arXiv:2603.07780v1 [econ.EM] 8 Mar 2026

内生性检验：基于矩的贝叶斯方法

*

Siddhartha Chib†
Minchul Shin‡
Anna Simoni§

2026年3月10日

摘要

在线性回归模型的贝叶斯估计中，一个标准假设是回归变量具有外生性，即它们与模型误差项不相关。然而，在实践中，这一假设可能不成立。本文运用指数倾斜经验似然框架，提出一种内生性贝叶斯因子检验，将一个在外生性成立时设定正确、但在存在内生性时设定错误的基准模型，与一个无论哪种情况下都设定正确的扩展模型进行比较。我们对数边际指数倾斜经验似然进行了全面研究。我们证明，从频率学派的角度看，该检验程序具有一致性：随着样本量增大，它几乎必然当且仅当回归变量具有外生性时选择基准模型，而当且仅当回归变量具有内生性时选择扩展模型。我们通过模拟数据，以及有关汽车价格对汽车需求的因果效应、可能具有内生性的机票价格对客运量的因果效应等问题，展示了这些方法的应用。

关键词：贝叶斯推断；因果推断；指数倾斜经验似然；内生性；外生性；工具变量；边际似然；后验一致性。

*本文所表达的观点仅代表作者本人，不一定代表费城联邦储备银行或联邦储备系统的观点。

†圣路易斯华盛顿大学奥林商学院，CampusBox1133，1BookingsDrive，密苏里州圣路易斯，邮编 63130。电子邮箱：chib@wustl.edu。

‡费城联邦储备银行研究部，10 Independence Mall，宾夕法尼亚州费城，邮编 19106，电子邮箱：visiblehand@gmail.com。§CREST、法国国家科学研究中心、ENSAE、巴黎综合理工学院、巴黎理工学院，5 Avenue Henry Le Chatelier，法国帕莱索，邮编 91120，电子邮箱：simoni.anna@gmail.com。

1 引言

考虑半参数线性回归模型

$$y = x'\beta + z_1'\gamma + \varepsilon,$$

其中， $y \in \mathbb{R}$ 是结果变量， $x \in \mathbb{R}^{d_x}$ 是我们感兴趣的处理变量向量， $z_1 \in \mathbb{R}^{d_{z_1}}$是控制变量向量， ε是未观测扰动项。贝叶斯分析中的一个常见假设是，回归变量 x是外生的，即它们与误差项 ε不相关。在许多实证情境中，这一假设值得怀疑。如果能够获得一组有效工具变量 $z_2 \in \mathbb{R}^{d_{z_2}}$，且其维度至少与 x相同，那么便可以开展能够正确处理内生性的贝叶斯分析。这类分析可在参数框架和半参数框架下进行，早期研究见 Drèze（1976），后续发展包括 Kleibergen 和 van Dijk（1998）、Chao 和 Phillips（1998）、Kleibergen 和 Zivot（2003）、以及 Schennach（2005）等。近期研究包括 Hoogerheide、Kleibergen 和 van Dijk（2007），Liao 和 Jiang（2011），Florens 和 Simoni （ 2012 ）、 Florens 和 Simoni （ 2016 ）、 Florens 和 Simoni （ 2021 ），Kato（2013），Shin（2014），以及 Chib、Shin 和 Simoni（2018），将这些思路拓展到了半参数和无似然设定。

贝叶斯文献中一个鲜少受到关注的重要问题是内生性检验。频率学派方法，例如经典的 Durbin-Wu-Hausman检验，提供了一种渐近程序，通过比较在不同假设下均具有一致性的估计量来评估外生性。然而，这些程序并不能自然地转化为贝叶斯框架。从贝叶斯的角度来看，将内生性检验视为模型之间的比较，而非参数之间的比较，概念上更为直接。具体而言，可以构建一种检验，根据数据对外生性模型和内生性模型的相对支持度进行比较。

为构建这一方法，并避免分布假设，我们在贝叶斯

矩条件模型框架下展开分析。我们考虑两种相互竞争的设定。第一种是由以下矩条件定义的基准模型

$$\mathbf{E}[\varepsilon(\theta)x] = 0, \quad \mathbf{E}[\varepsilon(\theta)z_1] = 0, \quad \mathbf{E}[\varepsilon(\theta)z_2] = 0,$$

其中， $\varepsilon(\theta) = y - x'\beta + z_1'\gamma$ 和 $\theta := (\beta, \gamma)$。第二种是一个放宽外生性限制的扩展模型，并允许

$$\mathbf{E}[\varepsilon(\theta)x] = v,$$

其中， v 刻画了误差项与内生变量 x 之间的协方差。我们通过非参数指数倾斜经验似然（ETEL）对每个模型进行先验-后验分析，然后根据边际似然和贝叶斯因子比较这两个模型。这种方法具有若干方法论优势。贝叶斯因子通过连续尺度衡量支持这两个模型的证据强度，而非采用严格的接受或拒绝规则。此外，使用 ETEL 能够提高对 (y, x, z_1, z_2) 联合分布模型设定错误的稳健性，并使我们无需指定扰动项的分布，也能获得仍然有效的结果。

利用 Chib（1995）边际似然恒等式，我们知道，对数边际似然可分解为三个部分：对数指数倾斜经验似然、对数先验以及负的对数后验密度值。我们证明，该表达式在渐近意义下等于一个依概率有界的项，加上一个与真概率分布和满足矩约束的最接近概率分布之间的 Kullback-Leibler 散度成比例的项，再加上一个对应于贝叶斯信息准则（BIC）惩罚项的惩罚项。该惩罚源于对在参数真值或伪真值处评价的后验密度进行变量变换。该变换的雅可比行列式的对数构成了惩罚项，而当 n 增加时，局部参数在零处的后验密度依概率有界。因此，当 x 具有外生性时，两个模型的对数指数倾斜经验似然在渐近意义下相同，但惩罚项不同。当x 具有内生性时，对数指数倾斜经验似然之差占主导地位，从而导致选择

扩展模型。因此，在大样本下，该检验能够正确区分两个数据生成过程。我们的构造与豪斯曼检验的逻辑相似：豪斯曼检验比较一个在存在内生性时不一致的估计量与一个在这种情况下仍然一致的估计量。这里的比较对象是矩条件数量不同的模型。从这个意义上说，我们的检验可以被视为豪斯曼设定检验的贝叶斯对应物。

与 Chib 等人（2018）相比，我们的研究建立在同一贝叶斯 ETEL 框架之上，但作出了几项重要贡献。首先，尽管 Chib 等人（2018）解释了如何在不同模型之间进行检验，但并未说明如何构建实际应用中检验感兴趣假设所需的特定模型，例如我们在此考察的内生性问题。本文明确构建了检验内生性所需的模型。其次，我们引入了一项假设，以保证 ETEL 函数存在；据我们所知，现有 ETEL 文献中并未包含这一假设。该假设确保以概率趋近于一的方式，ETEL 函数至少在真实参数值的某个合适邻域内存在。之所以会出现这一问题，是因为作为约束优化问题解的 ETEL 函数，对于某些参数值可能会出现可行集为空的情况 θ。若没有这一假设，就无法定义 ETEL 函数的导数，这会给频率学派和贝叶斯 ETEL 方法都带来困难。第三，我们给出了一个更直接的证明，表明 ETEL 函数渐近等价于一个二次函数。这一结果为我们建立伯恩斯坦–冯·米塞斯定理奠定了基础，我们进而利用该定理证明了检验程序的一致性。该直接证明利用了工具变量（IV）回归问题所隐含的矩约束关于 θ 的线性。基于同样的思路，本文的假设比 Chib 等人（2018）的假设更弱，因为这些假设利用了 IV 线性回归结构。

最后，作为证明我们的检验程序具有一致性的副产品，我们推导了对数边际 ETEL 函数的一种新的渐近表示，其定义为 ETEL 函数关于模型参数先验分布的积分。我们表明，每个模型的对数边际似然可以渐近分解为一个库尔贝克-莱布勒（KL散度）项（真实分布与满足该模型

矩约束）加上一个 BIC 型惩罚项。我们采用一种新颖的方法推导该惩罚项：通过局部参数的变量替换，重新表达真实（或伪真实）参数值处的后验密度值，所得的对数雅可比行列式给出惩罚项，而局部参数在零处取值的后验密度为 $\mathcal{O}_p(1)$（当 $n \to \infty$ 时）。这一表示阐明了此情形下贝叶斯因子检验的机制，并给出了比 Chib 等人（2018）更为清晰的模型选择一致性证明。值得注意的是，我们强调，仅当 x 为外生时，该惩罚项才在选择正确模型中发挥作用，此时两个模型均设定正确。

本文其余部分安排如下。第 2 节概述了使用指数倾斜经验似然（ETEL）对矩条件模型进行贝叶斯估计与比较。第 3 节介绍基准模型和扩展模型，并给出一个模拟示例以说明该程序。第 4 节提出内生性检验，并分析对数边际似然的大样本行为，从而确立该检验的一致性。第 5 节给出实证示例，第 6 节作结。附录包含主要结果的证明。

2 预备知识

本节简要介绍使用指数倾斜经验似然（ETEL）对矩条件模型进行贝叶斯估计的背景。进一步的细节可参见 Schennach（2005）和 Chib 等人（2018）。

2.1 矩约束与可行分布

设 $w \in \mathbb{R}^{d_w}$ 为一个随机向量，$\theta \in \Theta \subset \mathbb{R}^p$ 表示一般的参数向量。$\mathbb{M}$表示 $\mathbb{R}^{d_w}$上的所有概率分布的集合。对于一组已知的矩函数

$$g(w, \theta) : \mathbb{R}^{d_w} \times \Theta \to \mathbb{R}^d,$$

矩约束为

$$\mathbf{E}^{Q}[g(w,\theta)] = 0, \tag{2.1}$$

其中 $Q \in \mathbb{M}$ 是施加该限制的概率分布，且$\mathbf{E}^{Q}[\cdot]$ 表示在 Q下的期望。对每个 $\theta \in \Theta$，将满足矩约束的分布子集定义为

$$\mathcal{Q}(\theta) := \left\{ Q \in \mathbb{M} : \ \mathbf{E}^{Q}[g(w,\theta)] = 0 \right\}. \tag{2.2}$$

假设数据 $w_{1:n}$:=($w_1, \ldots, w_n$)独立抽取自真实分布 P，该分布不必属于某个 $\theta \in \Theta$下的 $\mathcal{Q}(\theta)$。关于真实分布 P所取的期望记为$\mathbf{E}[\cdot] \equiv \mathbf{E}^{P}[\cdot]$。

2.2 样本 ETEL 权重、倾斜样本分布与似然

(2.1)的经验对应物是加权限制

$$\sum_{i=1}^{n} q_i \, g(w_i,\theta) = 0, \qquad q_i > 0, \qquad \sum_{i=1}^{n} q_i = 1, \tag{2.3}$$

其中$\{q_i\}_{i=1}^{n}$ 是支撑于$\{w_i\}_{i=1}^{n}$的离散分布。等价地，任何这样的权重向量都会在 $\mathbb{R}^{d_w}$ 上诱导出一个以$\{w_1, \ldots, w_n\}$为支撑集的离散概率测度

$$Q_n(\cdot) := \sum_{i=1}^{n} q_i \, \delta_{w_i}(\cdot), \tag{2.4}$$

其中δ_{w_i} 表示 w_i处的点质量。

由于可能不存在满足条件的 $\theta \in \Theta$，使均匀经验分布 $q_i \equiv 1/n$ 满足（2.3），因此我们将 ETEL 权重定义为以下最接近 Kullback–Leibler（KL）散度的可行重新加权问题的解：

$$\{\widehat{q}_i(\theta)\}_{i=1}^{n} := \operatorname{argmax}_{q_1>0,\ldots,q_n>0} \sum_{i=1}^{n} [-q_i \log(n q_i)]$$

$$\text{subject to } \sum_{i=1}^{n} q_i = 1, \qquad \sum_{i=1}^{n} q_i g(w_i, \theta) = 0 \tag{2.5}$$

该问题取决于参数向量 $\theta \in \Theta$。令 $H_n \subset \Theta$ 为使规划问题（2.5）可行的 θ 值集合，即 使凸包 $\{g(w_i, \theta) : i = 1, \ldots, n\}$ 的内部包含零的 θ集合。下文的假设 3 保证，随着样本量增大， H_n 以趋近于 1 的概率非空。

对于 $\theta \in H_n$，ETEL（样本似然）定义为 ETEL 权重的乘积 s:

$$\widehat{q}(w_{1:n}|\theta) = \prod_{i=1}^{n} \widehat{q}_i(\theta).$$

ETEL 是通过对未知分布Q 进行积分而得到的积分似然；该积分基于一种特定的非参数先验，该先验在给定 $\theta \in H_n$ 的条件下施加了矩约束 (2.3)；参见 Schennach (2005)。

给定先验密度 $\pi(\theta)$，基于 ETEL 的后验分布为截断后验分布

$$\pi^n(\theta|w_{1:n}) \propto \pi(\theta)\, \widehat{q}(w_{1:n}|\theta)\, I[\theta \in H_n], \tag{2.6}$$

其中 $I[A]$ 表示示性函数。由于 (2.6) 没有闭形式解，后验汇总量通过专门设计的马尔可夫链蒙特卡罗 (MCMC) 方法获得。附录 D 介绍了 MCMC 抽样及相关计算的细节。

2.3 对偶表示与对数 ETEL 恒等式

一种便捷的计算 $\{\widehat{q}_i(\theta)\}$ 的方法是利用（2.5）的对偶表示。将 ETEL 乘子定义为，对于每个 $\theta \in H_n$

$$\widehat{\lambda}(\theta) \equiv \widehat{\lambda}(w_{1:n}, \theta) := \arg\min_{\lambda \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \exp\left(\lambda' g(w_i, \theta)\right).$$

那么，对于每个 $\theta \in H_n$：

$$\widehat{q}_i(\theta) = \frac{e^{\widehat{\lambda}(\theta)' g(w_i,\theta)}}{\sum_{j=1}^{n} e^{\widehat{\lambda}(\theta)' g(w_j,\theta)}}, \qquad i \le n. \tag{2.7}$$

将 (2.7) 视为对均匀经验分布 $P_n(\cdot)$ 的指数倾斜是很有用的：$=\frac{1}{n}\sum_{i=1}^{n}\delta_{w_i}(\cdot)$，该分布在每个观测值上赋予质量 $1/n$。对于固定的 $\theta \in H_n$，ETEL 权重$\{\widehat{q}_i(\theta)\}$ 因而定义了一个倾斜样本分布

$$\widehat{Q}_n(\cdot \mid \theta) := \sum_{i=1}^{n} \widehat{q}_i(\theta)\, \delta_{w_i}(\cdot), \tag{2.8}$$

即一个关于 P_n绝对连续的分布。特别地，对于每个支撑点 w_i ，

$$\frac{d\widehat{Q}_n(\cdot \mid \theta)}{dP_n}(w_i) = n\,\widehat{q}_i(\theta) = \frac{e^{\widehat{\lambda}(\theta)' g(w_i,\theta)}}{\frac{1}{n}\sum_{j=1}^{n} e^{\widehat{\lambda}(\theta)' g(w_j,\theta)}}, \tag{2.9}$$

因此，$Q_n(\cdot \mid \theta)$ 是 P_n 的指数倾斜，并满足样本矩约束 (2.3)。乘子 $\lambda(\theta)$ 满足样本一阶条件

$$\sum_{i=1}^{n} \widehat{q}_i(\theta)\, g(w_i,\theta) = 0, \tag{2.10}$$

该条件是在 ETEL 权重下 (2.1) 的样本类比。为便于后续使用，我们将对数 ETEL 记录为适合展开的形式。对 $\log\ \widehat{q}_i(\theta)$ 求和可得精确恒等式

$$\log \widehat{q}(w_{1:n} \mid \theta) = \sum_{i=1}^{n} \widehat{\lambda}(\theta)' g(w_i,\theta) - n \log\Big(\sum_{j=1}^{n} \exp\{\widehat{\lambda}(\theta)' g(w_j,\theta)\} \Big), \tag{2.11}$$

或者等价地，

$$\log \widehat{q}(w_{1:n} \mid \theta) = -n\log n + \sum_{i=1}^{n} \widehat{\lambda}(\theta)' g(w_i,\theta) - n \log\Big(\frac{1}{n}\sum_{j=1}^{n} \exp\{\widehat{\lambda}(\theta)' g(w_j,\theta)\} \Big). \tag{2.12}$$

2.4 总体 KL 投影、指数倾斜与伪真值

样本总体的对应形式是分布 $Q^*(\theta) \in \mathcal{Q}(\theta)$，它与 P的 KL 散度最小。对于每个满足 $\mathcal{Q}(\theta) \neq \emptyset$的 θ，定义

$$Q^*(\theta) := \arg \inf_{Q \in \mathcal{Q}(\theta)} \mathrm{KL}(Q||P),$$

其中

$$\mathrm{KL}(Q||P) := \int \log\left(\frac{dQ}{dP}\right) dQ$$

若 Q 关于 P 绝对连续，且 $\mathrm{KL}(Q||P) = +\infty$；否则。ETEL 乘子 $\lambda(\theta)$ 的总体对应量 $\widehat{}$ 为

$$\lambda_*(\theta) := \arg \min_{\lambda \in \mathbb{R}^d} \mathbf{E}[e^{\lambda' g(w_i,\theta)}]$$

对每个满足$\mathcal{Q}(\theta) \neq \emptyset$ 的 $\theta \in \Theta$ 均成立。这诱导出总体指数倾斜 lt

$$q(w;\theta) := \frac{\exp\{\lambda_*(\theta)' g(w,\theta)\}}{\mathbf{E}[\exp\{\lambda_*(\theta)' g(w,\theta)\}]}, \qquad \mathbf{E}[q(w;\theta)] = 1. \tag{2.13}$$

在温和的正则性条件下，KL 投影 $Q^*(\theta)$ 可表示为拉东-尼科迪姆导数：

$$\frac{dQ^*(\theta)}{dP}(w) = q(w;\theta). \tag{2.14}$$

因此，若 P 具有勒贝格密度 $p(w)$，则 $Q^*(\theta)$ 的密度可表示为对 $p(w)$ 进行指数倾斜所得：

$$q^*(w;\theta) = p(w)\,\frac{\exp\{\lambda_*(\theta)' g(w,\theta)\}}{\mathbf{E}[\exp\{\lambda_*(\theta)' g(w,\theta)\}]}.$$

通过测度变换，

$$\mathbf{E}^{Q^*(\theta)}\Big[g(w,\theta)\Big] = \mathbf{E}\Big[q(w;\theta)\, g(w,\theta)\Big]. \tag{2.15}$$

