arXiv:2611.30599v1 [econ.EM] 23 Nov 2026

实验应测量什么？结合短期实验与存在混杂的观察性数据，以识别长期处理效应*

Kentaro Kawato†

东京大学经济学研究生院

摘要

观察性数据通常会记录长期结果，但当选择取决于未记录特征时，无法识别长期处理效应；随机实验虽能消除这种选择，却需要多年随访。我们探讨短期实验应测量什么，才能与观察性数据结合，识别观察性总体中的平均长期处理效应。在我们的主要设定中，观察性数据记录了未记录特征的一种粗略的处理后测量，但没有记录能够消除混杂的详细测量。在此设定下，我们推导出关于实验应测量什么的三项启示。实验应当：（1）收集相同的粗略测量；（2）足够详细地测量这些特征，使得在给定该详细测量和已记录协变量的条件下，混杂得以消除，并且粗略测量不再能提供关于选择和处理特定的长期结果均值的信息；（3）区分的特征水平不应多于粗略测量所能区分的水平。我们基于排除限制和完备性，通过线性逆问题刻画处理效应，从而证明将此类实验与观察性数据结合能够识别该效应。我们提出一种基于有效影响函数的估计量，可消除干扰参数估计误差的一阶效应。我们给出该估计量渐近正态并达到半参数效率界的条件。交叉拟合放宽了对干扰参数估计的条件。在模拟和 Job Corps 示例中，所提出的 Tikhonov 惩罚级数估计量相比其他估计量具有更小的绝对偏差，覆盖率也更接近名义水平。

关键词： 长期处理效应；数据结合；实验设计；测量误差；代理变量；逆问题

*日期：2026年11月24日。感谢 Ryo Okui 一直以来的指导与支持。所有误差由本人承担。

†东京大学经济学研究生院，日本东京文京区本乡7-3-1，邮编113-0033。电子邮箱：kawato-kentaro380@g.ecc.u-tokyo.ac.jp

1 引言

许多政策决策取决于某项处理的长期影响，例如职业培训对数年后收入的影响，或小班教学对高中毕业的影响。对于许多政策，记录长期结果的观察性数据已经存在，但由于处理选择取决于数据中未记录的特征，这些数据无法识别长期效应。例如，行政记录将过去的职业培训经历与后续收入、过去的班级规模与高中毕业情况联系起来。收入前景不佳的劳动者可能会报名参加培训，而学校可能会将需要学业支持的学生安排到小班。于是，这些数据中的处理组和未处理组单位在也会影响长期结果的特征上存在差异，因此，即使调整已记录协变量，两组之间的比较仍有偏差。

本文探讨，实验应测量什么，才能与现有观察性数据结合后识别长期效应。此类实验面临的主要限制是时间。随机实验能够消除处理选择，但要持续跟踪参与者，直到长期结果出现，需要数年时间。因此，短期实验能提供什么，取决于它测量什么；研究者会根据观察性数据来选择实验中的测量项目。我们将结合这两个数据源视为一个实验设计问题，旨在克服观察性数据的局限，并给出关于测量项目的条件，使短期实验与观察性数据结合后能够识别长期结果变量的平均处理效应。图1概述了这一设计问题。近期研究也将短期实验与观察性数据结合起来，以估计长期影响（Athey、Chetty 和 Imbens 2025；Athey 等人 2026），我们将在本节末讨论这些研究与本文结果的关系。

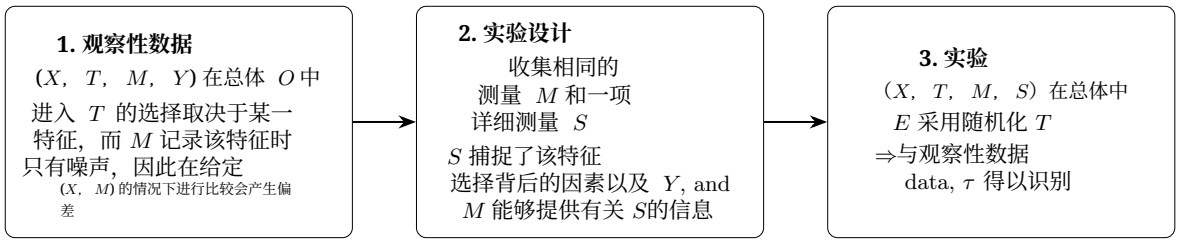

图1 观察性数据、实验设计与实验

注： X 表示处理前协变量， T 表示处理， M 表示两个数据源共有的粗略测量， S 表示对特征的详细测量， Y 表示长期结果。 O 和 E表示观察性总体和实验总体， τ 表示观察性总体中 Y 的平均处理效应。

实验应测量什么，取决于观察性数据在处理后记录了什么。我们考虑的观察性数据除了记录长期结果变量外，还记录每个单位的一项粗略处理后测量，例如常规标准化测试分数，或行政就业记录中的工作周占比。令 $G \in \{E, O\}$分别表示实验总体和观察性总体，X 表示处理前协变量，$T \in \{0, 1\}$表示处理，Y 表示长期结果变量。我们用 M 表示粗略的处理后测量，用 S 表示每个单位的详细处理后测量；只有实验会收集这项测量。实验数据记录（X，T，M，S），观察性数据记录（X，T，M，Y），因此详细测量与长期结果变量是在不同样本中记录的。

我们表明，实验应收集与观察性数据记录相同的粗略测量 M，此外还应收集详细测量 S。我们的主要识别结果给出了若干条件，在这些条件下，这些数据能够识别观察性总体中的平均处理效应，$\tau = \mathbb{E}[Y(1) - Y(0) \mid G = O]$，其中 $Y(t)$、$S(t)$ 和 $M(t)$ 表示处理 t下的潜在值。除了满足一致性和重叠性外，这些条件还要求在实验中给定 X 后，处理经过随机分配；并且给定 X 后，实验总体与观察性总体中 $(S(t), M(t))$ 的联合分布相同。另有三个条件共同构成识别的基础。

第一个条件是给定详细测量后的潜在无混杂性，

$$T \perp Y(t) \mid S(t), X, G = O \tag{1.1}$$

对每个处理 t，该条件指出，在观察性总体中具有相同协变量且 $S(t)$ 取值相同的单位，在不同处理组之间具有可比性。条件（1.1）揭示了 S 的一种解释。如果详细测量捕捉了每个单位的一种潜在特征，而这种特征同时影响处理选择和长期结果，那么具有相同特征值的单位便具有可比性，因此我们将 S 称为该特征。在职业培训示例中，S 是对工人完成项目后的收入能力进行的详细测量；条件（1.1）要求，在收入能力和协变量固定后，是否报名参加与潜在收入无关。在教育示例中，S 是对学生能力的详细评估；条件（1.1）要求，在能力和协变量固定后，是否被安排进入小班与潜在毕业结果无关。

另外两个条件将特征与粗略测量联系起来。第二个条件包含观察性总体中的两个排除限制。第一个排除限制指出，对于协变量和特征均相同的个体，在接受某项处理的情况下，接受处理的

概率不随粗略测量而变化。由于粗略测量是在处理之后进行的，选择不可能直接对其作出反应；该限制要求，在特征固定后，粗略测量不再包含关于选择的信息。在教育示例中，能力相同的两名学生，无论其常规测试分数如何，被安排进小班的可能性都相同。第二个排除限制指出，对于处理、协变量和特征均相同的个体，长期结果的均值不随粗略测量而变化。在教育示例中，一旦能力固定，常规测试分数便与毕业无关。当粗略测量是特征的含噪记录，且噪声既不影响选择，也不影响长期结果时，这两个限制均成立。第三个条件是完备性，它要求在观察性总体中，给定 $M(t)$ 时 $S(t)$ 的条件分布随 $M(t)$ 充分变化。当 $S(t)$ 和 $M(t)$ 取有限多个值时，当且仅当给定 $M(t)$ 的 $S(t)$ 各取值的条件概率矩阵具有满列秩时，完备性才成立；因此， $M(t)$ 至少必须取与 $S(t)$ 一样多的值。

测量误差模型给出了满足这些条件的一种特殊情形。给定 $X = x$ 和 $G = O$后，U_1、 U_2和 U_3 相互独立，其中 U_1 表示一个单位的潜在特征， U_2 是测量误差，而 U_3 是长期结果的扰动。对于每个 $t \in \{0, 1\}$，假设

$$S(t) = f(t, U_1),$$
$$M(t) = g(t, U_1, U_2),$$
$$Y(t) = h(t, U_1, U_3),$$

其中， $f(t, \cdot)$ 是具有可测逆函数的一一映射，并且观察性总体中的处理选择仅取决于该特征，即

$$P(T = 1 \mid U_1, U_2, U_3, X = x, G = O) = P(T = 1 \mid U_1, X = x, G = O).$$

该特征 U_1 随后同时驱动选择和长期结果，实验通过 S对其进行无误差测量，而观察性数据则通过M带噪声地记录该特征。由于 $M(t)$包含噪声 U_2，以 $M(t)$为条件后， U_1中仍存在变异，因此给定 $M(t)$的潜在无混杂性一般并不成立。以 $S(t)$为条件可以解决 U_1，从而潜在无混杂性（1.1）和两个排除限制均成立；此外，在对测量误差 U_2的分布施加条件时，完备性也成立。

识别依赖于两个通过粗略测量连接两个总体的矩方程。对于每个 t 和 x，将观察性总体中的选择概率和长期结果均值定义为该特征的函数：

$$\pi_t(s,x) = P(T = t \mid S(t) = s, X = x, G = O),$$
$$\mu_t(s,x) = \mathbb{E}[Y \mid T = t, S = s, X = x, G = O].$$

两个函数都描述观察性总体，但它们是特征的函数，而观察性数据并未记录该特征；根据排除限制，它们也不依赖于粗略测量。因此，仅凭观察性数据无法恢复这两个函数，而实验同时记录了特征和粗略测量，因而提供了缺失的联系。在我们的条件下，对于每个 t、$F_{X|G=O}$-几乎处处的 x，以及任意有界可测函数 h，这些函数满足

$$\mathbb{E}[\mathbf{1}\{T = t\}h(M) \mid X = x, G = O] = \mathbb{E}[\pi_t(S,x)h(M) \mid T = t, X = x, G = E],$$
$$\mathbb{E}[\mathbf{1}\{T = t\}Yh(M) \mid X = x, G = O] = \mathbb{E}[\pi_t(S,x)\mu_t(S,x)h(M) \mid T = t, X = x, G = E].$$

左侧是观察性数据的矩，右侧涉及实验数据中组别 t 的 (S, M) 的联合分布。每一种 h 都按照粗略测量的函数对样本单位进行加权。由于粗略测量被排除在两个未知函数之外，改变 h 会改变特征每个取值所获得的权重，而不会改变这些函数，因此不同的 h 选择会为同一个未知函数给出不同的方程。完备性要求这些方程只能有一个解。根据此类方程恢复函数属于线性逆问题，非参数工具变量模型就是其中一个例子（Newey and Powell 2003；Horowitz 2011）。两个解的比率给出 μ_t，再对相应实验组中关于特征的分布以及观察性总体中的协变量取平均，可得到 μ_1 和 μ_0，从而得到 τ。

这些条件为计划开展实验的研究人员提供了具体指导。假设观察性数据是学校记录，其中包含常规标准化测试分数和高中毕业信息，并且实验随机分配一项教育干预。那么，实验应设计如下。

(1) 进行与学校记录所采用的常规测试相同的测试，并以相同的方式计算得分函数，从而使两个数据源共享粗略测量 M。

(2) 比常规测试更详细地测量学生的能力，例如由研究人员在同一时间实施一项耗时更长的评估，从而一旦详细测量 S 确定下来，常规测试得分函数便不再携带有关班级安排或毕业的额外信息。

(3) 不要区分常规得分无法区分的更多能力等级。如果研究者将 S 记录为若干类别，那么在观察性数据中，常规得分至少必须取同样多的值。因此，只有当共享测量仍能区分这些类别时，更精细的详细测量才有帮助。

在估计与推断中，我们假设协变量取有限个值。由于结果均值是两个未知函数的比率，这两个函数的估计误差都会对 τ 产生一阶影响。我们的估计量通过加入由两个权重函数构造的校正项，消除了这种一阶效应。干扰函数满足线性逆问题，可通过级数方法进行估计（Aiand Chen 2003；Blundell, Chen,and Kristensen 2007；Chen 2007），模拟和应用中使用了带 Tikhonov 惩罚的级数估计。计量经济学中线性逆问题的 Tikhonov 正则化研究见Carrasco, Florens, andRenault (2007年）和Darolleset al. (2011年）；对级数估计系数施加惩罚，则与 Chen and Pouzo (2012）提出的惩罚筛选最小距离估计量有关。当权重函数唯一且满足切空间条件时，由此得到的影响函数就是 τ 的有效影响函数，其含义见vander Vaart (1998，第25章；这里所考虑的子模型中，两个未知函数在均方意义下均可微。

在关于干扰函数估计值的两类条件下，该估计量渐近正态，且其方差估计量是一致的；此外，当权重函数唯一且切空间条件成立时，该估计量还能达到效率界。第一类条件是对估计误差乘积的速率条件，其形式与Chernozhukov等人(2018年)提出的条件相同。例如，当每个误差的收敛速度都快于样本量倒数的四次方根时，这些条件便成立，因此可以使用收敛较慢的非参数估计。第二类条件控制在生成插件估计值的同一样本上对这些估计值取平均所产生的误差。当以趋近于一的概率，插件函数属于固定的 Donsker 类，且其平方误差属于固定的 Glivenko–Cantelli 类时，这些条件便成立（van der Vaart 1998，第19章）；而这些类可能过小，无法容纳非参数估计或机器学习估计量。交叉拟合版本在不包含用于求平均的观测值的折上估计干扰函数，仅需满足速率条件便能得出相同结论，这与Chernozhukov等人(2018年)提出的双重/去偏机器学习相同。

在一项模拟中，一种设计包含连续特征，另一种设计中的特征取三个值。随着样本量增加，采用 Tikhonov惩罚的所提出的估计量的绝对偏差趋近于零，其置信区间的覆盖率也趋近名义水平，而比较估计量仍然存在偏差。在两种设计中，观察性总体中的处理选择取决于该特征，粗略测量则会产生特征测量误差。第一个比较估计量是用粗略测量代替特征后计算的Athey、Chetty 和 Imbens(2025年)提出的实验选择校正；这两种设计都属于此类情形：这种替代并不足够，因为粗略测量相同的个体在驱动选择的特征上仍有差异。第二个比较估计量是在观察样本中对结果变量关于处理和协变量进行回归所得的估计量，该估计量会因选择而产生偏差。

我们使用 National Job Corps Study 的公开数据来说明该方法。在这项研究中，是否提供参加职业培训项目的机会是随机分配的。由于这些数据来自一项实验，我们据此构建了两种数据来源。我们从研究参与者中抽取实验样本，将第3年收入类别记录为特征；另抽取观察样本，将第4年的周收入记录为长期结果变量。两个样本都以第3年工作周占比的各个水平作为粗略测量，并设定观察样本中的处理概率取决于收入类别，从而在构建时便使观察性数据中的选择取决于该特征。与根据实验计算出的平均效应相比，在我们考察的三条选择规则下，采用 Tikhonov 惩罚的所提出的估计量，其绝对偏差小于比较估计量，且其置信区间的覆盖率更接近名义水平。

1.1 相关文献

本文与结合实验数据和观察性数据来估计长期处理效应的研究相关。Athey、Chetty 和 Imbens (2025年考虑短期结果在两个样本中均有观测、长期结果仅在观察样本中观测，且处理在实验中随机分配的情形，因此实验数据记录$(X,\ T,\ S)$而观察性数据记录 $(X,\ T,\ S,\ Y)$。其关键识别假设是给定短期结果的潜在无混杂性（Athey、Chetty 和 Imbens 2025的假设4），其形式为条件（1.1）。在该假设以及使两个总体具有可比性的条件下，他们识别观察性总体中的平均处理效应（Athey、Chetty 和 Imbens 2025的定理1），并通过实验选择校正（ESC）对其进行估计。Athey 等人的替代指标Athey 等人 (2026年则依赖替代性，即给定短期结果

和协变量后，处理与长期结果相互独立，而且其观察性数据不必记录处理。

在我们的设定中，只有实验中才能观察到详细测量。观察性数据仅记录粗略测量 M，而实验同时记录 M 和详细测量 S，因此，若要用 M 替代 S 来应用他们的方法，就需要在给定 $M(t)$ 的条件下满足潜在无混杂性。在上述测量误差模型中，这一条件不成立，因为 M 记录特征时带有噪声；而给定 $S(t)$ 的条件（1.1）成立，并且在我们的模拟和应用中，使用 M 计算的 ESC 存在偏差。我们不再对共同的短期结果进行条件化，而是将每种处理的结果均值表示为特征的函数，并按相应实验组中特征的分布对其取平均。

表1对比了这些数据结构。在Athey、Chetty 和 Imbens (2025年) 和 Athey 等人 (2026年) 的设定中，他们的识别假设所依据的短期结果 S同时记录在实验数据和观察性数据中。在我们的设定中，承载选择信息的变量 S 仅记录在实验数据中，而粗略测量 M 是两个数据来源都记录的唯一处理后变量。

用另一个样本中的测量来补充观察性数据所缺乏的特征信息，使本文与测量误差文献相关联。Schennach(2016年) 回顾了利用验证数据以及辅助变量（如重复测量、多个指标和工具变量）进行识别与估计的方法。在我们的设定中，提供详细测量与粗略测量联合分布的样本是一个随机实验，而且我们在每个处理组中分别使用这一联合分布。结合观察性总体中的处理选择和潜在无混杂性，这一联合分布能够识别长期因果均值。

根据排除测量的变异恢复特征的未知函数，是一个线性逆问题，如非参数工具变量模型中所示。

设定	X	T	M	S	Y
潜在无混杂性（Athey、Chetty 和 Imbens 2025年）	$E,\ O$	$E,\ O$	—	E, O	O
替代性（Athey 等人 2026年）	E, O	E	—	E, O	O
共享的治疗后测量 M （本文）	$E,\ O$	$E,\ O$	$E,\ O$	E	O

表1 既有研究和本文中实验（E）与观察性（O）数据所记录的变量

注：在每个单元格中，E、O 表示该变量在两个数据来源中均有记录，E 表示该变量仅记录在实验数据中，O 表示该变量仅记录在观察性数据中，破折号则表示该变量在该设定中未使用。在前两行中，S 是相应论文中的短期结果。

Newey 和 Powell (2003年)通过条件分布的完备性刻画这些模型中结构函数的识别，Carrasco、Florens 和 Renault (2007年)综述了结构计量经济学中的线性逆问题，而Horowitz(2011年)回顾了应用研究中的非参数工具变量估计。在我们的矩方程中，粗略测量充当工具变量，特征则充当内生变量；将特征的函数映射为给定测量条件下条件均值的算子，则由实验识别。d'Haultfœuille (2010年)将完备性与给定结果变量时工具变量和选择之间的独立性结合起来，以处理内生选择；我们的选择排除限制则要求给定特征时，粗略测量满足相同的条件。

近端因果推断同样通过观测代理变量中关于未观测变量的积分方程的解来识别因果效应（Miao、Geng 和 Tchetgen Tchetgen 2018年；Tchetgen Tchetgen 等人 2024年）。我们的权重函数求解同类型的积分方程，其中粗略测量是特征的代理变量。在我们的设定中，特征在实验样本中是可观测的，因此连接特征与代理变量的条件分布可以由第二个样本识别。Cui 等人 （2024年）推导了近端因果推断中平均处理效应的有效影响函数，而我们对有效影响函数的推导采用了类似的方法。

在估计中，我们将估计得到的干扰函数代入某个矩中，使其对干扰函数估计误差的一阶敏感性为零，这与Chernozhukov 等人 （2018年）的双重/去偏机器学习类似；他们将Neyman 正交矩与交叉拟合相结合。由于我们结合了两个样本的平均值，并对两个未知函数之比取平均值，因此对估计量的条件表现为：比率误差与两个权重函数误差的乘积需要满足相应的收敛速度。我们的主要估计量在估计干扰函数的同一样本上取平均值，并对这种平均过程施加条件；而我们的交叉拟合估计量则通过样本拆分取代这些条件。

1.2 概述

第2节介绍使用共享的处理后测量进行识别，第 3 节介绍估计与推断。第 4 节报告一项模拟研究，第 5 节使用 Job Corps数据阐释该方法。第6节总结全文。附录包含第2节和第 3 节的证明，以及关于模拟和实证设计的说明，最后还附有两份补充材料的指南。

2 基于共享的治疗后测量进行识别

本节在实验数据和观察性数据共享一个处理后测量时，识别长期处理效应。第 2.1 节设定模型和目标参数，第 2.2 节陈述假设，第 2.3 节给出识别结果。

2.1 设定

本小节介绍两个总体、目标参数，以及识别结果所恢复的特征函数。实验总体和观察性总体分别用 $G \in \{E、O\}$ 表示。实验数据记录为 $(X、T、S、M)$，其中 X 表示处理前协变量，$T \in \{0、1\}$ 表示二元处理，S 表示详细测量，M 表示粗略但共享的测量。观察性数据记录为 $(X、T、M、Y)$，其中 Y 是长期结果。与引言中一样，我们将 S 称为特征。我们将处理 t 下的潜在值记为 $S(t)$，$M(t)$，$Y(t)$。目标参数是观察性总体中的平均处理效应，

$$\tau = \mathbb{E}[Y(1) - Y(0) \mid O] \tag{2.1}$$

其中，此处及下文中，$\mid E$ 和 $\mid O$ 表示以总体为条件，$\mid x、G$ 还进一步以 $X = x$ 为条件。变量取值于欧几里得空间，且 Y 为实值变量。我们用 F 表示条件分布，用 $L^2(F)$ 表示关于 F 平方可积函数的等价类空间。

我们将观察性总体中的处理选择和长期结果描述为特征的函数，而观察性数据缺少这一特征，具体如下：

$$\begin{aligned} \pi_t(s,x) &= P(T = t \mid S(t) = s, x, O), \\ \mu_t(s,x) &= \mathbb{E}[Y \mid T = t, S = s, x, O], \\ \nu_t(s,x) &= \pi_t(s,x)\mu_t(s,x) \end{aligned} \tag{2.2}$$

这些函数是观察性总体分布的特征，其中涉及缺失的特征。在实验中同时测量 S 和 M，可以将观察样本中 M 的矩转化为关于它们的方程。

2.2 识别假设

本小节将假设分为三组。第一组使实验中特征与测量的联合分布能够提供有关观察性总体的信息；第二组将恢复出的结果变量均值与潜在

结果联系起来；第三组则允许通过共享测量恢复两个未知函数。

假设 2.1（一致性）。 $S = S(T)$， $M = M(T)$，以及 $Y = Y(T)$。

假设 2.1 表明，观测到的特征、测量值和结果变量等于实际接受处理下的潜在值，这将下文假设中的潜在值与观测数据联系起来。

假设 2.2（重叠性与正值性）。 $F_{X|O} \ll F_{X|E}$。对于 $F_{X|O}$-几乎处处的 x， $P(T = t \mid x, \ E) > 0$，且 $\pi_t(S(t), \ x) > 0$ 在观察性总体中几乎处处成立。

假设 2.2 要求，在目标总体的协变量取值范围内，实验的两个处理组都可用。选择概率为正，使我们能够将恢复的 ν_t 除以 π_t，从而得到结果均值。

假设 2.3（实验中的随机化）。对于每个 t， $T \perp (S(t), \ M(t)) \mid X, \ G = E$.

假设2.3指出，实验中的处理经过随机化，分配概率还可以依赖于 X。

假设 n2.4（联合分布的可迁移性）。 对于每个 t 和 $F_{X|O}$-几乎每个 x,

$$F_{S(t),M(t)|x,O} = F_{S(t),M(t)|x,E} \tag{2.3}$$

在假设 2.1、2.3 和 2.4 下，观察性总体中给定 $X = x$ 时，$(S(t), M(t))$ 的联合分布等于实验组 t 中给定 $X = x$ 时，(S, M) 的分布。随机化使实验总体中潜在值的分布与组别 t 中的分布相同，在组别 t 中这些值可以被观测到；方程 (2.3) 则将该分布推广至观察性总体。由此得到的联合分布是第 2.3 节中两个矩方程的已知系数。

假设 2.5（潜在无混杂性）。 对于每个 t,

$$T \perp Y(t) \mid S(t), X, G = O \tag{2.4}$$

假设 2.5 表明，给定治疗下的特征 t 和协变量后，观察性总体中的处理选择与长期潜在结果相互独立。它将下文恢复的结果均值 μ_t 与观察性总体中潜在结果的均值联系起来。与之配套使用的可积性条件见附录 A。

假设 2.1–2.5 是因果推断中的标准假设，与 Athey、Chetty 和 Imbens （2025年）的假设相似，只是用特征替代了他们的短期结果。接下来的三个假设是本文的关键假设。

假设 2.6 （选择排除限制）。 对每个 t,

$$P(T = t \mid S(t), M(t), X, O) = \pi_t(S(t), X) \tag{2.5}$$

几乎处处。

假设 2.7 （结果均值排除限制）。 对每个 t, $\mathbb{E}[Y \mid T = t, S, M, X, O] = \mu_t(S, X)$ 几乎处处。

假设 2.6 和 2.7 要求，在特征和协变量固定后，共享测量不再包含关于处理选择 t 或长期结果均值的额外信息。在假设 2.6 下，给定特征时 $M(t)$ 的变异识别出 π_t；而在假设 2.7 下，涉及长期结果的矩中的相同变异识别出 ν_t。在这两个例子中，这些假设要求 S捕捉驱动处理选择和长期结果的特征，因此，一旦详细测量固定，工作周占比或常规测试分数就不再提供额外信息。

假设 2.8 （完备性）。 对每个 t 和几乎每个 $F_{X|O}$- x，每个 $h \in L^2(F_{S(t)|x,\ O})$ 都满足

$$\mathbb{E}[h(S(t)) \mid M(t), x, O] = 0 \quad \Longrightarrow \quad h(S(t)) = 0 \quad \textit{almost surely} \tag{2.6}$$

假设 2.8 要求，特征的一个平方可积函数可以由 S（t） 的分布如何随 M（t）的取值变化来确定。它是Newey 和 Powell（2003年）针对非参数工具变量模型提出的完备性条件，其中，特征充当内生变量，共享测量充当工具变量。Newey 和 Powell （2003年，定理 2.2–2.4）根据给定工具变量的内生变量条件分布，为完备性给出了原始条件；在此处，即给定 M（t） 时 S（t）的条件分布，以及 $X = x$在指数族、均值关于工具变量呈线性的正态分布，以及取值有限的变量情形下的条件分布。将详细测量和粗略测量结合到实验中，可以提供完备性得以成立所需的信息，下面的例子将对此加以说明。

例 2.1 （测量误差模型）。 在观察性总体中固定 $X = x$，并在以下函数中省略x。设 U_1、 U_2 和 U_3 在给定 $X = x$ 和 $G = O$的条件下相互独立，

其中， U_1 表示潜在特征， U_2 是测量误差， U_3 是结果扰动。对于每个 $t \in \{0,1\}$，假设

$$S(t) = f(t, U_1),$$
$$M(t) = g(t, U_1, U_2),$$
$$Y(t) = h(t, U_1, U_3),$$

其中， $f(t, \cdot)$ 是一一映射且具有可测逆函数，并且观察性总体中的处理选择仅取决于该特征，即

$$P(T = 1 \mid U_1, U_2, U_3, x, O) = P(T = 1 \mid U_1, x, O).$$

当该结构对 $F_{X|O}$-几乎处处的 x 以及 t 都成立时，结合假设 2.1 和附录 A 中的可积性条件，它蕴含假设 2.5 中的潜在无混杂性，以及假设 2.6 和 2.7 中的排除限制。由于 $f(t, \cdot)$ 是一一映射，给定 $X = x$ 固定 $S(t)$ 会确定 U_1，因此，给定 $S(t)$、 $X = x$ 和 $G = O$ 时， T、 $M(t)$ 和 $Y(t)$ 相互独立。仅测量误差的独立性并不能推出假设 2.5，而上述选择条件提供了这一推论。由于 M 在处理之后测量，测量函数 g 可能取决于 t。

此外，如果对于每个 t 以及 $F_{X|O}$-几乎处处的 x，给定 $X = x$ 和 $G = O$ 时， $S(t)$ 在某个有界区间上服从均匀分布（例如， U_1 在某个有界区间上服从均匀分布，且 $f(t, u) = a_t + u$），并且 $M(t) = S(t) + U_2$，其中 $U_2 \sim \mathcal{N}(0, \sigma^2)$ 对于 $\sigma^2 > 0$ 成立，那么如附录 A 所证明的，假设 2.8 中的完备性也成立。更一般地，当 $M(t) = S(t) + U_2$，且具有实值的 $S(t)$ 和 U_2 在给定 $X = x$ 和 $G = O$ 的条件下相互独立时，只要 U_2 的特征函数在实数轴的稠密子集上非零，则对于 $S(t)$ 的任何分布（连续或离散），完备性都成立。正态误差满足这一条件，在区间上服从均匀分布的误差也满足，因为其特征函数仅在孤立点处为零。第 4 节模拟所采用的设计，其中该特征服从均匀分布或三点分布，且测量误差服从均匀分布，也属于此类（附录 A.2）。

2.3 识别结果

本小节给出了两个样本所识别的矩，并通过它们的唯一解获得平均处理效应。

定理 2.1（使用共享的处理后测量进行识别）。 假设假设 *2.1–2.8* 以及附录 A 中的可积性条件成立。那么，对于每个 t、 $F_{X|O}$-几乎处处的 x，以及每个有界可测函数 h，

$$\begin{aligned}\mathbb{E}[\mathbf{1}\{T=t\}h(M)\mid x,O] &= \mathbb{E}[\pi_t(S,x)h(M)\mid T=t,x,E],\\ \mathbb{E}[\mathbf{1}\{T=t\}Yh(M)\mid x,O] &= \mathbb{E}[\nu_t(S,x)h(M)\mid T=t,x,E]\end{aligned} \tag{2.7}$$

E每个方程在 $L^2(F_{S|T=t,x,E})$ 中都有唯一解，并且 $\mu_t=\nu_t/\pi_t$ 是可识别的。此外 ，

$$\tau=\int\left\{\int \mu_1(s,x)\,dF_{S|T=1,x,E}(s)-\int \mu_0(s,x)\,dF_{S|T=0,x,E}(s)\right\}dF_{X|O}(x) \tag{2.8}$$

方程 (2.7) 的左侧可由观察性数据得知，右侧的 (S,M) 的分布则可由实验得知。由于这些方程使用关于 M 的有界测试函数，因此对于 M 的任何分布，无论是离散分布还是连续分布，方程都成立。第一个方程恢复选择概率，第二个方程恢复选择概率与结果均值的乘积。对两个方程应用完备性，并除以正的选择概率，即可得到作为特征函数的结果均值。方程 (2.8) 对该函数在各实验组中特征的分布下取平均，并在观察性总体的协变量分布下取平均。证明见附录 A。

选取有限个检验函数后，方程 (2.7) 可给出估计准则。如果用一个基来近似特征的函数，右侧就变为一个由实验识别的矩阵与未知系数的乘积。第 3 节将由此得到的未知函数估计值，以及表示取平均所需权重的函数估计值，代入平均处理效应的估计量，并推导其渐近分布。

注 2.1（使用处理前代理变量进行识别）。 当观察性数据中长期结果变量是唯一的处理后变量，且两个样本都记录了特征的处理前代理变量W 时，同样的方法也适用；例如，职业培训示例中的随机分配前收入，或教育示例中上一学年的常规测试得分。随后，实验将特征与代理变量一同测量。给定代理变量和协变量（其分布可能因总体而异），实验中的随机化和两个总体之间的可迁移性成立，同时维持给定特征和协变量的潜在无混杂性。一旦特征和协变量固定，代理变量不会再为处理选择或长期结果的均值提供额外信息，而特征在给定代理变量和协变量条件下的分布在两个总体中相同，且代理变量的变化足以推动特征分布发生变化，从而满足完备性。

矩方程（2.7）随后被替换为关于代理变量和协变量的条件矩，τ 在给定代理变量和协变量的条件下，对每个实验组别中的结果均值关于特征取平均，然后再对观察性总体中代理变量和协变量的分布取平均。补充材料 1 陈述了这些条件和识别结果，给出了有效影响函数和估计量，报告了一项模拟研究，并给出了两个样本同时记录处理前代理变量和共享的治疗后测量时的识别条件。

3 估计与推断

本节根据第2节中的识别矩构造平均处理效应的估计量，并给出其效率、渐近分布和标准误。第 3.1 节陈述抽样假设，定义估计中使用的总体函数，并给出有效影响函数；第 3.2 节定义估计量；第 3.3 节陈述条件和主要定理。第3.4节讨论使用交叉拟合进行估计。

3.1 目标估计量与有效影响函数

本小节首先陈述抽样假设，然后将平均处理效应表示为对有限多个层的平均，通过观测值的函数表示干扰函数中误差的一阶效应，并证明所得影响函数就是有效影响函数。

设 n_E 和 n_O 分别为实验样本和观察性研究样本的大小，并设 $N = n_E + n_O$。接下来的三个假设描述抽样过程，将协变量限制为有限个取值，并从下方界定选择概率。

假设 3.1（抽样）。两个样本相互独立，每个样本中的观测值均为从相应总体中独立同分布抽取的样本，且 $n_E/N \to \rho$ 对于某个$\rho \in$（0，1）成立。

假设 3.2（有限个协变量取值）。

1. X 取值于一个固定的有限集合 $\mathcal{X}$。

2. 存在一个常数 $c > 0$，使得 $P(X = x \mid O) \geq c$ 且 $P(T = t,\ X = x \mid E) \geq c$ 对所有 t、x均成立。

假设 3.3（选择概率远离零）。存在常数 $a \in$（0、1），使得对于每个 t、x、$\pi_t \geq a$，在 $S \mid T = t$、x、E 的支持集上几乎处处成立。

