们进一步令 $2a-a^2=t$ 或 $a=1-\sqrt{1-t} \in [0,t]$。

工人度数也可以得到类似的界：

$$\Pr\left\{\left\|D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\right\| \geq \sqrt{1+t}\right\} \geq n\frac{\epsilon}{n+p} \quad \text{and} \quad \Pr\left\{\left\|I_n - D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\right\| \geq a\right\} \leq n\frac{2\epsilon}{n+p}.$$

因此，以至少 $1-\frac{2p\epsilon}{n+p}-\frac{p\epsilon}{n+p}-\frac{2n\epsilon}{n+p}=1-\frac{2+3\gamma}{1+\gamma}\epsilon$ 的概率，[8]

$$\left\|I_p - D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| + \left\|I_n - \mathfrak{D}_{w,\lambda}^{-1/2}D_{w,\lambda}^{1/2}\right\| \left\|D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \leq \left(1-\sqrt{1-t}\right)\left(1+\sqrt{t+1}\right) \leq 3t,$$

当 $1-\sqrt{1-t} \leq t$ 时，$t \in [0,1]$，且 $1+\sqrt{t}+1<3$。因此，以至少 $1-\frac{2+3\gamma}{1+\gamma}\epsilon-\epsilon=1-\frac{3+4\gamma}{1+\gamma}\epsilon$ 的概率，我们有 $\|E_\lambda - \mathfrak{E}_\lambda\| \leq 4t$。

拉普拉斯矩阵的界

对于拉普拉斯矩阵，乘以因子 2 后即可得到该不等式，因为我们有

$$\begin{aligned}\|L_{f,\lambda} - \mathfrak{L}_{f,\lambda}\| &= \left\|\mathfrak{E}_\lambda^\top\mathfrak{E}_\lambda - E_\lambda^\top E_\lambda\right\| = \left\|\mathfrak{E}_\lambda^\top(\mathfrak{E}_\lambda - E_\lambda) + (\mathfrak{E}_\lambda - E_\lambda)^\top E_\lambda\right\| \\ &\leq \left(\|\mathfrak{E}_\lambda\| + \|E_\lambda\|\right)\|\mathfrak{E}_\lambda - E_\lambda\| \leq 2\|\mathfrak{E}_\lambda - E_\lambda\|,\end{aligned}$$

以及 $\|L_{w,\lambda} - \mathfrak{L}_{w,\lambda}\| \leq 2\|\mathfrak{E}_\lambda - E_\lambda\|$.

### *A.5.* 定理 *4.3* 的证明

我们有

$$\begin{aligned}\|L_{f,\lambda} - \mathfrak{L}_{f,\lambda}\| &= \left\|\mathfrak{E}_\lambda^\top\mathfrak{E}_\lambda - E_\lambda^\top E_\lambda\right\| = \left\|\mathfrak{E}_\lambda^\top(\mathfrak{E}_\lambda - E_\lambda) + (\mathfrak{E}_\lambda - E_\lambda)^\top E_\lambda\right\| \\ &\leq \left(\|\mathfrak{E}_\lambda\| + \|E_\lambda\|\right)\|\mathfrak{E}_\lambda - E_\lambda\| \leq 2\|\mathfrak{E}_\lambda - E_\lambda\|,\end{aligned}$$

当 $\|\mathfrak{E}_\lambda\|, \|E_\lambda\| \leq 1$ 时。因此，

$$\begin{aligned}\left\|L_{f,\lambda}^{-1} - \mathfrak{L}_{f,\lambda}^{-1}\right\| &= \left\|L_{f,\lambda}^{-1}\left(L_{f,\lambda} - \mathfrak{L}_{f,\lambda}\right)\mathfrak{L}_{f,\lambda}^{-1}\right\| \leq \left\|L_{f,\lambda}^{-1}\right\|\left\|\mathfrak{L}_{f,\lambda}^{-1}\right\|\|L_{f,\lambda} - \mathfrak{L}_{f,\lambda}\| \\ &\leq \frac{\max_j d_{\cdot j}+\lambda_f}{\lambda_f}\frac{\max_j \theta_j C(\cdot,\ell_j)+\lambda_f}{\lambda_f}2\|\mathfrak{E}_\lambda - E_\lambda\|,\end{aligned}$$

[8] $P\left(\bigcup_k\{A_k > a_k\}\right) \geq 1-\sum_k \alpha_k$ $A_k \leq a_k\ \forall k$ $P\left(\bigcap_k\{A_k \leq a_k\}\right) = 1-$ 首先，$\prod_k A_k \leq \prod_k a_k$
() 对某些 成立意味着 。其次，， 意味着 。因此，$P\left(\prod_k A_k \leq \prod_k a_k\right) \geq P\left(\bigcap_k\{A_k \leq a_k\}\right) \geq 1-\sum_k \alpha_k$。

根据引理 2.3，$\text{eigmin}(L_{f,\lambda}) \geq \frac{\lambda_f}{\max_j d_{\cdot j}+\lambda_f}$，因此 $\text{eigmax}(L_{f,\lambda}^{-1}) \leq \frac{\max_j d_{\cdot j}+\lambda_f}{\lambda_f}$；根据引理 4.1，$\text{eigmin}(\mathfrak{L}_{f,\lambda}) \geq \frac{\lambda_f}{\max_j \theta_j C(\cdot,\ell_j)+\lambda_f} = \frac{\lambda_f}{\delta_f+\lambda_f}$。现在，我们已经证明，对于 $M_w \vee M_f \leq \frac{1}{3\ln\frac{n+p}{\epsilon}}$ 和 $t \equiv \sqrt{3(M_w \vee M_f)\ln\frac{n+p}{\epsilon}} \leq 1$，对所有 $j$，

$$\Pr\left\{\frac{d_{\cdot j}+\lambda_f}{\theta_j C(\cdot,\ell_j)+\lambda_f} \geq t+1\right\} \leq \frac{\epsilon}{n+p}.$$