右侧是总体倾斜矩条件。它是在 P 下而非在 $Q^*(\theta)$ 下表述的矩约束。如果一个或多个矩条件设定错误，那么对所有 $\theta \in \Theta$，$Q^*(\theta) \neq P$；而伪真值 θ_* 定义为反向 KL 散度的最小化点

$$\theta_* := \arg \min_{\theta;\mathcal{Q}(\theta)\neq\emptyset} \mathrm{KL}\Big(P \| Q^*(\theta)\Big), \tag{2.16}$$

其中

$$\mathrm{KL}\Big(P \| Q^*(\theta)\Big) := \int \log\Big(\frac{dP}{dQ^*(\theta)}\Big)\, dP \tag{2.17}$$

每当 P 关于 $Q^*(\theta)$绝对连续时。在正确设定下，存在 $\theta_\circ \in \Theta$使得 $P \in \mathcal{Q}(\theta_\circ)$，此时 $Q^*(\theta_\circ) = P$且 $\theta_* = \theta_\circ$。此外，在该情形下$\lambda_*(\theta_\circ) = 0$，因而 $q(w;\theta_\circ) \equiv 1$。最后，当对偶表示成立时，伪真值 θ_* 也可借助总体倾斜表示为

$$\theta_* = \arg \max_{\theta;\mathcal{Q}(\theta)\neq\emptyset} \mathbf{E}\left[\log\Big(\frac{\exp\{\lambda_*(\theta)' g(w,\theta)\}}{\mathbf{E}[\exp\{\lambda_*(\theta)' g(w,\theta)\}]}\Big)\right], \tag{2.18}$$

其中对数内部的项是(2.14)中的拉东-尼科迪姆导数 $[dQ^*(\theta)/dP](w)$。

3 模型

在本节中，我们将第 2 节中的一般矩约束框架具体化为引言中引入的半参数线性回归设定。

3.1 数据、回归结构与目标参数

令 $w := (y, x, z_1, z_2) \in \mathbb{R}^{d+1}$服从一个未知的概率分布 P，其中 $d := d_x + d_{z_1} + d_{z_2}$且 $d_w = d + 1$。全文中，$\mathbf{E}[\cdot] := \mathbf{E}^P[\cdot]$ 表示在P下的期望。我们假设，在 P 下，随机向量 w 服从如下回归模型

$$y = \beta_\circ' x + \gamma_\circ' z_1 + \varepsilon, \qquad \mathbf{E}[\varepsilon(\theta_\circ) z_j] = 0 \quad \text{for } j = 1, 2, \tag{3.1}$$

其中 $\theta_\circ := (\beta_\circ, \gamma_\circ) \in \Theta \subset \mathbb{R}^p$ 是真实的回归系数值，可视为 P: $\theta_\circ \equiv \theta_\circ(P)$ 和 $p = d_x + d_{z_1}$ 的函数。在模型（3.1）中，向量 z_1 包含外生控制变量（包括截距），而 z_2 包含工具变量。我们关注的是 x 对 y的因果效应，用 $\beta_\circ$表示。对于任意 $\theta := (\beta', \gamma')' \in \Theta \subset \mathbb{R}^p$，其中 $p = d_x + d_{z_1}$且 $\tilde{w}_1 := (x', z_1')'$，定义回归残差

$$\varepsilon(\theta) := y - \beta' x - \gamma' z_1 \equiv y - \theta' \tilde{w}_1.$$

如果在 P下 x 是内生的，那么$\mathbf{E}[\varepsilon(\theta_\circ)x] \neq 0$。当 $d_{z_2} \geq d_x$时，工具变量 z_2 有助于$\beta_\circ$，尽管存在内生性。

3.2 基准模型$\mathcal{M}_b$（可能设定错误）

基准模型记为$\mathcal{M}_b$，它施加以下矩约束

$$\mathbf{E}^Q[g_b(w, \theta)] = 0, \qquad Q \in \mathcal{Q}_b(\theta), \tag{3.2}$$

其中，基准模型矩函数为

$$g_b(w, \theta) := \varepsilon(\theta) \begin{pmatrix} x \\ z_1 \\ z_2 \end{pmatrix} \in \mathbb{R}^d, \qquad d = d_x + d_{z_1} + d_{z_2},$$

满足基准模型约束的分布集合为

$$\mathcal{Q}_b(\theta) = \left\{ Q \in \mathbb{M};\, \mathbf{E}^Q[g_b(w, \theta)] = 0 \right\}. \tag{3.3}$$

这里，$\mathbb{M}$ 表示 $\mathbb{R}^{d+1}$上所有概率分布的集合。在外生性条件下，$\mathbf{E}[\varepsilon(\theta_\circ)x] = 0$，且真实分布 P 在 $\theta_\circ$处满足基准模型矩，因此$\mathcal{Q}_b(\theta_\circ)$。在内生性条件下，$\mathbf{E}[\varepsilon(\theta_\circ)x] \neq 0$，因此对于每个 $\theta \in \Theta$，都有 $P \notin \mathcal{Q}_b(\theta)$；此时$\mathcal{M}_b$ 是

设定错误。在这种情况下，由样本 $w_{1:n}$构造的 ETEL 函数，是分布 $Q_b^*(\theta)$ 的经验对应物；该分布对每个 θ 都满足以下矩条件：

$$\mathbf{E}^{Q_b^*(\theta)}[g_b(w,\theta)] = 0$$

并且在集合$\mathcal{Q}_b(\theta)$中的所有分布中，该分布与 P之间的 KL 散度最小，即

$$Q_b^*(\theta) := \operatorname{arginf}_{Q\in\mathcal{Q}_b(\theta)} \mathrm{KL}(Q\|P).$$

注意，如果 Q 相对于 P不是绝对连续的，则 $\mathrm{KL}(Q\|P)$被设为 $+\infty$。此外，

$$\theta_* := \arg \max_{\theta\in\Theta;\mathcal{Q}_b(\theta)\neq\emptyset} \mathbf{E}\log\Big(\frac{e^{\lambda_*'(\theta)g_b(w,\theta)}}{\mathbf{E}[e^{\lambda_*'(\theta)g_b(W,\theta)}]}\Big) \tag{3.4}$$

表示基准模型中的伪真值。下文第 4.5 节给出的假设1保证了该值的存在。另一方面，如果 x 是外生的，则 $Q_b^*(\theta_*) = P$和 $\theta_* = \theta_\circ$，其中 $\theta_\circ$表示上文定义的 θ 的真值。在下文中，我们将基准模型的 ETEL 记为 $\widehat{q}(w_{1:n}|\theta,\ \mathcal{M}_b) := \prod_{i=1}^n \widehat{q}_i(\theta|\mathcal{M}_b)$，其中 $\widehat{q}_i(\theta|\mathcal{M}_b)$按照（2.5）构造，只是将$g(w_i,\ \theta)$替换为 $g_b(w_i,\ \theta)$。

3.3 扩展模型$\mathcal{M}_e$ （设定正确）

扩展模型记为 $\mathcal{M}_e$，通过对内生性部分进行显式参数化来扩充基准模型

$$v := \mathbf{E}[\varepsilon(\theta)x] \in \mathbb{R}^{d_x}.$$

令 $\mathcal{V}\subset\mathbb{R}^{d_x}$ 并定义扩展参数

$$\psi := (\theta, v) \in \Psi, \qquad \Psi := \Theta\times\mathcal{V}. \tag{3.5}$$

扩展模型矩函数为

$$g_e(w,\psi) := g_e(w,\theta,v) := \varepsilon(\theta)\begin{pmatrix} x \\ z_1 \\ z_2 \end{pmatrix} - \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix} = g_b(w,\theta) - \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix}, \tag{3.6}$$

并且该模型施加了矩约束

$$\mathbf{E}^Q\Big[g_e(w,\psi)\Big] = 0, \qquad Q \in \mathcal{Q}_e(\psi), \tag{3.7}$$

其中

$$\mathcal{Q}_e(\psi) := \Big\{Q \in \mathbb{M} : \ \mathbf{E}^Q\Big[g_e(w,\psi)\Big] = 0\Big\}. \tag{3.8}$$

根据构造，无论 x 是外生还是内生，$\mathcal{M}_e$ 的设定都是正确的。事实上，令

$$v_\circ := \mathbf{E}[\varepsilon(\theta_\circ)x]. \tag{3.9}$$

根据 (3.1)，我们有$\mathbf{E}[\varepsilon(\theta_\circ)z_1] = 0$ 以及$\mathbf{E}[\varepsilon(\theta_\circ)z_2] = 0$，因此

$$\mathbf{E}\Big[g_e(w,\psi_\circ)\Big] = 0, \qquad \psi_\circ := (\theta_\circ, v_\circ). \tag{3.10}$$

因此，$P \in \mathcal{Q}_e(\psi_\circ)$，且 KL 投影满足

$$Q_e^*(\psi_\circ) = P. \tag{3.11}$$

在扩展模型中，最小化点 $Q_e^*(\psi) = \arg\inf_{Q\in\mathcal{Q}_e(\psi)} \mathrm{KL}(Q\|P)$ 等于 P，而扩展模型中的总体矩条件为

$$\mathbf{E}^P[g_e(w,\psi_\circ)] = 0.$$

此外，

$$\psi_\circ := \arg \max_{\psi \in \mathcal{V}; \mathcal{Q}_e(\psi) \neq \emptyset} \mathbf{E} \log \left(\frac{e^{\lambda_*'(\psi) g_e(w,\psi)}}{\mathbf{E}[e^{\lambda_*'(\psi) g_e(W,\psi)}]} \right), \tag{3.12}$$

其中，对于满足 $\mathcal{Q}_e(\psi) \neq \emptyset$ 的每个 $\psi \in \mathcal{V}$，$\lambda_*(\psi) := \arg\min_{\lambda \in \mathbb{R}^d} \mathbf{E}[e^{\lambda' g_e(w,\psi)}]$。下文将扩展模型的 ETEL 记为 $\hat{q}(w_{1:n}|\theta, \mathcal{M}_e) := \prod_{i=1}^n \hat{q}_i(\theta|\mathcal{M}_e)$，其中 $\hat{q}_i(\theta|\mathcal{M}_e)$ 按照（2.5）构造，将其中的 $g(w_i, \theta)$ 替换为 $g_e(w_i, \psi)$。

3.4 数值示例

为说明基准模型和扩展模型的拟合，首先考虑存在内生性时的基准模型。设数据生成过程（DGP）为

$$\begin{aligned} y_i &= \gamma_0 + x_i \beta + z_{1i} \gamma_1 + \varepsilon_i \\ x_i &= \delta_0 + z_{1i} \delta_1 + z_{2i} \delta_2 + u_i \\ z_{1i} &= v_i \\ z_{2i} &= \omega_i \end{aligned}$$

对于 $i = 1, \ldots, n$，其中 $n \in \{250, 500, 1000, 2000\}$。假设 (u_i, v_i, ω_i) 的边际分布均为高斯分布，ε_i 的边际分布为偏态高斯混合分布 $0.5\mathrm{N}(0.5, 0.52) + 0.5\mathrm{N}(-0.5, 1.1182)$，$(\varepsilon_i, u_i, v_i)$ 的联合分布由协方差矩阵为 $R = \begin{pmatrix} 1 & 0.6 & 0 \\ 0.6 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ 的高斯 Copula 诱导，并且 ω_i 与其他各误差项的协方差为零。此外，假设每个参数均为 1（δ_1 除外，其值为 .5）。在此数据生成过程中，z_{1i} 与 ε_i 不相关，而与 x_i 相关；但由于 ω_i 与其他冲击均不相关，z_{2i} 是有效工具变量，且与 x_i 相关。对于四种样本量，计算 $\theta := (\beta, \gamma_0, \gamma_1)$ 的后验分布

来自四个矩条件

$$\mathbf{E}\Big[(y_i - x_i\,\beta - \gamma_0 - z_{1i}\,\gamma_1)\begin{pmatrix} x_i \\ 1 \\ z_{1i} \\ z_{2i} \end{pmatrix}\Big] = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

采用定制的单区块 Metropolis–Hastings（M-H）算法（Chib和Greenberg，1995），在 1,000 次预烧期迭代之后，再进行 20,000 次迭代，以抽取 ETEL后验。根据这些马尔可夫链蒙特卡罗抽样得到的样本，计算每种样本量下 β 的边际后验密度。图1给出了后验密度的核平滑版本。如图1所示，随着样本量增加，基准模型下 β 的后验密度集中在一个与真值 β 相差甚远的数值上，这表明忽略内生性导致了模型设定错误。在

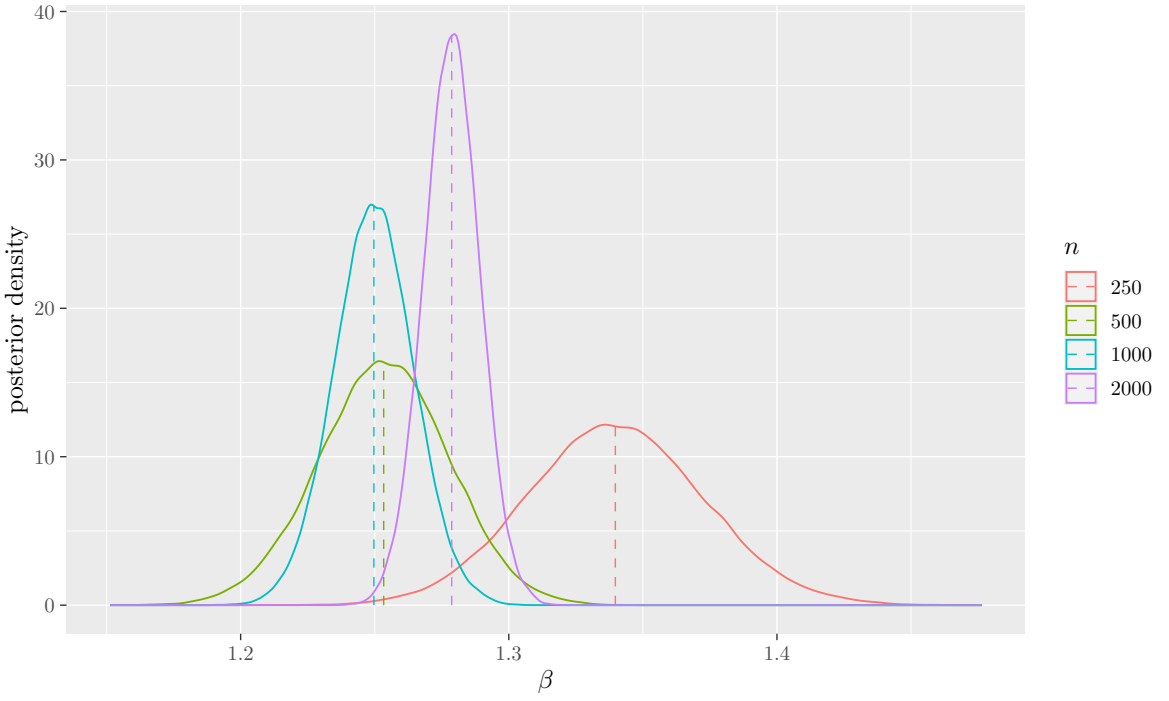

图1：忽略内生性时的基准模型：不同样本量下 β 的边际后验密度。后验均值以虚竖线标示。

扩展（设定正确的）模型中，我们有

$$\mathbf{E}\Big[(y_i - x_i\,\beta - \gamma_0 - z_{1i}\,\gamma_1)\begin{pmatrix} x_i \\ 1 \\ z_{1i} \\ z_{2i} \end{pmatrix}\Big] = \begin{pmatrix} v \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

现在感兴趣的参数为 $\psi := (\beta、\ \gamma_0、\ \gamma_1、\ v)$。我们对 v 采用默认的学生 t 先验，其中心设在广义矩估计（GMM）估计值处，离散程度为广义矩估计渐近方差的 4 倍。θ 的先验与基准模型中的相同。对于四种不同的样本量，均采用定制的单区块 Metropolis–Hastings 方法，在 1,000 次预烧迭代之后再进行 20,000 次迭代，以抽取 ETEL后验。 β 的边际后验密度见图 2， v 的边际后验密度见图3。可以看出，即使对于 $n = 250$， β的后验分布也接近 β的真值；而当 $n = 2{,}000$ 时，后验分布基本以真值为中心。此外， v （即 cov(x、 ε)）的后验分布趋于集中在真值 0.6 附近。

4 检验程序

4.1 贝叶斯因子

我们的内生性贝叶斯检验由模型$\mathcal{M}_e$ 相对于模型 $\mathcal{M}_b$的贝叶斯因子给出，定义如下：

$$\mathrm{BF}_{eb} = \frac{m(w_{1:n}|\mathcal{M}_e)}{m(w_{1:n}|\mathcal{M}_b)},$$

其中， $m(w_{1:n}|\mathcal{M}_b) := \int_{H_n} \widehat{q}(w_{1:n}|\theta, \mathcal{M}_b)\pi(\theta)d\theta$ 和 $m(w_{1:n}|\mathcal{M}_e) := \int_{H_n} \widehat{q}(w_{1:n}|\psi, \mathcal{M}_e)\pi(\psi)d\psi$是由指数倾斜经验似然函数（下文也称为边际指数倾斜经验似然函数）得到的模型边际似然。我们采用 Chib（1995）的方法计算这些值，该方法由 Chib 和 Jeliazkov（2001）扩展至一般的 Metropolis–Hastings 链。若 $\log(\mathrm{BF}_{eb}) > 0$，则选择$\mathcal{M}_e$ 而非 $\mathcal{M}_b$ ；否则选择 $\mathcal{M}_b$。

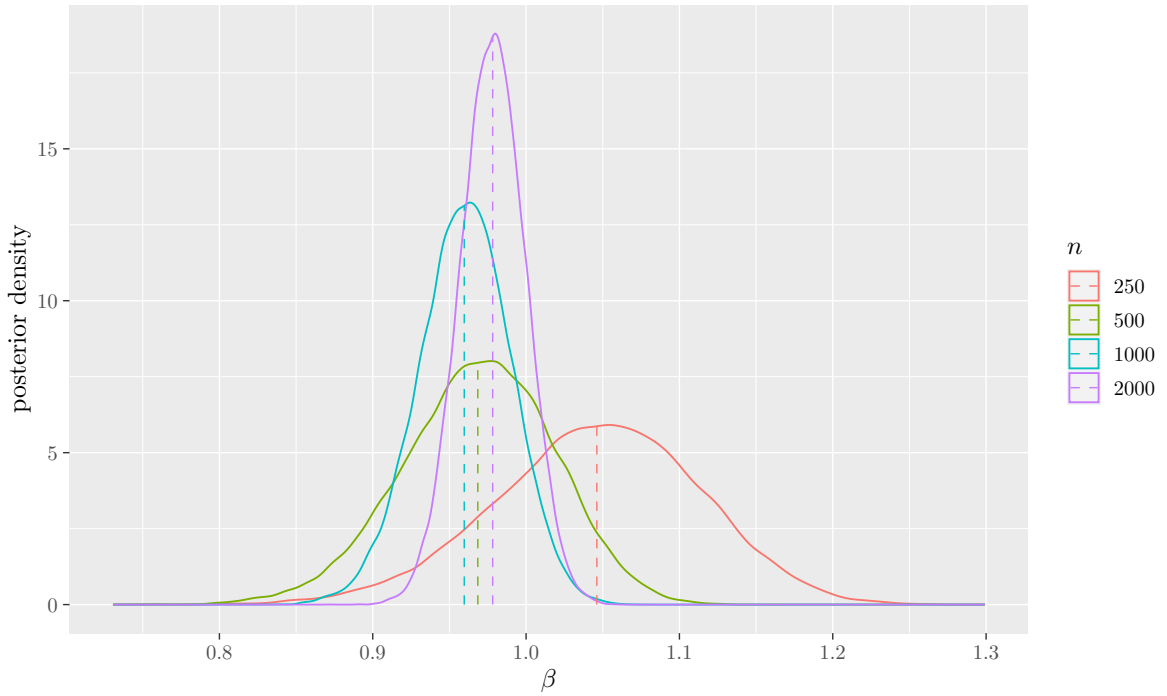

图 2：扩展模型（x_i 矩未激活）：不同样本量下 β 的边际后验密度。后验均值以虚竖线表示。

根据 Chib 等人（2018）的理论，要对矩条件模型进行有效比较，参与比较的模型必须来自同一个包容模型，且矩条件的数量应相同。我们通过在基准模型中纳入$\mathbf{E}[\varepsilon_i(\theta)z_{2,i}] = 0$ 限制条件，并且不从扩展模型中排除$\mathbf{E}[\varepsilon_i(\theta)x_i] = v$ 条件，确保满足这一要求。

