arXiv:2602.04751v1 [stat.CO] 4 Feb 2026

极端值下的多重插补方法

Enzo Porto Brasil[1]

摘要

缺失数据在实证数据库中无处不在，但统计分析通常需要完整的数据矩阵。多重插补为填补这些空白提供了一种有原则的解决方案。本研究使用R语言中的MICE软件包，评估了若干多重插补方法在存在和不存在极端值时的表现。通过蒙特卡罗模拟，我们生成了包含三个变量的不完整数据集，并在回归模型中评估了每种插补方法。结果表明，基于线性回归的插补方法展现出最佳的总体预测性能（交叉验证均方误差），而稀疏模型方法通常效率较低。我们的发现强调了在选择插补策略时极端值的重要性，并突出样本量、缺失比例、极端值的存在以及拟合模型的类型是影响性能的关键决定因素。尽管存在局限性，本研究仍为研究人员提供了实用建议，强调在选择插补方法之前，需要检查缺失机制以及极端值的出现情况。

关键词 缺失数据，多重插补，统计建模，极端观测值，离群值污染，蒙特卡洛模拟，软件R语言

1 引言

近年来，大规模数据以前所未有的方式得以获取，推动了不同学科领域的科学与技术进步。然而，信息量不断增长并不意味着数据完整或准确。缺失数据仍是普遍存在的挑战，其成因多种多样，例如调查中的无应答、传感器故障、数据录入错误或历史记录丢失。如果不加以处理，缺失可能会扭曲参数估计、降低统计功效，并使推断结论产生偏差；尤其是在未经适当调整便采用传统统计技术（这类技术通常假设数据完整）时更是如此（Nunes 2007；Rubin 1996）。

处理不完整数据的策略多种多样，从均值替代等临时方法到更为规范的统计框架不等。其中，多重插补（MI）已成为在完全随机缺失（MCAR）假设下、以及更一般地在随机缺失（MAR）假设下处理缺失的一种规范方法（Schafer 1999）。MI用从预测分布中抽取的多个合理估计值替换每个缺失值，从而生成若干完整数据集。这些数据集被分别分析，并将结果合并，以纳入与插补过程相关的不确定性。与单次插补或个案删除相比，MI能提供更为有效的统计推断，

[1]巴西利亚大学统计系，巴西。达西·里贝罗大学校园，A区，北区，L3北路，巴西联邦区，70910-900。通信请联系：<enzoportobrasil@gmail.com>。

在保持变异性的同时降低偏差（McKnight等人，2007；Kenward和Carpenter，2007；VanderHeijden等人，2006年）。
尽管多重插补具有诸多优势，但当存在极端值时，其表现可能会受到影响。尾部观测值无论是真实存在的，还是由误差引起的，都会对模型拟合以及用于插补的预测分布产生较大影响。Kotz 和 Nadarajah（2000年）指出，极端值可能自然产生于某些领域（例如金融、气象学和天体物理研究），也可能源于测量误差。无论属于哪种情形，它们都可能严重影响模型参数，扭曲用于插补的预测分布，并导致结果不稳定或产生偏差（Ferrari 和 Ozaki，2014年；Van Buuren，2018年；Li 等，2024年）。尽管缺失的影响已得到广泛研究，但缺失数据与极端值之间的相互作用仍相对缺乏研究。这一空白在应用研究情境中尤为重要，因为这两个问题可能同时出现，而插补模型的选择通常未考虑其对极端值及其他复杂分布特征（如多峰性和偏度）的敏感性（von Hippel，2013年；Templ，2024年）。

本研究在数据受污染和干净两种条件下，对广泛使用的多重插补（MI）程序进行了全面且可复现的评估。研究采用三变量正态设计（y、x_1、x_2），仅在 x_2中引入完全随机缺失，并通过逐案例污染，在 $\bar{v} \pm 3s_v$ 处替换选定的行，以生成垂直异常值和高杠杆点（Pukelsheim 1994；Robert 1995）。下游模型根据情形进行匹配：干净数据使用普通最小二乘法（OLS），污染数据则使用弹性网络（EN），以在存在杠杆效应和共线性的情况下稳定估计（Zou和Hastie 2005；Hastie等人 2009；Friedman等人 2010b）。在正确设定下，OLS作为效率基准，但对影响高度敏感（Rousseeuw和Leroy 1987）。

实验设计改变了样本量、缺失比例、极端值比例、协变量间相关性、 `mice`中的插补次数、蒙特卡洛重复次数以及分析模型类型。通过使用单一主种子和确定性子流、在每次重复中为各插补固定并重复使用交叉验证折，以及采用用于插补 x_2 的相容预测变量矩阵，确保配对比较和适当的方差核算（Van Buuren 2018；Van Buuren和Groothuis-Oudshoorn 2011；Arlot 和 Celisse 2010）。性能主要通过样本外交叉验证均方误差的均值、方差和分位数进行汇总。在干净情形下，使用Rubin规则（Rubin和Wiley 1987；White 等人 2011），通过合并偏差、均方根误差和（β_0、β_1、β_2）的95%覆盖率评估推断有效性；而在污染情形下，则在模型选择后如实报告系数。

本研究依次探讨四个问题：在存在极端值时，领先的多重插补方法在样本外预测上如何比较；在无污染情形下，推断会有何后果，以及这些指标在污染下如何表现；样本量与缺失程度如何调节表现和尾部风险；以及在污染下，参数多重插补在多大程度上能产生更紧的预测离散度，而供体法或机器学习（ML）程序又能在多大程度上降低斜率偏差。

结果表明，系数中持续存在污染方向性，表现为截距膨胀、 β_1中的系统性倾斜，以及 β_2中的收缩或符号偏移。增大样本量会降低变异性，但无法消除这些偏移。观察到稳定的方法权衡：参数多重插补通常能收窄预测分布尾部，而随着缺失程度增加，供体法和机器学习程序往往能减轻斜率偏差。这些证据表明，当缺失与极端值同时出现时，应审慎选择方法。所有计算均使用 R/RStudio 4.3.3（R Core Team 2023）完成。

本文结构如下。第 2 节介绍材料与方法。第 2.1–2.3 小节阐述数据生成过程、污染机制、统计模型和多重插补程序。第 2.4–2.5 小节说明比较标准、评估方案和可复现性控制。第 3 节报告主要结果。第 4 节阐述结论与启示。附录 A 提供透明性材料，并按样本量和情景列出扩展结果。

2 材料与方法

本节记录了在无需重新了解研究动机的情况下复现本研究所需的各项要素。所有情景均采用配对设计：一个基线清洁数据集及其污染对应数据集；在每次重复中，仅对 x_2 引入完全随机缺失；所有插补和方法均重复使用相同的折叠划分；下游模型按情景固定（干净数据采用普通最小二乘法，污染数据采用弹性网络）。其余设计因素和报告规范已在引言中说明，并在下文各小节中详细介绍。

2.1 数据集

所有数据集均由从正态分布中抽取的伪随机数生成。每个数据集包含三个连续变量，分别记为 y、x_1 和 x_2。在两种基线条件下评估性能：（i）从指定正态模型中抽取且未受污染的数据集；（ii）包含注入的极端值的数据集（污染）。为确保可复现性，所有模拟程序均使用固定随机种子（241103414）初始化，该随机种子不会对结果产生方法学影响。各变量定义如下。

· $\mathbf{y}$：最初生成自标准差为 1.5 的正态分布的连续响应变量，对于第 i个观测，其线性预测变量为 $\mu(y_i)$,

$$\mathbf{y_i} \sim N(\mu(\mathbf{y_i}),\, 1.5^2), \qquad \mu(\mathbf{y_i}) \;=\; 1 \;+\; 0.5\,\mathbf{x_{1i}} \;+\; 1.5\,\mathbf{x_{2i}}. \tag{1}$$

预测变量：

· $\mathbf{x_1}$：连续预测变量，服从 $x_1 \sim N(10, 2_2)$。

· $\mathbf{x_2}$：根据 x_1 有条件地生成的连续预测变量，以引入皮尔逊相关 ρ ：

$$\mathbf{x_{2i}} \;=\; 5 \;+\; \frac{\rho\,\sigma_2}{\sigma_1}(\mathbf{x_{1i}} - 10) \;+\; \sigma_2\sqrt{1-\rho^2}\,\mathbf{z_i}, \tag{2}$$

其中 $\sigma_1 = 2$、$\sigma_2 = 1.5$ 和 $\mathbf{z_i} \sim N(0, 1)$ 与 x_1 独立。

等价 地，（x_1，x_2）服从二元正态分布，均值向量为（10，5），协方差矩 x

$$\Sigma \;=\; \begin{pmatrix} 2^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & 1.5^2 \end{pmatrix}.$$

参数 ρ 控制 x_1 与 x_2 之间的诱导相关性（ $|\rho|$越大，线性关联越强）。

对于每种情景，均考虑两个参考数据集：一个是基线数据集（干净数据，无污染），另一个是含有极端值的受污染数据集。

2.1.1 极端值（污染机制）

令 F_0 表示第2.1节中 $(y、\ x_1、\ x_2)$ 的基线联合分布。对于给定的污染比例 $P_{\text{ext}} \in$（0，1）和样本量 n，令 $k =$ 为 round$(n \times P_{\text{ext}})$，并从中均匀且不放回地抽取一个大小为 k 的索引集合 $\mathcal{I}_{\text{ext}} \subset \{1、\ \ldots\ldots、\ n\}$。在基线（污染前）样本上，对每个 $v \in \{y、\ x_1、\ x_2\}$ 计算样本汇总统计量

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i, \qquad s_v = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(v_i - \bar{v})^2}.$$

任意排列 $\mathcal{I}_{\text{ext}} = \{i_1, \ldots, i_k\}$，并设奇数 $\eta_{i_j} = +1$ 对应的符号为 j，偶数 $\eta_{i_j} = -1$ 对应的符号为 j。对于每个 $i \in \mathcal{I}_{\text{ext}}$，以及每个 $v \in \{y, x_1, x_2\}$，执行分量替换

$$v_i \ \leftarrow\ \bar{v} \ +\ 3\,\eta_i\, s_v,$$

其余观测值保持不变。等价地，受污染抽样服从 $F = (1 - P_{\text{ext}})F_0 + P_{\text{ext}}\, H$，其中 $H = \frac{1}{2}\delta_{\bar{\mathbf{m}}-3\mathbf{s}} + \frac{1}{2}\delta_{\bar{\mathbf{m}}+3\mathbf{s}}$ 按分量作用，$\bar{\mathbf{m}} = (\bar{y}, x_1, x_2)$ 且 $\mathbf{s} = (s_y, s_{x_1}, s_{x_2})$。这种对称的、逐情形的“三西格玛”替换会产生垂直异常值和高杠杆点（Pukelsheim 1994；Kotz 和 Nadarajah 2000）。也可以采用其他基于尾部的设计（例如，超过 $\pm 3s$ 的截断正态）（Robert 1995），但本文报告的所有结果均采用上述 $\pm 3s$ 方案。

2.2 统计模型

如第2.4节所述，我们通过对响应变量 y withx_1 进行回归建模，并将 x_2 作为协变量，比较不同的插补方法。对于不含极端值的参考数据集，拟合普通最小二乘法（OLS）线性回归。对于含有极端值的数据集，则采用基于弹性网（EN）惩罚的稀疏回归，以便在存在污染和诱发共线性的情况下稳定估计。插补后采用相同的建模策略（模型形式和调参方案）：对基线（“标准值”）数据使用线性模型，对受污染（“极端值”）数据使用弹性网模型，从而使性能差异反映插补对后续分析的影响，而非分析模型的变化。

可比性依据。这两种设计（含极端值与不含极端值）在各项分析中保持估计目标和预测变量集一致：两种情况下，工作模型均关于 $(x_1,\ x_2)$ 呈线性，并通过 K 折交叉验证进行调参；每个重复内各次插补均重复使用固定的折分。弹性网（EN）求解惩罚最小二乘问题；当 $\lambda = 0$ 时，它退化为无惩罚最小二乘法。因此，在干净、低维的情形下，经交叉验证的EN通常会选择较小的 λ，其预测值也接近OLS；而其 ℓ_2 成分则能在多重共线性或污染导致方差膨胀时稳定估计（Zou和Hastie 2005年；Hastie等人 2009年；Friedman等人 2010b）。众所周知，普通最小二乘法对高杠杆点和异常值极为敏感（影响无界；崩溃点为 $1/n$），因此，在极端值污染下优先采用EN来控制方差；但需注意，标准EN由于仍采用平方损失，严格来说并不具备稳健性。真正稳健的变体会将损失函数替换为Huber型或最小绝对偏差（LAD）损失，并结合稀疏惩罚（Rousseeuw和Leroy 1987年；Huber和Ronchetti 2009年；Yu和Yao 2017年；Lambert-Lacroix和Zwald 2011年；Wang等人 2007年、2013年）。

符号约定。两个部分中的总体系数均使用相同的符号：$\boldsymbol{\beta} = (\beta_0,\ \beta_1,\ \beta_2)^\top$。其估计量始终简写为 $\hat{}$ β。如需区分，

我们会在文字中说明（例如，“OLS估计值 $\hat{}\ \beta$”与“弹性网络估计值 $\hat{}\ \beta$”）。所有系数均以原始数据尺度报告；为便于比较，惩罚拟合采用内部标准化（斜率标准化，截距不惩罚），并在输出时进行反变换。

无极端值数据：线性回归。 当不存在极端值时，使用 `stats::lm()`（R Core Team 2023）拟合普通最小二乘法（OLS）（Kutner等人 2005）：

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \qquad \epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2),\ i = 1, \ldots, n. \tag{3}$$

设 $\mathbf{y} = (y_1, \ldots, y_n)^\top$、$\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})^\top$、$\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})^\top$，以及 $\mathbf{X} = [\,\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2\,] \in \mathbb{R}^{n \times 3}$。$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ 的OLS估计量为

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}\right)^2 = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \right\}, \tag{4}$$

$\hat{}$ 当$\mathbf{X}$ 具有满列秩时；在实践中，`lm()` 通过带主元QR分解计算 $\boldsymbol{\beta}$，而不是构造 $(\mathbf{X}^\top \mathbf{X})^{-1}$。预测值为 $\hat{}\ \mathbf{y} = \mathbf{X}\hat{}\boldsymbol{\beta}$，即 $\hat{}\ y_i = \hat{}\ \beta_0 + \hat{}\ \beta_1 x_{1i} + \hat{}\ \beta_2 x_{2i}$.
残差方差为 $\hat{}\ \sigma^2 = \frac{1}{n-3} \sum_i (y_i - \hat{}\ y_i)^2$。如第2.1节所述，在 x_1 与 x_2 之间引入了预设相关性。

含极端值的数据：稀疏回归。 保留式（3）中的线性设定和式（4）中的平方误差损失，但对于存在极端值的数据集，使用 `glmnet::glmnet()`（Friedman等人，2021）通过弹性网络正则化估计 $\boldsymbol{\beta}$（Zou和Hastie，2005年；Bertsimas等人，2020年；Chang等人，2021年）。在这种情形下，预测变量会在惩罚前进行内部中心化和缩放；截距不受惩罚；所有报告的系数均经过反变换。估计值 $\hat{}\ \beta = (\hat{}\ \beta_0,\ \hat{}\ \beta_1,\ \hat{}\ \beta_2)^\top$ 由下式得到

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}\right)^2 \ + \ \lambda \left[\alpha \left(|\beta_1| + |\beta_2| \right) + \frac{1-\alpha}{2} \left(\beta_1^2 + \beta_2^2 \right) \right] \right\}, \tag{5}$$

其中，$\lambda \geq 0$ 控制整体收缩，$\alpha \in [0, 1]$ 用于混合 ℓ_1 惩罚和 ℓ_2 惩罚（α=1 最小绝对收缩和选择算子；α=0岭回归）。注意，令 λ=0 即可恢复式（4）中的普通最小二乘法目标函数。
在模拟中，α 固定为 0.5，λ 则在每个完整数据集中通过 K 折交叉验证选取；在每次重复的各次插补中重复使用相同的折分区；评估准则为留出折上的均方误差。预测值的形式与线性情形相同 $\hat{}\ y_i = \hat{}\ \beta_0 + \hat{}\ \beta_1 x_{1i} + \hat{}\ \beta_2 x_{2i}$，其中 $\hat{}\ \beta$ 由（5）得出。

备注。 使用弹性网络正则化可提高存在共线性时的稳定性，并降低污染数据下的预测方差；然而，按稳健统计学中严格意义上的定义，它并不能增强抗异常值能力。当异常值值得关注时，采用稳健损失函数进行拟合（例如 Huber 或 Tukey 损失）可作为一种互补策略（Hoerl and Kennard 1970年；Tibshirani 1996）。

2.3 数据插补。

缺失数据处理方法大致可分为两类：单次插补，即每个缺失值仅被替换一次，且该过程通常易于实现

；以及多重插补，它依赖迭代方案，因此计算成本较高（Rubin 1996；Nunes 等，2010）。根据研究背景，两种方法都可以应用于同一项分析中；R包中提供了高效的实现，例如 `MICE`、`Hmisc` 和 `mlr`（Frank 和 Jr，2023；Bischl 等，2016；van Buuren 和 Groothuis-Oudshoorn，2011）。

本研究的主要目的是通过蒙特卡罗模拟，揭示不同情景下若干多重插补（MI）方法之间可能存在的差异。每个情景包含三个变量（y、 x_1和 x_2），其中仅在 x_2 中引入缺失，如表1所示。尽管在选择最佳插补方法时，对变量数量与类型的所有可能组合进行全面考察是理想做法，但如此广泛的范围需要大量计算资源。

表1： 并排的示例数据集（各10行）。左侧：3个缺失值（红色），未引入极端值。右侧：3个缺失值（红色）以及选定的极端值（蓝色）

（a）无极端值

i	y	x_1	x_2
1	y_1	$x_{1,1}$	$x_{2,1}$
2	y_2	$x_{1,2}$	$x_{2,2}$
3	y_3	$x_{1,3}$	$\mathrm{NA}_{2,3}$
4	y_4	$x_{1,4}$	$x_{2,4}$
5	y_5	$x_{1,5}$	$x_{2,5}$
6	y_6	$x_{1,6}$	$\mathrm{NA}_{2,6}$
7	y_7	$x_{1,7}$	$x_{2,7}$
8	y_8	$x_{1,8}$	$x_{2,8}$
9	y_9	$x_{1,9}$	$x_{2,9}$
10	y_{10}	$x_{1,10}$	$\mathrm{NA}_{2,10}$

（b）含极端值

i	y	x_1	x_2
1	$y_1^\star$	$x_{1,1}^\star$	$x_{2,1}^\star$
2	$y_2^\star$	$x_{1,2}^\star$	$x_{2,2}^\star$
3	y_3	$x_{1,3}$	$\mathrm{NA}_{2,3}$
4	y_4	$x_{1,4}$	$x_{2,4}$
5	y_5	$x_{1,5}$	$x_{2,5}$
6	y_6	$x_{1,6}$	$\mathrm{NA}_{2,6}$
7	y_7	$x_{1,7}$	$x_{2,7}$
8	y_8	$x_{1,8}$	$x_{2,8}$
9	$y_9^\star$	$x_{1,9}^\star$	$x_{2,9}^\star$
10	y_{10}	$x_{1,10}$	$\mathrm{NA}_{2,10}$

2.3.1 多重插补

多重插补（MI）通过从基于观测数据拟合的插补模型中随机抽取值来替换缺失值，从而生成 M 插补完成数据集。每个插补完成数据集都采用相同的完整数据分析方法进行分析，并使用Rubin规则合并结果，以获得点估计和标准误。在随机缺失（MAR）假设成立且插补模型/方法设定正确的情况下，合并估计量是一致的，标准误则反映插补内部（及插补之间）的变异性。在实践中，多重插补通常通过联合建模（常采用马尔可夫链蒙特卡罗）或全条件设定实现，并可采用预测均值匹配等方法来保留分布特征（Rubin 1996；Schafer 和 Graham 2002年；Nunes 2007；Vinha 2016；Van Buuren 2018年）。尽管其本身较为复杂，Van Buuren（2018年）仍认为，多重插补是不完整数据最有效的处理策略。

Harrell 等人（2015）指出，基于回归的插补会根据模型预测值加入随机残差，以保留原变量的条件方差。每次重复都会生成一个插补完成数据集，并采用标准方法进行分析；最终估计值为各次插补结果的平均值。

插补次数 M 通常取决于缺失信息比例 P_{miss}。White 等人（2011）建议 $M = 100 \times P_{\text{miss}}$，而 Van Buuren（2018）和 Royston（2004）

建议至少进行 20 次插补。Bodner（2008）给出了 M 在 P_{miss}从 0.05 到 0.90 范围内的经验法则值，中间值可通过插值确定。最终， M 应反映数据的复杂性和潜在的缺失数据机制。

`mice` 包由 van Buuren 和 Groothuis-Oudshoorn（2011）提出，为多重插补提供参数选项和非参数选项。本研究设计了涵盖这两类方法的情景。参数法依赖明确的分布假设，例如用于连续变量的正态线性模型 `method="norm.predict"`。非参数法不依赖此类假设，包括：（i）基于树的方法，如分类回归树（`cart`）（Steinberg，2009）、预测均值匹配（`pmm`）（Allison，2015）和随机森林（`rf`）（Breiman，2001）；以及（ii）简单的供体法或重抽样方法，如 `sample` （热卡插补式）。基于树的插补方法无需指定参数形式即可近似条件分布，并从由终端节点确定的供体集合中抽取替代值，因此无需依赖参数假设，便可处理非线性关系和交互作用。

注。 完全随机缺失假设仍是合并估计无偏性的工作前提。当目标是对插补策略进行描述性比较时，可通过指标和图形评估分布拟合度（例如均方误差分布、箱线图和描述性统计，以及其他系数结果），正如模拟研究中所实施的那样（Rubin和Wiley 1987年；Van Buuren 2018年）。

2.3.2 插补方法

模拟研究比较了六种插补方法，记为{T1,.. . , T6}，并在图和表中以颜色加以区分。这些方法也可在 `mice::mice()`中使用，代表了参数法与非参数法的组合：

- T1： `norm.predict` –通过线性回归进行插补，属于预测型参数法，根据模型对数值进行插补，也称为回归插补。Van Buuren和Groothuis-Oudshoorn（2011）以及Little和Rubin（2019）也指出， `norm.predict`应避免用于正式数据分析，因为它忽略插补不确定性，并可能人为强化变量之间的关系。即使使用更丰富的模型，这些问题依然存在。随机方法（如 `pmm` 或 `norm`）通常更受青睐，不过在合理的正态性假设下，确定性预测仍可能提供合理的均值估计；

- T2： `lasso.select.norm` ——先进行最小绝对收缩和选择算子变量选择以实现稀疏性，再进行参数正态回归。从形式上说，该方法先进行预处理阶段的最小绝对收缩和选择算子变量选择，然后使用贝叶斯线性回归对单变量缺失数据进行插补。 `mice`函数中使用的方法基于 Deng 等人（2016）以及 Zhao 和 Long（2016）提出的正则化回归的间接使用（IURR）；

- T3： `norm.boot`——使用自助重抽样进行参数正态回归（Wu 1986），以在插补中纳入参数不确定性，因此使用自助法预测值对单变量缺失数据进行插补。在 `mice`中，该方法抽取一个自助样本，计算回归权重，并使用正态残差进行插补（van Buuren 和 Groothuis-Oudshoorn 2011）；

- T4： `pmm`– 预测均值匹配；一种供体法，根据不完整案例的预测均值，选取预测均值最接近的案例中的观测值来填补每个缺失值。该方法名称最初由 Little (1988) 提出，

其在 `mice` 中的实现遵循 Van Buuren 和 Groothuis- Oudshoorn (2011) 描述的方法。Morris等人 (2014) 提供了更多方法细节和调优指南；

- T5： `rf` – 使用随机森林（Breiman 2001）对单变量缺失数据进行插补。随机森林是一种非参数回归树集成，能够捕捉非线性关系和交互作用。在 `mice` 函数中使用的方法调用了 `randomForest::randomForest()`，后者实现了 Breiman 的随机森林算法（Liaw和Wiener 2002）。关于该方法的其他实现方式的更多详情及模拟方法，可参见 Shah等人（2014）和 Salman等人（2024）；以及
- T6： `midastouch` – 通过预测均值匹配进行插补，并利用距离辅助选择供体；供体选择会受到预测距离权重的影响。在 `mice` 中实现的方法基于 Rubin 和 Wiley（1987）以及 Siddique 和 Belin（2008）。更多详情可参见 Van Buuren等人（2006）和 Gaffert等人（2016）。

选择这些技术是为了体现不同的建模理念：完全参数化的回归模型（T1、T2、T3）、分布假设较弱的供体方法（T4、T6），以及非参数机器学习方法（T5）。这种多样性有助于评估在有极端值和无极端值的各种情景下，模型结构、预测变量选择和重抽样策略如何影响插补性能。关于所用插补方法及其他技术的更多详情，可参见 van Buuren和Groothuis-Oudshoorn（2011）。

2.4 多重插补方法比较

插补策略的质量通过其对后续分析的影响来评估，而不是直接将插补值与其未观测对应值进行比较。正如 Van Buuren（2018年）所强调的，插补并非预测，将插补视为预测的诊断往往会夸大结果或产生误导。因此，本比较聚焦于受控模拟设计下的预测性能和推断性能。

(i) 样本外预测误差：对于每个模拟重复和完整数据集，采用 K 折交叉验证均方误差（交叉验证均方误差）概括预测性能；在每次模拟重复中固定折分配，并在不同插补和方法中重复使用。这可以减少折分配带来的变异性，并确保不同方法之间进行配对的一一比较（Arlot 和 Celisse，2010）。在清洁数据情形下，使用普通最小二乘法（OLS）；存在污染时，则使用由 `cv.glmnet` 选择 $\alpha = 0.5$ 和 λ 的弹性网（EN）进行预测（Zou和Hastie，2005年；Friedman等人，2010年）。对于每种多重插补方法，在每次模拟重复中对 M 次插补的交叉验证均方误差取平均，再跨模拟重复概括其均值、方差以及第 2.5、第 50 和第 97.5 分位数。较低的分位数和较小的离散程度表示性能更好且更稳定。

（ii）推断标准与相容性：除预测外，清洁数据情形还评估多重插补能否为真实线性模型恢复名义推断：通过 Rubin规则报告（β_0、β_1、β_2）的合并偏差、均方根误差和95%覆盖率（Rubin和Wiley 1987年；White 等人 2011；Little和Rubin 2019）。 x_2 的插补模型纳入（y、x_1），并在预测变量矩阵中排除 x_2，以促进其与分析模型的相容性（Van Buuren 2018年；Van Buuren和Groothuis-Oudshoorn 2011）。在污染情形中，使用预先设定的选择规则（保留至少在50%的插补中被选中的变量；否则采用并集回退），在弹性网络之后通过选择后合并普通最小二乘法汇总系数。

这些汇总结果以变量被选中为条件，因此从构造上说，对覆盖率的估计偏于乐观；本文对此予以透明报告，而不将其视为考虑了选择过程的推断所提供的正式保证（Berk 等人 2013；Taylor 2015）。

(iii) 可视化摘要与校准：为补充表格摘要，重复层级的交叉验证均方误差密度图和箱线图呈现了分布形态与尾部行为。预测值与真实 Y 的分位数–分位数（QQ）曲线提供了校准视角（先计算每次重复中各插补的中位数，再汇总各次重复），其中 45° 线作为参考。

(iv) 公平性与可复现性控制：所有方法均采用同一个主随机种子进行评估，该种子仅初始化生成器一次。随机抽样按任务（折分配、MCAR掩码、污染、插补）组织为相互独立且确定性的子流，因此更改一项任务不会影响其他任务。根据 Arlot 和 Celisse（2010），每次重复内的交叉验证折均固定，并在所有插补和方法中重复使用，从而确保配对比较并减少折分配引入的噪声。对于受污染的数据，弹性网络使用这些固定折进行调参。各方法（T1–T6）的顺序在所有图和表中保持一致。如 Varma 和 Simon（2006）以及 Cawley 和 Talbot（2010）所述，应该承认，通过最小化交叉验证误差来选择 λ，相较于使用外部测试折或嵌套交叉验证，会引入轻微的乐观偏差。

2.5 仿真设计（方案与评估）

本小节仅记录蒙特卡罗方案 和评估 流程。所有数据生成细节（基线模型、污染方案）均在第2.1节（数据集）中定义，下游分析模型则在第2.2节（统计模型）中说明。此处不再重复介绍插补方法（第2.3.2小节）及其预测变量。

各情景的方案。对于由以下因素定义的每种情景：$\{n,\ P_{\text{miss}},\ P_{\text{ext}},\ \rho,\ M,\ n_{\text{sim}},$ 多重插补，分支$\}$:

表2：蒙特卡罗研究的设计因素和分析分支。情景由以下因素的笛卡尔积构成；使用普通最小二乘法分析干净数据，使用弹性网络（EN）分析受污染数据（含极端值）

因子	水平	定义
n	$\{20, 40, 80, 200, 500\}$	每个情景中的观测值数量
P_{miss}	$\{0.05, 0.10, 0.25, 0.30\}$	缺失比例
P_{ext}	$\{0.03, 0.04, 0.05, 0.10, 0.15, 0.30\}$	极端值比例
ρ	$\{0, 0.6\}$	x_1 与 x_2之间的相关性
M	$\{5, 10\}$	每种情景的插补次数（迭代次数）
n_{sim}	{50，300，1000，3000}	蒙特卡洛重复次数
MI	{T1, T2, T3, T4, T5, T6}	多重插补方法
分支	{干净/普通最小二乘法，受污染/弹性网络}	数据类型及相应的统计模型

*图例：除报告分层结果的小节另有说明外，各情景均包含所有因子水平
。本文还通过 $(iter)$ 符号标示每种情景的插补次数。

1. 生成一个基线（干净）数据集；根据前文所述的污染机制，生成与之配对的受污染副本。

2. 对每次重复 $r = 1, .., n_{\text{sim}}$，在 x_2 上为干净副本独立抽取一个大小为 k_{miss} =round($n \times P_{\text{miss}}$) 的完全随机缺失掩码，并为受污染

副本抽取另一个完全随机缺失掩码。为重复 r 固定一个 K 折分区（K=5）；在 r 内的所有插补和方法中重复使用这些折标识符。

3. 对于每种数据类型 $d \in$ {干净、受污染} 以及每种多重插补方法 $t \in \{T1$、.....、$T6\}$：

(a) 使用 M 次插补和适用于 x_2 的相容预测变量矩阵运行 `mice`（使用 y 和 x_1；不包括其自身），具体如第 2.3 节所述；(b) 拟合与完整数据所用的相同分析模型：干净数据使用普通最小二乘法；受污染数据使用弹性网（EN）及 $\alpha = 0.5$（第 2.2 节）。对于 EN，使用重复 r 的固定折，通过交叉验证选择 λ；(c) 将 m 个插补完成数据集上的 K 折交叉验证均方误差取平均，作为重复层面的预测误差（折标识符在重复 r 内固定，并在各方法间重复使用）；(d)（仅无污染情形）根据 Rubin 规则合并各次插补的系数估计，并记录 (β_0、β_1、β_2) 的偏差、均方根误差和 95%覆盖率；(e)（仅污染情形）按照固定规则记录 EN 之后变量筛选后的合并普通最小二乘法汇总结果：若某协变量在至少 50% 的插补中被选中，则予以保留（若无协变量入选，则采用并集作为后备方案）。

评估与汇总。主要比较标准是样本外预测均方误差（交叉验证均方误差）。对于每种多重插补方法 t，先在每个重复内对 M 次插补的交叉验证均方误差取平均，再通过均值、方差以及 2.5%、50% 和 97.5% 分位数对各次重复的结果进行汇总。这些统计量与表格中的“预测均方误差（样本外）”列以及图中的密度/箱线图面板完全对应。分布校准通过预测值与真实 Y的分位数–分位数曲线呈现：对每次重复，先计算各次插补结果的中位数，再对各次重复进行汇总。在清洁数据情形下，这些指标用于评估相容性条件下的推断有效性（Rubin and Wiley 1987年；White 等人 2011；Van Buuren 2018年）。在污染情形下，系数汇总明确报告为以入选为条件的结果，并遵循第 2.4 节所述的固定规则（Berk 等人 2013；Taylor 2015）。

此处考察的内容并非缺失机制。仅在 x_2 下引入缺失，且采用完全随机缺失。这是有意的安排，足以实现算法层面对多重插补程序的比较 这一目标，因为全条件设定每次只更新一个单变量条件分布；仅设置一个不完整变量，既能凸显各方法的核心行为，又能保证在完全随机缺失下进行干净的 Rubin 合并（Van Buuren 2018年；Van Buuren和Groothuis-Oudshoorn 2011；White 等人 2011；Little和Rubin 2019）。扩展到随机缺失/非随机缺失，或包含多个不完整变量，都需要重新设定条件模型、检查相容性并开展敏感性分析，相较于已经很大的网格（涵盖n、P_{miss}、P_{ext}、ρ、M、n_{sim}、多重插补和分支），计算负担将大幅增加。此外，探索 α 的其他取值或采用稳健损失函数，会增加一个调优维度（例如嵌套交叉验证），并改变分析情形（Rousseeuw和Leroy 1987年；Huber和Ronchetti 2009）。最后，生成极端值时并未强加基于极值理论的参数模型（例如广义极值分布或广义帕累托分布）；这样做可能会引入额外的选择和诊断，并降低所得见解的普适性（Fisher和Tippett 1928；Von Mises 1936；Gnedenko 1943；Jenkinson 1955）。

算法 1概述了蒙特卡罗方案：使用单个主种子，并为折、掩码、插补和模型调优设置确定性子流。

算法 1 使用单个主种子和确定性子流的蒙特卡罗方案

使用单个主种子 $S = 241103414$ 初始化一个伪随机数生成器，并使用
用于折、掩码、插补和调优的确定性子流。
对每个情景 $\{n,\ P_{\text{miss}},\ P_{\text{ext}},\ \rho,\ M,\ n_{\text{sim}}$，多重插补，分支$\}$ **执行**
  对 $r = 1, \ldots, n_{\text{sim}}$ **执行**
    数据生成（子流 A）。抽取一个干净基线样本$\{(y_i,\ x_{1i},\ x_{2i})\}_{i=}^{n}$ $_1 \sim$
    F_0 （第 2.1 节）；通过 $\pm 3s$机制（第 2.1.1 节）创建其污染孪生样本。
    折标识符（子流 B）。抽取一个 K折映射（$K = 5$） f_r： $\{1,\ \ldots\ldots,\ n\} \to$
    $\{1,\ \ldots\ldots,\ K\}$；将 f_r传递给重复 r 中的所有交叉验证调用。
    完全随机缺失掩码（子流 C、D）。独立抽取 $k_{\text{miss}} = \text{round}(n \times P_{\text{miss}})$
    干净副本和污染副本的索引；将这些索引处的 x_2 设置为 `NA` 。
    对于 数据类型 $d \in \{$干净，受污染$\}$ **执行**
      对于 方法 $t \in \{$`T1`，……，`T6`$\}$ **执行**
        插补（子流 E$_{r,\ t,\ d}$）。运行 `mice::mice()`使用 M 次插补以及
        面向 x_2 、仅使用 $(y,\ x_1)$ 的相容预测变量矩阵；获得完整数据集
        $\{D_{r,t,d}^{(m)}\}_{m=1}^{M}$。
        模型和预测误差。
        如果 $d =$干净 **则**
          在每个$\mathcal{D}$上拟合普通最小二乘法$_{r,t,d}^{(m)}$；使用 f_r计算交叉验证均方误差；设置 $e_{r,t,}^{(m)}{}_{d} =$ 交叉验证均方误差。
        否则如果 $d =$受污染 **则**
          使用 f_r 作为 `foldid`拟合弹性网络（$\alpha = 0.5$）；记录 $\lambda_{\min}$ 处的交叉验证均方误差为 $e_{r}^{(m)}$
          $_{,t,d}$.
        结束 if
        定义重复级误差 $\bar{e}_r$
          $_{,t,d} = \frac{1}{M}\sum_{m=1}^{M} e_{r,t,d}^{(m)}$.
        系数汇总。
        if $d =$干净 **则**
          通过 Rubin规则对 M 次插补的普通最小二乘结果进行合并；记录偏差、均方根误差和 95%
          （$\beta_0,\ \beta_1,\ \beta_2$）的覆盖率。
        else
          令 $S^\star = \{j \in \{x_1,\ x_2\} :$ 在 M 次弹性网络拟合中，$\Pr(j$ 已选中$) \geq$ `support_thresh`$\}$
          拟合；若 $S^\star = \varnothing$，则使用已选变量的并集（回退方案；若仍为空，则使用
          $\{x_1,\ x_2\}$）。对预测变量 $S^\star$ 使用合并普通最小二乘法。
        结束 if
      结束 for
    结束循环
  结束循环
  跨重复汇总（交叉验证均方误差）。对于每个 (t, d)，汇总重复的
  水平交叉验证均方误差值 $\{\bar{e}_{r,t,d}\}_{r=1}^{n_{\text{sim}}}$的均值、方差以及第 2.5%、50% 和第 97.5% 分位数；
  绘制交叉验证均方误差密度图和箱线图，并另外生成预测值与真实值的QQ校准曲线
  的**y**。
结束循环

3 结果

本节报告一组具有代表性的情景，以说明本文的主要结论。所有情景设置和重复的完整模拟表格见附录 A。

3.1 完整数据基准：结构与基线拟合

为了解完整数据集及相应模型调整的情况，考虑以下情形： $n = 500$ 个观测值，极端值污染比例为 $P_{\text{ext}} = 0.10$（因此每个变量中有 50 个值被替换）， x_1 与 x_2之间的相关性为 $\rho = 0.6$ 。图 1 和图 2 展示了两种数据状态下（y， x_1， x_2）的两两关系：第一行对应干净数据（蓝色色系），第二行对应污染数据（黄/红色色系）。坐标轴标记为 y， x_1， x_2；每一行都包含自己的颜色条（‘水平’），用于表示相对二元密度。