这意味着

$$\Pr\left\{\frac{d_{\cdot j}+\lambda_f}{\max_j \theta_j C(\cdot,\ell_j)+\lambda_f} \geq t+1\right\} \leq \frac{\epsilon}{n+p}.$$

因此，

$$\Pr\left\{\frac{\max_j d_{\cdot j}+\lambda_f}{\max_j \theta_j C(\cdot,\ell_j)+\lambda_f} \geq t+1\right\} \leq \frac{p\epsilon}{n+p} = \frac{\gamma}{1+\gamma}\epsilon.$$

（见定理 4.2的证明。）我们还有

$$\Pr\left\{\|E_\lambda - \mathfrak{E}_\lambda\| \geq 4t\right\} \leq \frac{3+4\gamma}{1+\gamma}\epsilon.$$

因此，以至少 $1-\frac{3+4\gamma}{1+\gamma}\epsilon-\frac{\gamma}{1+\gamma}\epsilon = 1-\frac{3+5\gamma}{1+\gamma}\epsilon$ 的概率有

$$\left\|L_{f,\lambda}^{-1}-\mathfrak{L}_{f,\lambda}^{-1}\right\| \leq 8t(t+1)\left(\frac{\max_j \theta_j C(\cdot,\ell_j)+\lambda_f}{\lambda_f}\right)^2 \leq 16t\left(\frac{\max_j \theta_j C(\cdot,\ell_j)+\lambda_f}{\lambda_f}\right)^2,$$

因为 $t \leq 1$。

同理，我们可以证明

$$\Pr\left\{\left\|D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\right\| \geq \sqrt{1+t}\right\} \leq \Pr\left\{\frac{\max_i d_{i\cdot}+\lambda_w}{\max_k \frac{1}{n_k}C(k,\cdot)+\lambda_w} \geq 1+t\right\} \leq \frac{n}{n+p}\epsilon = \frac{1}{1+\gamma}\epsilon.$$

因此，以至少 $1-\frac{3+4\gamma}{1+\gamma}\epsilon-\frac{1}{1+\gamma}\epsilon = 1-4\epsilon$ 的概率，我们有

$$\left\|L_{w,\lambda}^{-1}-\mathfrak{L}_{w,\lambda}^{-1}\right\| \leq 16t\left(\frac{\max_k \frac{1}{n_k}C(k,\cdot)+\lambda_w}{\lambda_w}\right)^2.$$

### *A.6.* 定理的证明*4.4*

我们有

$$\left\|\widetilde{L}_{f,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{f,\lambda}^{-1}\right\| \leq \left\|\mathfrak{D}_{f,\lambda}^{-1/2}\left(L_{f,\lambda}^{-1}-\mathfrak{L}_{f,\lambda}^{-1}\right)\mathfrak{D}_{f,\lambda}^{-1/2}\right\| + \left\|\widetilde{L}_{f,\lambda}^{-1}-\mathfrak{D}_{f,\lambda}^{-1/2}L_{f,\lambda}^{-1}\mathfrak{D}_{f,\lambda}^{-1/2}\right\|.$$

第一项有如下界：

$$\left\|\mathfrak{D}_{f,\lambda}^{-1/2}\left(L_{f,\lambda}^{-1}-\mathfrak{L}_{f,\lambda}^{-1}\right)\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \leq \left\|\mathfrak{D}_{f,\lambda}^{-1}\right\|\left\|L_{f,\lambda}^{-1}-\mathfrak{L}_{f,\lambda}^{-1}\right\| \leq M_f\left\|L_{f,\lambda}^{-1}-\mathfrak{L}_{f,\lambda}^{-1}\right\|.$$

第二项可写为:

$$\begin{aligned}\left\|\widetilde{L}_{f,\lambda}^{-1}-\mathfrak{D}_{f,\lambda}^{-1/2}L_{f,\lambda}^{-1}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| &= \left\|\widetilde{L}_{f,\lambda}^{-1}\left(I_p-D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right)+\left(I_p-\mathfrak{D}_{f,\lambda}^{-1/2}D_{f,\lambda}^{1/2}\right)\widetilde{L}_{f,\lambda}^{-1}D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \\ &\le \left\|\widetilde{L}_{f,\lambda}^{-1}\right\|\left\|I_p-D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\|\left(1+\left\|D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\|\right).\end{aligned}$$

我们已经证明

$$\begin{aligned}\Pr\left\{\left\|L_{f,\lambda}^{-1}-\mathfrak{L}_{f,\lambda}^{-1}\right\|>16t\left(\frac{\overline{\delta}_f+\lambda_f}{\lambda_f}\right)^2\right\} &\le \frac{3+5\gamma}{1+\gamma}\epsilon, \\ \Pr\left\{\frac{\max_j d_{\cdot j}+\lambda_f}{\overline{\delta}_f+\lambda_f}>t+1\right\} &\le \frac{p\epsilon}{n+p}=\frac{\gamma}{1+\gamma}\epsilon, \\ \Pr\left\{\left\|I_p-D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\|>1-\sqrt{1-t}\right\} &\le \frac{2p\epsilon}{n+p}=\frac{2\gamma}{1+\gamma}\epsilon, \\ \Pr\left\{\left\|D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\|>\sqrt{1+t}\right\} &\le \frac{p}{n+p}\epsilon=\frac{\gamma}{1+\gamma}\epsilon,\end{aligned}$$