直观而言，贝叶斯因子会选出正确模型，因为当 x 为外生时，$\mathcal{M}_b$ 设定正确；当 x 为内生时，$\mathcal{M}_b$ 设定错误；然而，$\mathcal{M}_e$ 在这两种情况下都设定正确。因此，根据 Chib 等人（2018），当 x 为外生时，贝叶斯因子会倾向于选择具有（$d-p$）个过度识别约束的 $\mathcal{M}_b$，而非具有（$d-p-d_x$）个过度识别约束的 M_e（因为它的过度识别约束比$\mathcal{M}_e$更多）；而当 x 为内生时，则会倾向于选择$\mathcal{M}_e$（因为此时$\mathcal{M}_b$ 会设定错误）。

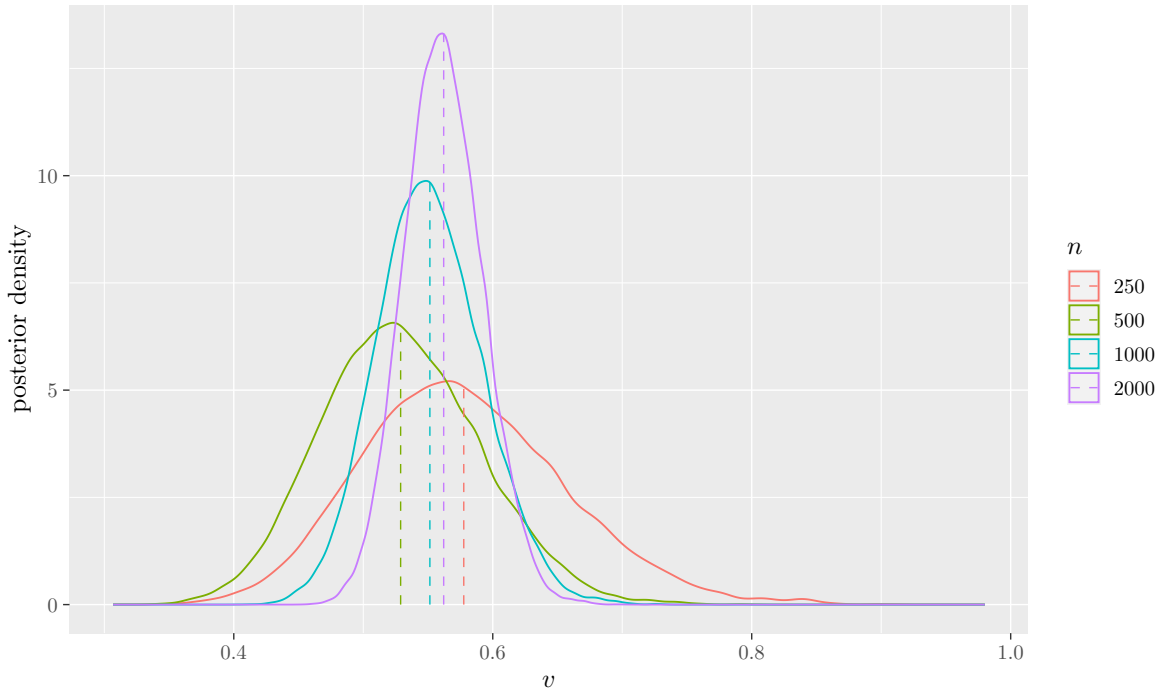

图3：忽略内生性时的扩展模型：不同样本量下 v =cov(x, ε) 的边际后验密度。后验均值由虚竖线标示。

4.2 原理

本节将阐述我们检验程序背后的原理。我们要检验的假设如下：

$$H_{miss}: \quad P \text{ is such that } \nexists \theta \in \Theta \subset \mathbb{R}^p \text{ such that } \mathbf{E}[\varepsilon_i(\theta)x_i] = 0 \qquad (endogeneity)$$

反对

$$H_{cs}: \quad P \text{ is such that } \exists \theta \in \Theta \subset \mathbb{R}^p \text{ such that } \mathbf{E}[\varepsilon_i(\theta)x_i] = 0 \qquad (exogeneity).$$

这里，下标 $miss$和 cs 分别表示模型设定错误和正确设定。前述假设也可等价地写作 H'_{miss}： $v \neq 0$ 和 H'_{cs}： $v = 0$。我们基于 $\mathbf{BF}_{eb}$ 的方法等价于一种贝叶斯检验：基于在 v 上的混合先验 $\pi_0\delta_0(v) + (1-\pi_0)\pi(v)$，检验 H'_{miss} 与 H'_{cs}，其中 $\delta_0(\cdot)$表示零点处的狄拉克质量， $\pi_0 \in [0,1]$，并且

$\pi(\cdot)$是连续分布。这两种方法的贝叶斯因子在数值上相同。检验程序如下：如果 $\text{BF}_{eb} \geq 1$，我们得出结论认为 x是内生的（即 接受 H_{miss}）；如果 $\text{BF}_{eb} < 1$，我们得出结论认为 x是外生的（即 接受 H_{cs}）。

下一个定理表明， H_{miss} 和 H_{cs} 可以用 P 与满足我们想要检验的矩约束以及已知对 P 成立的额外矩约束的分布集合$\mathcal{Q}_b(\theta)$之间的库尔贝克-莱布勒散度来表示。

定理 4.1 假设存在一个 $\theta \in \Theta$，使得$\mathbf{E}[\varepsilon_i(\theta)(z'_{1,i}z'_{2,i})'] = 0$，并且假设*1*成立。考虑以下陈述：

(i). P使得 $\nexists\theta \in \Theta$满足$\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$。*(ii).* $\text{KL}(P||Q_b^*(\theta_*)) > 0$。*(iii).* P使得 $\exists\theta \in \Theta$ 满足$\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$。*(iv).* $\text{KL}(P||Q_b^*(\theta_*)) = 0$。

那么，*(i)* 等价于 *(ii)*，而 *(iii)* 等价于 *(iv)*。

该定理清楚地表明，要检验 H_{miss}和 H_{cs}，可以等价地关注 Kullback-Leibler散度 KL($P||Q_b^*(\theta_*)$)。我们的贝叶斯检验基于贝叶斯因子和边际似然的比较。边际似然与 Kullback-Leibler散度之间存在严格联系：基准模型的对数边际似然渐近表现为 $-n$KL($P||Q_b^*(\theta_*)$) 加上一个惩罚项，其中惩罚项取决于待估参数的数量；扩展模型的对数边际似然也类似。我们将在本节余下部分阐明这一事实。

根据 Chib（1995）恒等式，对于基准模型，我们有： $\forall\theta \in H_n \subset \mathbb{R}^p$,

$$\log m(w_{1:n}|\mathcal{M}_b) = \log \pi(\theta|\mathcal{M}_b) + \log \widehat{q}(w_{1:n}|\theta, \mathcal{M}_b) - \log \pi^n(\theta|w_{1:n}, \mathcal{M}_b),$$

扩展模型也同样如此。由于这一恒等式对每个 $\theta \in H_n$都成立，因此在假设 1 和假设 3 下，它对 $\theta = \theta_*$ 成立：$\log\ m(w_{1:n}|\mathcal{M}_b) = \log\ \pi(\theta_*|\mathcal{M}_b) + \log\ \widehat{q}(w_{1:n}|\theta_*,\mathcal{M}_b) -$

log $\pi^n(\theta_*|w_{1:n}, \mathbf{M}_b)$。接下来，引入局部参数 $h_\theta := \sqrt{n}(\theta - \theta_*)$ 和 $h_\psi := \sqrt{n}(\psi - \psi_\circ)$，根据随机变量变换公式，有： $\pi^n(\theta|w_{1:n}, \mathbf{M}_b) = \pi^n_{h_\theta}(\sqrt{n}(\theta - \theta_*)|w_{1:n}, \mathbf{M}_b)n^{p/2}$ 以及 $\pi^n(\psi|w_{1:n}, \mathbf{M}_e) = \pi^n_{h_\psi}(\sqrt{n}(\psi - \psi_\circ)|w_{1:n}, \mathbf{M}_e)n^{(p+d_x)/2}$，其中 $\pi^n_{h_\theta}(\cdot|w_{1:n}, \underline{\mathbf{M}}_b)$ 和 $\pi^n_{h_\psi}(\cdot|w_{1:n}, \mathbf{M}_e)$ 分别表示 h_θ和 h_ψ的后验密度。将其代入边际似然的表达式，可得： $\forall\theta \in H_n \subset \mathbb{R}^p$。

$$\begin{aligned}\log m(w_{1:n}|\mathcal{M}_b) &= \log \pi(\theta|\mathcal{M}_b) + \log \widehat{q}(w_{1:n}|\theta, \mathcal{M}_b) - \log \pi^n_{h_\theta}(\sqrt{n}(\theta - \theta_*)|w_{1:n}, \mathcal{M}_b) - \frac{p}{2}\log(n) \\ &= \log \pi(\theta_*|\mathcal{M}_b) + \log \widehat{q}(w_{1:n}|\theta_*, \mathcal{M}_b) - \log \pi^n_{h_\theta}(0|w_{1:n}, \mathcal{M}_b) - \frac{p}{2}\log(n), \quad (4.1)\end{aligned}$$

以及， $\forall\psi \in H_n \subset \mathbb{R}^{p+d_x}$,

$$\begin{aligned}\log m(w_{1:n}|\mathcal{M}_e) &= \log \pi(\psi|\mathcal{M}_e) + \log \widehat{q}(w_{1:n}|\psi, \mathcal{M}_e) \\ &\quad - \log \pi^n_{h_\psi}(\sqrt{n}(\psi - \psi_\circ)|w_{1:n}, \mathcal{M}_e) - \frac{p + d_x}{2}\log(n) \\ &= \log \pi(\psi_\circ|\mathcal{M}_e) + \log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) - \log \pi^n_{h_\psi}(0|w_{1:n}, \mathcal{M}_e) - \frac{p + d_x}{2}\log(n). \quad (4.2)\end{aligned}$$

将 θ 的后验分布表示为局部参数的后验分布，其直观意义在于：变换的雅可比行列式明确体现了模型维度所发挥的作用，而局部参数的后验分布近似服从高斯分布。定理 4.1 的 *(i)*和*(iii)* 两种情形均如此。因此，雅可比行列式通过后验集中引入了一个明确依赖于维度的惩罚项。

因此，对数边际似然可分解为两项依概率有界的项 $n \to \infty$ （在伪真值处取值的先验密度值和局部参数的后验密度值），以及两项随 n发散的项：对数 ETEL 项和一个模型维度惩罚，其阶为每个参数$_{1\ 2}$ log n 。渐近地，边际似然的表现类似于带惩罚的对数 ETEL 准则，其中惩罚是通过局部重参数化，由后验集中内生地产生的，而非临时施加的。

当然，要使基于边际似然的检验程序有效，必须证明 $\pi^n_{h_\theta}$（$\sqrt{n}$（$\theta - \theta_*$）｜$w_{1:n}$，$\mathcal{M}_b$）和 $\pi^n_{h_\psi}$（$\sqrt{n}$（ψ— $\psi_\circ$）｜$w_{1:n}$，$\mathcal{M}_e$）依 $n \to \infty$概率有界。这一要求可能相当难以验证，尤其是在本文所考虑的这类非标准情形下，因为此时不存在参数似然，且模型可能设定错误。我们在在线附录的定理F.6和F.7中证明了这些结果，这两个定理对 Chib 等人（2018）的定理 1 和 2 作了细化。

证明这些结果的关键一步是表明对数 ETEL 函数满足随机局部渐近正态性（LAN）性质。尽管伯恩斯坦–冯·米塞斯论证的其余部分遵循标准思路，但建立随机局部渐近正态性具有挑战性，因为 ETEL 函数本身是一个随机似然。在本文中，我们为对数 ETEL 函数的局部渐近正态性性质提供了一个新的、更为直接的证明（见在线附录中的定理 F.1、F.2 和 F.3）。我们的证明利用了工具变量回归问题的特定结构：由于线性，对数 ETEL 函数在 θ_* 附近的中值定理展开中的每一项都可以在紧集上关于 h_θ 更直接且一致地加以控制。这使我们能够避免使用 Chib 等人（2018）中所用的经验过程理论。

理解边际似然渐近行为的最后一步由定理4.2和4.3给出，这两条定理推导了对数 ETEL 函数在基准模型和扩展模型中的随机展开。这些展开在 Chib 等人（2018）中并未明确给出，据我们所知是新的。

我们的出发点是对数ETEL的精确主恒等式(2.12)。在 $\theta = \theta_*$ 处对该恒等式 $\hat{}$ $\hat{}$求值，并记 $\lambda(\theta_*) = \lambda_*(\theta_*) + (\lambda(\theta_*) - \lambda_*(\theta_*))$，通过对 λ 在 $\lambda_*(\theta_*)$附近作二阶泰勒展开，我们得到该展开式。随机局部渐近正态性性质给出 $\sqrt{n}(\lambda(\theta_*) - \lambda_*(\theta_*))$的线性$\hat{}$表示；将其代回主恒等式后，便得到下文所报告的二次经验过程项。 $-$

结果成立所依据的假设汇总于第 4.5 节。我们用$\mathbf{E}_n[\cdot] = n^{-1}\sum_{i=1}^n(\cdot)$表示经验均值，用 $\mathbb{G}_n f = \sqrt{n}(\mathbf{E}_n[f] - \mathbf{E}[f])$表示中心化经验过程。 $-$

定理4.2（基准模型：对数ETEL的随机展开）假设假设*1–5*和*6(d)– (f)*成立，则

$$\log \widehat{q}(w_{1:n} \mid \theta_*, \mathcal{M}_b) = -n \log n + \sum_{i=1}^{n} \log\left(\frac{\exp\{\lambda_*(\theta_*)' g_b(w_i, \theta_*)\}}{\mathbf{E}_n[\exp\{\lambda_*(\theta_*)' g_b(w_j, \theta_*)\}]} \right)$$
$$- \mathbb{G}_n\left[\tau_i^\dagger(\lambda_*, \theta_*)\, g_b(w_i, \theta_*)'\right] \Omega_*^\dagger(\theta_*)^{-1} \mathbb{G}_n[g_b(w_i, \theta_*)] + n(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))' \mathbf{E}[g_b(w_i, \theta_*)]$$
$$+ \frac{1}{2}\, \mathbb{G}_n\left[\tau_i^\dagger(\lambda_*, \theta_*)\, g_b(w_i, \theta_*)'\right] \Omega_*^\dagger(\theta_*)^{-1} \mathbb{G}_n\left[\tau_i^\dagger(\lambda_*, \theta_*)\, g_b(w_i, \theta_*)\right] + o_p(1). \quad (4.3)$$

此外，

$$\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*, \theta_*)\, g_b(w_i, \theta_*)\right] \xrightarrow{d} \mathcal{N}\left(0, \Omega_*^\dagger(\theta_*)\right),$$

其中 $\tau_i^\dagger(\lambda_*, \theta_*) := [dQ^*(\theta_*)/dP](w_i)$ 且 $\Omega_*^\dagger(\theta_*) := \mathbf{E}^{Q^*(\theta_*)}[\varepsilon_i(\theta_*)^2 \widetilde{w}_i \widetilde{w}_i']$。此外，

$$\sqrt{n}(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*)) = -\Omega_*^\dagger(\theta_*)^{-1} \mathbb{G}_n\left[\tau_i^\dagger(\lambda_*, \theta_*)\, g_b(w_i, \theta_*)\right] + o_p(1).$$

该定理对基准模型$\mathcal{M}_b$的对数ETEL 函数进行了分解，刻画了其渐近行为。利用这一分解，我们证明了贝叶斯因子检验程序的一致性。定理4.2指出，$\log \widehat{q}(w_{1:n}|\theta_*,\mathcal{M}_b) + n \log n$ 可表示为四个随机分量之和，误差项为 $o_p(1)$。当 x_i 为内生变量时，等式(4.3)右侧的第三项和第五项均为$\mathcal{O}_P(1)$阶，而第二项和第四项分别为$\mathcal{O}_P(n)$阶和$\mathcal{O}_P(\sqrt{n})$阶；当 x_i 为外生变量时，这两项均为零。根据定理的最后一部分，第四项是线性的，其阶为$\mathcal{O}_P(\sqrt{n})$；而(4.3)中的第五项是二次的。项 $n \log n$ 不影响比较，因为它与扩展模型的对数ETEL 函数中的对应项相互抵消，正如下一个定理所示。

对于扩展模型，我们回顾一下，$\psi_\circ = (\theta_\circ' \; v_\circ')'$ 表示扩展模型中满足 $v_\circ = \mathbf{E}[\varepsilon(\theta_\circ)x]$的参数真值。

定理 4.3（扩展模型。）令假设 *2*、*3*、*4* 中的 Θ替换为 Ψ，并令假设

5 和 *6 (d)-(f)* 中的 θ_* 和 $\lambda_*(\theta_*)$ 分别替换为 $\theta_\circ$ 和 0，则

$$\log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) = -n\log n - \frac{1}{2}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] + o_p(1), \tag{4.4}$$

其中 $\Omega_{\psi_\circ} := \mathbf{E}[g_e(w_i,\psi_\circ)g_e(w_i,\psi_\circ)']$，且 $\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] \xrightarrow{d} \chi_d^2$，其中 χ_d^2表示自由度为 d 的卡方分布。