有三种明显的模式。(i) 在无污染情形下，散点图和线性平滑拟合显示变量间关系近似线性，且离散程度较小；此外，密度等高线近似呈椭圆形，这与 $\rho \approx 0.6$ 和二元正态生成器相符。(ii) 存在污染时，所有成对视图中的离散程度均有所增加，等值集向外扩展，高密度区域离中心质量更远，表明尾部极端值产生了杠杆效应。(iii) 不同情形下，拟合线性趋势线的方向相近，但存在极端值时，预测不确定性会大幅增加。

表 3 报告了基于完整数据（无缺失）拟合的参考模型的估计结果和预测误差：干净数据使用普通最小二乘法（OLS），受污染数据使用弹性网（EN）。线性模型在干净数据上的误差较低，而当污染导致方差膨胀并引入高杠杆点时，EN 的稳定性略高。针对这一情景的这些基准结果，将作为后续插补实验的基线。

表 3： 基于完整数据（$n = 500$, $P_{\text{ext}} = 0.10$, $\rho = 0.6$）拟合的线性（普通最小二乘法）和稀疏（弹性网络）回归模型的参数估计值及性能指标（均方误差）

Data	模型	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	均方误差
干净	OLS	1.170	0.517	1.440	2.190
受污染的	EN	1.190	0.532	1.410	2.020

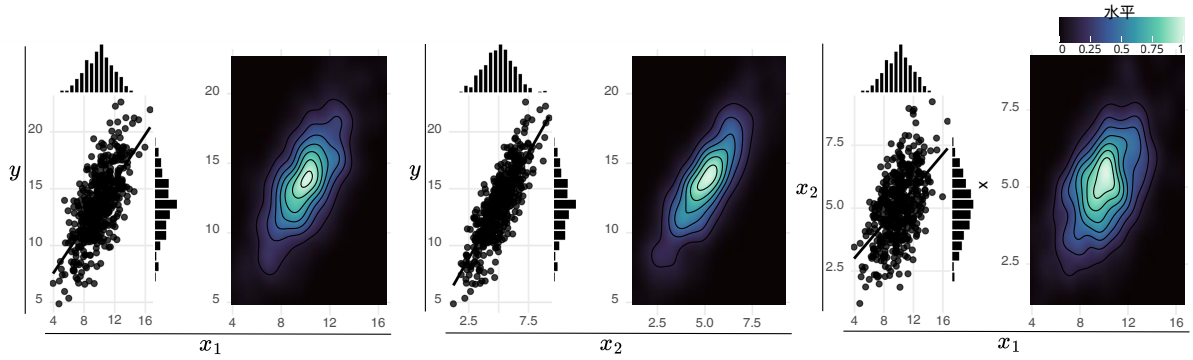

图1： 具有 $n = 500$、 $P_{\text{ext}} = 0.10$和 $\rho = 0.6$的**干净数据**集的关系（相关图、边缘直方图和经验密度）。坐标轴为 y， x_1， x_2。色标（“水平”）表示每行的相对二元密度

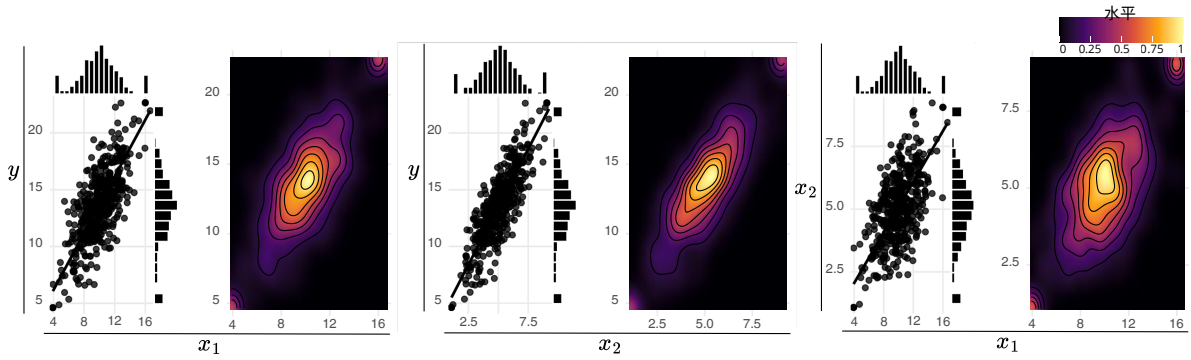

图2： 具有 $n = 500$、 $P_{\text{ext}} = 0.10$ 和 $\rho = 0.6$ 的**受污染数据**（含极端值）数据集的关系（相关图、边缘直方图和经验密度）。坐标轴为 y、x_1、x_2。色标（“水平”）表示每行对应的相对二元密度

3.2 模拟研究：缺失和极端值情况下的插补 s

报告了六种多重插补方法（T1–T6）在两种分析情形下的结果：对干净数据拟合普通最小二乘法（OLS），对含极端值的受污染数据拟合弹性网（EN）。比较按样本量区间组织：小（$n \in \{20, 40\}$）、中（$n \in \{80, 200\}$）和大（$n = 500$）。另按缺失率区块划分：低（5%、10%）和高（25%、30%）。推断性能通过偏差、均方根误差和 95%覆盖率进行汇总，对应（β_0, β_1, β_2）；预测性能则通过样本外交叉验证均方误差的均值 X、方差 σ^2和中位数 Q_{50}进行汇总。附录 A 还报告了尾部分位数 $Q_{2.5}$ 和 $Q_{97.5}$。

用于总体解读的方法分组为：参数法（T1–T3）、供体法（T4、T6）和非参数法/机器学习（T5）。有三种反复出现的模式可预示详细结果：（i）在污染情况下，系数呈现稳定的方向性（截距膨胀、β_1 倾斜、β_2收缩）；（ii）增加 n 主要会压缩变异性，而非消除这些偏移；（iii）在预测方面，参数多重插补（T1–T3）通常会产生较窄的交叉验证均方误差尾部，而供体/机器学习程序（T4–T6）往往能降低斜率偏差，但代价是离散程度更大。后续小节按顺序展示小、中、大样本的面板，附录则提供完整的数值细节。

小样本量（$n = 20$ 和 $n = 40$）。请参见表4和表5以及图3–6。此外，关于所介绍的模拟方法的更多细节，请参见附录中的表10–17。

系数（水平与偏差）。在使用普通最小二乘法且缺失率较低（$P_{\text{miss}} \le 0.10$）的干净数据下，T1–T3 和 T4–T6 的斜率偏差与均方根误差均较小，覆盖率接近名义水平n=20；在这种情况下，组间差异很小（表 4）。随着 P_{miss} 增加至{0.25，0.30}，所有方法的斜率均方根误差都增大；在若干单元格中，供体/机器学习组（T4–T6）的离散程度通常略高于 T1–T3，而合并汇总（T1–T6）则介于两组之间。在存在极端值（弹性网络）的污染情形下，两个 n=20 和 n=40都出现了截距水平偏移。在 n=20 和 $P_{\text{miss}} \le 0.10$下，T1–T3 的正向 β_0 偏差（例如 $[0.888, 1.662]$）大于供体/机器学习组 T4–T6（例如 $[0.081, 1.382]$），而 β_1 在 T1–T3 下通常比在 T4–T6 下更负（例如 $[-0.356, -0.225]$和 $[-0.264, -0.047]$）。因此，在污染情形下，参数多重插补往往会导致更强的系统性偏移，而供体/机器学习多重插补则能减轻系数偏差。β_2 的符号取决于设计：在 n=20下通常为正，在 n=40下为负，且两组的定性变化方向相同；失真会随.....

随着 P_{miss}加剧。汇总结果（T1–T6）呈现出介于 T1–T3 和 T4–T6 之间的模式。

覆盖率。 污染情况下覆盖率最为脆弱。在 n=20时，对于 $P_{\text{miss}} \leq 0.10$ ，弹性网络在各组中仍能保持不可忽略的覆盖率，但随着缺失增加，覆盖率会下降；在高P_{miss}单元中，T4–T6偏差的降低有时会使其覆盖率达到或优于T1–T3（表4）。在 n=40时，所有组的截距覆盖率在极端值影响下都可能崩溃（其中有若干数值接近于零），这表明在受污染的小n 情景中，组间差异次于整体区间可靠性不足的问题。

预测误差（交叉验证均方误差）与尾部风险。 在普通最小二乘法（干净）下，当 n=40且 P_{miss} 较低时，参数组T1–T3的X 和 σ^2 较低，且分布更集中于较低值，相比之下供体/机器学习组则较高：例如，T1–T3通常为 $X \approx 2.08$–2。28和 $\sigma^2 \approx 0.02$–0。06，而T4–T6的均值/方差则更高（例如，T4约为$X = [2.27, 2.36]$、 $\sigma^2 = [0.02, 0.13]$；T5约为 $[2.35, 2.56]$、$[0.05, 0.21]$）。在 n=20时，这种差异在T5中更为明显，其离散程度在干净情景中最大。

在受污染的弹性网络（EN）下，随着 P_{miss} 增加，交叉验证均方误差的水平 和离散程度 均有所上升，组间差异尤为明显。对于 n=20 和 $P_{\text{miss}} \in \{0.25,\ 0.30\}$，T4–T6位于$X$和 σ^2的上尾（例如，T4：$X = [3.258, 3.703]$、 $\sigma^2 = [1.237, 1.734]$；T5： $[4.243, 5.068]$、$[1.351, 2.908]$；T6： $[2.847, 3.197]$、$[1.182, 2.461]$），而T1–T3则明显更集中（例如，T1： $[1.180, 1.635]$、 $[0.072, 0.165]$；T2/T3也同样较低）。在 n=40处，这一排序依然存在，但离散程度有所减弱。随着供体/机器学习方法带来更厚的尾部，合并汇总（T1–T6）的范围也随之扩大（表5）。图3–6进一步印证了这些差异：在极端值条件下，均方误差密度右移且更为分散，QQ曲线则在高分位数处下弯至恒等线下方。

较小 n下的方法敏感性。 所有多重插补方法均受到极端值的影响。相对于T1–T3，供体/机器学习组T4–T6在 n=20时往往降低系数偏差（有时还能保持覆盖率），但代价是预测分布尾部风险略有增大（ σ^2更高，尤其是T5）。总体T1–T6反映了方差与尾部权衡中的这一偏差：中位数保持适中，而上尾随缺失增加而扩大。在 n=40时，差距缩小，但排序保持不变：T1–T3呈现更紧的尾部；T4–T6具有更高的离散程度；T1–T6介于两者之间（表4和表5）。

启示。 截距偏移和斜率扭曲是各组的主要污染信号。交叉验证均方误差离散度（σ^2）有助于在较小 n下区分参数法与供体/机器学习行为。对于伴有极端值的中高缺失率（$\geq$ 25%），尾部更紧的程序（T1–T3）能提供更稳定的样本外误差，而供体/机器学习方法（T4–T6）可能以更重的预测分布尾部为代价获得更小的系数偏差；总体行为（T1–T6）自然反映了这一折中。

表 4：在干净数据（使用普通最小二乘法）和含极端值的数据（使用弹性网络）下，$n = 20$（左）和 $n = 40$（右）的 $(\beta_0, \beta_1, \beta_2)$ 的偏差、均方根误差和 95%覆盖率。每个缺失率区块内的条目以最小值–最大值表示（低：$P_{\text{miss}} \in \{0.05, 0.10\}$；高：$P_{\text{miss}} \in \{0.25, 0.30\}$）。范围还汇总了所有设计组合 $P_{\text{ext}} \in \{0.03, 0.05, 0.10, 0.15, 0.30\}$、$n_{\text{sim}} \in \{50, 300, 1000\}$、迭代次数 $\in \{5, 10\}$ 以及 $\rho \in \{0, 0.6\}$。多重插补方法分为参数法（T1–T3）、供体/机器学习（T4–T6）和组合类别（T1–T6）

		n = 20			n = 40		
MI	指标	β_0	β_1	β_2	β_0	β_1	β_2
线性回归（普通最小二乘法）： 低缺失数据比例（5% 和 10%）							
T1–T3	Bias	[-0.424 , -0.097]	[-0.043, -0.007]	[0.038 , 0.158]	[1.940, 2.503]	[-0.203, -0.131]	[-0.226, -0.121]
T1–T3	RMSE	[0.315, 0.539]	[0.023 , 0.097]	[0.059 , 0.246]	[1.963, 2.526]	[0.135, 0.205]	[0.127, 0.228]
T1–T3	覆盖率	[1.000 , 1.000]	[0.990 , 1.000]	[0.980, 1.000]	[0.680, 0.980]	[0.620, 1.000]	[1.000 , 1.000]
T4–T6	Bias	[-0.286, -0.023]	[-0.025 , 0.002]	[0.012, 0.086]	[1.886 , 2.574]	[-0.193 , -0.102]	[-0.266, -0.167]
T4–T6	RMSE	[0.290, 0.478]	[0.026 , 0.097]	[0.050, 0.205]	[1.931, 2.619]	[0.121, 0.197]	[0.173, 0.276]
T4–T6	覆盖率	[1.000 , 1.000]	[0.993 , 1.000]	[0.990 , 1.000]	[0.720 , 0.980]	[0.880, 1.000]	[1.000 , 1.000]
T1–T6	Bias	[-0.424, -0.023]	[-0.043, 0.002]	[0.012, 0.158]	[1.886, 2.574]	[-0.203, -0.102]	[-0.266, -0.121]
T1–T6	RMSE	[0.290, 0.539]	[0.023 , 0.097]	[0.050, 0.246]	[1.931, 2.619]	[0.121, 0.205]	[0.127, 0.276]
T1–T6	覆盖率	[1.000 , 1.000]	[0.990 , 1.000]	[0.980, 1.000]	[0.680, 0.980]	[0.620, 1.000]	[1.000, 1.000]
线性回归（普通最小二乘法）： 高缺失数据比例（25%和30%）							
T1–T3	Bias	[-0.875 , 0.294]	[-0.113, -0.051]	[0.085 , 0.331]	[1.858 , 2.766]	[-0.232, -0.141]	[-0.228, 0.008]
T1–T3	RMSE	[0.726 , 1.151]	[0.145, 0.214]	[0.303 , 0.470]	[1.927, 2.830]	[0.154, 0.241]	[0.065, 0.245]
T1–T3	覆盖率	[0.980, 1.000]	[0.903, 0.987]	[0.799, 0.973]	[0.640, 0.820]	[0.320 , 1.000]	[1.000 , 1.000]
T4–T6	Bias	[-0.406, 0.191]	[-0.033 , 0.003]	[-0.016, 0.117]	[1.863, 3.087]	[-0.206 , -0.074]	[-0.365, -0.161]
T4–T6	RMSE	[0.643 , 0.837]	[0.116, 0.164]	[0.238, 0.353]	[1.946, 3.163]	[0.110, 0.217]	[0.182, 0.383]
T4–T6	覆盖率	[1.000 , 1.000]	[0.980, 1.000]	[0.940, 1.000]	[0.720 , 0.940]	[0.860 , 1.000]	[1.000 , 1.000]
T1–T6	Bias	[-0.875 , 0.294]	[-0.113, 0.003]	[-0.016, 0.331]	[1.858, 3.087]	[-0.232, -0.074]	[-0.365 , 0.008]
T1–T6	RMSE	[0.643 , 1.151]	[0.116 , 0.214]	[0.238, 0.470]	[1.927, 3.163]	[0.110, 0.241]	[0.065, 0.383]
T1–T6	覆盖率	[0.980, 1.000]	[0.903 , 1.000]	[0.799, 1.000]	[0.640, 0.940]	[0.320 , 1.000]	[1.000 , 1.000]
稀疏回归（弹性网络）：低缺失数据比例（5%和10%）							
T1–T3	Bias	[0.888 , 1.662]	[-0.356 , -0.225]	[0.197, 0.399]	[2.235 , 3.652]	[-0.239, -0.082]	[-0.433 , -0.245]
T1–T3	RMSE	[0.932, 1.815]	[0.229 , 0.361]	[0.201, 0.405]	[2.243, 3.669]	[0.092 , 0.241]	[0.248 , 0.438]
T1–T3	覆盖率	[0.900, 1.000]	[0.860 , 1.000]	[0.902, 1.000]	[0.000 , 0.340]	[0.040 , 1.000]	[0.460, 1.000]
T4–T6	Bias	[0.081 , 1.382]	[-0.264, -0.047]	[-0.007 , 0.237]	[2.197, 3.581]	[-0.221, -0.061]	[-0.454, -0.298]
T4–T6	RMSE	[0.558, 1.582]	[0.149, 0.275]	[0.115, 0.283]	[2.212, 3.617]	[0.079, 0.226]	[0.303, 0.460]
T4–T6	覆盖率	[0.950, 1.000]	[0.940, 1.000]	[0.950, 1.000]	[0.000 , 0.320]	[0.460, 1.000]	[0.480 , 1.000]
T1–T6	Bias	[0.081 , 1.662]	[-0.356 , -0.047]	[-0.007 , 0.399]	[2.197 , 3.652]	[-0.239, -0.061]	[-0.454 , -0.245]
T1–T6	RMSE	[0.558, 1.815]	[0.149, 0.361]	[0.115 , 0.405]	[2.212 , 3.669]	[0.079, 0.241]	[0.248 , 0.460]
T1–T6	覆盖率	[0.900, 1.000]	[0.860 , 1.000]	[0.902, 1.000]	[0.000 , 0.340]	[0.040 , 1.000]	[0.460, 1.000]
稀疏回归（弹性网络）：高缺失数据比例（25% 和 30%）							
T1–T3	Bias	[0.983, 1.809]	[-0.396 , -0.241]	[0.211, 0.423]	[2.037 , 3.599]	[-0.262, -0.068]	[-0.448 , -0.090]
T1–T3	RMSE	[1.158, 2.124]	[0.261, 0.429]	[0.242, 0.459]	[2.095, 3.661]	[0.102 , 0.272]	[0.151, 0.465]
T1–T3	覆盖率	[0.740 , 1.000]	[0.428, 1.000]	[0.660, 1.000]	[0.060, 0.640]	[0.100, 1.000]	[0.720, 1.000]
T4–T6	Bias	[-0.612 , 0.867]	[-0.197, 0.102]	[-0.228 , 0.191]	[1.833, 3.456]	[-0.193, 0.006]	[-0.520 , -0.314]
T4–T6	RMSE	[1.229, 1.707]	[0.144, 0.259]	[0.124, 0.276]	[1.903, 3.573]	[0.079, 0.213]	[0.329, 0.534]
T4–T6	覆盖率	[0.978, 1.000]	[0.980, 1.000]	[0.980, 1.000]	[0.220, 0.720]	[0.720, 1.000]	[0.600, 0.900]
T1–T6	Bias	[-0.612, 1.809]	[-0.396 , 0.102]	[-0.228 , 0.423]	[1.833, 3.599]	[-0.262 , 0.006]	[-0.520 , -0.090]
T1–T6	RMSE	[1.158, 2.124]	[0.144, 0.429]	[0.124, 0.459]	[1.903, 3.661]	[0.079, 0.272]	[0.151, 0.534]
T1–T6	覆盖率	[0.740 , 1.000]	[0.428 , 1.000]	[0.660 , 1.000]	[0.060 , 0.720]	[0.100, 1.000]	[0.600, 1.000]

* 1 6 1 norm.predict 2 lasso.select.norm 3 norm.boot 4 pmm 5 图例：多重插补方法（T –T ）：
rf 6 midastouch 95
T ; T ; T ; T ; T ; T 。覆盖率是包含真实系数的 % 区间所占的比例。

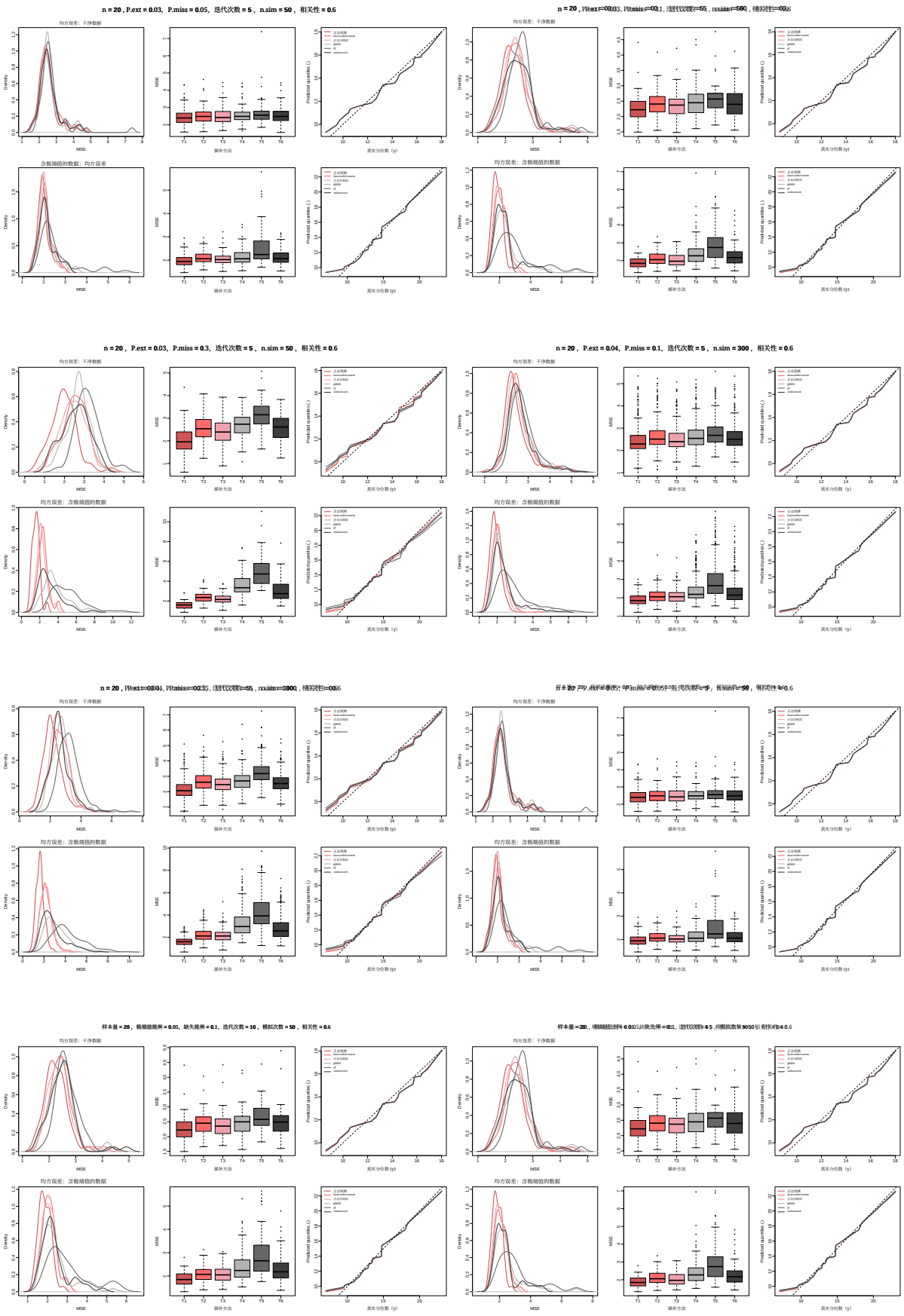

图 3： 预测均方误差密度（干净数据与极端值污染数据）、均方误差箱线图，以及六种多重插补方法（T1–T6）的 QQ图（ y 的预测分位数与真实分位数）。对于 $\boldsymbol{n = 20}$，按 P_{ext} 排序，并且 P_{miss} （第 1 个面板，共 4 个）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子面板展示设计值（n、 P_{ext}、 P_{miss}、迭代次数、模拟次数和 ρ）。

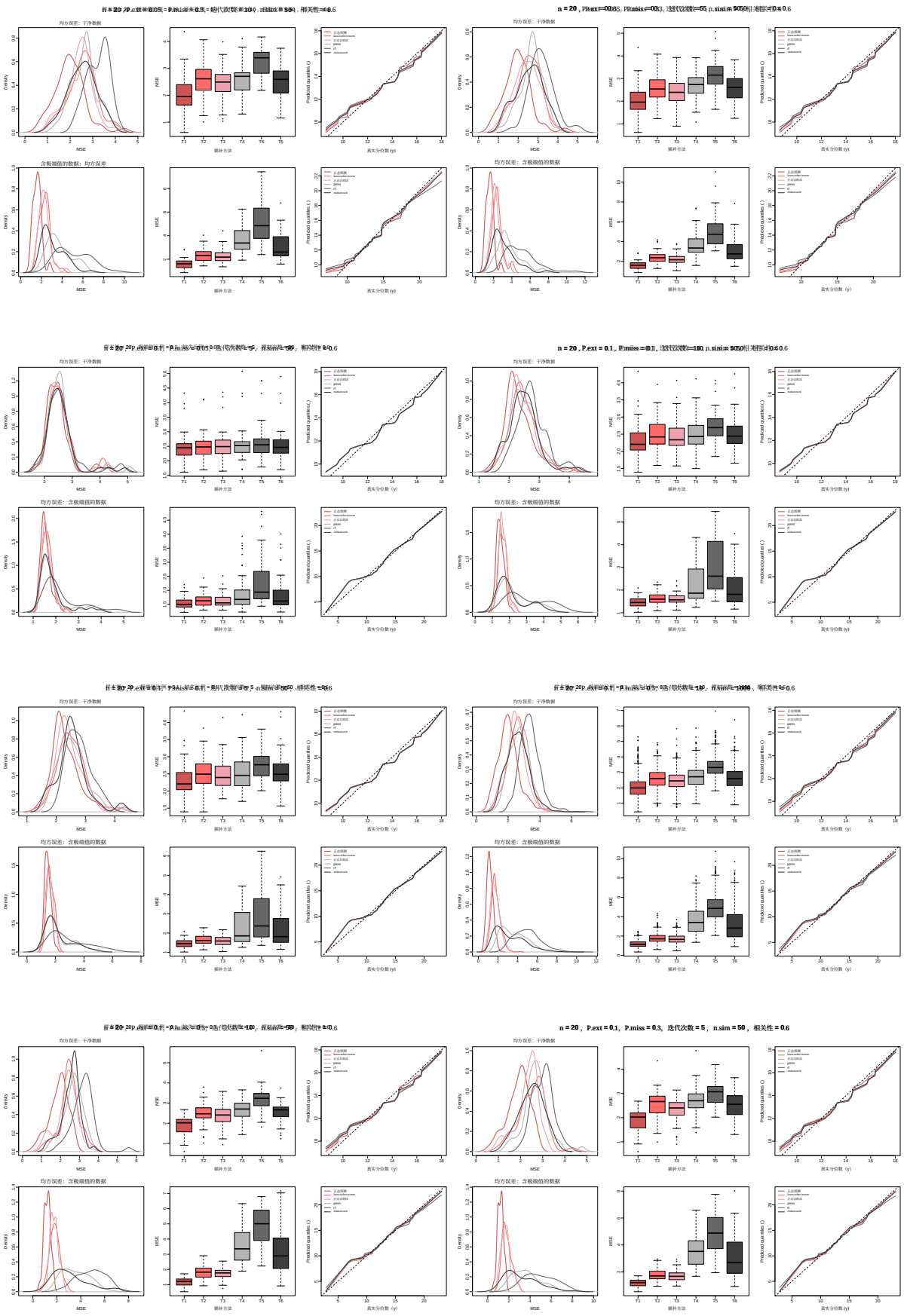

图 4： 预测均方误差密度（干净数据与极端值污染数据对比）、均方误差箱线图，以及六种多重插补方法（T1–T6）下的QQ图（ y 的预测分位数与真实分位数对比）。对于 $\boldsymbol{n = 20}$，按P_{ext}排序，且 P_{miss} （第 2/4 个面板）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子图均显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、n.sim 和 ρ）。

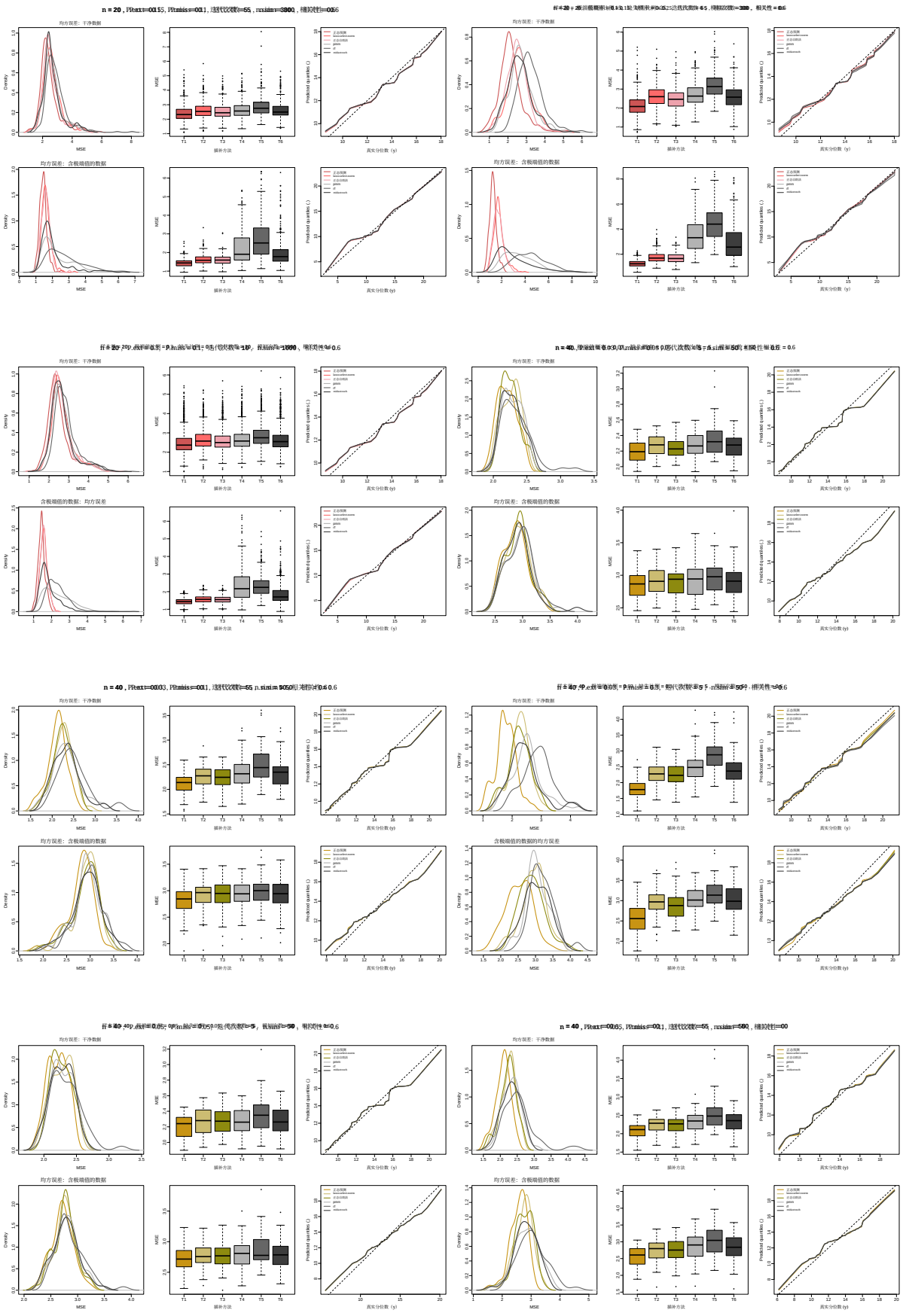

图 5： 预测均方误差密度（干净数据与极端值污染数据）、均方误差箱线图，以及六种多重插补方法（T1–T6）的 QQ图（ y 的预测分位数与真实分位数）。针对 $\boldsymbol{n = 20}$ 和 $\boldsymbol{n = 40}$，按 n、 P_{ext} 和 P_{miss} 排序（第 4 个面板中的第 3 个）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子面板均显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、n.sim 和 ρ）。

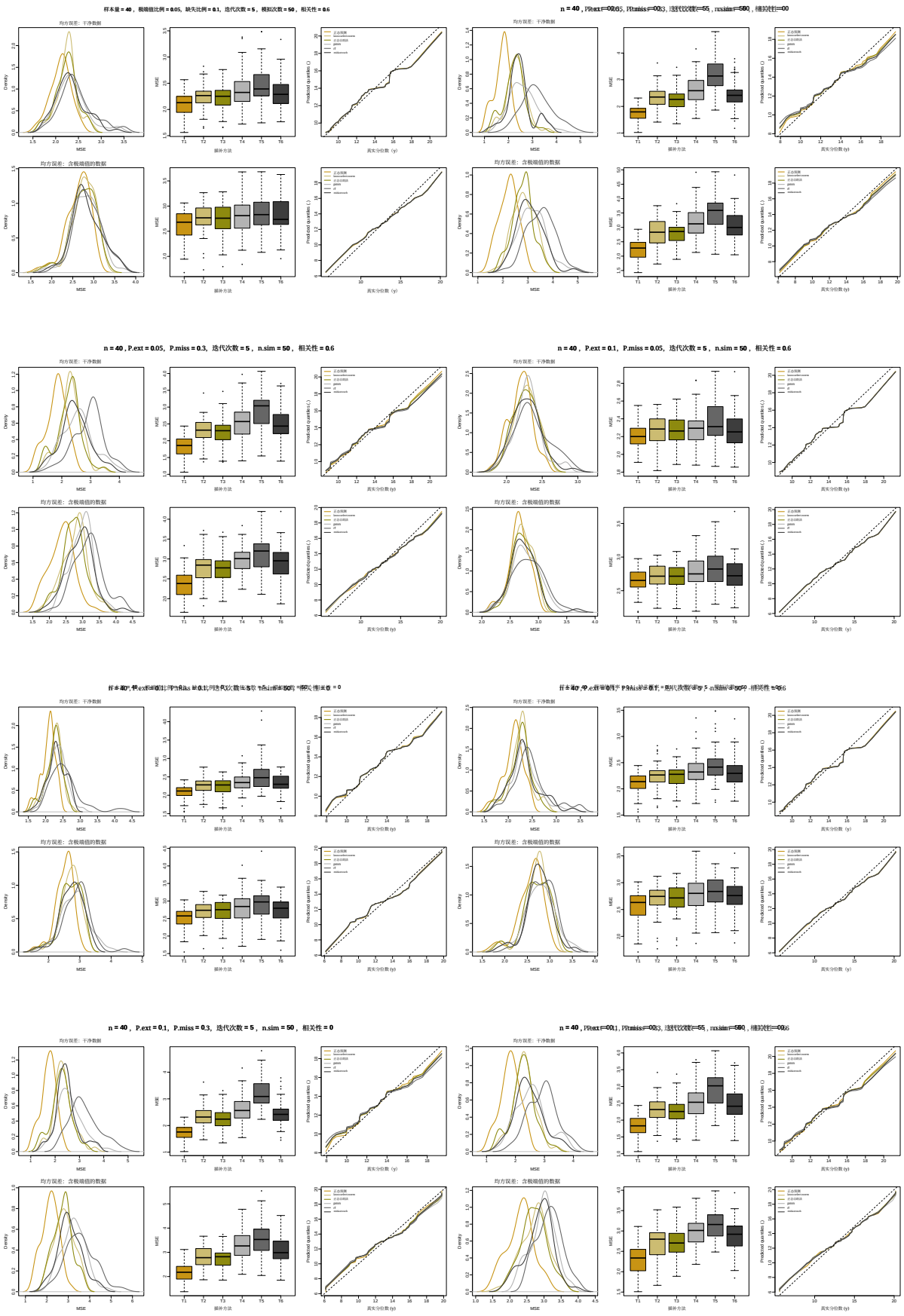

图6： 预测均方误差密度（干净数据与极端值污染数据）、均方误差箱线图，以及六种多重插补方法（T1–T6）的 QQ 图（ y 的预测分位数与真实分位数对比）。对于 $\boldsymbol{n = 40}$，按P_{ext}排序，并采用 P_{miss} （第 2 个面板，共 4 个）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子面板显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、n.sim 和 ρ）。