其中 $\overline{\delta}_f = \text{最大值}_j\,\theta_j C(\cdot,\ \ell_j)$ 。根据引理 2.3, $\left\|L_f^{-1}{}_{\cdot,\lambda}\right\| \le \frac{\text{最大值}_j\, d_{\cdot j}+\lambda_f}{\lambda_f}$. 因此，以至少 $1-\frac{\gamma}{1+\gamma}\epsilon$ 的概率,

$$\begin{aligned}&\left\|\widetilde{L}_{f,\lambda}^{-1}\right\| = \left\|D_{f,\lambda}^{-1/2}L_{f,\lambda}^{-1}D_{f,\lambda}^{-1/2}\right\| \le \left\|D_{f,\lambda}^{-1}\right\|\left\|L_{f,\lambda}^{-1}\right\| \\ &\le \frac{1}{\lambda_f}\frac{\max_j d_{\cdot j}+\lambda_f}{\lambda_f} = \frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}\frac{\max_j d_{\cdot j}+\lambda_f}{\overline{\delta}_f+\lambda_f} \le \frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}(t+1) \le 2\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}.\end{aligned}$$

并且，以至少

$$1-\frac{3+5\gamma}{1+\gamma}\epsilon-\frac{(1+2+1)\gamma}{1+\gamma}\epsilon = 1-\frac{3+9\gamma}{1+\gamma}\epsilon,$$

我们可以界定

$$\left\|\widetilde{L}_{f,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{f,\lambda}^{-1}\right\| \le M_f 16t\left(\frac{\overline{\delta}_f+\lambda_f}{\lambda_f}\right)^2+\frac{1}{\lambda_f}\frac{\overline{\delta}_f+\lambda_f}{\lambda_f}5t = \left(16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}+5\right)\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}t.$$

类似地，采用对称的记号，以至少 $1-\frac{1}{1+\gamma}\epsilon$, $\left\|\widetilde{L}_w^{-1}{}_{\cdot,\lambda}\right\| \le 2\frac{\delta_w+\lambda_w}{\lambda_w^2}$ 的概率,其中 $\delta_w = \text{最大值}_k \frac{1}{n_k} C(\bar{k},\ \cdot)$。并且,以至少$1-4\epsilon-\frac{4}{1+\gamma}\epsilon = 1-\frac{8+4\gamma}{1+\gamma}\epsilon$的概率，我们有

$$\left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\| \le \left(16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}+5\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t,$$

### *A.7.* 定理*5.1*

我们首先证明两个中间不等式。

**的集中界** $\left\|BD_f^{-1}{}_{,\lambda} - \mathfrak{B}\mathfrak{D}_f^{-1}{}_{,\lambda}\right\|$ **和** $\left\|BD_w^{-1}{}_{,\lambda} - \mathfrak{B}\mathfrak{D}_w^{-1}{}_{,\lambda}\right\|$ 我们可以界定

$$\begin{aligned}\left\|BD_{f,\lambda}^{-1} - \mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| &= \left\|(B-\mathfrak{B})\mathfrak{D}_{f,\lambda}^{-1} + B\left(D_{f,\lambda}^{-1} - \mathfrak{D}_{f,\lambda}^{-1}\right)\right\| \\ &\le \left\|(B-\mathfrak{B})\mathfrak{D}_{f,\lambda}^{-1}\right\| + \left\|BD_{f,\lambda}^{-1} - B\mathfrak{D}_{f,\lambda}^{-1}\right\|.\end{aligned}$$

首先， $E_\lambda = D_w^{-1/2}{}_{,\lambda}\ BD_{f\lambda}^{-1/2}{}_{,}$ ，

$$\begin{aligned}\left\|BD_{f,\lambda}^{-1} - B\mathfrak{D}_{f,\lambda}^{-1}\right\| &= \left\|BD_{f,\lambda}^{-1}\left(I_p - D_{f,\lambda}\mathfrak{D}_{f,\lambda}^{-1}\right)\right\| \le \left\|BD_{f,\lambda}^{-1}\right\|\left\|I_p - D_{f,\lambda}\mathfrak{D}_{f,\lambda}^{-1}\right\| \\ &\le \left\|D_{w,\lambda}^{1/2}E_\lambda D_{f,\lambda}^{-1/2}\right\|\left\|I_p - D_{f,\lambda}\mathfrak{D}_{f,\lambda}^{-1}\right\| \\ &\le \left\|D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\right\|\left\|\mathfrak{D}_{w,\lambda}^{1/2}E_\lambda D_{f,\lambda}^{-1/2}\right\|\left\|I_p - D_{f,\lambda}\mathfrak{D}_{f,\lambda}^{-1}\right\|,\end{aligned}$$

其中

$$\left\|\mathfrak{D}_{w,\lambda}^{1/2}E_\lambda D_{f,\lambda}^{-1/2}\right\| \le \left\|\mathfrak{D}_{w,\lambda}^{1/2}\right\|\|E_\lambda\|\left\|D_{f,\lambda}^{-1/2}\right\| \le \sqrt{\frac{\overline{\delta}_w + \lambda_w}{\lambda_f}},$$