定理 4.3 确立了扩展模型 $\mathcal{M}_e$的对数 ETEL 函数的渐近行为。与基准模型不同，对数 $\widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) + n\log n$ 除了一个渐近可忽略项外，等于一个依概率有界的二次随机项，其中 $n \to \infty_\circ$

若 $\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$ （外生情形），使得定理 4.2 中的假设在将 θ_*替换为 $\theta_\circ$ 且 $\lambda_*(\theta_*) = \lambda_*(\theta_\circ) = 0$时成立，则基准模型中的对数 ETEL 函数可简化为

$$\log \widehat{q}(w_{1:n}|\theta_\circ, \mathcal{M}_b) = -n\log n - \frac{1}{2}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)'\right]\Omega_\circ^{-1}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)\right] + o_p(1), \tag{4.5}$$

其中 $\Omega_\circ := \mathbf{E}[\varepsilon_i(\theta_\circ)^2\widetilde{w}_i\widetilde{w}_i']$，$\mathbb{G}_n\left[g_b(w_i,\ \theta_\circ)'\right]\Omega_\circ^{-1}\mathbb{G}_n\left[g_b(w_i,\ \theta_\circ)\right] \xrightarrow{d} \chi_d^2$，$\chi_d^2$ 表示自由度为 d 的卡方分布。对于扩展模型，若$\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$，则$g_e(w_i,\ \psi_\circ) = g_b(w_i,\ \theta_\circ)$，且对数 ETEL 函数可略作简化，如下：

$$\log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) = -n\log n - \frac{1}{2}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)'\right]\Omega_\circ^{-1}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)\right] + o_p(1), \tag{4.6}$$

其中 $\mathbb{G}_n\left[g_b(w_i,\ \psi_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_b(w_i,\ \psi_\circ)\right] \xrightarrow{d} \chi_{d\circ}^2$。因此，当 x 为外生变量时，$\log\ \widehat{q}(w_{1:n}|\theta_\circ,\ \mathcal{M}_b)$ 与 $\log\ \widehat{q}(w_{1:n}|\psi_\circ,\ \mathcal{M}_e)$ 渐近相等，并且在边际似然的比较中相互抵消。

相反，在存在内生性的情况下，$\log\ \widehat{q}(w_{1:n}|\theta_\circ,\ \mathcal{M}_b)$ 和 $\log\ \widehat{q}(w_{1:n}|\psi_\circ,\ \mathcal{M}_e)$ 并不相同，并且它们在比较边际似然时发挥着核心作用。在这种情况下，考察平均对数 ETEL 函数 $\frac{1}{n}\log\ \widehat{q}(w_{1:n}|\theta_*,\ \mathcal{M}_b)$ 的行为十分重要，因为该函数在渐近意义下保持有界。以下两个推论表明，渐近地，平均对数 ETEL 函数的表现与 $^-$

Kullback-Leibler散度相同，至多相差一项 $\log(n)$。虽然这些结果隐含在 ETEL 的定义中，我们仍在此给出正式且明确的表述，并在附录中给出证明。这使我们能够理解边际似然的行为。

推论4.1（基准模型。） 假设 *1–5* 和 *6 (d)–(f)* 成立。则当 $n \to \infty$ 时，

$$\frac{1}{n}\log \widehat{q}(w_{1:n}|\theta_*, \mathcal{M}_b) + \log(n) \xrightarrow{p} \mathbf{E}\left[\log(dQ_b^*(\theta_*)/dP)\right], \tag{4.7}$$

其中$\mathbf{E}\left[\log(dQ_b^*(\theta_*)/dP)\right] = \mathbf{E}\left[\log\left(\frac{e^{\lambda_*(\theta_*)' g_b(w,\theta_*)}}{[e^{\lambda_*(\theta_*)' g_b(w,\theta_*)}]}\right)\right] = -\mathbf{E}\mathrm{KL}(P||Q_b^*(\theta_*))$。

推论 4.2（扩展模型。） 假设 *2–4* 中的 Θ 被 Ψ 替换，*5* 和 *6 (d)–(f)* 中的 θ_* 和 $\lambda_*(\theta_*)$ 分别被 $\theta_\circ$ 和 0 替换时成立。则当 $n \to \infty$ 时，

$$\frac{1}{n}\log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) + \log(n) \xrightarrow{p} \mathbf{E}\left[\log(dQ_e^*(\psi_\circ)/dP)\right], \tag{4.8}$$

其中$\mathbf{E}\left[\log(dQ_e^*(\psi_\circ)/dP)\right] = \mathbf{E}\left[\log\left(\frac{e^{\lambda_*(\psi_\circ)' g(w,\psi_\circ)}}{[e^{\lambda_*(\psi_\circ)' g(w,\psi_\circ)}]}\right)\right] = \mathbf{E}\mathrm{KL}(P||Q_e^*(\psi_\circ))$。

注意，由于扩展模型设定正确，因此$\mathbf{E}\left[\log(dQ_e^*(\psi_\circ)/dP)\right] = 0\, dQ_e^*(\psi_\circ)/dP = 1$。

根据定理4.2和4.3、在线附录中的定理F.6和F.7，以及（4.1）—（4.2），存在一个 N，使得对每个 $n > N$:

$$\begin{aligned}
\log m(w_{1:n}|\mathcal{M}_b) &= -n\log n + \sum_{i=1}^{n} \log\left(\frac{e^{\lambda_*(\theta_*)' g_b(w_i,\theta_*)}}{\mathbf{E}_n[e^{\lambda_*(\theta_*)' g_b(w_j,\theta_*)}]}\right) \\
&\quad + n(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))' \mathbf{E}[g_b(w_i,\theta_*)] - \frac{p}{2}\log(n) + \mathcal{O}_p(1), \qquad (4.9) \\
\log m(w_{1:n}|\mathcal{M}_e) &= -n\log(n) - \frac{p+d_x}{2}\log(n) + o_p(1). \qquad (4.10)
\end{aligned}$$

这些对数边际似然量化了模型的整体有效性。事实上，从这些表达式可以看出，当 x_i 外生时，即 $\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$，则 $\lambda_*(\theta_*) = 0$并且

$\sum_{i=1}^{n} \log \left(\frac{e^{\lambda_*(\theta_*)' g_b(w_i,\theta_*)}}{n[e^{\lambda_*(\theta_*)' g_b(w_j,\theta_*)}]} \right) = 0$ **E**对于每个 $n \in \mathbb{N}$均成立。因此，很明显，渐近地，$\log m(w_{1:n}|\mathcal{M}_b)$ 大于 $\log m(w_{1:n}|\mathcal{M}_e)$。

另一方面，当不存在 $\theta \in \Theta$使得 $\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$时，则 $\lambda_*(\theta_*) \neq 0$ 和$\sum_{i=1}^{n} \log \left(\frac{e^{\lambda_*(\theta_*)' g_b(w_i,\theta_*)}}{n[e^{\lambda_*(\theta_*)' g_b(w_j,\theta_*)}]} \right) - n$ **E**$\log\ n$ 的发散速度 $-\infty$ 快于（4.9）中的后两项，因此渐近地，$\log\ m(w_{1:n}|\mathcal{M}_b)$ 小于 $\log m(w_{1:n}|\mathcal{M}_e)$。这就是下一节定理 4.4 和 4.5 中一致性结果的主要直觉。这些定理的证明见附录，其推导比上述论证更复杂，因为这些定理给出的是“当且仅当”命题，这比一致性更强。

4.3 检验程序的一致性

现在，我们运用上述理论，证明基于边际指数倾斜经验似然函数所构造的贝叶斯因子的检验程序具有一致性。下面的定理证明，随着样本量增加， BF_{eb} 当且仅当 x 为外生时选择 $\mathcal{M}_b$ ，而当且仅当 x 为内生时选择$\mathcal{M}_e$ ，且这两种情况发生的概率都趋近于 1。

定理 4.4 设假设*1—3*、*5*、*8*、*9*成立，并且假设*4*和*6*对于 θ_* 和$\lambda_*(\theta_*)$ 成立；同时，对于 θ_* 和 $\lambda_*(\theta_*)$ ，分别以 $\theta_\circ$和0替代。设 θ 和 ψ的先验分布为连续概率测度，其相对于勒贝格测度的密度存在，且分别在 θ_* 和 $\psi_\circ$的某个邻域内为正。考虑模型 M_b 和 M_e的比较。则，

$$\lim_{n\to\infty} P\left(\log m(w_{1:n}|\mathcal{M}_e) > \log m(w_{1:n}|\mathcal{M}_b)\right) = 1$$

当且仅当　不存在满足 $\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$ 的 $\theta \in \Theta$ ，否则极限为零　　*e.*

正如我们在证明中所示，对于任意 $\theta \in \Theta$，必要且充分条件$\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$ 不成立，等价于不等式 KL$(P||Q_e^*(\psi))$ $<$KL$(P||Q_b^*(\theta))$，其中 KL$(P||Q_e^*(\psi_\circ))$ $=$0。在这种情况下，对数 ETEL 函数的主导作用超过对数边际 ETEL

函数的其他组成部分，因此内置惩罚不起任何作用。因此，正如 Chib 等人 (2018, 定理 3.2) 针对矩条件模型给出的一般结果所示，通过比较基准模型和扩展模型的对数边际似然，并选择数值较高者，在极限情况下会选出与真实模型的 KL 散度最近的模型。在本文的框架下，这意味着比较边际似然能够正确得出 x_i是内生的结论。

接下来，我们考察变量 x_i 外生时会发生什么。在这种情况下，对于某个特定值 $\theta_\circ$，矩条件 $\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$ 成立，且所比较的两个模型均设定正确。下一个定理指出，此时会选择基准模型。可以从简约性的角度理解这一点：基准模型需要估计的参数较少，因此在其设定正确时更受青睐。

定理4.5 设假设 *1 - 3*、*5*、*8* 和 *9* 成立，并且对于 θ_*和 $\lambda_*(\theta_*)$，以及将 θ_* 和 $\lambda_*(\theta_*)$ 分别替换为 $\theta_\circ$ 和 0 后的情形，假设*4*和*6*均成立。设 θ 和ψ 的先验分布为连续概率测度，其相对于勒贝格测度的密度在 θ_* 和 $\psi_\circ$的某个邻域内分别为正。考虑模型 M_b 和 M_e的比较。则，

$$\lim_{n\to\infty} P\left(\log m(w_{1:n}|\mathcal{M}_b) > \log m(w_{1:n}|\mathcal{M}_e)\right) = 1$$

当且仅当存在一个 $\theta_\circ \in \Theta$，使得 $\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$成立。

当 x 外生时，如前一定理所示，对数 ETEL 函数和内置惩罚均在正确模型的选择中发挥作用。

讨论。 在本小节和上一小节中，我们证明了模型选择准则倾向于选择 Kullback-Leibler 信息准则（KLIC）较小的模型。当两个模型的 KLIC 相同时，我们的程序会选择过度识别约束数量较多的模型，即更简约、灵活性更低的模型。有趣的是，这与目标

一致，即 Sin 和 White（1996年）针对参数模型提出的惩罚似然准则的目标。因此，本文提出的模型选择程序以及 Chib 等人（2018）的程序，可以视为一致模型选择准则的完全贝叶斯半参数版本，专门用于内生性检验问题。与其他频率学派程序不同，保证一致性所需的“惩罚”项已内在地融入我们的贝叶斯计算中。Chib 等人（2018）并未强调这一点，而这正是本文的一项贡献。

Andrews（1999）、Andrews 和 Lu（2001），以及 Hong、Preston 和 Shum（2003）提出并研究了矩条件模型的模型选择准则，尽管这类模型并未定义正式的似然函数。这些准则并非针对似然函数，而是在广义矩估计法（GMM）以及更广义的广义经验似然（GEL）目标函数上附加惩罚项。在线附录 B 中可以找到基于 GMM估计的此类频率学派模型选择方法示例。然而，这些模型选择准则与 Sin 和 White（1996）针对可能设定错误的参数模型提出的 KLIC 最小化原理之间的关系并非一目了然。

值得注意的是，我们的方法展现出与 Hong 和 Preston（2012）的广义经验似然贝叶斯因子相同的渐近行为。他们为拉格朗日乘子单独设定了一个独立于 θ 的先验，但这并不能保证施加矩约束。相比之下，我们为“非活跃”矩约束引入了一个额外参数 v，从而确保我们对 θ 和 v 的先验符合矩约束。

4.4 两个以上模型之间的检验

我们的检验程序可以扩展到比较两个以上模型的情形。考虑 x 中仅有一部分变量是内生的情形。首先，假设$d_x = 2$，并且仅 x_1 是内生的，而 x_2 是外生的。也就是说，$\mathbf{E}[\varepsilon(\theta_\circ)(x_2, z_1')'] = 0$，而不存在 $\theta \in \mathbb{R}^p$ 使得$\mathbf{E}[\varepsilon(\theta)x_1] = 0$。如果我们只比较基准模型和扩展模型，$\mathcal{M}_b$ 和$\mathcal{M}_e$，我们可能会错误地得出 x_1 和 x_2 都是内生的结论。相反，更合适的做法是考虑以下模型：$\mathcal{M}_b$、$\mathcal{M}_e$、$\mathcal{M}_{e_1}$ 和$\mathcal{M}_{e_2}$，其中，对于

$i = 1,2$，$\mathcal{M}_{e_i}$ 是由矩条件定义的模型

$$\mathbf{E}^{Q}[g_{e_i}(w,\psi)] = 0, \qquad Q \in \mathcal{Q}_{e_i}(\psi), \tag{4.11}$$

其中

$$g_{e_i}(w,\psi_i) := \varepsilon(\theta)\begin{pmatrix} x \\ z_1 \\ z_2 \end{pmatrix} - \begin{pmatrix} v^{(i)} \\ 0 \\ 0 \end{pmatrix} = g_b(w,\theta) - \begin{pmatrix} v^{(i)} \\ 0 \\ 0 \end{pmatrix},$$

$\psi_i := (\theta, v^{(i)}) \in \Psi, \Psi := \Theta \times \mathbf{V}$，$\mathbf{V} \subset \mathbb{R}^2$ 满足 $v^{(i)} = (v_1^{(i)}, v_2^{(i)})$，且 $v_j^{(i)} \in \mathbb{R}$ 对于 $j = 1$、2 和 $\mathbf{Q}_{e_i}(\psi_i) := \left\{ Q \in \mathbb{M};\ \mathbf{E}^{Q}[g_{e_i}(w,\psi_i)] = 0 \right\}$。我们通过将 x的一个分量视为外生变量，并将 $v^{(i)}$ 中对应的元素设为零，来施加这一限制。具体而言，定义$v^{(1)} := (0, v_2^{(1)})$ 和 $v^{(2)} := (v_1^{(2)}, 0)$。模型 $\mathbf{M}_{e_1}$ 将 x_1 视为外生变量，并允许 x_2 为内生变量；而模型 $\mathbf{M}_{e_2}$ 将 x_2 视为外生变量，并允许 x_1 为内生变量。

这一构造使我们能够直接应用基准模型与扩展模型比较。将$\mathcal{M}_e$ 视为共同的参照扩展模型，并将$\mathcal{M}_{e_i}$ 与$\mathcal{M}_e$进行比较。若$\mathcal{M}_{e_2}$ 的边际似然超过$\mathcal{M}_e$的边际似然，则数据支持约束 $v_2 = 0$，表明x_2 是外生的，而 x_1 被视为内生。类似地，若$\mathcal{M}_{e_1}$的边际似然超过$\mathcal{M}_e$的边际似然，则数据支持 $v_1 = 0$，表明 x_1 是外生的，而 x_2 被视为内生。

更一般地，当考虑的模型多于两个时，可以通过它们的边际似然进行比较。在我们的情形中，这些候选模型是通过将 v 的一部分元素设为零而从扩展模型得到的。总共有 2^{d_x} 个模型，包括基准模型和扩展模型。每个模型对应于 x 中哪些元素被视为内生的一种配置：当相应的 v_i 不受约束时，x_i 被视为内生；而 $v_i = 0$ 则对应于 x_i 被视为外生。在这一有限模型集合上的边际似然比较会选出数据最支持的配置。

此外，我们的 通过比较不同的模型设定，可以丰富内生性检验 -

设。举例来说，假设 x是标量，并考虑线性设定与二次设定，

$$y_i = \gamma_0 + \beta_1 x_i + \beta_2 x_i^2 + \gamma_1 z_{1i} + \varepsilon_i.$$

然后，可以在每种设定下评估 x_i 的内生性，从而得到四个候选模型：线性-外生、线性-内生、二次-外生、二次-内生。边际似然比较可用于在这些候选模型中选出最佳模型。我们将这一思路应用于真实数据示例（BLP模型）。在该情形下，我们考虑四个同时改变函数形式与价格内生性状态的候选模型，即线性-外生、线性-内生、非线性-外生、非线性-内生。对这四个候选模型进行边际似然比较，可同时评估每种设定内的内生性，并给出跨设定的统一排序。

附录C报告了这两种应用情形的蒙特卡洛实验；参见C.1节和C.2节。

4.5 假设

我们给出用于证明前述各节结果的假设。第一条假设确保满足矩条件的分布集合非空，这对于指数倾斜经验似然的良好定义是必要的。它保证即使当 $P \notin \mathcal{Q}_{b,\theta}$ 对每个 $\theta \in \Theta$成立时，优化问题(2.5)的对偶表示仍然成立。事实上，在后一种情形下，$Q_b^*(\theta)$与 P 可能对任何 θ都不具有共同支撑集，此时(2.7)中的等式不成立；关于这一点的讨论参见Sueishi（2013）。

假设1（非空性。）当对每个 $\theta \in \Theta$，$\mathrm{E}[\varepsilon_i(\theta) x_i] \neq 0$ 时，存在 $Q \in \bigcup_{\theta \in \Theta} \mathrm{Q}_{b,\theta}$，使得 Q 相对于 P 相互绝对连续，其中 $\mathrm{Q}_{b,\theta}$ 在（3.3）中定义。该假设意味着存在一个 θ，使得 $\mathrm{Q}_{b,\theta}$ 非空，$dQ_b^*(\theta)/dP = \begin{pmatrix} e^{\lambda_*(\theta)' g(w,\theta)} \\ [e^{\lambda_*(\theta)' g(w,\theta)}] \end{pmatrix} \mathrm{E}$，并且 θ_* 可由（3.4）识别。此外，假设相互绝对连续，可确保 $\mathrm{KL}(Q_b^*(\theta)||P)$ 和 $\mathrm{KL}(P||Q_b^*(\theta))$ 都有良好定义。然后我们假设 θ_* 是唯一的。

假设 2（识别。）最大化点 θ_* 定义为使 $\mathrm{KL}(P||Q^*(\theta))$ 最小化的点。相对于 $\theta \in \Theta$，该点唯一且位于 Θ 的内部，其中内部是相对于 $\mathbb{R}^p$ 中的拓扑定义的。