假设 3.1 约束两个样本中平均值的变化。假设 3.2 允许在层x内使用样本均值来估计给定 $X = x$ 的每个条件均值，因此下文的估计量会在各层内取平均，其渐近理论依赖于 X的有限支撑。假设 3.3 中选择概率的下界使比率$\mu_t = \nu_t/\pi_t$ 及其估计量保持稳定，第 3.2 节中的估计量会将估计的选择概率截断在同一常数 a处。结果变量的有界性条件汇总于附录 B，估计量的有界性条件汇总于附录 C。

对于每个 $t \in \{0、1\}$ 和 $x \in \mathcal{X}$，令

$$\bar{\mu}_t(x) = \mathbb{E}[\mu_t(S,x) \mid T = t, x, E] = \mathbb{E}\left[\frac{\nu_t(S,x)}{\pi_t(S,x)} \,\middle|\, T = t, x, E\right] \tag{3.1}$$

表示在实验组 t 中，对 S 的分布取均值后得到的结果均值。根据（2.8）和（3.1），$\tau = \sum_{x\in\mathcal{X}} P(X = x \mid O)\{\bar{\mu}_1(x) - \bar{\mu}_0(x)\}$。（3.1）中的第二个表达式表明，$\bar{\mu}_t(x)$ 是（2.7）所确定的两个未知函数之比的均值，而这个比率决定了这两个函数的估计误差如何影响 τ 的估计量。

$\bar{\mu}_t(x)$ 的估计量同时使用 ν_t 和 π_t 的估计值，而它们的误差会以一阶影响 $\bar{\mu}_t(x)$。令 d 为 S 的一个有界函数，且满足 $|d| \le D$；再令 ε 为一个满足 $|\varepsilon|D \le a/2$ 的数。将 ν_t 替换为 $\nu_t + \varepsilon d$ 会使 $\bar{\mu}_t(x)$ 改变

$$\mathbb{E}\left[\frac{\nu_t(S,x) + \varepsilon d(S)}{\pi_t(S,x)} \,\middle|\, T = t, x, E\right] - \bar{\mu}_t(x) = \varepsilon\,\mathbb{E}\left[\frac{d(S)}{\pi_t(S,x)} \,\middle|\, T = t, x, E\right]$$

改用 $\pi_t$by $\pi_t + \varepsilon d$，则可使用恒等式 $\nu/(\pi + \varepsilon d) - \nu/\pi = -\varepsilon d\mu/\pi + \varepsilon^2 d^2\mu/\{\pi(\pi + \varepsilon d)\}$，其中 $\mu = \nu/\pi$，从而得到

$$\begin{aligned}
&\mathbb{E}\left[\frac{\nu_t(S,x)}{\pi_t(S,x) + \varepsilon d(S)} \,\middle|\, T = t, x, E\right] - \bar{\mu}_t(x) \\
&= -\varepsilon\,\mathbb{E}\left[\frac{d(S)\mu_t(S,x)}{\pi_t(S,x)} \,\middle|\, T = t, x, E\right] + \varepsilon^2\,\mathbb{E}\left[\frac{d(S)^2\mu_t(S,x)}{\pi_t(S,x)\{\pi_t(S,x) + \varepsilon d(S)\}} \,\middle|\, T = t, x, E\right]
\end{aligned}$$

由于 $\pi_t \ge a$ and $\pi_t + \varepsilon d \ge a/2$，最后一项的绝对值至多为 $2\varepsilon^2 D^2\mathbb{E}[|\mu_t(S,x)| \mid T = t, x, E]/a^2$。因此，对 ν_t or π_t 的估计值中存在估计误差 d，会使 $\bar{\mu}_t(x)$ 的代入估计发生偏移，偏移量为以 $1/\pi_t$ 加权的 d 的均值，或为 $-\mu_t/\pi_t$，误差阶为 ε^2。这些偏移与估计误差同阶；对于未知函数的非参数估计，其估计误差消失速度比 $N^{-1/2}$ 慢，因此估计量必须消除这些偏移。

为消除这些偏移，估计量减去或加上在真实函数处为零且可由两个样本计算的项。对于关于 M 和 x 的平方可积函数 ω_t^{ν}，

方程 (2.7) 的第二个方程将有界函数扩展至关于 M 的可积函数（附录 B），给出带有 $h = \omega_t^\nu(\cdot, x)$ 的结果

$$\mathbb{E}[\nu_t(S,x)\omega_t^\nu(M,x) \mid T=t,x,E] - \mathbb{E}[\mathbf{1}\{T=t\}Y\omega_t^\nu(M,x) \mid x,O] = 0$$

将 ν_t 替换为 $\nu_t + \varepsilon d$ 会使左侧改变 $\varepsilon\mathbb{E}[d(S)\omega_t^\nu(M, x) \mid T=t, x, E]$，因为第二个均值不涉及 ν_t。从 $\bar{\mu}_t(x)$ 的代入估计中减去这一项后，剩下的偏移为 $\varepsilon\mathbb{E}[d(S)\{1/\pi_t(S, x) - \omega_t^\nu(M, x)\} \mid T=t, x, E]$。由于 d 仅取决于 S，在期望中对给定的 S 取条件均值，可将该偏移写为

$$\varepsilon\,\mathbb{E}\left[d(S)\left\{\frac{1}{\pi_t(S,x)} - \mathbb{E}[\omega_t^\nu(M,x) \mid S, T=t,x,E]\right\} \middle| T=t,x,E\right]$$

因此，当且仅当 ω_t^ν满足以下条件时，对于每个有界的 d 和每个 ε，剩余偏移均为零

$$\mathbb{E}[\omega_t^\nu(M,x) \mid S, T=t,x,E] = \frac{1}{\pi_t(S,x)}$$

几乎处处成立。对 π_t中的误差，使用方程 (2.7) 的第一个方程可以得到相同的论证。对于关于 M和 x的函数 ω_t^π，将 π_t 替换为 $\pi_t + \varepsilon d$ 会使 $\mathbb{E}[\pi_t(S,x)\omega_t^\pi(M,x) \mid T=t,x,E] - \mathbb{E}[\mathbf{1}\{T=t\}\omega_t^\pi(M,x) \mid x,O]$ 改变为 $\varepsilon\mathbb{E}[d(S)\omega_t^\pi(M,x) \mid T=t,x,E]$；加上这一项后，将抵消偏移 $-\varepsilon\mathbb{E}[d(S)\mu_t(S,x)/\pi_t(S,x) \mid T=t,x,E]$。对于每个有界的 d，当且仅当$\mathbb{E}[\omega_t^\pi(M,x) \mid S,T=t,x,E] = \mu_t(S,x)/\pi_t(S,x)$ 几乎处处成立时，才能实现这一点。下一个假设要求此类函数存在。

假设 3.4 (权重的表示)。 对于每个 $t \in \{0,1\}$ 以及 $x \in \mathcal{X}$，存在关于 M 和 x 的实值函数 $\omega_t^\nu, \omega_t^\pi$，满足$\mathbb{E}[\omega_t^\nu(M,x)^2 \mid T=t,x,E] < \infty$ 以及 $\mathbb{E}[\omega_t^\pi(M,x)^2 \mid T=t,x,E] < \infty$，使得

$$\mathbb{E}[\omega_t^\nu(M,x) \mid S,T=t,x,E] = \frac{1}{\pi_t(S,x)}, \qquad \mathbb{E}[\omega_t^\pi(M,x) \mid S,T=t,x,E] = \frac{\mu_t(S,x)}{\pi_t(S,x)} \tag{3.2}$$

几乎处处。

假设 3.4 要求权重 $1/\pi_t$ 和 μ_t/π_t 是在实验组 t 中给定 S 时，关于 M 的函数的条件均值。我们将 ω_t^ν 和 ω_t^π 称为权重函数。在这一表示下，迭代期望给出 $\mathbb{E}[d(S)/\pi_t(S, x) \mid T=t, x, E] = \mathbb{E}[d(S)\omega_t^\nu(M, x) \mid T=t, x, E]$，对于有界的 d 成立，并且右侧通过（2.7）与观察样本中的一个均值相关联。方程（3.2）可能有多个解，因为可以将关于 M 且给定 S 时条件均值为零的函数加到任意一个解上。以下部分中的渐近分布

3.3 和 3.4 适用于权重估计所收敛到的解，我们将其记为 ω_t^{ν}， ω_t^{π}；效率结果利用了解的唯一性（注3.1）。

接下来，我们定义两个表示估计误差一阶部分的函数。对于实验样本中的一个观测值 $Z_E = (X, T, M, S)$ 和观察样本中的一个观测值 $Z_O = (X, T, M, Y)$，令

$$\begin{aligned}\psi_E(Z_E) = &\sum_{x\in\mathcal{X}} \frac{P(X=x \mid O)\mathbf{1}\{T=1, X=x\}}{P(T=1, X=x \mid E)}\{\mu_1(S,x) - \bar{\mu}_1(x) - \nu_1(S,x)\omega_1^{\nu}(M,x) + \pi_1(S,x)\omega_1^{\pi}(M,x)\} \\ &- \sum_{x\in\mathcal{X}} \frac{P(X=x \mid O)\mathbf{1}\{T=0, X=x\}}{P(T=0, X=x \mid E)}\{\mu_0(S,x) - \bar{\mu}_0(x) - \nu_0(S,x)\omega_0^{\nu}(M,x) + \pi_0(S,x)\omega_0^{\pi}(M,x)\}\end{aligned} \tag{3.3}$$

and

$$\begin{aligned}\psi_O(Z_O) = &\bar{\mu}_1(X) - \bar{\mu}_0(X) - \tau + \mathbf{1}\{T=1\}\{Y\omega_1^{\nu}(M,X) - \omega_1^{\pi}(M,X)\} \\ &- \mathbf{1}\{T=0\}\{Y\omega_0^{\nu}(M,X) - \omega_0^{\pi}(M,X)\}\end{aligned} \tag{3.4}$$

根据（3.2）， $\mathbb{E}[\nu_t(S,x)\omega_t^{\nu}(M,x) \mid T=t, x, E]$ 以及 $\mathbb{E}[\pi_t(S,x)\omega_t^{\pi}(M,x) \mid T=t, x, E]$ 都等于 $\bar{\mu}_t(x)$，而（2.7）连同这两个等式给出 $\mathbb{E}[\mathbf{1}\{T=t\}\{Y\omega_t^{\nu}(M,x) - \omega_t^{\pi}(M,x)\} \mid x, O] = 0$。因此，两个函数的均值都为零。函数 ψ_E 源自实验样本中的分层均值以及两个未知函数的估计，而 ψ_O 源自观察样本中的层份额和矩。

我们根据van der Vaart（1998，第25章）所述的意义，采用针对两个独立样本调整后的有效影响函数来刻画效率。模型 $\mathcal{P}$ 是 Z_E 和 Z_O 的概率律对所组成的集合，在该集合下， X 取值于 $\mathcal{X}$，且假设 2.1–2.8 以及附录 A 中的可积性条件均成立。附录 B 定义了带有有界得分函数的一维子模型；沿着这些子模型，根据概率律计算得到的函数 π_t 和 ν_t 关于子模型参数在均方意义下可微，同时还定义了在内积下的切空间，其中每个样本均按其份额加权，并给出了 τ的梯度以及切空间条件的形式化表述。切空间条件要求归一化对（ψ_E/ρ， $\psi_O/(1-\rho)$）属于切空间。

定理 3.1（有效影响函数与半参数效率界）。 假设假设 *2.1–2.8* 和 *3.1–3.4* 成立，附录 *A* 中的可积性条件以及附录 *B* 中的有界性和切空间条件也成立，并且对于每个 $t \in \{0、1\}$ 和 $x \in \mathcal{X}$，*(3.2)* 都有唯一的平方可积解（ω_t^{ν}、 ω_t^{π}）。那么，相对于附录 *B* 的子模型， τ 在模型$\mathcal{P}$ 中的有效影响函数为

$$\left(\frac{\psi_E(Z_E)}{\rho}, \frac{\psi_O(Z_O)}{1-\rho}\right)$$

ψ_E 和 ψ_O 由 *(3.3)* 和 *(3.4)* 给出，且它是 τ 唯一一个属于切空间的梯度。效率界为

$$\Sigma^2 = \frac{1}{\rho}\mathbb{E}[\psi_E(Z_E)^2 \mid E] + \frac{1}{1-\rho}\mathbb{E}[\psi_O(Z_O)^2 \mid O]$$

对于所有相对于切空间条件中得分函数子空间正则的估计量序列，$\sqrt{N}$ 倍估计误差的极限分布是 $\mathcal{N}(0, \Sigma^2)$ 与某个概率分布的卷积，因此，当其方差有限时，该方差至少为 Σ^2。

定理 3.1 给出了一个下界以及达到该下界的方法。对于所有相对于切空间条件中的得分函数正则的估计量，当方差有限时，$\sqrt{N}$ 倍估计误差的极限分布的方差至少为 Σ^2；而估计误差等于 $n_E^{-1}\sum_{i=1}^{n_E}\psi_E(Z_{E,i}) + n_O^{-1}\sum_{j=1}^{n_O}\psi_O(Z_{O,j})$（相差一个阶为 $o_p(N^{-1/2})$ 的项）的估计量，其渐近方差为 Σ^2，因此，有效影响函数告诉我们如何构造有效估计量。(3.3) 和 (3.4) 中涉及权重函数的项，是用于消除上述对 π_t 和 ν_t 的估计误差所产生的一阶效应的校正项。因此，第 3.2 节的估计量将干扰函数的估计值代入这些校正项，再将其加到$\bar{\mu}_t(x)$的代入估计上；下文的定理 3.2 和 3.3 表明，在满足定理 3.1 的条件以及关于干扰函数估计值的条件时，所提出的估计量达到了该下界。

沿着一个子模型，π_t 以及 ν_t 根据方程（2.7）随两个样本的概率律变化。由于这些函数沿附录 B 的子模型可微，沿一个子模型对（2.7）求导，可得 τ 的导数等于ψ_E 与 ψ_O 分别同该子模型的得分函数的乘积的均值之和，其中每个平方可积解均满足（3.2）。与Newey （1990）一样，我们通过沿一维子模型的导数研究效率，并将注意力限制在 π_t 和 ν_t 可微的那些子模型上。当 S 和 M 取给定的有限多个值 $T = t$ 以及 $X = x$时，如第 5 节的实证示例所示，（2.7）是一个线性方程组，其矩阵的列根据假设 2.8 线性独立，并且附录 B 的每条正则路径都是这样的子模型。切空间条件要求用 $\mathcal{P}$中的子模型得分函数逼近归一化影响函数。证明见附录 B。

近期研究推导了因果参数的有效影响函数，而这些参数的识别需要求解此类逆问题。定理 3.1 Cui等人(2024年) 在条件期望算子满足一定条件时，给出了近端因果推断中平均处理效应的有效影响函数；我们的定理 3.1 证明沿用了他们的方法。

注3.1（权重函数并非唯一）。影响函数 (3.3) 和 (3.4) 取决于 (3.2) 的解 $(\omega_t^\nu, \omega_t^\pi)$，而定理 3.1 以及定理 3.2 和 3.3 中关于效率的论断都假设该解唯一。当 M 比 S 更丰富时，解可能不唯一；此时，存在某个非零的平方可积函数，它是 M 的函数，且给定 S 时条件均值为零。例如，在实验中，若给定 $T=t$ 时 $M=S+U_2$，且 U_2 在 $[-1,1]$ 上均匀分布并独立于 S，如第 4 节的模拟设计所示，则 $\mathbb{E}[\sin(\pi M) \mid S, T=t, x, E]=0$。当给定 $T=t$ 和 $X=x$ 时，S 和 M 取有限个值，假设 2.8 要求 M 取值的数量至少与 S 一样多，而唯一性要求其至多与 S 一样多，因此只有二者取值数量相同时，这两个条件才能同时成立。

当解不唯一时，两个解仍可能给出相同的一对影响函数。由某个解构造出的每一对不同的影响函数都是 τ 的梯度；由于切空间中的两个梯度相同，切空间条件至多对其中一对不同的影响函数成立。在这种情况下，第 3.2 节和第 3.4 节的估计量仍具有定理 3.2 和 3.3 所述的渐近分布和方差估计量的一致性，其中 Σ^2 根据其权重估计所收敛到的解计算；而这些定理中的效率结论则要求解是唯一的。

3.2 估计量

本小节定义一种估计量：在同一个全样本上估计干扰函数并按层计算平均值，同时介绍样本均值的记号。两个独立样本分别为 $Z_{E,i}=(X_i, T_i, M_i, S_i)$，$i=1,\ldots,n_E$，以及 $Z_{O,j}=(X_j, T_j, M_j, Y_j)$，$j=1,\ldots,n_O$。我们用 $\mathbb{E}$ 和 P 表示条件均值和份额的全样本对应量。例如，$\mathbb{E}[f(S,M) \mid T=t, x, E]$ 表示在满足 $T_i=t$ 和 $X_i=x$ 的观测值中，$f(S_i, M_i)$ 的平均值；$\mathbb{E}[g(T, M, Y) \mid x, O]$ 表示在满足 $X_j=x$ 的观测值中，$g(T_j, M_j, Y_j)$ 的平均值；而 $P(T=t, X=x \mid E)$ 和 $P(X=x \mid O)$ 分别表示在各自样本中满足这些条件的观测值所占的份额。

对于每个 t，我们使用两个样本中的所有观测值来估计干扰函数 $\pi_t, \nu_t, \omega_t^\nu, \omega_t^\pi$。估计方法不作具体限定，且这些估计量关于样本和函数的自变量联合可测。我们将选择概率的估计值截断至 $[a,1]$，截断所用常数为假设 3.3 中的 a，并将结果记为 $\widehat{\pi}_t$；将乘积和两个权重函数的估计量分别记为$\widehat{\nu}_t, \widehat{\omega}_t^\nu, \widehat{\omega}_t^\pi$，将估计的比率记为 $\widehat{\mu}_t=\widehat{\nu}_t/\widehat{\pi}_t$。由于 π_t, ν_t 和权重函数是线性逆问题的解，因此级数估计量或筛估计量是自然的选择，这与非参数工具变量模型和条件矩模型中的做法一致（Newey and Powell 2003；Ai and Chen 2003；Blundell, Chen, and Kristensen 2007；Chen 2007；Chen and Pouzo 2012）。第 4 节中的模拟以及

第 5 节中的应用采用带有 Tikhonov惩罚的级数估计。固定这些估计量后，我们在同一个全样本上计算分层均值。

对于每个 t，x，令

$$\begin{aligned}\widehat{\bar{\mu}}_t(x) = \widehat{\mathbb{E}}[\widehat{\mu}_t(S,x) - \widehat{\nu}_t(S,x)\widehat{\omega}_t^{\nu}(M,x) + \widehat{\pi}_t(S,x)\widehat{\omega}_t^{\pi}(M,x) \mid T=t, x, E] \\ + \widehat{\mathbb{E}}[\mathbf{1}\{T=t\}\{Y\widehat{\omega}_t^{\nu}(M,x) - \widehat{\omega}_t^{\pi}(M,x)\} \mid x, O]\end{aligned} \tag{3.5}$$

并将平均处理效应的估计量定义为

$$\widehat{\tau} = \sum_{x\in\mathcal{X}} \widehat{P}(X=x \mid O)\{\widehat{\bar{\mu}}_1(x) - \widehat{\bar{\mu}}_0(x)\} \tag{3.6}$$

(3.5) 的第一项将两个校正项 $-\widehat{\nu}_t\widehat{\omega}_t^{\nu} + \widehat{\pi}_t\widehat{\omega}_t^{\pi}$ 加到估计比率的层均值上，第二项则在观察样本中估计相同校正项的均值。代入真实函数时，第一项中的预期校正项和第二项中的期望都为零，如第 3.1 节所示。代入估计量时，两项中的校正项会抵消估计比率误差的一阶部分，附录 C 将这一性质表述为一个恒等式。

$\widehat{\tau}$ 的渐近方差为 Σ^2（定理 3.2）；在该定理的条件下，它等于定理 3.1 的效率界。标准误则将 (3.3) 和 (3.4) 中未知的量 $\widehat{\ }$ 替换为估计值。令 $\psi_E(Z_{E,i})$ 表示在 $Z_{E,i}$ 处计算的 (3.3)，其中 $\mu_t, \nu_t, \pi_t, \omega_t^{\nu}, \omega_t^{\pi}$ 分别替换为 $\widehat{\mu}_t, \widehat{\nu}_t, \widehat{\pi}_t, \widehat{\omega}_t^{\nu}, \widehat{\omega}_t^{\pi}$，$\bar{\mu}_t(x)$ 替换为 $\widehat{\bar{\mu}}_t(x)$，而 $P(X=x \mid O)$ 和 $P(T=t, X=x \mid E)$ 分别替换为 $\widehat{P}(X=x \mid O)$ 和 $\widehat{P}(T=t, X=x \mid E)$。令 $\psi_O(Z_{O,j})$ 表示在 $Z_{O,j}$ 处计算的 (3.4)，采用相同的替换，并将 τ 替换为 $\widehat{\tau}$。我们回顾 Σ^2，并将其估计量定义为

$$\begin{aligned}\Sigma^2 &= \frac{1}{\rho}\mathbb{E}[\psi_E(Z_E)^2 \mid E] + \frac{1}{1-\rho}\mathbb{E}[\psi_O(Z_O)^2 \mid O], \\ \widehat{\Sigma}^2 &= \frac{N}{n_E}\frac{1}{n_E}\sum_{i=1}^{n_E}\widehat{\psi}_E(Z_{E,i})^2 + \frac{N}{n_O}\frac{1}{n_O}\sum_{j=1}^{n_O}\widehat{\psi}_O(Z_{O,j})^2\end{aligned} \tag{3.7}$$

并令 Σ 和 $\widehat{\Sigma}$ 为它们的非负平方根。标准误为 $\widehat{\Sigma}/\sqrt{N}$。

3.3 渐近正态性与效率

本小节陈述关于全样本估计量的两个条件以及主要定理。我们保留第2节中关于识别的条件以及假设 3.1–3.4。

涉及估计量的函数的期望是在样本固定的情况下，通过关于总体的条件分布进行积分得到的。例如，$\mathbb{E}[\widehat{\nu}_t(S,x)\widehat{\omega}_t^{\nu}(M,x) \mid T=t, x, E] = \int \widehat{\nu}_t(s,x)\widehat{\omega}_t^{\nu}(m,x)\, dF_{S,M\mid T=t,x,E}(s,m)$，并且 $\mathbb{E}[\cdot \mid x, O]$ 表示关于

$F_{T,M,Y|x,O}$的积分。给定 M 或 S 的条件期望同样是在样本固定的情况下，根据相同的条件分布得到的。对于关于 (S,M) 的函数 f，均方范数为 $\|f\|_{2,t,x}=\mathbb{E}[f(S,M)^2\mid T=t,x,E]^{1/2}$，我们将 $(\widehat{\mathbb{E}}-\mathbb{E})[\cdot\mid T=t,x,E]=\widehat{\mathbb{E}}[\cdot\mid T=t,x,E]-\mathbb{E}[\cdot\mid T=t,x,E]$ 记作样本均值与该期望之差，而 $(\mathbb{E}-\mathbb{E})[\cdot\mid x,O]$ 则是其在观察样本中的对应形式。

下面的假设 3.5 根据估计误差的速率，界定了在固定估计量，并将（3.5）中的两个样本均值替换为期望后，仍然存在的偏离 $\bar{\mu}_t(x)$ 的程度。假设 3.6 则界定了在用于估计这些函数的同一样本上，对代入后的函数进行平均所产生的误差。

假设 3.5（干扰函数的估计）。对于每个 t、x，第 *3.2* 节的全样本估计量 $\widehat{\pi}_t\ \widehat{\nu}_t\ \widehat{\omega}_t^{\nu}\ \widehat{\omega}_t^{\pi}$ 满足以下条件。

1. $\|\widehat{\pi}_t-\pi_t\|_{2,\ t,\ x}\to_p 0$、$\|\widehat{\nu}_t-\nu_t\|_{2,\ t,\ x}\to_p 0$、$\|\widehat{\omega}_t^{\nu}-\omega_t^{\nu}\|_{2,\ t,\ x}\to_p 0$，以及 $\|\widehat{\omega}_t^{\pi}-\omega_t^{\pi}\|_{2,\ t,\ x}\to_p 0$。

2. $\|\widehat{\pi}_t-\pi_t\|_{2,t,x}\times\{\|\widehat{\pi}_t-\pi_t\|_{2,t,x}+\|\widehat{\nu}_t-\nu_t\|_{2,t,x}\}=o_p(N^{-1/2})$.

3. ω_t^{ν} 的乘积满足

$$\min\Big\{\|\mathbb{E}[\widehat{\nu}_t(S,x)-\nu_t(S,x)\mid M,T=t,x,E]\|_{2,t,x}\times\|\widehat{\omega}_t^{\nu}-\omega_t^{\nu}\|_{2,t,x},\ \|\widehat{\nu}_t-\nu_t\|_{2,t,x}\times\|\mathbb{E}[\widehat{\omega}_t^{\nu}(M,x)-\omega_t^{\nu}(M,x)\mid S,T=t,x,E]\|_{2,t,x}\Big\}=o_p(N^{-1/2})$$

4. ω_t^{π} 的乘积满足

$$\min\Big\{\|\mathbb{E}[\widehat{\pi}_t(S,x)-\pi_t(S,x)\mid M,T=t,x,E]\|_{2,t,x}\times\|\widehat{\omega}_t^{\pi}-\omega_t^{\pi}\|_{2,t,x},\ \|\widehat{\pi}_t-\pi_t\|_{2,t,x}\times\|\mathbb{E}[\widehat{\omega}_t^{\pi}(M,x)-\omega_t^{\pi}(M,x)\mid S,T=t,x,E]\|_{2,t,x}\Big\}=o_p(N^{-1/2})$$

假设 3.6（对估计样本取平均）。对于每个 t,x，第 *3.2* 节的全样本估计量和 $\widehat{\mu}_t=\widehat{\nu}_t/\widehat{\pi}_t$ 满足以下条件。

1. $(\widehat{\mathbb{E}}-\mathbb{E})[\widehat{\mu}_t(S,x)\ -\mu_t(S,x)\ \mid T=t、\ x、\ E]=o_p(N^{-1/2})$。2. $(\widehat{\mathbb{E}}-\mathbb{E})[\widehat{\nu}_t(S,x)\widehat{\omega}_t^{\nu}(M,x)\ -\nu_t(S,x)\omega_t^{\nu}(M,x)\ \mid T=t、\ x、\ E]=o_p(N^{-1/2})$。3. $(\widehat{\mathbb{E}}-\mathbb{E})[\widehat{\pi}_t(S,x)\widehat{\omega}_t^{\pi}(M,x)\ -\pi_t(S,x)\omega_t^{\pi}(M,x)\ \mid T=t、\ x、\ E]=o_p(N^{-1/2})$。4. $(\widehat{\mathbb{E}}-\mathbb{E})[\mathbf{1}\{T=t\}\{\widehat{\omega}_t^{\nu}(M,x)\ -\omega_t^{\nu}(M,x)\}\ \mid x、\ O]=o_p(N^{-1/2})$。5. $(\widehat{\mathbb{E}}-\mathbb{E})[\mathbf{1}\{T=t\}\{\widehat{\omega}_t^{\pi}(M,x)\ -\omega_t^{\pi}(M,x)\}\ \mid x、\ O]=o_p(N^{-1/2})$。6. $\widehat{\mathbb{E}}[\{\widehat{\pi}_t(S,x)\ -\pi_t(S,x)\}^2\mid T=t、\ x、\ E]\to_p 0$。

7. $\widehat{\mathbb{E}}[\{\widehat{v}_t(S,x) - v_t(S,x)\}^2 \mid T = t、\ x、\ E] \to_p 0$。
8. $\widehat{\mathbb{E}}[\{\widehat{\omega}_t^{\nu}(M,x) - \omega_t^{\nu}(M,x)\}^2 \mid T = t、\ x、\ E] \to_p 0$。
9. $\widehat{\mathbb{E}}[\{\widehat{\omega}_t^{\pi}(M,x) - \omega_t^{\pi}(M,x)\}^2 \mid T = t、\ x、\ E] \to_p 0$。
10. $\widehat{\mathbb{E}}[\mathbf{1}\{T = t\}\{\widehat{\omega}_t^{\nu}(M,x) - \omega_t^{\nu}(M,x)\}^2 \mid x、$ $O] \to_p 0$。11. $\widehat{\mathbb{E}}[\mathbf{1}\{T = t\}\{\widehat{\omega}_t^{\pi}(M,x) - \omega_t^{\pi}(M,x)\}$ $^2 \mid x、\ O] \to_p 0$。

假设 3.5 的条件 (1) 要求干扰函数估计具有一致性，但不对收敛速率作任何具体要求。条件 (2)–(4) 采用了 Chernozhukov 等人 (2018 年的乘积速率形式。当估计值固定，并将 (3.5) 中的两个样本均值替换为期望时，这些用于衡量偏离 $\bar{\mu}_t(x)$的乘积以快于$N^{-1/2}$的速度趋于零。由于 (3.5) 中的矩对干扰函数的一阶误差不敏感，因此只出现误差的乘积；例如，当乘积的两个因子都为$o_p(N^{-1/4})$时，每个条件均成立。因此，可以使用以如此缓慢速率收敛的非参数估计量和机器学习估计量，对 τ进行估计与推断。

假设 3.6 包含条件 (1)–(5)，这些条件针对代入函数的样本均值与估计值固定时其期望之间的差异；还包含条件 (6)–(11)，这些条件针对估计误差平方的样本均值。条件 (1)–(5) 直接对估计量施加经验过程理论中的随机等度连续性要求（vander Vaart 1998，第 19 章）。根据 vander Vaart(1998) 第 19 章关于 Donsker 类的结果，并结合假设 3.5(1) 中的一致性，如果在概率趋于 1 的事件上，每个代入函数与其层的指示函数相乘后都属于相应总体的一个固定 Donsker 类，且该类包含对应的真实函数，那么这些条件就成立。

条件 (6)–(9) 要求实验分层内四个估计量的均方误差，以及条件 (10) 和 (11) 要求观测值中两个权重估计的均方误差 $T = t$ 在观察性分层内依概率收敛于零。它们用于保证方差估计量的一致性。根据 van derVaart (1998) 第 19 章中的一致大数定律以及假设 3.5(1)，如果在概率趋于 1 的事件上，每个乘以其所在层的指示函数的平方误差都属于相应总体的某个固定 Glivenko–Cantelli 类，则这些条件成立。

级数估计是满足假设 3.5 的乘积速率条件的估计量的一个例子，但还需要满足额外的正则条件。方程 (2.7) 中关于 π_t 和 ν_t 的方程具有非参数工具变量模型的形式，其中 M 充当工具变量；权重函数的方程 (3.2) 也具有相同的形式，只是 S 和 M的作用互换

，且右侧经过估计。此类模型的 L^2 级数估计和惩罚筛估计量的收敛速率（Blundell, Chen, andKristensen 2007；Chen and Pouzo 2012；Chenand Christensen 2018）构成了充分条件的基础。在未知函数具有足够光滑性且病态性程度适中的情况下，这些速率为o_p（$N^{-1/4}$），其中级数项数以适当方式增长；将这些结果应用于此处，还需要控制 (3.2) 的估计右侧中的误差，并确定 (3.2) 的解，使权重估计趋近于该解。对于不进行样本拆分的估计量，还必须单独检验假设 3.6。

定理 3.2（不进行样本拆分的估计量的渐近分布与效率）。假设假设 *2.1–2.8* 和 *3.1–3.6* 均成立，且满足附录 A 中的可积性条件以及附录 B 和 C 中的有界性条件。则

$$\sqrt{N}(\widehat{\tau} - \tau) \Rightarrow \mathcal{N}(0, \Sigma^2), \qquad \widehat{\Sigma}^2 \to_p \Sigma^2 \tag{3.8}$$

若 $\Sigma^2 > 0$，则 $\sqrt{N}(\widehat{\tau} - \tau)/\widehat{\Sigma} \Rightarrow \mathcal{N}(0, 1)$，并且对于 $\alpha \in (0, 1)$ 以及标准正态分布的 $1 - \alpha/2$ 分位数 $z_{1-\alpha/2}$，$\widehat{\tau} \pm z_{1-\alpha/2}\widehat{\Sigma}/\sqrt{N}$ 的覆盖概率收敛至 $1 - \alpha$。此外，若对于每个 t 和 x，*(3.2)* 都有唯一的平方可积解，且附录 B 中的切空间条件成立，则 $\widehat{\tau}$ 在模型 $\mathcal{P}$ 中相对于附录 B 的子模型是正则的，并达到定理 *3.1* 的效率界 Σ^2。

估计误差可近似表示为两个样本均值之和，$n_E^{-1}\sum_{i=1}^{n_E}\psi_E(Z_{E,\,i}) + n_O^{-1}\sum_{j=1}^{n_O}\psi_O(Z_{O,\,j})$，而 Σ^2 是这两个贡献之和。在定理 3.1 的条件下（其中包括解 (3.2) 的唯一性），这些加项构成有效影响函数，因此该估计量达到了效率界。当解不唯一时，渐近分布和方差估计量仍然有效（注 3.1）。证明见附录 C。

假设 3.6 背后的 Donsker 条件和 Glivenko–Cantelli条件可能不适用于非参数估计量或机器学习估计量类。 Chernozhukov等人 （2018年）提出了交叉拟合方法以避免这些条件，我们的估计量也可以采用相同的方式计算，正如第3.4节所述。

3.4 使用交叉拟合进行估计

本小节定义了一种估计量：将两个样本划分为若干折，并在每一层内对某一折中的观测值取平均，而在估计干扰函数时不使用该折中的观测值；同时，本小节还表明，在满足假设 3.7 的一致性和速率条件下，定理 3.2 的结论仍然成立