如 $\|E_\lambda\| \le 1$。其余项可按前述定理进行界定：

$$\begin{aligned}\Pr\left\{\left\|I_p - D_{f,\lambda}^{1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \ge 1 - \sqrt{1-t}\right\} &\le \frac{2p\epsilon}{n+p} = \frac{2\gamma}{1+\gamma}\epsilon, \\ \Pr\left\{\left\|D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\right\| \ge \sqrt{1+t}\right\} &\le \frac{n}{n+p}\epsilon = \frac{1}{1+\gamma}\epsilon.\end{aligned}$$

其次，我们转而处理项 $\left\|(B-\mathfrak{B})\mathfrak{D}_f^{-1}{}_{,\lambda}\right\|$ 并按定理 4.2中的方法对其进行界定。令

$$X_{ij} = \frac{d_{ij} - p_{ij}}{\theta_j C(\cdot,\ell_j) + \lambda_j}\Delta_{ij}.$$

于是， $\|X_{ij}\| \le M_f$，且

$$\begin{aligned}\left\|\sum\nolimits_{ij}\mathbb{E}(X_{ij}^\top X_{ij})\right\| &= \max_j \sum_i \frac{p_{ij}(1-p_{ij})}{\left(\theta_j C(\cdot,\ell_j) + \lambda_f\right)^2} \le \max_j \sum_i \frac{p_{ij}}{\left(\theta_j C(\cdot,\ell_j) + \lambda_f\right)^2} \\ &= \max_j \frac{\theta_j C(\cdot,\ell_j)}{\left(\theta_j C(\cdot,\ell_j) + \lambda_f\right)^2} \le M_f,\end{aligned}$$

而

$$\begin{aligned}&\left\|\sum\nolimits_{ij}\mathbb{E}(X_{ij}X_{ij}^\top)\right\| = \max_i \sum_j \frac{p_{ij}(1-p_{ij})}{\left(\theta_j C(\cdot,\ell_j) + \lambda_f\right)^2} \le \max_i \sum_j \frac{p_{ij}}{\left(\theta_j C(\cdot,\ell_j) + \lambda_f\right)^2} \\ &= \max_k \sum_j \frac{\frac{1}{n_k}\theta_j C(k,\ell_j)}{\left(\theta_j C(\cdot,\ell_j) + \lambda_f\right)^2} \le M_f^2 \max_k \sum_j \frac{1}{n_k}\theta_j C(k,\ell_j) = M_f^2 \max_k \frac{1}{n_k}C(k,\cdot).\end{aligned}$$

然后我们可以应用矩阵 Bernstein 不等式。对于所有 $t_f \geq 0$,

$$\Pr\left\{\left\|(B-\mathfrak{B})\mathfrak{D}_{f,\lambda}^{-1}\right\| \geq t_f\right\} \leq (n+p)\exp\left(\frac{-t_f^2/2}{M_f\chi_f + M_f t_f/3}\right)$$

其中 $\chi_f = \max\left(1, M_f\delta_{\underline{w}}\right) = \max\left(1, \frac{\delta_w}{\underline{\delta}_f+\lambda_f}\right)$。令

$$t_f = \frac{M_f}{3}\ln\frac{n+p}{\epsilon} + \sqrt{\left(\frac{M_f}{3}\ln\frac{n+p}{\epsilon}\right)^2 + 2\chi_f M_f \ln\frac{n+p}{\epsilon}}$$

为.....的正根

$$(n+p)\exp\left(\frac{-t_f^2/2}{M_f\chi_f + M_f t_f/3}\right) = \epsilon.$$

由于 $t = \sqrt{3M\ln\frac{n+p}{\epsilon}} \leq 1$ （因此 $t^4 \leq t^2 \leq t$），且由于 $\chi_f \geq 1$ （因此 $\chi_f \geq \sqrt{\chi_f}$），我们有

$$t_f \leq \frac{t^2}{9} + \sqrt{\frac{t^4}{9^2} + \frac{2t^2\chi_f}{3}} \leq \frac{t}{9}\left(1+\sqrt{1+6\chi_f}\right) \leq t\sqrt{\chi_f}.$$

最后，以至少.....的概率

$$1 - \epsilon - \frac{1+2\gamma}{1+\gamma}\epsilon = 1 - \frac{2+3\gamma}{1+\gamma}\epsilon,$$

我们有

$$\left\|BD_{f,\lambda}^{-1} - \mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| \leq t_f + \sqrt{1+t}\left(1-\sqrt{1-t}\right)\sqrt{\frac{\bar{\delta}_w+\lambda_w}{\lambda_f}} \leq \left(\sqrt{\chi_f} + 2\sqrt{\frac{\bar{\delta}_w+\lambda_w}{\lambda_f}}\right)t.$$

类似地定义 $t_w$,，则有 $t_w \leq t\sqrt{\chi_w}$，其中 $\chi_w := \max(1, M_w\delta_f)$,，并且以至少 $1 - \frac{3+2\gamma}{1+\gamma}\epsilon$, 的概率有

$$\left\|B^\top D_{w,\lambda}^{-1} - \mathcal{B}^\top\mathfrak{D}_{w,\lambda}^{-1}\right\| \leq t_w + \sqrt{1+t}\left(1-\sqrt{1-t}\right)\sqrt{\frac{\bar{\delta}_f+\lambda_f}{\lambda_w}} \leq \left(\sqrt{\chi_w} + 2\sqrt{\frac{\bar{\delta}_f+\lambda_f}{\lambda_w}}\right)t$$