由于在假设1下， θ_* 与(3.4)中的最小化点重合，因此前述假设也意味着后者的唯一性。

正如我们在第 2 节中所指出的，指数倾斜经验似然在使优化问题(2.5)不存在可行解的那些 θ处没有定义，也就是说，在不属于(2.6)中集合 H_n的那些 θ处没有定义。假设1保证优化问题(2.5)在θ_*处渐近可行。类似地，若模型设定正确，则优化问题(2.5)在$\theta_\circ$处可行（或在 $\psi_\circ$处可行，取决于我们所考虑的模型）。然而，这对于我们的渐近分析而言还不够。相反，我们需要假设：对于每一个充分接近 θ_* 的 θ ，(2.5)都以趋于1的概率存在解。所需的接近程度在下一个假设中加以规定，为此我们引入如下的球。对于任意序列 $M_n \to \infty$ （当 $n \to \infty$时）以及任意 $\widetilde{\theta} \in \Theta$，将围绕 $\widetilde{\theta}$ 的 $n^{-1/2}$-球定义为 $B(\widetilde{\theta}, M_n n^{-1/2}) := \{\theta \in \Theta : \ \|\theta - \widetilde{\theta}\| \leq M_n n^{-1/2}\}$。球 $B(\widetilde{\theta}, M_n n^{-1/2})$收缩至0的速度略慢于 $n^{-1/2}$ ，具体取决于 M_n 趋于 ∞的速度。围绕某个 $\widetilde{\psi} \in \Psi$ 的 $n^{-1/2}$-球 $B(\widetilde{\psi}, M_n n^{-1/2})$的定义与此类似。此外，用 $int\Delta_n := \{q := (q_1, \ldots, q_n)' ; \ \sum_{i=1}^n q_i = 1, \ q_1 > 0, \ldots, q_n > 0\}$表示$n-1$ 单纯形的内部。最后，对于任意序列 $M_n \to \infty$: $B_{*, \ n} := \bigcap_{\{M_n; M_n \to \infty\}} B(\theta_*, M_n n^{-1/2})$以及$B_{\circ, \ n} := \bigcap_{\{M_n; M_n \to \infty\}} B(\psi_\circ, M_n n^{-1/2})$。

假设 3 作为 $n \to \infty$，对于每个$\theta \in B_{*,n}$:

$$P\left(\sum_{i=1}^n g_b(w_i, \theta) q_i = 0 \textit{ for at least one } q \in int\Delta_n\right) \to 1 \tag{4.12}$$

且$\mathbf{E}^{Q_*}[g_b(w, \theta)] = 0$。类似地，当 $n \to \infty$ 且对每一个$\psi \in B_{\circ,n}$:

$$P\left(\sum_{i=1}^n g_e(w_i, \psi) q_i = 0 \textit{ for at least one } q \in int\Delta_n\right) \to 1 \tag{4.13}$$

且 $\mathbf{E}[g_e(w,\psi)]=0$。

该假设远弱于要求2.5对每一个 $\theta\in\Theta$ 都以趋近于1的概率存在解。即便文献中并未明确陈述，频率学派 ETEL、经验似然和指数倾斜估计量也需要类似的假设。

接下来的四个假设涉及模型。与 Chib 等人（2018）的假设相比，由于矩函数 $g_b(w,\theta)$ 和 $g_e(w,\theta)$ 关于 θ 是线性的，我们的假设较弱。因此，关于矩函数及其导数连续性的假设自动满足。回顾记号 $w_i:=(y_i,x_i',z_i')'$, $z_i:=(z_{1,i}',z_{2,i}')'$ 以及$\tilde{w}_{1,i}:=(x_i',z_{1,i}')'$。此外，我们记 $\tilde{w}_i:=(x_i',z_i')'$，$\|\cdot\|_2$ 表示欧几里得范数，$\|\cdot\|_{op}$ 表示算子范数。

假设 4 *(a)* w_i, $i=1,\dots,n$ 是独立同分布的可观测随机变量，每个变量的取值均属于完备概率空间 $(W,\mathfrak{B}_W,P)$，其中 $W\subseteq\mathbb{R}^{d+1}$, $\mathfrak{B}_W$是相应的σ-域，且 P 是满足模型 (3.1)的概率分布；*(b)* $\Theta\subset\mathbb{R}^p$ 紧致且连通；*(c)* 对于 $\lambda_*(\theta_*)$的某个邻域内的每个 λ，矩阵 $\mathbf{E}[e^{\lambda'\tilde{w}_i\varepsilon_i(\theta_*)}\varepsilon_i(\theta_*)^2\tilde{w}_i\tilde{w}_i']$ 的最小（分别为最大）特征值与零（分别与无穷大）保持正距离。

假设 5 *(a)* $\mathbf{E}[\tilde{w}_i\tilde{w}_{1,i}']<\infty$，秩为 p。

假设 4 和 5 在文献中是标准的，参见例如 Schennach（2007）。而以下假设是新的，用于证明边际似然的近似结果。我们用 $\tilde{w}_{i,k}$ 表示 $\tilde{w}_i$ 的第 k 个分量。此外，对于任意 $\delta>0$ 以及某个常数 $C>0$，我们将 $B_\delta(\lambda_*):=\{\lambda\in\mathbb{R}^d;\|\lambda-\lambda_*(\theta_*)\|_2\le C\delta\}$（分别将 $B_\delta(\theta_*):=\{\theta\in\mathbb{R}^p;\|\theta-\theta_*\|_2\le C\delta\}$）表示为 $\mathbb{R}^d$（分别为 $\mathbb{R}^p$）中的一个闭球，其中心为 $\lambda_*:=\lambda_*(\theta_*)$（分别为 θ_*），半径为 δ，其中 $\|\cdot\|_2$ 表示欧几里得范数。对于任意 $\delta>0$、$q\in\mathbb{N}_+$、集合 B 和函数 $\gamma:\mathbb{R}^{d+1}\times\mathbb{R}^p\to\mathbb{R}$，我们引入以下函数集：$\mathrm{F}_{q,\delta}(B;\gamma):=\{f:B\times\mathbb{R}^{d+1}\times B_\delta(\theta_*)\to\mathbb{R}^q;\ \|g(u,v,\theta)\|_2\le\gamma(v,\theta)\ \forall u\in B\}$。对于随机向量 w，我们用 w_k 表示其第 k 个分量。最后，$K_n\subset\mathbb{R}^p$ 表示一个以 0 为中心的紧集。

假设 6 对于任意 $\delta > 0$，存在定义在 $\mathbb{R}^{d+1} \times B_\delta(\theta_*)$ 上的实值函数 $\gamma_j(w,\theta)$、$j = 0$、1、2、3、4、5，并且满足 $\mathbf{E}[\gamma_j(w,\theta)] < \infty$，其中 $j = 0$、1、2、3、4、5 以及 $\forall \theta \in B_\delta(\theta_*)$，且满足：

(a) 函数 $(\lambda, w, \theta) \mapsto e^{\lambda' \widetilde{w}\varepsilon(\theta)}\, \widetilde{w}\varepsilon(\theta) \in F_{d,\delta}(B_\delta(\lambda_*); \gamma_0)$; *(b)* 函数 $(\lambda, w, \theta) \mapsto e^{\lambda' \widetilde{w}\varepsilon_i(\theta)} \in F_{1,\delta}(B_\delta(\lambda_*); \gamma_1)$; *(c)* 函数 $((\lambda', h')', w, \theta) \mapsto e^{\lambda' \widetilde{w}\varepsilon(\theta_*)} \varepsilon(\theta)^{j-1} \widetilde{w}_k^{\ell} (h'\widetilde{w}_1)^{\ell'} \widetilde{w}_k^{\ell''} \in F_{1,\delta}(B_\delta(\lambda_*) \times K_n; \gamma_2)$，对于 $(j,\ell,\ell',\ell'') = (1,1,1,0)$、$(j,\ell,\ell',\ell'') = (1,2,2,0)$ 或 $(j,\ell,\ell',\ell'') = (2,1,1,1)$ 三者之一；*(d)* 函数 $(\lambda, w, \theta) \mapsto e^{2\lambda' \widetilde{w}\varepsilon(\theta_*)} \varepsilon(\theta_*)^2 \left\|\widetilde{w}\right\|_2^2 \in F_{1,\delta}(B_\delta(\lambda_*); \gamma_3)$; *(e)* 函数 $(\lambda, w, \theta) \mapsto e^{\ell \lambda' g(w,\theta)}\, \widetilde{w}_j \widetilde{w}_k \varepsilon(\theta)^{\ell'} \widetilde{w}_{k'}^{i} \in F_{1,\delta}(B_\delta(\lambda_*); \gamma_4)$，对于每一个 $j, k, k' = 1, \ldots, d$，并且对于 $(\ell,\ell',i) = (1,1,1)$、$(\ell,\ell',i) = (1,2,0)$、$(\ell,\ell',i) = (1,3,1)$ 或 $(\ell,\ell',i) = (2,4,2)$ 四者之一，其中 γ_4 可以依赖于所有先前的指标；*(f)* 函数 $((\lambda', h')', w, \theta) \mapsto e^{\lambda' \widetilde{w}\varepsilon(\theta_*)} \varepsilon(\theta)\, \widetilde{w}\, \lambda' \widetilde{w}\, h' \widetilde{w}_1 \in F_{d,\delta}(B_\delta(\lambda_*) \times K_n; \gamma_5)$; *(g)* 函数 $((\lambda', h')', w, \theta) \mapsto e^{\lambda' \widetilde{w}\varepsilon(\theta_*)} \varepsilon(\theta)^j \widetilde{w}\, \lambda' \widetilde{w}\, (h'\widetilde{w}_1)^{\ell} \widetilde{w}_k \in F_{d,\delta}(B_\delta(\lambda_*) \times K_n; \gamma_6)$，对于每一个 $k = 1, \ldots, d$，并且对于 $(j,\ell) = (2,1)$ 或 $(j,\ell) = (1,2)$ 二者之一。

这一正则性假设是一项弱矩条件，要求对每个 $\theta \in B_\delta(\theta_*)$，都存在一个依赖于数据和 θ 的上界，且该上界的期望有限。该条件确保一致大数定律成立（例如，参见 Newey 和 McFadden （1994，引理 2.4））。具体而言，在证明中，我们利用这一假设来确立若干项依概率一致收敛，包括：$\mathrm{E}_n\left[e^{\widehat{\lambda}' g(w_i,\theta_*)} \varepsilon_i(\theta_*)\, \widetilde{w}_i\right]$、$\mathrm{E}_n\left[e^{\widehat{\lambda}' g(w_i,\theta_*)}\, \widetilde{w}_{1,i} \widetilde{w}_i\right]$，$h'\mathrm{E}_n\left[e^{\widehat{\lambda}' g(w_i,\theta_*)}\, h'\widetilde{w}_{1,i} \widetilde{w}_i' \widehat{\lambda} \varepsilon_i(\theta_*)\, \widetilde{w}_i\right]$，以及 $\mathrm{E}_n\left[e^{\widehat{\lambda}' g(w_i,\theta_*)} \varepsilon_i(\theta_*)\, \widetilde{w}_{1,i} \widetilde{w}_i\right]$。一致收敛^必须在闭球 $B_\delta(\lambda_*(\theta_*))$ 内对 λ 一致成立。此外，这一假设还保证满足应用控制收敛定理所需的支配条件；该定理也是证明中用到的结果。

前述假设的第*(e)* 部分意味着

$$\mathbf{E}\left[\sup_{\lambda \in B_\delta(\lambda_*(\theta_*))} \left\| e^{\ell \lambda' \widetilde{w}\varepsilon(\theta_*)} \varepsilon(\theta_*)^{\ell'} \widetilde{w}_k^{i} \widetilde{w} \widetilde{w}' \right\|_{op}\right]$$

对于每个 $k' = \{1 \ldots, d\}$，其值均有界，不会趋于无穷大；这正是证明所需的，因为该算子范数的上界为 $d \max_{\{j,k \in \mathbf{1}_{\{,\}} \ldots, d\}} |e^{\ell \lambda' g}{}_{(w,\theta_*)}\, \widetilde{w}_j \widetilde{w}_k \varepsilon(\theta_*)^{\ell} \widetilde{w}^i_{k'}|$，对每个 $k' \in \{1, \ldots, d\}$均成立。

要证明$\mathbf{E}_n [\varepsilon_i(\theta)\, \widetilde{w}_i]$ 和$\mathbf{E}_n [\varepsilon_i(\theta)\, \widetilde{w}_i \widetilde{w_i}]$ 在 $\theta \in B_\delta(\theta_*)$ 上一致收敛，必须作出类似于假设 6 的假设。为此，我们引入以下函数类

假设 7 对于任意 $\delta > 0$，存在定义在 $\mathbb{R}^{d+1}$上的实值函数 $\gamma_{\circ,\, j}(w)$、 $j = 1,2$，使得$\mathbf{E}[\gamma_j(w)] < \infty$，对于 $j = 1,2$ 以及：

(a) $\|\widetilde{w}\varepsilon(\theta)\| \leq \gamma_{\circ,\, 1}$ 对每个 $\theta \in B_\delta(\theta_*)$均成立；***(b)*** $|\varepsilon(\theta)^{\widetilde{w}_j \widetilde{w}_k}| \leq \gamma_{\circ,\, 2}$ 对每个 $\theta \in B_\delta(\theta_*)$均成立；***(c)*** $\|h' \widetilde{w}_1 \widetilde{w}'\| \leq \gamma_{\circ,\, 3}$ 对每个 $h \in K_n$均成立。

对于下一个假设，我们用 $\Theta_n := \{\|\theta - \theta_*\| \leq M_n/\sqrt{n}\}$ 表示以 θ_* 为中心、半径不超过 $M_n/\sqrt{n}$的球，其中 M_n 是任意一个趋于发散的正数序列。我们用 $\ell_{n,\theta}(w_i)$ 表示单个观测值 w_i: ℓ_n 的对数似然函数，即 $\ell_\theta(w_i) := \log \widehat{q}_i(\theta|\mathbf{M}_b)$；用 $\ell_{n,\theta}(w_{1:n}) := \sum_{i=1}^n \ell_{n,\theta}(w_i) = \log \widehat{q}(w_{1:n}|\theta, \mathbf{M}_b)$ 表示对数 ETEL 函数。注意，$\ell_{n,\theta}(w_i)$ 和 $\ell_{n,\theta}(w_{1:n})$ 都仅在 $\theta \in H_n$ 上有定义，而根据假设 3，它们至少在 $B_{*,n}$上有定义。下一个假设约束了 ETEL 函数 $\theta \mapsto \ell_{n,\theta}(w_i)$ 在远离 θ_* 处的行为，并确保 θ_* 与距离它一定距离的 θs 之间有明确的分离。

As假设 8（基准模型。）假设存在一个常数 $C > 0$，使得 *t*

$$P\left(\sup_{\theta \in H_n \cap \Theta_n^c} \frac{1}{n} \sum_{i=1}^n (\ell_{n,\theta}(w_i) - \ell_{n,\theta_*}(w_i)) \leq -\frac{C M_n^2}{n}\right) \to 1 \text{ , } as\ n \to \infty, \tag{4.14}$$

其中 M_n 是用于定义 Θ_n的同一序列。

与假设 8 类似的条件见 Kleijn 和 van der Vaart（2012，引理 4.2），它也与经典条件有关，例如 Lehmann 和 Casella（1998，假设 6.B.3）以及 Chernozhukov 和 Hong（2003，假设 3）中的条件。然而，在我们的情形中，假设中的上确界是在一个更小的集合上取值的，

该集合为 $H_n \cap \Theta_n^c$，而不是上述文献中的 Θ_n^c。为更好地理解这一假设的含义，请注意，从渐近角度看，对数 ETEL 函数在伪真值 θ_* 处达到最大值。因此，假设（4.14）要求：如果参数 θ 远离伪真值 θ_*，即 $\|\theta - \theta_*\| > M_n/\sqrt{n}$，那么在这样的 θ 处求值的和 $\sum_{i=1}^n \ell_{n,\,\theta}(w_i)$ 必须相对于和 $\sum_{i=1}^n \ell_{n,\,\theta_*}(w_i)$（即最大值）足够小。控制这种行为很重要，因为后验分布涉及对 θ 的整个支撑集进行积分。基于 ETEL 函数，可以一致地区分于θ_* 的 Θ 子集（随着 $n \to \infty$，概率趋近于 1）将获得渐近可忽略的后验概率。该条件的一个替代方案是要求存在渐近一致检验 ϕ_n，能够以一致的方式与真实分布P区分开来；也就是说，对于每个 $\epsilon > 0$，都存在一列检验$\{\phi_n\}$，使得当 $n \to$趋于 0 时，

$$\mathbf{E}[\phi_n] \to 0, \qquad \text{and} \qquad \sup_{\{\theta;\|\theta-\theta_*\|\geq\epsilon\}} \mathbf{E}\left[e^{\ell_{n,\theta}(w_i)-\ell_{n,\theta_*}(w_i)}(1-\phi_n)\right] \to 0. \tag{4.15}$$

类似地，对于扩展模型，我们将单个观测值 w_i 的对数似然函数记为 $\ell_{n,\psi}(w_i)$：$\ell_{n,\psi}(w_i) := \log\, \widehat{q}_i(\psi|\mathcal{M}_e)$；将对数 ETEL 函数记为 $\ell_{n,\psi}(w_{1:n}) := \sum_{i=1}^n \ell_{n,\psi}(w_i) = \log\, \widehat{q}(w_{1:n}|\psi,\mathcal{M}_e)$。下一个假设的含义与假设 8 相同，但适用于扩展模型。

假设 9（扩展模型。）假设存在常数 $C > 0$，使得当 $n \to \infty$ 时，

$$P\left(\sup_{\substack{\|\psi-\psi_\circ\|>M_n/\sqrt{n};\\ \psi\in H_n\times\mathcal{V}}} \frac{1}{n}\sum_{i=1}^n \left(\ell_{n,\psi}(w_i) - \ell_{n,\psi_\circ}(w_i)\right) \leq -\frac{CM_n^2}{n}\right) \to 1, \tag{4.16}$$

其中，M_n 是任意一个趋于无穷大的正数序列。

4.6 实验

考虑与第3.4节相同的生成过程，并假设 $(\varepsilon_i,\ u_i,\ v_i)$ 的联合分布由协方差矩阵为 $R = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ 的高斯 Copula 诱导。参数

ρ 控制内生性的程度。我们令 ρ 取值于从 -.5 到 .5、步长为 0.1 的集合。对于该集合中的每个 ρ 值，我们生成 100 个样本量为 n 的样本。对于每个样本，我们计算基准模型和扩展模型，并计算对数边际似然。随后，我们统计$\mathcal{M}_e$ 的对数边际似然超过$\mathcal{M}_b$的次数。结果见表1。从表中可以看出，即使 ρ的值很小，我们的内生性检验也能正确判断正确模型为$\mathcal{M}_e$。