对于固定整数 $K \geq 2$，我们将 $\{1, \ldots, n_E\}$ 划分为 K 个大小至多相差一个元素的不相交集合 $\mathrm{I}_{E,1}, \ldots, \mathrm{I}_{E,K}$，并以相同方式将 $\{1, \ldots, n_O\}$ 划分为 $\mathrm{I}_{O,1}, \ldots, \mathrm{I}_{O,K}$。这些索引集是非随机的，且不依赖于观测值。第 k 折由 $Z_{E,i}, i \in \mathrm{I}_{E,k}$ 和 $Z_{O,j}, j \in \mathrm{I}_{O,k}$ 组成。我们将折 k 内条件均值和份额的相应量分别记为 $\mathbb{E}_k$ 和 P_k。例如，$\mathbb{E}_k[f(S,M) \mid T=t, x, E]$ 是对满足 $i \in \mathrm{I}_{E,k}$，$T_i = t$ 的观测值中的 $f(S_i, M_i)$ 取平均，而 $X_i = x$ 和 $P_k(T=t, X=x \mid E)$ 则是在 $i \in \mathrm{I}_{E,k}$ 中同时满足 $T_i = t$ 和 $X_i = x$ 的观测值所占的份额。

对于每个 k，我们仅使用折 k 之外的观测值来估计 π_t，ν_t，ω_t^{ν}，ω_t^{π}，并将这些估计量记为 $\widehat{\pi}_t^{(-k)}$，$\widehat{\nu}_t^{(-k)}$，$\widehat{\omega}^{\nu,(-k)}_{\cdot,t}$，$\widehat{\omega}^{\pi,(-k)}_{\cdot,t}$，其中上标 $(-k)$ 表示省略了折 k。与第 3.2 节中一样，$\widehat{\pi}_t^{(-k)}$ 截断为 $[a, 1]$，且 $\widehat{\mu}_t^{(-k)} = \widehat{\nu}_t^{(-k)} / \widehat{\pi}_t^{(-k)}$。固定这些估计量后，我们计算折 k 上的分层均值。对于每个 t，x，k，令

$$\begin{aligned}\widehat{\tilde{\mu}}_{t,k}(x) = \widehat{\mathbb{E}}_k[\widehat{\mu}_t^{(-k)}(S,x) - \widehat{\nu}_t^{(-k)}(S,x)\widehat{\omega}_t^{\nu,(-k)}(M,x) + \widehat{\pi}_t^{(-k)}(S,x)\widehat{\omega}_t^{\pi,(-k)}(M,x) \mid T=t, x, E] \\ + \widehat{\mathbb{E}}_k[\mathbf{1}\{T=t\}\{Y\widehat{\omega}_t^{\nu,(-k)}(M,x) - \widehat{\omega}_t^{\pi,(-k)}(M,x)\} \mid x, O]\end{aligned} \tag{3.9}$$

并通过将 $\widehat{\tilde{\mu}}_t(x) = K^{-1}\sum_{k=1}^{K}\widehat{\tilde{\mu}}_{t,k}(x)$ 代入 (3.6)，定义平均处理效应的估计量 $\widehat{\tau}$。对于标准误，我们按第 3.2 节中的替换方式，对 $i \in \mathrm{I}_{E,k}$ 计算 $\psi_E(Z_{E,i})$，并对 $j \in \mathrm{I}_{O,k}$ 计算 $\psi_O(Z_{O,j})$；用 $\widehat{\mu}_t^{(-k)}, \widehat{\nu}_t^{(-k)}, \widehat{\pi}_t^{(-k)}, \widehat{\omega}^{\nu,(-k)}_{\cdot,t}, \widehat{\omega}^{\pi,(-k)}_{\cdot,t}$ 替代全样本估计量，并在 (3.7) 中使用 $\widehat{\Sigma}^2$。在本小节中，$\widehat{\tilde{\mu}}_t(x), \widehat{\tau}, \widehat{\psi}_E, \widehat{\psi}_O$ 和 $\widehat{\Sigma}^2$ 表示这些交叉拟合量。涉及估计量的函数的期望和范数是在样本固定的条件下计算的，与第 3.3 节中一样。

假设 3.7 (留出一个折的估计量)。对于每个 k、t 和 x，估计量 $\widehat{\pi}_t^{(-k)}$ $\widehat{\nu}_t^{(-k)}$ $\widehat{\omega}^{\nu,(-k)}_{\cdot,t}$ $\widehat{\omega}^{\pi,(-k)}_{\cdot,t}$ 满足以下条件。

1. $\|\widehat{\pi}_t^{(-k)} - \pi_t\|_{2,\,t,\,x} \to_p 0$、$\|\widehat{\nu}_t^{(-k)} - \nu_t\|_{2,\,t,\,x} \to_p 0$、$\|\widehat{\omega}^{\nu,(-k)}_{\cdot,t} - \omega_t^{\nu}\|_{2,\,t,\,x} \to_p 0$，以及 $\|\widehat{\omega}^{\pi,(-k)}_{\cdot,t} - \omega_t^{\pi}\|_{2,\,t,\,x} \to_p 0$。

2. $\|\widehat{\pi}_t^{(-k)} - \pi_t\|_{2,\,t,\,x} \times \{\|\widehat{\pi}_t^{(-k)} - \pi_t\|_{2,\,t,\,x} + \|\widehat{\nu}_t^{(-k)} - \nu_t\|_{2,\,t,\,x}\} = o_p(N^{-1/2})$。

3. ω_t^{ν} 的乘积满足

$$\begin{aligned}\min\Big\{&\|\mathbb{E}[\widehat{\nu}_t^{(-k)}(S,x) - \nu_t(S,x) \mid M, T=t, x, E]\|_{2,t,x} \times \|\widehat{\omega}_t^{\nu,(-k)} - \omega_t^{\nu}\|_{2,t,x}, \\ &\|\widehat{\nu}_t^{(-k)} - \nu_t\|_{2,t,x} \times \|\mathbb{E}[\widehat{\omega}_t^{\nu,(-k)}(M,x) - \omega_t^{\nu}(M,x) \mid S, T=t, x, E]\|_{2,t,x}\Big\} = o_p(N^{-1/2})\end{aligned}$$

4. ω_t^π的乘积满足

$$\min\Big\{\|\mathbb{E}[\widehat{\pi}_t^{(-k)}(S,x)-\pi_t(S,x)\mid M,T=t,x,E]\|_{2,t,x}\times\|\widehat{\omega}_t^{\pi,(-k)}-\omega_t^\pi\|_{2,t,x},$$
$$\|\widehat{\pi}_t^{(-k)}-\pi_t\|_{2,t,x}\times\|\mathbb{E}[\widehat{\omega}_t^{\pi,(-k)}(M,x)-\omega_t^\pi(M,x)\mid S,T=t,x,E]\|_{2,t,x}\Big\}=o_p(N^{-1/2})$$

假设 3.7 的条件 (1)–(4) 是针对省略每一折后的估计量而表述的假设 3.5 的条件 (1)–(4)，下一定理将使用这些条件来替代假设 3.5 和假设 3.6。

定理 3.3（交叉拟合估计量的渐近分布与效率）。 假设假设 *2.1–2.8*、*3.1–3.4* 和 *3.7* 均成立，同时满足附录 A 中的可积性条件以及附录 B 和 C 中的有界性条件。则对于估计量 $\widehat{\tau}$ 以及第*3.4* 节的方差估计量 $\widehat{\Sigma}^2$，*(3.8)* 成立。若 $\Sigma^2>0$，则 $\sqrt{N}\,(\widehat{\tau}-\tau)/\widehat{\Sigma}\Rightarrow\mathcal{N}(0,1)$；对于 $\alpha\in(0,1)$ 以及标准正态分布的$1-\alpha/2$ 分位数 $z_{1-\alpha/2}$，$\widehat{\tau}\pm z_{1-\alpha/2}\widehat{\Sigma}/\sqrt{N}$的覆盖概率收敛于 $1-\alpha$。此外，若对于每个 t 和 x，*(3.2)* 都有唯一的平方可积解，并且附录 B 中的切空间条件成立，则 $\widehat{\tau}$ 在模型 $\mathcal{P}$ 中相对于附录 B 的子模型是正则的，并达到定理 *3.1* 的效率界 Σ^2。

在定理 3.3 中，每一折与不包含该折数据所计算出的估计量之间的独立性，取代了假设 3.6。在估计量固定的条件下，假设 3.6 中在第 k 折上计算的每个量，都是独立观测值的平均值与其期望之差，而假设 3.7(1) 中的一致性使这一差异足够小。将基于其他折估计的干扰函数代入对其中一阶误差不敏感的矩条件中，这遵循了 Chernozhukov等人（2018年）提出的双重/去偏机器学习方法；他们将此类矩条件称为 Neyman 正交矩条件，并将其与交叉拟合相结合。证明见附录 C。

4 模拟研究

本节将所提出的估计量的三个版本，与三种使用共享测量代替特征或仅使用观察样本的估计量进行比较。所考察的设计中，观察性总体中的处理选择取决于特征，而共享测量则带有误差地记录特征。第 4.1 节介绍这些设计，第 4.2 节介绍估计量，第 4.3 节报告结果。补充材料 2 详述了计算过程。

4.1 设计

本小节介绍两种数据生成设计。在每次重复实验中，我们分别独立抽取一个样本量为 n_E 的实验样本和一个样本量为 n_O 的观察样本，其中样本量取值为$n_E = n_O \in \{1，000，2，000、4，000、8，000\}$，且每种样本量均进行 1,000 次重复。协变量$X$ 在实验总体中以相同的 1/4 概率分别取值 0、1、2、3；在观察性总体中，取值概率依次为 0.3、0.3、0.2、0.2。在两个总体中，X、U_1、U_2和 U_3 相互独立，且 U_2 和 U_3 在 $[-1, 1]$上服从均匀分布。两个总体中的潜在值以及观察性总体中的处理由以下方式生成：

$$
\begin{aligned}
S(t) &= U_1 + (1 + 0.5X)t, \\
M(t) &= S(t) + U_2, \\
Y(t) &= 2S(t) + 5U_1 + U_3, \\
P(T = 1 \mid U_1, U_2, U_3, X, O) &= 1/2 + \kappa U_1
\end{aligned}
\tag{4.1}
$$

并且在实验总体中，T与（X，U_1，U_2，U_3）相互独立，且 $P(T = 1 \mid E) = 1/2$。实验样本包含（X，T，M，S），观察样本包含（X，T，M，Y）。

两种设计在 U_1 的分布以及选择强度 κ 方面存在差异。在连续设计中，U_1 在 $[-1, 1]$ 和 $\kappa = 1/10$上服从均匀分布。在离散设计中，U_1 取 −1、0和1三个值，每个值的概率均为1/3，且 $\kappa = 1/4$，因此该特征在每个单元格内取三个值，即 X和组别的每个取值。处理对长期结果的影响为 $Y(1) - Y(0) = 2\{S(1) - S(0)\}$，对每个个体而言，等于其对该特征影响的两倍，因此两种设计中的目标参数（2.1）均为 $\tau = 2 + \mathbb{E}[X \mid O] = 2 + (0。3 + 0。4 + 0。6) = 3。30$。

两种设计都是例 2.1 中测量误差模型的特殊情形。因此，它们满足假设 2.1–2.7；由于 U_2 的特征函数仅在孤立点处为零，假设 2.8 中的完备性成立（附录 A.2）。选择概率介于 $[1/2 - \kappa, 1/2 + \kappa]$ 之间，因此假设 3.3 在截断常数为 $a = 0.05$ 时成立。在这两种设计中，都存在满足（3.2）的有界权重函数，但它们并不唯一（附录 D.1）。

4.2 估计量

我们报告所提出的估计量的三个版本，均使用干扰函数的级数估计计算。Proposed（Tikhonov）是第 3.2 节的估计量，其级数估计包含 Tikhonov惩罚。Proposed（Tikhonov，交叉拟合）是第3.4节的估计量，其中相同的级数估计使用每一折之外的观测值计算；我们报告了采用 $K = 2$和 $K = 4$ 折的结果。Proposed（未惩罚）使用相同的级数估计计算第 3.2 节的估计量，并将惩罚项设为零，定义如下。该

基 b_J 在 S 中由连续设计中次数至多为 4 的单项式，以及离散设计中次数至多为 2 的单项式构成；在离散设计中，它张成了该特征三个取值上的所有函数，因此$J = 5$ 和 $J = 3$。检验函数 φ_L 是 M中次数至多为 5 的单项式，因此 $L = 6$；两个变量均在实验样本中进行了标准化。

我们通过级数方法，以 φ_L 的各个分量作为（2.7）中的检验函数，以相同方式估计选择概率 π_t 和乘积 ν_t 。对于 $\pi_t(s, x) = b_J(s)^\top \beta$ 和 $\nu_t(s, x) = b_J(s)^\top \gamma$，（2.7）的两个方程变为 $\mathbb{E}[\mathbf{1}\{T = t\}\varphi_L(M) \mid x, O] = Q_t(x)\beta$ 以及 $\mathbb{E}[\mathbf{1}\{T = t\}Y\varphi_L(M) \mid x, O] = Q_t(x)\gamma$，其中 $Q_t(x) = \mathbb{E}[\varphi_L(M)b_J(S)^\top \mid T = t、 x、 E]$ 是两个具有相同矩阵的线性$\widehat{\quad}$ $\widehat{\quad}$ $\widehat{\quad}$系统。利用样本均值 $Q_t(x)$ 和 $V(x) = \mathbb{E}[\varphi_L(M)\varphi_L(M)^\top \mid x, O]$，Tikhonov版本使用正则化矩阵估计两个系数向量

$$\widehat{A}_{t,\lambda}(x) = \{\widehat{Q}_t(x)^\top \widehat{V}(x)^{-1}\widehat{Q}_t(x) + \lambda I\}^{-1}\widehat{Q}_t(x)^\top \widehat{V}(x)^{-1} \tag{4.2}$$

$\widehat{\quad}$其中， I 是 $J \times J$ 单位矩阵，而 $\lambda > 0$ 是一个惩罚项。 β 和 γ 的估计量为$A_{t,\ \lambda}(x)$乘以单元格 x中观测单元的$\mathbf{1}\{T = t\}\varphi_L(M)$和$\mathbf{1}\{T = t\}Y\varphi_L(M)$的样本均值；同一个矩阵也给出了两个权重函数的估计值。这两个估计值分别通过最小化其矩方程的加权残差平方来求得，权重为 $\widehat{V}(x)^{-1}$，并加上系数平方范数的 λ 倍。惩罚项从 10^{-6}， 10^{-5}， $\ldots$， 10^{0}中选取，方法是对矩条件进行二折验证，并对所有单元格和组别使用同一个取值；交叉拟合版本则根据每一折之外的观测值来选择该取值。Proposed（未惩罚）使用含有 $\lambda = 0$的矩阵（4.2），这要求 $\widehat{Q}_t(x)^\top \widehat{V}(x)^{-1}\widehat{Q}_t(x)$可逆。三个版本的标准误均采用（3.7）中的标准误；对于交叉拟合版本，则按照第3.4节中的方法计算。

三个比较估计量使用共享测量来替代特征或单独使用观察样本，并使用所提出的估计量的单元格和检验函数计算这些估计量φ_L 。ESC 是 Athey、Chetty 和 Imbens （2025年）提出的实验选择校正方法，是一种控制函数方法，其中以 M 替代其短期结果变量。其控制变量是实验样本中相同单元格和组别内 M 的经验分布函数，并在观测到的 M处进行评估。在观察样本的每个单元格和组别内，ESC 将 Y 对该控制变量的基进行回归。ESC 插补将 Y 对φ_L（M）在观察样本的每个单元格和组别内进行回归，并在实验样本的相同单元格和组别上对拟合值取平均。观测回归是观察样本中 T 对 Y 截距、 T和单元指示变量进行回归时对应的最小二乘系数。与所提出的估计量一样，它们的标准误根据其影响函数的估计值计算。

4.3 结果

表 2 和表 3 分别报告了连续设计和离散设计下，各估计量在为其保留的重复试验中的误差，以及其 95% 区间的覆盖率。表下注释定义了各列。

在连续设计中（表2），Tikhonov版本的绝对偏差从样本量为 $n_E = n_O = 1,000$ 时的 0.519 降至 8,000 时的 0.074，而 ESC 和 ESC 插补的偏差始终约为 0.47，观测回归在每个样本量下的偏差均约为 0.94。其 95% 区间的覆盖率在样本量为 1,000 时为 78.8%，在较大的样本量下介于 86.5% 和 89.6% 之间；相比之下，比较估计量的覆盖率随着样本量增加而趋近于零。在离散设计中（表3），Tikhonov版本在每个样本量下都具有七个估计量中最小的绝对偏差：样本量为 1,000 时为 0.422，在较大的样本量下小于 0.04；其区间的覆盖率则从样本量为 1,000 时的 83.1% 上升至 8,000 时的 92.8%。比较估计量的偏差绝对值至少为 1.54，其覆盖率最高为 2.1%。图 2 展示了两种设计下 Tikhonov版本的估计值分布。大多数估计值位于 τ 附近，而在较小的样本量下，少数远低于 τ 的估计值将估计值的均值拉低。

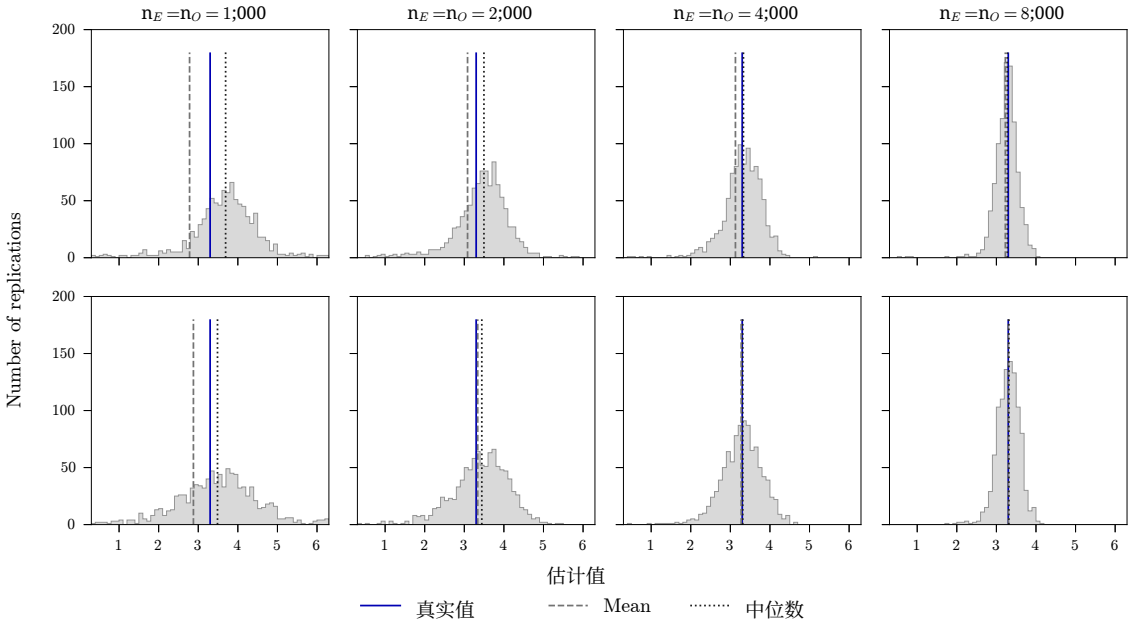

图 2 共享测量设计中 Proposed（Tikhonov）估计量的分布注：每个面板展示了某一样本量下，在连续设计（表2）或离散设计（表3）中对 Proposed（Tikhonov）进行的 1,000 次估计。直方图统计了介于 0.30 和 6.30 之间、位于真实效应 $\tau = 3$ 的 3 个单位以内的估计量，组距为 0.1。实线表示 τ，虚线和点线分别表示全部 1,000 个估计量的均值和中位数。

另外两个版本相较于 Tikhonov 版本几乎没有改进。交叉拟合并未一致地提升其表现。在两种设计中，采用两折和四折的交叉拟合版本

$n_E = n_O$　估计量	Bias	标准差	均方根误差	标准误	覆盖率
1,000 所提出的方法（Tikhonov）	−0.519	9.233	9.243	3.849	78.8
1,000 所提出的方法（Tikhonov，交叉拟合，2折）	6.819	87.249	87.472	10.720	70.0
1,000 所提出的方法（Tikhonov、交叉拟合、4折）	1.175	42.690	42.684	7.963	83.3
1,000 ESC	0.479	0.270	0.550	0.266	56.9
1,000 ESC 插补	0.484	0.273	0.555	0.269	56.8
1,000　观测回归	0.946	0.261	0.981	0.259	4.8
2,000 Proposed（Tikhonov）	−0.218	3.106	3.112	0.943	87.0
2,000 Proposed（Tikhonov，交叉拟合，2折）	1.667	25.745	25.786	2.560	70.6
2,000 Proposed（Tikhonov，交叉拟合，4折）	1.272	7.484	7.587	1.914	86.8
2,000 ESC	0.468	0.188	0.504	0.187	31.8
2,000 ESC 插补	0.469	0.188	0.506	0.187	30.6
2,000　观测回归	0.936	0.181	0.953	0.183	0.0
4,000 所提出的方法（Tikhonov）	−0.171	1.684	1.692	0.559	86.5
4,000 所提出的方法（Tikhonov、交叉拟合、2折）	0.813	5.551	5.607	0.915	76.6
4,000 所提出的方法（Tikhonov、交叉拟合、4折）	0.775	5.774	5.823	0.902	86.6
4,000 ESC	0.470	0.134	0.489	0.132	5.5
4,000 ESC 插补	0.470	0.134	0.489	0.132	5.5
4,000　观测回归	0.929	0.129	0.938	0.129	0.0
8,000 Proposed（Tikhonov）	−0.074	0.562	0.567	0.254	89.6
8,000 Proposed（Tikhonov，交叉拟合，2折）	0.406	2.114	2.151	0.430	77.7
8,000 Proposed（Tikhonov，交叉拟合，4折）	0.454	2.905	2.938	0.576	86.0
8,000 ESC	0.469	0.094	0.479	0.093	0.1
8,000 ESC 插补	0.469	0.094	0.479	0.093	0.1
8,000　观测回归	0.935	0.095	0.940	0.091	0.0

表2 连续设计中估计量的误差和覆盖率

注：真实效应为 $\tau = 3$。30，且每个样本量均进行了 1,000 次重复。偏差是估计量的均值减去 τ，标准差是估计量的标准差，均方根误差是估计量与 τ 之差的平方的均值的平方根。标准误是标准误的均值，覆盖率是指在多少比例的重复试验中，估计值加减 1.96 个标准误的区间包含 τ。所提出的估计量的三个版本的标准误由公式 (3.7) 给出，而比较估计量的标准误则根据其影响函数的估计值计算。未显示 Proposed（未惩罚），因为在此设计中，它所求逆的矩阵 $\widehat{Q}_t(x)^\top \widehat{V}(x)^{-1} \widehat{Q}_t(x)$ 近乎奇异。在 $n_E = n_O = 1$，000、2,000、4,000 和 8,000 的样本量下，该矩阵在 1,000 次重复试验中，分别有 990、999、999 和 1,000 次在某个单元格和组别中的条件数至少为 10^{10}。

$n_E = n_O$	估计量	Bias	标准差	均方根误差	标准误	覆盖率
1,000	所提出的方法（Tikhonov）	−0.422	5.548	5.562	1.259	83.1
1,000	所提出的方法（Tikhonov，交叉拟合，2折）	3.500	10.067	10.653	2.274	48.6
1,000	所提出的方法（Tikhonov、交叉拟合、4折）	1.940	3.826	4.288	1.724	63.6
1,000	所提出的方法（未惩罚）	−15.701	22.725	27.612	20.968	99.5
1,000	ESC	1.551	0.403	1.602	0.395	2.1
1,000	ESC 插补	1.558	0.429	1.616	0.376	1.4
1,000	观测回归	4.669	0.332	4.681	0.334	0.0
2,000	Proposed（Tikhonov）	0.039	1.118	1.119	0.591	84.0
2,000	Proposed（Tikhonov，交叉拟合，2折）	1.670	8.613	8.770	1.094	64.0
2,000	Proposed（Tikhonov，交叉拟合，4折）	1.067	2.814	3.008	0.939	71.1
2,000	Proposed（未惩罚）	−3.841	5.992	7.114	5.546	99.9
2,000	ESC	1.556	0.279	1.581	0.269	0.0
2,000	ESC 插补	1.553	0.271	1.576	0.260	0.0
2,000	观测回归	4.682	0.237	4.688	0.236	0.0
4,000	Proposed（Tikhonov）	−0.026	0.608	0.608	0.381	85.7
4,000	所提出的方法（Tikhonov、交叉拟合、2折）	0.727	1.700	1.848	0.491	64.6
4,000	所提出的方法（Tikhonov、交叉拟合、4折）	0.437	1.053	1.140	0.430	78.8
4,000	Proposed（未惩罚）	−1.076	1.519	1.861	1.381	99.8
4,000	ESC	1.556	0.190	1.568	0.189	0.0
4,000	ESC 插补	1.546	0.184	1.557	0.183	0.0
4,000	观测回归	4.675	0.164	4.678	0.167	0.0
8,000	Proposed（Tikhonov）	0.000	0.295	0.295	0.255	92.8
8,000	Proposed（Tikhonov，交叉拟合，2折）	0.292	0.346	0.453	0.270	77.6
8,000	Proposed（Tikhonov，交叉拟合，4折）	0.205	0.448	0.493	0.272	86.1
8,000	Proposed（未惩罚）	−0.376	0.452	0.588	0.498	98.2
8,000	ESC	1.550	0.131	1.556	0.133	0.0
8,000	ESC 插补	1.541	0.128	1.546	0.129	0.0
8,000	观测回归	4.659	0.115	4.661	0.118	0.0

表3 离散设计中估计量的误差和覆盖率

注：各项的定义与表2相同。对每个估计量和样本量均使用全部1,000 次重复。

在每个样本量下，其绝对偏差都大于未进行样本拆分的版本，且在8,000时其覆盖率也更低，介于77.6%与86.1%之间。无惩罚版本仅在离散设计的4,000和8,000时，绝对偏差小于全部三个比较估计量。Tikhonov版本在每次重复实验中均可计算，并且在两种设计的每个样本量下，其绝对偏差都小于全部三个比较估计量，唯独连续设计中样本量为1,000时除外：此时少数极端估计量拉动了其均值（图 2），而ESC和ESC 插补的绝对偏差略小。因此，这是我们推荐的版本。

5 使用 JobCorps 数据进行实证示例

本节使用 National Job Corps Study 的数据，在第2节的设计下说明第 3 节中的估计量。第5.1节介绍数据以及实验样本和观察性研究样本的构建，第 5 节.2介绍估计量，并将其误差与根据实验计算出的伪真实值进行比较。补充材料 2详细介绍样本的构建和计算过程。

5.1 数据与样本构建

本小节介绍数据、变量以及两个样本的抽取方式。我们的数据来自 Job Corps 的随机评估公开数据。Job Corps 是美国面向弱势青年的培训项目，其入学提供方案采用随机分配（Schochet、Burghardt 和 Glazerman，2001年）。这些数据随 R软件包causalweight 一同分发。在公开样本的 9,240 名人员中，我们使用其中教育年限不缺失的 9,084 名人员；教育年限是第 5 节.2比较回归中的一个协变量。在这些人员中，5,483 人被分配接受处理，3,601 人被分配到对照组。

我们将第4年周收入作为结果变量，并通过对数据中记录的两个变量进行离散化，获得特征和共享测量。目标参数是提供方案 $T \in \{0, 1\}$ 对第4年周收入的平均效应 Y。特征 $S \in \{1, 2, 3\}$ 是某人记录的第3年周收入类别，其中，$S = 1$表示第3年没有收入，$S = 2$表示每周收入为正且不超过9,084 人中第3年正周收入的中位数（约188 美元），$S = 3$ 表示每周收入高于该中位数。共享测量 M 是记录的第3年工作周占比水平，用于记录某人在第3年没有工作过任何一周、工作周占比为正且不超过9,084 人中正工作周占比的中位数（约73%），或工作周占比更高。因而，这两个变量都是将每个人的记录值粗化为三个类别。处理前变量 X 取六个值，即由随机分配时的性别和三个年龄组（16–17岁、18–19岁和20–24岁）形成的单元格，这些年龄组是

对于这些年龄组，Schochet、Burghardt 和 Glazerman（2001年）分别报告了影响估计量。三个收入类别分别包含5,483名处理组人员中的905人、2,260人和2,318人，以及3,601名对照组人员中的667人、1,496人和1,438人；的三个水平分别包含处理组中的898人、2,284人和2,301人，以及对照组人员中的663人、1,515人和1,423人。

由于 Job Corps数据来自一项随机实验，我们构造了一个观察样本，其中处理选择取决于特征。在每个单元格内，提供方案的分配独立于潜在值，因此，从9,084 人中以相等概率抽取的样本将构成另一项实验；在该实验中，通过比较处理组和对照组人员，无需任何关于 S的信息即可估计处理效应。因此，我们按照取决于处理和收入类别 S的概率抽取观察样本，而观察样本不记录该收入类别。在由此形成的观察性总体中，与第2节的观察性数据一样，选择概率取决于某人在提供方案下本会处于的收入类别；在给定 X 和 M 的条件下比较处理组和对照组人员会产生偏差。

我们将9,084 人视为总体，并在1,000 次重复中分别从该总体抽取两个样本。实验样本由以相等概率有放回抽取的4,500 人组成，并记录（X， T， S， M）。观察样本由独立于实验样本、有放回抽取的4,500 人组成，并记录（X， T， M， Y）。其抽样保持每个单元格在9,084 人中的份额不变，但在每个单元格内扭曲各收入类别中处理组和对照组人员被抽中的概率，因此，在观察性总体中，某人在单元格 x 且其在提供方案下的收入类别为 k 时，选择概率 $\pi_1(k, x) = P(T = 1 \mid S(1) = k, x, O)$取决于 k，具体取决于三个选择规则中的哪一个。

图3绘出了每个单元格的选择概率 $\pi_1(k, x)$ 与提供方案下的收入类别k 之间的关系，同时还绘出了实验总体中的相同概率 $P(T = 1 \mid S(1) = k, x, E)$；对于每个 k，该概率等于单元格 x 中提供方案的份额。在实验总体中，选择概率不取决于提供方案下的收入类别，并且在各单元格中介于0.56和0.66之间。在观察性总体中，根据全部三种选择规则，该概率取决于收入类别。在一个单元格内，根据选择规则A，随着收入类别提高，该概率以相等的步长上升；根据选择规则B，则以相等的步长下降。根据选择规则 C，对于没有第3年收入的人员和收入高于中位数的人员，该概率取相同的值；对于收入为正且不高于中位数的人员，则取较低的值。

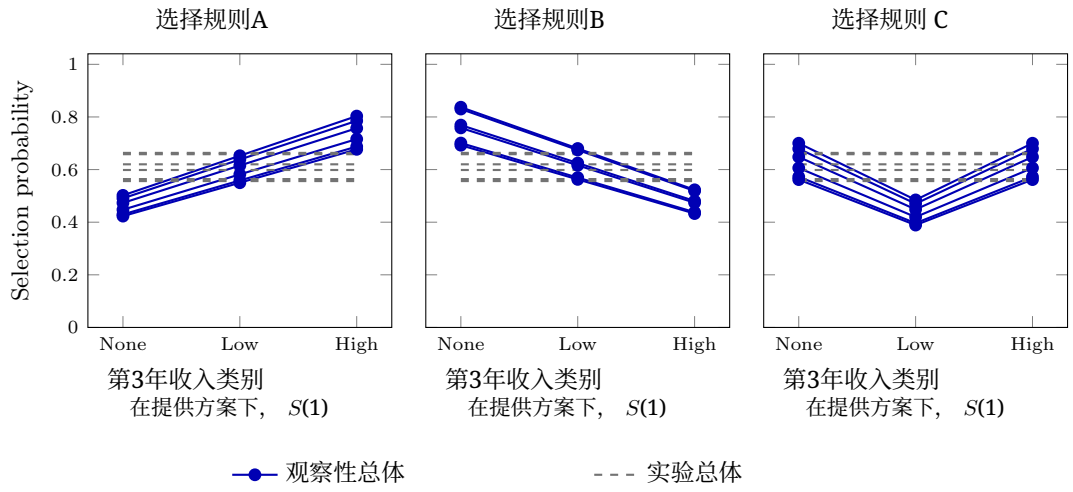

图3 标准误 观察性总体和实验总体中按第3年收入类别划分的抽样概率s

注：每个面板展示一种选择规则。每条实线连接六个单元格之一 x 中的 $\pi_1(k,\ x) = P(T = 1 \mid S(1) = k,\ x,\ O)$、$k = 1,\ 2,\ 3$，而每条虚线表示 $P(T = 1 \mid S(1) = k,\ x,\ E)$，对于每个 k，该值等于单元格 x 中人员的提供方案占比。在每条规则下，不同 k 对应的 $\pi_1(k,\ x)$ 之比率在每个单元格中都相同，因此一个面板中的实线形状相同，而不同单元格之间仅相差一个共同因子。取值几乎相等的单元格的线条会重叠。None、Low和High是第5.1节中提供方案下的收入类别 $k = 1,\ 2,\ 3$。

我们将估计值与伪真实值进行比较

$$\tilde{\tau} = \sum_x \bar{p}(x)\{\bar{Y}_1(x) - \bar{Y}_0(x)\} = 19.50$$

其中，$\bar{p}(x)$ 是单元格 x 在 9,084 人中的份额，而 $\bar{Y}_t(x)$ 是满足 $T = t$ 和 $X = x$的人员中 Y 的均值。这一数值对三条规则和所有估计量都相同。

在这一构造中，对于抽取样本所依据的总体，假设2.1–2.6、3.2和3.3按设计成立；将独立抽样方案扩展为样本量相等的情形，可得出假设 3.1；并且伪真实值 $\tilde{\tau}$ 等于观察性总体的平均处理效应，因此下文报告的覆盖率对应于从 9,084 人中重复抽样的结果。由于 S 和 M 各取三个值，假设 2.8 中的完备性以及假设 3.4 中权重函数的存在性，可由收入类别与 M各水平的联合概率所满足的秩条件推出；该条件在每个单元格和组别中均成立，且权重函数唯一。