**.....的界** $\left\|\mathfrak{B}\mathfrak{D}_f^{-1}{}_{,\lambda}\right\|$ **和** $\left\|BD_{f\lambda}^{-1}\right\|$**，以及** $\left\|\mathfrak{B}^\top\mathfrak{D}_w^{-1}{}_{,\lambda}\right\|$ **和** $\left\|B^\top D_w^{-1}{}_{,\lambda}\right\|$我们有

$$\left\|\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| = \left\|\mathfrak{D}_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1/2}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \leq \left\|\mathfrak{D}_{w,\lambda}^{1/2}\right\|\|\mathfrak{E}_\lambda\|\left\|\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \leq \sqrt{\frac{\bar{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}},$$

and

$$\begin{aligned}\left\|BD_{f,\lambda}^{-1}\right\| &= \left\|D_{w,\lambda}^{1/2}E_\lambda D_{f,\lambda}^{-1/2}\right\| = \left\|D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\mathfrak{D}_{w,\lambda}^{1/2}E_\lambda D_{f,\lambda}^{-1/2}\right\| \\ &\leq \left\|D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\right\|\left\|\mathfrak{D}_{w,\lambda}^{1/2}\right\|\|E_\lambda\|\left\|D_{f,\lambda}^{-1/2}\right\| \leq \left\|D_{w,\lambda}^{1/2}\mathfrak{D}_{w,\lambda}^{-1/2}\right\|\sqrt{\frac{\bar{\delta}_w+\lambda_w}{\lambda_f}},\end{aligned}$$

其中 $\Pr\left\{\left\|D_w^{1/2},_\lambda \mathfrak{D}_w^{-1/2},_\lambda\right\| \geq \sqrt{1+t}\right\} \leq \frac{1}{1+\gamma}\epsilon$。因此，以至少$1-\frac{1}{1+\gamma}\epsilon$的概率，

$$\left\|BD_{f,\lambda}^{-1}\right\| \leq \sqrt{1+t}\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}} \leq 2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}.$$

类似地，我们有 $\left\|\mathfrak{B}^\top \mathfrak{D}_w^{-1},_\lambda\right\| \leq \sqrt{\frac{\delta_f+\lambda_f}{\underline{\delta}_w+\lambda_w}}$，并且以至少 $1-\frac{\gamma}{1+\gamma}\epsilon$ 的概率，$\left\|B^\top D_w^{-1},_\lambda\right\| \leq 2\sqrt{\frac{\delta_f+\lambda_f}{\lambda_w}}$。

**定理 5.1 的主要论证** 关于 $\widehat{\mu}$的偏差为

$$b_{\mu,\lambda} = \mathbb{E}\left(\widehat{\mu}-\mu \mid X,Z,\beta^*\right) = \widetilde{L}_{w,\lambda}^{-1}\left(-\lambda_w Z_w\mu^* + \lambda_f BD_{f,\lambda}^{-1} Z_f\phi^*\right).$$

$\mu$ 上的“总体”偏差为

$$\mathfrak{b}_{\mu,\lambda} = \widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\left(-\lambda_w Z_w\mu^* + \lambda_f \mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1} Z_f\phi^*\right).$$

那么（对于欧几里得范数），

$$\|b_{\mu,\lambda}-\mathfrak{b}_{\mu,\lambda}\| \leq \lambda_w\left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|\|Z_w\mu^*\| + \lambda_f\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|\|Z_f\phi^*\|.$$

我们已经有

$$\Pr\left\{\left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\| > \left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t\right\} \leq \frac{8+4\gamma}{1+\gamma}\epsilon,$$

$$\Pr\left\{\left\|\widetilde{L}_{w,\lambda}^{-1}\right\| > 2\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}\right\} \leq \frac{1}{1+\gamma}\epsilon,$$

$$\Pr\left\{\left\|BD_{f,\lambda}^{-1}-\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| > \left(\sqrt{\chi_f}+2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right)t\right\} \leq \frac{2+3\gamma}{1+\gamma}\epsilon,$$

and

$$\left\|\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| = \left\|\mathfrak{D}_{w,\lambda}^{1/2}\mathfrak{E}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| < \sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}.$$

此外，

$$\begin{aligned}\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| &\leq \left\|\left(\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right)\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}+\widetilde{L}_{w,\lambda}^{-1}\left(BD_{f,\lambda}^{-1}-\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right)\right\| \\ &\leq \left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|\left\|\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| + \left\|\widetilde{L}_{w,\lambda}^{-1}\right\|\left\|BD_{f,\lambda}^{-1}-\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|.\end{aligned}$$

因此，以至少 $1-\frac{8+4\gamma}{1+\gamma}\epsilon-\frac{1}{1+\gamma}\epsilon-\frac{2+3\gamma}{1+\gamma}\epsilon = 1-\frac{11+7\gamma}{1+\gamma}\epsilon$的概率，

$$\|b_{\mu,\lambda}-\mathfrak{b}_{\mu,\lambda}\| \leq \left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t\left(\lambda_w\sqrt{n}\|\mu^*\| + \lambda_f\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}\sqrt{p}\|\phi^*\|\right)$$
$$+ 2\lambda_f\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}\left(\sqrt{\chi_f}+2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right)t\sqrt{p}\|\phi^*\|.$$

最后，请注意

$$\left\|\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\| = \left\|\mathfrak{D}_{w,\lambda}^{-1/2}\mathfrak{L}_{w,\lambda}^{-1}\mathfrak{D}_{w,\lambda}^{-1/2}\right\| \leq \left\|\mathfrak{D}_{w,\lambda}^{-1}\right\|\left\|\mathfrak{L}_{w,\lambda}^{-1}\right\| = \frac{1}{\underline{\delta}_w+\lambda_w}\frac{\overline{\delta}_w+\lambda_w}{\lambda_w}.$$