表5：按模型（干净数据使用普通最小二乘法；污染数据使用弹性网（EN））和缺失分块（低：$P_{\text{miss}} \in \{0.05, 0.10\}$；高：$P_{\text{miss}} \in \{0.25, 0.30\}$）分层的 $\boldsymbol{n = 20}$（左）和$\boldsymbol{n = 40}$（右）各多重插补方法的样本外预测误差汇总（交叉验证均方误差）。对于每种方法，报告均方误差的最小值至最大值：均值 X、方差 σ^2和中位数 Q_{50}，均基于相应设计单元计算。范围还汇总了所有设计组合 $P_{\text{ext}} \in \{0.03, 0.05, 0.10, 0.15, 0.30\}$、$n_{\text{sim}} \in \{50, 300, 1000\}$、迭代次数 $\in \{5, 10\}$和 $\rho \in \{0, 0.6\}$。

	n = 20			**n = 40**		
MI	$\overline{X}$	σ^2	Q_{50}	$\overline{X}$	σ^2	Q_{50}
线性回归（普通最小二乘法）： 低缺失数据比例（5%和10%）						
T1	[2.283 , 2.524]	[0.236, 0.461]	[2.206 , 2.450]	[2.077, 2.207]	[0.020 , 0.050]	[2.110, 2.241]
T2	[2.441, 2.690]	[0.200, 0.413]	[2.407, 2.576]	[2.242, 2.271]	[0.019, 0.057]	[2.265, 2.285]
T3	[2.399, 2.620]	[0.196, 0.369]	[2.336 , 2.500]	[2.236, 2.275]	[0.018, 0.061]	[2.228 , 2.280]
T4	[2.473, 2.693]	[0.240, 0.429]	[2.433 , 2.578]	[2.273, 2.357]	[0.023, 0.125]	[2.260, 2.346]
T5	[2.611, 2.909]	[0.173, 0.736]	[2.553, 2.766]	[2.351, 2.560]	[0.053, 0.213]	[2.312, 2.479]
T6	[2.432 , 2.672]	[0.226 , 0.465]	[2.399 , 2.547]	[2.273, 2.337]	[0.024, 0.112]	[2.253, 2.354]
线性回归（普通最小二乘法）： 高缺失数据比例（25% 和 30%）						
T1	[1.912, 2.166]	[0.252, 0.496]	[1.956 , 2.081]	[1.725, 1.819]	[0.093 , 0.121]	[1.757, 1.858]
T2	[2.484 , 2.628]	[0.292, 0.419]	[2.470 , 2.666]	[2.256 , 2.334]	[0.123 , 0.176]	[2.275 , 2.351]
T3	[2.296, 2.494]	[0.265 , 0.415]	[2.390, 2.491]	[2.236, 2.261]	[0.158, 0.193]	[2.230 , 2.297]
T4	[2.635, 2.712]	[0.246, 0.428]	[2.631, 2.737]	[2.516, 2.605]	[0.274, 0.312]	[2.480, 2.589]
T5	[3.015, 3.355]	[0.204, 0.524]	[3.070 , 3.390]	[2.890, 3.201]	[0.248 , 0.403]	[2.878 , 3.140]
T6	[2.506 , 2.616]	[0.235 , 0.501]	[2.530, 2.666]	[2.434, 2.470]	[0.221, 0.340]	[2.364, 2.435]
稀疏回归（弹性网络）： 低缺失数据比例（5% 和 10%）						
T1	[1.441, 1.995]	[0.038, 0.095]	[1.438, 1.943]	[2.505, 2.867]	[0.032, 0.091]	[2.574, 2.867]
T2	[1.577, 2.118]	[0.046 , 0.162]	[1.566, 2.072]	[2.687, 2.923]	[0.033 , 0.113]	[2.716 , 2.964]
T3	[1.558, 2.083]	[0.045, 0.142]	[1.549 , 2.056]	[2.704 , 2.911]	[0.035 , 0.116]	[2.715, 2.948]
T4	[1.906 , 2.553]	[0.175, 1.002]	[1.675, 2.274]	[2.771, 2.925]	[0.057, 0.199]	[2.750, 2.944]
T5	[2.275, 3.103]	[0.302, 1.721]	[1.935 , 2.740]	[2.822, 3.045]	[0.059 , 0.246]	[2.776 , 3.042]
T6	[1.846 , 2.380]	[0.150, 0.877]	[1.618, 2.180]	[2.735 , 2.914]	[0.056, 0.160]	[2.722 , 2.943]
稀疏回归（弹性网络）： 缺失数据比例较高（25% 和 30%）						
T1	[1.180, 1.635]	[0.072, 0.165]	[1.146, 1.616]	[2.162, 2.529]	[0.136, 0.155]	[2.177, 2.558]
T2	[1.780 , 2.403]	[0.174, 0.361]	[1.698, 2.361]	[2.734, 2.948]	[0.134 , 0.217]	[2.779, 2.969]
T3	[1.703, 2.270]	[0.139, 0.317]	[1.655, 2.179]	[2.710, 2.896]	[0.137, 0.201]	[2.704 , 2.878]
T4	[3.258 , 3.703]	[1.237, 1.734]	[2.974, 3.524]	[2.961, 3.252]	[0.108, 0.334]	[3.009 , 3.264]
T5	[4.243, 5.068]	[1.351, 2.908]	[3.928 , 5.008]	[3.153, 3.560]	[0.123 , 0.505]	[3.136, 3.590]
T6	[2.847, 3.197]	[1.182 , 2.461]	[2.572 , 2.891]	[2.857 , 3.079]	[0.144, 0.322]	[2.919, 2.997]

中等样本量（$\boldsymbol{n = 80}$ 和 $\boldsymbol{n = 200}$）。参见表 6（偏差/均方根误差/覆盖率）和表 7（交叉验证均方误差），以及图 7–10。更多详情请参见附录中的表 18–25。

系数（水平与偏差）。在 $n = 80$、低缺失率的干净普通最小二乘法下，T1–T3的斜率已经表现良好（β_1 偏差 $[0.021, 0.037]$，均方根误差 $[0.030, 0.046]$；β_2 偏差 $[-0.199, -0.115]$，均方根误差 $[0.124, 0.203]$），T4–T6与之相当，但离散程度略高（β_2 均方根误差最高达0.218）。在 $n = 200$，斜率偏差进一步向0收缩（例如，T1–T3 β_1 偏差 $[-0.018, 0.003]$，均方根误差 $[0.011, 0.022]$；β_2 偏差 $[0.002, 0.067]$，均方根误差 $[0.015, 0.071]$），供体/机器学习组也类似地收敛（T4–T6 β_1 均方根误差 $[0.010, 0.022]$，β_2 均方根误差 $[0.014, 0.028]$）。在更高的缺失率（$\geq$ 25%）下，两组在 $n = 80$ 均可见均方根误差膨胀（例如，T4–T6 β_2均方根误差 $[0.209, 0.280]$），但这种上升在 $n = 200$ 被大幅抑制（T4–T6 β_2 均方根误差$[0.037, 0.064]$）。汇总区间（T1–T6）在所有面板中均介于两组之间。在极端值（弹性网络）污染下，三种典型扭曲依然存在：截距水平偏移、β_1 倾斜以及 β_2 收缩，并表现出明显的 n效应。在 $n = 80$ 且低 P_{miss}时，T1–T3显示出小到中等的 β_0 偏差（$[-0.285, 0.047]$）以及正的 β_1 偏差（$[0.083, 0.136]$），而 β_2被拉向负值（$[-0.227, -0.144]$）；T4–T6表现类似，负向略强

对 β_2 的拉动（$[-0.252, -0.196]$）。在 $n = 200$，符号 依然保持（截距膨胀、正的 β_1、负的 β_2），但离散程度急剧收缩：例如，T1–T3 β_1 均方根误差降至 $[0.010, 0.091]$，β_2 均方根误差降至 $[0.017, 0.145]$；T4–T6呈现类似的收紧。因此，提高 n 会降低波动性，但并不能消除极端值所引起的结构性偏移。

覆盖率。 在干净数据上使用普通最小二乘法时，低缺失率下各组的覆盖率基本达到名义水平（所有系数均为 ≈ 1.00 ），并且在 $n = 200$ 时仍保持较高水平。在较高缺失率下，覆盖率下降主要集中于 β_2：在 $n = 80$ 时，T4–T6 达到 0.84–0.96；在 $n = 200$ 时，参数法 区块在某些单元格中的覆盖率可能显著不足（最低为 0.10），尽管偏差较小，这也是均方根误差与覆盖率因方差低估而出现分歧的一个实例。在污染条件下（弹性网络），截距覆盖率仍不可靠——即使在 $n = 200$ 时，也有多个单元格接近于零 —— 而斜率覆盖率相较于 $n = 20$ 和 $n = 40$ 有所改善，但在多个面板中仍低于名义水平。与 T1–T3 相比，供体/机器学习 T4–T6 方法有时能保持覆盖率（尤其是在 β_2 且 P_{miss} 较高时），这与其略低的系数偏差一致。

预测误差（交叉验证均方误差）和尾部风险。 在普通最小二乘法（干净）下，样本量对水平 和离散程度都有显著的稳定作用：对于 $P_{\text{miss}} \leq 0.10$， $n = 80$ 的方差较小（例如，T1 $\sigma^2 = [0.009, 0.019]$、T5最高为 0.030）；而在 $n = 200$，各方法的方差几乎可以忽略（通常为 ≤ 0.01）。与 T1–T3 相比，供体 / 机器学习仍呈右移（例如，在较高的 P_{miss}下， $n = 80$：T5 $X = [2.811, 2.905]$、 $\sigma^2 = [0.049, 0.074]$ 对比 T2 $X = [2.551, 2.600]$、$\sigma^2 = [0.046, 0.056]$），但在 $n = 200$ 时差距缩小（唯一的差异在于 T1，其结果低于其他方法）。

在受污染的弹性网络条件下，均值水平（X）和离散程度（σ^2）都高于普通最小二乘法对应值，表明预测变异性更高；图中也显示密度呈右偏。对于 $n = 80$ 在 P_{miss}较高时，供体/机器学习方法的离散程度更大（例如，T5$\sigma^2 \in [0.058, 0.116]$，T4最高达0.088），而参数法T1–T3的离散程度较小（例如，T1 $\sigma^2 \in [0.023, 0.033]$）。在 $n = 200$时，各方法的离散程度均有所收敛（例如，T5 $\sigma^2 \in [0.022, 0.049]$；T1 $\sigma^2 \in [0.017, 0.022]$），但排序不变：T1–T3的离散度低于T4–T6，不过绝对差距有所缩小。

综合比较n（相较于较小的 $n = 20$，40）。(i) 在干净数据下采用普通最小二乘法时，从 $n = 20$，40增至$n = 80$，200会使斜率偏差向零收敛，并显著降低方差；当 P_{miss} 目标值较高时，残余覆盖失败仍然存在，且取决于方法。(ii) 在污染条件下（弹性网络），失真方向在 n 中保持稳定（截距上升， β_1 上升， β_2 下降）；增大 n 主要降低离散程度和上尾风险，而非消除偏差。(iii) 方法方面，在 $n = 20$，40时观察到的T1–T3与T4–T6之间的权衡依然存在：供体/机器学习方法的系数偏差通常较小（在某些情形下有助于提高覆盖率），但预测分布尾部更厚；参数多重插补的交叉验证—均方误差离散度更小。总体方法（T1–T6）表现居中，其尾部行为受供体/机器学习方法的贡献驱动。

实践指南。 对于中等程度的 n 和干净数据，在低P_{miss}下，方法选择的影响较小；对于 $P_{\text{miss}} \gtrsim 0.25$，应监测 β_2 覆盖率：参数多重插补（T1–T3）即使均方根误差较小，也可能覆盖不足。在存在污染时，若目标是获得稳定的预测（T1–T3），应优先选择尾部更紧的多重插补程序；同时也应认识到，供体/机器学习（T4–T6）可能产生更小的系数偏差，但尾部风险较高。在 $n = 200$下，绝对差异较小，但当控制上尾至关重要时，这些差异仍与决策相关。

表6：在干净数据下使用普通最小二乘法，以及在含极端值的数据下使用弹性网络时，($\beta_0, \beta_1, \beta_2$) 的偏差、均方根误差和95%覆盖率，分别对应 $n = 80$ （左）和 $n = 200$ （右）。每个缺失率区块内的结果以最小值–最大值表示（低： $P_{\text{miss}} \in \{0.05, 0.10\}$；高： $P_{\text{miss}} \in \{0.25, 0.30\}$）。范围还汇总了所有设计组合 $P_{\text{ext}} \in \{0.03, 0.05, 0.10, 0.15, 0.30\}$、 $n_{\text{sim}} \in \{50, 300, 1000\}$、迭代次数 $\in \{5, 10\}$ 和 $\rho \in \{0, 0.6\}$ 下的结果。多重插补方法分为参数法（T1–T3）、供体/机器学习（T4–T6）和组合类别（T1–T6）

		n = 80			**n = 200**		
MI	指标	β_0	β_1	β_2	β_0	β_1	β_2
线性回归（普通最小二乘法）： 低缺失数据比例（5% 和 10%）							
T1–T3	Bias	[0.147, 0.425]	[0.021, 0.037]	[-0.199, -0.115]	[-0.268, 0.034]	[-0.018, 0.003]	[0.002, 0.067]
T1–T3	RMSE	[0.301, 0.487]	[0.030 , 0.046]	[0.124 , 0.203]	[0.128 , 0.350]	[0.011, 0.022]	[0.015, 0.071]
T1–T3	覆盖率	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]
T4–T6	Bias	[0.359 , 0.475]	[0.026 , 0.037]	[-0.213 , -0.183]	[-0.047, 0.144]	[-0.009, 0.001]	[-0.010 , 0.010]
T4–T6	RMSE	[0.411, 0.533]	[0.037 , 0.045]	[0.189 , 0.218]	[0.127, 0.295]	[0.010, 0.022]	[0.014 , 0.028]
T4–T6	覆盖率	[1.000 , 1.000]	[1.000 , 1.000]	[0.980, 1.000]	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]
T1–T6	Bias	[0.147, 0.475]	[0.021, 0.037]	[-0.213, -0.115]	[-0.268, 0.144]	[-0.018, 0.003]	[-0.010, 0.067]
T1–T6	RMSE	[0.301, 0.533]	[0.030, 0.046]	[0.124, 0.218]	[0.127, 0.350]	[0.010, 0.022]	[0.014 , 0.071]
T1–T6	覆盖率	[1.000 , 1.000]	[1.000 , 1.000]	[0.980, 1.000]	[1.000, 1.000]	[1.000, 1.000]	[1.000, 1.000]
线性回归（普通最小二乘法）： 高缺失数据比例（25%和30%）							
T1–T3	Bias	[-0.368, 0.444]	[-0.007 , 0.039]	[-0.207 , 0.047]	[-0.872 , -0.004]	[-0.045, 0.018]	[-0.001 , 0.196]
T1–T3	RMSE	[0.485, 0.683]	[0.041, 0.059]	[0.075, 0.230]	[0.336 , 0.990]	[0.029 , 0.055]	[0.035 , 0.201]
T1–T3	覆盖率	[0.980, 1.000]	[0.993 , 1.000]	[0.900, 1.000]	[0.680, 1.000]	[0.940, 1.000]	[0.100, 1.000]
T4–T6	Bias	[0.354, 0.632]	[0.032, 0.049]	[-0.259 , -0.188]	[-0.175, 0.473]	[-0.020 , 0.015]	[-0.051, 0.015]
T4–T6	RMSE	[0.527, 0.796]	[0.050, 0.062]	[0.209, 0.280]	[0.385 , 0.656]	[0.033 , 0.042]	[0.037 , 0.064]
T4–T6	覆盖率	[1.000 , 1.000]	[1.000 , 1.000]	[0.840, 0.960]	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]
T1–T6	Bias	[-0.368, 0.632]	[-0.007 , 0.049]	[-0.259 , 0.047]	[-0.872, 0.473]	[-0.045, 0.018]	[-0.051 , 0.196]
T1–T6	RMSE	[0.485, 0.796]	[0.041, 0.062]	[0.075, 0.280]	[0.336 , 0.990]	[0.029 , 0.055]	[0.035 , 0.201]
T1–T6	覆盖率	[0.980, 1.000]	[0.993 , 1.000]	[0.840, 1.000]	[0.680, 1.000]	[0.940, 1.000]	[0.100, 1.000]
稀疏回归（弹性网络）： 缺失数据比例较低（5% 和 10%）							
T1–T3	Bias	[-0.285, 0.047]	[0.083, 0.136]	[-0.227 , -0.144]	[0.663, 1.738]	[-0.090 , 0.011]	[-0.144, 0.007]
T1–T3	RMSE	[0.113, 0.346]	[0.086, 0.139]	[0.145, 0.228]	[0.666, 1.743]	[0.010, 0.091]	[0.017, 0.145]
T1–T3	覆盖率	[1.000, 1.000]	[0.980, 1.000]	[0.188, 1.000]	[0.000, 1.000]	[0.560, 1.000]	[0.100, 1.000]
T4–T6	Bias	[-0.352, -0.010]	[0.109, 0.155]	[-0.252, -0.196]	[0.682 , 1.830]	[-0.090 , 0.011]	[-0.151, -0.058]
T4–T6	RMSE	[0.100, 0.437]	[0.112, 0.158]	[0.197, 0.255]	[0.684 , 1.833]	[0.013 , 0.090]	[0.061 , 0.152]
T4–T6	覆盖率	[1.000 , 1.000]	[0.920, 1.000]	[0.170, 1.000]	[0.000 , 1.000]	[0.800, 1.000]	[0.040 , 1.000]
T1–T6	Bias	[-0.352 , 0.047]	[0.083, 0.155]	[-0.252, -0.144]	[0.663, 1.830]	[-0.090 , 0.011]	[-0.151, 0.007]
T1–T6	RMSE	[0.100, 0.437]	[0.086, 0.158]	[0.145, 0.255]	[0.666, 1.833]	[0.010, 0.091]	[0.017, 0.152]
T1–T6	覆盖率	[1.000 , 1.000]	[0.920, 1.000]	[0.170, 1.000]	[0.000 , 1.000]	[0.560 , 1.000]	[0.040 , 1.000]
稀疏回归（弹性网络）： 高缺失数据比例（25% 和 30%）							
T1–T3	Bias	[-0.506, 0.039]	[0.037 , 0.138]	[-0.221, -0.014]	[0.604, 1.717]	[-0.122 , 0.013]	[-0.148, 0.151]
T1–T3	RMSE	[0.261, 0.707]	[0.057, 0.148]	[0.055, 0.228]	[0.665, 1.733]	[0.027 , 0.125]	[0.037 , 0.155]
T1–T3	覆盖率	[0.980, 1.000]	[0.880, 1.000]	[0.673, 1.000]	[0.000 , 0.900]	[0.000 , 1.000]	[0.380, 1.000]
T4–T6	Bias	[-0.707 , -0.016]	[0.108, 0.215]	[-0.334 , -0.199]	[0.715 , 1.976]	[-0.106, 0.018]	[-0.175 , -0.066]
T4–T6	RMSE	[0.233, 0.939]	[0.116, 0.228]	[0.207, 0.342]	[0.751, 1.983]	[0.024, 0.111]	[0.076 , 0.177]
T4–T6	覆盖率	[0.920, 1.000]	[0.660 , 0.967]	[0.380, 0.920]	[0.000 , 0.940]	[0.500, 1.000]	[0.180, 1.000]
T1–T6	Bias	[-0.707, 0.039]	[0.037 , 0.215]	[-0.334 , -0.014]	[0.604, 1.976]	[-0.122, 0.018]	[-0.175, 0.151]
T1–T6	RMSE	[0.233 , 0.939]	[0.057, 0.228]	[0.055, 0.342]	[0.665, 1.983]	[0.024, 0.125]	[0.037 , 0.177]
T1–T6	覆盖率	[0.920, 1.000]	[0.660, 1.000]	[0.380, 1.000]	[0.000 , 0.940]	[0.000 , 1.000]	[0.180, 1.000]

* 1 6 1 norm.predict 2 lasso.select.norm 3 norm.boot 4 pmm 5 图例：多重插补方法（T –T ）：
rf 6 midastouch 95
T ; T ; T ; T ; T ; T 。覆盖率是包含真实系数的 % 区间所占的比例。

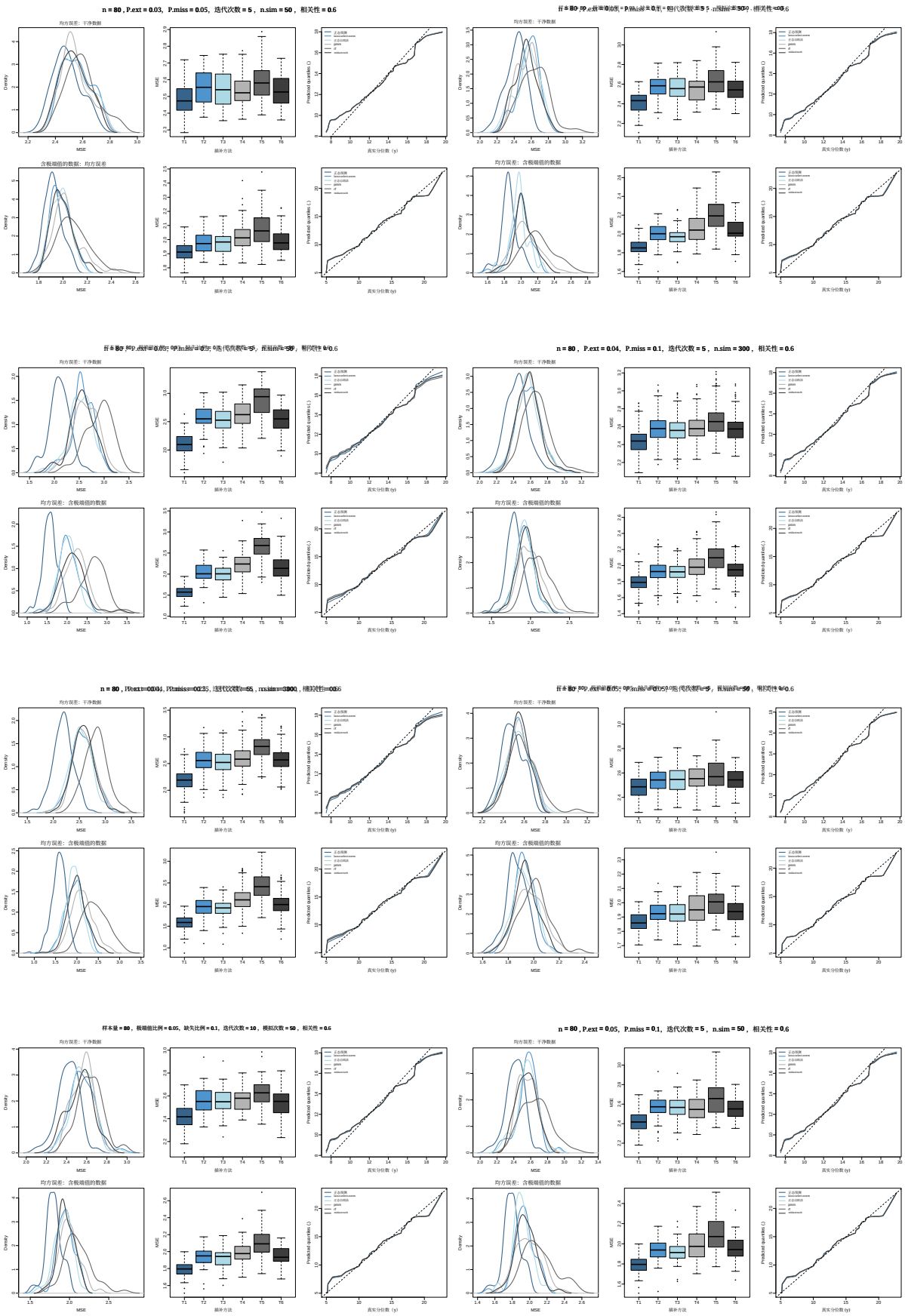

图 7： 预测均方误差密度（干净数据与极端值污染数据）、均方误差箱线图，以及六种多重插补方法（T1–T6）的 QQ 图（ y 的预测分位数与真实分位数对比）。对于 $\boldsymbol{n = 80}$，按 P_{ext} 排序，且 P_{miss} （第 1 个面板，共 4 个）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子图均显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、n.sim 和 ρ）。

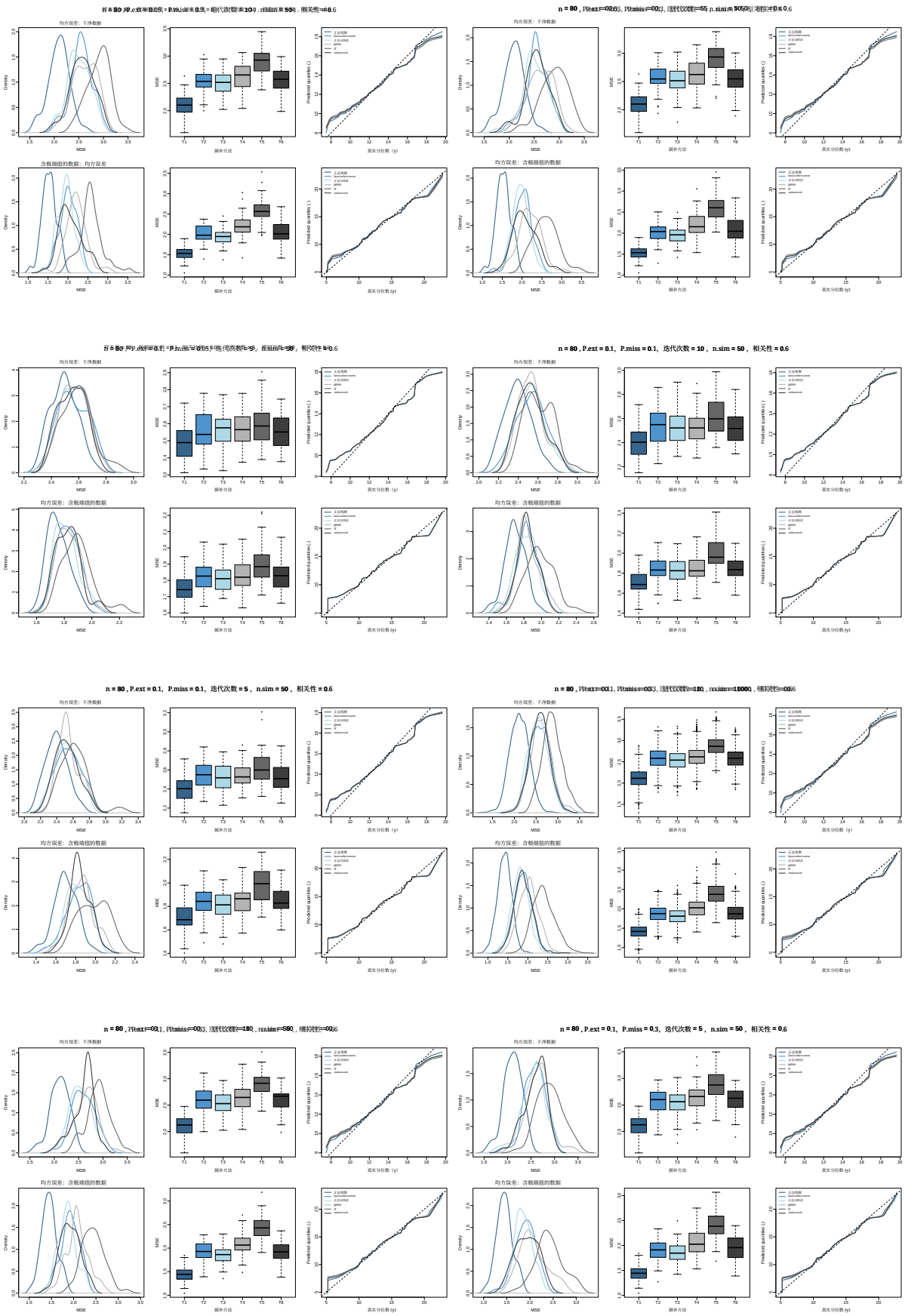

图 8： 预测均方误差密度（干净数据与极端值污染数据对比）、均方误差箱线图，以及六种多重插补方法（T1–T6）的 QQ 图（ y 的预测分位数与真实分位数对比）。对于 $n = 80$，按 P_{ext} 排序，且 P_{miss} （第 2 个，共 4 个面板）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子面板均显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、n.sim 和 ρ）。

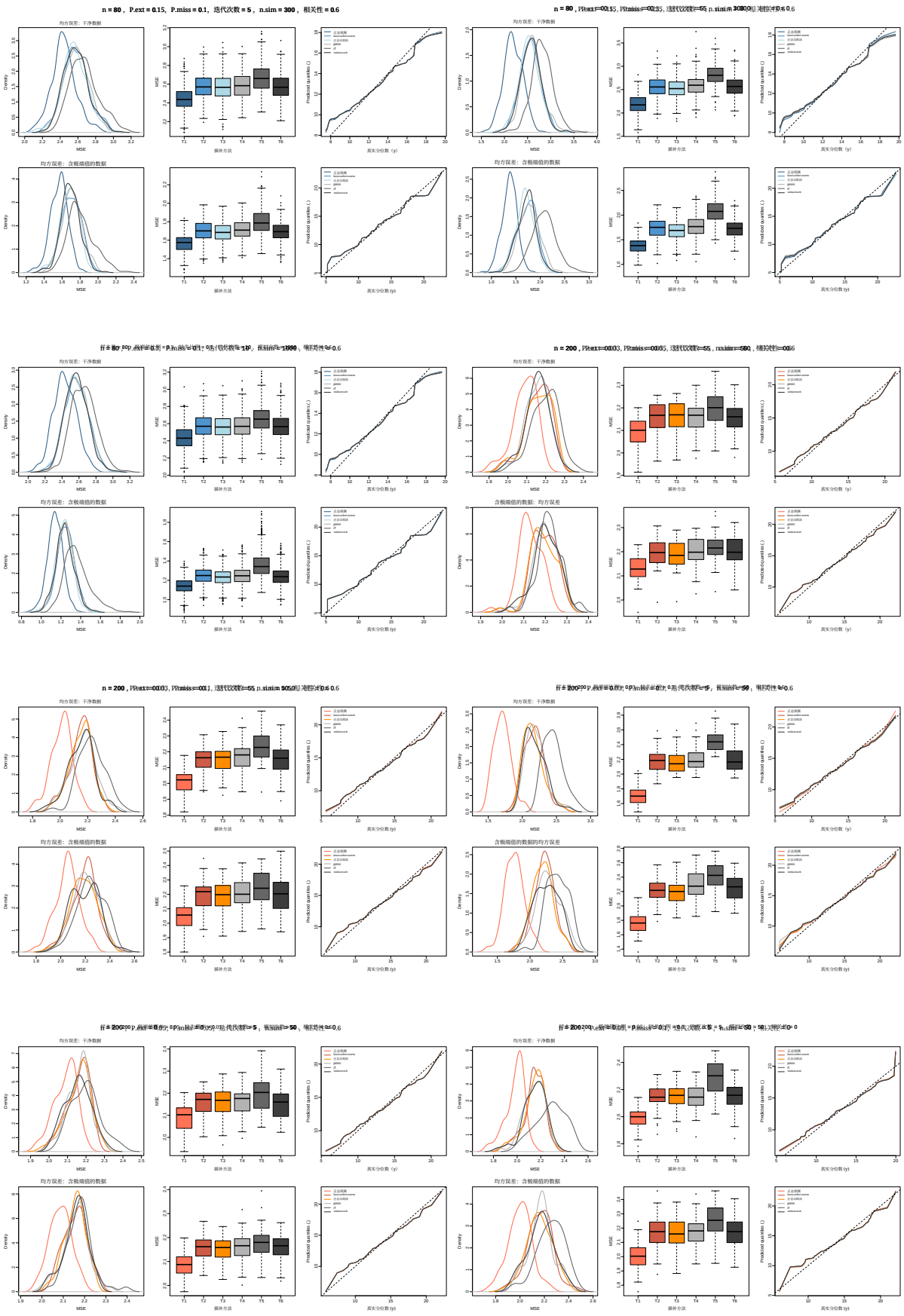

图 9： 预测均方误差密度（干净数据与极端值污染数据）、均方误差箱线图，以及六种多重插补方法（T1–T6）的 QQ图（ y 的预测分位数与真实分位数对比）。对于 $\boldsymbol{n = 80}$ 和 $\boldsymbol{n = 200}$，按 n、 P_{ext} 和 P_{miss} 排序（共 4 个面板中的第 3 个）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子面板均显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、模拟次数和 ρ）。

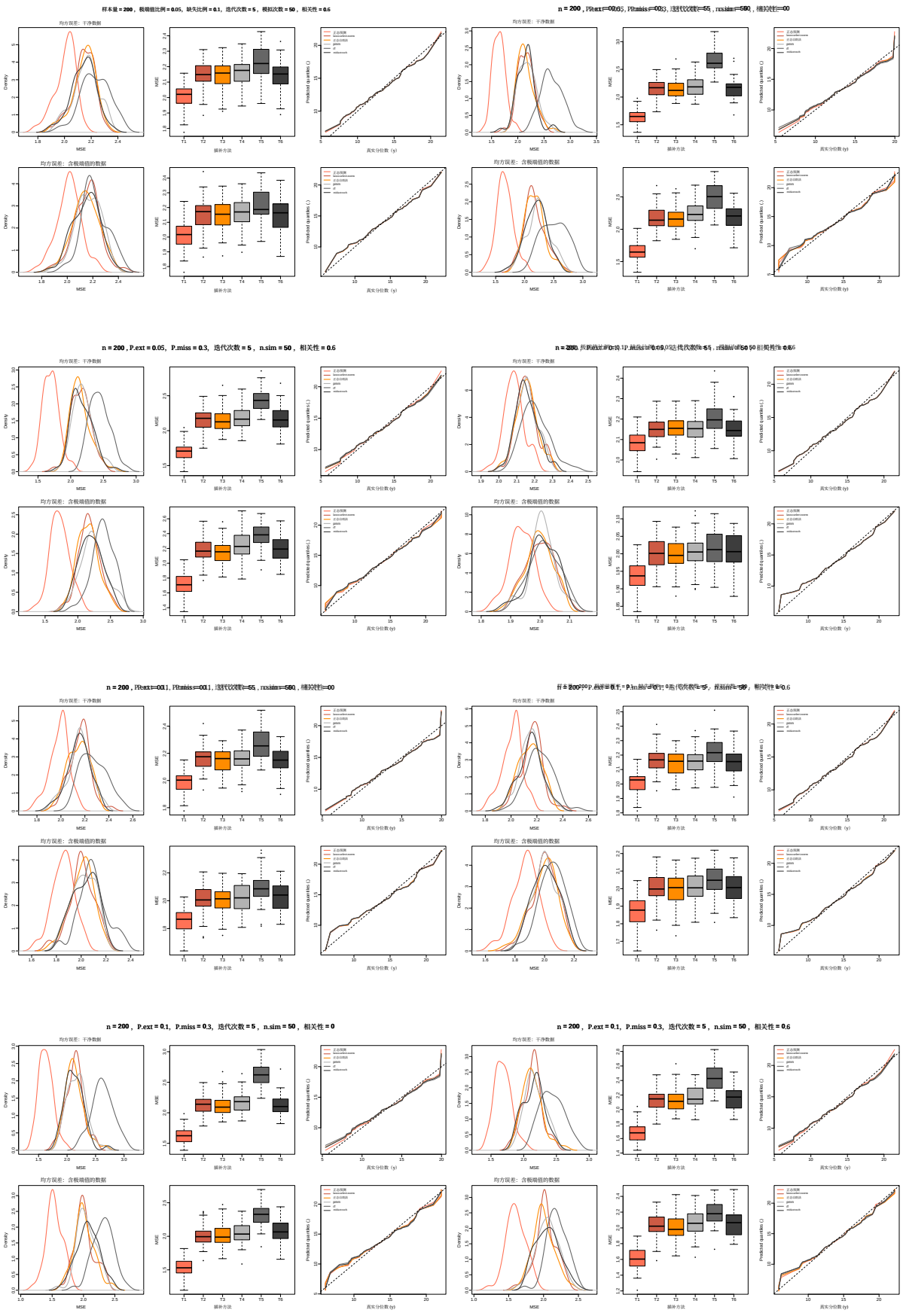

图10：预测均方误差密度（干净数据与极端值污染数据）、均方误差箱线图，以及六种多重插补方法（T1–T6）下 y的预测分位数与真实分位数的QQ图。对于 $\boldsymbol{n = 200}$，按P_{ext}排序，且 P_{miss} （第4个共4个面板）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子面板均显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、n.sim 和 ρ）。预印本。*2026*年*2*月*5*日 26

表7：按模型（干净数据使用普通最小二乘法；污染数据使用弹性网（EN））和缺失分块（低：$P_{\text{miss}} \in \{0.05, 0.10\}$；高：$P_{\text{miss}} \in \{0.25, 0.30\}$）分层的 $\boldsymbol{n = 80}$ (左侧)和$\boldsymbol{n = 200}$ (右侧)各多重插补方法的样本外预测误差汇总（交叉验证均方误差）。对于每种方法，报告均方误差的最小值至最大值：均值 X、方差 σ^2和中位数 Q_{50}，均基于相应的设计单元计算。范围还汇总了所有设计组合 $P_{\text{ext}} \in \{0.03, 0.05, 0.10, 0.15, 0.30\}$、$n_{\text{sim}} \in \{50, 300, 1000\}$、迭代次数 $\in \{5, 10\}$和 $\rho \in \{0, 0.6\}$。