假设 2.7 在这一构造中仅近似成立，但偏离程度很小。该假设要求：对于具有相同单元格、组别和收入类别的人员，第4年收入的均值在 M 的每个水平上都相同。在这 9,084 人中，这些均值并不相同，但与收入类别之间的差异相比，这些差异很小。在单元格和组别内部，收入类别解释了 Y 方差的 26.7%，而

M 在收入类别内部的水平又增加了 0.6 个百分点。在排除附录 D.2 所述内容后，仍保留至少两个 M 水平的、具有相同单元格、组别和收入类别的人员组中， M 某一水平下 Y 的均值与该组均值之间的绝对差异，在这些组内按人员计算的平均值约为相邻收入类别均值之间平均差异的 7%。附录 D.2 给出了相关检验以及这种偏离对估计量的影响。

5.2 估计与结果

本小节介绍这些估计量，并报告它们在三条规则下的误差和覆盖率。我们计算第 4.2 节所提出的估计量的三个版本。基 b_J in$S$ 由三个收入类别的指示函数组成，检验函数 φ_L 则是 M三个水平的指示函数，因此 $J = L = 3$。与第 4.2 节一样，交叉拟合版本使用两折和四折。我们还计算第 4.2 节的三个比较估计量，它们使用M 或处理前协变量代替 S。ESC 对其单个控制变量使用二次函数，ESC 插补使用指示函数检验函数 $\varphi_L(M)$，而观测回归则在第 4.2 节的回归中加入公开数据的 26 个处理前协变量。

表 4 报告了在每种选择规则下，1,000 次重复中各估计量的误差及其 95% 区间的覆盖率。在所有三种选择规则下，所提出的方法（Tikhonov）的绝对偏差均小于三个比较估计量，覆盖率也更高。在选择规则 A 下，其偏差为–4.09；在选择规则 B 下为 3.33；在选择规则 C 下为 –5.21；其覆盖率分别为 92.9%、92.4% 和 93.6%。交叉拟合版本的表现也很好。使用四折时，其绝对偏差较小，覆盖率接近名义水平；使用两折时，其估计量的离散程度更大，覆盖率接近 90%。无惩罚版本与 Tikhonov版本相近。两个使用 M 替代收入类别的比较估计量的偏差会随选择方向改变符号：在选择规则 A 下为正，在选择规则 B 下为负，在选择规则 C 下为正；其覆盖率介于 60.3% 和 77.4% 之间。观测回归在选择规则 A 和 B 下都远离 $\tilde{\tau}$ ；在选择规则 C 下，两个外部类别具有相同的选择概率，此时其偏差为 10.74，覆盖率为 51.1%。

选择规则	估计量	Bias	标准差	均方根误差	标准误	覆盖率
A	Proposed（Tikhonov）	−4.09	9.10	9.97	8.28	92.9
A	Proposed（Tikhonov、交叉拟合、2折）	6.09	28.00	28.64	11.45	90.1
A	所提出的方法（Tikhonov、交叉拟合、4折）	1.87	10.60	10.76	9.43	94.6
A	所提出的方法（未惩罚）	−4.22	9.22	10.14	9.16	95.1
A	ESC	6.37	6.09	8.81	5.73	77.0
A	ESC 插补	7.47	6.26	9.75	5.85	72.6
A	观测回归	27.32	5.81	27.93	5.61	0.5
B	Proposed（Tikhonov）	3.33	7.84	8.51	7.52	92.4
B	Proposed（Tikhonov、交叉拟合、2折）	−0.51	22.99	22.99	8.85	90.4
B	Proposed（Tikhonov、交叉拟合、4折）	1.97	7.80	8.04	7.90	94.9
B	Proposed（未惩罚）	3.54	7.90	8.66	8.03	93.8
B	ESC	−8.21	5.58	9.92	5.53	69.4
B	ESC 插补	−7.09	5.81	9.16	5.77	77.4
B	观测回归	−28.76	5.71	29.32	5.59	0.2
C	Proposed（Tikhonov）	−5.21	9.20	10.57	8.37	93.6
C	Proposed（Tikhonov，交叉拟合，2折）	6.46	30.26	30.92	11.44	89.5
C	Proposed（Tikhonov，交叉拟合，4折）	1.43	13.33	13.40	9.50	95.8
C	Proposed（未惩罚）	−5.41	10.17	11.51	9.23	95.3
C	ESC	8.45	5.66	10.17	5.57	66.3
C	ESC 插补	9.88	5.80	11.46	5.76	60.3
C	观测回归	10.74	5.71	12.16	5.52	51.1

表 4 选择规则 A、B 和 C 下估计量的误差和覆盖率

注：伪真实值为 $\tilde{\tau} = 19.50$，每种选择规则均有 1,000 次重复，在每次重复中都计算了每个估计值和标准误。交叉拟合版本将每个样本分成两折和四折。偏差是估计值的均值减去 $\tilde{\tau}$，标准差是估计值的标准差，均方根误差是估计值与 $\tilde{\tau}$ 之间平方差均值的平方根。标准误是标准误的均值，覆盖率是估计值加减 1.96 个标准误后包含 $\tilde{\tau}$ 的重复所占百分比。估计值的单位为每周美元。

6 结论

本文探讨了短期实验应测量哪些内容，以便与观察性数据结合后，识别观察性总体中的平均长期处理效应。当观察性数据记录了处理后的粗略测量值，而该测量值反映了驱动选择的特征时，实验应收集相同的测量值，并足够详细地测量这些特征以消除混杂，同时区分出的水平数不应多于该粗略测量值所能区分的数量。在第2节的识别假设下（其中包括两个排除限制和完备性），通过共享测量值连接两个总体的两个线性逆问题可以识别该效应。根据定理 3.2 和 3.3 的条件，所提出的估计量具有渐近正态性；此外，当每个组别和单元格中的权重方程均具有唯一的平方可积解，且附录 B 的切空间条件成立时，它们还能够达到附录 B 中子模型的效率界。在模拟研究中，随着样本规模增大，不进行样本拆分的 Tikhonov估计量的偏差逐渐减小，而使用共享测量值替代特征的比较估计量仍然存在偏差；在 Job Corps 示例中，在三种选择规则下，Tikhonov版本的绝对偏差均较小，覆盖率均高于比较估计量。

关于在手稿准备过程中使用生成式人工智能和人工智能辅助技术的声明

在准备本文的过程中，作者使用了 Claude（Anthropic）、Codex（OpenAI）和 Gemini（Google）来起草和修订文本，将手稿的部分内容翻译成英文，为模拟和实证示例编写并检查代码，以及检查推导和证明。在使用这些工具后，作者根据需要对内容进行了审阅和编辑，并对已发表文章的内容承担全部责任。

参考文献

Ai, C. 和 X. Chen（2003）。含有未知函数的条件矩约束模型的有效估计。*Econometrica* 71，1795–1843。Andrews, D. W. K.（2017）。 L^2-完备与有界完备分布的例子。*Journal ofEconometrics*199，213–220。Athey, S.、R. Chetty 和 G. W. Imbens（2025）。实验选择校正估计量：利用实验消除观察性估计中的偏差。NBER工作论文33817。https://www.nber.org/papers/w33817

Athey, S.、R. Chetty、G. W. Imbens 和 H. Kang（2026）。替代指标：结合短期代理变量，更快速、更精确地估计长期处理效应。《经济研究评论》93，2284–2312。Blundell, R.、X. Chen 和 D. Kristensen（2007）。形状不变恩格尔曲线的半非参数工具变量估计。Econometrica 75，1613–1669。Carrasco, M.、J.-P. Florens 和 E. Renault（2007）。结构计量经济学中的线性逆问题：基于谱分解与正则化的估计。载于 J. J. Heckman 和 E. E. Leamer（编），《计量经济学手册》，第6B卷，5633–5751。爱思唯尔。Chen, X.（2007）。半非参数模型的大样本筛估计。载于 J. J. Heckman 和 E. E. Leamer（编），《计量经济学手册》，第6B卷，5549–5632。爱思唯尔。Chen, X. 和 T. M. Christensen（2018）。非参数工具变量回归非线性泛函的最优上确界范数收敛率与一致推断。Quantitative Economics 9，39–84。Chen, X. 和 D. Pouzo（2012）。具有可能不光滑广义残差的非参数条件矩模型估计。Econometrica 80，277–321。Chernozhukov, V.、D. Chetverikov、M. Demirer、E. Duflo、C. Hansen、W. Newey 和 J. Robins（2018）。用于处理效应和结构参数的双重/去偏机器学习。《计量经济学杂志》21(1)，C1–C68。https://doi.org/10.1111/ectj.12097 Cui, Y.、H. Pu、X. Shi、W. Miao 和 E. Tchetgen Tchetgen（2024）。半参数近端因果推断。《美国统计学会会刊》119，1348–1359。Darolles, S.、Y. Fan、J.-P. Florens 和 E. Renault（2011）。非参数工具变量回归。Econometrica 79，1541–1565。d’Haultfœuille, X.（2010）。一种处理内生选择的新工具变量方法。Journal of Econometrics 154，1–15。d’Haultfœuille, X.（2011）。关于非参数工具变量问题中的完备性条件。《计量经济理论》27，460–471。Horowitz, J. L.（2011）。应用非参数工具变量估计。Econometrica 79(2)，347–394。Miao, W.、Z. Geng 和 E. J. Tchetgen Tchetgen（2018）。利用未测量混杂因素的代理变量识别因果效应。《生物计量学》105，987–993。Newey, W. K.（1990）。半参数效率界。《应用计量经济学杂志》5，99–135。Newey, W. K. 和 J. L. Powell（2003）。非参数模型的工具变量估计。Econometrica 71，1565–1578。Schennach, S. M.（2016）。测量误差文献的最新进展。《经济学年度评论》8，341–377。Schochet, P. Z.、J. Burghardt 和 S. Glazerman（2001）。National Job Corps Study：Job Corps 对参与者就业及相关结果的影响。最终报告，2001年6月。Mathematica政策研究。Tchetgen Tchetgen, E. J.、A. Ying、Y. Cui、X. Shi 和 W. Miao（2024）。近端因果推断导论。《统计科学》39，375–390。van der Vaart, A. W.（1998）。《渐近统计学》。剑桥大学出版社。

附录

实验应测量什么？结合短期实验与存在混杂的观察性数据识别长期处理效应

Kentaro Kawato 东京大学经济学研究生院

内容

附录 A 第2节的证明 .39

A.1 定理 2.1 的证明 .40A.2 加性测量误差下的完备性 .42

附录 B 定理 3.1 的证明 .43

B.1 模型、光滑子模型与切空间 .45B.2 引理与定理 3.1 的证明 .48B.3 引理B.1–B.3 的证明 .52B.4 引理 B.4 的证明 .54

附录 C 定理 3.2 和 3.3 的证明 .57

C.1 定理 3.2 的证明 .58C.2 引理与定理 3.3 的证明 .70

附录 D 关于模拟和实证设计的说明 .76

D.1 模拟设计中的权重函数 .76D.2 实证示例中的假设 2.7 .77

附录 E 补充材料 .78

附录 A 第2节证明

附录 A.1 通过推导两个矩方程、证明其解的唯一性，并将这些解与潜在结果的均值联系起来，证明了定理 2.1，即第 2 节中带有共享的治疗后测量的识别结果。附录 A.2 给出了加性测量误差下完备性的一般充分条件，并验证了其作为特殊情形在第 2.2 节例 2.1 的第二部分以及第 4 节的模拟设计中的完备性。

A.1 定理 2.1 的证明

A.1.1 可积性条件

始终假设 $\mathbb{E}[|Y(t)| \mid O] < \infty$，并且对于 $F_{X|O}$-几乎处处的 x，有 $\mathbb{E}[Y^2 \mid T = t, x, O] < \infty$。这些就是附录 A 中的可积性条件，正文对此有所引用。第一个条件使目标估计量以及潜在结果的条件均值得到良好定义，第二个条件确保第二个方程 (2.7) 的解属于施加完备性条件的函数类。在附录 B 的假设 B.1 下，这些条件对于第 3 节所使用的矩均成立。

A.1.2 定理 *2.1* 的证明

证明表明，方程 (2.7) 的解为 π_t 和 ν_t，并且按照 (2.8) 对它们的比率取平均，得到 τ。假设 2.6 和 2.7 首先将两个观察性矩转化为观察性总体中关于 $S(t)$ 的函数的平均值，而假设 2.3 和 2.4 则将这些平均值推广到实验中。随后，完备性适用于两个解之差，因为该差的条件期望为零。最后，假设 2.5 将这些解与潜在结果的均值联系起来。

回顾方程 (2.7)。对于每个 t，对于 $F_{X|O}$-几乎处处的 x，以及任意有界可测函数 h，

$$\mathbb{E}[\mathbf{1}\{T = t\}h(M) \mid x, O] = \mathbb{E}[\pi_t(S, x)h(M) \mid T = t, x, E],$$
$$\mathbb{E}[\mathbf{1}\{T = t\}Yh(M) \mid x, O] = \mathbb{E}[\nu_t(S, x)h(M) \mid T = t, x, E]$$

固定一个有界可测函数 h，并固定 t 和 x。一致性、迭代期望以及 (2.5) 中关于选择的排除限制给出

$$\begin{aligned}\mathbb{E}[\mathbf{1}\{T = t\}h(M) \mid x, O] &= \mathbb{E}[\mathbf{1}\{T = t\}h(M(t)) \mid x, O] \\ &= \mathbb{E}[\pi_t(S(t), x)h(M(t)) \mid x, O] \\ &= \mathbb{E}[\pi_t(S, x)h(M) \mid T = t, x, E]\end{aligned} \tag{A.1}$$

其中，最后一个等式由联合分布的可迁移性和实验中的随机化得出。根据假设 2.1 和 2.2，观察性总体中满足 $T = t$ 的单位里，S 的分布相对于$F_{S(t)|x,O}$的密度为 $\pi_t(\cdot, x)/P(T = t \mid x, O)$，且该密度$F_{S(t)|x,O}$-几乎处处为正，因此这两个分布具有相同的零测集。因此，$\mu_t(\cdot, x)$（其定义在 $F_{S|T=t,x,O}$-几乎处处成立）也定义在$F_{S(t)|x,O}$-几乎处处成立，并且下文的 $\mu_t(S(t), x)$ 在观察性总体中几乎处处有定义

中。对于含结果变量的方程，假设 2.7 首先将观察到的结果变量替换为其条件均值，随后进行相同的计算可得

$$\begin{aligned}\mathbb{E}[\mathbf{1}\{T=t\}Yh(M)\mid x,O] &= \mathbb{E}[\mathbf{1}\{T=t\}\mu_t(S,x)h(M)\mid x,O]\\ &= \mathbb{E}[\pi_t(S(t),x)\mu_t(S(t),x)h(M(t))\mid x,O]\\ &= \mathbb{E}[\nu_t(S,x)h(M)\mid T=t,x,E]\end{aligned} \tag{A.2}$$

每个积分 在上述可积性条件以及有界性条件下都有良好定义 h.

由于选择概率取值于 $[0,1]$、 $\pi_t \in L^2$（$F_{S|T=t、x、E}$）。根据条件詹森不等式和 $\pi_t^2 \le \pi_t$，可得

$$\begin{aligned}\mathbb{E}[\nu_t(S(t),x)^2\mid x,O] &\le \mathbb{E}[\pi_t(S(t),x)\mu_t(S(t),x)^2\mid x,O]\\ &= \mathbb{E}[\mathbf{1}\{T=t\}\mu_t(S,x)^2\mid x,O]\\ &\le \mathbb{E}[\mathbf{1}\{T=t\}Y^2\mid x,O]<\infty\end{aligned}$$

并且根据可迁移性， ν_t 也属于 L^2（$F_{S|T=t、x、E}$）。

设 d 为方程 (A.1) 的两个 L^2 解之差。由于对于每个有界可测函数 $h(M)$， $\mathbb{E}[d(S)h(M)\mid T=t、x、E]=0$，条件期望的定义性质给出 $\mathbb{E}[d(S)\mid M、T=t、x、E]=0$。根据假设 2.3 和 2.4，这等价于$\mathbb{E}[d(S(t))\mid M(t)、x、O]=0$，而假设 2.8 中的 (2.6) 给出 $d=0$。对方程 (A.2) 的两个解之差也可应用同样的论证。因此， π_t 和 ν_t 是唯一确定的，而假设 2.2 中的正值性在 $F_{S(t)|x、O}$的支持集上确定了 $\mu_t=\nu_t/\pi_t$。

一致性，以及由假设 2.5 中的条件独立性和附录 A.1.1 中的 $\mathbb{E}[|Y(t)|\mid O]<\infty$ 推出的条件均值相等，在 $F_{S|T=t,\ x,\ O}$ 的支持集上成立，

$$\mu_t(S(t),x)=\mathbb{E}[Y(t)\mid T=t,S(t),x,O]=\mathbb{E}[Y(t)\mid S(t),x,O]$$

而正值性将这一等式扩展到 $F_{S(t)|x,\ O}$ 的支持集上。依次应用迭代期望、可迁移性以及实验中的随机化，可得

$$\mathbb{E}[Y(t)\mid x,O]=\int \mu_t(s,x)\,dF_{S(t)|x,O}(s)=\int \mu_t(s,x)\,dF_{S|T=t,x,E}(s)$$

最后，对 $F_{X|O}$ 取平均，并计算 $t=1$ 与 $t=0$ 之差，得到

$$\tau=\int\left\{\int \mu_1(s,x)\,dF_{S|T=1,x,E}(s)-\int \mu_0(s,x)\,dF_{S|T=0,x,E}(s)\right\}dF_{X|O}(x)$$

即为（2.8）。 □

A.2 加性测量误差下的完备性

本小节以误差 U_2 的分布为条件，给出完备性的一个充分条件，其中例 2.1 中的测量采取 $M(t) = S(t) + U_2$ 的形式。作为特殊情形，它推导出例 2.1 的第二部分以及第 4 节中设计的完备性。固定 $(t,\ x)$，并省略观察性总体中关于 $X = x$ 的条件化。假设 $S(t)$ 和 U_2 为实值，U_2 与 $S(t)$ 独立，并且 U_2 的特征函数 $u \mapsto \mathbb{E}[e^{iuU_2}]$ 不为零的点集在实数轴上稠密。不对 $S(t)$ 的分布作任何条件限制。那么，如果 $h \in L^2(F_{S(t)|x,\ O})$ 满足 $\mathbb{E}[h(S(t)) \mid M(t)] = 0$ 几乎处处成立，则 $h(S(t)) = 0$ 几乎处处成立。

该结果将d'Haultfœuille (2011年) 的命题2.4应用于实值模型 $M(t) = S(t) + U_2$，并放宽了对误差的条件。该命题针对这样一种模型确立了可积函数的完备性：测量的某个函数等于潜变量的某个函数加上一个误差；给定共同变量时，该误差与潜变量独立，将潜变量进行变换的函数为一对一函数，且误差的特征函数的零点是孤立的。在 $M(t) = S(t) + U_2$中，两个函数都是恒等函数，因此满足一对一条件。如果零点是孤立的，那么特征函数非零的点集在实数轴上稠密，因此上述对误差的条件更弱。我们的结论仅适用于平方可积函数，证明采用相同的论证。按照 -完备性这一术语，平方可积函数的完备性是 Andrews(2017年) 所说的 -完备性，它介于可积函数的完备性与有界函数的完备性之间。假设 2.8 要求满足这种形式的完备性。我们在此证明该结果，以确保论文内容自成一体。

柯西–施瓦茨不等式给出 $\mathbb{E}[|h(S(t))|] \le \mathbb{E}[h(S(t))^2]^{1/2} < \infty$。固定一个实数 u。由于 $e^{iuM(t)}$ 是 $M(t)$ 的有界函数，根据迭代期望以及 U_2 与 $S(t)$ 的独立性可得

$$0 = \mathbb{E}\left[\mathbb{E}[h(S(t)) \mid M(t)]\, e^{iuM(t)}\right] = \mathbb{E}[h(S(t))e^{iuM(t)}] = \mathbb{E}[h(S(t))e^{iuS(t)}]\,\mathbb{E}[e^{iuU_2}]$$

因此，对每个 u（满足 $\mathbb{E}[e^{iuU_2}] \neq 0$），都有 $\mathbb{E}[h(S(t))e^{iuS(t)}] = 0$。由于 $\mathbb{E}[|h(S(t))|] < \infty$，控制收敛定理表明 $u \mapsto \mathbb{E}[h(S(t))e^{iuS(t)}]$ 连续；又因为它在一个稠密集上为零，所以对每个 u 都为零。因此，有限符号测度$B \mapsto \mathbb{E}[h(S(t))\mathbf{1}\{S(t) \in B\}]$ 的傅里叶变换恒为零，而傅里叶变换的唯一性

意味着该测度为零。取 $B = \{s : h(s) > 0\}$ 和 $B = \{s : h(s) < 0\}$ 可知，$h(S(t))$ 的正部和负部的期望均为零，因此 $h(S(t)) = 0$ 几乎处处成立。 □

在例 2.1 的第二部分中，U_2 独立于 $S(t)$，且服从均值为零、方差为 σ^2的正态分布。其特征函数 $\exp(-\sigma^2 u^2/2)$ 对每个 u都非零，因此上述结果对每个 t 和 x给出了假设 2.8。

在第 4 节的两种设计中，给定 $X = x$，$S(t) = U_1 + (1 + 0.5x)t$ 且 $M(t) = S(t) + U_2$，其中 U_2 在 $[-1, 1]$ 上均匀分布，并且独立于 (U_1, X)。U_2 的特征函数等于 $\sin(u)/u$（当 $u \neq 0$ 时），并在 $u = 0$处等于 1，因此其零点为 $u = k\pi$，其中 k 为非零整数；其非零点集合是稠密的。因此，上述结果对每个 t 和 x给出了假设 2.8：这既适用于连续设计（其中 U_1 服从均匀分布），也适用于离散设计（其中 U_1 取三个值）。

附录 B 定理 3.1 的证明

本附录证明第 3 节中的定理 3.1，该定理给出了有效影响函数和半参数效率界。表 B.1 列出了各小节的内容，图 B.1 展示了证明中所用结果之间的依赖关系。第 3 节中关于估计量的定理 3.2 和 3.3 在附录 C 中证明。

图 B.1 展示了用于证明定理 3.1 的各项结果之间的依赖关系，其中，从一项结果指向另一项结果的箭头表示前者用于证明后者。定理 3.1 由引理 B.2、B.3 和 B.4 推出。引理 B.2 表明，影响函数 ψ_E, ψ_O的均值为零且具有有限二阶矩；引理 B.3 表明，双样本对数似然比是局部渐近正态的；引理 B.4 表明，τ 沿光滑子模型的右导数等于影响函数与得分函数的期望乘积之和。根据引理 B.2 和 B.4，归一化影响函数（$\psi_E/\rho, \psi_O/(1-\rho)$）是$\tau$的梯度；根据假设 B.2，该梯度属于切空间。引理 B.3 通过卷积定理发挥作用，该定理使效率界成为正则估计量渐近

小节	内容
B.1	定义模型、光滑子模型、切空间和梯度，并陈述假设 B.2
B.2	陈述引理 B.1–B.4，并利用它们证明定理 3.1
B.3	证明引理B.1–B.3
B.4	证明引理 B.4

表 B.1 附录 B 的小节及其内容

方差的下界。引理 B.2 和 B.4 的证明使用了引理 B.1，该引理将观测矩方程（2.7）扩展到了无界函数。

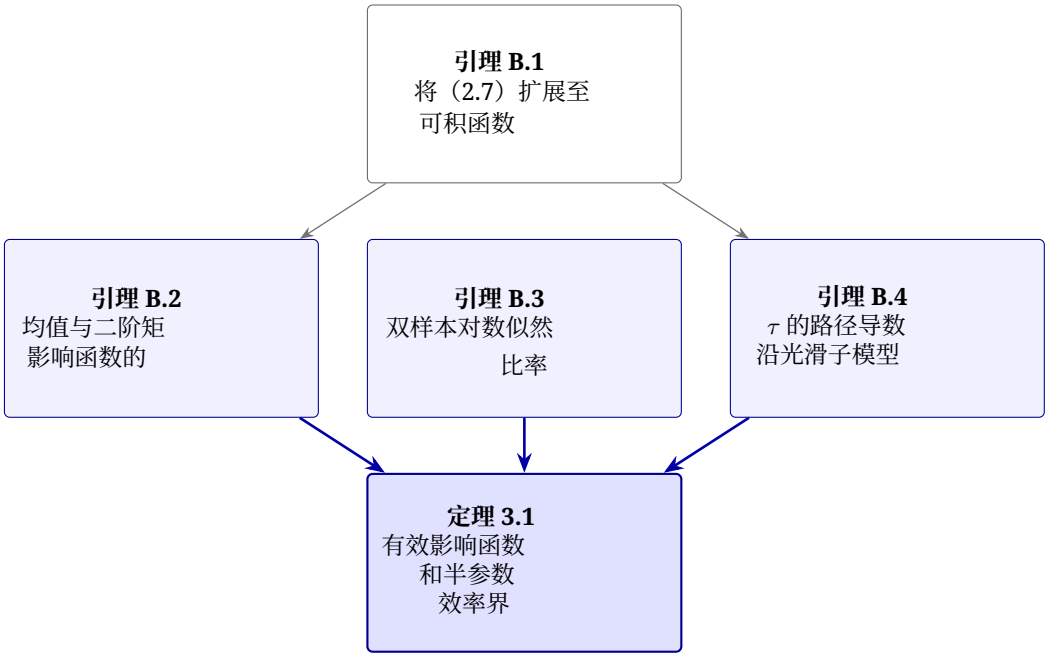

图 B.1 用于证明定理 3.1 的各项结果之间的依赖关系注：粗箭头从直接推出定理 3.1 的引理出发。

假设 B.1 (有界性)。 存在一个常数 $C > 0$，使得对于每个 $t \in \{0、1\}$ 以及 $x \in \mathcal{X}$，以下条件均成立。

1. $|\mu_t| \le C$ 在 $S \mid T = t$的支持集上几乎处处成立， x、 E。
2. $|Y| \le C$ 在观察性总体中几乎处处成立。

假设 B.1 对真实函数和结果变量给出了界，并且结合 $\pi_t \le 1$ 可得 $|\nu_t| \le C$。估计量的有界性在附录 C 的假设 C.1 中陈述。下文中，我们取 $\pi_t(\cdot,\ x)$和 $\mu_t(\cdot,\ x)$的版本，使得 $a \le \pi_t(s,\ x) \le 1$ 且 $|\mu_t(s,\ x)| \le C$ 对每个 s均成立，其中 a 是假设 3.3 中的常数。由于不同版本仅在 $S \mid T = t,\ x,\ E$的零测集上有所不同，因此这一选择不会改变（2.7）、（3.2）–（3.4）以及下文中的期望，并且 $|\nu_t(s,\ x)| \le C$ 对每个 s均成立。

我们回顾正文中的四个方程，附录 B 和 C 中的证明会反复用到它们。方程 (2.7) 将两个观察性矩表示为实验平均值。对于每个 t，对于 $F_{X|O}$-几乎处处的 x，以及任意有界可测函数 h，

$$\begin{aligned}\mathbb{E}[\mathbf{1}\{T = t\}h(M) \mid x, O] &= \mathbb{E}[\pi_t(S, x)h(M) \mid T = t, x, E],\\ \mathbb{E}[\mathbf{1}\{T = t\}Yh(M) \mid x, O] &= \mathbb{E}[\nu_t(S, x)h(M) \mid T = t, x, E]\end{aligned} \tag{2.7}$$

方程 (3.2) 给出了假设 3.4 对权重函数 ω_t^{ν}， ω 施加的条件 π_t，

$$\mathbb{E}[\omega_t^{\nu}(M,x) \mid S, T=t, x, E] = \frac{1}{\pi_t(S,x)}, \qquad \mathbb{E}[\omega_t^{\pi}(M,x) \mid S, T=t, x, E] = \frac{\mu_t(S,x)}{\pi_t(S,x)} \tag{3.2}$$

方程（3.6）根据（3.5）中的分层均值 $\widehat{\mu}_t(x)$定义估计量，

$$\widehat{\tau} = \sum_{x \in \mathcal{X}} \widehat{P}(X=x \mid O)\{\widehat{\mu}_1(x) - \widehat{\mu}_0(x)\} \tag{3.6}$$

方程（3.7）定义渐近方差 Σ^2 及其估计量 $\widehat{\Sigma}^2$，

$$\begin{aligned} \Sigma^2 &= \frac{1}{\rho}\mathbb{E}[\psi_E(Z_E)^2 \mid E] + \frac{1}{1-\rho}\mathbb{E}[\psi_O(Z_O)^2 \mid O], \\ \widehat{\Sigma}^2 &= \frac{N}{n_E}\frac{1}{n_E}\sum_{i=1}^{n_E} \widehat{\psi}_E(Z_{E,i})^2 + \frac{N}{n_O}\frac{1}{n_O}\sum_{j=1}^{n_O} \widehat{\psi}_O(Z_{O,j})^2 \end{aligned} \tag{3.7}$$

B.1 模型、光滑子模型和切空间

本小节针对两个独立样本定义光滑子模型、切空间、有效影响函数和正则估计量，这些内容由定理 3.1 使用；同时陈述切空间条件，即假设 B.2。这些定义将 van der Vaart (1998) 在第25章 P_E中的单样本定义调整为双样本内积。记 P_O 为观测分布，并记 $P = (P_E, P_O)$ 为它们构成的二元组。主文中的模型 $\mathcal{P}$ 是所有观测映像二元组的集合；这些观测映像来自完整数据律，并且 X 的值域包含于 $\mathcal{X}$ 中，同时满足假设 2.1–2.8 以及附录 A 中的可积性条件。根据定理 2.1， τ 是定义在 $\mathcal{P}$上的函数。

我们用 $L_2^0(P_E)$， $L_2^0(P_O)$ 表示观测值的均值为零且平方可积的实值函数集合。在假设 3.1 的 ρ 下，双样本内积为

$$\langle (a_E, a_O), (b_E, b_O) \rangle_{\rho} = \rho\mathbb{E}[a_E b_E \mid E] + (1-\rho)\mathbb{E}[a_O b_O \mid O] \tag{B.1}$$

且 $\|\cdot\|_{\rho}$ 表示其范数。路径 $\{P_{\varepsilon} = (P_{E,\,\varepsilon},\ P_{O,\,\varepsilon}) : 0 \le \varepsilon < \varepsilon_0\}$ 属于 $\mathcal{P}$，且满足 $\varepsilon_0 > 0$，是一个族，其中每个 P_{ε} 都属于 $\mathcal{P}$ 且满足 $P_0 = P$。如果一个路径的似然比具有带有界得分函数的一阶展开，则称其为正则路径；也就是说，对于每个 $G \in \{E,\ O\}$， $P_{G,\,\varepsilon}$ 相对于 P_G 绝对连续，并且存在 $a_G \in L_2^0(P_G)$（在 P_G 下有界），以及函数 $r_{G,\,\varepsilon}$，其在 P_G 下的本质上确界 $\|r_{G,\,\varepsilon}\|_{\infty}$ 随着 $\varepsilon \downarrow 0$趋于零，使得密度 $dP_{G,\,\varepsilon}/dP_G$ 等于

$1 + \varepsilon a_G + \varepsilon r_{G,\,\varepsilon}$。我们称 $(a_E,\ a_O)$ 为该路径的得分函数。与 van derVaart (1998)第25章中具有有界得分的路径一样，正则路径在二次均值意义下可微，

$$\int \left\{ dP_{G,\varepsilon}^{1/2} - dP_G^{1/2} - \frac{\varepsilon}{2} a_G \, dP_G^{1/2} \right\}^2 = o(\varepsilon^2) \quad (\varepsilon \downarrow 0), \qquad G \in \{E, O\}, \tag{B.2}$$

(B.2) 中的积分是根据一个支配 P_G 和 $P_{G,\,\varepsilon}$ 的 σ有限测度，对密度的平方根进行积分；该积分不依赖于测度的选择。沿着一条正则路径，$P_{G,\,\varepsilon}$ 相对于 P_G 绝对连续，因此对于来自 P_G 的观测值，密度 $dP_{G,\,\varepsilon}/dP_G$ 以概率 1 有定义。

沿路径的分布下，选择概率和乘积也被唯一确定为 (2.7) 的解。取 $\mathcal{P}$ 中的一条正则路径。对于每个 ε，我们将完整数据律的 (2.2) 中的函数记为 $\pi_{t,\varepsilon}, \mu_{t,\varepsilon}$，其观测映像为 P_ε；并令 $\nu_{t,\,\varepsilon} = \pi_{t,\varepsilon}\mu_{t,\varepsilon}$。当 ε 足够小时，两个样本的密度 $dP_{G,\varepsilon}/dP_G = 1 + \varepsilon a_G + \varepsilon r_{G,\varepsilon}$ 均介于 1/2 和 3/2 之间，因此对于所有 t, x，$P_\varepsilon(X = x \mid O) \geq c/2$，且 $P_\varepsilon(T = t, X = x \mid E) \geq c/2$。由于 P_ε 属于 $\mathcal{P}$，将定理 2.1 应用于 P_ε 可知，对于所有 t, x, $\pi_{t,\varepsilon}(\cdot, x)$ 和 $\nu_{t,\varepsilon}(\cdot,\ x)$ 都是在 P_ε 下 (2.7) 的唯一解，且相对于 P_ε 下 $S \mid T = t, x, E$ 的分布平方可积；特别是，它们不依赖于完整数据律的选择。密度的界意味着，在 P_ε 下和 P 下，$S \mid T = t, x, E$ 的分布相互绝对连续，且具有相同的平方可积函数空间。因此，$\pi_{t,\varepsilon}(\cdot, x)$ 和 $\nu_{t,\varepsilon}(\cdot,\ x)$ 属于 $L^2(F_{S|T=t,x,E})$，可以与真实分布下的 $\pi_t(\cdot, x)$ 和 $\nu_t(\cdot, x)$ 进行比较。与第 3.3 节一样，我们用 $\|f\|_{2,t,x} = \mathbb{E}[f(S, M)^2 \mid T = t, x, E]^{1/2}$ 来衡量差异的大小，其中 f 是 (S, M) 的函数。

我们将注意力限制在选择概率和乘积均可微的正则路径上。若对于每个 $t \in \{0, 1\}$ 和 $x \in \mathcal{X}$，都存在 S 的函数 $\dot{\pi}_t(\cdot,\ x)$ 和 $\dot{\nu}_t(\cdot,\ x)$，满足 $\|\dot{\pi}_t\|_{2,t,x} < \infty$ 以及 $\|\dot{\nu}_t\|_{2,t,x} < \infty$，使得$\|(\pi_{t,\varepsilon} - \pi_t)/\varepsilon - \dot{\pi}_t\|_{2,t,x} \to 0$ 以及 $\|(\nu_{t,\varepsilon} - \nu_t)/\varepsilon - \dot{\nu}_t\|_{2,t,x} \to 0$ 在 $\varepsilon \downarrow 0$时成立，则 $\mathcal{P}$ 中的正则路径是一个光滑子模型。由于沿该路径的 π_t 和 ν_t 是该路径上概率律下方程 (2.7) 的解，它们的变化无法直接从观测分布的变化中得出；因此，遵循 Newey (1990) 根据子模型上的导数推导效率界的做法，我们将注意力限制在这些函数可微的子模型上。第 3 节中的效率结论是相对于光滑子模型而言的。附录B.2.1中的引理 B.4表明，对于每个得分函数为 (a_E, a_O) 的光滑子模型，$\tau(P_\varepsilon)$ 在 $\varepsilon = 0$处右可微，且

$$\left. \frac{d\tau(P_\varepsilon)}{d\varepsilon} \right|_{\varepsilon = 0+} = \mathbb{E}[\psi_E(Z_E) a_E(Z_E) \mid E] + \mathbb{E}[\psi_O(Z_O) a_O(Z_O) \mid O] \tag{B.3}$$

对于　由 (3.2) 的任意平方可积解构造的 (3.3) 和 (3.4) 中的函数　　.