因此，

$$\begin{aligned}\|\mathfrak{b}_{\mu,\lambda}\| &\leq \left\|\mathfrak{L}_{w,\lambda}^{-1}\right\|\left\|-\lambda_w Z_w\mu^* + \lambda_f\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}Z_f\phi^*\right\| \\ &\leq \frac{1}{\underline{\delta}_w+\lambda_w}\frac{\overline{\delta}_w+\lambda_w}{\lambda_w}\left(\lambda_w\sqrt{n}\,\|\mu^*\| + \lambda_f\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}\sqrt{p}\,\|\phi^*\|\right).\end{aligned}$$

类似地，以至少 $1-\frac{3+9\gamma}{1+\gamma}\epsilon-\frac{\gamma}{1+\gamma}\epsilon-\frac{3+2\gamma}{1+\gamma}\epsilon = 1-\frac{6+12\gamma}{1+\gamma}\epsilon$ 的概率，有

$$\begin{aligned}\|b_{\phi,\lambda}-\mathfrak{b}_{\phi,\lambda}\| \leq{}& \left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}t\left(\lambda_f\sqrt{p}\,\|\phi^*\| + \lambda_w\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_w+\lambda_w}}\sqrt{n}\,\|\mu^*\|\right) \\ &+ 2\lambda_w\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}\left(\sqrt{\chi_w}+2\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\lambda_w}}\right)t\sqrt{n}\,\|\mu^*\|.\end{aligned}$$

此外，

$$\left\|\widetilde{\mathfrak{L}}_{f,\lambda}^{-1}\right\| = \left\|\mathfrak{D}_{f,\lambda}^{-1/2}\mathfrak{L}_{f,\lambda}^{-1}\mathfrak{D}_{f,\lambda}^{-1/2}\right\| \leq \left\|\mathfrak{D}_{f,\lambda}^{-1}\right\|\left\|\mathfrak{L}_{f,\lambda}^{-1}\right\| = \frac{1}{\underline{\delta}_f+\lambda_f}\frac{\overline{\delta}_f+\lambda_f}{\lambda_f}.$$

因此，我们还可以得到

$$\begin{aligned}\|\mathfrak{b}_{\phi,\lambda}\| &\leq \left\|\mathfrak{L}_{f,\lambda}^{-1}\right\|\left\|-\lambda_f Z_f\phi^* + \lambda_w\mathfrak{B}^\top\mathfrak{D}_{w,\lambda}^{-1}Z_w\mu^*\right\| \\ &\leq \frac{1}{\underline{\delta}_f+\lambda_f}\frac{\overline{\delta}_f+\lambda_f}{\lambda_f}\left(\lambda_f\sqrt{p}\,\|\phi^*\| + \lambda_w\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_w+\lambda_w}}\sqrt{n}\,\|\mu^*\|\right).\end{aligned}$$

### *A.8.* 定理 *5.2* 的证明

$\widehat{\mu}-\mu$ 的方差为

$$V_{w,\lambda} = \sigma^2\widetilde{L}_{w,\lambda}^{-1} + \left(\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right)\widetilde{L}_{w,\lambda}^{-2} + \left(\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right)\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-2}B^\top\widetilde{L}_{w,\lambda}^{-1}.$$

“总体”方差为

$$\mathfrak{V}_{w,\lambda} = \sigma^2\widetilde{\mathfrak{L}}_{w,\lambda}^{-1} + \left(\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right)\widetilde{\mathfrak{L}}_{w,\lambda}^{-2} + \left(\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right)\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-2}\mathfrak{B}^\top\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}.$$

两边相减，得到

$$\begin{aligned}\|V_{w,\lambda}-\mathfrak{V}_{w,\lambda}\| \leq{}& \sigma^2\left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\| + \left|\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right|\left\|\widetilde{L}_{w,\lambda}^{-2}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-2}\right\| \\ &+ \left|\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right|\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-2}B^\top\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-2}\mathfrak{B}^\top\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|.\end{aligned}$$

我们已经知道

$$\Pr\left\{\left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|>\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t\right\}\leq\frac{8+4\gamma}{1+\gamma}\epsilon,$$

$$\Pr\left\{\left\|\widetilde{L}_{w,\lambda}^{-1}\right\|>2\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}\right\}\leq\frac{1}{1+\gamma}\epsilon,$$

$$\Pr\left\{\left\|BD_{f,\lambda}^{-1}-\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|>\left(\sqrt{\chi_f}+2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right)t\right\}\leq\frac{2+3\gamma}{1+\gamma}\epsilon,$$

$$\Pr\left\{\left\|BD_{f,\lambda}^{-1}\right\|>2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right\}\leq\frac{1}{1+\gamma}\epsilon,$$

并且

$$\left\|\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|=\left\|\mathfrak{D}_{w,\lambda}^{-1/2}\mathfrak{L}_{w,\lambda}^{-1}\mathfrak{D}_{w,\lambda}^{-1/2}\right\|\leq\left\|\mathfrak{D}_{w,\lambda}^{-1}\right\|\left\|\mathfrak{L}_{w,\lambda}^{-1}\right\|=\frac{1}{\underline{\delta}_w+\lambda_w}\frac{\overline{\delta}_w+\lambda_w}{\lambda_w},$$

$$\left\|\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|=\left\|\mathfrak{D}_{w,\lambda}^{1/2}\mathfrak{E}\mathfrak{D}_{f,\lambda}^{-1/2}\right\|<\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}.$$

由此可知，以至少

$$1-\frac{8+4\gamma}{1+\gamma}\epsilon-\frac{1}{1+\gamma}\epsilon-\frac{2+3\gamma}{1+\gamma}\epsilon-\frac{1}{1+\gamma}\epsilon=1-\frac{12+7\gamma}{1+\gamma}\epsilon,$$