	$\mathbf{n = 80}$			$\mathbf{n = 200}$		
MI	$\overline{X}$	σ^2	Q_{50}	$\overline{X}$	σ^2	Q_{50}
线性回归（普通最小二乘法）： 低缺失数据比例（5% 和 10%）						
T1	[2.397, 2.493]	[0.009, 0.019]	[2.403, 2.490]	[1.983, 2.089]	[0.003, 0.007]	[2.000, 2.103]
T2	[2.536, 2.580]	[0.010, 0.023]	[2.537, 2.583]	[2.141, 2.159]	[0.004, 0.008]	[2.144, 2.175]
T3	[2.523, 2.568]	[0.010, 0.021]	[2.518 , 2.576]	[2.146 , 2.161]	[0.004, 0.009]	[2.155, 2.169]
T4	[2.534, 2.585]	[0.009, 0.019]	[2.522, 2.584]	[2.142 , 2.164]	[0.003, 0.010]	[2.144, 2.180]
T5	[2.589, 2.673]	[0.013 , 0.030]	[2.569, 2.660]	[2.198, 2.284]	[0.005, 0.016]	[2.196, 2.302]
T6	[2.522 , 2.580]	[0.009, 0.020]	[2.509, 2.574]	[2.144, 2.156]	[0.004, 0.009]	[2.144, 2.160]
线性回归（普通最小二乘法）： 高缺失数据比例（25% 和 30%）						
T1	[2.093, 2.189]	[0.036 , 0.047]	[2.095 , 2.191]	[1.628, 1.707]	[0.015, 0.017]	[1.624 , 1.707]
T2	[2.551, 2.600]	[0.046 , 0.056]	[2.537 , 2.601]	[2.131, 2.170]	[0.020 , 0.024]	[2.140 , 2.185]
T3	[2.499 , 2.540]	[0.045, 0.064]	[2.514, 2.561]	[2.132, 2.162]	[0.021, 0.027]	[2.092 , 2.140]
T4	[2.598, 2.652]	[0.049, 0.065]	[2.582, 2.659]	[2.168 , 2.203]	[0.023 , 0.029]	[2.143, 2.184]
T5	[2.811 , 2.905]	[0.049, 0.074]	[2.807, 2.941]	[2.433, 2.648]	[0.021, 0.042]	[2.427, 2.621]
T6	[2.552, 2.600]	[0.043, 0.059]	[2.547, 2.669]	[2.125, 2.189]	[0.024, 0.032]	[2.102 , 2.176]
稀疏回归（弹性网络）：低缺失数据比例（5% 和 10%）						
T1	[1.141 , 1.919]	[0.006, 0.013]	[1.140, 1.912]	[1.856, 2.129]	[0.002, 0.010]	[1.867, 2.128]
T2	[1.249, 2.002]	[0.006 , 0.017]	[1.250 , 2.002]	[1.999 , 2.194]	[0.002 , 0.011]	[1.998, 2.219]
T3	[1.231, 1.979]	[0.006 , 0.015]	[1.234 , 1.983]	[1.991, 2.192]	[0.002 , 0.011]	[1.995, 2.198]
T4	[1.246, 2.058]	[0.007, 0.024]	[1.247, 2.041]	[2.004 , 2.203]	[0.002 , 0.011]	[2.005, 2.201]
T5	[1.363, 2.206]	[0.012, 0.034]	[1.343 , 2.192]	[2.014 , 2.252]	[0.003, 0.014]	[2.012, 2.254]
T6	[1.240 , 2.035]	[0.008 , 0.015]	[1.238 , 2.008]	[2.008, 2.202]	[0.002 , 0.013]	[2.006 , 2.203]
稀疏回归（弹性网络）：缺失数据比例较高（25% 和 30%）						
T1	[1.369, 1.578]	[0.023, 0.033]	[1.381, 1.586]	[1.536, 1.761]	[0.017, 0.022]	[1.531, 1.763]
T2	[1.726, 2.047]	[0.040, 0.059]	[1.751, 2.038]	[2.017, 2.217]	[0.024, 0.034]	[1.994 , 2.216]
T3	[1.679, 2.010]	[0.035 , 0.048]	[1.689, 2.007]	[2.001 , 2.193]	[0.025, 0.029]	[1.981, 2.201]
T4	[1.772, 2.237]	[0.044, 0.088]	[1.767, 2.240]	[2.047, 2.287]	[0.024, 0.043]	[2.034, 2.275]
T5	[2.080, 2.639]	[0.058, 0.116]	[2.075, 2.680]	[2.190, 2.506]	[0.022, 0.049]	[2.180, 2.504]
T6	[1.716 , 2.149]	[0.036 , 0.099]	[1.732 , 2.137]	[2.059, 2.251]	[0.030 , 0.034]	[2.065, 2.265]

相对较大的样本量（$\boldsymbol{n = 500}$）。请参见表 8–9 和图 12–11。详情请参见附录中的表 26–28。

系数（水平和偏差）。在干净数据（普通最小二乘法）且缺失率较低（$\leq$ 10%）时，所有方法组在$n = 500$下的斜率基本无偏。对于 T1–T3，β_1 偏差位于 $[-0.007, 0.018]$，均方根误差为 $[0.008, 0.020]$；β_2 偏差位于 $[-0.058, 0.013]$，均方根误差为 $[0.017, 0.059]$。T4–T6 的结果相近，斜率均方根误差略大（例如，β_2 均方根误差 $[0.055, 0.066]$），且有轻微的负向 β_2 偏差 $[-0.064, -0.053]$。缺失率较高（$\geq$ 25%）时，两个方法组的普通最小二乘法斜率均方根误差均适度增加（例如，T1–T3：β_1 均方根误差 $[0.024, 0.067]$，β_2 均方根误差 $[0.060, 0.176]$；T4–T6：$[0.025, 0.033]$ 和 $[0.060, 0.088]$），而 T4–T6 的偏差较小且稳定（例如，β_2 偏差$[-0.083, -0.055]$）。在极端值污染下（弹性网络），大n面板保留了相同的方向性模式：截距膨胀、β_1负向偏移，以及 β_2正向偏移。对于$P_{\text{miss}} \leq 0.10$，T1–T3 的 β_0 偏差为 $[0.076, 0.270]$，β_1 偏差为 $[-0.108, -0.060]$，且 β_2 偏差为 $[0.080,0.162]$；T4–T6 的方向相似，斜率偏差略小（例如，β_1 $[-0.083,-0.056]$、β_2 $[0.076, 0.109]$）。当 $P_{\text{miss}} \geq 0.25$时，T1–T3 的 β_1 倾斜加剧（$[-0.168, -0.066]$），且 β_2 偏差范围扩大（$[0.076, 0.284]$）；而 T4–T6 的斜率偏差仍相对较小

（例如， β_2 [0.046, 0.112]），且截距偏差稍大（β_0 最高达 0.367）。考虑到样本量，这些差异的幅度仍然较小。

覆盖率。对于干净数据上的普通最小二乘法，所有方法在 $P_{\text{miss}} \leq 0.10$ 下的覆盖率均达到名义水平。在 $P_{\text{miss}} \in \{0.25, 0.30\}$ 时，覆盖率下降主要集中在斜率的参数块中，尤以 β_2 最为明显（T1–T3： β_1 覆盖率 [0.580, 1.000], β_2 [0.000, 1.000]）；而供体/机器学习块对所有系数的覆盖率均接近名义水平（T4–T6： β_1 [0.980, 1.000],β_2 [0.960, 1.000]）。在弹性网络下，截距覆盖率通常较高（T1–T3： [0.912, 1.000]；T4–T6： [0.920, 1.000]），而斜率覆盖率，尤其是 β_1，即使在 $n = 500$ 时也可能低于名义水平（例如，缺失率较高时，T1–T3： β_1 覆盖率 [0.000, 0.840]）。随着缺失率上升，供体/机器学习方法通常能比 T1–T3 提高β_2 覆盖率（例如， [0.420, 1.000] 与 [0.000, 0.980]相比），这与它们在污染情况下较小的 β_2 偏差一致。

预测误差（交叉验证均方误差）和离散程度。预测分布高度集中在 $n = 500$。对于 $P_{\text{miss}} \leq 0.10$，普通最小二乘法下的方法均值$X$ 集中在狭窄区间内（例如，T1：[2.081, 2.153]；T5： [2.224, 2.255]），方差 σ^2 接近 10^{-3}。缺失程度较高时，离散程度略有增加，但仍然较小（例如，T1： σ^2 = [0.005, 0.008]；T5： [0.006, 0.010]）。在弹性网络下，尽管存在污染，水平和方差仍保持稳定。对于较低的 P_{miss}，T1–T3达到最低或接近最低的X （例如，T1： [1.342, 2.080]），且 $\sigma^2 \leq 0.004$；而T5往往右移（例如， [1.499, 2.203]）。对于 $P_{\text{miss}} \geq 0.25$，这一排序仍然成立（T1–T3的水平和离散程度均低于T4–T6）：T1显示 X =[1.494, 1.694] ，且σ^2 = [0.005, 0.006]；T2， X =[1.939, 2.172]，且 σ^2 = [0.006, 0.009]；T5， X =[2.124, 2.354]，且 σ^2 = [0.008, 0.015]。中位数 Q_{50} （以及 $Q_{2.5}$ 和 $Q_{97.5}$）也呈现出相同的排序。

大样本下的方法敏感性与权衡 n。在较小 n 下观察到的参数法（T1–T3）与供体/机器学习法（T4–T6）之间的差异依然存在，但有所减弱：(i) 对于预测，T1–T3的交叉验证均方误差水平和离散度仍较低，包括在污染情形下；(ii) 在高缺失率下使用普通最小二乘法进行推断时，T4–T6的斜率覆盖率更可靠，这与较小的β_2 偏差有关；(iii) 综合方法块（T1–T6）的表现仍处于中间水平，其尾部行为主要由T5驱动。

交叉n 视角。与小样本面板 $n = 20$、40及中等情形$n = 80$、200相比， $n = 500$ 结果显示：(a) 在无污染和受污染情形下，随着 $n = 20 \to 40 \to 80 \to 200 \to 500$ 增加，均方根误差和交叉验证均方误差离散度逐步收窄；(b) 污染的方向性 在所有 n下均持续存在，其特征为截距膨胀、 β_1负向偏移以及 β_2正向偏移；随着 n 增加，变异性会收窄，而偏差并未消失。(c) 不同样本量下的方法排序相同：参数法T1–T3的预测离散度和交叉验证均方误差水平均低于供体/机器学习法T4–T6，这在小 n 下最为明显（例如 $n = 20$、40），在 $n = 80$、200和$n = 500$下仍然存在，但绝对差距较小；(d) 就覆盖率而言，污染情形下 $n = 20$、40时的覆盖率极其脆弱（若干单元中的截距覆盖率接近零），随着 $n = 80$、200而有所改善，并在 $n = 500$时进一步改善；然而β_1 覆盖率在较高 P_{miss}下仍最容易受到弹性网络影响；(e) 在干净数据的高缺失率情形下，从中等 n起，供体/机器学习块的斜率覆盖率保持接近名义水平，而参数法往往低估斜率覆盖率，尤其是 β_2；这种差异在 $n = 500$下仍然存在，但有所减弱。

表 8：按多重插补方法列出的 $n = 500$ 的样本外预测误差汇总（交叉验证均方误差），并按模型（干净数据使用普通最小二乘法；污染数据使用弹性网（EN））和缺失分块（低：$P_{\text{miss}} \in \{0.05, 0.10\}$；高：$P_{\text{miss}} \in \{0.25, 0.30\}$）进行分层。对于每种方法，报告均方误差的最小值–最大值：均值 X、方差 σ^2 和中位数 Q_{50}，涵盖相应的设计单元。取值范围进一步汇总了所有设计组合 $P_{\text{ext}} \in \{0.03, 0.05, 0.10, 0.15, 0.30\}$、$n_{\text{sim}} \in \{50, 300, 1000\}$、迭代次数 $\in \{5, 10\}$ 和 $\rho \in \{0, 0.6\}$。

		低缺失率（5% 和 10%）			高缺失率（25% 和 30%）		
MI	**模型**	$\overline{X}$	σ^2	Q_{50}	$\overline{X}$	σ^2	Q_{50}
T1	OLS	[2.081 , 2.153]	[0.001, 0.002]	[2.079 , 2.154]	[1.784 , 1.797]	[0.005, 0.008]	[1.786 , 1.796]
T1	EN	[1.342, 2.080]	[0.001, 0.003]	[1.343 , 2.083]	[1.494, 1.694]	[0.005, 0.006]	[1.495, 1.694]
T2	OLS	[2.208, 2.222]	[0.001 , 0.003]	[2.202, 2.222]	[2.211, 2.230]	[0.007, 0.011]	[2.213 , 2.229]
T2	EN	[1.443 , 2.154]	[0.001, 0.003]	[1.446, 2.158]	[1.939 , 2.172]	[0.006 , 0.009]	[1.940, 2.165]
T3	OLS	[2.209, 2.220]	[0.001, 0.003]	[2.206 , 2.221]	[2.208, 2.221]	[0.007, 0.012]	[2.194 , 2.220]
T3	EN	[1.442, 2.151]	[0.001 , 0.003]	[1.444, 2.149]	[1.924 , 2.153]	[0.007, 0.009]	[1.923 , 2.143]
T4	OLS	[2.211, 2.224]	[0.001, 0.003]	[2.204 , 2.226]	[2.221, 2.235]	[0.005, 0.011]	[2.226 , 2.246]
T4	EN	[1.450 , 2.155]	[0.001, 0.004]	[1.450 , 2.160]	[1.931, 2.160]	[0.007, 0.010]	[1.934, 2.151]
T5	OLS	[2.224, 2.255]	[0.001, 0.003]	[2.221, 2.255]	[2.323 , 2.335]	[0.006 , 0.010]	[2.314 , 2.340]
T5	EN	[1.499 , 2.203]	[0.001, 0.004]	[1.491 , 2.213]	[2.124, 2.354]	[0.008 , 0.015]	[2.122, 2.340]
T6	OLS	[2.207, 2.225]	[0.001, 0.003]	[2.211, 2.226]	[2.221, 2.237]	[0.006 , 0.012]	[2.232, 2.235]
T6	EN	[1.453, 2.150]	[0.001, 0.004]	[1.454, 2.156]	[1.936, 2.147]	[0.008, 0.009]	[1.931, 2.149]

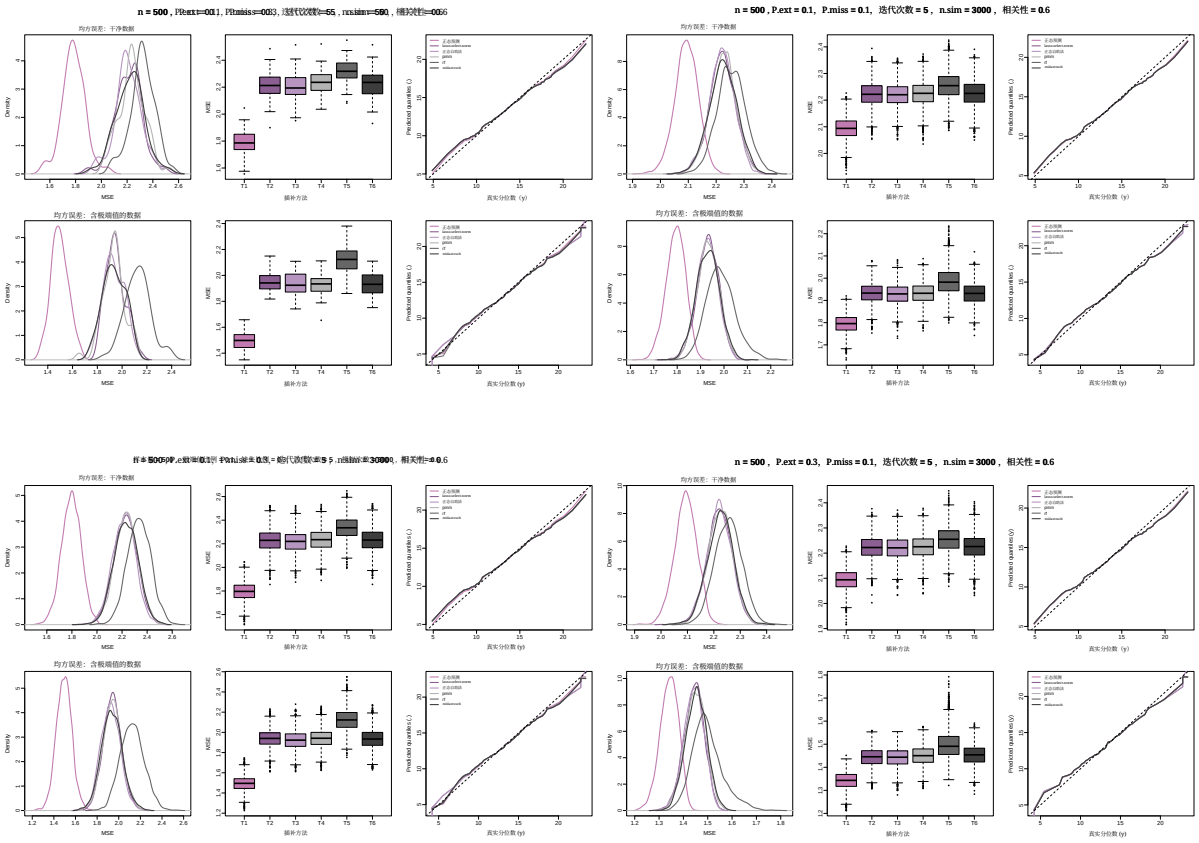

图11： 六种多重插补方法（T1–T6）下的预测均方误差密度（干净数据与极端值污染数据对比）、均方误差箱线图和QQ图（ y的预测分位数与真实分位数对比）。对于 $n = 500$，按P_{ext}排序，并显示 P_{miss} （第1个面板，共2个）。干净数据采用普通最小二乘法进行分析，污染数据采用弹性网络进行分析。每个子面板展示设计值（n、 P_{ext}、 P_{miss}、迭代次数、 n_{sim}和 ρ）

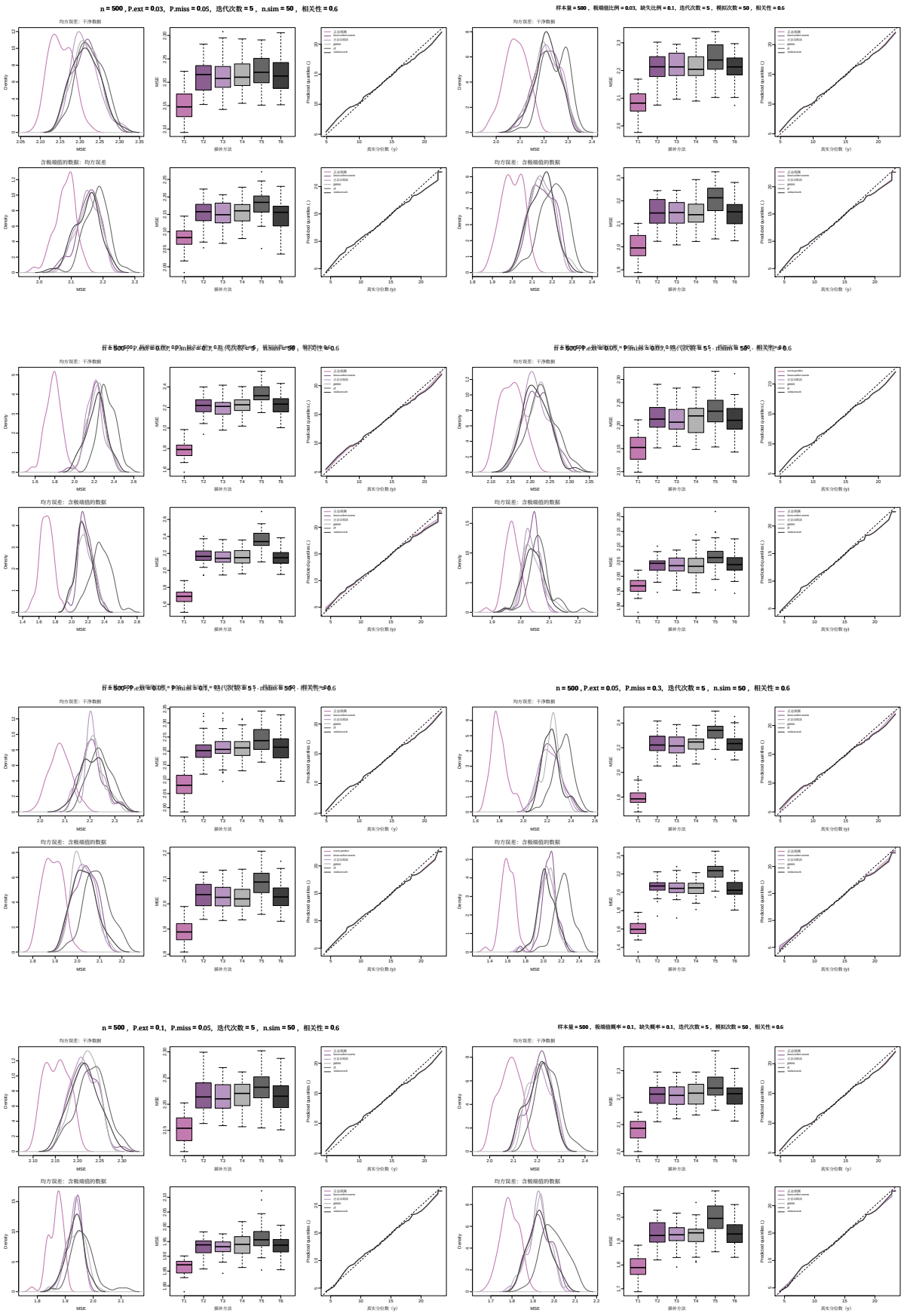

图12： 预测均方误差密度（干净数据与极端值污染数据）、均方误差箱线图，以及六种多重插补方法（T1–T6）的QQ图（ y的预测分位数与真实分位数）。对于 $\boldsymbol{n = 500}$，按P_{ext}排序，以及 P_{miss} （第2个面板，共2个）。干净数据采用普通最小二乘法分析，污染数据采用弹性网络分析。每个子面板显示设计值（n、 P_{ext}、 P_{miss}、迭代次数、 n_{sim}和 ρ）预印本。*2026*年*2*月*5*日 30

表9：在干净数据下使用普通最小二乘法，以及含极端值的数据下使用弹性网络时，针对 $\boldsymbol{n = 20}$（左）和 $\boldsymbol{n = 40}$（右）的 $(\beta_0, \beta_1, \beta_2)$ 的偏差、均方根误差和 95%覆盖率。每个缺失率区块中的结果以最小值–最大值表示（低： $P_{\text{miss}} \in \{0.05, 0.10\}$；高： $P_{\text{miss}} \in \{0.25, 0.30\}$）。范围还汇总了所有设计组合 $P_{\text{ext}} \in \{0.03, 0.05, 0.10, 0.15, 0.30\}$、$n_{\text{sim}} \in \{50, 300, 1000\}$、迭代次数 $\in \{5, 10\}$ 以及 $\rho \in \{0, 0.6\}$ 的结果。多重插补方法分为参数法（T1–T3）、供体/机器学习（T4–T6）和合并集合（T1–T6）。覆盖率是包含真实系数的 95%区间所占的比例

		n = 500		
Case	指标	β_0	β_1	β_2
线性回归（普通最小二乘法）		$P_{miss} \in \{0.05,0.10\}$		
T1–T3	Bias	[0.057, 0.177]	[-0.007 , 0.018]	[-0.058 , 0.013]
T1–T3	RMSE	[0.109, 0.206]	[0.008 , 0.020]	[0.017, 0.059]
T1–T3	覆盖率	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]
T4–T6	Bias	[0.158, 0.188]	[0.016, 0.019]	[-0.064, -0.053]
T4–T6	RMSE	[0.179, 0.215]	[0.018, 0.022]	[0.055, 0.066]
T4–T6	覆盖率	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]
T1–T6	Bias	[0.057, 0.188]	[-0.007 , 0.019]	[-0.064 , 0.013]
T1–T6	RMSE	[0.109, 0.215]	[0.008 , 0.022]	[0.017, 0.066]
T1–T6	覆盖率	[1.000 , 1.000]	[1.000 , 1.000]	[1.000 , 1.000]
线性回归（普通最小二乘法）		$P_{miss} \in \{0.25,0.30\}$		
T1–T3	Bias	[-0.200, 0.206]	[-0.064 , 0.017]	[-0.061 , 0.174]
T1–T3	RMSE	[0.251 , 0.294]	[0.024, 0.067]	[0.060 , 0.176]
T1–T3	覆盖率	[0.960, 1.000]	[0.580, 1.000]	[0.000, 1.000]
T4–T6	Bias	[0.138, 0.227]	[0.017 , 0.026]	[-0.083 , -0.055]
T4–T6	RMSE	[0.222, 0.299]	[0.025, 0.033]	[0.060 , 0.088]
T4–T6	覆盖率	[0.997, 1.000]	[0.980, 1.000]	[0.960, 1.000]
T1–T6	Bias	[-0.200, 0.227]	[-0.064 , 0.026]	[-0.083 , 0.174]
T1–T6	RMSE	[0.222 , 0.299]	[0.024, 0.067]	[0.060 , 0.176]
T1–T6	覆盖率	[0.960 , 1.000]	[0.580, 1.000]	[0.000 , 1.000]
稀疏回归（弹性网络）		$P_{miss} \in \{0.05,0.10\}$		
T1–T3	Bias	[0.076 , 0.270]	[-0.108 , -0.060]	[0.080, 0.162]
T1–T3	RMSE	[0.087 , 0.285]	[0.061 , 0.108]	[0.081 , 0.163]
T1–T3	覆盖率	[1.000, 1.000]	[0.000, 0.860]	[0.000, 1.000]
T4–T6	Bias	[0.082, 0.289]	[-0.083, -0.056]	[0.076 , 0.109]
T4–T6	RMSE	[0.092 , 0.300]	[0.057, 0.084]	[0.076 , 0.110]
T4–T6	覆盖率	[0.997, 1.000]	[0.000, 0.900]	[0.300, 1.000]
T1–T6	Bias	[0.076, 0.289]	[-0.108, -0.056]	[0.076 , 0.162]
T1–T6	RMSE	[0.087 , 0.300]	[0.057, 0.108]	[0.076 , 0.163]
T1–T6	覆盖率	[0.997 , 1.000]	[0.000 , 0.900]	[0.000 , 1.000]
稀疏回归（弹性网络）		$P_{miss} \in \{0.25,0.30\}$		
T1–T3	Bias	[0.007, 0.294]	[-0.168, -0.066]	[0.076 , 0.284]
T1–T3	RMSE	[0.150, 0.334]	[0.068 , 0.169]	[0.079, 0.285]
T1–T3	覆盖率	[0.912, 1.000]	[0.000 , 0.840]	[0.000 , 0.980]
T4–T6	Bias	[0.206, 0.367]	[-0.086, -0.055]	[0.046 , 0.112]
T4–T6	RMSE	[0.236, 0.385]	[0.058 , 0.088]	[0.050, 0.115]
T4–T6	覆盖率	[0.920, 1.000]	[0.272 , 0.940]	[0.420, 1.000]
T1–T6	Bias	[0.007, 0.367]	[-0.168, -0.055]	[0.046 , 0.284]
T1–T6	RMSE	[0.150, 0.385]	[0.058 , 0.169]	[0.050, 0.285]
T1–T6	覆盖率	[0.912, 1.000]	[0.000, 0.940]	[0.000, 1.000]

* 1 6 1 norm.predict 2 lasso.select.norm 3 图例：多 norm.boot 5 rf 6 midastouch 95 重插补方法（T–T）：T；T；T；T4 pmm；T；T。覆盖率是包含真实系数的 % 区间所占的比例。

4 结论

本研究比较了在 `mice` 中实现的六种多重插补方法：参数回归类方法（T1： `norm.predict`，T2： `lasso.select.norm`，T3： `norm.boot`）、基于供体的方法（T4： `pmm`，T6： `midastouch`），以及一种非参数机器学习方法（T5： `rf`）。研究评估了在干净数据上采用普通最小二乘法分析，以及在逐案例极端污染下采用弹性网络分析时，各方法在不同缺失程度下的表现。

整个设计中呈现出三种稳定的模式。第一，在干净数据且缺失率较低时，斜率偏差可忽略不计；随着样本量增加，覆盖率接近名义水平，且不确定性随 n 显著降低。第二，在干净数据且缺失率较高时，主要差异体现在推断方面：基于供体的方法和灵活方法（T4–T6）更稳定地使斜率覆盖率保持在接近名义水平，而参数块（T1–T3）的斜率覆盖率低于名义水平，其中 β_2 的情况尤为明显，即使偏差和均方根误差都很小。第三，在极端污染下，系数估计相较于干净数据基准呈现系统性偏移。增加 n 主要降低了离散程度和尾部风险，但并未消除污染导致的这些偏移。

预测与推断之间存在明显的权衡。在预测方面，参数块（T1–T3）通常取得更低且更集中的交叉验证均方误差，即使在污染情形下也是如此；而 T5（`rf`）往往产生更大且更分散的误差，尤其是在抽样量较小、缺失率较高时。在干净数据缺失率较高的情况下，对于推断，基于供体的方法（T4 和 T6）以及灵活方法（T5）更常能实现可靠的斜率覆盖率，并减少斜率扭曲。总的来说，多重插补是一种建模选择，而非中立且纯技术性的预处理步骤。插补方法所依据的建模假设会实质性地影响预测性能和推断有效性，也可能显著左右后续结论，尤其是在存在极端值和缺失的情况下。

参考文献

Paul Allison。预测均值匹配插补：前景与风险。*Statistical Horizons*，2015年。

Sylvain Arlot 和 Alain Celisse。模型选择的交叉验证程序综述。*Statistics Surveys*，4:40–79，2010年。 doi:10.1214/09-SS054。URL https://projecteuclid.org/euclid.ssu/1268143839。

Richard Berk、Lawrence Brown、Andreas Buja、Kai Zhang 和 Linda Zhao。有效的选择后推断。*Annals of Statistics*，2013年。参见 Berk*et al.*（2013）关于 PoSI 框架的研究。

Dimitris Bertsimas、Jean Pauphilet 和 Bart Van Parys。稀疏回归。*Statistical Science*，35(4):555–578，2020年。

Bernd Bischl、Michel Lang、Lars Kotthoff、Julia Schiffner、Jakob Richter、Erich Studerus、Giuseppe Casalicchio 和 Zachary M. Jones。mlr：R语言中的机器学习。*Journal of MachineLearning Research*，17(170):1–5，2016。URL https://jmlr.org/papers/v17/15-066.html。

Todd E Bodner。增加缺失数据插补次数后有哪些改进？*Structural equationmodeling: a multidisciplinary journal*，15(4):651–675，2008年。

Leo Breiman. 随机森林。机器学习，45(1):5–32，2001年。

Gavin C. Cawley 和 Nicola L. C. Talbot。论模型选择中的过拟合及其在性能评估中导致的选择偏差。*Journal of Machine Learning Research*，11:2079– 2107，2010。网址 https://www.jmlr.org/papers/v11/cawley10a.html。Andersen Chang、Minjie Wang 和 Genevera I Allen。极端值稀疏回归。《电子统计学杂志》，15(2):5995–6035，2021。Yi Deng、Changgee Chang、Moges Seyoum Ido 和 Qi Long。高维数据背景下适用于一般缺失数据模式的多重插补。《科学报告》，6(1):21689，2016。Gláucia Tatiana Ferrari 和 Vitor Ozaki。气候数据集的缺失数据插补：对极端干旱事件建模的影响。《巴西气象学杂志》，29:21–28，2014。Ronald Aylmer Fisher 和 Leonard Henry Caleb Tippett。抽样中最大或最小成员的频率分布的极限形式。载于《剑桥哲学学会数学学报》，第24卷，第180–190页。剑桥大学出版社，1928。Frank 和 Harrell Jr. aregimpute函数，2023。访问日期：2023年11月27日。Jerome Friedman、Trevor Hastie 和 Robert Tibshirani。通过坐标下降求解广义线性模型的正则化路径。*Journal of Statistical Software*，33(1):1–22，2010年。网址 https://www.jstatsoft.org/v33/i01/。Jerome Friedman、Trevor Hastie 和 Robert Tibshirani。通过坐标下降求解广义线性模型的正则化路径。*Journal of Statistical Software*，33(1):1–22，2010b。Jerome Friedman、Trevor Hastie、Rob Tibshirani、Balasubramanian Narasimhan、Kenneth Tay、Noah Simon 和 Junyang Qian。glmnet软件包。*CRAN R Repositary*，595:874，2021。Philipp Gaffert、Florian Meinfelder 和 Volker Bosch。迈向适当的多重插补预测均值匹配。调查研究方法分部，*JSM 2018*，2016。Boris Gnedenko。关于随机序列最大项的极限分布。《数学年刊》，44(3):423–453，1943。Frank E Harrell 等人。回归建模策略：线性模型、逻辑回归、有序回归和生存分析的应用，第3卷。施普林格，2015。Trevor Hastie、Robert Tibshirani 和 Jerome Friedman。《统计学习基础》。施普林格，第2版，2009。Arthur E Hoerl 和 Robert W Kennard。岭回归：针对非正交问题的有偏估计。《技术计量学》，12(1):55–67，1970年。Peter J. Huber 和 Elvezio M. Ronchetti。《稳健统计学》。Wiley，第2版，2009。Arthur F Jenkinson。气象要素年最大值（或最小值）的频率分布。《英国皇家气象学会季刊》，81(348): 158–171，1955。Michael G Kenward 和 James Carpenter。多重插补：当前观点。《医学研究中的统计方法》，16(3):199–218，2007。

Samuel Kotz and Saralees Nadarajah。《极值分布：理论与应用》。world scientific，2000。Michael H Kutner、Christopher J Nachtsheim、John Neter、William Li等。《应用线性统计模型》，2005。Sophie Lambert-Lacroix和Laurent Zwald。基于Huber准则和自适应Lasso惩罚的稳健回归。《电子统计学杂志》，5:1015–1053，2011。JiaHang Li、ShuXia Guo、RuLin Ma、Jia He、XiangHui Zhang、DongSheng Rui、YuSong Ding、Yu Li、LeYao Jian、Jing Cheng等。队列研究数据集中缺失数据的插补方法在预测建模中的效果比较。《BMC医学研究方法学》，24(1):41，2024。Andy Liaw和Matthew Wiener。基于随机森林的分类与回归。《R新闻》，2(3): 18–22，2002。URL https://CRAN.R-project.org/doc/Rnews/。Roderick JA Little。大型调查中的缺失数据调整。《商业与经济统计学杂志》，6(3):287–296，1988。Roderick JA Little和Donald B Rubin。《缺失数据的统计分析》。John Wiley & Sons，2019。Patrick E McKnight、Katherine M McKnight、Souraya Sidani和Aurelio Jose Figueredo。《缺失数据：简明入门》。Guilford Press，2007。Tim P Morris、Ian R White和Patrick Royston。通过预测均值匹配和局部残差抽取调整多重插补。《BMC医学研究方法学》，14(1):75，2014。Luciana Neves Nunes。《卫生领域数据插补方法》。流行病学博士论文，南里奥格兰德联邦大学，巴西阿雷格里港，2007。医学院。Luciana Neves Nunes、Mariza Machado Klück和Jandyra Maria Guimarães Fachel。以手术死亡风险模型为例比较单次插补与多重插补方法。《巴西流行病学杂志》，13:596–606，2010。Friedrich Pukelsheim。三西格玛规则。《美国统计学家》，48(2):88–91，1994。R核心团队。《R：统计计算语言与环境》。R统计计算基金会，奥地利维也纳，2023。URL https://www.R-project.org/。Christian P Robert。截断正态变量的模拟。Statistics and Computing，5(2): 121–125，1995。Peter J. Rousseeuw和Annick M. Leroy。《稳健回归与离群值检测》。Wiley，1987。Patrick Royston。缺失值的多重插补。《Stata杂志》，4(3):227–241，2004。DB Rubin和J Wiley。纽约、奇切斯特、布里斯班、多伦多、新加坡。1987年《调查无应答的多重插补》。Donald B Rubin。多重插补之后 18+ 年。《美国统计学会杂志》，91(434):473–489，1996。

Hasan Ahmed Salman、Ali Kalakech和Amani Steiti。随机森林算法概述。《Babylonian Journal of Machine Learning》，2024:69–79，2024。Joseph L Schafer。多重插补：入门。《医学研究中的统计方法》，8(1): 3–15，1999。Joseph L Schafer和John W Graham。缺失数据：我们对现状的看法。《Psycho- logical methods》，7(2):147，2002。Anoop D Shah、Jonathan W Bartlett、James Carpenter、Owen Nicholas和Harry Hemingway。使用mice插补缺失数据时随机森林与参数插补模型的比较：CALIBER研究。《American Journal of Epidemiology》，179(6):764–774，2014。Juned Siddique和Thomas R Belin。采用基于距离的供体选择的迭代热卡插补。《Statistics in Medicine》，27(1):83–102，2008。Dan Steinberg。分类回归树。载于《数据挖掘十大算法》，第193–216页。Chapman and Hall/CRC，2009。Jonathan Taylor。统计学习与选择性推断。《美国国家科学院院刊》，2015。选择性推断概述；PNAS作者介绍页。Matthias Templ。采用广义加性模型进行稳健多重插补。Statistics and Computing，34(4):121，2024。Robert Tibshirani。通过Lasso进行回归收缩与变量选择。《Journal of the Royal Statistical Society Series B: Statistical Methodology》，58(1):267–288，1996。Stef Van Buuren。《缺失数据的灵活插补》。CRC出版社，2018。Stef van Buuren和Karin Groothuis-Oudshoorn。mice：R语言中的链式方程多变量插补。《Journal of Statistical Software》，45(3):1–67，2011。doi:10.18637/jss.v045.i03。Stef Van Buuren和Karin Groothuis-Oudshoorn。mice：R语言中的链式方程多变量插补。《Journal of Statistical Software》，45:1–67，2011。Stef Van Buuren、Jaap PL Brand、Catharina GM Groothuis-Oudshoorn和Donald B Rubin。多变量插补中的全条件设定。《Journal of Statistical Computation and Simulation》，76(12):1049–1064，2006。Geert JMG Van der Heijden、A Rogier T Donders、Theo Stijnen和Karel GM Moons。在多变量诊断研究中，插补缺失值优于完整病例分析和缺失指示变量法：临床示例。《Journal of Clinical Epidemiology》，59 (10):1102–1109，2006。Sudhir Varma和Richard Simon。使用交叉验证进行模型选择时的误差估计偏差。《BMC Bioinformatics》，7:91，2006。doi:10.1186/1471-2105-7-91。Luís Gustavo do Amaral Vinha。《教育评估中的纵向研究与缺失数据处理》。博士论文，巴西利亚大学，心理学研究所，巴西利亚，2016。葡萄牙语。Paul T von Hippel。是否应修改正态插补模型以插补偏态变量？Sociological Methods & Research，42(1):105–138，2013。

Richard Von Mises。n 个观测值中最大值的分布。《数学评论：跨巴尔干联盟》，1:141–160，1936。

Han Wang、Bo Li 和 Guohua Jiang。通过 LAD-Lasso 实现稳健回归收缩和一致的变量选择。《商业与经济统计学杂志》，25(3):347–355，2007。