当 S 和 M 在给定 $T = t$ 和 $X = x$的条件下只取有限个值时，$\mathcal{P}$ 中的每条正则路径都是一个光滑子模型。在这种情况下，以 M 各取值的指示函数作为h 时，(2.7)在路径上的每个分布下都是关于$\pi_{t,\varepsilon}(\cdot, x)$各取值的线性方程组，以及关于$\nu_{t,\varepsilon}(\cdot, x)$各取值的另一个线性方程组；两个方程组有相同的系数矩阵，其列根据假设 2.8线性无关。矩阵的元素以及等式右侧的项都是 P_ε下的条件概率和均值；由于路径的密度是右可微的，这些项在 $\varepsilon = 0$处也右可微。因此，将矩阵的左逆 $(A^\top A)^{-1}A^\top$ 作用于等式右侧 A 得到的解也同样右可微。

切空间是所有光滑子模型得分函数的线性组合在 $\|\cdot\|_\rho$ 下的闭包。若一对 $(b_E, b_O) \in L_2^0(P_E) \times L_2^0(P_O)$ 满足：对于每个光滑子模型，$\tau(P_\varepsilon)$ 在 $\varepsilon = 0$ 处右可微，且其右导数等于 (b_E, b_O), $(a_E, a_O)_\rho$ 与得分函数的内积，则称该对为 τ 的梯度。切空间中的梯度称为 τ的有效影响函数（EIF）。当两个样本均值的系数为 (ψ_E, ψ_O) 时，在此内积下对应的元素为 $(\psi_E/\rho, \psi_O/(1-\rho))$。有效影响函数的平方范数称为效率界。为定义估计量的正则性，取一个光滑子模型；当 N足够大时，让两个样本分别来自 $P_{1/\sqrt{N}} = (P_{E,1/\sqrt{N}}, P_{O,1/\sqrt{N}})$，并固定 $L_2^0(P_E) \times L_2^0(P_O)$的一个线性子空间。若对于每个得分函数属于该子空间的光滑子模型，$\sqrt{N}$ 乘以估计量与 $\tau(P_{1/\sqrt{N}})$之差在 $P_{1/\sqrt{N}}$ 下依分布收敛至一个不依赖于该子模型的极限，则称估计量序列相对于该子空间是正则的。若这一性质对每个光滑子模型都成立，则称该估计量序列在模型 $\mathcal{P}$中是正则的。van der Vaart（1998）第25章也将这些概念定义为相对于给定的路径集合。

假设 B.2（切空间条件）。存在一个由 L_2^0（P_E）$\times L_2^0$（P_O）构成的线性子空间，满足以下两个条件。

1. 每个元素都是某个光滑子模型的得分函数。

2. 它在 $\|\cdot\|_\rho$ 下的闭包包含（$\psi_E/\rho, \psi_O/$（$1-\rho$）。

假设 B.2 要求归一化影响函数可以由光滑子模型的得分函数逼近。条件（1）要求存在一个得分函数线性空间，这是卷积定理将效率界转化为下界所需要的。条件（2）要求由（3.3）和（3.4）构成的归一化对可以由这些得分函数逼近。由于条件（1）中的子空间元素都是得分函数，其闭包包含于切空间中，因此条件（2）意味着 $(\psi_E/\rho, \psi_O/(1-\rho))$ 位于切空间中。如果所有光滑子模型的得分函数

构成一个线性空间，则该空间可以作为条件（1）中的子空间，此时条件（1）和（2）等价于 $(\psi_E/\rho, \psi_O/(1-\rho))$ 位于切空间中。（3.3）和（3.4）中的函数取决于（3.2）的固定解 $\omega_t^\nu, \omega_t^\pi$，假设 B.2 施加于由该解构成的对上（见注3.1）。下文中，我们固定一个满足假设 B.2 的线性子空间，并将其称为假设 B.2 中的子空间。

B.2 定理 3.1 的引理与证明

本小节证明第 3.1 节中的定理 3.1，其在附录 B 中的有界性和切空间条件分别为假设 B.1 和 B.2。附录B.2.1 陈述证明所使用的引理 B.1–B.4，附录B.2.2 证明该定理。引理B.1–B.3 在附录 B.3 中证明，引理 B.4 在附录 B.4 中证明。

B.2.1 证明中使用的引理

本部分陈述用于证明定理 3.1 的四个引理。引理 B.1 将（2.7）扩展到无界函数，引理 B.2 涉及影响函数的均值和二阶矩，引理 B.3 涉及双样本对数似然比，而引理 B.4 涉及沿光滑子模型的 τ 导数。

引理 B.1 （将 (2.7) 扩展至可积函数）。 假设假设 *2.1–2.8*、附录 *A* 中的可积性条件、假设 *3.2* 和 *3.3* 以及假设 *B.1* 均成立，并固定$t \in \{0, 1\}$ 和 $x \in \mathcal{X}$。

1. 对于每个非负可测函数 g, $\mathbb{E}[\mathbf{1}\{T = t\}g(M) \mid x, O] = \mathbb{E}[\pi_t(S, x)g(M) \mid T = t, x, E]$，其中两边都可以为无穷大。

2. 若可测函数 g 满足 $\mathbb{E}[|g(M)| \mid T = t, x, E] < \infty$，则 $\mathbb{E}[\mathbf{1}\{T = t\}g(M) \mid x, O] = \mathbb{E}[\pi_t(S, x)g(M) \mid T = t, x, E]$，且 $\mathbb{E}[\mathbf{1}\{T = t\}Yg(M) \mid x, O] = \mathbb{E}[\nu_t(S, x)g(M) \mid T = t, x, E]$，并且所有这些期望均为有限值。

3. 如果假设 *3.4* 也成立，则 $\mathbb{E}[\nu_t(S, x)\omega_t^\nu(M, x) \mid T = t、\ x、\ E]$ 和 $\mathbb{E}[\pi_t(S, x)\omega_t^\pi(M, x) \mid T = t、\ x、\ E]$ 均等于 $\bar{\mu}_t(x)$，且 $\mathbb{E}[\mathbf{1}\{T = t\}\{Y\omega_t^\nu(M, x) - \omega_t^\pi(M, x)\} \mid x、\ O] = 0$。因此，$\mathbb{E}[\psi_E(Z_E) \mid E] = \mathbb{E}[\psi_O(Z_O) \mid O] = 0$。

引理 B.1 的主要结论是断言 (3)，它指出 (3.3) 和 (3.4) 中的函数均值为零。由于权重函数不一定有界，断言 (1) 和 (2) 首先通过采用截断和收敛定理的标准论证，将针对有界函数陈述的 (2.7) 推广到可积函数。

引理 B.2（影响函数的均值与二阶矩）。 假设假设 *2.1–2.8*、附录 *A* 中的可积性条件、假设 *3.1–3.4* 以及假设 *B.1* 均成立。则 $\mathbb{E}[\psi_E(Z_E) \mid E] = \mathbb{E}[\psi_O(Z_O) \mid O] = 0$, $\mathbb{E}[\psi_E(Z_E)^2 \mid E] < \infty$，以及 $\mathbb{E}[\psi_O(Z_O)^2 \mid O] < \infty$。

引理 B.2 确保 ψ_E/ρ 和 $\psi_O/(1-\rho)$ 分别属于 $L_2^0(P_E)$ 和 $L_2^0(P_O)$，因此它们是附录 B.1 内积下梯度的候选项。

引理 B.3 (双样本对数似然比)。 假设假设 *3.1* 成立，并取 $\mathcal{P}$ 中得分函数为 (a_E, a_O) 的正则路径。那么，当 $N \to \infty$ 时，在真实分布 P 下，

$$\begin{aligned} &\sum_{i=1}^{n_E} \log \frac{dP_{E,1/\sqrt{N}}}{dP_E}(Z_{E,i}) + \sum_{j=1}^{n_O} \log \frac{dP_{O,1/\sqrt{N}}}{dP_O}(Z_{O,j}) \\ &= \frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^{n_E} a_E(Z_{E,i}) + \sum_{j=1}^{n_O} a_O(Z_{O,j}) \right\} - \frac{1}{2} \|(a_E, a_O)\|_\rho^2 + o_p(1) \end{aligned} \tag{B.4}$$

且右侧花括号中的项除以 $\sqrt{N}$ 后，依分布收敛于均值为零、方差为 $\|(a_E, a_O)\|_\rho^2$ 的正态分布。

引理 B.3 用得分函数的标准化和减去附录 B.1 中平方范数的一半，近似沿路径移动 $1/\sqrt{N}$ 后的分布与真实分布之间的双样本对数似然比。定理 3.1 的证明在卷积定理中用到该引理，而附录 C 将其与 Le Cam 第三引理结合，以证明该估计量是正则的。

引理 B.4（ 的路径导数 τ）。假设假设 *2.1–2.8*、附录 *A* 中的可积性条件、假设 *3.2–3.4* 以及假设 *B.1* 均成立，并取一个得分函数为（a_E, a_O）的光滑子模型。那么， τ（P_ε）在 $\varepsilon = 0$ 处右可微；对于由 *(3.2)* 的任意平方可积解构造的 *(3.3)* 和 *(3.4)* 中的函数，

$$\left.\frac{d\tau(P_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0+} = \mathbb{E}[\psi_E(Z_E)a_E(Z_E) \mid E] + \mathbb{E}[\psi_O(Z_O)a_O(Z_O) \mid O]$$

即附录 *B.1* 中的 *(B.3)*。

引理 B.4 给出了附录 B.1 中所述的 (B.3)，且该结论沿每条光滑子模型均成立。附录 B.4 通过将 $\tau(P_\varepsilon)$ 表示为层份额和层均值的函数，并对 ε 求导，证明了这一点。

B.2.2 定理 *3.1* 的证明

本部分证明定理 3.1 的四个断言。断言 (1) 表明，$(\psi_E/\rho,\ \psi_O/(1-\rho))$ 是 τ 的梯度；断言 (2) 表明它是切空间中的唯一梯度；断言 (3) 表明

效率界等于 Σ^2；断言 (4) 表明，正则估计量的极限分布是与 $\mathcal{N}(0,\ \Sigma^2)$ 的卷积。引理 B.2 和 B.4 给出断言 (1)，假设 B.2 以及内积的性质给出断言 (2)，内积的定义给出断言 (3)，而引理 B.3 和卷积定理给出断言 (4)。

定理 3.1 的证明。 我们证明论断（1）。引理 B.2 给出 $\psi_E \in L_2^0(P_E)$和 $\psi_O \in L_2^0(P_O)$，并且由于根据假设 3.1， ρ 和 $1-\rho$ 为正， $\psi_E/\rho \in L_2^0(P_E)$ 和 $\psi_O/(1-\rho) \in L_2^0(P_O)$。任取一个得分函数为 (a_E, a_O) 的光滑子模型。定义（B.1）给出

$$\begin{aligned}\left\langle \left(\frac{\psi_E}{\rho}, \frac{\psi_O}{1-\rho}\right), (a_E, a_O)\right\rangle_\rho &= \rho\mathbb{E}\left[\frac{\psi_E(Z_E)}{\rho} a_E(Z_E) \mid E\right] + (1-\rho)\mathbb{E}\left[\frac{\psi_O(Z_O)}{1-\rho} a_O(Z_O) \mid O\right] \\ &= \mathbb{E}[\psi_E(Z_E) a_E(Z_E) \mid E] + \mathbb{E}[\psi_O(Z_O) a_O(Z_O) \mid O]\end{aligned} \tag{B.5}$$

引理 B.4 表明， $\tau(P_\varepsilon)$ 在 $\varepsilon = 0$ 处右可微，且其右导数等于（B.3）的右端，而该右端正是（B.5）的右端。因此，右导数等于（B.5）左端的内积。由于路径是任意的，$(\psi_E/\rho, \psi_O/(1-\rho))$ 是 τ 的梯度。

我们证明论断 (2)。由于假设 B.2 中子空间的元素都是得分函数，因此它们属于所有得分函数线性组合构成的集合，所以其闭包包含于切空间中。根据假设 B.2(2)，$(\psi_E/\rho,\ \psi_O/(1-\rho))$ 属于该闭包，因而属于切空间；结合论断 (1)，根据定义，它就是有效影响函数。为证明唯一性，取切空间中的一个梯度 $(b_E,\ b_O)$。对于每个光滑子模型，这两个梯度与得分函数的内积都等于同一个右导数，因此对于得分函数 $(a_E,\ a_O)$，

$$\left\langle \left(b_E - \frac{\psi_E}{\rho}, b_O - \frac{\psi_O}{1-\rho}\right), (a_E, a_O)\right\rangle_\rho = 0 \tag{B.6}$$

由内积的线性性可知，(B.6) 对得分函数的任意线性组合也成立；又由内积在 $\|\cdot\|_\rho$ 下的连续性可知，对于这些线性组合的闭包中的任意元素，(B.6) 也成立，而该闭包就是切空间。$(b_E,\ b_O)$ 和 $(\psi_E/\rho,\ \psi_O/(1-\rho))$ 均属于切空间，因此它们的差也属于切空间；在 (B.6) 中用该差替换 $(a_E,\ a_O)$，得到

$$\begin{aligned}0 &= \left\langle \left(b_E - \frac{\psi_E}{\rho}, b_O - \frac{\psi_O}{1-\rho}\right), \left(b_E - \frac{\psi_E}{\rho}, b_O - \frac{\psi_O}{1-\rho}\right)\right\rangle_\rho \\ &= \rho\mathbb{E}\left[\left(b_E - \frac{\psi_E}{\rho}\right)^2 \mid E\right] + (1-\rho)\mathbb{E}\left[\left(b_O - \frac{\psi_O}{1-\rho}\right)^2 \mid O\right]\end{aligned}$$

两个系数均为正，且两个期望均为非负，因此两个期望均为零。因此，在 P_E 下， $b_E = \psi_E/\rho$ 几乎处处成立；在 P_O 下， $b_O = \psi_O/(1-\rho)$ 几乎处处成立。

我们证明论断 (3)。根据论断 (2)，有效影响函数为 $(\psi_E/\rho, \psi_O/(1-\rho))$，根据定义 (B.1)，其平方范数为

$$\begin{aligned}\left\|\left(\frac{\psi_E}{\rho}, \frac{\psi_O}{1-\rho}\right)\right\|_\rho^2 &= \rho\mathbb{E}\left[\left(\frac{\psi_E(Z_E)}{\rho}\right)^2 \mid E\right] + (1-\rho)\mathbb{E}\left[\left(\frac{\psi_O(Z_O)}{1-\rho}\right)^2 \mid O\right] \\ &= \frac{1}{\rho}\mathbb{E}[\psi_E(Z_E)^2 \mid E] + \frac{1}{1-\rho}\mathbb{E}[\psi_O(Z_O)^2 \mid O] = \Sigma^2\end{aligned} \tag{B.7}$$

其中，最后一个等式是 (3.7) 中 Σ^2 的定义。根据引理 B.2，Σ^2 是有限的。

我们证明论断 (4)。我们使用正则估计的标准卷积定理（van der Vaart 1998，第25章），即其适用于实验的局部渐近正态序列的形式。取假设 B.2 中子空间的有限个线性独立元素 $(a_E, 1, a_O, 1), \ldots, (a_{E,k}, a_{O,k})$。对于 $\mathrm{h} = (\mathrm{h}_1, \ldots, \mathrm{h}_k) \in \mathbb{R}^k$，由子空间的线性性质可知，$\sum_{l=1}^k \mathrm{h}_l(a_{E,l}, a_{O,l})$ 也属于该子空间，因此根据假设 B.2(1)，存在一个具有此得分函数的光滑子模型。对每个 h 选取这样一条路径，并令 $\mathrm{h} = 0$ 时路径为常值路径 $P_\varepsilon = P$，再令 $\varepsilon = 1/\sqrt{N}$ 处的概率律定义一个以 h 为索引的实验。这个实验序列具有三项性质。第一，将引理 B.3 应用于得分函数 $\sum_l \mathrm{h}_l(a_{E,l}, a_{O,l})$，可知对数似然比等于一个趋于正态分布的随机向量与 h 的内积，减去 $\frac{1}{2}\sum_{l,l'} \mathrm{h}_l \mathrm{h}_{l'}(a_{E,l}, a_{O,l}), (a_{E,l'}, a_{O,l'})_\rho$，误差为 $o_p(1)$；并且极限协方差矩阵的 (l, l') 项为 $(a_{E,l}, a_{O,l}), (a_{E,l'}, a_{O,l'})_\rho$。因此，该实验序列是局部渐近正态的，其信息由内积 (B.1) 给出。第二，沿着 $\varepsilon = 1/\sqrt{N}$ 计算论断 (1) 中的右导数，可知 $\sqrt{N}\{\tau(P_{1/\sqrt{N}}) - \tau(P)\}$ 收敛到 $\sum_l \mathrm{h}_l(\psi_E/\rho, \psi_O/(1-\rho)), (a_{E,l}, a_{O,l})_\rho$。第三，如果估计量序列相对于假设 B.2 中的子空间是正则的，那么在 $P_{1/\sqrt{N}}$ 下，$\sqrt{N}$ 乘以估计量与 $\tau(P_{1/\sqrt{N}})$ 之差的极限分布不依赖于 h。根据这三项性质，卷积定理表明，极限分布是某个概率分布与一个正态分布的卷积；后者的方差是 $(\psi_E/\rho, \psi_O/(1-\rho))$ 在 $(a_E, 1, a_O, 1), \ldots, (a_{E,k}, a_{O,k})$ 张成空间上的正交投影的平方范数。令 σ_k^2 表示该平方范数，则对每个实数 u，极限分布的特征函数等于 $\exp(-\sigma_k^2 u^2/2)$ 与卷积中另一因子的特征函数的乘积。

我们通过扩大元素的有限集合，将这一结论推广到 Σ^2。根据假设 B.2(2)，存在一个子空间中元素构成的序列，在 $\|\cdot\|_\rho$ 下收敛于（ψ_E/ρ，$\psi_O/(1-\rho)$），我们取其前 k 个元素的张成空间的一个基作为有限集合（如果该张成空间仅包含零元素，则 $\sigma_k^2 = 0$，上述结论可直接成立）。根据正交投影的性质，（ψ_E/ρ，$\psi_O/(1-\rho)$）的平方范数减去其投影的平方范数 σ_k^2

等于它们之差的平方范数。由于投影是该张成空间中距离最近的元素，这一差值不超过到序列中第 k个元素的平方距离，而该元素属于该张成空间；当 $k \to \infty$时，该差值趋于零。根据（B.7），$\sigma_k^2 \to \Sigma^2$。卷积中另一因子的特征函数等于极限分布的特征函数乘以 $\exp(\sigma_k^2 u^2/2)$，因此当 $k \to \infty$ 时，对于每个u，它都收敛于极限分布的特征函数乘以 $\exp(\Sigma^2 u^2/2)$。该函数在 $u = 0$ 处等于一且连续，因此根据Lévy连续性定理，它是某个概率分布的特征函数。因此，极限分布的特征函数等于 $\exp(-\Sigma^2 u^2/2)$ 与该概率分布的特征函数的乘积，而极限分布是均值为零、方差为 Σ^2 的正态分布与该概率分布的卷积。如果极限分布具有有限方差，则该方差至少为 Σ^2，因为两个独立随机变量之和的方差等于它们方差之和。这证明了定理 3.1 的四个断言。 □

本部分及附录 B.3.3 中的 (B.9) 所使用的卷积定理，是 vander Vaart （1998）第 25 章关于正则估计的标准结果。

B.3 引理 B.1–B.3 的证明

本小节依次证明附录 B.2.1 中的引理 B.1、B.2 和 B.3。

B.3.1 引理 *B.1* 的证明

引理 B.1 的证明。方程 (2.7) 对于 $F_{X|O}$-几乎处处的 x 成立，并且根据假设 3.2，$\mathcal{X}$ 中的每个点都具有正概率，因此它对每个 $x \in \mathcal{X}$ 都成立。论断 (1) 可通过将 (2.7) 的第一个方程应用于有界截断 $\min\{g,\ n\}$，并利用单调收敛定理令 $n \to \infty$ 而得出。论断 (2) 可通过将 (2.7) 的两个方程应用于双侧截断 $\max\{-n,\ \min\{g,\ n\}\}$，并利用控制收敛定理令 $n \to \infty$ 而得出。在实验侧和观察侧，控制函数分别为 $\max\{C,\ 1\}|g(M)|$ 和 $\max\{C,\ 1\}\mathbf{1}\{T = t\}|g(M)|$，其中常数来自假设 B.1 中的 $|Y| \le C$ 以及 $|\nu_t| \le C$。根据论断 (1) 针对 $|g|$ 给出的方程以及 $\pi_t \le 1$，观察侧的控制函数是可积的。

接下来证明论断（3）。根据迭代期望和（3.2），可得

$$
\begin{aligned}
&\mathbb{E}[\nu_t(S,x)\omega_t^{\nu}(M,x) \mid T = t, x, E] \\
&= \mathbb{E}[\nu_t(S,x)\mathbb{E}[\omega_t^{\nu}(M,x) \mid S, T = t, x, E] \mid T = t, x, E] = \mathbb{E}[\mu_t(S,x) \mid T = t, x, E], \\
&\mathbb{E}[\pi_t(S,x)\omega_t^{\pi}(M,x) \mid T = t, x, E] \\
&= \mathbb{E}[\pi_t(S,x)\mathbb{E}[\omega_t^{\pi}(M,x) \mid S, T = t, x, E] \mid T = t, x, E] = \mathbb{E}[\mu_t(S,x) \mid T = t, x, E]
\end{aligned}
\tag{B.8}
$$

右侧均等于 $\bar{\mu}_t(x)$，其中第一个方程中的最后一个等号使用了 $\nu_t/\pi_t = \mu_t$。根据 $|\nu_t| \le C$、$\pi_t \le 1$ 以及 $\omega_t^\nu, \omega_t^\pi$ 的平方可积性，每个期望都是有限的。由于平方可积函数是可积的，将论断（2）应用于 $g = \omega_t^\nu(\cdot, x)$ 和 $g = \omega_t^\pi(\cdot, x)$，再结合（B.8），可得

$$\mathbb{E}[\mathbf{1}\{T = t\}\{Y\omega_t^\nu(M, x) - \omega_t^\pi(M, x)\} \mid x, O] = \mathbb{E}[\nu_t(S, x)\omega_t^\nu(M, x) - \pi_t(S, x)\omega_t^\pi(M, x) \mid T = t, x, E]$$

$$= \bar{\mu}_t(x) - \bar{\mu}_t(x) = 0$$

最后，我们计算这两个函数的均值。由于观测到的（T, X）等于（t, x）的概率为 P（$T = t$, $X = x \mid E$），因此（3.3）中每一项的期望为 P（$X = x \mid O$）$\mathbb{E}[\mu_t$（S, x）$- \bar{\mu}_t$（x）$- \nu_t$（S, x）ω_t^ν（M, x）$+ \pi_t$（S, x）ω_t^π（M, x）$\mid T = t, x, E] = P$（$X = x \mid O$）$\{\bar{\mu}_t$（x）$- \bar{\mu}_t$（x）$- \bar{\mu}_t$（x）$+ \bar{\mu}_t$（x）$\} = 0$，因此 $\mathbb{E}[\psi_E$（Z_E）$\mid E] = 0$。在（3.4）中，$\mathbb{E}[\bar{\mu}_1$（X）$- \bar{\mu}_0$（X）$\mid O] = \tau$，并且在给定 $X = x$ 的情况下，上述等式给出 $\mathbb{E}[\psi_O$（Z_O）$\mid O] = 0$。这证明了论断（3）。 □

B.3.2 引理 *B.2* 的证明

引理 B.2 的证明。均值为零性质由引理 B.1(3) 得出，因此我们证明二阶矩有限。当观测到的 (T, X) 等于 (t, x) 时，(3.3) 中的求和仅剩一项，其权重 $P(X = x \mid O)/P(T = t, X = x \mid E)$ 根据假设 3.2(2) 至多为 c^{-1}。假设 B.1 给出 $|\mu_t|, |\bar{\mu}_t|, |\nu_t| \le C$ 和 $\pi_t \le 1$，而假设 3.4 使 $\omega_t^\nu, \omega_t^\pi$ 平方可积，因此剩余项平方可积，且 $\mathbb{E}[\psi_E(Z_E)^2 \mid E] < \infty$。对于 (3.4)，引理 B.1(1) 给出 $\mathbb{E}[\mathbf{1}\{T = t\}\omega_t^\nu(M, x)^2 \mid x, O] = \mathbb{E}[\pi_t(S, x)\omega_t^\nu(M, x)^2 \mid T = t, x, E] < \infty$，对 ω_t^π 也同样成立；结合 $|Y| \le C$ 和 $|\bar{\mu}_t| \le C$，可得 $\mathbb{E}[\psi_O(Z_O)^2 \mid O] < \infty$。 □

B.3.3 引理 *B.3* 的证明

引理 B.3 的证明。我们对每个样本应用二次均值可微路径的局部渐近正态性，并将两个展开式相加。对于附录 B.1 中的正则路径，$P_{G,\varepsilon}$ 相对于 P_G 绝对连续，因此似然比 $dP_{G,\varepsilon}/dP_G$ 是 $P_{G,\varepsilon}$ 相对于 P_G 的密度，且没有奇异部分。在 $P_{1/\sqrt{N}}$ 下两个样本的联合分布同样相对于在 P 下的联合分布绝对连续，其似然比是各个观测值的似然比的乘积。因此，左侧的 (B.4) 中的对数似然比就是两个样本联合分布的对数似然比。在附录 C.1.6 中应用 Le Cam 第三引理时会用到这一点。

—

我们将使用以下关于二次均值可微路径的标准结果（van der Vaart 1998，第25章）。设 W_1, ……, W_n 为独立同分布的

样本，其大小为 n，来自分布 Q；设 Q_ε 为具有得分函数$b \in L_2^0(Q)$的、形式为（B.2）的路径。对于满足当 n 足够大时 $s_n \geq 0$并收敛于 $s \geq 0$的实数序列 s_n，在 Q下，

$$\sum_{k=1}^{n} \log \frac{dQ_{s_n/\sqrt{n}}}{dQ}(W_k) = \frac{s_n}{\sqrt{n}} \sum_{k=1}^{n} b(W_k) - \frac{s^2}{2}\mathbb{E}_Q[b^2] + o_p(1) \tag{B.9}$$

我们将此应用于具有 $n = n_E$、$Q = P_E$、$b = a_E$ 和 $s_{n_E} = \sqrt{n_E/N}$ 的实验样本。于是 $s_{n_E}/\sqrt{n_E} = 1/\sqrt{N}$，而假设 3.1 给出 $s_{n_E} \to \sqrt{\rho}$。根据假设 3.1，实验样本是独立同分布的，因此由 (B.9) 得

$$\sum_{i=1}^{n_E} \log \frac{dP_{E,1/\sqrt{N}}}{dP_E}(Z_{E,i}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{n_E} a_E(Z_{E,i}) - \frac{\rho}{2}\mathbb{E}[a_E^2 \mid E] + o_p(1) \tag{B.10}$$

对于观察样本，$s_{n_O} = \sqrt{n_O/N} \to \sqrt{1-\rho}$，因此同理可得

$$\sum_{j=1}^{n_O} \log \frac{dP_{O,1/\sqrt{N}}}{dP_O}(Z_{O,j}) = \frac{1}{\sqrt{N}} \sum_{j=1}^{n_O} a_O(Z_{O,j}) - \frac{1-\rho}{2}\mathbb{E}[a_O^2 \mid O] + o_p(1) \tag{B.11}$$

将 (B.10) 和 (B.11) 相加，两个二次项之和为 $-\rho\mathbb{E}[a_E^2 \mid E]/2 - (1-\rho)\mathbb{E}[a_O^2 \mid O]/2 = -\|(a_E,\ a_O)\|_\rho^2/2$，由此得到 (B.4)。对于依分布收敛，将大括号中的项除以 $\sqrt{N}$ 可写为

$$\frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^{n_E} a_E(Z_{E,i}) + \sum_{j=1}^{n_O} a_O(Z_{O,j}) \right\} = \sqrt{\frac{n_E}{N}} \frac{1}{\sqrt{n_E}} \sum_{i=1}^{n_E} a_E(Z_{E,i}) + \sqrt{\frac{n_O}{N}} \frac{1}{\sqrt{n_O}} \sum_{j=1}^{n_O} a_O(Z_{O,j}) \tag{B.12}$$

根据中心极限定理，两个标准化和分别依分布收敛于均值为零、方差为 $\mathbb{E}[a_E^2 \mid E]$ 和 $\mathbb{E}[a_O^2 \mid O]$ 的正态分布，并且根据假设 3.1，它们相互独立。根据 $\sqrt{n_E/N} \to \sqrt{\rho}$、$\sqrt{n_O/N} \to \sqrt{1-\rho}$和 Slutsky 引理，该表达式依分布收敛于均值为零、方差为$\rho\mathbb{E}[a_E^2 \mid E] + (1-\rho)\mathbb{E}[a_O^2 \mid O] = \|(a_E,\ a_O)\|_\rho^2$的正态分布。 □

B.4 引理 B.4 的证明

引理 B.4 的证明。 我们沿光滑子模型对层份额和层均值求导，并将结果与影响函数和得分函数的乘积的期望进行比较。我们假设假设 2.1–2.8、附录 A 中的可积性条件、假设 3.2–3.4 以及假设 B.1 成立，并固定一个得分函数为 $(a_E,\ a_O)$、余项为 $r_{E,\,\varepsilon},\ r_{O,\,\varepsilon}$ 的光滑子模型 $\{P_\varepsilon\}$，如附录 B.1 所述。我们取 ε 足够小，使得密度

$dP_{G,\,\varepsilon}/dP_G$ 在两个样本中均介于 1/2 和 3/2 之间，记 $\mathbb{E}_{P_\varepsilon}$ 为在 P_ε 下的期望，并且当 $t,\ x$ 固定时，省略函数的自变量 $(S,\ x)$ 和 $(M,\ x)$。

沿着该路径，τ 再次表示为层份额和层均值的组合。对于$\bar{\mu}_{t,\ \varepsilon}(x) = \mathbb{E}_{P_\varepsilon}[\mu_{t,\varepsilon}(S,x) \mid T=t,\ x,\ E]$，将定理 2.1 的 (2.8) 应用于 P_ε，得到 $\tau(P_\varepsilon) = \sum_{x\in\mathcal{X}} P_\varepsilon(X=x \mid O)\{\bar{\mu}_{1,\ \varepsilon}(x) - \bar{\mu}_{0,\ \varepsilon}(x)\}$。我们通过条件协方差 $\mathrm{Cov}(f, a_E \mid T=t,x,E) = \mathbb{E}[f a_E \mid T=t,x,E] - \mathbb{E}[f \mid T=t,x,E]\,\mathbb{E}[a_E \mid T=t,x,E]$表示导数；对于任意满足 $\mathbb{E}[|f| \mid T=t,x,E] < \infty$ 的函数 $f(S,M)$，该协方差都是有限的，因为 a_E 有界；对于关于(T,M,Y)的函数 g，我们以同样的方式定义 $\mathrm{Cov}(g, a_O \mid x, O)$。给定 $T=t,\ X=x$，在 P_ε 下的实验分布相对于 P 下实验分布的密度，是$1 + \varepsilon a_E + \varepsilon r_{E,\ \varepsilon}$与其条件均值之比；该比值等于 $1 + \varepsilon\{a_E - \mathbb{E}[a_E \mid T=t,x,E]\}$，差一个本质上确界为 $o(\varepsilon)$ 的函数。因此，对于每个这样的 f，以及每个满足 $\mathbb{E}[|(f_\varepsilon - f)/\varepsilon - \dot{f}| \mid T=t,x,E] \to 0$ 的、对应于某个可积 $\dot{f}$ 的族 f_ε，

$$
\begin{aligned}
\lim_{\varepsilon\downarrow 0} \frac{\mathbb{E}_{P_\varepsilon}[f \mid T=t,x,E] - \mathbb{E}[f \mid T=t,x,E]}{\varepsilon} &= \mathrm{Cov}(f, a_E \mid T=t,x,E),\\
\lim_{\varepsilon\downarrow 0} \frac{\mathbb{E}_{P_\varepsilon}[f_\varepsilon \mid T=t,x,E] - \mathbb{E}[f \mid T=t,x,E]}{\varepsilon} &= \mathbb{E}[\dot{f} \mid T=t,x,E] + \mathrm{Cov}(f, a_E \mid T=t,x,E)
\end{aligned}
\tag{B.13}
$$