我们有以下不等式。首先,

$$\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|\leq\left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|\left\|\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|+\left\|\widetilde{L}_{w,\lambda}^{-1}\right\|\left\|BD_{f,\lambda}^{-1}-\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|$$

$$\leq\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}+2\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}\left(\sqrt{\chi_f}+2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right)t.$$

其次,

$$\left\|\widetilde{L}_{w,\lambda}^{-2}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-2}\right\|\leq\left(\left\|\widetilde{L}_{w,\lambda}^{-1}\right\|+\left\|\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|\right)\left\|\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|$$

$$\leq\frac{\overline{\delta}_w+\lambda_w}{\lambda_w}\left(\frac{2}{\lambda_w}+\frac{1}{\underline{\delta}_w+\lambda_w}\right)\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t.$$

第三,

$$\begin{aligned}
&\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-2}B^{\top}\widetilde{L}_{w,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-2}\mathfrak{B}^{\top}\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\| \\
&\quad\leq\left(\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-1}\right\|+\left\|\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|\right)\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| \\
&\quad\leq\left(\left\|\widetilde{L}_{w,\lambda}^{-1}\right\|\left\|BD_{f,\lambda}^{-1}\right\|+\left\|\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\right\|\left\|\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\|\right)\left\|\widetilde{L}_{w,\lambda}^{-1}BD_{f,\lambda}^{-1}-\widetilde{\mathfrak{L}}_{w,\lambda}^{-1}\mathfrak{B}\mathfrak{D}_{f,\lambda}^{-1}\right\| \\
&\quad\leq\left(2\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}+\frac{1}{\underline{\delta}_w+\lambda_w}\frac{\overline{\delta}_w+\lambda_w}{\lambda_w}\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}\right) \\
&\times\left[\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}+2\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}\left(\sqrt{\chi_f}+2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right)\right]t.
\end{aligned}$$

最后,

$$\begin{aligned}
\|V_{w,\lambda}-\mathfrak{V}_{w,\lambda}\|&\leq\sigma^2\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\overline{\delta}_w+\lambda_w}{\lambda_w^2}t \\
&+\left|\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right|\left(\frac{2}{\lambda_w}+\frac{1}{\underline{\delta}_w+\lambda_w}\right)\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\frac{\left(\overline{\delta}_w+\lambda_w\right)^2}{\lambda_w^3}t \\
&+\left|\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right|\left(4\frac{1}{\lambda_w}\sqrt{\frac{1}{\lambda_f}}+\frac{1}{\underline{\delta}_w+\lambda_w}\sqrt{\frac{1}{\underline{\delta}_f+\lambda_f}}\right)\frac{\left(\overline{\delta}_w+\lambda_w\right)^{5/2}}{\lambda_w^3} \\
&\times\left[\left(5+16\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_w+\lambda_w}\right)\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\underline{\delta}_f+\lambda_f}}+2\left(\sqrt{\chi_f}+2\sqrt{\frac{\overline{\delta}_w+\lambda_w}{\lambda_f}}\right)\right]t.
\end{aligned}$$

类似地，以至少

$$1-\frac{3+9\gamma}{1+\gamma}\epsilon-\frac{\gamma}{1+\gamma}\epsilon-\frac{3+2\gamma}{1+\gamma}\epsilon-\frac{\gamma}{1+\gamma}\epsilon=1-\frac{5+13\gamma}{1+\gamma}\epsilon,$$

我们有

$$\begin{aligned}
\|V_{f,\lambda}-\mathfrak{V}_{f,\lambda}\|&\leq\sigma^2\left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\frac{\overline{\delta}_f+\lambda_f}{\lambda_f^2}t \\
&+\left|\lambda_f^2\sigma_f^2-\lambda_f\sigma^2\right|\left(\frac{2}{\lambda_f}+\frac{1}{\underline{\delta}_f+\lambda_f}\right)\left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\frac{\left(\overline{\delta}_f+\lambda_f\right)^2}{\lambda_f^3}t \\
&+\left|\lambda_w^2\sigma_w^2-\lambda_w\sigma^2\right|\left(4\frac{1}{\lambda_f}\sqrt{\frac{1}{\lambda_w}}+\frac{1}{\underline{\delta}_f+\lambda_f}\sqrt{\frac{1}{\underline{\delta}_w+\lambda_w}}\right)\frac{\left(\overline{\delta}_f+\lambda_f\right)^{5/2}}{\lambda_f^3} \\
&\times\left[\left(5+16\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_f+\lambda_f}\right)\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\underline{\delta}_w+\lambda_w}}+2\left(\sqrt{\chi_w}+2\sqrt{\frac{\overline{\delta}_f+\lambda_f}{\lambda_w}}\right)\right]t.
\end{aligned}$$

证明至此结束。

**致谢**

作者谨向匿名审稿人、一位副编辑和编辑致谢，感谢他们提出建设性意见，提升了本文的质量。本文的初稿曾于2023年在布里斯托尔举行的计量经济学研究小组会议上报告。较为完善的版本曾在2024年里约计量经济学会议以及2025年5月纪念Thierry Magnac的会议上报告，Koen Jochmans在会上提出了有益的意见。我们感谢2025年5月芝加哥大学研讨会的与会者提出宝贵意见和建议。