Xin Wang、Yang Jiang 等。采用指数平方损失的稳健变量选择。《美国统计学会杂志》，108(502):632–643，2013。

Ian R White、Patrick Royston 和 Angela M Wood。使用链式方程进行多重插补：实践中的问题与指导。*Statistics in Medicine*，30(4):377–399，2011。

Chien-Fu Jeff Wu。回归分析中的刀切法、自助法及其他重抽样方法。《统计年刊》，14(4):1261–1295，1986。

Chong Yu和Weixin Yao。稳健线性回归：综述与比较。《统计通讯——模拟与计算》，46(8):6261–6282，2017年。

Yize Zhao和Qi Long。高维数据情形下的多重插补。《医学研究中的统计方法》，25(5):2021–2035，2016。

Hui Zou和Trevor Hastie。基于弹性网络的正则化与变量选择。*Journalof the Royal Statistical Society Series B: Statistical Methodology*，67(2):301–320，2005年。

致谢：
感谢 Eduardo Yoshio Nakano 博士就数据插补、建模以及蒙特卡罗模拟的设计提供了富有见地的讨论和建议。我也感谢 Cira Etheowalda Guevara Otiniano 博士（巴西利亚大学统计系）就极端事件理论和概率进行讨论，并感谢 Fidel Ernesto Castro Morales 博士（北里奥格兰德联邦大学统计系）就极端事件及相关计算方面进行了有益的讨论。

A 概览与表格指南

本附录旨在提高透明度，并更详细地展示结果部分报告的结果，重点关注所考虑的不同样本量和情景。表格安排如下：

1. 小样本量 2. 中等样本量 3.
相对较大的样本量

每张表都报告：情景描述（设置）；所使用的多重插补方法（方法）；以及每个拟合回归模型（线性或稀疏）的描述性统计（均值、中位数、分位数和方差）。有意查看具体数值并进行更精确的跨情景比较的读者，可参阅后续页面中的相关表格。

表10： 普通最小二乘法（干净）和弹性网络（弹性网络，受污染）在不同多重插补方法下的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差，其中 $\boldsymbol{n = 20}$, $n_{\text{sim}} = 50$，迭代次数为 $\text{iter} = 5$, $\rho = 0.6$，污染程度为 $P_{\text{ext}} \in \{0.03, 0.05\}$，缺失率为 $P_{\text{miss}} \in \{0.05, 0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	-0.287	-0.014	0.074	0.367	0.024	0.083	1.000	1.000	1.000	2.520	0.344	1.680	2.397	4.182
	T2	-0.176	-0.007	0.039	0.323	0.024	0.059	1.000	1.000	1.000	2.581	0.323	1.728	2.485	4.070
	T3	-0.199	-0.008	0.046	0.315	0.024	0.064	1.000	1.000	1.000	2.565	0.322	1.835	2.423	4.134
	T4	-0.154	-0.004	0.031	0.319	0.026	0.061	1.000	1.000	1.000	2.624	0.358	1.780	2.487	4.206
	T5	-0.152	-0.002	0.027	0.290	0.030	0.061	1.000	1.000	1.000	2.741	0.736	1.936	2.555	4.629
	T6	-0.241	-0.006	0.050	0.349	0.033	0.061	1.000	1.000	1.000	2.576	0.339	1.743	2.483	4.226
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	1.457	-0.288	0.242	1.471	0.290	0.247	1.000	1.000	1.000	1.995	0.092	1.518	1.943	2.709
	T2	1.452	-0.272	0.210	1.472	0.274	0.215	1.000	1.000	1.000	2.111	0.092	1.600	2.060	2.765
	T3	1.452	-0.276	0.219	1.471	0.278	0.224	1.000	1.000	1.000	2.077	0.106	1.551	2.023	2.900
	T4	1.347	-0.264	0.218	1.443	0.273	0.225	1.000	1.000	1.000	2.170	0.175	1.575	2.059	3.314
	T5	1.382	-0.238	0.155	1.491	0.251	0.170	1.000	1.000	1.000	2.575	0.787	1.841	2.241	4.921
	T6	1.366	-0.264	0.212	1.453	0.271	0.219	1.000	1.000	1.000	2.147	0.150	1.595	2.063	3.081
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.411	-0.025	0.117	0.533	0.067	0.175	1.000	1.000	0.980	2.283	0.243	1.521	2.221	3.317
	T2	-0.118	-0.017	0.046	0.417	0.063	0.138	1.000	1.000	1.000	2.441	0.203	1.660	2.407	3.290
	T3	-0.185	-0.019	0.064	0.436	0.070	0.147	1.000	1.000	1.000	2.399	0.226	1.623	2.372	3.461
	T4	-0.198	-0.002	0.032	0.427	0.058	0.132	1.000	1.000	1.000	2.487	0.333	1.627	2.450	4.147
	T5	-0.133	-0.007	0.033	0.336	0.057	0.126	1.000	1.000	1.000	2.611	0.260	1.945	2.569	3.874
	T6	-0.280	-0.000	0.044	0.434	0.062	0.129	1.000	1.000	1.000	2.432	0.226	1.639	2.399	3.436
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	1.650	-0.314	0.267	1.778	0.323	0.274	0.900	0.880	0.980	1.866	0.095	1.349	1.849	2.415
	T2	1.635	-0.287	0.213	1.776	0.297	0.223	0.960	0.980	1.000	2.118	0.162	1.487	2.058	2.964
	T3	1.662	-0.298	0.230	1.815	0.309	0.241	0.940	0.980	0.980	2.052	0.142	1.447	1.969	2.872
	T4	1.189	-0.238	0.201	1.582	0.275	0.221	1.000	1.000	1.000	2.553	1.002	1.623	2.274	4.937
	T5	1.215	-0.199	0.107	1.563	0.233	0.143	1.000	1.000	1.000	3.103	1.640	1.825	2.740	6.595
	T6	1.240	-0.246	0.207	1.573	0.272	0.218	0.980	0.940	1.000	2.380	0.594	1.510	2.172	4.495
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.640	-0.112	0.331	1.063	0.214	0.470	0.980	0.920	0.800	1.980	0.459	0.837	1.956	3.319
	T2	0.277	-0.084	0.098	1.003	0.187	0.358	1.000	0.960	0.920	2.526	0.389	1.349	2.536	3.600
	T3	0.113	-0.090	0.141	0.951	0.199	0.353	1.000	0.980	0.920	2.410	0.415	1.192	2.390	3.341
	T4	-0.317	-0.012	0.089	0.731	0.144	0.322	1.000	1.000	0.960	2.669	0.335	1.539	2.737	3.572
	T5	0.058	-0.015	0.021	0.783	0.142	0.273	1.000	1.000	1.000	3.185	0.397	2.259	3.152	4.651
	T6	-0.235	-0.026	0.094	0.663	0.151	0.351	1.000	0.980	0.940	2.561	0.479	1.273	2.610	3.761
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	1.748	-0.396	0.423	2.124	0.429	0.459	0.740	0.560	0.660	1.594	0.165	1.046	1.608	2.656
	T2	1.577	-0.291	0.236	1.921	0.318	0.274	0.960	0.980	1.000	2.403	0.361	1.517	2.361	3.984
	T3	1.809	-0.323	0.255	2.122	0.348	0.290	0.940	1.000	1.000	2.228	0.299	1.403	2.163	3.598
	T4	0.271	-0.106	0.123	1.665	0.222	0.187	0.980	0.980	1.000	3.688	1.703	1.980	3.345	7.045
	T5	0.507	-0.015	-0.132	1.416	0.145	0.236	1.000	1.000	1.000	5.065	2.908	3.066	4.735	9.210
	T6	0.675	-0.177	0.191	1.707	0.256	0.248	0.980	0.980	0.980	3.126	1.581	1.641	2.752	5.728
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	-0.287	-0.014	0.074	0.367	0.024	0.083	1.000	1.000	1.000	2.520	0.344	1.680	2.397	4.182
	T2	-0.176	-0.007	0.039	0.323	0.024	0.059	1.000	1.000	1.000	2.581	0.323	1.728	2.485	4.070
	T3	-0.199	-0.008	0.046	0.315	0.024	0.064	1.000	1.000	1.000	2.565	0.322	1.835	2.423	4.134
	T4	-0.154	-0.004	0.031	0.319	0.026	0.061	1.000	1.000	1.000	2.624	0.358	1.780	2.487	4.206
	T5	-0.152	-0.002	0.027	0.290	0.030	0.061	1.000	1.000	1.000	2.741	0.736	1.936	2.555	4.629
	T6	-0.241	-0.006	0.050	0.349	0.033	0.061	1.000	1.000	1.000	2.576	0.339	1.743	2.483	4.226
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	1.457	-0.288	0.242	1.471	0.290	0.247	1.000	1.000	1.000	1.995	0.092	1.518	1.943	2.709
	T2	1.452	-0.272	0.210	1.472	0.274	0.215	1.000	1.000	1.000	2.111	0.092	1.600	2.060	2.765
	T3	1.452	-0.276	0.219	1.471	0.278	0.224	1.000	1.000	1.000	2.077	0.106	1.551	2.023	2.900
	T4	1.347	-0.264	0.218	1.443	0.273	0.225	1.000	1.000	1.000	2.170	0.175	1.575	2.059	3.314
	T5	1.382	-0.238	0.155	1.491	0.251	0.170	1.000	1.000	1.000	2.575	0.787	1.841	2.241	4.921
	T6	1.366	-0.264	0.212	1.453	0.271	0.219	1.000	1.000	1.000	2.147	0.150	1.595	2.063	3.081

表11： 在 OLS（干净）和 EN（受污染）情况下，各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差，设置为 $\boldsymbol{n = 20}$、$n_{\text{sim}} = 50$，迭代次数 $= 5$、$\rho = 0.6$，污染程度 $P_{\text{ext}} \in \{0.05、0.10\}$，缺失率 $P_{\text{miss}} \in \{0.05、0.10、0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.411	-0.025	0.117	0.533	0.067	0.175	1.000	1.000	0.980	2.283	0.243	1.521	2.221	3.317
	T2	-0.118	-0.017	0.046	0.417	0.063	0.138	1.000	1.000	1.000	2.441	0.203	1.660	2.407	3.290
	T3	-0.185	-0.019	0.064	0.436	0.070	0.147	1.000	1.000	1.000	2.399	0.226	1.623	2.372	3.461
	T4	-0.198	-0.002	0.032	0.427	0.058	0.132	1.000	1.000	1.000	2.487	0.333	1.627	2.450	4.147
	T5	-0.133	-0.007	0.033	0.336	0.057	0.126	1.000	1.000	1.000	2.611	0.260	1.945	2.569	3.874
	T6	-0.280	-0.000	0.044	0.434	0.062	0.129	1.000	1.000	1.000	2.432	0.226	1.639	2.399	3.436
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	1.650	-0.314	0.267	1.778	0.323	0.274	0.900	0.880	0.980	1.866	0.095	1.349	1.849	2.415
	T2	1.635	-0.287	0.213	1.776	0.297	0.223	0.960	0.980	1.000	2.118	0.162	1.487	2.058	2.964
	T3	1.662	-0.298	0.230	1.815	0.309	0.241	0.940	0.980	0.980	2.052	0.142	1.447	1.969	2.872
	T4	1.189	-0.238	0.201	1.582	0.275	0.221	1.000	1.000	1.000	2.553	1.002	1.623	2.274	4.937
	T5	1.215	-0.199	0.107	1.563	0.233	0.143	1.000	1.000	1.000	3.103	1.640	1.825	2.740	6.595
	T6	1.240	-0.246	0.207	1.573	0.272	0.218	0.980	0.940	1.000	2.380	0.594	1.510	2.172	4.495
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.640	-0.112	0.331	1.063	0.214	0.470	0.980	0.920	0.800	1.980	0.459	0.837	1.956	3.319
	T2	0.277	-0.084	0.098	1.003	0.187	0.358	1.000	0.960	0.920	2.526	0.389	1.349	2.536	3.600
	T3	0.113	-0.090	0.141	0.951	0.199	0.353	1.000	0.980	0.920	2.410	0.415	1.192	2.390	3.341
	T4	-0.317	-0.012	0.089	0.731	0.144	0.322	1.000	1.000	0.960	2.669	0.335	1.539	2.737	3.572
	T5	0.058	-0.015	0.021	0.783	0.142	0.273	1.000	1.000	1.000	3.185	0.397	2.259	3.152	4.651
	T6	-0.235	-0.026	0.094	0.663	0.151	0.351	1.000	0.980	0.940	2.561	0.479	1.273	2.610	3.761
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	1.748	-0.396	0.423	2.124	0.429	0.459	0.740	0.560	0.660	1.594	0.165	1.046	1.608	2.656
	T2	1.577	-0.291	0.236	1.921	0.318	0.274	0.960	0.980	1.000	2.403	0.361	1.517	2.361	3.984
	T3	1.809	-0.323	0.255	2.122	0.348	0.290	0.940	1.000	1.000	2.228	0.299	1.403	2.163	3.598
	T4	0.271	-0.106	0.123	1.665	0.222	0.187	0.980	0.980	1.000	3.688	1.703	1.980	3.345	7.045
	T5	0.507	-0.015	-0.132	1.416	0.145	0.236	1.000	1.000	1.000	5.065	2.908	3.066	4.735	9.210
	T6	0.675	-0.177	0.191	1.707	0.256	0.248	0.980	0.980	0.980	3.126	1.581	1.641	2.752	5.728
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	-0.292	-0.014	0.074	0.369	0.023	0.083	1.000	1.000	1.000	2.480	0.236	1.847	2.450	3.925
	T2	-0.149	-0.009	0.038	0.326	0.027	0.061	1.000	1.000	1.000	2.565	0.239	1.992	2.479	4.104
	T3	-0.197	-0.009	0.048	0.331	0.025	0.064	1.000	1.000	1.000	2.552	0.239	1.943	2.490	4.119
	T4	-0.156	-0.010	0.041	0.350	0.033	0.056	1.000	1.000	1.000	2.573	0.336	1.789	2.527	4.297
	T5	-0.144	-0.006	0.032	0.294	0.028	0.057	1.000	1.000	1.000	2.644	0.353	1.957	2.553	4.651
	T6	-0.206	-0.002	0.034	0.348	0.034	0.050	1.000	1.000	1.000	2.561	0.318	1.818	2.464	4.245
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：外部。	T1	0.939	-0.241	0.220	0.949	0.242	0.223	1.000	1.000	1.000	1.544	0.047	1.256	1.498	2.065
	T2	0.913	-0.227	0.197	0.935	0.229	0.201	1.000	1.000	1.000	1.636	0.052	1.305	1.625	2.108
	T3	0.912	-0.229	0.200	0.932	0.231	0.204	1.000	1.000	1.000	1.625	0.062	1.309	1.553	2.185
	T4	0.594	-0.189	0.178	1.051	0.218	0.194	1.000	1.000	1.000	1.906	0.451	1.290	1.675	3.659
	T5	0.561	-0.164	0.128	1.060	0.196	0.145	1.000	1.000	1.000	2.275	0.765	1.473	1.935	4.607
	T6	0.644	-0.196	0.184	1.053	0.221	0.196	1.000	1.000	1.000	1.848	0.388	1.260	1.618	3.611
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.420	-0.025	0.120	0.539	0.067	0.177	1.000	1.000	0.980	2.331	0.265	1.604	2.206	3.439
	T2	-0.121	-0.018	0.047	0.520	0.067	0.135	1.000	1.000	1.000	2.503	0.207	1.793	2.492	3.401
	T3	-0.190	-0.016	0.059	0.391	0.069	0.149	1.000	1.000	1.000	2.460	0.196	1.819	2.395	3.313
	T4	-0.234	-0.007	0.052	0.419	0.058	0.143	1.000	1.000	1.000	2.535	0.260	1.763	2.457	3.533
	T5	-0.023	-0.011	0.019	0.346	0.063	0.141	1.000	1.000	1.000	2.776	0.173	2.179	2.766	3.638
	T6	-0.224	-0.005	0.041	0.401	0.068	0.139	1.000	1.000	1.000	2.550	0.297	1.656	2.490	4.018
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	1.081	-0.262	0.242	1.156	0.269	0.251	1.000	0.860	1.000	1.457	0.047	1.083	1.452	1.866
	T2	0.979	-0.233	0.202	1.059	0.239	0.209	1.000	1.000	1.000	1.657	0.062	1.283	1.615	2.061
	T3	1.037	-0.243	0.210	1.145	0.252	0.221	1.000	0.960	1.000	1.604	0.069	1.139	1.589	2.144
	T4	0.474	-0.168	0.165	1.167	0.218	0.194	1.000	0.980	1.000	2.200	0.782	1.346	1.855	4.294
	T5	0.415	-0.123	0.076	1.254	0.198	0.155	1.000	1.000	1.000	2.902	1.721	1.471	2.370	5.749
	T6	0.518	-0.173	0.165	1.155	0.220	0.195	1.000	0.960	1.000	2.175	0.877	1.218	1.817	4.397

表 12： 普通最小二乘法（干净）和弹性网络（受污染）下，混合 $n = 20$（区块 $P_{\text{ext}} = 0.10$、$P_{\text{miss}} = 0.30$）与 $n = 40$（区块 $P_{\text{ext}} = 0.03$、$P_{\text{miss}} \in \{0.05、0.10、0.30\}$）时，各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差；$n_{\text{sim}} = 50$，迭代次数 $= 5$、$\rho = 0.6$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.617	-0.113	0.329	1.031	0.213	0.469	0.980	0.920	0.800	1.912	0.252	0.917	2.023	2.649
	T2	0.294	-0.085	0.096	1.063	0.204	0.351	0.980	0.960	0.960	2.512	0.374	1.357	2.666	3.324
	T3	0.039	-0.077	0.133	0.861	0.195	0.362	1.000	0.960	0.920	2.296	0.265	1.099	2.399	3.108
	T4	-0.245	-0.013	0.077	0.767	0.139	0.306	1.000	1.000	0.980	2.692	0.321	1.595	2.698	3.734
	T5	0.063	-0.020	0.028	0.674	0.143	0.298	1.000	1.000	1.000	3.015	0.204	2.222	3.070	3.837
	T6	-0.311	-0.026	0.114	0.713	0.142	0.334	1.000	1.000	0.940	2.506	0.301	1.410	2.554	3.362
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	1.294	-0.338	0.361	1.519	0.362	0.394	0.780	0.540	0.660	1.202	0.072	0.696	1.210	1.692
	T2	1.117	-0.265	0.241	1.350	0.289	0.277	1.000	0.960	0.980	1.803	0.202	1.175	1.706	2.799
	T3	1.178	-0.267	0.233	1.428	0.296	0.278	0.960	0.960	0.960	1.723	0.212	1.026	1.678	2.730
	T4	-0.493	-0.010	0.029	1.440	0.180	0.154	1.000	1.000	1.000	3.629	1.619	1.834	3.524	6.175
	T5	-0.320	0.085	-0.217	1.311	0.175	0.276	1.000	1.000	1.000	4.919	2.066	2.735	4.882	7.577
	T6	-0.167	-0.070	0.094	1.419	0.210	0.199	1.000	0.980	1.000	3.189	2.461	1.333	2.703	6.326
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	2.021	-0.147	-0.189	2.028	0.149	0.191	0.900	1.000	1.000	2.205	0.020	1.969	2.192	2.455
	T2	2.051	-0.134	-0.221	2.059	0.136	0.223	0.920	1.000	1.000	2.263	0.019	2.027	2.280	2.482
	T3	2.058	-0.135	-0.221	2.066	0.137	0.223	0.960	1.000	1.000	2.259	0.018	2.076	2.228	2.548
	T4	2.045	-0.134	-0.219	2.050	0.136	0.221	0.960	1.000	1.000	2.280	0.023	2.020	2.266	2.500
	T5	2.104	-0.131	-0.239	2.116	0.133	0.242	0.960	1.000	1.000	2.351	0.053	2.094	2.320	2.960
	T6	2.055	-0.136	-0.219	2.062	0.138	0.222	0.960	1.000	1.000	2.277	0.024	2.070	2.278	2.557
稀疏回归（弹性网络）															
样本量: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	2.733	-0.117	-0.386	2.738	0.121	0.388	0.140	1.000	1.000	2.867	0.045	2.541	2.867	3.271
	T2	2.772	-0.100	-0.427	2.777	0.105	0.429	0.120	1.000	0.980	2.923	0.048	2.564	2.907	3.355
	T3	2.790	-0.104	-0.424	2.795	0.108	0.426	0.060	1.000	1.000	2.911	0.050	2.592	2.939	3.404
	T4	2.770	-0.102	-0.422	2.776	0.107	0.424	0.080	1.000	0.980	2.925	0.057	2.533	2.944	3.359
	T5	2.783	-0.097	-0.436	2.791	0.104	0.439	0.120	1.000	0.960	2.962	0.059	2.576	2.978	3.440
	T6	2.770	-0.103	-0.419	2.777	0.110	0.423	0.100	1.000	0.980	2.913	0.068	2.559	2.910	3.406
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	1.948	-0.158	-0.150	1.972	0.162	0.157	0.820	1.000	1.000	2.117	0.050	1.613	2.139	2.496
	T2	2.038	-0.131	-0.221	2.063	0.137	0.228	0.880	1.000	1.000	2.259	0.054	1.822	2.274	2.646
	T3	2.025	-0.136	-0.209	2.052	0.140	0.216	0.840	1.000	1.000	2.242	0.053	1.784	2.244	2.639
	T4	1.906	-0.113	-0.231	1.951	0.127	0.238	0.940	1.000	1.000	2.345	0.115	1.728	2.316	3.146
	T5	1.973	-0.102	-0.266	2.019	0.121	0.276	0.940	1.000	1.000	2.496	0.144	1.947	2.442	3.538
	T6	1.917	-0.116	-0.226	1.966	0.131	0.234	0.880	1.000	1.000	2.337	0.102	1.813	2.349	3.128
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	2.625	-0.124	-0.354	2.640	0.131	0.359	0.340	0.980	1.000	2.788	0.083	2.195	2.844	3.172
	T2	2.752	-0.097	-0.433	2.769	0.106	0.438	0.320	0.980	0.880	2.897	0.082	2.343	2.964	3.296
	T3	2.765	-0.099	-0.431	2.778	0.109	0.437	0.340	1.000	0.900	2.892	0.097	2.178	2.948	3.368
	T4	2.711	-0.093	-0.434	2.729	0.105	0.439	0.320	1.000	0.920	2.916	0.087	2.351	2.941	3.380
	T5	2.711	-0.085	-0.447	2.733	0.102	0.452	0.320	1.000	0.880	2.984	0.098	2.444	3.001	3.598
	T6	2.704	-0.093	-0.430	2.722	0.109	0.437	0.300	1.000	0.900	2.914	0.102	2.222	2.943	3.475
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	1.869	-0.227	0.003	1.941	0.237	0.079	0.640	0.640	1.000	1.791	0.121	1.135	1.780	2.478
	T2	2.178	-0.141	-0.226	2.229	0.156	0.244	0.820	1.000	1.000	2.256	0.123	1.544	2.275	2.860
	T3	2.132	-0.144	-0.213	2.194	0.160	0.234	0.760	0.960	1.000	2.236	0.158	1.423	2.230	2.945
	T4	1.873	-0.096	-0.253	1.963	0.134	0.279	0.840	0.980	1.000	2.516	0.308	1.602	2.480	3.770
	T5	2.185	-0.075	-0.358	2.256	0.116	0.378	0.820	1.000	1.000	2.890	0.286	1.922	2.878	4.080
	T6	1.863	-0.105	-0.232	1.946	0.139	0.257	0.820	1.000	1.000	2.434	0.340	1.539	2.364	3.993
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	2.209	-0.178	-0.160	2.328	0.199	0.199	0.640	0.880	1.000	2.529	0.138	1.819	2.558	3.201
	T2	2.691	-0.082	-0.448	2.772	0.116	0.465	0.640	1.000	0.760	2.948	0.134	2.165	2.969	3.624
	T3	2.693	-0.090	-0.433	2.749	0.124	0.457	0.580	0.960	0.860	2.896	0.137	2.299	2.878	3.743
	T4	2.450	-0.056	-0.447	2.542	0.098	0.460	0.600	1.000	0.780	3.042	0.108	2.345	3.014	3.655
	T5	2.636	-0.048	-0.503	2.709	0.090	0.515	0.680	1.000	0.740	3.199	0.123	2.643	3.136	4.065
	T6	2.562	-0.065	-0.450	2.642	0.112	0.468	0.680	0.980	0.780	3.015	0.144	2.290	2.986	3.699



表13： 普通最小二乘法和弹性网络在干净（普通最小二乘法）和受污染（弹性网络）条件下，各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差，其中 $\boldsymbol{n = 40}$, $n_{\text{sim}} = 50$，迭代次数 $= 5$, $\rho = 0.6$，污染 $P_{\text{ext}} \in \{0.05, 0.10\}$，缺失率 $P_{\text{miss}} \in \{0.05, 0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	2.021	-0.148	-0.188	2.027	0.149	0.190	0.860	1.000	1.000	2.207	0.023	1.914	2.241	2.442
	T2	2.051	-0.134	-0.220	2.058	0.136	0.222	0.960	1.000	1.000	2.271	0.027	1.967	2.280	2.527
	T3	2.048	-0.134	-0.221	2.054	0.135	0.223	0.980	1.000	1.000	2.275	0.026	2.016	2.272	2.566
	T4	2.046	-0.135	-0.218	2.052	0.136	0.220	0.960	1.000	1.000	2.273	0.030	1.952	2.260	2.564
	T5	2.078	-0.129	-0.236	2.084	0.132	0.240	0.980	1.000	1.000	2.352	0.053	2.069	2.349	2.780
	T6	2.052	-0.136	-0.218	2.057	0.137	0.221	0.980	1.000	1.000	2.273	0.030	1.930	2.262	2.561
稀疏回归（弹性网络）															
样本量: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	2.574	-0.103	-0.373	2.579	0.108	0.375	0.000	1.000	1.000	2.712	0.043	2.357	2.716	3.107
	T2	2.600	-0.086	-0.412	2.604	0.092	0.414	0.020	1.000	0.960	2.769	0.043	2.385	2.757	3.207
	T3	2.617	-0.090	-0.409	2.621	0.095	0.411	0.000	1.000	0.980	2.757	0.045	2.408	2.769	3.189
	T4	2.559	-0.083	-0.411	2.569	0.094	0.415	0.140	1.000	0.920	2.796	0.063	2.342	2.809	3.255
	T5	2.566	-0.072	-0.433	2.577	0.088	0.438	0.160	1.000	0.860	2.858	0.076	2.469	2.776	3.388
	T6	2.556	-0.082	-0.411	2.566	0.095	0.415	0.160	1.000	0.940	2.782	0.060	2.366	2.785	3.255
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据: 干净	T1	1.941	-0.159	-0.149	1.963	0.162	0.156	0.780	1.000	1.000	2.109	0.050	1.635	2.130	2.524
	T2	2.054	-0.134	-0.220	2.076	0.138	0.225	0.780	1.000	1.000	2.246	0.057	1.702	2.265	2.710
	T3	2.030	-0.136	-0.210	2.051	0.140	0.216	0.800	1.000	1.000	2.237	0.061	1.685	2.252	2.699
	T4	1.921	-0.114	-0.231	1.960	0.126	0.239	0.940	1.000	1.000	2.357	0.125	1.776	2.325	3.294
	T5	1.963	-0.107	-0.255	2.014	0.124	0.263	0.900	1.000	1.000	2.457	0.147	1.813	2.391	3.410
	T6	1.904	-0.115	-0.225	1.949	0.131	0.233	0.940	1.000	1.000	2.331	0.112	1.781	2.288	2.944
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	2.512	-0.113	-0.344	2.523	0.120	0.349	0.020	0.960	1.000	2.628	0.085	1.951	2.680	3.011
	T2	2.616	-0.085	-0.420	2.625	0.095	0.425	0.100	0.980	0.820	2.750	0.096	1.991	2.767	3.178
	T3	2.589	-0.083	-0.420	2.598	0.093	0.425	0.100	1.000	0.860	2.742	0.094	2.033	2.759	3.209
	T4	2.508	-0.071	-0.427	2.529	0.089	0.432	0.280	1.000	0.920	2.805	0.105	2.192	2.811	3.323
	T5	2.548	-0.061	-0.454	2.567	0.086	0.460	0.220	1.000	0.780	2.879	0.112	2.253	2.830	3.581
	T6	2.514	-0.070	-0.429	2.534	0.094	0.437	0.220	1.000	0.820	2.830	0.128	2.170	2.736	3.565
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	1.858	-0.229	0.008	1.927	0.238	0.075	0.660	0.640	1.000	1.814	0.121	1.129	1.858	2.417
	T2	2.181	-0.141	-0.228	2.235	0.154	0.245	0.740	1.000	1.000	2.292	0.154	1.476	2.314	2.838
	T3	2.187	-0.152	-0.208	2.236	0.164	0.225	0.760	0.960	1.000	2.244	0.188	1.379	2.297	3.047
	T4	1.874	-0.092	-0.264	1.955	0.128	0.290	0.940	0.980	1.000	2.572	0.307	1.738	2.570	3.682
	T5	2.219	-0.076	-0.365	2.275	0.110	0.383	0.900	1.000	1.000	2.931	0.271	1.858	3.038	3.912
	T6	1.903	-0.110	-0.231	1.978	0.138	0.255	0.860	0.980	1.000	2.464	0.259	1.456	2.435	3.575
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	2.232	-0.175	-0.157	2.308	0.195	0.193	0.400	0.820	1.000	2.365	0.136	1.674	2.382	3.009
	T2	2.522	-0.068	-0.431	2.567	0.102	0.447	0.500	1.000	0.800	2.771	0.151	2.023	2.845	3.559
	T3	2.539	-0.079	-0.415	2.597	0.114	0.431	0.460	0.980	0.820	2.754	0.140	2.017	2.776	3.534
	T4	2.203	-0.026	-0.450	2.268	0.079	0.461	0.600	1.000	0.740	2.961	0.121	2.334	3.009	3.616
	T5	2.383	-0.010	-0.520	2.446	0.084	0.534	0.580	1.000	0.600	3.181	0.200	2.502	3.200	4.176
	T6	2.298	-0.035	-0.451	2.359	0.092	0.471	0.540	1.000	0.780	2.909	0.171	2.163	2.953	3.636
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	2.015	-0.146	-0.190	2.022	0.148	0.192	0.900	1.000	1.000	2.195	0.027	1.912	2.201	2.528
	T2	2.076	-0.134	-0.226	2.084	0.136	0.228	0.880	1.000	1.000	2.270	0.029	1.940	2.285	2.554
	T3	2.062	-0.134	-0.223	2.069	0.137	0.225	0.980	1.000	1.000	2.267	0.028	2.009	2.263	2.564
	T4	2.024	-0.128	-0.227	2.035	0.132	0.230	0.940	1.000	1.000	2.294	0.041	1.925	2.294	2.797
	T5	2.092	-0.128	-0.242	2.104	0.131	0.245	0.980	1.000	1.000	2.364	0.059	2.026	2.312	2.860
	T6	2.041	-0.133	-0.221	2.054	0.137	0.224	0.940	1.000	1.000	2.278	0.042	1.918	2.253	2.661
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	2.282	-0.108	-0.304	2.285	0.113	0.307	0.000	1.000	1.000	2.657	0.032	2.223	2.653	2.957
	T2	2.316	-0.091	-0.344	2.320	0.098	0.347	0.020	1.000	1.000	2.718	0.033	2.295	2.716	2.988
	T3	2.294	-0.089	-0.345	2.298	0.097	0.349	0.040	1.000	1.000	2.724	0.035	2.406	2.716	3.030
	T4	2.224	-0.077	-0.354	2.234	0.092	0.360	0.240	1.000	1.000	2.771	0.058	2.264	2.750	3.181
	T5	2.267	-0.072	-0.372	2.278	0.089	0.378	0.140	1.000	1.000	2.822	0.073	2.327	2.822	3.352
	T6	2.283	-0.087	-0.345	2.290	0.100	0.351	0.060	1.000	0.980	2.735	0.056	2.297	2.722	3.110

表14： 普通最小二乘法（干净）和弹性网络（受污染）中各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。第一组： $\boldsymbol{n = 40}$, $n_{\text{sim}} = 50$, 迭代次数 $= 5$, $\rho = 0.6$,$P_{\text{ext}} = 0.10$, $P_{\text{miss}} \in \{0.10,0.30\}$。第二组： $\boldsymbol{n = 20}$, $n_{\text{sim}} = 300$, 迭代次数 $= 5$, $\rho = 0.6$, $P_{\text{ext}} \in \{0.04\}$,$P_{\text{miss}} \in \{0.10,0.25\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	1.940	-0.158	-0.150	1.964	0.162	0.157	0.820	1.000	1.000	2.106	0.044	1.635	2.138	2.444
	T2	2.062	-0.134	-0.221	2.089	0.140	0.227	0.760	1.000	1.000	2.255	0.057	1.702	2.265	2.737
	T3	2.030	-0.135	-0.212	2.053	0.140	0.218	0.800	1.000	1.000	2.240	0.055	1.685	2.278	2.613
	T4	1.910	-0.114	-0.227	1.950	0.127	0.234	0.920	1.000	1.000	2.339	0.101	1.773	2.322	3.060
	T5	1.972	-0.105	-0.258	2.025	0.122	0.266	0.940	1.000	1.000	2.464	0.139	1.831	2.415	3.429
	T6	1.886	-0.113	-0.226	1.931	0.129	0.234	0.940	1.000	1.000	2.332	0.103	1.781	2.296	2.944
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	2.235	-0.119	-0.276	2.243	0.126	0.283	0.020	0.960	1.000	2.572	0.077	1.865	2.631	2.964
	T2	2.312	-0.088	-0.352	2.319	0.098	0.359	0.040	0.980	1.000	2.687	0.084	1.920	2.745	3.090
	T3	2.262	-0.082	-0.354	2.271	0.095	0.362	0.120	0.980	1.000	2.704	0.091	1.943	2.715	3.122
	T4	2.197	-0.069	-0.368	2.212	0.090	0.377	0.180	1.000	1.000	2.777	0.114	2.073	2.801	3.429
	T5	2.251	-0.064	-0.386	2.263	0.079	0.392	0.180	1.000	0.980	2.823	0.072	2.317	2.836	3.288
	T6	2.260	-0.078	-0.361	2.268	0.092	0.369	0.140	1.000	0.980	2.749	0.087	2.070	2.760	3.258
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	1.888	-0.232	0.008	1.949	0.241	0.080	0.640	0.640	1.000	1.819	0.114	1.149	1.830	2.415
	T2	2.199	-0.144	-0.227	2.256	0.159	0.245	0.700	1.000	1.000	2.301	0.150	1.547	2.314	2.956
	T3	2.213	-0.154	-0.209	2.251	0.166	0.226	0.780	0.960	1.000	2.261	0.176	1.389	2.256	3.064
	T4	1.885	-0.095	-0.258	1.964	0.134	0.288	0.920	0.960	1.000	2.565	0.307	1.738	2.531	3.704
	T5	2.205	-0.074	-0.364	2.268	0.112	0.382	0.900	1.000	1.000	2.939	0.248	1.983	3.020	3.912
	T6	1.899	-0.112	-0.227	1.974	0.144	0.255	0.840	0.960	1.000	2.470	0.251	1.611	2.414	3.589
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	2.037	-0.190	-0.090	2.095	0.210	0.154	0.260	0.760	1.000	2.307	0.148	1.585	2.336	3.044
	T2	2.223	-0.068	-0.370	2.260	0.107	0.392	0.440	0.980	0.880	2.734	0.163	2.046	2.799	3.484
	T3	2.235	-0.081	-0.349	2.277	0.118	0.372	0.400	0.980	0.960	2.710	0.146	1.979	2.704	3.470
	T4	1.833	-0.009	-0.411	1.903	0.094	0.428	0.720	1.000	0.800	2.964	0.136	2.296	3.012	3.672
	T5	2.043	0.006	-0.483	2.121	0.089	0.497	0.540	1.000	0.840	3.153	0.138	2.537	3.159	3.923
	T6	2.048	-0.037	-0.395	2.101	0.108	0.425	0.460	0.980	0.900	2.857	0.174	2.036	2.919	3.535
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.423	-0.042	0.157	0.527	0.096	0.244	1.000	0.990	0.983	2.455	0.421	1.589	2.292	4.360
	T2	-0.142	-0.035	0.089	0.441	0.092	0.210	1.000	0.990	0.990	2.633	0.413	1.644	2.510	4.449
	T3	-0.213	-0.032	0.098	0.418	0.094	0.214	1.000	0.990	0.990	2.550	0.369	1.719	2.396	4.183
	T4	-0.221	-0.020	0.079	0.478	0.090	0.203	1.000	0.993	0.990	2.645	0.401	1.651	2.543	4.366
	T5	-0.120	-0.025	0.070	0.410	0.092	0.180	1.000	1.000	1.000	2.807	0.368	1.987	2.679	4.501
	T6	-0.266	-0.020	0.084	0.476	0.090	0.194	1.000	0.997	0.990	2.603	0.379	1.633	2.504	4.387
稀疏回归（弹性网络）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	1.526	-0.303	0.266	1.607	0.309	0.274	0.957	0.913	1.000	1.894	0.092	1.381	1.843	2.575
	T2	1.516	-0.277	0.215	1.591	0.283	0.225	0.997	0.983	1.000	2.099	0.137	1.559	2.051	2.896
	T3	1.533	-0.284	0.225	1.620	0.291	0.235	0.990	0.967	1.000	2.076	0.134	1.487	2.056	2.871
	T4	1.136	-0.234	0.202	1.426	0.257	0.213	1.000	0.987	1.000	2.428	0.523	1.635	2.186	4.476
	T5	1.163	-0.203	0.126	1.430	0.228	0.159	1.000	0.997	1.000	2.934	1.018	1.725	2.659	5.622
	T6	1.201	-0.240	0.202	1.463	0.260	0.214	0.997	0.993	0.997	2.364	0.536	1.541	2.146	4.479
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.25 ρ: 0.6 数据：干净	T1	-0.715	-0.075	0.277	0.937	0.160	0.393	0.993	0.960	0.880	2.166	0.442	1.121	2.063	3.923
	T2	0.048	-0.053	0.085	0.818	0.147	0.310	1.000	0.987	0.963	2.628	0.419	1.438	2.614	4.018
	T3	-0.138	-0.051	0.120	0.726	0.153	0.303	1.000	0.987	0.973	2.494	0.358	1.424	2.460	3.942
	T4	-0.338	-0.007	0.082	0.778	0.118	0.266	1.000	1.000	0.973	2.712	0.414	1.688	2.697	4.064
	T5	0.041	-0.005	0.005	0.691	0.118	0.238	1.000	1.000	1.000	3.249	0.524	2.105	3.177	4.922
	T6	-0.376	-0.007	0.086	0.744	0.120	0.277	1.000	0.993	0.973	2.597	0.446	1.442	2.530	4.117
稀疏回归（弹性网络）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.25 ρ: 0.6 数据：外部	T1	1.680	-0.361	0.359	1.982	0.383	0.383	0.797	0.557	0.827	1.635	0.141	0.994	1.616	2.446
	T2	1.635	-0.289	0.217	1.916	0.311	0.247	0.960	0.970	0.990	2.225	0.310	1.411	2.129	3.520
	T3	1.642	-0.301	0.236	1.941	0.323	0.267	0.950	0.943	0.993	2.171	0.294	1.238	2.129	3.313
	T4	0.583	-0.155	0.151	1.503	0.236	0.208	0.997	1.000	0.997	3.258	1.419	1.657	2.974	6.326
	T5	0.786	-0.090	-0.038	1.424	0.171	0.163	1.000	1.000	1.000	4.243	2.132	2.134	3.928	7.804
	T6	0.867	-0.197	0.183	1.580	0.259	0.230	0.987	0.980	0.990	2.847	1.196	1.506	2.580	5.603