其中，第二个极限由第一个极限以及条件密度的有界性推出。在观察样本中给定 $X=x$ 时也有相同的极限，此时用 a_O 替代 a_E。

我们首先将权重函数作为检验函数，沿路径对 (2.7) 求导。由于 $|Y| \le C$ 在 P_ε 下也成立，而 P_ε 相对于 P 绝对连续，并且 $\mu_{t,\ \varepsilon}$ 是 Y 的条件均值，因此界 $|\mu_{t,\ \varepsilon}| \le C$ 和 $|\nu_{t,\ \varepsilon}| \le C$ 在真实分布下几乎处处成立。方程 (2.7) 在 P_ε 下成立，因为 P_ε 属于 $\mathcal{P}$；将引理 B.1 的证明中的截断论证应用于 P_ε，并使用这些界，可将其推广至 $h = \omega_t^\nu$ 和 $h = \omega_t^\pi$。这两个函数在 P 下平方可积，因此在 P_ε 下也平方可积。我们用 (B.13) 对推广后方程的两边求导。在左边，函数 $\nu_{t,\ \varepsilon}\omega_t^\nu$，$\pi_{t,\ \varepsilon}\omega_t^\pi$ 和实验分布都会变化；由柯西–施瓦茨不等式可知，这些函数的差商在绝对均值意义下收敛于 $\dot{\nu}_t\omega_t^\nu$，$\dot{\pi}_t\omega_t^\pi$。在右边，只有观察分布发生变化。令两边的导数相等，得到

$$
\begin{aligned}
\mathbb{E}[\dot{\nu}_t\omega_t^\nu \mid T=t,x,E] &= \mathrm{Cov}(\mathbf{1}\{T=t\}Y\omega_t^\nu, a_O \mid x, O) - \mathrm{Cov}(\nu_t\omega_t^\nu, a_E \mid T=t,x,E),\\
\mathbb{E}[\dot{\pi}_t\omega_t^\pi \mid T=t,x,E] &= \mathrm{Cov}(\mathbf{1}\{T=t\}\omega_t^\pi, a_O \mid x, O) - \mathrm{Cov}(\pi_t\omega_t^\pi, a_E \mid T=t,x,E)
\end{aligned}
\tag{B.14}
$$

其中，根据引理 B.1(1)，观测协方差是有限的。

接下来，我们对分层均值求微分。恒等式 $\nu_{t,\ \varepsilon} = \pi_{t,\ \varepsilon}\mu_{t,\ \varepsilon}$ 和 $\nu_t = \pi_t\mu_t$ 给出$\pi_t(\mu_{t,\ \varepsilon} - \mu_t) = (\nu_{t,\ \varepsilon} - \nu_t) - \mu_{t,\ \varepsilon}(\pi_{t,\ \varepsilon} - \pi_t)$。由于 $\pi_t \ge a, |\mu_{t,\ \varepsilon}| \le C$，并且.....的差商

π_t和 ν_t在 $\|\cdot\|_{2,\,t,\,x}$中收敛，因此该恒等式给出 $\|\mu_{t,\,\varepsilon}-\mu_t\|_{2,\,t,\,x}\le C\varepsilon$。将 $\mu_{t,\,\varepsilon}=\mu_t+(\mu_{t,\,\varepsilon}-\mu_t)$写在其右端，并除以 $\varepsilon\pi_t$，得到

$$\frac{\mu_{t,\varepsilon}-\mu_t}{\varepsilon}=\frac{1}{\pi_t}\cdot\frac{\nu_{t,\varepsilon}-\nu_t}{\varepsilon}-\frac{\mu_t}{\pi_t}\cdot\frac{\pi_{t,\varepsilon}-\pi_t}{\varepsilon}-\frac{(\mu_{t,\varepsilon}-\mu_t)(\pi_{t,\varepsilon}-\pi_t)}{\varepsilon\pi_t}$$

且柯西–施瓦茨不等式将最后一项的绝对值均值界定为 $C\varepsilon$。因此， μ_t 的差商在绝对值均值意义下收敛于 $(\dot\nu_t-\mu_t\dot\pi_t)/\pi_t$，而（B.13）中的第二个极限给出

$$\left.\frac{d\bar\mu_{t,\varepsilon}(x)}{d\varepsilon}\right|_{\varepsilon=0+}=\operatorname{Cov}(\mu_t,a_E\mid T=t,x,E)+\mathbb{E}\left[\frac{\dot\nu_t}{\pi_t}-\frac{\mu_t\dot\pi_t}{\pi_t}\,\middle|\,T=t,x,E\right]$$

由于 $\dot\nu_t$ 和 $\dot\pi_t$ 是 S的函数，迭代期望和（3.2）将 $1/\pi_t$ 和 μ_t/π_t 分别替换为 ω_t^ν和 ω_t^π，因此最后一个期望等于 $\mathbb{E}[\dot\nu_t\omega_t^\nu-\dot\pi_t\omega_t^\pi\mid T=t,x,E]$，而（B.14）给出

$$\begin{aligned}\left.\frac{d\bar\mu_{t,\varepsilon}(x)}{d\varepsilon}\right|_{\varepsilon=0+}&=\operatorname{Cov}(\mu_t-\nu_t\omega_t^\nu+\pi_t\omega_t^\pi,a_E\mid T=t,x,E)\\&\quad+\operatorname{Cov}(\mathbf{1}\{T=t\}(Y\omega_t^\nu-\omega_t^\pi),a_O\mid x,O)\end{aligned}\tag{B.15}$$

最后，我们对 $\tau(P_\varepsilon)$求导，并将结果与期望的乘积进行比较。层份额满足 $\{P_\varepsilon(X=x\mid O)-P(X=x\mid O)\}/\varepsilon=\mathbb{E}[\mathbf{1}\{X=x\}(a_O+r_{O,\varepsilon})\mid O]\to P(X=x\mid O)\mathbb{E}[a_O\mid x,O]$。由于 $\mathcal{X}$是有限的，乘积法则和（B.15）给出

$$\begin{aligned}\left.\frac{d\tau(P_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0+}&=\sum_{x\in\mathcal{X}}P(X=x\mid O)\{\bar\mu_1(x)-\bar\mu_0(x)\}\,\mathbb{E}[a_O\mid x,O]\\&\quad+\sum_{x\in\mathcal{X}}P(X=x\mid O)\sum_{t\in\{0,1\}}(2t-1)\operatorname{Cov}(\mu_t-\nu_t\omega_t^\nu+\pi_t\omega_t^\pi,a_E\mid T=t,x,E)\\&\quad+\sum_{x\in\mathcal{X}}P(X=x\mid O)\sum_{t\in\{0,1\}}(2t-1)\operatorname{Cov}(\mathbf{1}\{T=t\}(Y\omega_t^\nu-\omega_t^\pi),a_O\mid x,O)\end{aligned}$$

第二个求和等于 $\mathbb{E}[\psi_E(Z_E)a_E(Z_E)\mid E]$，而第一个和第三个求和相加得到$\mathbb{E}[\psi_O(Z_O)a_O(Z_O)\mid O]$，从而得到（B.3）。事实上，在（3.3）中，带符号计入的(t,x)项仅当$(T,X)=(t,x)$时才非零，并且根据引理 B.1(3)，花括号中的表达式在给定 $T=t,X=x$ 时条件均值为零，因此该项与 a_E 的乘积的期望等于$(2t-1)P(X=x\mid O)$乘以第二个求和中的条件协方差，由此常数 $\bar\mu_t(x)$消去。在（3.4）中， $\mathbb{E}[\tau a_O(Z_O)\mid O]=0$，因为得分函数的均值为零；项 $\bar\mu_1(X)-\bar\mu_0(X)$给出第一个求和；而项$\mathbf{1}\{T=t\}\{Y\omega_t^\nu(M,X)-\omega_t^\pi(M,X)\}$根据引理 B.1(3)在给定 $X=x$ 时条件均值为零，因此它们与 a_O 在给定$X=x$ 时的乘积期望就是第三个求和中的条件协方差。 □

附录 C 定理 3.2 和定理 3.3 的证明

本附录证明第 3 节的定理 3.2 和 3.3，这些定理给出了不进行样本拆分的估计量和交叉拟合估计量的渐近分布与效率。表 C.1 列出了各小节的内容，图 C.1 展示了证明中所使用的结果之间的相互依赖关系。

图 C.1 中箭头的含义与图 B.1 中相同。定理 3.2 的证明核心是引理 C.2，该引理将层均值中的误差分解为在真实函数处计算的样本均值，以及一个以快于 $N^{-1/2}$ 的速度趋于零的余项。在该余项的各项中，当估计量固定不变并将样本均值替换为期望时产生的差异，可通过引理 C.1 加以界定；该引理将其表示为估计误差乘积的形式。引理 C.2 给出了估计量的渐近线性表示（附录 C.1.4）；结合引理 C.3 中方差估计量的一致性，可得到学生化估计量的极限。附录 C.1.6 利用引理 B.2–B.4 和定理 3.1 证明了效率，其中定理 3.1 在附录 B 中得到证明。

定理 3.3 的证明再次针对交叉拟合估计量证明了引理 C.2 和 C.3 的结论，并运用了附录 C.1.4 和 C.1.6 中的论证。根据引理 C.4，在给定某一折之外的观测值的条件下，由代入留出一个折的估计量所产生的项以及方差估计量的均方误差均有界；引理 C.5（附录 C.2.4）则将各折平均值转化为全样本上的求和。

假设 C.1 (估计量的有界性)。 对于假设 *B.1* 中的常数 C，以下结论对每个 $t \in \{0、1\}$ 以及 $x \in \mathcal{X}$ 均成立。

1. 在概率趋于 *1* 的事件上，第 *3.2* 节的全样本估计量满足$|\widehat{v}_t(s,\ x)| \leq C$，对所有 s 均成立。

2. 对于每个 k，在概率趋于 *1* 的事件上，第*3.4*节中留出 k 折的估计量满足 $|\widehat{v}_t^{(-k)}\ (s,\ x)| \leq C$ 对所有 s 均成立。

小节	内容
C.1	通过概率趋于 1 的事件（C.1.1）证明定理 3.2， 引理 C.1（C.1.2）、引理 C.2（C.1.3）以及渐近线性表示（C.1.4）， 引理 C.3（C.1.5），以及学生化和效率（C.1.6）
C.2	定义概率趋于 1 的事件（C.2.1），陈述并证明引理 C.4（C.2.2）和引理 C.5（C.2.3），随后证明定理 3.3（C.2.4）

表 C.1 附录 C 的小节及其内容

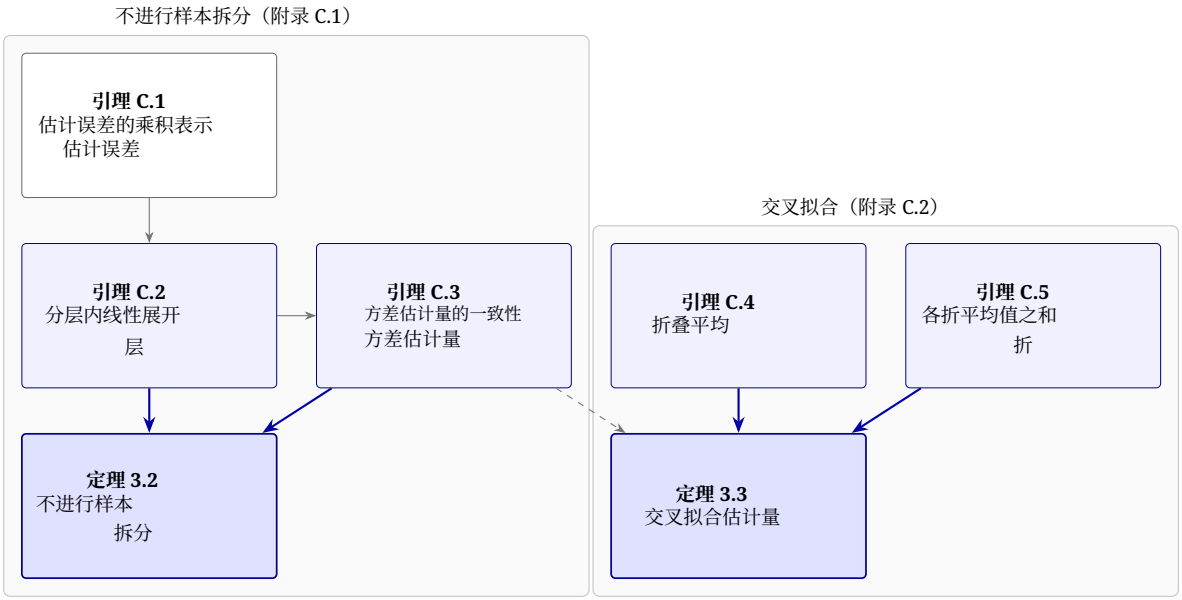

图 C.1 用于证明定理 3.2 和 3.3 的各项结果之间的依赖关系注：粗箭头从直接导出这些定理的引理出发。虚线箭头表示，定理 3.3 的证明逐折应用引理 C.2 和 C.3 的论证。概率趋于 1 的事件（附录 C.1.1 和 C.2.1）在相应定理的全部证明中均被使用。

假设 C.1 的条件 (1) 对全样本估计量进行界定，而条件 (2) 对留出一个折的估计量进行界定。定理 3.2 使用假设 B.1 和条件 (1)，定理 3.3 使用假设 B.1 和条件 (2)。本附录中的证明使用了附录 B 开头选定的 π_t，μ_t 的有界版本，以及正文中的方程 (2.7)、(3.2)、(3.6) 和 (3.7)；这些方程已在附录 B 开头重新列出。

在本附录中，C 表示一个不依赖于 N 的正数常数，其值在每次出现时可能有所不同。常数 C 可以依赖于 K、ρ、a、c以及各假设中的常数，并且取为至少不小于假设 B.1 和 C.1 中的常数，因此这些假设中的不等式对该 C同样成立。

C.1 定理 3.2 的证明

本小节证明第 3.3 节的定理 3.2。定理 3.2 的目标是证明 (3.8) 中的两个收敛结果以及该区间的覆盖概率。我们首先将估计量与真实值之差拆分为三项，即观测分层份额的误差、分层均值的误差，以及二者的乘积。层份额的误差是观察样本的样本均值，对应 (3.4) 中的部分 $\bar{\mu}_1(X) - \bar{\mu}_0(X) - \tau$。引理 C.2 将分层均值的误差拆分为五项。在真实函数处计算的两个样本均值给出了 (3.3) 中函数的样本均值以及 (3.4) 中剩余的

部分。根据假设 3.6(1)–(5)，由代入估计值产生的两项以快于 $N^{-1/2}$ 的速度趋于零；根据附录 C.1.2 的引理 C.1 以及假设 3.5(2)–(4)，当估计值固定、并以期望替代样本均值时产生的偏差 $\bar{\mu}_t(x)$ 也以同样的速度趋于零。两项误差的乘积为 $O_p(N^{-1})$。接着，我们对由此得到的两个样本均值之和应用中心极限定理，并对方差估计量应用引理 C.3，该引理使用假设 3.6(6)–(11)。最后，Slutsky 引理给出学生化估计量的极限以及该区间的覆盖概率。附录 C.2 中关于交叉拟合的证明有两处不同：由代入估计值产生的两项的界，以及方差估计量的界。

C.1.1 概率趋近于一的事件

在下文的证明中，我们只考虑以下三个条件同时对所有 t 和 x 成立的事件$\widehat{\ }$，并将其称为附录 C.1.1 的事件。(E1) $P(T = t, X = x \mid E) \gtrsim c/2$，且 $P(X = x \mid O) \geq c/2$。(E2) 假设 C.1(1) 中的事件发生，也就是说，某个概率趋近于一的事件，在该事件上 $|\widehat{\nu}_t(s, x)| \leq C$ 对所有 s 均成立。(E3) $\|\widehat{\omega}_t^{\nu}\|_{2,t,x}$ 和 $\|\widehat{\omega}_t^{\pi}\|_{2,t,x}$ 均有限。这里，c 是假设 3.2(2) 中的常数，C 是假设$\widehat{\ }$C.1(1) 中界的常数。(E1) 的概率趋近于一。例如，$P(T = t, X = x \mid E)$ 是一个样本均值，其均值为 $P(T = t, X = x \mid E) \geq c$，方差至多为 $1/(4n_E)$；假设 3.1 将 n_E 下界限定为一个正数常数乘以 N，因此根据切比雪夫不等式，$\widehat{\ }$ $\widehat{\ }P(T = t, X = x \mid E) < c/2$ 的概率趋于零。对 $P(X = x \mid O)$ 也同样成立。根据假设 C.1(1)，(E2) 的概率趋近于一。对于 (E3)，闵可夫斯基不等式给出 $\|\widehat{\omega}_t^{\nu}\|_{2,t,x} \leq \|\omega_t^{\nu}\|_{2,t,x} + \|\widehat{\omega}_t^{\nu} - \omega_t^{\nu}\|_{2,t,x}$；右侧第一项根据假设 3.4 有限，第二项根据假设 3.5(1) 依概率趋于零，因此范数有限的概率趋近于一。对 $\widehat{\omega}_t^{\pi}$ 也同样成立。由于有限个事件的概率都趋近于一，它们同时成立的概率也趋近于一。在该事件上，(3.5)、(3.6) 以及方差估计量取均值时所涉及的所有组均非空。由于其补事件的概率趋于零，在该事件上得到的依概率收敛和依分布收敛，在不限制于该事件时仍然成立。

C.1.2 乘积余项

本部分陈述并证明引理 C.1。该引理通过估计误差的乘积，表示在估计量保持不变、并将式 (3.5) 中的两个样本均值替换为期望时，由此产生的与 $\bar{\mu}_t(x)$ 的差异。该引理用于附录 C.1.3 中引理 C.2 的证明。

引理 C.1（估计误差的乘积表示）。假设假设 *2.1–2.8*、附录 *A* 中的可积性条件、假设 *3.1–3.4* 以及假设 *B.1* 均成立，并固定 t、x。在样本值保持不变的情况下，假设 $|\widehat{\nu}_t(s,\ x)| \le C$ 对所有 s 均成立，并且 $\|\widehat{\omega}_t^{\nu}\|_{2,\ t,\ x}$ 和 $\|\widehat{\omega}_t^{\pi}\|_{2,\ t,\ x}$ 均为有限值。在式 *(3.5)* 中，

$$\begin{aligned}\widehat{\bar{\mu}}_t(x) = \widehat{\mathbb{E}}[\widehat{\mu}_t(S,x) - \widehat{\nu}_t(S,x)\widehat{\omega}_t^{\nu}(M,x) + \widehat{\pi}_t(S,x)\widehat{\omega}_t^{\pi}(M,x) \mid T=t,x,E] \\ + \widehat{\mathbb{E}}[\mathbf{1}\{T=t\}\{Y\widehat{\omega}_t^{\nu}(M,x) - \widehat{\omega}_t^{\pi}(M,x)\} \mid x,O]\end{aligned}$$

$\widehat{\ }$将两个样本均值 $\mathbb{E}$ 替换为相应期望 $\mathbb{E}$ 并减去 $\bar{\mu}_t(x)$ 得到

$$\begin{aligned}&\mathbb{E}\left[\frac{\{\pi_t(S,x) - \widehat{\pi}_t(S,x)\}\{\widehat{\mu}_t(S,x) - \mu_t(S,x)\}}{\pi_t(S,x)}\,\middle|\, T=t,x,E\right] \\ &\quad + \mathbb{E}[\{\nu_t(S,x) - \widehat{\nu}_t(S,x)\}\{\widehat{\omega}_t^{\nu}(M,x) - \omega_t^{\nu}(M,x)\} \mid T=t,x,E] \\ &\quad - \mathbb{E}[\{\pi_t(S,x) - \widehat{\pi}_t(S,x)\}\{\widehat{\omega}_t^{\pi}(M,x) - \omega_t^{\pi}(M,x)\} \mid T=t,x,E]\end{aligned} \tag{C.1}$$

引理 C.1 的证明所进行的计算将估计量视为固定函数，因此该引理适用于全样本估计量的每一种实现。右侧第一项是选择概率误差与比率误差的乘积，第二项和第三项分别是干扰函数误差与权重函数误差乘积的均值。

引理 C.1 的证明。固定 t、x 并保持样本值不变。由于 $\widehat{\omega}_t^{\nu}(\cdot,\ x)$ 和 $\widehat{\omega}_t^{\pi}(\cdot,\ x)$ 是平方可积的，因而也是可积的，故引理 B.1(2) 将观测期望$\mathbb{E}[\mathbf{1}\{T = t\}\{Y\widehat{\omega}_t^{\nu}(M,x) - \widehat{\omega}_t^{\pi}(M,x)\} \mid x,O]$ 转化为 $\mathbb{E}[\nu_t(S,x)\widehat{\omega}_t^{\nu}(M,x) - \pi_t(S,x)\widehat{\omega}_t^{\pi}(M,x) \mid T=t,x,E]$。将乘以相同 $\widehat{\omega}$ 的各项收集起来可知，将 (3.5) 中的两个样本均值替换为期望后所得的量，减去 $\bar{\mu}_t(x)$，等于

$$\begin{aligned}&\mathbb{E}[\widehat{\mu}_t(S,x) \mid T=t,x,E] - \bar{\mu}_t(x) + \mathbb{E}[\{\nu_t(S,x) - \widehat{\nu}_t(S,x)\}\widehat{\omega}_t^{\nu}(M,x) \mid T=t,x,E] \\ &\quad - \mathbb{E}[\{\pi_t(S,x) - \widehat{\pi}_t(S,x)\}\widehat{\omega}_t^{\pi}(M,x) \mid T=t,x,E]\end{aligned} \tag{C.2}$$

根据 $|\widehat{\nu}_t| \le C$，由 $\widehat{\pi}_t \ge a$、$|\nu_t| \le C$、$\widehat{\pi}_t$ 所蕴含的界 $|\widehat{\mu}_t| \le C/a$，$\pi_t \le 1$，以及 $\widehat{\omega}_t^{\nu}$ 的平方可积性，$\widehat{\omega}_t^{\pi}$，ω_t^{ν}，ω_t^{π}，每个期望都是有限的。

我们将 $\widehat{\omega}_t^{\nu} = \omega_t^{\nu} + (\widehat{\omega}_t^{\nu} - \omega_t^{\nu})$ 和 $\widehat{\omega}_t^{\pi} = \omega_t^{\pi} + (\widehat{\omega}_t^{\pi} - \omega_t^{\pi})$ 写在 (C.2) 的最后两项中。$\widehat{\omega}_t^{\nu} - \omega_t^{\nu}$ 和 $\widehat{\omega}_t^{\pi} - \omega_t^{\pi}$ 部分给出 (C.1) 的最后两项。对于 ω_t^{ν}和 ω_t^{π} 部分，省略参数 (S,x) 和 (M,x) 后，由迭代期望和 (3.2) 可得 $\mathbb{E}[(\nu_t - \widehat{\nu}_t)\omega_t^{\nu} \mid T=t,x,E] =\mathbb{E}[(\nu_t - \widehat{\nu}_t)/\pi_t \mid T=t,x,E]$ 和 $\mathbb{E}[(\pi_t - \widehat{\pi}_t)\omega_t^{\pi} \mid T=t,x,E] = \mathbb{E}[(\pi_t - \widehat{\pi}_t)\mu_t/\pi_t \mid T=t,x,E]$。

加上 $\bar{\mu}_t(x) = \mathbb{E}[\mu_t \mid T = t, x, E]$ 后，其余各项是关于 S 的某个函数的条件均值，$\nu_t = \pi_t \mu_t$ 和 $\widehat{\nu}_t = \widehat{\pi}_t \widehat{\mu}_t$ 将其简化为

$$\begin{aligned}&\widehat{\mu}_t - \mu_t + \frac{\pi_t \mu_t - \widehat{\pi}_t \widehat{\mu}_t}{\pi_t} - \frac{(\pi_t - \widehat{\pi}_t)\mu_t}{\pi_t} \\ &= \frac{(\pi_t - \widehat{\pi}_t)\widehat{\mu}_t}{\pi_t} - \frac{(\pi_t - \widehat{\pi}_t)\mu_t}{\pi_t} = \frac{(\pi_t - \widehat{\pi}_t)(\widehat{\mu}_t - \mu_t)}{\pi_t}\end{aligned}$$

这是 (C.1) 中第一项的函数。 □

C.1.3 分层内线性展开

引理 C.2 （分层内线性展开）。 假设假设 *2.1–2.8*、附录 *A* 中的可积性条件、假设 *3.1–3.5*、假设*3.6(1)–(5)*、假设 *B.1* 以及假设 *C.1(1)*均成立。那么，对于每个 $t \in \{0、1\}$ 以及 $x \in \mathcal{X}$，

$$\begin{aligned}&\widehat{\bar{\mu}}_t(x) - \bar{\mu}_t(x) \\ &= \frac{1}{n_E} \sum_{i=1}^{n_E} \frac{\mathbf{1}\{T_i = t, X_i = x\}}{P(T = t, X = x \mid E)} \{\mu_t(S_i, x) - \bar{\mu}_t(x) - \nu_t(S_i, x)\omega_t^{\nu}(M_i, x) + \pi_t(S_i, x)\omega_t^{\pi}(M_i, x)\} \\ &\quad + \frac{1}{n_O} \sum_{j=1}^{n_O} \frac{\mathbf{1}\{X_j = x\}}{P(X = x \mid O)} \mathbf{1}\{T_j = t\}\{Y_j \omega_t^{\nu}(M_j, x) - \omega_t^{\pi}(M_j, x)\} + o_p(N^{-1/2})\end{aligned} \tag{C.3}$$

右侧的两个和均为 O_p（$N^{-1/2}$）。

引理 C.2 的证明。 固定 t， x。目标是证明 (C.3)。我们将$\widehat{\bar{\mu}}_t(x)$ 与$\bar{\mu}_t(x)$ 之差拆分为五项，即在真实函数处求得的两个样本均值、由于代入估计量而导致的两个样本均值的变化，以及在估计量保持不变并将样本均值替换为期望时，由 $\bar{\mu}_t(x)$ 产生的差异。真实函数的样本均值构成 (C.3) 中的两个和。根据假设 3.6(1)–(5)，由估计量引起的变化以比 $N^{-1/2}$ 更快的速度趋于零；根据附录 C.1.2 的引理 C.1，该差异也以更快的速度趋于零，因为该引理将其改写为估计误差乘积的均值，而这又由假设 3.5(2)–(4) 得出。

下文中，我们在附录 C.1.1 的事件上进行讨论。在该事件上，下面的所有期望均为有限值，因此可以逐项拆分。事实上，(E2) 和 $\widehat{\pi}_t \geq a$ 给出 $|\widehat{\mu}_t| \leq C/a$，而实验分层中的期望由 (E2)、(E3)、假设 3.4、假设 B.1 以及柯西–施瓦茨不等式可知是有限的。观察分层中的期望则由引理 B.1(2) 应用于 $\widehat{\omega}_t^{\nu}(\cdot, x)$、 $\widehat{\omega}_t^{\pi}(\cdot, x)$、 $\omega_t^{\nu}(\cdot, x)$ 和 $\omega_t^{\pi}(\cdot, x)$ 得到，它们均为平方可积

因而也是可积的。在式 (3.5) 的每一项中减去并加上包含真实函数的项，可得

$$
\begin{aligned}
\widehat{\bar{\mu}}_t(x) - \bar{\mu}_t(x) = & \underbrace{\widehat{\mathbb{E}}[\mu_t(S,x) - \bar{\mu}_t(x) - \nu_t(S,x)\omega_t^{\nu}(M,x) + \pi_t(S,x)\omega_t^{\pi}(M,x) \mid T=t,x,E]}_{\text{(I)}} \\
& + \underbrace{\widehat{\mathbb{E}}[\mathbf{1}\{T=t\}\{Y\omega_t^{\nu}(M,x) - \omega_t^{\pi}(M,x)\} \mid x,O]}_{\text{(II)}} \\
& \underbrace{\begin{array}{l}(\widehat{\mathbb{E}}-\mathbb{E})[\widehat{\mu}_t(S,x) - \mu_t(S,x) \mid T=t,x,E] \\ -(\widehat{\mathbb{E}}-\mathbb{E})[\widehat{\nu}_t(S,x)\widehat{\omega}_t^{\nu}(M,x) - \nu_t(S,x)\omega_t^{\nu}(M,x) \mid T=t,x,E] \\ +(\widehat{\mathbb{E}}-\mathbb{E})[\widehat{\pi}_t(S,x)\widehat{\omega}_t^{\pi}(M,x) - \pi_t(S,x)\omega_t^{\pi}(M,x) \mid T=t,x,E]\end{array}}_{\text{(III)}} \\
& + \underbrace{\begin{array}{l}(\widehat{\mathbb{E}}-\mathbb{E})[\mathbf{1}\{T=t\}Y\{\widehat{\omega}_t^{\nu}(M,x) - \omega_t^{\nu}(M,x)\} \mid x,O] \\ -(\widehat{\mathbb{E}}-\mathbb{E})[\mathbf{1}\{T=t\}\{\widehat{\omega}_t^{\pi}(M,x) - \omega_t^{\pi}(M,x)\} \mid x,O]\end{array}}_{\text{(IV)}} \\
& + \underbrace{\begin{array}{l}\mathbb{E}[\widehat{\mu}_t(S,x) - \widehat{\nu}_t(S,x)\widehat{\omega}_t^{\nu}(M,x) + \widehat{\pi}_t(S,x)\widehat{\omega}_t^{\pi}(M,x) \mid T=t,x,E] \\ +\mathbb{E}[\mathbf{1}\{T=t\}\{Y\widehat{\omega}_t^{\nu}(M,x) - \widehat{\omega}_t^{\pi}(M,x)\} \mid x,O] - \bar{\mu}_t(x)\end{array}}_{\text{(V)}}
\end{aligned}
\tag{C.4}
$$

为验证 (C.4)，注意到根据引理 B.1(3) 和 $\mathbb{E}[\mu_t \mid T=t,x,E] = \bar{\mu}_t(x)$，函数$\mu_t - \nu_t\omega_t^{\nu} + \pi_t\omega_t^{\pi}$ 在给定 $T=t, X=x$时的条件均值为 $\bar{\mu}_t(x)$；而函数 $\mathbf{1}\{T=t\}\{Y\omega_t^{\nu} - \omega_t^{\pi}\}$ 在给定 $X=x$ 时的条件均值为零。这里省略了自变量 (S,x) 和$\widehat{\ }(M,x)$。因此，项 (I) 与 (III) 之和为 $(\widehat{\mathbb{E}}-\mathbb{E})[\widehat{\mu}_t - \widehat{\nu}_t\widehat{\omega}_t^{\nu} + \widehat{\pi}_t\widehat{\omega}_t^{\pi} \mid T=t,x,E]$，项 (II) 与 (IV) 的$\widehat{\ }$和为 $(\widehat{\mathbb{E}}-\mathbb{E})[\mathbf{1}\{T=t\}\{Y\widehat{\omega}_t^{\nu} - \widehat{\omega}_t^{\pi}\} \mid x,O]$。加上项 (V) 后，两个期望相互抵消，剩下 (3.5) 右侧减去 $\bar{\mu}_t(x)$，这正是 (C.4) 的左侧。项 (I) 和 (II) 是在真实函数处计算的样本均值，项 (III) 和 (IV) 是由于代入估计量而导致的样本均值变化，而项 (V) 是附录 C.1.2 中引理 C.1 的左侧。

对于项（I）和（II），我们使用以下等式。假设实验观测的函数 f 满足 $\mathbb{E}[f(Z_E) \mid T=t, x, E] = 0$ 且 $\mathbb{E}[f(Z_E)^2 \mid T=t, x, E] < \infty$。

根据样本均值的定义，$\widehat{\mathbb{E}}[f(Z_E) \mid T=t,\ x,\ E] = n_E^{-1}\sum_{i=1}^{n_E}\mathbf{1}\{T_i=t,\ X_i=x\}f(Z_{E,i})/P(T=t,\ X=x \mid E)$，因此

$$\begin{aligned}&\widehat{\mathbb{E}}[f(Z_E)\mid T=t,x,E]-\frac{1}{n_E}\sum_{i=1}^{n_E}\frac{\mathbf{1}\{T_i=t,X_i=x\}f(Z_{E,i})}{P(T=t,X=x\mid E)}\\&=\frac{1}{n_E}\sum_{i=1}^{n_E}\mathbf{1}\{T_i=t,X_i=x\}f(Z_{E,i})\cdot\frac{P(T=t,X=x\mid E)-\widehat{P}(T=t,X=x\mid E)}{\widehat{P}(T=t,X=x\mid E)P(T=t,X=x\mid E)}\end{aligned}\tag{C.5}$$

由于各项 $\mathbf{1}\{T_i=t,X_i=x\}f(Z_{E,i})$ 相互独立且同分布，其均值为$P(T=t,X=x\mid E)\mathbb{E}[f(Z_E)\mid T=t,x,E]=0$，方差为 $P(T=t,X=x\mid E)\mathbb{E}[f(Z_E)^2\mid T=t,x,E]$，因此假设 3.1 和切比雪夫不等式表明，右侧的第一个因子为 $O_p(N^{-1/2})$。类似地，$\widehat{P}(T=t,X=x\mid E)-P(T=t,X=x\mid E)$ 为 $O_p(N^{-1/2})$，且在附录 C.1.1 的事件上，分母至少为 $c^2/2$，因此 (C.5) 的右侧为 $O_p(N^{-1})$。出于同样的原因，(C.5) 左侧的和为 $O_p(N^{-1/2})$。