**资助**

第二作者感谢欧洲研究理事会的支持（资助编号：ERC-2020-ADG-101018130）。

**参考文献**

Abowd, John M., Francis Kramarz, and David N. Margolis(1999): “高工资工人与高工资企业”，《计量经济学》，67 (2), 251–334.Abowd, John M., Kevin L. McKinney, and Ian M. Schmutte(2019): “收入决定中的内生流动建模”，《商业与经济统计学杂志》，37 (3), 405–418. Andrews, M. J., L. Gill, T. Schank, and R. Upward(2008): “高工资工人与低工资企业：负向择偶匹配还是有限流动性偏误？”《英国皇家统计学会会刊：*A*辑（社会统计学）》，171 (3), 673–697.Azkarate-Askasua, Miren and Miguel Zerecero(2022): “校正含有众多协变量的线性模型中的小样本偏误”，Crc tr 224讨论论文系列，波恩大学和曼海姆大学，德国。Babet, Damien, Olivier Godechot, and Marco G Palladino(2022): “在AKM之乡：解释法国工资不平等的动态”，工作论文或预印本。Bigot, Jérémie, Issa-Mbenard Dabo, and Camille Male(2024): “方差轮廓下非独立同分布数据的岭回归高维分析”，。Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa(2019): “雇主—雇员匹配数据的分布框架”，《计量经济学》，87 (3), 699–739.Chaudhuri, Kamalika, Fan Chung Graham, and Alexander Tsiatas(2012): “扩展植入划分模型中具有一般度数的图的谱聚类”，载于计算学习理论年会。Chung, F.R.K. (1997): 谱图论，CBMS区域会议丛书第92号，数学科学会议委员会。

Chung, Fan and Mary Radcliffe (2011): “关于一般随机图的谱”，电子组合学杂志，18 (1).Dall’Amico, Lorenzo, Romain Couillet, and Nicolas Tremblay(2021): “稀疏图谱聚类的统一框架”，机器学习研究杂志，22 (1).Dasgupta, A., J.E. Hopcroft, and F. McSherry (2004): “具有偏斜度分布的随机图的谱分析”，载于第45届IEEE计算机科学基础年会，602–610.Dicker, Lee H. (2013): “具有任意预测变量协方差的高维线性模型的最优等变预测”，电子统计学杂志，7 (none), 1806 – 1834.Dobriban, Edgar and Stefan Wager(2018): “预测的高维渐近理论：岭回归与分类”，统计年鉴，46 (1), 247 – 279.El Karoui, Noureddine (2018): “预测变量几何结构对高维岭正则化广义稳健回归估计量性能的影响”，概率论及其相关领域，170 (1), 95–175.Gaure, Simen (2014): “双向固定效应线性回归中的相关偏差校正”，Stat，3 (1), 379–390.Hastie, Trevor (2020): “岭正则化：数据科学中的基本概念”，Technometrics，62 (4), 426–433, pMID: 36033922.
Jochmans, Koen and Martin Weidner (2019): “网络数据上的固定效应回归”，《计量经济学》，87 (5), 1543–1560.Karrer, Brian and M. E. J. Newman (2011): “网络中的随机块模型与社区结构”，Physical Review E，83 (1).Kline, Patrick, Raffaele Saggio, and Mikkel Soelvsten(2020): “方差分量的留出估计”，《计量经济学》，88(5),1859– 1898.Larremore, Daniel B., Aaron Clauset, and Abigail Z. Jacobs(2014): “高效推断二部网络中的社区结构”，Phys. Rev. E，90, 012805.Lentz, Rasmus, Suphanit Piyapromdee, and Jean-Marc Robin(2023): “排序的剖析——丹麦数据的证据”，《计量经济学》，91 (6), 2409–2455.Nimczik, Jan Sebastian(2017): “工作流动网络与内生劳动力市场”，VfS 2017年年会（维也纳）：货币与银行业的替代结构168147，Verein fur Socialpolitik / 德国经济学会。Patil, Pratik, Jin-Hong Du, and Ryan J. Tibshirani(2025): “分布外预测的最优岭正则化”，载于第41届国际机器学习会议论文集，JMLR.org，ICML’24.Qin, Tai and Karl Rohe (2013): “度修正随机块模型下的正则化谱聚类”，载于第26届神经信息处理系统国际会议论文集——第2卷，美国纽约州雷德胡克：Curran Associates Inc.，NIPS’13, 3120–3128.Razaee, Zahra S., Arash A. Amini, and Jingyi Jessica Li(2019):

“带协变量的匹配二部块模型”，机器学习研究期刊，20（34），1–44。Richards, Dominic, Jaouad Mourtada, and Lorenzo Rosasco（2020）：“一般源条件下岭（无）回归的渐近性”，载于国际人工智能与统计会议。Tropp, Joel A.（2015）：“矩阵集中不等式导论”，机器学习基础与趋势，8（1-2），1–230。von Luxburg, Ulrike（2007）：“谱聚类教程。”统计计算，17（4），395–416。Wu, Denny and Ji Xu（2020）：“过参数化线性回归中的最优加权 \ell 2 正则化”，载于神经信息处理系统进展，H. Larochelle、M. Ranzato、R. Hadsell、M.F. Balcan 和 H. Lin 编，Curran Associates, Inc.，第 33 卷，10112–10123。Zhang, Yilin and Karl Rohe（2018）：“通过图导通率理解正则化谱聚类”，载于神经信息处理系统进展，S. Bengio、H. Wallach、H. Larochelle、K. Grauman、N. Cesa-Bianchi 和 R. Garnett 编，Curran Associates, Inc.，第 31 卷。Zhou, Tao, Jie Ren, Matus Medo, and Yi-Cheng Zhang（2007）：“二部网络投影与个性化推荐”，物理评论*E*，76，046115。