ρ	-0.5	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.3	0.4	0.5
$n = 250$	99	96	82	48	12	2	18	54	93	100	100
$n = 500$	100	100	98	76	17	1	29	87	99	100	100
$n = 1000$	100	100	100	96	46	1	46	100	100	100	100
$n = 2000$	100	100	100	100	80	1	70	100	100	100	100

表1：根据示例1中的设计模拟生成的数据，进行100次重复所得的模型选择频率。对于 n 和 Cov(ε, u) $= \rho$的每种组合，表中数值表示在100次数据重复中，模型$\mathcal{M}_e$ 的对数边际似然超过模型$\mathcal{M}_b$的对数边际似然的次数。

5 真实数据示例

5.1 价格对汽车需求的因果效应

我们考察 Berry, Levinsohn and Pakes (1995) 研究的汽车需求经典问题。Chernozhukov, Hansen and Spindler (2015) 最近重新研究了这一问题，以下分别简称为 BLP 和 CHS。除了问题本身的价值外，值得分析这一问题的原因还在于它涉及数量较多且符合现实情况的控制变量和工具变量。

为设定问题，令 y_{ijt} 表示市场 j 在 t时产品 i相对于外部选项的市场份额之比的对数，并令 x_{ijt} 表示可能存在内生性的汽车价格变量。在样本数据中，该变量已去均值。对于控制变量，令 z_{ijt} 表示产品的可观测特征。在 BLP 中，这些特征被设定为常数、空调虚拟变量（ y_{ijt} ）、马力重量比（ i）、每美元行驶里程（ j ）和车辆尺寸（ t）。在我们的记号中，

其中 $y_{ijt} = x_{ijt}\beta + z'_{1ijt}\gamma + \varepsilon_i$， z_{1ijt} =(1， mpd_{ijt}， $space_{ijt}$， $hpwt_{ijt}$， air_{ijt})。BLP 使用了十个工具变量，其中五个是将同一企业生产的其他汽车的这五项特征分别求和得到的，另外五个是将其他企业生产的汽车的上述特征分别求和得到的。这些构成了 z_{2ijt}。在重新审视这一分析时，CHS 在原有控制变量的基础上，加入了 $trend$、mpd、 $space$和 $hpwt$的二次项和三次项，以及所有一阶交互项，随后使用这些特征之和作为潜在工具变量。

在我们的分析中，我们同时考虑这两种表述；但在扩展版本中，我们通过自然三次样条基函数分别对 $trend$、 $hpwt$、 mpd 和 $space$ 进行变换，从而引入非线性控制变量。每个基函数均以五个等距分布的分位数节点为中心（这些三次样条基函数取自 Chib 和 Greenberg（2010））。我们采用这种方法，是为了避免这些协变量的参数二次项和三次项导致协变量取值差异过大。施加基函数展开的识别约束后，每个连续协变量对应的非线性项数量减少至四个；此时，扩展结果模型右边定义为 x（价格）和 z_1 （由一个截距、十六个非线性协变量以及空调虚拟变量构成）。对于，...，4，这些非线性协变量分别记为 $trend_Bj$、 mpd_{Bj}、 $space_{Bj}$ 和 $hpwt_{Bj}$。随后，按照 BLP 的方法构造扩展工具变量集合，该集合在此扩展模型中构成 z_2。

我们针对这些数据拟合了四个模型：采用BLP中的控制变量和工具变量的基准模型和扩展模型，以及采用增广的控制变量和工具变量集合的基准模型和扩展模型。在BLP版本中，基准模型和扩展模型分别包含六个和七个参数以及十个工具变量；而在增广变体中，基准模型和扩展模型分别具有十九个和二十个参数以及53个矩约束。我们假设关于（y_{ijt}, x_{ijt}, z_{1ijt}）的$n = 2217$ 个观测值是来自跨市场和时间的汽车产品总体的随机样本。由于难以通过先验考量来设定参数的先验分布，我们随机选取样本的15%来构建训练样本先验。具体而言，我们使用在训练数据上拟合的GMM估计值及其标准误（按模型

	原始 BLP（线性）	扩展 BLP（非线性）
基准模型（价格外生）	-14386.81	-14431.86
扩展模型（价格内生）	-14364.59	-14397.67

表2：所提出的内生性贝叶斯检验结果。原始 BLP 模型及其扩展版本下，基准模型和扩展模型的对数边际似然。结果基于训练样本先验（随机选取 15% 的数据）以及采用定制单区块M-H算法进行的 10,000 次马尔可夫链蒙特卡罗迭代（预烧期为 1,000 次）。边际似然的对数采用 Chib（1995）以及 Chib和Jeliazkov（2001）的方法计算。

模型）作为先验均值，并将两倍的GMM标准误作为先验标准差（sd）。指数倾斜经验似然由剩余数据构建，每个模型的后验分布通过Chib和Greenberg（1995）的单区块M-H算法进行抽样。尽管参数和工具变量的数量相对较多，该算法仍然快速且高效。结果表明，$price$ 的系数后验均值为-0.14，95%后验可信区间为（-.16，-.13）。该后验均值的绝对值大于BLP最初报告的OLS估计值。请注意，协方差参数 υ的后验分布集中在零的右侧，表明 $price$很可能是内生的。

为确认这一点，我们采用内生性的正式检验。结果见表2。可以看出，在原始 BLP 和扩展 BLP 设定下，扩展模型的边际似然均较大，这支持价格为内生变量这一结论。

我们通过绘制各模型中价格系数的后验分布来结束这项分析。考虑价格的内生性后，价格对汽车需求的估计影响更大（绝对值更大）。值得注意的是，在扩展模型中，价格效应更小且分布更集中，这表明原始 BLP 模型中观察到的部分过度价格敏感性源于遗漏了非线性控制变量。此外，值得注意的是，如果我们只拟合基准模型（边际似然证实该模型在此情形下设定错误），就会忽略纳入非线性项会影响后验分布这一事实。

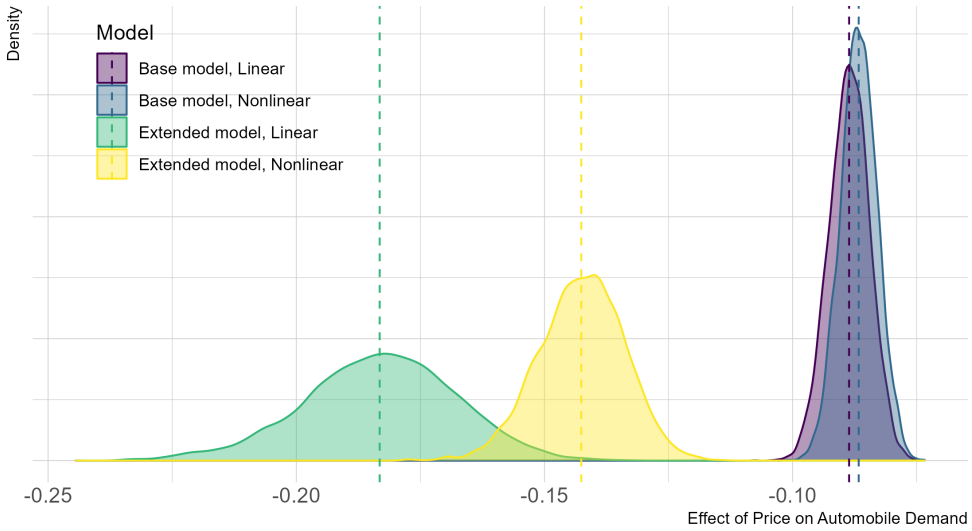

图4：BLP模型：价格变量系数的边际后验分布，β。对于采用原始BLP（线性）设定的基准模型，β 的后验均值和标准差分别为-0.089和0.004；而在扩展BLP（非线性）设定下，它们分别为-0.087和0.004。对于扩展模型，线性设定下 β 的后验均值和标准差分别为-0.183和0.015，非线性设定下分别为-0.143和0.009。

5.2 航空票价对客运量的影响

本文理论与应用的重点是只有一个结果变量的情形；不过，我们的框架适用范围更广。一个重要的例子是聚类纵向数据。令 $y_i = (y_{i1}, \ldots, y_{iT})$ 表示受试对象 i 上 T可能相关且具有异方差的测量值。因此，结果是一个 $T \times 1$ 向量，而非标量。分别调整控制变量和工具变量的维度后，假设在 i之间，聚类结果相互独立且服从线性模型 $y_i = X_i\beta + Z_{1,i}\gamma + \varepsilon_i$，其中 X_i 为 $T \times d_x$, $Z_{1,i}$ 为$T \times d_{z_1}$, $Z_{2,i}$ 为 $T \times d_{z_2}$，且 ε_i 为 $T \times 1$。现在假设 $Z_{1,i}$ 和 $Z_{2,i}$ 满足聚类数据外生性限制 $\mathrm{E}[Z'_{j,i}\varepsilon_i(\theta)] = 0, j = 1, 2$，但与 X_i 相关的聚类数据外生性限制 $\mathrm{E}[X'_i\varepsilon_i(\theta)] = 0$ 存在疑问。我们可以通过定义一个施加后者限制的基准模型，以及一个包含未启用限制 $\mathrm{E}[X'_i\varepsilon_i(\theta)] = v$ 的扩展模型，将我们的框架应用于这一问题；其中 v 现在是一个 $d_x \times 1$ 未知参数向量。与上文提出的方法类似，这些模型的边际似然比较是一种

检验 X的外生性。

为了说明这一扩展框架，我们考虑一个 $T = 4$平衡纵向数据集，其中包含Wooldridge（2010）提供的1997、1998、1999和2000年的航空票价和客运量数据。对于每一年 t，$t \leq 4$，数据按航线 i、$i \leq n = 1149$进行聚类。对于由始发城市和目的地城市界定的每条航班航线，数据包含日均乘客人数的对数（$lpassen$）、以美元计的平均单程票价的对数（$lfare$）、以英里计的距离的对数（$ldist$），以及最大承运商所占市场份额（$concen$）。我们关注的模型为 $lpassen_{it} = \beta\, lfare_{it} + \gamma_1 trend_t + \gamma_2 ldist_{it} + \varepsilon_{it}$，其中 $trend$是一个取值为1、2、3、和4的趋势变量，并且该回归中的每个变量都以均值为中心。我们的目标是估计价格弹性参数 β，但我们担心 $lfare$ 可能是内生的。在估计中，我们假设 $concen$ 是一个有效工具变量（它不会直接出现在结果模型中，并且会影响 $lfare$；这两个假设都是合理的）。

按航线聚类 i，我们有

$$\begin{pmatrix} lpassen_{i1} \\ lpassen_{i2} \\ lpassen_{i3} \\ lpassen_{i4} \end{pmatrix} = \begin{pmatrix} lfare_{i1} & 1 & ldist_{i1} \\ lfare_{i2} & 2 & ldist_{i2} \\ lfare_{i3} & 3 & ldist_{i3} \\ lfare_{i4} & 4 & ldist_{i4} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{pmatrix},$$

~或简写为 $y_i = W_{1,\ i}\theta + \varepsilon_i, i = 1$，2，.....，1149，其中 θ：7×1 是我们关注的未知参数。在此模型中，ε_i 的分布未作规定。此外，ε_i 的各个元素可以以任意且未知的方式呈现序列相关和异方差。

现在令 $Z_i := \left(\widetilde{W}_{1,i}, 1, concen_i\right), i \leq n$, 为一个 4×5 矩阵，其中 1 是全1向量，而 $concen_i = (concen_{i1},\ \ldots,\ concen_{i4})'$：$4 \times 1$ 是航线 i的 $concen$ 值向量。在基准模型中，$lfare$ 是外生的。该模型由以下五个矩条件定义

$$\mathcal{M}_b: \quad \mathbf{E}[Z_i'(y_i - X_i\theta)] = 0_{5\times 1}.$$

在扩展模型中，$lfare$ 矩条件不活跃。具体而言 ，

$$\mathcal{M}_e: \quad \mathbf{E}[Z_i'(y_i - X_i\theta)] = \begin{pmatrix} v \\ 0_{4\times 1} \end{pmatrix}.$$

基于 ETEL 的这两个模型的估计，不对簇级误差的联合分布作任何假设。

我们根据训练样本设定先验分布。我们将样本随机分为训练样本（例如包含 115 个簇，相当于簇总数的 10%）和估计样本（包含其余 1,034 个簇）。随后，我们使用一个以训练数据得出的系统范围两阶段最小二乘估计为中心、标准差为 10 且自由度为 2.5 的学生 t 先验，在训练样本上估计基准模型。在该先验下，根据这些训练数据计算后验均值和标准差。然后，我们将训练样本拟合所得的后验均值和两倍标准差，分别作为先验的均值和标准差。根据训练样本确定先验有助于模型拟合；但由于先验分布具有厚尾，先验所提供的信息与估计样本所提供的信息相比微不足道。

我们采用单块定制 MCMC 算法对每个模型中的后验分布进行抽样。在基准模型中，基于预烧期 1,000 次后的 10,000 次 MCMC 抽样，我们发现 β 的后验均值为 -0.551，其 95% 后验可信区间为 (-0.683, -0.419)。此外，计算结果表明，$\log(m(w_{1:n}|\mathcal{M}_b) = -7190.222$和 $\log(m(w_{1:n}|\mathcal{M}_e) = -7191.06$，这表明在此问题中，$lprice$ 可视为外生变量。

6 结论

本文针对线性均值回归模型中所关注处理变量向量的外生性/内生性，开发了一种贝叶斯检验。在贝叶斯文献中，通常会假定不存在这一内生性问题，但由于内生性在实践中是常态而非例外，这会导致严重的模型设定错误问题。为避免分布设定错误的风险，我们所建立的框架仅依赖矩约束。本文的分析围绕两个模型展开：

基准模型，其中强制施加外生性假设；以及扩展模型，其中包含外生性矩条件，但使其处于非活动状态。

外生性/内生性的检验程序基于贝叶斯因子，通过比较基准模型与扩展模型的边际 ETEL 来进行检验。我们从频率学派的角度验证了该程序，因为我们确立了贝叶斯因子检验的大样本相合性。此外，我们全面研究了对数边际 ETEL 函数，并确定了根据协变量 x 是外生还是内生，该函数的哪些部分会在检验程序中发挥作用。

真实数据示 例展示了这些方法的实际意义 .

值得一提的是，本文提出的方法可以扩展到假设控制变量以非参数方式进入模型的情形。对于这类模型，在通过（例如）样条基展开方法逼近未知函数后，其有限样本分析将与本文讨论的方法大致相同；但先验的设定和大样本分析则需要新的理论进展，以处理随样本量增加而不断增多的基函数参数。我们计划在未来的论文中阐述相关理论。

另一个有趣的延拓方向是放宽我们目前的强识别假设。该假设基于相关工具变量数量固定，而放宽后将允许弱工具变量和大量工具变量。这需要对我们的渐近结果做出重大调整，因为还必须针对 ETEL 开发局部渐近贝叶斯因子，并纳入工具变量数量增长惩罚或收缩，以防止过拟合。

A 证明

A.1 定理 4.1 的证明

我们首先证明*(i)* 等价于*(ii)*。令 $H_\infty := \{\theta \in \Theta; \mathcal{Q}_b(\theta) \neq \emptyset\}$。假设*(i)* 成立。那么，对于每个 $\theta \in \Theta$，都有 $P \notin \mathcal{Q}_b(\theta)$，且 I-投影到集合 $\mathcal{Q}_b(\theta)$ 上的 P 不同于

P，也就是说：对每个 $\theta \in H_\infty$， $Q_b^*(\theta) \neq P$且 $KL(Q_b^*(\theta)||P) > 0$；而对于每个$\theta \in \Theta \setminus H_\infty$，它都没有定义，因为在这些 θ 取值下，集合 $\mathcal{Q}_b^*(\theta)$ 为空。因此，根据假设1，反向Kullback-Leibler散度（即交换两个概率所起的作用）也严格为正：对每个 $\theta \in H_\infty$，$\mathrm{KL}(P||Q_b^*(\theta)) > 0$。由于这对每个 $\theta \in H_\infty$ 都成立，因此对 θ_*也成立。于是*(ii)* 成立。

现在，假设（*ii*）成立。由于 $\mathrm{KL}(P||Q_b^*(\theta_*)) > 0$，因此 $P \neq Q_b^*(\theta_*)$，从而$P \notin \mathcal{Q}_b(\theta_*)$。由于对于满足$\mathcal{Q}_b(\theta) \neq \emptyset$的所有 θ， θ_* 使 $\mathrm{KL}(P||Q_b^*(\theta))$ 最小，因此对于满足$\mathcal{Q}_b(\theta) \neq \emptyset$的每个 $\theta \in \Theta$，我们也有 $P \notin \mathcal{Q}_b(\theta)$。对于满足$\mathcal{Q}_b(\theta) = \emptyset$ 的 θ，条件（*i*）显然成立，因此我们证明了（*i*）对每个 $\theta \in \Theta$均成立。

接下来，我们证明（*iii*）等价于（*iv*）。假设（*iii*）成立。那么，存在一个 $\theta \in \Theta$，记为θ_*，使得 $P \in \mathcal{Q}_b(\theta_*)$。因此， $Q_b^*(\theta_*) = P$，且 $\mathrm{KL}(P||Q_b^*(\theta)) = 0$。所以（*iv*）成立。

现在，假设*(iv)*成立。根据Kullback-Leibler散度的性质，KL散度$(P||Q_b^*(\theta_*))= 0$ 当且仅当 $P = Q_b^*(\theta_*)$。由此可得 $P \in \mathcal{Q}_b(\theta_*)$，因为 $Q_b^*(\theta_*) \in \mathcal{Q}_b(\theta_*)$，因此$P$满足矩约束$\mathbf{E}^P[\varepsilon_i(\theta_*) x_i] = 0$。故*(iii)* 成立。

□

A.2 定理4.2的证明

我们来考察在 θ_*处取值的似然表达式：

$$\begin{aligned}\log \widehat{q}(w_{1:n}|\theta_*, \mathcal{M}_b) &= -n\log n + \sum_{i=1}^n \widehat{\lambda}(\theta_*)' g_b(w_i, \theta_*) - n\log \frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\theta_*)' g_b(w_j, \theta_*)} \\ &= -n\log n + \sum_{i=1}^n \widehat{\lambda}(\theta_*)' \widetilde{g}_b(w_i, \theta_*) - n\log \frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\theta_*)' \widetilde{g}_b(w_j, \theta_*)}, \\ &= -n\log n + \sum_{i=1}^n \widehat{\lambda}(\theta_*)' \widetilde{g}_b(w_i, \theta_*) + n\widehat{\lambda}(\theta_*)' \mathbf{E}[g_b(w_i, \theta_*)] - n\log \frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\theta_*)' g_b(w_j, \theta_*)}, \quad \text{(A.1)}\end{aligned}$$