项 (I) 是 $f(Z_E)=\mu_t(S,\ x)-\bar{\mu}_t(x)-\nu_t(S,\ x)\omega_t^{\nu}(M,\ x)+\pi_t(S,\ x)\omega_t^{\pi}(M,\ x)$ 的样本均值。根据引理 B.1(3)，其均值为零；根据 $|\mu_t|,\ |\bar{\mu}_t|,\ |\nu_t|\le C,\ \pi_t\le 1$ 和假设 3.4，它是平方可积的。因此，由 (C.5)，项 (I) 等于 (C.3) 中的第一项求和加上 $O_p(N^{-1})$。

项 (II) 是 $g(Z_O)=\mathbf{1}\{T=t\}\{Y\omega_t^{\nu}(M,x)-\omega_t^{\pi}(M,x)\}$ 的样本均值，根据引理 B.1(3)，其均值为零。根据 $|Y|\le C$ 和引理 B.1(1)，$\mathbb{E}[g(Z_O)^2\mid x,O]\le 2C^2\mathbb{E}[\pi_t(S,x)\omega_t^{\nu}(M,x)^2\mid T=t,x,E]+2\mathbb{E}[\pi_t(S,x)\omega_t^{\pi}(M,x)^2\mid T=t,x,E]<\infty$。将 n_E、$\mathbf{1}\{T_i=t,X_i=x\}$ 和 $P(T=t,X=x\mid E)$ 分别替换为 n_O、$\mathbf{1}\{X_j=x\}$ 和 $P(X=x\mid O)$，即可得到 (C.5) 的对应版本；该版本表明，项 (II) 等于 (C.3) 中的第二项求和加上 $O_p(N^{-1})$，且该和为 $O_p(N^{-1/2})$。

项 (III) 的三行和项 (IV) 的两行，在符号意义下正是假设 3.6(1)–(5) 要求为 $o_p(N^{-1/2})$ 的量，因此项 (III) 和项 (IV) 是 $o_p(N^{-1/2})$。

我们对项 (V) 进行界定。在附录 C.1.1 的事件上，$|\widehat{\nu}_t(s,\ x)|\le C$ 对所有 s 都成立，且 $\|\widehat{\omega}_t^{\nu}\|_{2,\ t,\ x}$和 $\|\widehat{\omega}_t^{\pi}\|_{2,\ t,\ x}$ 均为有限值，因此，对于每一组样本值，附录 C.1.2 的引理 C.1 的前提均成立。项 (V) 是将 (3.5) 中的两个样本均值替换为期望后，再减去 $\bar{\mu}_t(x)$ 所得到的量，也就是引理 C.1 的左侧，因此它等于 (C.1) 中三个项之和。对于第一项，我们对比率中的误差进行界定。将 $\widehat{\mu}_t=\widehat{\nu}_t/\widehat{\pi}_t$ 与 $\mu_t=\nu_t/\pi_t$ 之差通分，并利用 $\widehat{\pi}_t\ge a$、$\pi_t\ge a$ 和 $|\nu_t|\le C$，逐点可得；其中省略了参数 $(S,\ x)$：

$$|\widehat{\mu}_t-\mu_t|=\left|\frac{\widehat{\nu}_t-\nu_t}{\widehat{\pi}_t}+\frac{\nu_t(\pi_t-\widehat{\pi}_t)}{\widehat{\pi}_t\pi_t}\right|\le\frac{|\widehat{\nu}_t-\nu_t|}{a}+\frac{C|\widehat{\pi}_t-\pi_t|}{a^2}\tag{C.6}$$

由闵可夫斯基不等式可得 $\|\widehat{\mu}_t-\mu_t\|_{2,\ t,\ x}\le a^{-1}\|\widehat{\nu}_t-\nu_t\|_{2,\ t,\ x}+Ca^{-2}\|\widehat{\pi}_t-\pi_t\|_{2,\ t,\ x}$，因此 $\pi_t\ge a$；再由柯西–施瓦茨不等式可得

$$\begin{aligned}&\left|\mathbb{E}\left[\frac{\{\pi_t(S,x)-\widehat{\pi}_t(S,x)\}\{\widehat{\mu}_t(S,x)-\mu_t(S,x)\}}{\pi_t(S,x)}\,\middle|\,T=t,x,E\right]\right|\\&\le\frac{\|\widehat{\pi}_t-\pi_t\|_{2,t,x}\|\widehat{\mu}_t-\mu_t\|_{2,t,x}}{a}\le C\|\widehat{\pi}_t-\pi_t\|_{2,t,x}\{\|\widehat{\pi}_t-\pi_t\|_{2,t,x}+\|\widehat{\nu}_t-\nu_t\|_{2,t,x}\}\end{aligned}\tag{C.7}$$

根据假设 3.5(2)，其为 $o_p(N^{-1/2})$。对于第二项，在给定 M 的条件下，利用迭代期望和柯西–施瓦茨不等式可得

$$\begin{aligned}&|\mathbb{E}[\{\nu_t(S,x)-\widehat{\nu}_t(S,x)\}\{\widehat{\omega}_t^{\nu}(M,x)-\omega_t^{\nu}(M,x)\}\mid T=t,x,E]|\\&=|\mathbb{E}[\mathbb{E}[\nu_t(S,x)-\widehat{\nu}_t(S,x)\mid M,T=t,x,E]\{\widehat{\omega}_t^{\nu}(M,x)-\omega_t^{\nu}(M,x)\}\mid T=t,x,E]|\\&\le\|\mathbb{E}[\widehat{\nu}_t(S,x)-\nu_t(S,x)\mid M,T=t,x,E]\|_{2,t,x}\|\widehat{\omega}_t^{\nu}-\omega_t^{\nu}\|_{2,t,x}\end{aligned}\tag{C.8}$$

以 S 为条件化对象时，同样可通过 $\|\widehat{\nu}_t-\nu_t\|_{2,t,x}\|\mathbb{E}[\widehat{\omega}_t^{\nu}(M,x)-\omega_t^{\nu}(M,x)\mid S,T=t,x,E]\|_{2,t,x}$ 对其进行界定，因此其绝对值至多为两个乘积中较小者。根据假设 3.5(3)，该较小者为 $o_p(N^{-1/2})$。第三项根据同样的界（将 $\nu_t,\widehat{\nu}_t,\omega_t^{\nu},\widehat{\omega}_t^{\nu}$ 替换为 $\pi_t,\widehat{\pi}_t,\omega_t^{\pi},\widehat{\omega}_t^{\pi}$）以及假设 3.5(4)，为 $o_p(N^{-1/2})$。因此，项 (V) 为 $o_p(N^{-1/2})$。

(C.4) 中的项 (I) 和 (II) 等于 (C.3) 中的两个求和加上 $O_p(N^{-1})$，而项 (III)、(IV) 和 (V) 是 $o_p(N^{-1/2})$，因此 (C.3) 成立。 □

C.1.4 渐近线性表示

按照附录 C.1 中的思路，本部分通过分解误差、界定各项并依次证明依分布收敛，来证明 (3.8) 中的第一个收敛结果。我们在附录 C.1.1 的事件上进行推导。将 $P(X=x\mid O)=P(X=x\mid O)+\{\widehat{P}(X=x\mid O)-P(X=x\mid O)\}$ 和 $\widehat{\bar{\mu}}_1(x)-\widehat{\bar{\mu}}_0(x)=\bar{\mu}_1(x)-\bar{\mu}_0(x)+[\{\widehat{\bar{\mu}}_1(x)-\bar{\mu}_1(x)\}-\{\widehat{\bar{\mu}}_0(x)-\bar{\mu}_0(x)\}]$ 代入 (3.6) 中的 $\widehat{\tau}$，并将 $\tau=\sum_{x\in\mathcal{X}}P(X=x\mid O)\{\bar{\mu}_1(x)-\bar{\mu}_0(x)\}$ 展开，得到

$$\begin{aligned}\widehat{\tau}-\tau=&\underbrace{\sum_{x\in\mathcal{X}}\{\widehat{P}(X=x\mid O)-P(X=x\mid O)\}\{\bar{\mu}_1(x)-\bar{\mu}_0(x)\}}_{\text{(i)}}\\&+\underbrace{\sum_{x\in\mathcal{X}}P(X=x\mid O)[\{\widehat{\bar{\mu}}_1(x)-\bar{\mu}_1(x)\}-\{\widehat{\bar{\mu}}_0(x)-\bar{\mu}_0(x)\}]}_{\text{(ii)}}\\&+\underbrace{\sum_{x\in\mathcal{X}}\{\widehat{P}(X=x\mid O)-P(X=x\mid O)\}[\{\widehat{\bar{\mu}}_1(x)-\bar{\mu}_1(x)\}-\{\widehat{\bar{\mu}}_0(x)-\bar{\mu}_0(x)\}]}_{\text{(iii)}}\end{aligned}\tag{C.9}$$

项 (i) 是层份额的估计误差，项 (ii) 是层均值的估计误差，项 (iii) 则是二者的乘积。

由于 $\widehat{P}(X = x \mid O) = n_O^{-1} \sum_{j=1}^{n_O} \mathbf{1}\{X_j = x\}$，对于定义在 $\mathcal{X}$ 上的每个函数 g，我们都有 $\sum_{x \in \mathcal{X}} \widehat{P}(X = x \mid O) g(x) = n_O^{-1} \sum_{j=1}^{n_O} g(X_j)$。将其应用于 $g = \bar{\mu}_1 - \bar{\mu}_0$ 并减去 $\sum_{x \in \mathcal{X}} P(X = x \mid O)\{\bar{\mu}_1(x) - \bar{\mu}_0(x)\} = \tau$ gives

$$\text{(i)} = \frac{1}{n_O} \sum_{j=1}^{n_O} \{\bar{\mu}_1(X_j) - \bar{\mu}_0(X_j) - \tau\} \tag{C.10}$$

我们在 $t = 1$ 和 $t = 0$处应用引理 C.2，分别乘以 $P(X = x \mid O)$，对 x求和，然后用 $t = 0$ 对应的结果减去 $t = 1$对应的结果。(C.3) 中的第一组求和根据 (3.3) 得到 $n_E^{-1} \sum_{i=1}^{n_E} \psi_E(Z_{E,i})$。在第二组求和中，$\mathbf{1}\{X_j = x\}$ 只留下含有 $x = X_j$ 的项；结合 (C.10)，根据 (3.4) 可得 $n_O^{-1} \sum_{j=1}^{n_O} \psi_O(Z_{O,j})$。引理 C.2 中的余项在乘以 $P(X = x \mid O) \leq 1$ 并对有限项求和后仍为 $o_p(N^{-1/2})$，因此

$$\text{(i)} + \text{(ii)} = \frac{1}{n_E} \sum_{i=1}^{n_E} \psi_E(Z_{E,i}) + \frac{1}{n_O} \sum_{j=1}^{n_O} \psi_O(Z_{O,j}) + o_p(N^{-1/2}) \tag{C.11}$$

差值 $\widehat{P}(X = x \mid O) - P(X = x \mid O)$ 是独立同分布的项 $\mathbf{1}\{X_j = x\} - P(X = x \mid O)$ 的样本均值，这些项的均值为零，且其方差至多为 $1/(4n_O)$。假设 3.1 将 n_O 的下界限定为正数常数乘以 N，因此根据切比雪夫不等式，该差值为 $O_p(N^{-1/2})$。根据引理 C.2，$\widehat{\bar{\mu}}_t(x) - \bar{\mu}_t(x) = O_p(N^{-1/2})$，并且由于 $\mathcal{X}$ 有限，项 (iii) 为 $O_p(N^{-1})$。

方程 (C.9) 和 (C.11) 以及对项 (iii) 的界给出

$$\begin{aligned} \sqrt{N}(\widehat{\tau} - \tau) = &\sqrt{\frac{N}{n_E}} \frac{1}{\sqrt{n_E}} \sum_{i=1}^{n_E} \psi_E(Z_{E,i}) \\ &+ \sqrt{\frac{N}{n_O}} \frac{1}{\sqrt{n_O}} \sum_{j=1}^{n_O} \psi_O(Z_{O,j}) + o_p(1) \end{aligned} \tag{C.12}$$

根据引理 B.2，$\psi_E(Z_{E,\,i})$独立同分布，均值为零且方差有限，$\psi_O(Z_{O,\,j})$也如此。两个样本相互独立，因此，分别对每个样本应用中心极限定理、假设 3.1 的 $N/n_E \to 1/\rho$ 和 $N/n_O \to 1/(1-\rho)$，以及 Slutsky 引理，可知 (C.12) 右侧依分布收敛到均值为零、方差为 $\rho^{-1}\mathbb{E}[\psi_E(Z_E)^2 \mid E] + (1-\rho)^{-1}\mathbb{E}[\psi_O(Z_O)^2 \mid O]$ 的正态分布；根据 (3.7)，该方差为Σ^2。这证明了 (3.8) 中的第一个收敛。特别地，$\widehat{\tau} - \tau = O_p(N^{-1/2})$，这是一个无需借助引理 C.3 得到的界。

C.1.5 方差估计量的一致性

引理 C.3 (方差估计量的一致性)。 假设假设 *2.1–2.8*、附录 *A* 中的可积性条件、假设 *3.1–3.6*、假设 *B.1* 以及假设 *C.1(1)* 均成立。那么

$$\begin{aligned}\frac{1}{n_E}\sum_{i=1}^{n_E}\{\widehat{\psi}_E(Z_{E,i})-\psi_E(Z_{E,i})\}^2 &\to_p 0,\\ \frac{1}{n_O}\sum_{j=1}^{n_O}\{\widehat{\psi}_O(Z_{O,j})-\psi_O(Z_{O,j})\}^2 &\to_p 0\end{aligned} \tag{C.13}$$

以及 $\widehat{\Sigma}^2 \to_p \Sigma^2$。

引理 C.3 的证明。 目标是证明 (C.13) 和 $\widehat{\Sigma}^2 \to_p \Sigma^2$。我们首先证明 (C.13) 蕴含 $\widehat{\Sigma}^2$ 的一致性，然后分别针对实验样本和观察样本证明 (C.13)。对于两个样本，我们利用假设 3.6(6)–(11) 给出的界来限制估计误差平方的样本均值。

均方根的三角不等式给出

$$\left|\left\{\frac{1}{n_E}\sum_{i=1}^{n_E}\widehat{\psi}_E(Z_{E,i})^2\right\}^{1/2}-\left\{\frac{1}{n_E}\sum_{i=1}^{n_E}\psi_E(Z_{E,i})^2\right\}^{1/2}\right| \le \left\{\frac{1}{n_E}\sum_{i=1}^{n_E}\{\widehat{\psi}_E(Z_{E,i})-\psi_E(Z_{E,i})\}^2\right\}^{1/2} \tag{C.14}$$

根据引理 B.2 和大数定律，$n_E^{-1}\sum_{i=1}^{n_E}\psi_E(Z_{E,\,i})^2 \to_p \mathbb{E}[\psi_E(Z_E)^2 \mid E]$，因此由 (C.14)、(C.13) 的第一个方程以及连续映射定理可得$n_E^{-1}\sum_{i=1}^{n_E}\widehat{\psi}_E(Z_{E,\,i})^2 \to_p \mathbb{E}[\psi_E(Z_E)^2 \mid E]$。对 (C.13) 的第二个方程进行同样的论证，可得 $n_O^{-1}\sum_{j=1}^{n_O}\widehat{\psi}_O(Z_{O,\,j})^2 \to_p \mathbb{E}[\psi_O(Z_O)^2 \mid O]$；再结合假设 3.1 的 $N/n_E \to 1/\rho$ 和 $N/n_O \to 1/(1-\rho)$ 以及 (3.7)，即可得 $\widehat{\Sigma}^2 \to_p \Sigma^2$。

我们在附录 C.1.1 的事件上，证明实验样本的方程 (C.13) 的第一个等式。对于一条观测 $(T_i, X_i) = (t, x)$，(3.3) 及其

$\widehat{\ }$ $\widehat{\ }$插件版本中只剩下一项。将 $P(X = x \mid O)/P(T = t, X = x \mid E)$ 乘以 (3.3) 中大括号内的表达式，再将其加上并减去，可知若 $t = 1$，则 $\psi_E(Z_{E,i}) - \psi_E(Z_{E,i})$ 等于

$$\begin{aligned}
&\underbrace{\frac{\widehat{P}(X = x \mid O)}{\widehat{P}(T = t, X = x \mid E)}[\{\widehat{\mu}_t - \mu_t\} - \{\widehat{\nu}_t\widehat{\omega}_t^{\nu} - \nu_t\omega_t^{\nu}\} + \{\widehat{\pi}_t\widehat{\omega}_t^{\pi} - \pi_t\omega_t^{\pi}\}]}_{\text{(I)}} \\
&\quad \underbrace{- \frac{\widehat{P}(X = x \mid O)}{\widehat{P}(T = t, X = x \mid E)}\{\widehat{\bar{\mu}}_t - \bar{\mu}_t\}}_{\text{(II)}} \\
&\quad \underbrace{+ \left\{\frac{\widehat{P}(X = x \mid O)}{\widehat{P}(T = t, X = x \mid E)} - \frac{P(X = x \mid O)}{P(T = t, X = x \mid E)}\right\}\{\mu_t - \bar{\mu}_t - \nu_t\omega_t^{\nu} + \pi_t\omega_t^{\pi}\}}_{\text{(III)}}
\end{aligned} \tag{C.15}$$

并且若 $t = 0$，则它等于符号相反的同一表达式。这里，$\widehat{\ }$函数的自变量为 $(S_i,\ x)$ 或 $(M_i,\ x)$；我们将它们连同 $\bar{\mu}_t$，$\bar{\mu}_t$ 的自变量 x 一并省略。(I) −(II) + (III) 的平方至多为 $3\{(\text{I})^2 + (\text{II})^2 + (\text{III})^2\}$。在附录 C.1.1 的事件上，$\widehat{P}(T = t,\ X = x \mid E) \geq c/2$ 且 $\widehat{P}(X = x \mid O) \leq 1$，因此该比率的平方至多为 $4c^{-2}$。

对于项 (I)，我们在 (S_i, M_i) 处计算方括号内的函数，将其平方，乘以 $\mathbf{1}\{T_i = t, X_i = x\}\widehat{\ }$，对 i 求和，再除以 n_E。由 $P(T = t, X = x \mid E) \leq 1$，结果至多为该函数平方的 $\mathbb{E}[\cdot \mid T = t, x, E]\ \widehat{\ }$。方括号内的表达式是三个差值之和，因此其平方至多为这三个差值平方和的三倍；所以只需证明每个平方差的 $\mathbb{E}[\cdot \mid T = t, x, E]$ 依概率趋于零。首先，由 (C.6)，并且由于两个实数之和的平方至多为它们平方和的两倍，在每个观测值处都有 $\{\widehat{\mu}_t - \mu_t\}^2 \leq 2a^{-2}\{\widehat{\nu}_t - \nu_t\}^2 + 2C^2a^{-4}\{\widehat{\pi}_t - \pi_t\}^2\ \widehat{\ }$，而根据假设 3.6(6)(7)，右侧的 $\mathbb{E}[\cdot \mid T = t, x, E]$ 依概率趋于零。其次，$\widehat{\nu}_t\widehat{\omega}_t^{\nu} - \nu_t\omega_t^{\nu} = \widehat{\nu}_t(\widehat{\omega}_t^{\nu} - \omega_t^{\nu}) + (\widehat{\nu}_t - \nu_t)\omega_t^{\nu}$ 和 $|\widehat{\nu}_t| \leq C$ 给出在每个观测值处的 $\{\widehat{\nu}_t\widehat{\omega}_t^{\nu} - \nu_t\omega_t^{\nu}\}^2 \leq 2C^2(\widehat{\omega}_t^{\nu} - \omega_t^{\nu})^2 + 2(\widehat{\nu}_t - \nu_t)^2(\omega_t^{\nu})^2$。对于正数 r，$|\widehat{\nu}_t - \nu_t| \leq 2C$ 在每个观测值处给出，

$$(\widehat{\nu}_t - \nu_t)^2(\omega_t^{\nu})^2 \leq r^2(\widehat{\nu}_t - \nu_t)^2 + 4C^2(\omega_t^{\nu})^2\mathbf{1}\{|\omega_t^{\nu}| > r\} \tag{C.16}$$

其结论源于：当 $|\omega_t^{\nu}| \leq r$ 时，用 r^2 对左侧的 $(\omega_t^{\nu})^2$ 进行界定；当 $|\omega_t^{\nu}| > r$ 时，用 $4C^2$ 对 $(\widehat{\nu}_t - \nu_t)^2$ 进行界定。对于任意 $\varepsilon > 0$，根据假设 3.4 和控制收敛定理，我们可以选择 r，使得 $8C^2\mathbb{E}[\omega_t^{\nu}(M, x)^2\mathbf{1}\{|\omega_t^{\nu}(M, x)| > r\} \mid T = t, x, E] < \varepsilon$。对于这个$r$，根据大数定律，$8C^2\widehat{\mathbb{E}}[\omega_t^{\nu}(M, x)^2\mathbf{1}\{|\omega_t^{\nu}(M, x)| > r\} \mid T = t, x, E] < \varepsilon$ 以趋近于 1 的概率成立；并且根据假设3.6(7)(8)，$2C^2\widehat{\mathbb{E}}[\{\widehat{\omega}_t^{\nu}(M, x) - \omega_t^{\nu}(M, x)\}^2 \mid T =$

$t, x, E] + 2r^2\widehat{\mathbb{E}}[\{\widehat{\nu}_t(S,x) - \nu_t(S,x)\}^2 \mid T = t, x, E] < \varepsilon$ 以趋近于 1 的概率成立。因此$\widehat{\mathbb{E}}[\{\widehat{\nu}_t(S,x)\widehat{\omega}_t^{\nu}(M,x) - \nu_t(S,x)\omega_t^{\nu}(M,x)\}^2 \mid T = t, x, E] < 2\varepsilon$ 以趋近于 1 的概率成立。第三，我们写作 $\widehat{\pi}_t\widehat{\omega}_t^{\pi} - \pi_t\omega_t^{\pi} = \widehat{\pi}_t(\widehat{\omega}_t^{\pi} - \omega_t^{\pi}) + (\widehat{\pi}_t - \pi_t)\omega_t^{\pi}$，并使用相同的论证，其中$\widehat{\pi}_t \le 1$、$|\widehat{\pi}_t - \pi_t| \le 1$，以及假设 3.6(6)(9)。

将项 (II) 的平方乘以 $\mathbf{1}\{T_i = t, X_i = x\}$，对 i 求和，再除以 n_E，所得量至多为 $4c^{-2}\{\widehat{\bar{\mu}}_t(x) - \bar{\mu}_t(x)\}^2$,，根据引理 C.2，该量依概率趋于零。对于项 (III)，大数定律和假设 3.2(2)表明，比率之差依概率趋于零；而 $n_E^{-1}\sum_{i=1}^{n_E}\mathbf{1}\{T_i = t, X_i = x\}\{\mu_t - \bar{\mu}_t - \nu_t\omega_t^{\nu} + \pi_t\omega_t^{\pi}\}^2$依概率收敛于一个有限极限，因此项 (III) 的平方的平均值依概率趋于零。对有限多个 (t, x) 求和，即得到方程 (C.13) 的第一式。

我们给出观察样本的（C.13）中的第二个方程。对于观测值 j，（3.4）及其插件版本给出

$$\begin{aligned}\widehat{\psi}_O(Z_{O,j}) - \psi_O(Z_{O,j}) = {} & \underbrace{\{\widehat{\bar{\mu}}_1(X_j) - \bar{\mu}_1(X_j)\} - \{\widehat{\bar{\mu}}_0(X_j) - \bar{\mu}_0(X_j)\} - (\widehat{\tau} - \tau)}_{\text{(IV)}} \\ & + \underbrace{\mathbf{1}\{T_j = 1\}[Y_j\{\widehat{\omega}_1^{\nu} - \omega_1^{\nu}\}(M_j, X_j) - \{\widehat{\omega}_1^{\pi} - \omega_1^{\pi}\}(M_j, X_j)]}_{\text{(V)}} \\ & - \underbrace{\mathbf{1}\{T_j = 0\}[Y_j\{\widehat{\omega}_0^{\nu} - \omega_0^{\nu}\}(M_j, X_j) - \{\widehat{\omega}_0^{\pi} - \omega_0^{\pi}\}(M_j, X_j)]}_{\text{(VI)}}\end{aligned} \tag{C.17}$$

这里，$\{\widehat{\omega}_t^{\nu} - \omega_t^{\nu}\}(M_j, X_j)$ 表示 $\widehat{\omega}_t^{\nu}(M_j, X_j) - \omega_t^{\nu}(M_j, X_j)$。由于 $\mathcal{X}$ 是有限的，引理 C.2 以及附录 C.1.4 中在不使用引理 C.3 的情况下建立的界 $\widehat{\tau} - \tau = O_p(N^{-1/2})$ 表明，项 (IV) 关于 j 一致地依概率收敛于零。我们将项 (V) 和 (VI) 的平方乘以 $\mathbf{1}\{X_j = x\}$，对 j 求和，再除以 n_O。根据 $|Y_j| \le C$（由假设 B.1，对于所有 j 以概率 1 成立），并且由于两个实数之和的平方至多是其平方和的两倍，该结果至多为 $2C^2\widehat{P}(X = x \mid O)\widehat{\mathbb{E}}[\mathbf{1}\{T = t\}\{\widehat{\omega}_t^{\nu}(M,x) - \omega_t^{\nu}(M,x)\}^2 \mid x, O] + 2\widehat{P}(X = x \mid O)\widehat{\mathbb{E}}[\mathbf{1}\{T = t\}\{\widehat{\omega}_t^{\pi}(M,x) - \omega_t^{\pi}(M,x)\}^2 \mid x, O]$，其中项 (V) 对应 $t = 1$，项 (VI) 对应 $t = 0$；根据假设 3.6(10)(11)，其依概率收敛于零。(C.17) 左侧等于 (IV) +(V) −(VI)，而三个实数之和的平方至多是其平方和的三倍，因此其平方至多为 $3\{(\text{IV})^2 + (\text{V})^2 + (\text{VI})^2\}$。将其乘以 $\mathbf{1}\{X_j = x\}$，对 j 求和，除以 n_O，再对有限多个 x 求和，即得到 (C.13) 的第二个方程。 □

C.1.6 学生化与效率

(3.8) 中的第二个收敛是引理 C.3。设 $\Sigma^2 > 0$。则根据连续映射定理，$\widehat{\Sigma} \to_p \Sigma > 0$，因此 $P(\widehat{\Sigma} > 0) \to 1$；在 $\{\widehat{\Sigma} > 0\}$ 上，比率 $\sqrt{N}(\widehat{\tau} - \tau)/\widehat{\Sigma}$ 等于 $\{\sqrt{N}(\widehat{\tau} - \tau)/\Sigma\}(\Sigma/\widehat{\Sigma})$，根据 Slutsky 引理，其依分布收敛于 $\mathcal{N}(0, 1)$ 的正态分布。在同一事件上，当且仅当 $|\sqrt{N}(\widehat{\tau} - \tau)/\widehat{\Sigma}| \le z_{1-\alpha/2}$ 时，该区间包含 τ；因此，由于标准正态分布的分布函数是连续的，覆盖概率收敛于 $1 - \alpha$。 □

本部分的其余内容将证明定理 3.2 和 3.3 的效率断言。在真实分布 P 下，两个估计量具有相同的线性展开（C.12）：定理 3.2 的估计量的线性展开见附录 C.1.4，定理 3.3 的估计量的线性展开见附录 C.2.4。以下论证仅使用（C.12）、引理 B.2 和 B.3，以及（B.3）；引理 B.4 表明，（B.3）沿每个光滑子模型均成立，因此该论证适用于两个估计量。

取任意一个得分为 (a_E, a_O) 的光滑子模型。我们求（C.12）右侧两个和以及（B.12）中数量的联合极限。根据引理 B.2 和得分的定义，$(\psi_E(Z_{E,i}), a_E(Z_{E,i}))$ 是均值为零且具有有限二阶矩的独立同分布二维向量，$(\psi_O(Z_{O,j}), a_O(Z_{O,j}))$ 也同样如此。根据多元中心极限定理，$n_E^{-1/2} \sum_i (\psi_E(Z_{E,i}), a_E(Z_{E,i}))$ 依分布收敛于一个正态分布，其协方差矩阵的元素为 $\mathbb{E}[\psi_E^2 \mid E]$、$\mathbb{E}[\psi_E a_E \mid E]$ 和 $\mathbb{E}[a_E^2 \mid E]$；观察组求和也以相同方式依分布收敛。两个样本相互独立，且假设 3.1 给出 $N/n_E \to 1/\rho$ 和 $N/n_O \to 1/(1-\rho)$，因此根据 Slutsky 引理，（C.12）右侧两项之和与（B.12）中的数量联合依分布收敛于二维正态分布。对于（C.12）右侧各项之和，其极限方差为 $\rho^{-1}\mathbb{E}[\psi_E^2 \mid E] + (1-\rho)^{-1}\mathbb{E}[\psi_O^2 \mid O] = \Sigma^2$；对于（B.12）中的数量，其极限方差为 $\rho\mathbb{E}[a_E^2 \mid E] + (1-\rho)\mathbb{E}[a_O^2 \mid O] = \|(a_E, a_O)\|_\rho^2$。实验组求和的系数收敛于 $\sqrt{N/n_E} \to 1/\sqrt{\rho}$ 和 $\sqrt{n_E/N} \to \sqrt{\rho}$，观察组求和的系数收敛于 $1/\sqrt{1-\rho}$ 和 $\sqrt{1-\rho}$，因此极限协方差为

$$\frac{1}{\sqrt{\rho}}\sqrt{\rho}\,\mathbb{E}[\psi_E a_E \mid E] + \frac{1}{\sqrt{1-\rho}}\sqrt{1-\rho}\,\mathbb{E}[\psi_O a_O \mid O] = \mathbb{E}[\psi_E a_E \mid E] + \mathbb{E}[\psi_O a_O \mid O] \tag{C.18}$$

根据 Slutsky 引理，(C.12) 和 (B.4) 中的 $o_p(1)$ 项不会改变极限。因此，$\sqrt{N}(\widehat{\tau} - \tau)$ 与 (B.4) 左侧的对数似然比联合依分布收敛至二维正态分布，其均值为 $(0,\ -\|(a_E,\ a_O)\|_\rho^2/2)$，方差为上述方差，协方差为 (C.18)。

Le Cam 第三引理（van der Vaart 1998）随后表明，在 $P_{1/\sqrt{N}}$，$\sqrt{N}\,(\widehat{\tau}-\overline{\tau})$ 下，依分布收敛至均值等于协方差 (C.18)、方差为Σ^2的正态分布。另一方面，将引理 B.4 的 (B.3) 应用于 $\varepsilon = 1/\sqrt{N}$ 可得

$$\sqrt{N}\{\tau(P_{1/\sqrt{N}}) - \tau(P)\} \to \mathbb{E}[\psi_E a_E \mid E] + \mathbb{E}[\psi_O a_O \mid O] \tag{C.19}$$

我们将 $P_{1/\sqrt{N}}$ 下的估计误差分解为

$$\sqrt{N}\{\widehat{\tau} - \tau(P_{1/\sqrt{N}})\} = \sqrt{N}\{\widehat{\tau} - \tau(P)\} - \sqrt{N}\{\tau(P_{1/\sqrt{N}}) - \tau(P)\}$$

右侧第一项依分布收敛于上述正态分布，第二项收敛于（C.19）中的常数，因此根据 Slutsky 引理，左侧在 $P_{1/\sqrt{N}}$ 下依分布收敛于均值为零、方差为 Σ^2的正态分布。该极限不依赖于子模型，因此 $\widehat{\tau}$ 在模型 $\mathcal{P}$ 中是正则的，特别地，相对于假设 B.2 中的子空间也是正则的。这个均值为零、方差为 Σ^2 的正态极限，是附录B.2.2中论断（4）的卷积，其中另一个概率分布是零点处的点质量，并且其方差等于定理 3.1 的效率界 Σ^2。因此 $\widehat{\tau}$ 达到了效率界。 □

C.2 引理与定理3.3的证明

本小节证明第3.4节中的定理 3.3。我们首先回顾每一折的分层均值。

定理 3.3 的估计量对每一折 k 使用主文中公式（3.9）给出的层均值，

$$\begin{aligned}\widehat{\mu}_{t,k}(x) = \widehat{\mathbb{E}}_k[&\widehat{\mu}_t^{(-k)}(S,x) - \widehat{\nu}_t^{(-k)}(S,x)\widehat{\omega}_t^{\nu,(-k)}(M,x) + \widehat{\pi}_t^{(-k)}(S,x)\widehat{\omega}_t^{\pi,(-k)}(M,x) \mid T=t,x,E] \\ &+ \widehat{\mathbb{E}}_k[\mathbf{1}\{T=t\}\{Y\widehat{\omega}_t^{\nu,(-k)}(M,x) - \widehat{\omega}_t^{\pi,(-k)}(M,x)\} \mid x,O]\end{aligned}$$

除效率之外各项断言的证明遵循附录 C.1 中定理 3.2 的证明。附录 C.1.4 以及附录 C.1.6 前半部分的证明，在附录 C.1.1 的事件上仅使用引理 B.2、引理 C.2 的 (C.3) 以及引理 C.3 的 (C.13)。如果能够针对交叉拟合的 $\widehat{\mu}_t(x)$和 $\widehat{\Sigma}^2$证明 (C.3) 和 (C.13)，那么附录 C.1.4 以及附录 C.1.6 前半部分的论证便可不加修改地得到定理 3.3。我们通过逐折执行附录 C.1.3 中引理 C.2 的证明来证明 (C.3)，并在附录 C.1.5 中引理 C.3 的证明里用各折的量替换全样本量，从而证明 (C.13)。附录 C.2.1 给出了本证明中所使用的概率趋于 1 的事件。为了界定由代入估计量和方差估计量所导致的各项，