表15： 在含有 $n = 20$ 的情况下，普通最小二乘法（干净）和弹性网络（受污染）各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。区块：(i) $n_{\text{sim}} = 300$，迭代次数 $= 5$, $\rho = 0.6$,$P_{\text{ext}} = 0.15$, $P_{\text{miss}} \in \{0.10,0.25\}$；(ii) $n_{\text{sim}} = 50$，迭代次数 $= 10$, $\rho = 0.6$, $P_{\text{ext}} = 0.05$, $P_{\text{miss}} \in \{0.10,0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.424	-0.043	0.158	0.528	0.097	0.246	1.000	0.990	0.983	2.480	0.461	1.606	2.322	4.545
	T2	-0.114	-0.038	0.087	0.452	0.095	0.212	1.000	0.990	0.990	2.647	0.392	1.724	2.538	4.206
	T3	-0.211	-0.032	0.097	0.413	0.093	0.214	1.000	0.990	0.990	2.576	0.367	1.743	2.445	4.219
	T4	-0.211	-0.024	0.084	0.454	0.091	0.205	1.000	0.993	0.990	2.680	0.429	1.725	2.565	4.376
	T5	-0.124	-0.019	0.060	0.449	0.097	0.191	1.000	0.997	0.993	2.909	0.569	2.005	2.756	4.661
	T6	-0.286	-0.018	0.086	0.458	0.088	0.191	1.000	0.997	0.993	2.658	0.465	1.654	2.494	4.536
稀疏回归（弹性网络）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.959	-0.251	0.244	1.004	0.256	0.252	0.997	0.943	0.997	1.447	0.055	1.051	1.439	1.971
	T2	0.888	-0.225	0.203	0.947	0.231	0.212	1.000	0.997	1.000	1.628	0.081	1.209	1.583	2.285
	T3	0.905	-0.230	0.209	0.965	0.236	0.220	1.000	0.993	1.000	1.599	0.069	1.187	1.592	2.080
	T4	0.322	-0.138	0.135	0.956	0.190	0.175	1.000	1.000	1.000	2.255	0.790	1.298	1.906	4.449
	T5	0.311	-0.096	0.046	0.929	0.156	0.130	1.000	1.000	1.000	2.760	1.133	1.475	2.512	5.358
	T6	0.531	-0.170	0.158	0.977	0.206	0.189	1.000	0.993	1.000	2.039	0.744	1.232	1.779	4.875
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.25 ρ: 0.6 数据：干净	T1	-0.725	-0.075	0.277	0.944	0.160	0.393	0.993	0.960	0.880	2.159	0.424	1.071	2.081	3.692
	T2	0.039	-0.056	0.091	0.767	0.145	0.303	1.000	0.977	0.960	2.619	0.419	1.298	2.600	3.960
	T3	-0.142	-0.052	0.121	0.739	0.150	0.305	1.000	0.977	0.963	2.484	0.341	1.406	2.464	3.794
	T4	-0.332	-0.008	0.082	0.742	0.116	0.263	1.000	0.997	0.983	2.697	0.428	1.618	2.631	4.266
	T5	0.000	-0.003	0.010	0.748	0.122	0.244	1.000	1.000	1.000	3.224	0.441	2.186	3.132	4.888
	T6	-0.367	-0.008	0.086	0.718	0.125	0.282	1.000	0.987	0.973	2.616	0.452	1.487	2.575	3.966
稀疏回归（弹性网络）															
n: 20 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.25 ρ: 0.6 数据：外部	T1	1.137	-0.308	0.327	1.290	0.327	0.354	0.893	0.590	0.817	1.266	0.092	0.756	1.244	1.989
	T2	0.983	-0.241	0.217	1.158	0.261	0.250	0.987	0.987	0.987	1.780	0.225	1.058	1.698	2.941
	T3	1.016	-0.253	0.233	1.194	0.274	0.268	0.970	0.977	0.963	1.716	0.186	0.990	1.662	2.666
	T4	-0.553	0.012	-0.011	1.375	0.183	0.166	0.993	0.997	0.997	3.516	1.697	1.641	3.320	6.568
	T5	-0.454	0.090	-0.199	1.271	0.186	0.264	0.993	1.000	1.000	4.462	1.769	2.250	4.403	7.146
	T6	-0.091	-0.073	0.075	1.270	0.207	0.204	0.997	0.997	0.997	3.016	2.135	1.216	2.572	6.718
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.411	-0.025	0.117	0.533	0.067	0.175	1.000	1.000	0.980	2.283	0.243	1.521	2.221	3.317
	T2	-0.136	-0.016	0.047	0.446	0.066	0.140	1.000	1.000	1.000	2.452	0.200	1.710	2.451	3.465
	T3	-0.247	-0.012	0.062	0.412	0.061	0.144	1.000	1.000	1.000	2.411	0.241	1.739	2.353	3.649
	T4	-0.216	-0.005	0.044	0.432	0.066	0.140	1.000	1.000	1.000	2.473	0.260	1.627	2.498	3.916
	T5	-0.090	-0.004	0.020	0.334	0.063	0.126	1.000	1.000	1.000	2.678	0.201	1.988	2.580	3.521
	T6	-0.259	0.002	0.035	0.437	0.062	0.124	1.000	1.000	1.000	2.499	0.326	1.654	2.491	4.020
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	1.650	-0.314	0.267	1.778	0.323	0.274	0.900	0.880	0.980	1.866	0.095	1.349	1.849	2.415
	T2	1.652	-0.291	0.219	1.786	0.301	0.228	0.980	0.940	1.000	2.087	0.131	1.508	2.072	2.885
	T3	1.638	-0.293	0.223	1.743	0.301	0.232	0.980	0.980	1.000	2.083	0.132	1.518	2.049	2.951
	T4	1.160	-0.237	0.204	1.490	0.264	0.216	1.000	0.980	1.000	2.476	0.653	1.615	2.237	4.320
	T5	1.192	-0.201	0.117	1.459	0.224	0.138	1.000	1.000	1.000	3.013	1.118	1.821	2.653	5.465
	T6	1.285	-0.251	0.208	1.570	0.272	0.219	1.000	0.940	1.000	2.363	0.504	1.525	2.180	4.062
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.640	-0.112	0.331	1.063	0.214	0.470	0.980	0.920	0.800	1.980	0.459	0.837	1.956	3.319
	T2	0.294	-0.084	0.095	1.030	0.191	0.341	1.000	0.960	0.920	2.545	0.417	1.272	2.614	3.869
	T3	-0.008	-0.075	0.138	0.920	0.192	0.347	1.000	0.960	0.920	2.394	0.354	1.156	2.491	3.456
	T4	-0.298	-0.014	0.087	0.738	0.133	0.295	1.000	1.000	1.000	2.635	0.346	1.581	2.704	3.764
	T5	0.191	-0.019	0.004	0.643	0.137	0.266	1.000	1.000	1.000	3.224	0.234	2.276	3.390	3.946
	T6	-0.294	-0.028	0.113	0.744	0.164	0.353	1.000	0.980	0.940	2.541	0.388	1.257	2.592	3.652
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	1.748	-0.396	0.423	2.124	0.429	0.459	0.740	0.560	0.660	1.594	0.165	1.046	1.608	2.656
	T2	1.642	-0.288	0.217	1.951	0.314	0.252	0.940	1.000	0.980	2.318	0.271	1.499	2.317	3.473
	T3	1.746	-0.319	0.257	2.067	0.350	0.295	0.920	0.980	0.940	2.270	0.317	1.402	2.179	3.470
	T4	0.188	-0.098	0.117	1.594	0.214	0.177	1.000	1.000	1.000	3.703	1.348	2.092	3.376	6.020
	T5	0.518	-0.007	-0.148	1.323	0.144	0.252	1.000	1.000	1.000	5.068	2.619	2.927	4.846	8.192
	T6	0.626	-0.165	0.174	1.533	0.233	0.220	1.000	0.980	0.980	3.117	1.182	1.997	2.611	5.219

表16： 普通最小二乘法（干净）和弹性网络（受污染）在不同多重插补方法下的模型系数偏差、均方根误差、95%覆盖率及样本外预测均方误差。第一部分： $\boldsymbol{n = 20}$, $n_{\text{sim}} = 50$, 迭代次数 $= 10$, $\rho = 0.6$,$P_{\text{ext}} = 0.10$, $P_{\text{miss}} \in \{0.10,0.30\}$。第二部分： $\boldsymbol{n = 40}$, $n_{\text{sim}} = 50$, 迭代次数 $= 5$, $\rho = 0$, $P_{\text{ext}} = 0.05$,$P_{\text{miss}} \in \{0.10,0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.420	-0.025	0.120	0.539	0.067	0.177	1.000	1.000	0.980	2.331	0.265	1.604	2.206	3.439
	T2	-0.141	-0.018	0.054	0.443	0.064	0.144	1.000	1.000	1.000	2.491	0.205	1.758	2.420	3.404
	T3	-0.225	-0.016	0.066	0.426	0.063	0.150	1.000	1.000	1.000	2.443	0.227	1.699	2.336	3.531
	T4	-0.241	-0.002	0.043	0.409	0.057	0.134	1.000	1.000	1.000	2.501	0.240	1.668	2.433	3.536
	T5	-0.108	0.001	0.012	0.371	0.063	0.135	1.000	1.000	1.000	2.741	0.229	1.941	2.692	3.921
	T6	-0.258	0.001	0.037	0.412	0.060	0.122	1.000	1.000	1.000	2.494	0.259	1.688	2.444	3.743
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	1.081	-0.262	0.242	1.156	0.269	0.251	1.000	0.860	1.000	1.457	0.047	1.083	1.452	1.866
	T2	1.021	-0.237	0.202	1.102	0.244	0.211	1.000	1.000	1.000	1.629	0.064	1.268	1.599	2.207
	T3	1.049	-0.242	0.205	1.126	0.249	0.214	1.000	1.000	1.000	1.617	0.060	1.205	1.570	2.190
	T4	0.433	-0.159	0.153	1.111	0.207	0.179	1.000	1.000	1.000	2.256	0.823	1.263	1.863	4.260
	T5	0.391	-0.106	0.044	1.183	0.176	0.115	1.000	1.000	1.000	3.036	1.289	1.641	2.615	4.995
	T6	0.497	-0.174	0.171	1.140	0.217	0.193	1.000	0.980	1.000	2.146	0.761	1.238	1.815	3.951
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.617	-0.113	0.329	1.031	0.213	0.469	0.980	0.920	0.800	1.912	0.252	0.917	2.023	2.649
	T2	0.256	-0.083	0.101	0.911	0.186	0.363	1.000	0.980	0.940	2.484	0.292	1.298	2.470	3.473
	T3	0.061	-0.075	0.125	0.894	0.189	0.359	1.000	0.940	0.920	2.330	0.312	1.226	2.410	3.260
	T4	-0.266	-0.010	0.075	0.666	0.139	0.314	1.000	1.000	0.960	2.661	0.246	1.603	2.707	3.442
	T5	0.160	-0.017	0.006	0.694	0.130	0.249	1.000	1.000	1.000	3.188	0.345	2.090	3.238	3.990
	T6	-0.274	-0.033	0.117	0.672	0.153	0.339	1.000	0.980	0.960	2.541	0.235	1.425	2.666	3.256
稀疏回归（弹性网络）															
n: 20 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	1.294	-0.338	0.361	1.519	0.362	0.394	0.780	0.540	0.660	1.202	0.072	0.696	1.210	1.692
	T2	1.196	-0.262	0.220	1.417	0.290	0.266	0.960	0.960	0.960	1.832	0.174	1.103	1.827	2.685
	T3	1.194	-0.273	0.239	1.425	0.298	0.276	0.960	1.000	0.960	1.763	0.139	1.015	1.774	2.445
	T4	-0.506	-0.006	0.023	1.413	0.162	0.124	1.000	1.000	1.000	3.583	1.237	2.003	3.358	5.400
	T5	-0.398	0.095	-0.221	1.229	0.164	0.261	1.000	1.000	1.000	4.850	1.351	2.527	5.008	6.669
	T6	-0.166	-0.068	0.092	1.304	0.193	0.192	1.000	1.000	1.000	3.180	2.296	1.401	2.891	6.975
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.0 数据：干净	T1	2.253	-0.203	-0.121	2.277	0.205	0.127	0.740	0.620	1.000	2.081	0.049	1.582	2.110	2.468
	T2	2.490	-0.199	-0.177	2.512	0.201	0.181	0.680	0.780	1.000	2.242	0.053	1.763	2.284	2.636
	T3	2.449	-0.200	-0.167	2.474	0.201	0.172	0.700	0.820	1.000	2.236	0.060	1.671	2.267	2.636
	T4	2.367	-0.191	-0.167	2.399	0.194	0.173	0.840	0.900	1.000	2.336	0.072	1.931	2.346	2.806
	T5	2.538	-0.190	-0.204	2.579	0.195	0.211	0.800	0.920	1.000	2.547	0.212	2.012	2.479	3.874
	T6	2.405	-0.193	-0.172	2.434	0.196	0.178	0.800	0.900	1.000	2.320	0.078	1.831	2.354	2.758
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.0 数据：外部	T1	3.436	-0.238	-0.279	3.455	0.241	0.282	0.020	0.160	0.920	2.566	0.091	1.897	2.606	2.999
	T2	3.652	-0.234	-0.329	3.669	0.238	0.333	0.020	0.360	0.460	2.736	0.113	2.097	2.797	3.347
	T3	3.580	-0.225	-0.334	3.595	0.229	0.337	0.020	0.540	0.460	2.746	0.115	1.994	2.753	3.215
	T4	3.547	-0.221	-0.338	3.572	0.226	0.342	0.060	0.460	0.480	2.854	0.187	2.049	2.905	3.654
	T5	3.581	-0.209	-0.366	3.617	0.218	0.369	0.140	0.560	0.500	3.045	0.246	2.278	3.042	3.965
	T6	3.531	-0.219	-0.337	3.561	0.225	0.341	0.060	0.520	0.480	2.826	0.160	2.049	2.837	3.532
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.0 数据：干净	T1	1.884	-0.223	-0.009	1.958	0.229	0.065	0.720	0.340	1.000	1.725	0.100	1.101	1.794	2.271
	T2	2.733	-0.215	-0.190	2.798	0.223	0.203	0.660	0.840	1.000	2.325	0.176	1.476	2.351	3.098
	T3	2.577	-0.214	-0.163	2.631	0.220	0.176	0.680	0.800	1.000	2.239	0.193	1.392	2.264	3.102
	T4	2.422	-0.196	-0.169	2.486	0.203	0.185	0.900	0.920	1.000	2.605	0.312	1.651	2.589	3.660
	T5	3.070	-0.205	-0.276	3.142	0.214	0.290	0.740	0.920	1.000	3.193	0.403	2.234	3.140	4.428
	T6	2.333	-0.192	-0.161	2.400	0.198	0.182	0.820	0.880	1.000	2.459	0.277	1.472	2.415	3.614
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.0 数据：外部	T1	2.939	-0.250	-0.149	3.038	0.261	0.171	0.140	0.260	1.000	2.225	0.139	1.561	2.281	2.872
	T2	3.599	-0.223	-0.337	3.661	0.234	0.349	0.220	0.620	0.720	2.840	0.189	2.195	2.832	3.573
	T3	3.442	-0.214	-0.323	3.528	0.229	0.334	0.200	0.700	0.720	2.803	0.178	1.948	2.856	3.550
	T4	3.226	-0.185	-0.337	3.316	0.205	0.350	0.360	0.720	0.680	3.181	0.334	2.183	3.123	4.371
	T5	3.456	-0.175	-0.403	3.573	0.199	0.412	0.360	0.820	0.660	3.515	0.315	2.565	3.590	4.534
	T6	3.317	-0.192	-0.342	3.393	0.208	0.355	0.320	0.820	0.780	3.079	0.277	2.153	2.997	3.976



表17： 普通最小二乘法（干净）和弹性网络（受污染）中，各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。 $\boldsymbol{n = 40}$, $n_{\text{sim}} = 50$, 迭代次数 $= 5$, $P_{\text{ext}} = 0.10$, $P_{\text{miss}} \in \{0.10, 0.30\}$; 随后是一个附加区块，其中包含 $\boldsymbol{n = 20}$, $n_{\text{sim}} = 1000$, 迭代次数 $= 10$, $P_{\text{ext}} = 0.30$, $P_{\text{miss}} \in \{0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
样本量: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.0 数据：干净	T1	2.253	-0.203	-0.122	2.279	0.205	0.128	0.720	0.620	1.000	2.077	0.043	1.582	2.116	2.389
	T2	2.503	-0.201	-0.176	2.526	0.203	0.181	0.680	0.800	1.000	2.251	0.050	1.763	2.283	2.623
	T3	2.449	-0.199	-0.168	2.475	0.201	0.173	0.720	0.860	1.000	2.239	0.055	1.671	2.280	2.624
	T4	2.374	-0.189	-0.172	2.408	0.193	0.177	0.860	0.900	1.000	2.351	0.068	1.931	2.346	2.852
	T5	2.574	-0.191	-0.207	2.619	0.197	0.215	0.720	0.900	1.000	2.560	0.213	2.012	2.479	3.891
	T6	2.380	-0.191	-0.173	2.415	0.194	0.178	0.800	0.880	1.000	2.314	0.070	1.831	2.299	2.718
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.0 数据：外部	T1	3.277	-0.239	-0.245	3.286	0.241	0.248	0.000	0.040	1.000	2.505	0.084	1.839	2.574	2.927
	T2	3.428	-0.229	-0.296	3.436	0.232	0.300	0.000	0.240	0.820	2.698	0.106	2.017	2.735	3.205
	T3	3.337	-0.220	-0.296	3.347	0.223	0.300	0.000	0.380	0.800	2.708	0.116	1.934	2.751	3.153
	T4	3.322	-0.214	-0.304	3.337	0.219	0.308	0.020	0.520	0.900	2.842	0.199	2.016	2.843	3.644
	T5	3.371	-0.212	-0.318	3.384	0.217	0.322	0.020	0.580	0.860	2.906	0.185	2.020	2.973	3.575
	T6	3.367	-0.221	-0.298	3.375	0.225	0.303	0.000	0.460	0.840	2.743	0.136	1.922	2.794	3.367
线性回归（普通最小二乘法）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.0 数据：干净	T1	1.920	-0.227	-0.009	1.984	0.233	0.069	0.680	0.320	1.000	1.728	0.093	1.103	1.757	2.272
	T2	2.766	-0.218	-0.190	2.830	0.226	0.205	0.660	0.820	1.000	2.334	0.167	1.517	2.318	3.133
	T3	2.609	-0.217	-0.163	2.651	0.222	0.175	0.720	0.820	1.000	2.254	0.177	1.398	2.239	3.102
	T4	2.477	-0.202	-0.166	2.540	0.210	0.185	0.880	0.880	1.000	2.605	0.274	1.763	2.563	3.660
	T5	3.087	-0.206	-0.276	3.163	0.217	0.290	0.720	0.900	1.000	3.201	0.376	2.243	3.089	4.428
	T6	2.357	-0.193	-0.161	2.424	0.201	0.182	0.800	0.860	1.000	2.467	0.221	1.595	2.418	3.614
稀疏回归（弹性网络）															
n: 40 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.0 数据：外部	T1	2.936	-0.262	-0.125	2.997	0.272	0.151	0.060	0.100	1.000	2.162	0.155	1.434	2.177	2.828
	T2	3.401	-0.221	-0.300	3.436	0.230	0.314	0.060	0.580	0.780	2.809	0.217	2.010	2.779	3.627
	T3	3.239	-0.212	-0.287	3.295	0.225	0.299	0.180	0.580	0.860	2.755	0.201	1.941	2.818	3.626
	T4	2.939	-0.169	-0.314	3.015	0.192	0.329	0.240	0.780	0.780	3.252	0.299	2.344	3.264	4.220
	T5	3.114	-0.151	-0.383	3.190	0.183	0.402	0.260	0.860	0.760	3.560	0.505	2.273	3.530	5.010
	T6	3.178	-0.193	-0.314	3.248	0.213	0.332	0.220	0.740	0.880	3.072	0.322	2.001	2.986	4.296
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.30 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.413	-0.036	0.143	0.518	0.085	0.217	1.000	0.996	0.991	2.524	0.419	1.607	2.368	4.287
	T2	-0.097	-0.031	0.071	0.404	0.084	0.182	1.000	0.997	0.996	2.690	0.345	1.820	2.576	4.237
	T3	-0.208	-0.025	0.083	0.402	0.081	0.183	1.000	0.996	0.996	2.620	0.335	1.804	2.500	4.145
	T4	-0.221	-0.016	0.070	0.415	0.075	0.170	1.000	0.998	0.996	2.693	0.381	1.768	2.578	4.347
	T5	-0.140	-0.014	0.052	0.401	0.083	0.162	1.000	0.998	0.998	2.883	0.397	2.007	2.749	4.577
	T6	-0.264	-0.012	0.069	0.436	0.077	0.162	1.000	0.997	0.996	2.672	0.405	1.760	2.547	4.476
稀疏回归（弹性网络）															
n: 20 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.30 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	1.170	-0.356	0.399	1.214	0.361	0.405	0.900	0.980	0.902	1.441	0.038	1.045	1.438	1.865
	T2	1.033	-0.305	0.323	1.107	0.315	0.339	0.921	0.997	0.921	1.577	0.046	1.183	1.566	2.025
	T3	1.049	-0.313	0.337	1.119	0.323	0.350	0.913	0.999	0.913	1.558	0.045	1.150	1.549	2.018
	T4	0.081	-0.047	-0.007	0.786	0.201	0.251	1.000	1.000	1.000	2.337	0.648	1.290	2.168	4.096
	T5	0.271	-0.070	0.002	0.558	0.149	0.179	1.000	1.000	1.000	2.318	0.302	1.518	2.245	3.555
	T6	0.797	-0.238	0.237	0.929	0.266	0.283	0.950	1.000	0.950	1.846	0.289	1.206	1.700	3.117
线性回归（普通最小二乘法）															
n: 20 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.875	-0.087	0.328	1.151	0.181	0.452	0.989	0.903	0.799	2.045	0.496	0.877	1.995	3.702
	T2	0.046	-0.058	0.093	0.851	0.156	0.333	0.999	0.979	0.950	2.585	0.402	1.283	2.593	3.818
	T3	-0.211	-0.051	0.133	0.819	0.160	0.336	0.999	0.973	0.937	2.453	0.397	1.228	2.451	3.692
	T4	-0.406	-0.004	0.088	0.837	0.125	0.297	1.000	0.998	0.968	2.710	0.409	1.544	2.717	3.955
	T5	0.074	0.003	-0.016	0.738	0.124	0.240	1.000	1.000	1.000	3.355	0.389	2.264	3.312	4.753
	T6	-0.402	-0.006	0.088	0.807	0.135	0.316	1.000	0.988	0.952	2.614	0.501	1.399	2.594	4.158
稀疏回归（弹性网络）															
n: 20 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	1.198	-0.320	0.341	1.411	0.340	0.369	0.832	0.428	0.700	1.180	0.106	0.634	1.146	1.882
	T2	1.069	-0.248	0.211	1.281	0.266	0.242	0.994	0.982	0.991	1.793	0.228	1.005	1.715	2.804
	T3	1.078	-0.255	0.226	1.306	0.276	0.258	0.985	0.962	0.984	1.703	0.215	0.917	1.655	2.714
	T4	-0.612	0.001	0.022	1.632	0.188	0.147	0.978	0.999	1.000	3.608	1.734	1.741	3.410	6.421
	T5	-0.476	0.102	-0.228	1.485	0.186	0.266	0.996	1.000	1.000	4.961	1.837	2.758	4.861	8.013
	T6	-0.306	-0.059	0.090	1.540	0.204	0.183	0.990	0.993	0.999	3.197	2.359	1.254	2.828	6.813



表18： 普通最小二乘法（干净）和弹性网络（弹性网络，受污染）在不同多重插补方法下的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差，其中 $\boldsymbol{n = 80}$、 $n_{\text{sim}} = 50$、迭代次数 $= 5$、 $\rho = 0.6$、污染程度 $P_{\text{ext}} \in \{0.03, 0.05\}$ 以及缺失率 $P_{\text{miss}} \in \{0.05, 0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	0.282	0.029	-0.159	0.311	0.032	0.161	1.000	1.000	1.000	2.488	0.011	2.330	2.473	2.682
	T2	0.389	0.035	-0.192	0.417	0.037	0.194	1.000	1.000	1.000	2.557	0.010	2.413	2.555	2.724
	T3	0.371	0.036	-0.192	0.397	0.038	0.194	1.000	1.000	1.000	2.551	0.012	2.406	2.540	2.747
	T4	0.387	0.034	-0.191	0.415	0.037	0.194	1.000	1.000	1.000	2.538	0.011	2.369	2.522	2.750
	T5	0.425	0.036	-0.202	0.453	0.039	0.205	1.000	1.000	1.000	2.592	0.014	2.396	2.581	2.855
	T6	0.401	0.036	-0.197	0.428	0.038	0.199	1.000	1.000	1.000	2.539	0.010	2.368	2.527	2.723
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	-0.250	0.110	-0.170	0.280	0.111	0.171	1.000	1.000	1.000	1.919	0.006	1.781	1.912	2.073
	T2	-0.215	0.121	-0.199	0.253	0.122	0.199	1.000	1.000	1.000	1.984	0.006	1.875	1.972	2.153
	T3	-0.230	0.122	-0.197	0.259	0.123	0.198	1.000	1.000	0.980	1.979	0.006	1.843	1.983	2.118
	T4	-0.284	0.128	-0.199	0.336	0.129	0.200	1.000	0.980	1.000	2.029	0.014	1.849	2.012	2.281
	T5	-0.279	0.132	-0.208	0.336	0.135	0.210	1.000	1.000	1.000	2.080	0.018	1.844	2.062	2.340
	T6	-0.240	0.123	-0.196	0.280	0.124	0.197	1.000	1.000	1.000	1.994	0.008	1.868	1.976	2.210
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.154	0.021	-0.116	0.306	0.034	0.124	1.000	1.000	1.000	2.412	0.014	2.185	2.432	2.577
	T2	0.383	0.035	-0.192	0.474	0.046	0.198	1.000	1.000	1.000	2.565	0.015	2.313	2.583	2.762
	T3	0.380	0.032	-0.184	0.456	0.040	0.190	1.000	1.000	1.000	2.553	0.020	2.277	2.555	2.775
	T4	0.383	0.032	-0.184	0.466	0.041	0.189	1.000	1.000	1.000	2.554	0.015	2.343	2.571	2.761
	T5	0.437	0.035	-0.203	0.517	0.044	0.211	1.000	1.000	1.000	2.634	0.026	2.361	2.623	2.952
	T6	0.441	0.026	-0.184	0.504	0.037	0.190	1.000	1.000	1.000	2.538	0.016	2.316	2.542	2.790
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	-0.230	0.095	-0.144	0.314	0.098	0.145	1.000	1.000	1.000	1.859	0.010	1.633	1.852	2.055
	T2	-0.206	0.121	-0.200	0.322	0.125	0.202	1.000	0.980	0.980	2.002	0.013	1.787	2.002	2.198
	T3	-0.166	0.114	-0.194	0.276	0.117	0.196	1.000	1.000	1.000	1.979	0.013	1.725	1.971	2.232
	T4	-0.294	0.129	-0.200	0.423	0.134	0.201	1.000	0.920	0.960	2.058	0.024	1.794	2.041	2.349
	T5	-0.161	0.131	-0.229	0.337	0.136	0.232	1.000	1.000	0.880	2.206	0.034	1.930	2.192	2.559
	T6	-0.232	0.126	-0.204	0.330	0.129	0.206	1.000	1.000	0.940	2.035	0.015	1.786	2.008	2.275
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.353	-0.002	0.039	0.634	0.041	0.109	1.000	1.000	0.980	2.100	0.043	1.653	2.095	2.460
	T2	0.418	0.036	-0.200	0.650	0.052	0.225	1.000	1.000	0.960	2.565	0.051	2.058	2.549	2.938
	T3	0.351	0.039	-0.192	0.637	0.057	0.215	1.000	1.000	0.900	2.527	0.059	2.060	2.526	2.973
	T4	0.373	0.042	-0.207	0.580	0.055	0.232	1.000	1.000	0.900	2.628	0.058	2.171	2.625	3.022
	T5	0.536	0.049	-0.254	0.729	0.058	0.277	1.000	1.000	0.860	2.893	0.070	2.284	2.941	3.295
	T6	0.375	0.034	-0.188	0.559	0.050	0.211	1.000	1.000	0.960	2.552	0.057	1.993	2.549	2.943
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	-0.506	0.057	-0.014	0.707	0.078	0.055	0.980	0.940	1.000	1.576	0.033	1.250	1.573	1.906
	T2	-0.283	0.128	-0.196	0.545	0.139	0.204	1.000	0.960	0.940	2.047	0.059	1.687	2.007	2.514
	T3	-0.378	0.130	-0.182	0.584	0.140	0.189	1.000	0.960	0.920	2.010	0.048	1.645	2.007	2.391
	T4	-0.692	0.172	-0.206	0.936	0.183	0.216	0.980	0.760	0.820	2.237	0.085	1.769	2.240	2.757
	T5	-0.437	0.194	-0.299	0.733	0.204	0.306	1.000	0.820	0.780	2.639	0.116	1.894	2.680	3.179
	T6	-0.513	0.151	-0.199	0.779	0.162	0.207	0.960	0.900	0.920	2.149	0.099	1.671	2.137	2.831
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	0.272	0.030	-0.159	0.301	0.033	0.161	1.000	1.000	1.000	2.489	0.010	2.301	2.486	2.661
	T2	0.374	0.036	-0.192	0.399	0.038	0.194	1.000	1.000	1.000	2.543	0.011	2.342	2.543	2.717
	T3	0.362	0.037	-0.192	0.391	0.040	0.193	1.000	1.000	1.000	2.544	0.012	2.349	2.547	2.738
	T4	0.393	0.034	-0.193	0.420	0.037	0.195	1.000	1.000	1.000	2.557	0.011	2.346	2.553	2.738
	T5	0.404	0.037	-0.201	0.440	0.040	0.205	1.000	1.000	1.000	2.592	0.019	2.393	2.569	2.860
	T6	0.384	0.036	-0.194	0.411	0.038	0.196	1.000	1.000	1.000	2.544	0.010	2.356	2.543	2.711
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	-0.226	0.119	-0.198	0.254	0.120	0.198	1.000	1.000	1.000	1.861	0.006	1.707	1.858	1.982
	T2	-0.227	0.133	-0.224	0.265	0.134	0.225	1.000	1.000	0.560	1.926	0.006	1.768	1.923	2.073
	T3	-0.235	0.133	-0.223	0.270	0.134	0.224	1.000	1.000	0.560	1.922	0.008	1.769	1.921	2.056
	T4	-0.260	0.137	-0.227	0.299	0.139	0.228	1.000	0.980	0.540	1.962	0.012	1.773	1.950	2.181
	T5	-0.246	0.140	-0.234	0.283	0.142	0.236	1.000	1.000	0.640	2.007	0.012	1.827	2.006	2.200
	T6	-0.243	0.135	-0.225	0.268	0.136	0.226	1.000	1.000	0.500	1.937	0.008	1.769	1.938	2.114

表19： 普通最小二乘法（干净）和弹性网络（受污染）在不同多重插补方法下的模型系数偏差、均方根误差、95%覆盖率及样本外预测均方误差，其中 $\boldsymbol{n = 80}$, $n_{\text{sim}} = 50$, 迭代次数 $= 5$, $\rho = 0.6$, 污染程度 $P_{\text{ext}} \in \{0.05, 0.10\}$, 缺失率 $P_{\text{miss}} \in \{0.05, 0.10, 0.30\}$

设置	MI	Bias			RMSE			覆盖率			预测均方误差（样本外）				
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.159	0.022	-0.118	0.309	0.033	0.125	1.000	1.000	1.000	2.416	0.012	2.191	2.417	2.565
	T2	0.397	0.035	-0.195	0.479	0.043	0.200	1.000	1.000	1.000	2.561	0.016	2.266	2.573	2.733
	T3	0.367	0.034	-0.187	0.443	0.041	0.192	1.000	1.000	1.000	2.561	0.016	2.309	2.566	2.768
	T4	0.397	0.032	-0.189	0.473	0.040	0.194	1.000	1.000	0.980	2.559	0.013	2.381	2.546	2.758
	T5	0.459	0.035	-0.210	0.530	0.043	0.217	1.000	1.000	1.000	2.660	0.030	2.368	2.657	2.994
	T6	0.440	0.029	-0.191	0.513	0.040	0.196	1.000	1.000	0.980	2.550	0.013	2.357	2.551	2.756
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	-0.181	0.101	-0.171	0.263	0.104	0.172	1.000	1.000	1.000	1.798	0.010	1.581	1.795	1.985
	T2	-0.156	0.125	-0.223	0.267	0.128	0.225	1.000	0.980	0.640	1.941	0.014	1.763	1.939	2.148
	T3	-0.166	0.125	-0.221	0.265	0.128	0.222	1.000	1.000	0.580	1.916	0.012	1.675	1.913	2.151
	T4	-0.260	0.138	-0.230	0.373	0.143	0.232	1.000	0.920	0.500	1.992	0.021	1.725	1.974	2.238
	T5	-0.196	0.144	-0.252	0.317	0.148	0.255	1.000	0.960	0.620	2.110	0.033	1.832	2.072	2.497
	T6	-0.189	0.133	-0.232	0.283	0.137	0.234	1.000	0.980	0.500	1.953	0.015	1.733	1.943	2.164
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.359	-0.002	0.040	0.650	0.045	0.113	1.000	1.000	0.980	2.093	0.045	1.653	2.103	2.461
	T2	0.410	0.035	-0.198	0.647	0.053	0.223	1.000	1.000	0.960	2.551	0.052	2.058	2.546	2.938
	T3	0.357	0.038	-0.192	0.647	0.059	0.217	1.000	1.000	0.900	2.521	0.061	2.060	2.514	2.973
	T4	0.386	0.040	-0.204	0.579	0.053	0.230	1.000	1.000	0.860	2.620	0.063	2.171	2.625	3.012
	T5	0.540	0.049	-0.255	0.787	0.062	0.280	1.000	1.000	0.840	2.900	0.072	2.284	2.934	3.339
	T6	0.380	0.034	-0.188	0.606	0.051	0.215	1.000	1.000	0.920	2.552	0.058	1.993	2.547	2.943
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	-0.328	0.052	-0.044	0.563	0.075	0.069	0.980	0.980	1.000	1.533	0.032	1.236	1.535	1.853
	T2	-0.236	0.133	-0.221	0.521	0.145	0.228	1.000	0.900	0.920	2.014	0.053	1.623	2.038	2.431
	T3	-0.344	0.138	-0.209	0.559	0.148	0.215	1.000	0.920	0.840	1.955	0.045	1.585	1.961	2.340
	T4	-0.707	0.190	-0.242	0.939	0.201	0.250	0.940	0.660	0.660	2.199	0.083	1.670	2.153	2.734
	T5	-0.495	0.215	-0.334	0.807	0.228	0.342	0.960	0.660	0.540	2.592	0.097	2.059	2.604	3.260
	T6	-0.446	0.157	-0.228	0.740	0.170	0.236	0.960	0.800	0.780	2.077	0.089	1.632	2.046	2.762
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	0.273	0.030	-0.159	0.302	0.032	0.161	1.000	1.000	1.000	2.493	0.009	2.339	2.490	2.673
	T2	0.370	0.037	-0.192	0.390	0.038	0.194	1.000	1.000	1.000	2.560	0.012	2.379	2.537	2.755
	T3	0.373	0.036	-0.192	0.397	0.038	0.193	1.000	1.000	1.000	2.564	0.010	2.396	2.576	2.755
	T4	0.390	0.036	-0.196	0.417	0.038	0.198	1.000	1.000	1.000	2.566	0.009	2.381	2.565	2.752
	T5	0.395	0.036	-0.198	0.426	0.039	0.200	1.000	1.000	1.000	2.589	0.013	2.405	2.587	2.846
	T6	0.389	0.036	-0.195	0.414	0.038	0.198	1.000	1.000	1.000	2.551	0.009	2.383	2.552	2.723
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：外部。	T1	-0.205	0.114	-0.191	0.222	0.115	0.192	1.000	1.000	0.960	1.755	0.006	1.645	1.743	1.911
	T2	-0.223	0.128	-0.216	0.245	0.129	0.217	1.000	1.000	0.500	1.826	0.007	1.689	1.827	1.964
	T3	-0.219	0.128	-0.216	0.238	0.129	0.216	1.000	1.000	0.560	1.817	0.007	1.704	1.811	1.998
	T4	-0.221	0.127	-0.213	0.237	0.128	0.214	1.000	1.000	0.520	1.830	0.007	1.701	1.820	1.994
	T5	-0.227	0.132	-0.222	0.249	0.134	0.223	1.000	1.000	0.580	1.896	0.012	1.745	1.885	2.193
	T6	-0.208	0.127	-0.215	0.223	0.127	0.215	1.000	1.000	0.540	1.828	0.008	1.706	1.829	2.046
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.147	0.022	-0.115	0.301	0.036	0.125	1.000	1.000	1.000	2.397	0.017	2.171	2.403	2.667
	T2	0.394	0.033	-0.191	0.486	0.044	0.197	1.000	1.000	1.000	2.547	0.023	2.308	2.549	2.812
	T3	0.383	0.032	-0.185	0.449	0.040	0.191	1.000	1.000	1.000	2.523	0.021	2.253	2.518	2.770
	T4	0.392	0.032	-0.186	0.466	0.043	0.193	1.000	1.000	1.000	2.549	0.017	2.365	2.527	2.802
	T5	0.421	0.035	-0.199	0.516	0.045	0.209	1.000	1.000	1.000	2.615	0.029	2.367	2.599	3.067
	T6	0.411	0.029	-0.185	0.493	0.042	0.192	1.000	1.000	1.000	2.522	0.020	2.270	2.509	2.789
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	-0.183	0.099	-0.166	0.248	0.102	0.168	1.000	1.000	1.000	1.698	0.013	1.455	1.687	1.901
	T2	-0.200	0.124	-0.214	0.274	0.128	0.216	1.000	1.000	0.640	1.833	0.016	1.580	1.843	2.024
	T3	-0.234	0.128	-0.214	0.288	0.131	0.216	1.000	1.000	0.660	1.817	0.015	1.561	1.813	2.019
	T4	-0.231	0.131	-0.220	0.291	0.135	0.222	1.000	0.940	0.480	1.855	0.015	1.629	1.865	2.082
	T5	-0.232	0.142	-0.241	0.284	0.144	0.243	1.000	1.000	0.460	1.982	0.022	1.727	1.993	2.237
	T6	-0.181	0.126	-0.220	0.232	0.129	0.222	1.000	0.980	0.560	1.844	0.011	1.653	1.828	2.024