其中 $\widetilde{g}_b(w_i, \theta_*) := g_b(w_i, \theta_*) - \mathbf{E}[g_b(w_i, \theta_*)]$。我们首先处理（A.1）右边的第二项。利用引理 F.4 的结果：

$$\sum_{i=1}^{n}\widehat{\lambda}(\theta_*)'\widetilde{g}_b(w_i,\theta_*)=\sqrt{n}(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{g}_b(w_i,\theta_*)+\lambda_*(\theta_*)'\sum_{i=1}^{n}\widetilde{g}_b(w_i,\theta_*)$$
$$=-\mathbb{G}_n\left[\tau_i^{\dagger}(\lambda_*,\theta_*)g_b(w_i,\theta_*)'\right]\Omega_*^{\dagger}(\theta_*)^{-1}\mathbb{G}_n\left[g_b(w_i,\theta_*)\right]+\lambda_*(\theta_*)'\sum_{i=1}^{n}\widetilde{g}_b(w_i,\theta_*). \quad \text{(A.2)}$$

˜接下来，我们分析（A.1）右边的第四项。设 λ 位于连接 $\lambda_*(\theta_*)$ 和 $\widehat{\lambda}(\theta_*)$ 的直线上，则对 $\lambda\mapsto\frac{1}{n}\sum_{j=1}^{n}e^{\lambda' g_b(w_j,\theta_*)}$ 在 $\lambda_*(\theta_*)$ 处进行二阶项泰勒展开可得

$$\frac{1}{n}\sum_{j=1}^{n}e^{\widehat{\lambda}(\theta_*)'g_b(w_j,\theta_*)}=\frac{1}{n}\sum_{j=1}^{n}e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}+(\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\frac{1}{n}\sum_{i=1}^{n}e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}g_b(w_i,\theta_*)$$
$$+\frac{1}{2}(\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\frac{1}{n}\sum_{j=1}^{n}e^{\tilde{\lambda}'g_b(w_j,\theta_*)}g_b(w_j,\theta_*)g_b(w_j,\theta_*)'(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*)). \quad \text{(A.3)}$$

根据假设 4，并且由于 $\|\tilde{\lambda}-\lambda_*(\theta_*)\|_2=\mathcal{O}_p(n^{-1/2})$（因为由引理 F.3 $\|\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*)\|_2=\mathcal{O}_p(n^{-1/2})$，且 $\tilde{\lambda}=\tau(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))+\lambda_*(\theta_*)$ 对某个 $\tau\in[0,1]$ 成立），我们可以应用引理 F.6 的证明中相同的论证，得到：

$$\frac{1}{n}\sum_{j=1}^{n}e^{\tilde{\lambda}'g_b(w_j,\theta_*)}g_b(w_j,\theta_*)g_b(w_j,\theta_*)'\xrightarrow{p}\Omega_*^{\diamond}(\theta_*):=\mathbf{E}[e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}\varepsilon_i(\theta_*)^2\widetilde{w}_i\widetilde{w}_i']. \quad \text{(A.4)}$$

将此代入 (A.3)，并利用引理 F.3 得到项 $o_p(1/n)$项的收敛速率，我们得到：

$$\frac{1}{n}\sum_{j=1}^{n}e^{\widehat{\lambda}(\theta_*)'g_b(w_j,\theta_*)}=\frac{1}{n}\sum_{j=1}^{n}e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}+(\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\frac{1}{n}\sum_{i=1}^{n}e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}g_b(w_i,\theta_*)$$
$$+\frac{1}{2}(\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\Omega_*^{\diamond}(\theta_*)(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))+o_p\left(\frac{1}{n}\right) \quad \text{(A.5)}$$

现在，我们对函数 $u\mapsto\log(u)$ 在 v 处进行一阶泰勒展开：$\log(u)=\log(v)+\frac{u-v}{v}+o(|u-v|)$，并将 (A.5) 代入其中，得到：

$$\log\left(\frac{1}{n}\sum_{i=1}^{n}e^{\widehat{\lambda}(\theta_*)'g_b(w_i,\theta_*)}\right)=\log\left(\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]\right)+(\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\frac{1}{n}\sum_{i=1}^{n}\tau_i(\lambda_*,\theta_*)g_b(w_i,\theta_*)$$

$$+\frac{1}{2}(\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\Omega_*^\dagger(\theta_*)(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))+o_p\left(\frac{1}{n}\right)$$
$$+o\left(\left|\frac{1}{n}\sum_{i=1}^n e^{\widehat{\lambda}(\theta_*)'g_b(w_i,\theta_*)}-\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]\right|\right), \quad \text{(A.6)}$$

其中，为了得到项 $o_p\left(\frac{1}{n}\right)$，我们使用了引理 F.3，以及以下事实：$\left|\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]-\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]\right|=\mathcal{O}_p(1/\sqrt{n})$。再次利用后一收敛性、引理 F.4，以及以下事实：$\mathbf{E}[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)]=0$ 以及

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)=\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)\right],$$

then

$$\begin{aligned}
&n\left((\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\frac{1}{n}\sum_{i=1}^n\tau_i(\lambda_*,\theta_*)g_b(w_i,\theta_*)+\frac{1}{2}(\widehat{\lambda}(\theta_*)'-\lambda_*(\theta_*)')\Omega_*^\dagger(\theta_*)(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))\right)\\
&\quad=-\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)\right]\\
&\quad+\frac{1}{2}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)\right]+o_p(1)\\
&\quad=-\frac{1}{2}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)\right]+o_p(1). \quad \text{(A.7)}
\end{aligned}$$

最后，我们必须处理 $\left|\frac{1}{n}\sum_{i=1}^n e^{\widehat{\lambda}(\theta_*)'g_b(w_i,\theta_*)}-\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]\right|$。利用 (A.5) 和 (A.7)，我们得到 $\left|\frac{1}{n}\sum_{i=1}^n e^{\widehat{\lambda}(\theta_*)'g_b(w_i,\theta_*)}-\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]\right|=O_p(1/n)$。

将此结果以及 (A.2)、(A.6) 和 (A.7) 代入 (A.1)，我们得到：

$$\begin{aligned}
\log\widehat{q}(w_{1:n}|\theta_*,M_b)&=-n\log n-\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n\left[g_b(w_i,\theta_*)\right]\\
&+\lambda_*(\theta_*)'\sum_{i=1}^n\widetilde{g}_b(w_i,\theta_*)+n\widehat{\lambda}(\theta_*)'\mathbf{E}[g_b(w_i,\theta_*)]-n\log\left(\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]\right)\\
&+\frac{1}{2}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)\right]+o_p(1), \quad \text{(A.8)}
\end{aligned}$$

其中 $\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)g_b(w_i,\theta_*)\right]\xrightarrow{d}$ 服从 $N(0,\Omega_*^\dagger(\theta_*))$。注意到 $\lambda_*(\theta_*)'\sum_{i=1}^n\widetilde{g}_b(w_i,\theta_*)-n\log\left(\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}\right]\right)=\sum_{i=1}^n\log\left(\tau_i(\lambda_*,\theta_*)\right)-n\lambda_*(\theta_*)'\mathbf{E}[g_b(w_i,\theta_*)]$,，即可证明 (4.3)。

定理的最后一部分由引理 F.4 得出

□

A.3 定理 4.3 的证明

由于我们处于扩展模型中，因此存在一个 $\psi_\circ := (\theta'_\circ \ v'_\circ)'$，使得 $\mathbf{E}[\varepsilon_i(\theta_\circ)\ \widetilde{w}_i] = (v'_\circ, 0')'$ 且 $\lambda_*(\psi_\circ) = 0_\circ$ 考虑在 $\psi_\circ$ 处取值的似然表达式：

$$\begin{aligned}\log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) &= -n\log n + \sum_{i=1}^n \widehat{\lambda}(\psi_\circ)' g_b(w_i, \theta_\circ) - n\log \frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\psi_\circ)' g_b(w_j,\theta_\circ)} \\ &= -n\log n + \sum_{i=1}^n \widehat{\lambda}(\psi_\circ)' g_e(w_i, \psi_\circ) - n\log \frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\psi_\circ)' g_e(w_j,\psi_\circ)}. \quad \text{(A.9)}\end{aligned}$$

我们先处理 (A.9) 右边的第二项。利用引理 F.17 的结果：

$$\begin{aligned}\sum_{i=1}^n \widehat{\lambda}(\psi_\circ)' g_e(w_i, \psi_\circ) &= \sqrt{n}\widehat{\lambda}(\psi_\circ)' \frac{1}{\sqrt{n}}\sum_{i=1}^n g_e(w_i, \psi_\circ) \\ &= -\mathbb{G}_n\left[g_e(w_i, \psi_\circ)'\right] \Omega_{\psi_\circ}^{-1} \mathbb{G}_n\left[g_e(w_i, \psi_\circ)\right] + o_p(1). \quad \text{(A.10)}\end{aligned}$$

令 λ 位于连接 0 和 $\widehat{\lambda}(\psi_\circ)$ 的直线上，则函数 $\lambda \mapsto \frac{1}{n}\sum_{j=1}^n e^{\lambda' g_e(w_j,\ \theta_\circ)}$ 在 0 处的二阶泰勒展开为

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\psi_\circ)' g_e(w_j,\psi_\circ)} &= 1 + \frac{1}{n}\sum_{i=1}^n \widehat{\lambda}(\psi_\circ)' g_e(w_i, \psi_\circ) \\ &\quad + \frac{1}{2}\widehat{\lambda}(\psi_\circ)' \frac{1}{n}\sum_{j=1}^n e^{\tilde{\lambda}' g_e(w_j,\psi_\circ)} g_e(w_j, \psi_\circ) g_e(w_j, \psi_\circ)' \widehat{\lambda}(\psi_\circ). \quad \text{(A.11)}\end{aligned}$$

在假设 4 下，并且由于 $\tilde{\lambda} = \mathcal{O}_p(n^{-1/2})$（因为根据引理 F.18 $\widehat{\lambda}(\psi_\circ) = \mathcal{O}_p(n^{-1/2})$，且 $\tilde{\lambda}$ 介于 0 和 $\widehat{\lambda}(\psi_\circ)$ 之间），我们可以沿用引理 F.6 证明中的相同论证，得到：

$$\frac{1}{n}\sum_{j=1}^n e^{\tilde{\lambda}' g_e(w_j,\theta_\circ)} g_e(w_j, \theta_\circ) g_e(w_j, \theta_\circ)' \xrightarrow{p} \Omega_{\psi_\circ} := \mathbf{E}[g_e(w_i, \psi_\circ) g_e(w_i, \psi_\circ)']. \quad \text{(A.12)}$$

将其代入 (A.11)，并利用引理 F.18 求出 $o_p(1/n)$ 项的收敛速率，可得：

$$\frac{1}{n}\sum_{j=1}^{n} e^{\widehat{\lambda}(\psi_\circ)' g_e(w_j,\psi_\circ)} = 1 + \frac{1}{n}\sum_{i=1}^{n} \widehat{\lambda}(\psi_\circ)' g_e(w_i,\psi_\circ) + \frac{1}{2}\widehat{\lambda}(\psi_\circ)'\Omega_{\psi_\circ}\widehat{\lambda}(\psi_\circ) + o_p(1/n). \quad \text{(A.13)}$$

现在，我们使用函数 $u \mapsto \log(u)$ 在 1 处的一阶泰勒展开：$\log(u) = u - 1 + o(|u-1|)$，并将其应用于由 (A.13) 左侧给出的 u，得到：

$$\begin{aligned}\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\widehat{\lambda}(\psi_\circ)' g_e(w_i,\psi_\circ)}\right) &= \frac{1}{n}\sum_{i=1}^{n} \widehat{\lambda}(\psi_\circ)' g_e(w_i,\psi_\circ) \\ &\quad + \frac{1}{2}\widehat{\lambda}(\psi_\circ)'\Omega_{\psi_\circ}\widehat{\lambda}(\psi_\circ) + o_p(1/n) + o\left(\left|\frac{1}{n}\sum_{i=1}^{n} e^{\widehat{\lambda}(\psi_\circ)' g_e(w_i,\psi_\circ)} - 1\right|\right). \quad \text{(A.14)}\end{aligned}$$

利用引理 F.17 的结果以及 $\mathbf{E}[g_e(w_i,\psi_\circ)] = 0$ 这一事实，可得

$$\begin{aligned} & n\left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\lambda}(\psi_\circ)' g_e(w_i,\psi_\circ) + \frac{1}{2}\widehat{\lambda}(\psi_\circ)'\Omega_{\psi_\circ}\widehat{\lambda}(\psi_\circ)\right) \\ &\qquad = \sqrt{n}\widehat{\lambda}(\psi_\circ)'\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] + \frac{1}{2}\sqrt{n}\widehat{\lambda}(\psi_\circ)'\Omega_{\psi_\circ}\sqrt{n}\widehat{\lambda}(\psi_\circ) \\ &= -\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] + \frac{1}{2}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] + o_p(1) \\ &\qquad = -\frac{1}{2}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right]\Omega_*^{-1}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] + o_p(1). \quad \text{(A.15)}\end{aligned}$$

最后，我们需要处理 $\left|\frac{1}{n}\sum_{i=1}^{n} e^{\widehat{\lambda}(\psi_\circ)' g_e(w_i,\psi_\circ)} - 1\right|$。由 (A.14) 和 (A.15) 可得 $\left|\frac{1}{n}\sum_{i=1}^{n} e^{\widehat{\lambda}(\psi_\circ)' g_e(w_i,\psi_\circ)} - 1\right| = \mathcal{O}_p(1/n)$。

将此结果和（A.15）代入（A.14），再将（A.10）和（A.14）代入（A.9），可得：

$$\begin{aligned}\log\widehat{q}(w_{1:n}|\psi_\circ, M_b) &= -n\log n - \frac{1}{2}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] + o_p(1) \\ &= -n\log n - \frac{1}{2}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)\right] - 2\sqrt{n}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)'\right]\Omega_{\psi_\circ}^{-1}\tilde{v}_\circ + n\tilde{v}_\circ'\Omega_{\psi_\circ}^{-1}\tilde{v}_\circ + o_p(1). \\ & \quad \text{(A.16)}\end{aligned}$$

此外，根据中心极限定理，

$$\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] \xrightarrow{d} \mathcal{N}(0,\Omega_{\psi_\circ})$$

and

$$\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right]\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] \xrightarrow{d} \chi^2_d.$$

□

A.4 推论4.1的证明

根据定理4.2中的结果（4.3），我们有

$$\frac{1}{n}\log\widehat{q}(w_{1:n}|\theta_*,\mathcal{M}_b)+\log(n)=\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}}{\mathbf{E}_n[e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}]}\right)$$
$$+(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))'\mathbf{E}[g_b(w_i,\theta_*)]+\mathcal{O}_p(1/n). \quad \text{(A.17)}$$

根据在线附录中的引理 F.3，$\|\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*)\|_2 \xrightarrow{p} 0_\circ$ 根据大数定律

$$\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}}{\mathbf{E}_n[e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}]}\right)\xrightarrow{p}\mathbf{E}\left[\log\left(\frac{e^{\lambda_*(\theta_*)'g_b(w_i,\theta_*)}}{\mathbf{E}[e^{\lambda_*(\theta_*)'g_b(w_j,\theta_*)}]}\right)\right].$$

至此，证明完毕。

□

A.5 推论4.2的证明

根据定理 4.3 中的结果 (4.4)，我们有

$$\frac{1}{n}\log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) + \log(n) = \mathcal{O}_p(1/n).$$

由于 $\lambda_*(\psi_\circ) = 0$，因此，$\frac{e^{\lambda_*(\psi_\circ)'\sum_{i=1}^n g_b(w_{i,}\ \psi_\circ)}}{\mathbf{E}[e^{\lambda_*(\psi_\circ)'\sum_{i=1}^n g_b(w_{j,}\ \psi_\circ)}]} = 1$，因此我们可以等价地写成：

$$\frac{1}{n}\log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) + \log(n) = \frac{1}{n}\sum_{i=1}^n \log\left(\frac{e^{\lambda_*(\psi_\circ)'g_b(w_i,\psi_\circ)}}{\mathbf{E}[e^{\lambda_*(\psi_\circ)'g_b(w_j,\psi_\circ)}]}\right) + \mathcal{O}_p(1/n).$$

根据大数定律

$$\frac{1}{n}\sum_{i=1}^n \log\left(\frac{e^{\lambda_*(\psi_\circ)'g_b(w_i,\psi_\circ)}}{\mathbf{E}[e^{\lambda_*(\psi_\circ)'g_b(w_j,\psi_\circ)}]}\right) \xrightarrow{p} \mathbf{E}\left[\log\left(\frac{e^{\lambda_*(\psi_\circ)'g_b(w_i,\psi_\circ)}}{\mathbf{E}[e^{\lambda_*(\psi_\circ)'g_b(w_j,\psi_\circ)}]}\right)\right].$$

至此，证明完毕。

□

A.6 定理4.4的证明

证明分为两部分。第一部分，我们证明 $\mathrm{KL}(P||Q_e^*(\psi_\circ)) < \mathrm{KL}(P||Q_b^*(\theta_*))$ 当且仅当不存在满足以下条件的 θ： $\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$。第二部分，我们证明

$$P\left(\log m(w_{1:n}|\mathcal{M}_e) > \log m(w_{1:n}|\mathcal{M}_b)\right) \to 1$$

当且仅当 $\mathrm{KL}(P||Q_e^*(\psi_\circ)) < \mathrm{KL}(P||Q_b^*(\theta_*))$。

第一部分。我们首先证明，KL散度$(P||Q_e^*(\psi_\circ)) <$ KL散度$(P||Q_b^*(\theta_*))$ 当且仅当不存在 $\theta \in \Theta$ 使得$\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$。注意，KL散度$(P||Q_e^*(\psi_\circ)) = 0$。假设 KL散度$(P||Q_e^*(\psi_\circ)) <$KL散度$(P||Q_b^*(\theta_*))$，并假设存在一个 $\theta \in \Theta$ 使得$\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$，从而 $P \in \mathcal{Q}_b(\theta)$。根据假设 2，将 θ_* 替换为 $\theta_\circ$，则这个 θ 必定等于 $\theta_\circ$，而后者又等于 $\theta_{*\circ}$

由此可得 $P \in \mathcal{Q}_b(\theta_*)$。根据 $Q_b^*(\theta_*)$的定义： $Q_b^*(\theta_*) = P$，因为在 KL 散度意义上，在$\mathcal{Q}_b(\theta_*)$中的所有分布里， $Q_b^*(\theta_*)$最接近P。因此，KL散度$(P||Q_b^*(\theta_*)) = 0$。但这与 KL散度$(P||Q_e^*(\psi_\circ)) <$KL散度$(P||Q_b^*(\theta_*))$ 的假设矛盾。因此，不存在 $\theta \in \Theta$ 使得$\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$。