我们使用引理 C.4 替代假设 3.6，而引理 C.5 随后将各折上的平均值转换为全样本上的求和。

C.2.1 概率趋于 *1* 的事件

我们定义定理 3.3 的证明中所用的概率趋于 1 的事件。在附录 C.1.1 的 $\widehat{\ }$(E1) 之外，即对所有 t, x 有 $P(T = t, X = x \mid E) \geq c/2$ 且 $P(X = x \mid O) \geq c/2$，我们对所有 t, x, k 再加入以下三个条件。(E4) $P_k(T = t, X = x \mid E) \geq c/2\widehat{\ }$且 $P_k(X = x \mid O) \geq c/2$。(E5) 假设 C.1(2) 的事件发生，即一个概率趋于 1 的事件，在该事件上对所有 s 有 $|\widehat{v}_t^{(-k)}(s, x)| \leq C$，其中 C 是假设 C.1(2) 的界中的常数。(E6) $\|\widehat{\omega}^{\nu}{}_{,t}^{(-k)}\|_{2,t,x}$ 与 $\|\widehat{\omega}^{\pi}{}_{,t}^{(-k)}\|_{2,t,x}$ 均为有限。我们将 (E1)、(E4)、(E5) 与 (E6) 同时成立的事件称为附录 C.2.1 的事件。由附录 C.1.1，(E1) 的概率趋于 1。对于 (E4)，各折的样本量至多相差 1，因此$|I_{E,k}| \geq n_E/K - 1$，且假设 3.1 将 $|I_{E,k}|$ 从下方界定为正数常数乘以 $N.\widehat{\ }$由于 $P_k(T = t, X = x \mid E)$ 是均值为 $P(T = t, X = x \mid E) \geq c$、方差至$\widehat{\ }$多为 $1/(4|I_{E,k}|)$ 的样本均值，由切比雪夫不等式，$P_k(T = t, X = x \mid E) < c/2$ 的概率趋于$\widehat{\ }$零。对 $P_k(X = x \mid O)$ 同样成立。由假设 C.1(2)，(E5) 的概率趋于 1。对于 (E6)，闵可夫斯基不等式给出 $\|\widehat{\omega}^{\nu}{}_{,t}^{(-k)}\|_{2,t,x} \leq \|\omega_t^{\nu}\|_{2,t,x} + \|\widehat{\omega}^{\nu}{}_{,t}^{(-k)} - \omega_t^{\nu}\|_{2,t,x}$，其中右端第一项由假设 3.4 为有限，第二项由假设 3.7 依概率趋于零，因此该范数为有限的概率趋于 1。对 $\widehat{\omega}^{\pi}{}_{,t}^{(-k)}$ 同样成立。由于有限多个事件的概率均趋于 1，它们同时成立的概率也趋于 1。在该事件上，(3.9)、(3.6) 以及方差估计量所取平均的所有组均为非空。由于补集的概率趋于零，在该事件上建立的依概率收敛与依分布收敛在不限于该事件时仍然成立。下文中，在附录 C.1.4 的证明以及附录 C.1.6 前半部分的证明中，我们也用该事件代替附录 C.1.1 的事件；这些证明仅使用附录 C.1.1 的事件的性质 (E1)。

C.2.2 折叠平均的条件界

引理 C.4 (折叠平均)。 假设假设 *2.1–2.8*、附录 *A* 中的可积性条件、假设 *3.1–3.4*、假设 *3.7*、假设 *B.1* 以及假设 *C.1(2)*均成立。则$\widehat{\ }$对于每个 t、x、k，以下结论均成立，其中 $(\mathbb{E}_k - \mathbb{E})[\cdot \mid T = t、\ x、\ E] = \mathbb{E}_k[\cdot \mid T = t、\ x、\ E] - \mathbb{E}[\cdot \mid T = t、\ x、\ E]\widehat{\ }$，且 $(\mathbb{E}_k - \mathbb{E})[\cdot \mid x、\ O]$ 在观察样本中以相同方式定义。

1. $(\widehat{\mathbb{E}}_k - \mathbb{E})[\widehat{\mu}_t^{(-k)}(S,\ x) - \mu_t(S,\ x) \mid T = t,\ x,\ E]$ 为 $o_p(N^{-1/2})$。

2. $(\widehat{\mathbb{E}}_k - \mathbb{E})[\widehat{v}_t^{(-k)}(S,\ x)\widehat{\omega}^{\nu}{}_{,t}^{(-k)}(M,\ x) - v_t(S,\ x)\omega_t^{\nu}(M,\ x) \mid T = t,\ x,\ E]$ 是$o_p(N^{-1/2})$。

3. $(\widehat{\mathbb{E}}_k - \mathbb{E})[\widehat{\pi}_t^{(-k)}(S,\ x)\widehat{\omega}^{\pi,\ (-k)}_{t}(M,\ x) - \pi_t(S,\ x)\omega_t^{\pi}(M,\ x) \mid T = t,\ x,\ E]$ 是 $o_p(N^{-1/2})$.

4. $(\widehat{\mathbb{E}}_k - \mathbb{E})[\mathbf{1}\{T = t\}\{\widehat{\omega}^{\nu,\ (-k)}_{t}(M, x) - \omega_t^{\nu}(M, x)\} \mid x, O]$ 为 $o_p(N^{-1/2})$。5. $(\widehat{\mathbb{E}}_k - \mathbb{E})[\mathbf{1}\{T = t\}\{\widehat{\omega}^{\pi,\ (-k)}_{t}(M, x) - \omega_t^{\pi}(M, x)\} \mid x, O]$ 为 $o_p(N^{-1/2})$。6. $\widehat{\mathbb{E}}_k[\{\widehat{\pi}_t^{(-k)}(S, x) - \pi_t(S, x)\}^2 \mid T = t, x, E]$ 依概率趋于零。7. $\widehat{\mathbb{E}}_k[\{\widehat{\nu}_t^{(-k)}(S, x) - \nu_t(S, x)\}^2 \mid T = t, x, E]$ 依概率趋于零。8. $\widehat{\mathbb{E}}_k[\{\widehat{\omega}^{\nu,\ (-k)}_{t}(M, x) - \omega_t^{\nu}(M, x)\}^2 \mid T = t, x, E]$ 依概率趋于零。9. $\widehat{\mathbb{E}}_k[\{\widehat{\omega}^{\pi,\ (-k)}_{t}(M, x) - \omega_t^{\pi}(M, x)\}^2 \mid T = t, x, E]$ 依概率趋于零。10. $\widehat{\mathbb{E}}_k[\mathbf{1}\{T = t\}\{\widehat{\omega}^{\nu,\ (-k)}_{t}(M, x) - \omega_t^{\nu}(M, x)\}^2 \mid x, O]$ 依概率趋于零。11. $\widehat{\mathbb{E}}_k[\mathbf{1}\{T = t\}\{\widehat{\omega}^{\pi,\ (-k)}_{t}(M, x) - \omega_t^{\pi}(M, x)\}^2 \mid x, O]$ 依概率趋于零。

引理 C.4 的证明。固定 $t,\ x,\ k$，并令 $\mathcal{G}_k$ 表示由折 k 之外的观测值生成的 σ-域。留出折 k 的估计量关于 $\mathcal{G}_k$ 可测。根据假设 3.1，折 k 中的观测值与 $\mathcal{G}_k$ 独立，且服从原始分布，因此，给定 $\mathcal{G}_k$，折 k 中某个观测值及这些估计量的函数的条件期望，等于第 3.3 节中定义的、将估计量固定后的期望。附录 C.2.1 中的事件涉及折 k，未必关于 $\mathcal{G}_k$ 可测，因此我们只对不受该事件限制的量取给定 $\mathcal{G}_k$ 的条件期望，并且仅在条件期望之外使用 (E4) 中的份额界和 (E5) 中的界。

对于断言 (1)–(5)，令 f_k 为实验观测及留出折 k 的估计量的函数，其均方 $\mathbb{E}[f_k(Z_E)^2 \mid T = t,\ x,\ E]$ 是在将估计量固定后计算的，并依概率趋于零。我们证明 $(\widehat{\mathbb{E}}_k - \mathbb{E})[f_k(Z_E) \mid T = t,\ x,\ E] = o_p(N^{-1/2})$。根据样本均值的定义，

$$
\begin{aligned}
&(\widehat{\mathbb{E}}_k - \mathbb{E})[f_k(Z_E) \mid T = t, x, E] \\
&= \frac{1}{\widehat{P}_k(T = t, X = x \mid E)} \cdot \frac{1}{|\mathcal{I}_{E,k}|} \sum_{i \in \mathcal{I}_{E,k}} \mathbf{1}\{T_i = t, X_i = x\}\{f_k(Z_{E,i}) - \mathbb{E}[f_k(Z_E) \mid T = t, x, E]\}
\end{aligned}
\tag{C.20}
$$

均方仅由估计量决定，因此其有限这一事件相对于 $\mathcal{G}_k$ 可测，且其概率趋近于一。在该事件上，给定 $\mathcal{G}_k$，(C.20) 中的加项相互独立、均值为零，且二阶矩为 $P(T = t, X = x \mid E)\ \mathrm{Var}(f_k(Z_E) \mid T = t, x, E) \le \mathbb{E}[f_k(Z_E)^2 \mid T = t, x, E]$，因此给定 $\mathcal{G}_k$ 的条件期望，即 N 乘以它们在 $\mathrm{I}_{E,k}$ 上平均值的平方，不超过 $(N/|\mathrm{I}_{E,k}|)\mathbb{E}[f_k(Z_E)^2 \mid T = t, x, E]$。根据假设 3.1，$N/|\mathrm{I}_{E,k}| \to K/\rho$，因此该上界依概率趋于零。一个非负

随机变量 U 若给定 $\mathcal{G}_k$ 的条件期望依概率趋于零，则其本身也依概率趋于零，因为对于 $\varepsilon > 0$ 和 $\delta \in (0,1)$，条件马尔可夫不等式给出 $P(U > \varepsilon) \le \mathbb{E}[\min\{1, \mathbb{E}[U \mid \mathcal{G}_k]/\varepsilon\}] \le \delta + P(\mathbb{E}[U \mid \mathcal{G}_k] > \delta\varepsilon)$。将其应用于 N 乘以该平方平均值再乘以事件的指示函数，可知 $\sqrt{N}$ 乘以该$\widehat{\ }$平均值依概率趋于零。在 (E4) 中将 $P_k(T = t, X = x \mid E) \ge c/2$ 移至条件期望之外，(C.20) 得到 $(\widehat{\mathbb{E}}_k - \mathbb{E})[f_k(Z_E) \mid T = t, x, E] = o_p(N^{-1/2})$。对于观察样本中某个观测的函数 g_k，若 $\mathbb{E}[g_k(Z_O)^2 \mid x, O]$ 依概率趋于零，则同样的论证给出 $(\widehat{\mathbb{E}}_k - \mathbb{E})[g_k(Z_O) \mid x, O] = o_p(N^{-1/2})$。

对于断言 (1)–(3)，根据附录 C.1.3 和 C.1.5 的界，并使用排除第 k 折后的估计量，三个函数的均方依概率趋于零。在断言 (1) 中，由 (C.6) 和闵可夫斯基不等式可得

$$\|\widehat{\mu}_t^{(-k)} - \mu_t\|_{2,t,x} \le a^{-1}\|\widehat{\nu}_t^{(-k)} - \nu_t\|_{2,t,x} + Ca^{-2}\|\widehat{\pi}_t^{(-k)} - \pi_t\|_{2,t,x}$$

右侧根据假设 3.7 依概率趋于零。在断言 (2) 中，我们写成

$$\widehat{\nu}_t^{(-k)}\widehat{\omega}_t^{\nu,(-k)} - \nu_t\omega_t^{\nu} = \widehat{\nu}_t^{(-k)}(\widehat{\omega}_t^{\nu,(-k)} - \omega_t^{\nu}) + (\widehat{\nu}_t^{(-k)} - \nu_t)\omega_t^{\nu}$$

在 (E5) 中的事件上，第一部分的范数至多为 $C\|\widehat{\omega}^{\nu,\,(-k)}_{t,\,t,\,x} - \omega_t^{\nu}\|_{2,\,t,\,x}$；在 (C.16) 中用 $\widehat{\nu}_t^{(-k)}$ 替代 $\widehat{\nu}_t$ 并取期望，可得对每个 $r > 0$,

$$\|(\widehat{\nu}_t^{(-k)} - \nu_t)\omega_t^{\nu}\|_{2,t,x}^2 \le r^2\|\widehat{\nu}_t^{(-k)} - \nu_t\|_{2,t,x}^2 + 4C^2\mathbb{E}[\omega_t^{\nu}(M,x)^2\mathbf{1}\{|\omega_t^{\nu}(M,x)| > r\} \mid T = t, x, E]$$

对于任意 $\varepsilon > 0$，根据假设 3.4 和控制收敛定理，我们可以选择 r，使得第二项小于 $\varepsilon/2$；对于该 r，根据假设 3.7，第一项以趋近于一的概率小于 $\varepsilon/2$。由于 (E5) 的概率趋近于一，断言 (2) 中函数的均方依概率趋于零。断言 (3) 以同样方式成立，其中使用 $\widehat{\pi}_t^{(-k)} \le 1$ 和 $|\widehat{\pi}_t^{(-k)} - \pi_t| \le 1$。将上述结论应用于这三个函数，即可得到断言 (1)–(3)。

对于断言 (4) 和 (5)，$|Y| \le C$、引理 B.1(1) 和 $\pi_t \le 1$ 分别将两个观测函数的均方限定为 $C^2\|\widehat{\omega}^{\nu,\,(-k)}_{t,\,t,\,x} - \omega_t^{\nu}\|_{2,\,t,\,x}^2$ 和 $\|\widehat{\omega}^{\pi,\,(-k)}_{t,\,t,\,x} - \omega_t^{\pi}\|_{2,\,t,\,x}^2$；根据假设 3.7，这些量依概率趋于零，因此上述结论的观测版本给出了断言 (4) 和 (5)。

我们证明断言 (6)–(9)。对于断言 (6)，令 $f_k(Z_E) = \{\widehat{\pi}_t^{(-k)}(S,x) - \pi_t(S,x)\}^2$。给定 $\mathcal{G}_k$ 的非负随机变量 $|\mathrm{I}_{E,k}|^{-1}\sum_{i\in\mathrm{I}_{E,k}}\mathbf{1}\{T_i = t, X_i = x\}f_k(Z_{E,i})$ 的条件期望为 $P(T = t, X = x \mid E)\|\widehat{\pi}_t^{(-k)} - \pi_t\|_{2,t,x}^2$；根据假设 3.7，该值依概率趋于零，因此根据

上面的条件马尔可夫论证，该随机变量依概率趋于零。由于 $\widehat{\mathbb{E}}_k[f_k(Z_E) \mid T=t,X=x,E]$ 是该随机变量除以 $P_k(T=t,X=x\mid E)$所得的结果，根据 (E4)，它依概率趋于零。断言 (7)–(9)可通过将 $\widehat{\pi}_t^{(-k)}-\pi_t$ 替换为$\widehat{\nu}_t^{(-k)}-\nu_t$, $\widehat{\omega}^{\nu}_{,t}{}^{(-k)}-\omega_t^{\nu}$以及 $\widehat{\omega}^{\pi}_{,t}{}^{(-k)}-\omega_t^{\pi}$，采用相同的论证得到。对于断言 (10)和 (11)，我们利用给定 $\mathcal{G}_k$时 $|\mathcal{I}_{O,k}|^{-1}\sum_{j\in\mathcal{I}_{O,k}}\mathbf{1}\{X_j=x\}\mathbf{1}\{T_j=t\}\{\widehat{\omega}^{\nu}_{,t}{}^{(-k)}(M_j,x)-\omega_t^{\nu}(M_j,x)\}^2$的条件期望至多为$P(X=x\mid O)\mathbb{E}[\pi_t(S,x)\{\widehat{\omega}^{\nu}_{,t}{}^{(-k)}(M,x)-\omega_t^{\nu}(M,x)\}^2\mid T=t,x,E]\le\|\widehat{\omega}^{\nu}_{,t}{}^{(-k)}-\omega_t^{\nu}\|_{2,t,x}^2$，根据引理 B.1(1)以及 $\pi_t\le 1$，并采用相同的论证。对于 $\widehat{\omega}^{\pi}_{,t}{}^{(-k)}$ 同样成立。 □

C.2.3 各折平均值

引理 C.5（各折平均值之和）。 假设假设 *3.1* 和 *3.2* 成立，并固定 t,x。若实验观测的函数 f 满足 $\mathbb{E}[f(Z_E)\mid T=t,x,E]=0$ 以及$\mathbb{E}[f(Z_E)^2\mid T=t,x,E]<\infty$，则

$$\frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbb{E}}_k[f(Z_E)\mid T=t,x,E]=\frac{1}{n_E}\sum_{i=1}^{n_E}\frac{\mathbf{1}\{T_i=t,X_i=x\}f(Z_{E,i})}{P(T=t,X=x\mid E)}+O_p(N^{-1}) \tag{C.21}$$

且右侧的和为 $O_p(N^{-1/2})$。若观察样本中观测值的函数 g 满足 $\mathbb{E}[g(Z_O)\mid x,O]=0$ 以及 $\mathbb{E}[g(Z_O)^2\mid x,O]<\infty$，则将（*C.21*）中的 $\mathcal{I}_{E,k}$, n_E、$\mathbf{1}\{T_i=t,X_i=x\}$ 和 $P(T=t,X=x\mid E)$ 上的样本均值分别替换为 $\mathcal{I}_{O,k}$, n_O、$\mathbf{1}\{X_j=x\}$ 和 $P(X=x\mid O)$ 上的样本均值后，所得形式也成立。

引理 C.5 的证明。我们在 (E4) 成立的事件上进行讨论；根据附录 C.2.1，该事件的概率趋近于 1。将 (C.5) 中的论证应用于折 k、 中的观测值；根据假设 3.1，该折中的观测值数量至少为 N 乘以一个正数常数。这表明 $\mathbb{E}_k[f(Z_E)\mid T=t、\ X=x、\ E]$ 等于 $|\mathcal{I}_{E、k}|^{-1}\sum_{i\in\mathcal{I}_{E、k}}\mathbf{1}\{T_i=t、X_i=x\}f(Z_{E、i})/P(T=t、\ X=x\mid E)$ 加上 $O_p(N^{-1})$。各折的大小至多相差一，因此对于足够大的 N，$|n_E-K|\mathcal{I}_{E、k}|\|<K$，并且 $|1/(K|\mathcal{I}_{E、k}|)-1/n_E|\le 1/(|\mathcal{I}_{E、k}|n_E)\le CN^{-2}$。根据切比雪夫不等式，$\sum_{i\in\mathcal{I}_{E、k}}\mathbf{1}\{T_i=t、\ X_i=x\}f(Z_{E、i})=O_p(N^{1/2})$，因此对 k 求平均，会将这些按折归一化的和替换为 (C.21) 右侧的和（误差为 $O_p(N^{-3/2})$），从而得到 (C.21)。根据切比雪夫不等式，(C.21) 右侧的和为$O_p(N^{-1/2})$，观测版本同理可得。 □

C.2.4 定理*3.3*的证明

定理 3.3 的证明。 我们首先证明 (C.3)。固定 t，x 并在附录 C.2.1 的事件上进行分析。 在附录 C.1.3 的 (C.4) 中，将全样本 $\mathbb{E}$和全样本估计量替换为 $\mathbb{E}_k$ 在折 k 上的值以及排除折 k 的估计量，将$\bar{\mu}_{t,\ k}(x)-\bar{\mu}_t(x)$拆分为五项，我们也将其称为 (I)–(V)。验证 (C.4) 的计算只使用了样本均值和

期望的线性性、引理 B.1(3) 以及 $\mathbb{E}[\mu_t(S,\ x)\ |T=t,\ x,\ E]=\bar{\mu}_t(x)$，因此同样适用于折 k，且在附录 C.2.1 的事件上所有期望均为有限值。

将这五项对 $k=1,\ \ldots\ldots,\ K$ 取平均，在左端得到 $\widehat{\bar{\mu}}_t(x)\ -\bar{\mu}_t(x)$。项 (I) 是附录 C.1.3 中函数 $f(Z_E)\ =\mu_t(S,x)-\bar{\mu}_t(x)-\nu_t(S,x)\omega_t^{\nu}(M,x)+\pi_t(S,x)\omega_t^{\pi}(M,x)$ 在折 k 上的样本均值，该函数均值为零且具有有限的二阶矩，因此由引理 C.5，项 (I) 对 k 的平均等于 (C.3) 中的第一个和加上 $O_p(N^{-1})$。项 (II) 是附录 C.1.3 中函数 $g(Z_O)\ =\mathbf{1}\{T=t\}\{Y\omega_t^{\nu}(M,x)\ -\omega_t^{\pi}(M,x)\}$ 在折 k 上的样本均值，因此由引理 C.5 的观测版本，其对 k 的平均等于 (C.3) 中的第二个和加上 $O_p(N^{-1})$。这两个和均为 $O_p(N^{-1/2})$。项 (III) 的三行在不计符号的意义下是引理 C.4(1)–(3) 中的三个量，项 (IV) 的两行在不计符号的意义下是引理 C.4(4)(5) 中的两个量，因此项 (III) 与项 (IV) 为 $o_p(N^{-1/2})$。

项 (V) 是将 (3.9) 中的两个样本均值替换为期望并减去 $\bar{\mu}_t(x)$ 后得到的量。在附录 C.2.1 的事件上，(E5) 和 (E6) 使得对于排除折 k 的估计量，针对每一组样本值，均满足引理 C.1 的前提。由于引理 C.1 的证明由将估计量视为固定函数所进行的计算组成，因此，项 (V) 等于将 (C.1) 中的 $\widehat{\pi}_t,\widehat{\nu}_t,\widehat{\mu}_t,\widehat{\omega}_t^{\nu},\widehat{\omega}_t^{\pi}$ 替换为 $\widehat{\pi}_t^{(-k)},\widehat{\nu}_t^{(-k)},\widehat{\mu}_t^{(-k)},\widehat{\omega}^{\nu}{}_{,t}^{(-k)},\widehat{\omega}^{\pi}{}_{,t}^{(-k)}$ 后所得的三个期望之和。根据附录 C.1.3 中的界 (C.7) 和 (C.8)，这三个期望的绝对值分别至多为附录 C.1.3 中三个量的某个常数倍；这些量是将估计量写成排除折 k 的形式，并依据假设 3.7 等于 $o_p(N^{-1/2})$。由于 K 是固定的，项 (III)、(IV) 和 (V) 在 k 上的平均值也为 $o_p(N^{-1/2})$，并且 (C.3) 成立，右侧的两个和分别为上文所示的 $O_p(N^{-1/2})$。将附录 C.1.4 的论证应用于该 (C.3)，即可得到 (3.8) 中的第一个收敛结果以及交叉拟合估计量的 $\widehat{\tau}-\tau=O_p(N^{-1/2})$；后者将在下文用于界定 (C.17) 中的项 (IV)。

最后，我们证明 (C.13)。附录 C.1.5 中从 (C.13) 推导出 $\widehat{\Sigma}^2\to_p\Sigma^2$ 的证明部分无需修改。对于实验样本中的观测值 $i\in\mathrm{I}_{E,\ k}$，将附录 C.1.5 中 (C.15) 的 $\widehat{\mu}_t,\ \widehat{\nu}_t,\ \widehat{\pi}_t,\ \widehat{\omega}_t^{\nu},\ \widehat{\omega}_t^{\pi}$ 替换为 $\widehat{\mu}_t^{(-k)},\ \widehat{\nu}_t^{(-k)},\ \widehat{\pi}_t^{(-k)},\ \widehat{\omega}^{\nu}{}_{,\ t}^{(-k)},\ \widehat{\omega}^{\pi}{}_{,\ (-k)}^{\ t}$ 后，结论仍然成立。我们用 f_k 表示 (C.15) 中项 (I) 方括号内的函数。由 $|\mathrm{I}_{E,\ k}|\le n_E$ 和 $\sum_t P_k(T=t,\ X=x\mid E)\ \le 1$，

$$\frac{1}{n_E}\sum_{i\in\mathcal{I}_{E,k}}\mathbf{1}\{T_i=t,X_i=x\}f_k(S_i,M_i)^2$$
$$=\frac{|\mathcal{I}_{E,k}|}{n_E}\widehat{P}_k(T=t,X=x\mid E)\widehat{\mathbb{E}}_k[f_k(S,M)^2\mid T=t,x,E]\le\widehat{\mathbb{E}}_k[f_k(S,M)^2\mid T=t,x,E]$$

与附录 C.1.5 中一样，我们将 f_k 的平方拆分为三个差的平方，使用引理 C.4(6)–(9) 替代假设 3.6(6)–(9)，并将大数定律应用于

折 k 内 (C.16) 中 $\omega_t^{\nu}(M,x)^2\mathbf{1}\{|\omega_t^{\nu}(M,x)|>r\}$ 的样本均值，这表明 $\widehat{\mathbb{E}}_k[f_k(S,M)^2 \mid T=t,x,E]$ 依概率趋于零。(C.15) 中项 (II) 和 (III) 的界与附录 C.1.5 中相同，对有限多个 (t,x,k) 求和可得 (C.13) 的第一个方程。对于观察样本，对于观测值 $j \in \mathrm{I}_{O,k}$，将 $\widehat{\omega}_t^{\nu}, \widehat{\omega}_t^{\pi}$ 替换为 $\widehat{\omega}^{\nu}{}_{,t}^{(-k)}, \widehat{\omega}^{\pi}{}_{,t}^{(-k)}$ 后的 (C.17) 版本成立，并且 (C.17) 中项 (IV) 的界与附录 C.1.5 中相同。我们将 (C.17) 中项 (V) 和 (VI) 的平方乘以 $\mathbf{1}\{X_j=x\}$，对 $j \in \mathrm{I}_{O,k}$ 求和，再除以 n_O。根据 $|\mathrm{I}_{O,k}| \le n_O$ 和 $P_k(X=x \mid O) \le 1$，结果至多为

$$2C^2\widehat{\mathbb{E}}_k[\mathbf{1}\{T=t\}\{\widehat{\omega}^{\nu}{}_{,t}^{(-k)}(M,\ x)-\omega_t^{\nu}(M,\ x)\}^2 \mid x,\ O]+2\widehat{\mathbb{E}}_k[\mathbf{1}\{T=t\}\{\widehat{\omega}^{\pi}{}_{,t}^{(-k)}(M,\ x)-\omega_t^{\pi}(M,\ x)\}^2 \mid x,\ O]$$

对于项 (V) 的 $t=1$ 和项 (VI) 的 $t=0$，根据引理 C.4(10)(11)，它依概率趋于零。对有限多个 $(x,\ k)$ 求和，得到 (C.13) 的第二个方程。由于 (C.3) 和 (C.13) 成立，附录 C.1.4 以及附录 C.1.6 前半部分中的论证给出了定理 3.3 关于渐近分布和方差估计量的断言。定理 3.3 的效率论断见附录 C.1.6。 □

附录 D 模拟与实证设计的说明

本附录表明，在第 4 节的模拟设计中，权重函数并非唯一；并衡量第 5 节实证示例中偏离假设 2.7 的程度。

D.1 模拟设计中的权重函数

在第 4.1 节的两种设计中，都存在满足 (3.2) 的有界权重函数，补充材料 2 的第 S2.1.1 节构造了这些函数。这些函数并非唯一。给定实验中的 $T=t$ 和 $X=x$，$M=S+U_2$，其中 U_2 在 $[-1,1]$ 上服从均匀分布，且与 S 独立，因此对每个 s,

$$\mathbb{E}[\sin(\pi M) \mid S=s,T=t,x,E]=\frac{1}{2}\int_{s-1}^{s+1}\sin(\pi m)\,dm=\frac{\cos(\pi s-\pi)-\cos(\pi s+\pi)}{2\pi}=0$$

因此，将带有 $c \neq 0$ 的 $c\sin(\pi M)$ 加到 (3.2) 的一个解上，会得到另一个解。由于 M 具有由 $T=t$ 和 $X=x$ 给出的密度，$\sin(\pi M)$ 以概率 1 满足 $\neq 0$，所以这两个解不同，在这些设计中，权重函数并不唯一（注3.1）。

D.2 实证示例中的假设 2.7

在 9,084 人中，假设 2.7 不成立，但它排除的差异与收入类别之间的差异相比很小。在这一构造中，观察性总体中给定 $T = t$、$S = k$ 和 $X = x$ 时，(M, Y) 的条件分布等于满足 $T_i = t$、$S_i = k$ 和 $X_i = x$ 的人员中的经验分布，因为选择概率仅通过 $S(t)$ 依赖于潜在值（补充材料 2 的第 S2.2.3 节）。因此，假设 2.7 要求在这些组内，Y 的均值在 M 的每个水平上都相同。在单元格和组别内，收入类别解释了 9,084 人中 Y 方差的 26.7%；在收入类别内加入 M 的各水平后，这一份额增加了 0.61 个百分点，相当于类别所解释份额的 2.3%。对于以下检验和规模度量，我们剔除在其组内某个 M 水平上仅有的四人，剩下 9,080 人。在每个组内检验 M 各水平间均值相等的置换检验，将各组内 M 的水平置换 2,000 次，并以各组水平间平方和之和作为统计量，得到的 p 值为 0.001。其参考分布将组内人员视为在 M 各水平之间可交换。进行 Wald 检验时，我们还进一步剔除 M 的一个水平，因为该水平上的人员 Y 的取值都相同。采用异方差稳健（HC3）标准误、允许 Y 的方差在各水平间不同，对剩余水平间均值相等进行 Wald 检验，得到的 p 值低于 0.0001。为衡量差异的大小，我们保留至少有两个 M 水平的组，这些组共包含 7,987 人，并基于这些人员计算组均值。在这些人员中，M 某一水平上的 Y 均值与其所在组的均值平均相差 9.4 美元；使用相同的稳健标准误计算这些绝对差异的单侧 95% 置信上界，其平均值为 20.8 美元。在单元格和组别内，相邻收入类别的均值平均相差 135.5 美元，因此这些上界相当于该差异的 15%。

偏离仅通过这些差异影响估计量。考虑根据总体矩计算的 Proposed（未惩罚），其中 9,084 人作为实验总体，观察性总体采用第5.1节中的总体。在此计算中，选择概率、求解矩方程的矩阵以及 ω_t^{ν} 对 $\varphi_L(M)$ 的系数均不涉及 Y；由于 $\pi_t > 0.20$，选择概率始终高于截断常数。因此，Y 仅通过 $\mathbb{E}[\mathbf{1}\{T = t\}Y\varphi_L(M) \mid x, O]$ 进入方程（2.7）的第二个方程和方程（3.5）的观察性项，所以该值关于 Y 呈线性，并且仅通过观察性总体中给定 (X, T, S, M) 时的均值依赖于 Y。将 Y 替换为给定 T、X 时的均值，并在 9,084 人中保持 S 不变；对于这些人，假设 2.7 成立

且每个单元格中的组别均值保持不变，计算得出 $\tilde{\tau}$。因此，实际 Y 与 $\tilde{\tau}$ 下的值之差，是 9,084 人的所有组中差异 $\mathbb{E}[Y \mid T=t, S=k, M=m, X=x, O] - \mathbb{E}[Y \mid T=t, S=k, X=x, O]$ 的线性函数，而上述度量对此进行了总结。在选择规则 A、B 和 C 下，该差异分别为 -2。61、2。49 和 -3。33；这些数值分别是表 4 中 Proposed（Tikhonov）标准误的 0.29、0.32 和 0.36 倍，并且超过其偏差的一半。

与估计量的抽样变异相比，这一偏移很小。如果某个估计值服从正态分布，标准差为 σ，且其中心偏离 $\tilde{\tau}$ $\delta\sigma$，则以该估计值为中心、半宽为 1.96σ 的区间以 $\Phi(1.96-\delta) - \Phi(-1.96-\delta)$ 的概率覆盖 $\tilde{\tau}$，其中 Φ 为标准正态分布函数。该概率随 $\delta \geq 0$ 增大而降低，并且当 $\delta < 0.4$ 时超过 93%。因此，假设 2.7 在以下意义下近似成立。上述检验可以检测出该假设的失效，但它所导致的所提出的方法（未惩罚）的总体值偏移量约为来自 4,500 个样本的估计量标准差的三分之一；在正态近似下，这种大小的偏移会使 95% 区间的覆盖率降低不到两个百分点。

附录 E 补充材料

本附录介绍了随本文一同提供的两份补充材料，相关材料将在 https://kkawato.github.io/ 上提供。

补充材料 1 讨论了注 2.1 的设计：在该设计中，两个样本共享的是该特征的处理前代理变量，而不是处理后测量。文中陈述了该设计的假设和识别结果，给出了有效影响函数和不进行样本拆分的估计量，报告了一项模拟研究，证明了识别结果及有效影响函数的形式，并在其示例的加性误差情形和模拟设计中验证了完备性。此外，文中还给出了在两个样本均记录处理前代理变量和共享的治疗后测量时的识别条件。

补充材料 2 描述了如何计算估计量。对于第 4 节和补充材料 1 的模拟研究，其中给出了各设计的总体函数、每次重复实验中所提出的方法和比较估计量的计算过程，以及评估结果。对于第 5 节的实证示例，其中给出了选择规则、样本构建、估计量的计算过程，以及该构建所满足的假设。此外，还报告了其他结果，即图 2 中的极端估计值、模拟研究和实证示例中选定的惩罚项、三种选择规则下选择概率的范围、Job

Corps变量的描述性统计、实证示例中完备性背后的秩条件，以及包含两类、四类和五类收入类别时的结果。最后给出了实现这些计算的 R代码。