表 20： 普通最小二乘法（干净）和弹性网络（受污染）各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。上部区块： $\boldsymbol{n = 80}$、 $n_{\text{sim}} = 50$，迭代次数为 $= 5$、 $\rho = 0.6$，包含 $P_{\text{ext}} = 0.10$ 和 $P_{\text{miss}} = 0.30$；随后是一个附加区块，包含 $\boldsymbol{n = 200}$、 $n_{\text{sim}} = 50$，迭代次数为 $= 5$、$\rho = 0.6$、 $P_{\text{ext}} = 0.03$ 和 $P_{\text{miss}} \in \{0.05, 0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
样本量：80	T1	-0.368	-0.004	0.047	0.672	0.046	0.101	0.980	1.000	1.000	2.112	0.036	1.694	2.126	2.410
n_{sim}：50	T2	0.381	0.037	-0.194	0.683	0.058	0.216	1.000	1.000	0.900	2.582	0.046	2.162	2.601	2.954
迭代次数：5	T3	0.328	0.037	-0.183	0.577	0.054	0.202	1.000	1.000	0.960	2.540	0.064	1.964	2.561	2.953
P_{ext}：0.10	T4	0.357	0.039	-0.196	0.584	0.055	0.218	1.000	1.000	0.900	2.652	0.065	2.171	2.659	3.194
P_{miss}：0.30	T5	0.560	0.044	-0.249	0.753	0.056	0.269	1.000	1.000	0.900	2.896	0.074	2.494	2.876	3.415
ρ：0.6	T6	0.385	0.034	-0.191	0.623	0.053	0.213	1.000	1.000	0.900	2.586	0.053	2.141	2.627	2.938
数据：干净															
稀疏回归（弹性网络）															
n：80	T1	-0.193	0.040	-0.047	0.391	0.064	0.071	0.980	0.980	1.000	1.452	0.030	1.154	1.443	1.791
n_{sim}：50	T2	-0.203	0.126	-0.212	0.419	0.137	0.219	1.000	0.900	0.780	1.907	0.049	1.503	1.911	2.272
迭代次数：5	T3	-0.305	0.131	-0.202	0.467	0.140	0.209	1.000	0.880	0.800	1.864	0.045	1.462	1.846	2.226
P_{ext}：0.10	T4	-0.598	0.179	-0.239	0.778	0.189	0.247	0.960	0.740	0.640	2.069	0.088	1.628	2.026	2.678
P_{miss}：0.30	T5	-0.442	0.201	-0.312	0.671	0.211	0.319	0.960	0.860	0.620	2.418	0.090	1.895	2.385	3.030
ρ：0.6	T6	-0.299	0.140	-0.220	0.552	0.151	0.227	0.960	0.920	0.760	1.930	0.062	1.523	1.957	2.369
数据：外部															
线性回归（普通最小二乘法）															
n：200	T1	-0.050	-0.011	0.034	0.128	0.015	0.036	1.000	1.000	1.000	2.089	0.004	1.930	2.100	2.179
n_{sim}：50	T2	0.023	-0.004	0.006	0.128	0.012	0.015	1.000	1.000	1.000	2.158	0.005	1.986	2.167	2.254
迭代次数：5	T3	0.019	-0.004	0.005	0.132	0.012	0.016	1.000	1.000	1.000	2.161	0.005	2.001	2.169	2.251
P_{ext}：0.03	T4	0.003	-0.003	0.007	0.127	0.011	0.016	1.000	1.000	1.000	2.156	0.004	2.021	2.167	2.271
P_{miss}：0.05	T5	0.057	-0.005	-0.001	0.163	0.015	0.015	1.000	1.000	1.000	2.198	0.005	2.045	2.201	2.317
ρ：0.6	T6	0.005	-0.003	0.006	0.132	0.011	0.016	1.000	1.000	1.000	2.156	0.004	2.021	2.160	2.255
数据：干净															
稀疏回归（弹性网络）															
样本量：200	T1	0.874	-0.047	-0.027	0.878	0.049	0.030	0.960	1.000	1.000	2.129	0.003	2.046	2.128	2.224
n_{sim}：50	T2	0.931	-0.037	-0.058	0.935	0.039	0.060	0.920	1.000	1.000	2.194	0.003	2.120	2.196	2.282
迭代次数：5	T3	0.909	-0.035	-0.058	0.914	0.037	0.060	0.940	1.000	1.000	2.192	0.003	2.112	2.185	2.287
P_{ext}：0.03	T4	0.910	-0.034	-0.059	0.915	0.036	0.061	0.960	1.000	1.000	2.203	0.003	2.088	2.198	2.296
P_{miss}：0.05	T5	0.948	-0.035	-0.065	0.954	0.037	0.067	0.920	1.000	1.000	2.219	0.004	2.116	2.215	2.338
ρ：0.6	T6	0.931	-0.036	-0.059	0.935	0.038	0.061	0.960	1.000	1.000	2.202	0.004	2.077	2.198	2.296
数据：外部															
线性回归（普通最小二乘法）															
n：200	T1	-0.143	-0.018	0.065	0.224	0.022	0.068	1.000	1.000	1.000	2.010	0.006	1.833	2.022	2.142
n_{sim}：50	T2	0.019	-0.004	0.004	0.187	0.014	0.025	1.000	1.000	1.000	2.148	0.007	1.959	2.162	2.299
迭代次数：5	T3	0.013	-0.004	0.006	0.166	0.013	0.022	1.000	1.000	1.000	2.150	0.007	1.977	2.166	2.290
P_{ext}：0.03	T4	-0.016	-0.002	0.008	0.187	0.014	0.023	1.000	1.000	1.000	2.164	0.010	1.957	2.180	2.342
P_{miss}：0.10	T5	0.052	-0.003	-0.004	0.213	0.016	0.025	1.000	1.000	1.000	2.229	0.009	2.097	2.227	2.405
ρ：0.6	T6	-0.018	-0.002	0.008	0.191	0.014	0.025	1.000	1.000	1.000	2.153	0.009	1.953	2.159	2.353
数据：干净															
稀疏回归（弹性网络）															
n：200	T1	0.872	-0.064	0.007	0.884	0.066	0.019	0.840	1.000	1.000	2.054	0.009	1.886	2.056	2.249
n_{sim}：50	T2	0.981	-0.041	-0.060	0.992	0.043	0.063	0.740	1.000	1.000	2.193	0.011	1.968	2.219	2.377
迭代次数：5	T3	0.949	-0.039	-0.057	0.961	0.042	0.060	0.820	1.000	1.000	2.187	0.011	2.005	2.198	2.375
P_{ext}：0.03	T4	0.932	-0.036	-0.060	0.945	0.040	0.063	0.860	1.000	1.000	2.203	0.011	2.022	2.201	2.392
P_{miss}：0.10	T5	0.991	-0.037	-0.068	1.008	0.041	0.072	0.680	1.000	1.000	2.246	0.012	2.041	2.242	2.414
ρ：0.6	T6	0.955	-0.039	-0.058	0.968	0.043	0.062	0.760	1.000	1.000	2.197	0.013	2.025	2.203	2.389
数据：外部															
线性回归（普通最小二乘法）															
n：200	T1	-0.514	-0.045	0.194	0.628	0.055	0.198	0.840	0.940	0.100	1.707	0.015	1.515	1.702	1.987
n_{sim}：50	T2	-0.004	0.003	-0.001	0.336	0.030	0.038	1.000	1.000	1.000	2.170	0.020	1.954	2.185	2.462
迭代次数：5	T3	-0.009	0.002	0.002	0.372	0.031	0.044	1.000	1.000	1.000	2.162	0.021	1.961	2.140	2.491
P_{ext}：0.03	T4	-0.146	0.013	0.008	0.385	0.034	0.044	1.000	1.000	1.000	2.203	0.023	1.993	2.170	2.575
P_{miss}：0.30	T5	0.096	0.013	-0.043	0.391	0.038	0.063	1.000	1.000	1.000	2.454	0.021	2.240	2.437	2.759
ρ：0.6	T6	-0.133	0.011	0.008	0.388	0.034	0.047	1.000	1.000	1.000	2.189	0.024	1.993	2.164	2.523
数据：干净															
稀疏回归（弹性网络）															
n：200	T1	0.654	-0.119	0.151	0.735	0.123	0.155	0.800	0.440	0.700	1.761	0.022	1.513	1.763	2.056
n_{sim}：50	T2	0.962	-0.040	-0.062	1.012	0.050	0.075	0.800	1.000	1.000	2.217	0.025	1.903	2.216	2.526
迭代次数：5	T3	0.943	-0.038	-0.061	0.987	0.048	0.070	0.860	1.000	1.000	2.193	0.027	1.880	2.201	2.518
P_{ext}：0.03	T4	0.790	-0.021	-0.066	0.872	0.042	0.076	0.880	1.000	1.000	2.287	0.034	1.962	2.275	2.597
P_{miss}：0.30	T5	0.939	-0.017	-0.101	1.003	0.040	0.109	0.840	1.000	1.000	2.418	0.028	2.172	2.424	2.657
ρ：0.6	T6	0.910	-0.031	-0.067	0.980	0.047	0.078	0.840	1.000	1.000	2.251	0.034	1.922	2.265	2.540
数据：外部															



表21： 在普通最小二乘法（干净）和弹性网络（受污染）下，各多重插补方法的模型系数偏差、均方根误差、95%覆盖率以及样本外预测均方误差，其中 $\boldsymbol{n = 200}$、 $n_{\text{sim}} = 50$，迭代次数 $= 5$、 $\rho = 0.6$，污染 $P_{\text{ext}} \in \{0.05, 0.10\}$ 以及缺失 $P_{\text{miss}} \in \{0.05, 0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	-0.051	-0.011	0.034	0.141	0.015	0.036	1.000	1.000	1.000	2.088	0.004	1.951	2.103	2.182
	T2	0.020	-0.004	0.006	0.136	0.011	0.015	1.000	1.000	1.000	2.156	0.004	2.012	2.172	2.241
	T3	0.018	-0.003	0.004	0.143	0.012	0.015	1.000	1.000	1.000	2.159	0.005	2.014	2.168	2.269
	T4	0.007	-0.003	0.006	0.134	0.010	0.015	1.000	1.000	1.000	2.161	0.004	2.028	2.176	2.278
	T5	0.053	-0.005	0.000	0.160	0.014	0.014	1.000	1.000	1.000	2.198	0.006	2.075	2.204	2.365
	T6	0.009	-0.003	0.004	0.151	0.012	0.016	1.000	1.000	1.000	2.155	0.005	2.027	2.160	2.284
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	0.753	-0.030	-0.034	0.757	0.031	0.036	1.000	1.000	1.000	2.087	0.003	1.989	2.087	2.176
	T2	0.792	-0.018	-0.064	0.795	0.021	0.066	1.000	1.000	1.000	2.157	0.003	2.064	2.162	2.262
	T3	0.773	-0.016	-0.064	0.779	0.020	0.065	1.000	1.000	1.000	2.154	0.002	2.049	2.158	2.234
	T4	0.786	-0.017	-0.065	0.790	0.020	0.066	1.000	1.000	1.000	2.159	0.004	2.035	2.166	2.260
	T5	0.814	-0.017	-0.070	0.819	0.020	0.071	0.940	1.000	1.000	2.178	0.004	2.083	2.180	2.313
	T6	0.800	-0.018	-0.066	0.803	0.020	0.067	1.000	1.000	1.000	2.158	0.003	2.033	2.166	2.248
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.155	-0.018	0.067	0.239	0.022	0.071	1.000	1.000	1.000	2.005	0.006	1.832	2.021	2.119
	T2	0.005	-0.003	0.006	0.191	0.014	0.025	1.000	1.000	1.000	2.141	0.008	1.957	2.149	2.274
	T3	0.012	-0.004	0.007	0.187	0.015	0.025	1.000	1.000	1.000	2.148	0.008	1.937	2.159	2.280
	T4	-0.033	-0.001	0.010	0.207	0.015	0.028	1.000	1.000	1.000	2.159	0.009	1.955	2.176	2.312
	T5	0.053	-0.004	-0.002	0.202	0.016	0.025	1.000	1.000	1.000	2.231	0.011	2.005	2.220	2.412
	T6	-0.037	-0.001	0.010	0.208	0.014	0.027	1.000	1.000	1.000	2.144	0.009	1.940	2.152	2.306
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.755	-0.047	-0.000	0.764	0.049	0.017	0.960	1.000	1.000	2.014	0.009	1.839	2.017	2.208
	T2	0.830	-0.021	-0.065	0.838	0.025	0.068	0.960	1.000	1.000	2.152	0.011	1.933	2.172	2.336
	T3	0.800	-0.019	-0.062	0.810	0.024	0.065	0.960	1.000	1.000	2.146	0.011	1.963	2.155	2.339
	T4	0.810	-0.019	-0.064	0.819	0.024	0.067	0.980	1.000	1.000	2.165	0.010	1.962	2.170	2.317
	T5	0.867	-0.020	-0.074	0.879	0.025	0.076	0.860	1.000	1.000	2.214	0.010	2.018	2.187	2.412
	T6	0.837	-0.022	-0.063	0.845	0.027	0.066	0.920	1.000	1.000	2.149	0.011	1.984	2.164	2.321
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.518	-0.045	0.196	0.645	0.055	0.200	0.840	0.940	0.140	1.698	0.016	1.502	1.707	1.972
	T2	-0.015	0.003	0.001	0.370	0.030	0.043	1.000	1.000	1.000	2.155	0.022	1.920	2.177	2.462
	T3	-0.031	0.002	0.006	0.396	0.030	0.048	1.000	1.000	1.000	2.153	0.025	1.943	2.126	2.491
	T4	-0.159	0.013	0.011	0.418	0.033	0.045	1.000	1.000	1.000	2.188	0.024	1.949	2.165	2.554
	T5	0.103	0.010	-0.038	0.418	0.036	0.064	1.000	1.000	1.000	2.442	0.023	2.205	2.433	2.742
	T6	-0.152	0.011	0.015	0.415	0.033	0.051	1.000	1.000	1.000	2.173	0.027	1.962	2.153	2.486
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	0.604	-0.106	0.141	0.665	0.111	0.145	0.800	0.560	0.820	1.724	0.022	1.467	1.710	2.030
	T2	0.806	-0.019	-0.068	0.856	0.036	0.079	0.860	1.000	1.000	2.178	0.027	1.867	2.164	2.500
	T3	0.784	-0.017	-0.067	0.823	0.033	0.074	0.900	1.000	1.000	2.155	0.027	1.854	2.154	2.467
	T4	0.715	-0.008	-0.073	0.775	0.035	0.081	0.940	1.000	1.000	2.245	0.042	1.882	2.227	2.641
	T5	0.879	-0.004	-0.112	0.922	0.031	0.117	0.900	1.000	1.000	2.383	0.027	2.051	2.383	2.660
	T6	0.836	-0.021	-0.069	0.884	0.039	0.079	0.820	1.000	1.000	2.202	0.031	1.897	2.192	2.531
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	-0.057	-0.011	0.034	0.145	0.015	0.036	1.000	1.000	1.000	2.086	0.003	1.988	2.084	2.193
	T2	0.018	-0.004	0.006	0.143	0.012	0.016	1.000	1.000	1.000	2.156	0.004	2.064	2.150	2.271
	T3	0.011	-0.003	0.005	0.147	0.013	0.016	1.000	1.000	1.000	2.155	0.004	2.038	2.155	2.285
	T4	0.007	-0.003	0.006	0.144	0.011	0.016	1.000	1.000	1.000	2.157	0.003	2.063	2.153	2.262
	T5	0.052	-0.003	-0.003	0.178	0.014	0.017	1.000	1.000	1.000	2.204	0.006	2.089	2.196	2.376
	T6	-0.014	-0.001	0.006	0.134	0.011	0.015	1.000	1.000	1.000	2.155	0.004	2.054	2.144	2.296
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	0.663	-0.004	-0.074	0.666	0.010	0.075	0.980	1.000	1.000	1.936	0.002	1.850	1.937	2.011
	T2	0.670	0.011	-0.104	0.672	0.015	0.105	1.000	1.000	1.000	1.999	0.002	1.915	2.001	2.092
	T3	0.671	0.010	-0.103	0.674	0.014	0.104	1.000	1.000	1.000	1.996	0.002	1.907	1.995	2.075
	T4	0.685	0.010	-0.105	0.687	0.013	0.105	1.000	1.000	1.000	2.008	0.002	1.903	2.005	2.104
	T5	0.694	0.011	-0.108	0.697	0.015	0.109	0.960	1.000	1.000	2.014	0.003	1.925	2.012	2.103
	T6	0.682	0.011	-0.106	0.684	0.014	0.106	1.000	1.000	1.000	2.008	0.002	1.890	2.006	2.081

表22： 普通最小二乘法（干净）和弹性网络（受污染）在不同多重插补方法下的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。第一个区块： $\boldsymbol{n = 200}$, $n_{\text{sim}} = 50$, 迭代次数 $= 5$, $\rho = 0.6$, 取值为 $P_{\text{ext}} = 0.10$ 和 $P_{\text{miss}} \in \{0。10, 0。30\}$；随后是一个附加区块，其中 $\boldsymbol{n = 80}$, $n_{\text{sim}} = 300$, 迭代次数 $= 5$, $\rho = 0.6$, $P_{\text{ext}} = 0.04$ 取值为 $P_{\text{miss}} \in \{0。10, 0。25\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	-0.127	-0.018	0.062	0.202	0.022	0.065	1.000	1.000	1.000	2.008	0.006	1.840	2.028	2.153
	T2	0.022	-0.004	0.003	0.160	0.013	0.021	1.000	1.000	1.000	2.159	0.008	2.016	2.167	2.333
	T3	0.034	-0.004	0.003	0.175	0.014	0.021	1.000	1.000	1.000	2.146	0.008	1.976	2.157	2.285
	T4	0.010	-0.003	0.006	0.177	0.015	0.023	1.000	1.000	1.000	2.152	0.007	1.975	2.158	2.321
	T5	0.072	-0.004	-0.006	0.189	0.017	0.026	1.000	1.000	1.000	2.221	0.010	2.053	2.217	2.381
	T6	0.011	-0.002	0.004	0.174	0.014	0.023	1.000	1.000	1.000	2.150	0.009	1.989	2.155	2.340
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.676	-0.022	-0.041	0.684	0.026	0.044	0.840	1.000	1.000	1.871	0.007	1.677	1.878	2.015
	T2	0.701	0.007	-0.103	0.711	0.017	0.104	0.880	1.000	1.000	2.003	0.008	1.830	1.998	2.141
	T3	0.690	0.007	-0.100	0.698	0.017	0.102	0.920	1.000	1.000	1.991	0.009	1.802	2.006	2.153
	T4	0.715	0.005	-0.100	0.721	0.015	0.102	0.900	1.000	1.000	2.004	0.008	1.849	2.005	2.160
	T5	0.730	0.007	-0.108	0.738	0.018	0.110	0.880	1.000	1.000	2.052	0.008	1.871	2.048	2.201
	T6	0.707	0.006	-0.101	0.713	0.016	0.102	0.920	1.000	1.000	2.008	0.007	1.853	2.006	2.155
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.567	-0.041	0.196	0.693	0.052	0.201	0.780	0.960	0.180	1.683	0.017	1.464	1.679	1.959
	T2	-0.050	0.006	0.001	0.384	0.029	0.043	1.000	1.000	1.000	2.131	0.022	1.831	2.147	2.436
	T3	-0.104	0.007	0.009	0.405	0.029	0.053	1.000	1.000	1.000	2.132	0.025	1.878	2.113	2.480
	T4	-0.143	0.013	0.007	0.442	0.033	0.045	1.000	1.000	1.000	2.174	0.023	1.940	2.143	2.468
	T5	0.119	0.008	-0.036	0.431	0.034	0.064	1.000	1.000	1.000	2.433	0.033	2.149	2.427	2.797
	T6	-0.175	0.015	0.008	0.448	0.036	0.048	1.000	1.000	1.000	2.152	0.026	1.878	2.171	2.457
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部。	T1	0.637	-0.088	0.092	0.665	0.092	0.098	0.640	0.720	0.980	1.602	0.019	1.365	1.604	1.844
	T2	0.655	0.013	-0.110	0.685	0.030	0.117	0.900	1.000	0.960	2.027	0.024	1.714	2.022	2.313
	T3	0.667	0.011	-0.107	0.693	0.027	0.112	0.900	1.000	1.000	2.001	0.025	1.752	1.981	2.314
	T4	0.751	0.002	-0.106	0.771	0.024	0.111	0.860	1.000	0.980	2.052	0.027	1.759	2.059	2.335
	T5	0.764	0.018	-0.141	0.787	0.032	0.146	0.800	1.000	0.940	2.190	0.022	1.963	2.180	2.465
	T6	0.731	0.006	-0.109	0.751	0.030	0.116	0.840	1.000	0.980	2.059	0.030	1.801	2.066	2.374
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.200	0.021	-0.125	0.305	0.030	0.131	1.000	1.000	1.000	2.438	0.018	2.188	2.439	2.725
	T2	0.421	0.034	-0.199	0.484	0.041	0.203	1.000	1.000	1.000	2.580	0.020	2.321	2.578	2.869
	T3	0.406	0.033	-0.193	0.472	0.039	0.197	1.000	1.000	1.000	2.565	0.021	2.286	2.558	2.862
	T4	0.407	0.033	-0.193	0.463	0.039	0.198	1.000	1.000	1.000	2.584	0.018	2.353	2.577	2.871
	T5	0.475	0.035	-0.212	0.533	0.041	0.217	1.000	1.000	1.000	2.666	0.025	2.367	2.656	3.024
	T6	0.414	0.033	-0.195	0.465	0.039	0.201	1.000	1.000	1.000	2.575	0.018	2.350	2.574	2.877
稀疏回归（弹性网络）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	-0.285	0.113	-0.173	0.346	0.115	0.174	1.000	1.000	0.997	1.790	0.012	1.549	1.791	1.992
	T2	-0.244	0.136	-0.227	0.319	0.139	0.228	1.000	0.983	0.610	1.926	0.016	1.680	1.925	2.194
	T3	-0.266	0.136	-0.223	0.336	0.138	0.224	1.000	0.990	0.643	1.918	0.014	1.668	1.921	2.144
	T4	-0.352	0.149	-0.231	0.437	0.152	0.232	1.000	0.933	0.563	1.992	0.021	1.730	1.979	2.279
	T5	-0.301	0.155	-0.252	0.416	0.158	0.254	1.000	0.923	0.553	2.100	0.027	1.818	2.096	2.418
	T6	-0.287	0.141	-0.228	0.355	0.143	0.229	1.000	0.967	0.567	1.945	0.015	1.691	1.946	2.203
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.25 ρ: 0.6 数据：干净	T1	-0.192	0.001	-0.000	0.492	0.045	0.077	0.997	0.993	1.000	2.189	0.038	1.801	2.191	2.554
	T2	0.405	0.033	-0.193	0.607	0.056	0.209	1.000	1.000	0.977	2.562	0.049	2.130	2.557	2.983
	T3	0.359	0.033	-0.183	0.573	0.054	0.198	1.000	0.997	0.977	2.528	0.047	2.103	2.521	2.987
	T4	0.369	0.036	-0.193	0.527	0.051	0.209	1.000	1.000	0.950	2.600	0.049	2.225	2.582	3.031
	T5	0.568	0.038	-0.238	0.706	0.054	0.253	1.000	1.000	0.907	2.811	0.049	2.368	2.820	3.268
	T6	0.381	0.035	-0.192	0.548	0.051	0.209	1.000	1.000	0.947	2.574	0.043	2.193	2.568	2.979
稀疏回归（弹性网络）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.04 P_{miss}: 0.25 ρ: 0.6 数据：外部	T1	-0.340	0.069	-0.077	0.516	0.083	0.090	0.990	0.977	0.997	1.578	0.026	1.241	1.586	1.869
	T2	-0.251	0.134	-0.221	0.479	0.142	0.226	0.997	0.910	0.693	1.948	0.047	1.505	1.953	2.339
	T3	-0.290	0.134	-0.214	0.470	0.141	0.219	0.997	0.920	0.747	1.912	0.038	1.521	1.921	2.267
	T4	-0.570	0.171	-0.234	0.743	0.180	0.240	0.997	0.733	0.663	2.122	0.066	1.644	2.108	2.678
	T5	-0.400	0.188	-0.299	0.635	0.197	0.305	1.000	0.800	0.597	2.427	0.086	1.922	2.412	3.031
	T6	-0.403	0.152	-0.227	0.581	0.160	0.233	1.000	0.903	0.720	2.008	0.053	1.600	2.000	2.498



表23： OLS（干净）和EN（受污染）中，不同多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。上部区块： $\boldsymbol{n = 80}$、 $n_{\text{sim}} = 300$，迭代次数为 $= 5$、 $\rho = 0.6$，其中 $P_{\text{ext}} = 0.15$ 和 $P_{\text{miss}} \in \{0.10, 0.25\}$；随后是附加区块，其中 $\boldsymbol{n = 80}$、 $n_{\text{sim}} = 50$，迭代次数为 $= 10$、 $\rho = 0.6$， $P_{\text{ext}} = 0.05$ 和 $P_{\text{miss}} \in \{0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.200	0.021	-0.126	0.306	0.030	0.131	1.000	1.000	1.000	2.442	0.019	2.174	2.436	2.752
	T2	0.425	0.034	-0.199	0.487	0.040	0.203	1.000	1.000	1.000	2.578	0.022	2.293	2.570	2.910
	T3	0.405	0.033	-0.193	0.472	0.040	0.198	1.000	1.000	1.000	2.568	0.021	2.271	2.566	2.871
	T4	0.404	0.034	-0.195	0.460	0.040	0.200	1.000	1.000	1.000	2.585	0.019	2.325	2.584	2.874
	T5	0.475	0.035	-0.213	0.531	0.042	0.218	1.000	1.000	0.987	2.673	0.027	2.406	2.660	3.022
	T6	0.418	0.033	-0.196	0.477	0.039	0.201	1.000	1.000	0.997	2.580	0.020	2.332	2.567	2.893
稀疏回归（弹性网络）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	-0.031	0.086	-0.164	0.143	0.089	0.166	1.000	1.000	0.960	1.566	0.009	1.366	1.575	1.739
	T2	-0.067	0.113	-0.210	0.170	0.116	0.212	1.000	0.997	0.547	1.699	0.013	1.477	1.700	1.903
	T3	-0.081	0.114	-0.209	0.175	0.116	0.211	1.000	0.997	0.537	1.680	0.011	1.439	1.684	1.867
	T4	-0.055	0.114	-0.214	0.140	0.116	0.216	1.000	1.000	0.467	1.712	0.012	1.502	1.709	1.942
	T5	-0.074	0.122	-0.227	0.160	0.125	0.229	1.000	0.997	0.547	1.809	0.022	1.559	1.786	2.150
	T6	-0.049	0.112	-0.210	0.129	0.114	0.212	1.000	1.000	0.520	1.694	0.011	1.507	1.692	1.901
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.25 ρ: 0.6 数据：干净	T1	-0.183	0.000	-0.002	0.485	0.046	0.075	0.997	0.993	1.000	2.187	0.041	1.805	2.173	2.572
	T2	0.405	0.034	-0.196	0.608	0.058	0.210	1.000	1.000	0.960	2.559	0.049	2.130	2.557	2.983
	T3	0.353	0.034	-0.184	0.581	0.057	0.200	1.000	0.997	0.963	2.527	0.049	2.077	2.522	2.939
	T4	0.389	0.034	-0.194	0.556	0.052	0.210	1.000	1.000	0.937	2.598	0.056	2.146	2.595	3.087
	T5	0.564	0.039	-0.238	0.705	0.056	0.252	1.000	1.000	0.930	2.815	0.052	2.391	2.807	3.284
	T6	0.411	0.032	-0.193	0.565	0.051	0.210	1.000	1.000	0.953	2.570	0.047	2.156	2.568	2.971
稀疏回归（弹性网络）															
n: 80 n_{sim}: 300 迭代次数: 5 P_{ext}: 0.15 P_{miss}: 0.25 ρ: 0.6 数据：外部	T1	0.039	0.037	-0.082	0.261	0.057	0.095	1.000	1.000	0.993	1.369	0.023	1.062	1.381	1.650
	T2	-0.052	0.108	-0.205	0.272	0.117	0.211	1.000	0.950	0.673	1.726	0.040	1.313	1.751	2.081
	T3	-0.098	0.111	-0.201	0.293	0.120	0.207	1.000	0.957	0.680	1.679	0.035	1.309	1.689	2.040
	T4	-0.114	0.124	-0.223	0.309	0.132	0.229	0.997	0.937	0.513	1.772	0.044	1.378	1.767	2.160
	T5	-0.141	0.148	-0.266	0.324	0.157	0.272	1.000	0.920	0.633	2.080	0.058	1.628	2.075	2.590
	T6	-0.016	0.108	-0.210	0.233	0.116	0.216	1.000	0.967	0.607	1.716	0.036	1.327	1.732	2.079
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.159	0.022	-0.118	0.309	0.033	0.125	1.000	1.000	1.000	2.416	0.012	2.191	2.417	2.565
	T2	0.381	0.034	-0.190	0.456	0.041	0.195	1.000	1.000	1.000	2.556	0.015	2.334	2.551	2.752
	T3	0.382	0.032	-0.186	0.466	0.041	0.190	1.000	1.000	1.000	2.556	0.014	2.339	2.548	2.736
	T4	0.392	0.032	-0.188	0.467	0.041	0.192	1.000	1.000	1.000	2.561	0.010	2.390	2.579	2.730
	T5	0.448	0.034	-0.204	0.515	0.042	0.209	1.000	1.000	1.000	2.632	0.017	2.397	2.627	2.914
	T6	0.402	0.032	-0.188	0.485	0.041	0.194	1.000	1.000	1.000	2.536	0.017	2.332	2.552	2.772
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	-0.181	0.101	-0.171	0.263	0.104	0.172	1.000	1.000	1.000	1.798	0.010	1.581	1.795	1.985
	T2	-0.161	0.127	-0.226	0.263	0.130	0.227	1.000	1.000	0.640	1.944	0.015	1.657	1.949	2.150
	T3	-0.173	0.127	-0.223	0.250	0.129	0.225	1.000	1.000	0.620	1.922	0.013	1.692	1.943	2.087
	T4	-0.251	0.138	-0.231	0.353	0.142	0.232	1.000	0.980	0.480	1.993	0.018	1.772	1.977	2.226
	T5	-0.192	0.143	-0.252	0.288	0.147	0.254	1.000	0.980	0.460	2.106	0.032	1.782	2.092	2.471
	T6	-0.179	0.131	-0.230	0.256	0.134	0.232	1.000	1.000	0.480	1.956	0.011	1.764	1.934	2.157
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.359	-0.002	0.040	0.650	0.045	0.113	1.000	1.000	0.980	2.093	0.045	1.653	2.103	2.461
	T2	0.417	0.039	-0.207	0.649	0.057	0.230	1.000	1.000	0.900	2.552	0.050	2.080	2.537	2.932
	T3	0.322	0.038	-0.184	0.584	0.057	0.208	1.000	1.000	0.940	2.499	0.052	2.035	2.519	2.886
	T4	0.366	0.039	-0.199	0.609	0.056	0.222	1.000	1.000	0.900	2.626	0.058	2.125	2.655	2.992
	T5	0.613	0.044	-0.259	0.796	0.057	0.278	1.000	1.000	0.840	2.899	0.053	2.417	2.927	3.273
	T6	0.410	0.034	-0.194	0.610	0.050	0.218	1.000	1.000	0.960	2.568	0.059	2.036	2.576	2.980
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	-0.328	0.052	-0.044	0.563	0.075	0.069	0.980	0.980	1.000	1.533	0.032	1.236	1.535	1.853
	T2	-0.232	0.133	-0.221	0.519	0.143	0.228	1.000	0.900	0.800	2.002	0.046	1.640	1.981	2.362
	T3	-0.304	0.137	-0.214	0.513	0.144	0.219	1.000	0.900	0.800	1.941	0.043	1.609	1.948	2.308
	T4	-0.703	0.189	-0.243	0.931	0.199	0.249	0.920	0.680	0.720	2.206	0.073	1.797	2.187	2.817
	T5	-0.472	0.212	-0.331	0.772	0.224	0.337	1.000	0.680	0.380	2.576	0.088	2.010	2.563	3.229
	T6	-0.481	0.164	-0.233	0.734	0.175	0.240	0.960	0.820	0.740	2.071	0.081	1.681	2.016	2.623

表24： 各多重插补方法下，普通最小二乘法（干净）与弹性网络（受污染）的模型系数偏差、均方根误差、95%覆盖率及样本外预测均方误差。上部区块： $\boldsymbol{n = 80}$、 $n_{\text{sim}} = 50$，迭代次数 $= 10$、 $\rho = 0.6$，以及 $P_{\text{ext}} = 0.10$ 和 $P_{\text{miss}} \in \{0.10, 0.30\}$；随后是一个附加区块，含 $\boldsymbol{n = 200}$、 $n_{\text{sim}} = 50$，迭代次数 $= 5$、 $\rho = 0$、 $P_{\text{ext}} = 0.05$ 和 $P_{\text{miss}} \in \{0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.147	0.022	-0.115	0.301	0.036	0.125	1.000	1.000	1.000	2.397	0.017	2.171	2.403	2.667
	T2	0.366	0.035	-0.188	0.452	0.045	0.194	1.000	1.000	1.000	2.536	0.023	2.247	2.549	2.807
	T3	0.353	0.034	-0.184	0.436	0.044	0.190	1.000	1.000	1.000	2.533	0.020	2.294	2.522	2.839
	T4	0.359	0.033	-0.183	0.442	0.043	0.190	1.000	1.000	1.000	2.534	0.018	2.318	2.522	2.805
	T5	0.426	0.033	-0.196	0.517	0.043	0.204	1.000	1.000	1.000	2.611	0.021	2.389	2.597	2.888
	T6	0.399	0.032	-0.187	0.461	0.041	0.194	1.000	1.000	1.000	2.531	0.019	2.311	2.517	2.799
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	-0.183	0.099	-0.166	0.248	0.102	0.168	1.000	1.000	1.000	1.698	0.013	1.455	1.687	1.901
	T2	-0.203	0.127	-0.217	0.266	0.129	0.219	1.000	0.980	0.540	1.837	0.017	1.517	1.832	2.049
	T3	-0.214	0.126	-0.215	0.278	0.129	0.216	1.000	0.980	0.620	1.824	0.015	1.615	1.825	2.050
	T4	-0.238	0.132	-0.220	0.297	0.135	0.222	1.000	0.940	0.400	1.852	0.016	1.656	1.824	2.111
	T5	-0.223	0.141	-0.241	0.277	0.143	0.243	1.000	1.000	0.480	1.985	0.024	1.722	1.960	2.297
	T6	-0.171	0.125	-0.218	0.228	0.127	0.220	1.000	1.000	0.460	1.842	0.012	1.644	1.836	2.035
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.368	-0.004	0.047	0.672	0.046	0.101	0.980	1.000	1.000	2.112	0.036	1.694	2.126	2.410
	T2	0.403	0.039	-0.202	0.645	0.057	0.219	1.000	1.000	0.960	2.600	0.053	2.220	2.596	3.013
	T3	0.311	0.036	-0.176	0.612	0.056	0.197	1.000	1.000	0.920	2.532	0.045	2.176	2.529	2.956
	T4	0.354	0.039	-0.194	0.565	0.055	0.215	1.000	1.000	0.940	2.637	0.065	2.105	2.646	3.053
	T5	0.632	0.040	-0.255	0.779	0.053	0.273	1.000	1.000	0.880	2.905	0.056	2.451	2.914	3.310
	T6	0.368	0.037	-0.193	0.592	0.055	0.216	1.000	1.000	0.920	2.600	0.048	2.134	2.669	2.953
稀疏回归（弹性网络）															
n: 80 n_{sim}: 50 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	-0.193	0.040	-0.047	0.391	0.064	0.071	0.980	0.980	1.000	1.452	0.030	1.154	1.443	1.791
	T2	-0.200	0.127	-0.213	0.380	0.135	0.220	1.000	0.940	0.740	1.918	0.043	1.531	1.929	2.246
	T3	-0.283	0.132	-0.208	0.421	0.139	0.213	1.000	0.900	0.740	1.851	0.042	1.507	1.862	2.252
	T4	-0.580	0.176	-0.240	0.766	0.186	0.246	0.960	0.780	0.540	2.083	0.063	1.653	2.065	2.577
	T5	-0.428	0.199	-0.312	0.599	0.206	0.317	0.980	0.840	0.500	2.435	0.063	1.965	2.429	2.883
	T6	-0.293	0.140	-0.220	0.503	0.149	0.227	0.960	0.920	0.740	1.924	0.049	1.569	1.919	2.362
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.0 数据：干净	T1	-0.268	0.003	0.049	0.350	0.016	0.053	1.000	1.000	1.000	1.983	0.007	1.793	2.000	2.099
	T2	-0.007	-0.001	0.004	0.238	0.017	0.021	1.000	1.000	1.000	2.141	0.008	1.938	2.144	2.280
	T3	-0.005	-0.002	0.006	0.229	0.017	0.021	1.000	1.000	1.000	2.147	0.009	1.922	2.158	2.269
	T4	-0.033	-0.000	0.009	0.261	0.019	0.024	1.000	1.000	1.000	2.142	0.008	1.980	2.144	2.293
	T5	0.120	-0.008	-0.007	0.295	0.021	0.023	1.000	1.000	1.000	2.284	0.016	2.027	2.302	2.470
	T6	-0.047	0.001	0.007	0.272	0.018	0.025	1.000	1.000	1.000	2.148	0.009	1.927	2.158	2.291
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.0 数据：外部	T1	1.478	-0.090	-0.057	1.485	0.091	0.058	0.000	0.620	1.000	2.000	0.010	1.824	2.003	2.188
	T2	1.587	-0.079	-0.099	1.595	0.081	0.100	0.000	0.940	1.000	2.163	0.011	1.949	2.175	2.368
	T3	1.577	-0.078	-0.099	1.585	0.080	0.099	0.000	0.940	1.000	2.160	0.011	1.963	2.159	2.361
	T4	1.664	-0.088	-0.096	1.670	0.089	0.097	0.000	0.880	0.980	2.175	0.011	1.982	2.180	2.365
	T5	1.671	-0.082	-0.107	1.683	0.084	0.108	0.000	0.940	0.960	2.252	0.013	1.978	2.254	2.431
	T6	1.698	-0.090	-0.100	1.701	0.090	0.101	0.000	0.860	0.980	2.170	0.012	1.965	2.175	2.383
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.0 数据：干净	T1	-0.819	0.013	0.140	0.935	0.037	0.145	0.740	1.000	0.240	1.642	0.015	1.467	1.650	1.902
	T2	-0.019	0.004	0.001	0.453	0.035	0.036	1.000	1.000	1.000	2.155	0.023	1.928	2.169	2.473
	T3	-0.045	0.004	0.004	0.487	0.036	0.039	1.000	1.000	1.000	2.153	0.027	1.931	2.120	2.496
	T4	-0.066	0.004	0.009	0.493	0.036	0.042	1.000	1.000	1.000	2.186	0.029	1.935	2.184	2.549
	T5	0.473	-0.020	-0.051	0.656	0.040	0.064	1.000	1.000	1.000	2.648	0.042	2.286	2.610	3.009
	T6	-0.050	0.005	0.005	0.426	0.034	0.037	1.000	1.000	1.000	2.156	0.032	1.897	2.176	2.592
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.0 数据：外部。	T1	1.219	-0.113	0.034	1.265	0.117	0.046	0.200	0.220	1.000	1.660	0.020	1.414	1.648	1.982
	T2	1.572	-0.078	-0.103	1.603	0.084	0.107	0.220	0.940	0.880	2.174	0.034	1.856	2.140	2.543
	T3	1.558	-0.077	-0.099	1.591	0.083	0.104	0.300	0.840	0.960	2.172	0.029	1.867	2.157	2.556
	T4	1.638	-0.090	-0.091	1.668	0.095	0.096	0.100	0.820	0.880	2.233	0.043	1.875	2.229	2.613
	T5	1.835	-0.088	-0.134	1.871	0.095	0.137	0.140	0.840	0.760	2.506	0.049	2.154	2.504	2.858
	T6	1.843	-0.106	-0.098	1.865	0.111	0.102	0.080	0.600	0.860	2.194	0.034	1.848	2.205	2.547



表25： 普通最小二乘法（干净）和弹性网络（受污染）在不同多重插补方法下的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。上部区块： $\boldsymbol{n = 200}$, $n_{\text{sim}} = 50$, 迭代次数为 $= 5$, $\rho = 0$, 其中 $P_{\text{ext}} = 0.10$ 和 $P_{\text{miss}} \in \{0.10,0.30\}$；随后是附加区块，其中 $\boldsymbol{n = 80}$, $n_{\text{sim}} = 1000$, 迭代次数为 $= 10$, $\rho = 0.6$, $P_{\text{ext}} = 0.30$ 和 $P_{\text{miss}} \in \{0.10,0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.0 数据：干净	T1	-0.230	0.000	0.045	0.298	0.015	0.048	1.000	1.000	1.000	1.986	0.006	1.816	2.003	2.127
	T2	0.009	-0.002	0.003	0.199	0.015	0.018	1.000	1.000	1.000	2.158	0.008	2.012	2.175	2.326
	T3	0.027	-0.003	0.002	0.209	0.016	0.017	1.000	1.000	1.000	2.146	0.008	1.971	2.161	2.287
	T4	0.002	-0.001	0.002	0.229	0.018	0.019	1.000	1.000	1.000	2.156	0.010	1.946	2.159	2.339
	T5	0.144	-0.009	-0.010	0.295	0.022	0.023	1.000	1.000	1.000	2.274	0.012	2.118	2.255	2.500
	T6	-0.006	-0.001	0.003	0.218	0.017	0.017	1.000	1.000	1.000	2.150	0.009	1.947	2.150	2.312
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.0 数据：外部	T1	1.683	-0.089	-0.105	1.688	0.091	0.106	0.000	0.560	0.720	1.856	0.008	1.659	1.867	2.019
	T2	1.738	-0.074	-0.144	1.743	0.076	0.145	0.000	0.940	0.100	2.006	0.011	1.758	2.005	2.191
	T3	1.735	-0.075	-0.143	1.740	0.076	0.144	0.000	0.920	0.100	2.006	0.009	1.807	2.014	2.179
	T4	1.796	-0.082	-0.140	1.799	0.083	0.141	0.000	0.840	0.080	2.018	0.009	1.831	2.020	2.163
	T5	1.820	-0.079	-0.151	1.827	0.080	0.152	0.000	0.880	0.040	2.094	0.014	1.854	2.085	2.333
	T6	1.830	-0.085	-0.140	1.833	0.087	0.141	0.000	0.800	0.100	2.028	0.010	1.843	2.041	2.189
线性回归（普通最小二乘法）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.0 数据: 干净	T1	-0.872	0.018	0.140	0.990	0.040	0.145	0.680	1.000	0.280	1.628	0.016	1.424	1.624	1.888
	T2	-0.053	0.007	-0.000	0.472	0.035	0.035	1.000	1.000	1.000	2.134	0.024	1.820	2.140	2.469
	T3	-0.131	0.011	0.007	0.505	0.035	0.043	1.000	1.000	1.000	2.132	0.027	1.881	2.092	2.498
	T4	-0.080	0.006	0.007	0.504	0.036	0.040	1.000	1.000	1.000	2.168	0.027	1.906	2.182	2.489
	T5	0.395	-0.015	-0.047	0.643	0.042	0.063	1.000	1.000	1.000	2.624	0.035	2.255	2.621	2.980
	T6	-0.112	0.010	0.007	0.484	0.038	0.037	1.000	1.000	1.000	2.125	0.028	1.859	2.102	2.406
稀疏回归（弹性网络）															
n: 200 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.0 数据：外部	T1	1.572	-0.122	-0.023	1.586	0.125	0.037	0.000	0.000	1.000	1.536	0.017	1.306	1.531	1.765
	T2	1.705	-0.071	-0.148	1.720	0.076	0.152	0.000	0.960	0.380	2.017	0.027	1.773	1.994	2.352
	T3	1.717	-0.073	-0.146	1.733	0.077	0.148	0.020	0.920	0.420	2.014	0.027	1.730	1.988	2.374
	T4	1.911	-0.100	-0.133	1.918	0.102	0.135	0.000	0.620	0.420	2.047	0.024	1.806	2.034	2.323
	T5	1.948	-0.082	-0.175	1.960	0.086	0.177	0.000	0.860	0.180	2.311	0.028	2.039	2.328	2.604
	T6	1.976	-0.104	-0.136	1.983	0.107	0.139	0.000	0.500	0.460	2.072	0.032	1.747	2.065	2.432
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.30 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.185	0.022	-0.125	0.301	0.032	0.131	1.000	1.000	1.000	2.435	0.019	2.172	2.431	2.713
	T2	0.403	0.035	-0.198	0.468	0.042	0.202	1.000	1.000	1.000	2.572	0.020	2.300	2.569	2.842
	T3	0.391	0.034	-0.193	0.460	0.041	0.198	1.000	1.000	1.000	2.562	0.019	2.289	2.560	2.828
	T4	0.393	0.034	-0.194	0.447	0.040	0.198	1.000	1.000	1.000	2.575	0.019	2.331	2.569	2.846
	T5	0.460	0.036	-0.211	0.515	0.042	0.216	1.000	1.000	0.999	2.657	0.022	2.400	2.653	2.953
	T6	0.406	0.034	-0.197	0.462	0.040	0.202	1.000	1.000	1.000	2.568	0.019	2.321	2.565	2.856
稀疏回归（弹性网络）															
n: 80 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.30 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.047	0.083	-0.181	0.114	0.086	0.183	1.000	0.998	0.476	1.141	0.006	0.974	1.140	1.296
	T2	-0.025	0.110	-0.220	0.118	0.112	0.222	1.000	0.991	0.228	1.249	0.008	1.071	1.250	1.421
	T3	-0.032	0.111	-0.220	0.113	0.113	0.222	1.000	0.991	0.188	1.231	0.007	1.063	1.234	1.389
	T4	-0.015	0.111	-0.225	0.100	0.114	0.226	1.000	0.993	0.170	1.246	0.008	1.070	1.247	1.430
	T5	-0.062	0.125	-0.242	0.130	0.127	0.244	1.000	0.987	0.229	1.363	0.017	1.143	1.343	1.681
	T6	-0.010	0.109	-0.221	0.106	0.112	0.223	1.000	0.988	0.191	1.240	0.009	1.056	1.238	1.431
线性回归（普通最小二乘法）															
n: 80 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.304	-0.007	0.037	0.611	0.050	0.088	0.996	0.997	1.000	2.115	0.047	1.698	2.114	2.525
	T2	0.444	0.034	-0.206	0.670	0.058	0.220	1.000	1.000	0.949	2.584	0.056	2.125	2.583	3.021
	T3	0.377	0.034	-0.191	0.614	0.057	0.206	1.000	1.000	0.962	2.539	0.053	2.130	2.541	2.984
	T4	0.399	0.035	-0.201	0.588	0.054	0.218	1.000	1.000	0.931	2.617	0.060	2.154	2.617	3.155
	T5	0.630	0.040	-0.257	0.773	0.057	0.271	1.000	1.000	0.876	2.872	0.055	2.437	2.864	3.371
	T6	0.422	0.033	-0.200	0.615	0.052	0.217	1.000	1.000	0.930	2.581	0.049	2.154	2.584	3.027
稀疏回归（弹性网络）															
n: 80 n_{sim}: 1000 迭代次数: 10 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	-0.137	0.039	-0.056	0.384	0.063	0.075	0.994	0.992	0.998	1.414	0.030	1.082	1.418	1.750
	T2	-0.180	0.125	-0.216	0.394	0.134	0.222	0.999	0.921	0.726	1.871	0.049	1.426	1.876	2.279
	T3	-0.240	0.128	-0.211	0.415	0.137	0.217	0.999	0.915	0.710	1.810	0.045	1.388	1.811	2.248
	T4	-0.495	0.170	-0.245	0.645	0.179	0.250	0.989	0.736	0.495	2.021	0.061	1.565	2.026	2.502
	T5	-0.355	0.193	-0.315	0.527	0.201	0.320	0.994	0.831	0.479	2.399	0.076	1.918	2.375	2.996
	T6	-0.213	0.134	-0.226	0.365	0.142	0.232	1.000	0.941	0.607	1.884	0.049	1.450	1.872	2.304



表26： 普通最小二乘法（干净）和弹性网络（弹性网络，受污染）在不同多重插补方法下的模型系数偏差、均方根误差、95%覆盖率及样本外预测均方误差； $\boldsymbol{n = 500}$, $n_{\text{sim}} = 50$，迭代次数 $= 5$,$\rho = 0.6$，污染程度 $P_{\text{ext}} \in \{0.03, 0.05\}$ 和缺失率 $P_{\text{miss}} \in \{0.05, 0.10, 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
样本量: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	0.117	0.005	-0.023	0.130	0.008	0.025	1.000	1.000	1.000	2.152	0.001	2.110	2.147	2.203
	T2	0.176	0.016	-0.057	0.185	0.018	0.058	1.000	1.000	1.000	2.214	0.001	2.157	2.216	2.278
	T3	0.167	0.017	-0.056	0.179	0.018	0.057	1.000	1.000	1.000	2.213	0.001	2.162	2.208	2.289
	T4	0.167	0.017	-0.056	0.179	0.018	0.057	1.000	1.000	1.000	2.215	0.001	2.158	2.211	2.282
	T5	0.177	0.017	-0.059	0.189	0.019	0.060	1.000	1.000	1.000	2.224	0.001	2.162	2.221	2.282
	T6	0.176	0.017	-0.057	0.187	0.018	0.058	1.000	1.000	1.000	2.215	0.001	2.154	2.213	2.279
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	0.165	-0.077	0.115	0.172	0.077	0.116	1.000	0.060	0.200	2.080	0.001	2.018	2.083	2.133
	T2	0.203	-0.066	0.086	0.210	0.066	0.087	1.000	0.860	0.980	2.154	0.001	2.073	2.158	2.211
	T3	0.194	-0.065	0.086	0.200	0.065	0.087	1.000	0.860	0.960	2.151	0.001	2.081	2.149	2.205
	T4	0.205	-0.065	0.085	0.210	0.065	0.085	1.000	0.900	1.000	2.155	0.001	2.083	2.160	2.225
	T5	0.204	-0.063	0.082	0.212	0.064	0.082	1.000	0.900	0.980	2.179	0.001	2.118	2.184	2.242
	T6	0.202	-0.065	0.085	0.207	0.066	0.086	1.000	0.720	0.940	2.150	0.002	2.084	2.156	2.220
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.058	-0.007	0.013	0.114	0.012	0.019	1.000	1.000	1.000	2.082	0.002	1.985	2.082	2.163
	T2	0.165	0.017	-0.055	0.197	0.019	0.057	1.000	1.000	1.000	2.213	0.003	2.127	2.213	2.300
	T3	0.162	0.017	-0.054	0.191	0.020	0.057	1.000	1.000	1.000	2.213	0.003	2.101	2.213	2.292
	T4	0.161	0.016	-0.053	0.195	0.020	0.055	1.000	1.000	1.000	2.211	0.003	2.107	2.204	2.311
	T5	0.185	0.019	-0.062	0.213	0.021	0.064	1.000	1.000	1.000	2.245	0.003	2.165	2.238	2.322
	T6	0.164	0.017	-0.055	0.195	0.020	0.057	1.000	1.000	1.000	2.209	0.003	2.103	2.213	2.294
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.140	-0.091	0.150	0.169	0.092	0.150	1.000	0.040	0.000	2.000	0.003	1.903	1.995	2.088
	T2	0.215	-0.067	0.088	0.237	0.068	0.089	1.000	0.740	0.860	2.151	0.003	2.048	2.147	2.243
	T3	0.205	-0.067	0.089	0.230	0.068	0.091	1.000	0.680	0.880	2.144	0.003	2.049	2.149	2.231
	T4	0.205	-0.067	0.090	0.222	0.068	0.091	1.000	0.660	0.840	2.146	0.004	2.051	2.140	2.284
	T5	0.214	-0.064	0.081	0.236	0.065	0.082	1.000	0.840	0.940	2.203	0.004	2.065	2.213	2.294
	T6	0.207	-0.067	0.089	0.226	0.068	0.091	1.000	0.700	0.880	2.146	0.004	2.039	2.152	2.256
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.186	-0.063	0.170	0.269	0.066	0.172	0.980	0.620	0.000	1.786	0.006	1.664	1.791	1.929
	T2	0.206	0.012	-0.055	0.279	0.025	0.062	1.000	1.000	1.000	2.216	0.008	2.051	2.220	2.372
	T3	0.192	0.014	-0.055	0.266	0.026	0.062	1.000	1.000	1.000	2.208	0.008	2.066	2.210	2.391
	T4	0.158	0.017	-0.055	0.238	0.027	0.060	1.000	1.000	1.000	2.221	0.007	2.076	2.226	2.375
	T5	0.225	0.025	-0.083	0.298	0.033	0.088	1.000	0.980	0.980	2.329	0.007	2.177	2.314	2.476
	T6	0.164	0.018	-0.057	0.249	0.031	0.066	1.000	1.000	1.000	2.221	0.010	2.013	2.232	2.408
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.03 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	0.007	-0.144	0.280	0.150	0.145	0.281	1.000	0.000	0.000	1.694	0.006	1.534	1.694	1.862
	T2	0.262	-0.071	0.085	0.307	0.073	0.090	1.000	0.700	0.880	2.172	0.009	1.964	2.165	2.359
	T3	0.223	-0.066	0.082	0.268	0.068	0.087	1.000	0.840	0.920	2.153	0.009	1.972	2.143	2.351
	T4	0.206	-0.066	0.086	0.236	0.067	0.089	1.000	0.740	0.820	2.160	0.010	1.968	2.151	2.325
	T5	0.292	-0.055	0.049	0.324	0.058	0.057	1.000	0.940	1.000	2.354	0.015	2.115	2.340	2.544
	T6	0.232	-0.067	0.084	0.274	0.069	0.088	0.980	0.820	0.840	2.147	0.009	1.971	2.149	2.305
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	0.112	0.006	-0.023	0.126	0.008	0.025	1.000	1.000	1.000	2.152	0.001	2.106	2.153	2.203
	T2	0.169	0.018	-0.058	0.180	0.019	0.058	1.000	1.000	1.000	2.218	0.001	2.160	2.214	2.287
	T3	0.172	0.017	-0.056	0.181	0.018	0.057	1.000	1.000	1.000	2.213	0.001	2.157	2.207	2.265
	T4	0.170	0.017	-0.056	0.180	0.018	0.057	1.000	1.000	1.000	2.216	0.001	2.158	2.221	2.270
	T5	0.175	0.018	-0.060	0.186	0.019	0.061	1.000	1.000	1.000	2.230	0.001	2.166	2.231	2.306
	T6	0.172	0.017	-0.058	0.184	0.019	0.059	1.000	1.000	1.000	2.214	0.001	2.154	2.211	2.267
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	0.236	-0.081	0.109	0.240	0.082	0.109	1.000	0.000	0.200	1.968	0.001	1.926	1.968	2.010
	T2	0.269	-0.071	0.081	0.274	0.071	0.081	1.000	0.460	1.000	2.039	0.001	1.985	2.043	2.082
	T3	0.255	-0.069	0.080	0.261	0.069	0.081	1.000	0.480	1.000	2.036	0.001	1.996	2.035	2.077
	T4	0.268	-0.070	0.080	0.271	0.070	0.081	1.000	0.440	1.000	2.041	0.001	1.991	2.033	2.119
	T5	0.270	-0.068	0.076	0.274	0.068	0.076	1.000	0.700	1.000	2.067	0.002	1.995	2.063	2.148
	T6	0.268	-0.070	0.079	0.272	0.070	0.080	1.000	0.540	0.980	2.041	0.001	1.983	2.039	2.101

表 27： 普通最小二乘法（干净）和弹性网络（受污染）下，各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差，参数为 $\boldsymbol{n = 500}$, $n_{\text{sim}} = 50$，迭代次数为 $= 5$, $\rho = 0.6$。前两个区块报告 $P_{\text{ext}} = 0.05$，对应 $P_{\text{miss}} \in \{0.10, 0.30\}$；后两个区块报告 $P_{\text{ext}} = 0.10$，对应$P_{\text{miss}} \in \{0.05, 0.10\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.065	-0.007	0.012	0.113	0.011	0.018	1.000	1.000	1.000	2.081	0.002	2.013	2.079	2.167
	T2	0.177	0.016	-0.056	0.206	0.019	0.058	1.000	1.000	1.000	2.208	0.002	2.135	2.202	2.319
	T3	0.175	0.016	-0.055	0.196	0.019	0.057	1.000	1.000	1.000	2.211	0.002	2.123	2.206	2.313
	T4	0.174	0.016	-0.055	0.201	0.019	0.057	1.000	1.000	1.000	2.212	0.002	2.133	2.211	2.308
	T5	0.188	0.017	-0.060	0.215	0.020	0.062	1.000	1.000	1.000	2.240	0.002	2.163	2.237	2.326
	T6	0.179	0.016	-0.057	0.201	0.019	0.059	1.000	1.000	1.000	2.212	0.003	2.123	2.214	2.319
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.218	-0.096	0.141	0.234	0.096	0.142	1.000	0.000	0.000	1.888	0.002	1.817	1.888	1.955
	T2	0.270	-0.071	0.083	0.285	0.072	0.085	1.000	0.440	0.900	2.035	0.002	1.946	2.036	2.113
	T3	0.261	-0.070	0.084	0.275	0.071	0.085	1.000	0.520	0.920	2.027	0.002	1.939	2.026	2.098
	T4	0.266	-0.072	0.085	0.278	0.073	0.086	1.000	0.440	0.960	2.026	0.002	1.950	2.020	2.128
	T5	0.289	-0.069	0.076	0.300	0.070	0.077	1.000	0.560	1.000	2.088	0.003	1.974	2.086	2.203
	T6	0.266	-0.071	0.084	0.276	0.072	0.085	1.000	0.500	0.960	2.031	0.003	1.947	2.027	2.139
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.189	-0.063	0.169	0.286	0.066	0.171	0.960	0.600	0.000	1.797	0.005	1.685	1.786	1.948
	T2	0.199	0.015	-0.061	0.294	0.028	0.068	1.000	1.000	1.000	2.230	0.007	2.069	2.222	2.402
	T3	0.191	0.016	-0.059	0.271	0.025	0.065	1.000	1.000	1.000	2.221	0.007	2.087	2.214	2.381
	T4	0.138	0.020	-0.058	0.228	0.028	0.064	1.000	1.000	1.000	2.231	0.005	2.083	2.246	2.362
	T5	0.220	0.024	-0.081	0.299	0.032	0.086	1.000	1.000	0.960	2.330	0.006	2.187	2.340	2.464
	T6	0.154	0.020	-0.061	0.260	0.028	0.066	1.000	1.000	1.000	2.237	0.006	2.121	2.233	2.409
稀疏回归（弹性网络）															
样本量: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.05 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	0.127	-0.148	0.263	0.195	0.149	0.264	0.980	0.000	0.000	1.602	0.006	1.485	1.597	1.734
	T2	0.294	-0.070	0.077	0.334	0.073	0.079	0.980	0.600	0.980	2.063	0.007	1.907	2.068	2.210
	T3	0.283	-0.069	0.076	0.316	0.071	0.079	1.000	0.700	0.960	2.047	0.009	1.921	2.043	2.232
	T4	0.301	-0.071	0.078	0.323	0.073	0.081	1.000	0.620	0.960	2.039	0.007	1.882	2.048	2.200
	T5	0.367	-0.062	0.046	0.385	0.063	0.050	0.980	0.840	1.000	2.229	0.010	2.029	2.235	2.415
	T6	0.301	-0.071	0.077	0.322	0.073	0.079	0.980	0.720	0.980	2.036	0.008	1.861	2.023	2.215
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：干净	T1	0.112	0.006	-0.023	0.126	0.008	0.025	1.000	1.000	1.000	2.153	0.001	2.112	2.154	2.192
	T2	0.167	0.017	-0.057	0.179	0.019	0.058	1.000	1.000	1.000	2.215	0.001	2.166	2.214	2.261
	T3	0.165	0.017	-0.056	0.175	0.019	0.057	1.000	1.000	1.000	2.213	0.001	2.166	2.210	2.256
	T4	0.173	0.017	-0.057	0.184	0.018	0.057	1.000	1.000	1.000	2.217	0.001	2.161	2.221	2.259
	T5	0.176	0.019	-0.061	0.187	0.020	0.062	1.000	1.000	1.000	2.231	0.001	2.166	2.232	2.298
	T6	0.172	0.017	-0.057	0.184	0.018	0.058	1.000	1.000	1.000	2.215	0.001	2.156	2.215	2.272
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.05 ρ: 0.6 数据：外部	T1	0.253	-0.093	0.132	0.255	0.094	0.132	1.000	0.000	0.000	1.866	0.001	1.828	1.871	1.899
	T2	0.263	-0.081	0.106	0.266	0.081	0.106	1.000	0.000	0.360	1.932	0.001	1.876	1.939	1.976
	T3	0.253	-0.080	0.105	0.256	0.080	0.105	1.000	0.000	0.400	1.932	0.001	1.881	1.933	1.971
	T4	0.264	-0.080	0.105	0.266	0.081	0.105	1.000	0.040	0.300	1.938	0.001	1.883	1.940	1.990
	T5	0.271	-0.079	0.101	0.274	0.079	0.101	1.000	0.100	0.520	1.963	0.002	1.898	1.956	2.080
	T6	0.261	-0.080	0.105	0.263	0.080	0.105	1.000	0.000	0.360	1.935	0.001	1.875	1.938	1.989
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.057	-0.007	0.013	0.113	0.012	0.017	1.000	1.000	1.000	2.081	0.002	2.008	2.086	2.146
	T2	0.165	0.017	-0.056	0.195	0.020	0.057	1.000	1.000	1.000	2.208	0.002	2.115	2.213	2.289
	T3	0.167	0.016	-0.055	0.199	0.020	0.056	1.000	1.000	1.000	2.209	0.002	2.124	2.208	2.290
	T4	0.158	0.018	-0.055	0.188	0.020	0.057	1.000	1.000	1.000	2.213	0.002	2.140	2.217	2.289
	T5	0.182	0.018	-0.062	0.209	0.022	0.063	1.000	1.000	1.000	2.244	0.002	2.155	2.235	2.345
	T6	0.170	0.016	-0.055	0.198	0.020	0.056	1.000	1.000	1.000	2.207	0.002	2.122	2.214	2.279
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.251	-0.108	0.162	0.258	0.108	0.163	1.000	0.000	0.000	1.790	0.002	1.715	1.789	1.856
	T2	0.265	-0.081	0.106	0.272	0.081	0.107	1.000	0.080	0.400	1.932	0.003	1.839	1.924	2.017
	T3	0.257	-0.080	0.107	0.265	0.081	0.108	1.000	0.140	0.400	1.925	0.002	1.835	1.927	2.013
	T4	0.265	-0.083	0.109	0.270	0.083	0.110	1.000	0.080	0.320	1.928	0.003	1.820	1.935	2.010
	T5	0.281	-0.078	0.097	0.289	0.078	0.098	1.000	0.280	0.760	1.998	0.003	1.901	1.995	2.104
	T6	0.269	-0.083	0.109	0.275	0.084	0.110	1.000	0.080	0.340	1.931	0.003	1.838	1.930	2.034

表28： 普通最小二乘法（干净）和弹性网络（受污染）下，各多重插补方法的模型系数偏差、均方根误差、95%覆盖率和样本外预测均方误差。上部区块： $\boldsymbol{n = 500}$, $n_{\text{sim}} = 50$, 迭代次数 $= 5$,$\rho = 0.6$, $P_{\text{ext}} = 0.10$, $P_{\text{miss}} = 0.30$。其余区块： $\boldsymbol{n = 500}$, $n_{\text{sim}} = 3000$, 迭代次数 $= 5$, $\rho = 0.6$, 其中$P_{\text{ext}} \in \{0.10,\ 0.30\}$和 $P_{\text{miss}} \in \{0.10,\ 0.30\}$

		Bias			RMSE			覆盖率			预测均方误差（样本外）				
设置	MI	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	$\overline{X}$	σ^2	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.192	-0.064	0.174	0.273	0.067	0.176	1.000	0.580	0.000	1.784	0.008	1.593	1.786	1.954
	T2	0.206	0.012	-0.053	0.280	0.024	0.061	1.000	1.000	1.000	2.211	0.011	2.027	2.213	2.402
	T3	0.195	0.013	-0.053	0.268	0.024	0.060	1.000	1.000	1.000	2.210	0.012	1.979	2.194	2.406
	T4	0.150	0.018	-0.056	0.222	0.025	0.063	1.000	1.000	1.000	2.235	0.011	2.040	2.236	2.393
	T5	0.227	0.022	-0.078	0.279	0.028	0.082	1.000	1.000	1.000	2.323	0.010	2.112	2.325	2.467
	T6	0.163	0.017	-0.055	0.242	0.025	0.063	1.000	1.000	0.980	2.222	0.012	2.025	2.235	2.424
稀疏回归（弹性网络）															
n: 500 n_{sim}: 50 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部	T1	0.225	-0.167	0.282	0.257	0.168	0.283	0.920	0.000	0.000	1.497	0.005	1.371	1.498	1.636
	T2	0.275	-0.082	0.105	0.301	0.084	0.108	0.980	0.360	0.600	1.948	0.006	1.825	1.942	2.093
	T3	0.255	-0.079	0.103	0.289	0.081	0.105	1.000	0.520	0.720	1.931	0.007	1.775	1.924	2.083
	T4	0.287	-0.086	0.112	0.305	0.088	0.115	0.940	0.280	0.420	1.931	0.008	1.788	1.934	2.101
	T5	0.324	-0.069	0.071	0.347	0.072	0.075	0.920	0.760	0.980	2.136	0.008	1.982	2.127	2.328
	T6	0.277	-0.081	0.103	0.298	0.083	0.106	0.980	0.340	0.600	1.936	0.008	1.772	1.931	2.074
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 3000 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.060	-0.006	0.011	0.109	0.011	0.018	1.000	1.000	1.000	2.093	0.002	2.007	2.094	2.173
	T2	0.175	0.017	-0.058	0.201	0.020	0.059	1.000	1.000	1.000	2.222	0.002	2.125	2.222	2.316
	T3	0.172	0.017	-0.057	0.197	0.020	0.059	1.000	1.000	1.000	2.220	0.002	2.122	2.220	2.306
	T4	0.165	0.018	-0.057	0.190	0.021	0.059	1.000	1.000	1.000	2.224	0.002	2.127	2.226	2.314
	T5	0.185	0.019	-0.064	0.211	0.022	0.066	1.000	1.000	1.000	2.255	0.003	2.156	2.255	2.357
	T6	0.171	0.018	-0.058	0.198	0.021	0.060	1.000	1.000	1.000	2.225	0.002	2.124	2.225	2.321
稀疏回归（弹性网络）															
n: 500 n_{sim}: 3000 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.247	-0.108	0.162	0.253	0.108	0.162	1.000	0.000	0.000	1.794	0.002	1.707	1.796	1.869
	T2	0.261	-0.081	0.107	0.268	0.082	0.108	1.000	0.094	0.409	1.932	0.002	1.840	1.933	2.018
	T3	0.254	-0.080	0.107	0.260	0.081	0.108	1.000	0.108	0.424	1.928	0.002	1.830	1.929	2.016
	T4	0.264	-0.082	0.108	0.269	0.083	0.109	1.000	0.069	0.370	1.932	0.002	1.839	1.932	2.022
	T5	0.279	-0.079	0.098	0.286	0.079	0.099	0.997	0.215	0.694	1.986	0.004	1.877	1.981	2.123
	T6	0.263	-0.082	0.107	0.268	0.082	0.108	1.000	0.083	0.400	1.931	0.002	1.832	1.932	2.027
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 3000 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：干净	T1	-0.200	-0.061	0.170	0.280	0.065	0.172	0.984	0.682	0.007	1.795	0.006	1.635	1.796	1.943
	T2	0.192	0.016	-0.058	0.273	0.026	0.065	0.999	1.000	0.994	2.228	0.009	2.039	2.229	2.415
	T3	0.169	0.017	-0.056	0.251	0.026	0.063	0.999	1.000	0.994	2.216	0.008	2.032	2.220	2.392
	T4	0.139	0.021	-0.057	0.235	0.029	0.064	1.000	1.000	0.993	2.234	0.008	2.053	2.235	2.418
	T5	0.208	0.026	-0.081	0.283	0.033	0.086	0.997	0.999	0.970	2.335	0.009	2.156	2.336	2.519
	T6	0.154	0.019	-0.058	0.252	0.029	0.066	0.999	0.999	0.988	2.232	0.009	2.050	2.232	2.414
稀疏回归（弹性网络）															
n: 500 n_{sim}: 3000 迭代次数: 5 P_{ext}: 0.10 P_{miss}: 0.30 ρ: 0.6 数据：外部。	T1	0.227	-0.168	0.284	0.258	0.169	0.285	0.912	0.000	0.000	1.494	0.005	1.351	1.495	1.635
	T2	0.272	-0.083	0.108	0.296	0.085	0.111	0.979	0.370	0.584	1.939	0.007	1.773	1.940	2.113
	T3	0.252	-0.081	0.107	0.280	0.083	0.111	0.981	0.442	0.606	1.924	0.008	1.753	1.923	2.106
	T4	0.281	-0.086	0.111	0.298	0.087	0.114	0.972	0.272	0.506	1.941	0.008	1.760	1.942	2.127
	T5	0.324	-0.072	0.077	0.347	0.075	0.082	0.943	0.665	0.922	2.124	0.011	1.924	2.122	2.340
	T6	0.282	-0.085	0.109	0.300	0.087	0.113	0.971	0.323	0.564	1.936	0.009	1.749	1.934	2.134
线性回归（普通最小二乘法）															
n: 500 n_{sim}: 3000 迭代次数: 5 P_{ext}: 0.30 P_{miss}: 0.10 ρ: 0.6 数据：干净	T1	0.060	-0.006	0.011	0.109	0.012	0.018	1.000	1.000	1.000	2.093	0.002	2.007	2.094	2.169
	T2	0.176	0.017	-0.058	0.201	0.020	0.059	1.000	1.000	1.000	2.222	0.002	2.128	2.222	2.312
	T3	0.173	0.017	-0.057	0.197	0.020	0.059	1.000	1.000	1.000	2.220	0.002	2.126	2.221	2.310
	T4	0.165	0.018	-0.058	0.191	0.021	0.060	1.000	1.000	1.000	2.224	0.002	2.132	2.225	2.314
	T5	0.185	0.019	-0.064	0.211	0.022	0.066	1.000	1.000	1.000	2.254	0.003	2.154	2.255	2.353
	T6	0.171	0.018	-0.058	0.198	0.021	0.060	1.000	1.000	1.000	2.225	0.002	2.128	2.226	2.320
稀疏回归（弹性网络）															
n: 500 n_{sim}: 3000 迭代次数: 5 P_{ext}: 0.30 P_{miss}: 0.10 ρ: 0.6 数据：外部	T1	0.125	-0.093	0.150	0.131	0.094	0.151	1.000	0.000	0.000	1.342	0.001	1.263	1.343	1.407
	T2	0.076	-0.060	0.095	0.087	0.061	0.097	1.000	0.743	0.652	1.443	0.002	1.358	1.446	1.519
	T3	0.077	-0.061	0.096	0.087	0.062	0.097	1.000	0.740	0.649	1.442	0.002	1.355	1.444	1.516
	T4	0.086	-0.063	0.098	0.093	0.064	0.100	1.000	0.710	0.626	1.450	0.002	1.362	1.450	1.530
	T5	0.082	-0.056	0.087	0.092	0.057	0.088	1.000	0.884	0.855	1.499	0.004	1.393	1.491	1.651
	T6	0.085	-0.062	0.098	0.092	0.063	0.099	1.000	0.699	0.631	1.453	0.002	1.367	1.454	1.535