我们现在证明逆向蕴含。假设不存在值 $\theta \in \Theta$ 使得$\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$。因此，对每一个 $\theta \in \Theta$都有 $P \notin \mathcal{Q}_b(\theta)$，这意味着 $P \notin \mathcal{Q}_b(\theta_*)$且KL散度$(P||Q_b^*(\theta_*)) >0$。另一方面，由于$\mathcal{M}_e$ 总是被正确设定，存在唯一的 $\psi_\circ \in \mathbb{R}^{d_x}$ 使得 $P \in \mathcal{Q}_e(\psi_\circ)$。这意味着KL散度$(P||Q_e^*(\psi_\circ)) = 0$，从而KL散度$(P||Q_e^*(\psi_\circ)) <$KL散度$(P||Q_b^*(\theta_*))$。

第二部分。 我们证明， $P(\log\ m(w_{1:n}|\mathcal{M}_e) >\log\ m(w_{1:n}|\mathcal{M}_b)) \to 1$ 当且仅当KL散度$(P||Q_e^*(\psi_\circ))$ <KL散度$(P||Q_b^*(\theta_*))$。根据定理4.2和4.3以及在线附录中的定理F.6和F.7，并由(4.1)-(4.2)，可知(4.9)-(4.10)成立。根据(4.9)，$\log m(w_{1:n}|\mathcal{M}_b)$等于

$$-n\log n + \sum_{i=1}^{n} \log\left(\frac{e^{\lambda_*(\theta_*)' g_b(w_i,\theta_*)}}{\mathbf{E}_n[e^{\lambda_*(\theta_*)' g_b(w_j,\theta_*)}]}\right) + n(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))'\mathbf{E}[g_b(w_i,\theta_*)] - \frac{p}{2}\log(n) + \mathcal{O}_p(1)$$

并且根据（4.10），$\log\ m(w_{1:n}|\mathcal{M}_e) = -n\log(n) - \frac{p+d_x}{2}\log(n) + \mathcal{O}_p(1)$。因此，由大数定律可知

$$\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{e^{\lambda_*(\theta_*)' \sum_{i=1}^{n} g_b(w_i,\theta_*)}}{\mathbf{E}_n[e^{\lambda_*(\theta_*)' \sum_{i=1}^{n} g_b(w_j,\theta_*)}]}\right) \xrightarrow{p} \mathbf{E}\left[\log\left(\frac{e^{\lambda_*(\theta_*)' \sum_{i=1}^{n} g_b(w_i,\theta_*)}}{\mathbf{E}[e^{\lambda_*(\theta_*)' \sum_{i=1}^{n} g_b(w_j,\theta_*)}]}\right)\right] \equiv -\mathrm{KL}(P||Q_b^*(\theta_*)),$$

由此可得

$$\begin{aligned} P\left(\log m(w_{1:n}|\mathcal{M}_e) > \log m(w_{1:n}|\mathcal{M}_b)\right) &= P\left(\frac{1}{n}\log m(w_{1:n}|\mathcal{M}_e) > \frac{1}{n}\log m(w_{1:n}|\mathcal{M}_b)\right) \\ &= P\left(0 > -\mathrm{KL}(P||Q_b^*(\theta_*)) + \mathcal{O}_p(\log(n)/\sqrt{n})\right), \end{aligned}$$

其中，我们使用在线附录中的引理 F.3 来控制 $\sqrt{n}(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))$。假设 KL$(P||Q_b^*(\theta_*)) > 0$，则前述概率收敛于 1。另一方面，sup-

pose that $P(0 > -\mathrm{KL}(P||Q_b^*(\theta_*)) + \mathcal{O}_P(\log(n)/\sqrt{n})) \to 1$当 $n \to \infty$时成立。这只有在 $\mathrm{KL}(P||Q_b^*(\theta_*)) > 0$时才可能。根据证明的第一部分，$\mathrm{KL}(P||Q_b^*(\theta_*)) > 0$ 当且仅当不存在 θ 使得 $\mathbf{E}[\varepsilon_i(\theta)x_i] = 0$。

我们现在证明该定理的最后一个断言。在存在一个 $\theta_\circ$ 使得$\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$的情形下，则KL散度$(P||Q_b^*(\theta_\circ)) = 0$，并且概率 $P(0 > -$KL散度$(P||Q_b^*(\theta_*)) + \mathcal{O}_P(\log(n)/\sqrt{n}))$在 $n \to \infty$时等于零。–至此证明完毕。

□

A.7 定理4.5的证明

我们首先假设$\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$。在这种情况下，$\theta_* = \theta_\circ$，$\lambda_*(\theta_*) = \lambda_*(\theta_\circ) = 0$，并根据定理4.2和4.3：

$$\begin{aligned}\log \widehat{q}(w_{1:n}|\theta_\circ, \mathcal{M}_b) - \log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) \\ = -n\log n - \frac{1}{2}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)'\right]\Omega_\circ^{-1}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)\right] + n\log n \\ + \frac{1}{2}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)'\right]\Omega_\circ^{-1}\mathbb{G}_n\left[g_b(w_i,\theta_\circ)\right] + o_p(1) = o_p(1). \quad \text{(A.18)}\end{aligned}$$

设 $\pi_{h_\theta}^n(\cdot|w_{1:n}, \mathcal{M}_b)$和 $\pi_{h_\psi}^n(\cdot|w_{1:n}, \mathcal{M}_e)$分别表示 h_θ和 h_ψ的后验密度。根据下面的推论 F.1（若$\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$ 成立，则该推论有效）

$$\begin{aligned}\log \pi_{h_\theta}^n(\sqrt{n}(\theta-\theta_\circ)|w_{1:n}, \mathcal{M}_b)\Big|_{\theta=\theta_\circ} = -\frac{p}{2}\log(2\pi) + \frac{1}{2}\log|V_{\theta_\circ}^{-1}| \\ - \frac{1}{2}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i'\right]\Omega_\circ^{-1}\mathbf{E}[\widetilde{w}_i\widetilde{w}_{1,i}']V_{\theta_\circ}\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\Omega_\circ^{-1}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i\right] + o_p(1), \quad \text{(A.19)}\end{aligned}$$

并由下文的定理 F.7

$$\begin{aligned}\log \pi_{h_\psi}^n(\sqrt{n}(\psi-\psi_\circ)|w_{1:n}, \mathcal{M}_e)\Big|_{\psi=\psi_\circ} = -\frac{(p+d_x)}{2}\log(2\pi) + \frac{1}{2}\log|V_{\psi_\circ}^{-1}| \\ - \frac{1}{2}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i'\right]\Omega_\circ^{-1}\left[\frac{dg_e(w_i,\psi_\circ)'}{d\psi}\right]V_{\psi_\circ}\left[\frac{dg_e(w_i,\psi_\circ)}{d\psi'}\right]\Omega_\circ^{-1}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i\right] + o_p(1), \quad \text{(A.20)}\end{aligned}$$

其中 $V_{\theta_\circ}$ 与 $V_{\psi_\circ}$ 的定义见下文推论 F.1 和定理 F.7。因此，利用 (4.1)-(4.2) 所给的对数边际似然表达式，并将 θ_* 替换为 $\theta_\circ$，对 $\log m(w_{1:n}|\mathcal{M}_b) - \log m(w_{1:n}|\mathcal{M}_e)$ 中的 (A.20)、(A.19) 和 (A.18) 进行替换，我们得到：

$$
\begin{aligned}
P\left(\log m(w_{1:n}|\mathcal{M}_b) > \log m(w_{1:n}|\mathcal{M}_e)\right) = P\Big(& \log \pi(\theta_\circ|\mathcal{M}_b) + \frac{p}{2}\log(2\pi) - \frac{1}{2}\log|V_{\theta_\circ}^{-1}| \\
& + \frac{1}{2}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i'\right]\Omega_\circ^{-1}\mathbf{E}[\widetilde{w}_i\widetilde{w}_{1,i}']V_{\theta_\circ}\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\Omega_\circ^{-1}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i\right] - \frac{p}{2}\log(n) \\
& > \log \pi(\psi_\circ|\mathcal{M}_e) + \frac{(p+d_x)}{2}\log(2\pi) - \frac{1}{2}\log|V_{\psi_\circ}^{-1}| \\
+\frac{1}{2}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i'\right]\Omega_\circ^{-1} & \left[\frac{dg(w_i,\psi_\circ)'}{d\psi}\right] V_{\psi_\circ}\left[\frac{dg(w_i,\psi_\circ)}{d\psi'}\right]\Omega_\circ^{-1}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i\right] - \frac{p+d_x}{2}\log(n) + o_p(1)\Big).
\end{aligned}
\tag{A.21}
$$

由于 $\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\,\widetilde{w}_i\right] = \mathcal{O}_p(1)$、$|V_{\theta_\circ}^{-1}| = \mathcal{O}(1)$ 和 $|V_{\psi_\circ}^{-1}| = \mathcal{O}(1)$（因为在假设 5 下，$V_{\theta_\circ}$ 和 $V_{\psi_\circ}$ 是正定的），因此我们可以将 (A.21) 中的 $\log(n)$ 因式分解，得到：

$$
\begin{aligned}
& P\left(\log m(w_{1:n}|\mathcal{M}_b) > \log m(w_{1:n}|\mathcal{M}_e)\right) = \\
& \quad = P\Big(0 > \log(n)\Big[\frac{1}{\log(n)}\log\frac{\pi(\psi_\circ|\mathcal{M}_e)}{\pi(\theta_\circ|\mathcal{M}_b)} + \frac{d_x\log(2\pi)}{2\log(n)} - \frac{1}{2\log(n)}\log\frac{|V_{\theta_\circ}|}{|V_{\psi_\circ}|} - \frac{d_x}{2} \\
& + \frac{1}{2\log(n)}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i'\right]\Omega_\circ^{-1}\left(\left[\frac{dg(w_i,\psi_\circ)'}{d\psi}\right]V_{\psi_\circ}\left[\frac{dg(w_i,\psi_\circ)}{d\psi'}\right] - \mathbf{E}[\widetilde{w}_i\widetilde{w}_{1,i}']V_{\theta_\circ}\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\right) \\
& \quad \times \Omega_\circ^{-1}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i\right]\Big] + o_p(1)\Big) = P\left(0 > \log(n)\left[o_p(1) - \frac{d_x}{2}\right] + o_p(1)\right) \to 1
\end{aligned}
\tag{A.22}
$$

因为 $n \to \infty_\circ$ 这证明了第一个蕴含。

现在我们证明逆向蕴含。假设 $P(\log m(w_{1:n}|\mathcal{M}_b) > \log m(w_{1:n}|\mathcal{M}_e)) \to 1$。由 (4.1)-(4.2) 可得：

$$
\begin{aligned}
P\left(\log m(w_{1:n}|\mathcal{M}_b) > \log m(w_{1:n}|\mathcal{M}_e)\right) = P\Big(& \log \pi(\theta_*|\mathcal{M}_b) + \log \widehat{q}(w_{1:n}|\theta_*,\mathcal{M}_b) \\
& - \log \pi_{h_\theta}^n(0|w_{1:n},\mathcal{M}_b) - \frac{p}{2}\log(n) > \log \pi(\psi_\circ|\mathcal{M}_e) + \log \widehat{q}(w_{1:n}|\psi_\circ,\mathcal{M}_e) \\
& - \log \pi_{h_\psi}^n(0|w_{1:n},\mathcal{M}_e) - \frac{p+d_x}{2}\log(n)\Big).
\end{aligned}
\tag{A.23}
$$

利用定理4.2和4.3，我们得到：

$$\begin{aligned}\log \widehat{q}(w_{1:n}|\theta_*, \mathcal{M}_b) - \log \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) &= -\mathcal{A}_n' \Omega_*^\dagger(\theta_*)^{-1} \mathcal{B}_n \\ &+ \sum_{i=1}^n \log\left(\frac{e^{\lambda_*(\theta_*)' g_b(w_i,\theta_*)}}{\mathbf{E}_n[e^{\lambda_*(\theta_*)' g_b(w_j,\theta_*)}]}\right) + n(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))' \mathbf{E}[g_b(w_i,\theta_*)] \\ &+ \frac{1}{2}\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*) g_b(w_i,\theta_*)'\right] \Omega_*^\dagger(\theta_*)^{-1} \mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*) g_b(w_i,\theta_*)\right] \\ &+ \frac{1}{2}\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right] \Omega_{\psi_\circ}^{-1} \mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] + o_p(1), \quad \text{(A.24)}\end{aligned}$$

其中 $\mathcal{A}_n := \mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*) g_b(w_i,\theta_*)'\right] \xrightarrow{d} \mathcal{N}(0, \Omega_*^\dagger(\theta_*))$, $\mathcal{B}_n := \mathbb{G}_n\left[g_b(w_i,\theta_*)'\right] \xrightarrow{d} \mathcal{N}(0, \mathbf{E}[\varepsilon_i(\theta_\circ)]\widetilde{w}_i\widetilde{w}_i')$，并且 $\mathbb{G}_n\left[g_e(w_i,\psi_\circ)'\right] \Omega_{\psi_\circ}^{-1} \mathbb{G}_n\left[g_e(w_i,\psi_\circ)\right] \xrightarrow{d} \chi_d^2$，因此它们依概率有界。此外，根据大数定律

$$\left|\frac{1}{n}\sum_{i=1}^n \log\left(\frac{e^{\lambda_*(\theta_*)' g_b(w_i,\theta_*)}}{\mathbf{E}_n[e^{\lambda_*(\theta_*)' g_b(w_j,\theta_*)}]}\right) - \mathbf{E}\left[\log(dQ_b^*(\theta_*)/dP)\right]\right| = \mathcal{O}_p(1/\sqrt{n}),$$

其中$\mathbf{E}\left[\log(dQ_b^*(\theta_*)/dP)\right] = -$KL散度$(P||Q_b^*(\theta_*))$，并根据下面的引理 F.4,

$$\sqrt{n}(\widehat{\lambda}(\theta_*) - \lambda(\theta_*))' \mathbf{E}[g_b(w_i,\theta_*)] = -\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i'] \Omega_*^\dagger(\theta_*)^{-1} \mathbf{E}[g_b(w_i,\theta_*)] + o_p(1).$$

因此,

$$\begin{aligned}&\text{对数 } \widehat{q}(w_{1:n}|\theta_*, \mathcal{M}_b) - \text{对数 } \widehat{q}(w_{1:n}|\psi_\circ, \mathcal{M}_e) \\ &\qquad = \mathcal{O}_p(1) + n\left(\mathcal{O}_p(1/\sqrt{n}) - \mathrm{KL}(P||Q_b^*(\theta_*))\right) + \sqrt{n}\mathcal{O}_p(1). \quad \text{(A.25)}\end{aligned}$$

将(A.25)代入(A.23)，并利用下面的定理F.6和F.7来证明$\log \pi_{h_\theta}^n(0|w_{1:n}, \mathcal{M}_b) = \mathcal{O}_P(1)$以及$\log \pi_{h_\psi}^n(0|w_{1:n}, \mathcal{M}_e) = \mathcal{O}_P(1)$，则(A.23)中的表达式等于：

$$\begin{aligned}&P(\log m(w_{1:n}|\mathcal{M}_b) > \log m(w_{1:n}|\mathcal{M}_e)) \\ &\qquad = P\Big(\mathcal{O}_p(1)\sqrt{n} - n\mathrm{KL}(P||Q_b^*(\theta_*)) > \mathcal{O}_p(1) - \frac{d_x}{2}\log(n)\Big), \quad \text{(A.26)}\end{aligned}$$

其中，左边的$\mathcal{O}_p(1)\sqrt{n} - n\mathrm{KL}(P||Q_b^*(\theta_*))$收敛于 $-\infty$，当且仅当$\mathrm{KL}(P||Q_b^*(\theta_*)) > 0$（因为 n 中的项比 $\sqrt{n}$中的项发散得更快），而右边的项也收敛于 $-\infty$。因此，只有当$\mathrm{KL}(P||Q_b^*(\theta_*))$ $= 0$时，该不等式才以趋近于1的概率成立。这等价于$\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$（根据定理4.4证明的第一部分），由此我们证明了“当且仅当”陈述的第二部分。

□

参考文献

Andrews, D. W. K. (1999), ‘广义矩估计法的一致矩选择程序’, *Econometrica* **67**(3),543–563.

Andrews, D.W. K.andLu, B.(2001), ‘用于GMM估计的一致模型与矩选择程序及其在动态面板数据模型中的应用’, *Journalof Econometrics***101**(1), 123–164.

Berry, S., Levinsohn, J. and Pakes, A. (1995), ‘市场均衡中的汽车价格’, *Econo-metrica* **63**(4), 841–890.

Borwein, J. M. and Lewis, A. S. (1991), ‘熵类最小化问题的对偶关系’, *SIAM*控制与优化杂志**29**(2), 325–338.

Chao, J. and Phillips, P. (1998), ‘使用杰弗里斯先验对联立方程模型进行有限信息分析的后验分布’, *Journal of Econometrics* **87**(1), 49–86.

Chernozhukov, V.、Hansen, C. 和 Spindler, M.（2015），“具有大量控制变量和工具变量的线性模型中的后选择与后正则化推断”，《美国经济评论》**105**(5)，486–490。

Chernozhukov, V. 和 Hong, H.（2003），《经典估计的马尔可夫链蒙特卡罗方法》，*JournalofEconometrics* **115**(2), 293–346。

Chib, S.（1995），《Gibbs 输出的边际似然》，美国统计学会会刊 **90**(432),1313–1321。

Chib, S. 和 Greenberg, E.（1995），“理解 Metropolis-Hastings算法”，*The American Statistician***49**（4），327–335。

Chib, S. 和 Greenberg, E.（2010），“带狄利克雷过程混合误差的加性三次样条回归”，*Journal of Econometrics* **156**（2），322–336。

Chib, S. 和 Jeliazkov, I.（2001年），“根据 Metropolis-Hastings算法输出计算边际似然”，*Journal of the American Statistical Association* **96**（453），270–281。

Chib, S.、Shin, M. 和 Simoni, A.（2018），“矩条件模型的贝叶斯估计与比较”，*Journal of theAmerican Statistical Association* **113**（524），1656–1668。

Dieudonné, J.（1969），现代分析基础，Read Books。

Drèze, J. H. (1976), ‘联立方程模型的贝叶斯有限信息分析’,*Econometrica* **44**(5), 1045–1075.

Florens, J.-P. and Simoni, A. (2012), ‘工具变量回归的非参数估计：基于正则化后验分布的准贝叶斯方法’, 计量经济学杂志**170**(2), 458 – 475.

Florens, J.-P. and Simoni, A. (2016), ‘线性逆问题的正则化先验’,计量经济学理论**32**(1), 71–121.

Florens, J.-P. and Simoni, A. (2021), ‘高斯过程与贝叶斯矩估计’, 商业与经济统计杂志 **39**(2), 482–492.

Hong, H. 和 Preston, B.（2012），《贝叶斯平均、预测与非嵌套模型选择》，*Journalof Econometrics* **167**(2)，358–369。

Hong, H.、Preston, B. 和 Shum, M.（2003），《基于广义经验似然的矩条件模型选择准则》，*Econometric Theory* 第923–943页。

Hoogerheide, L., Kleibergen, F. and van Dijk, H. K. (2007), '应用于Angrist-Krueger数据的工具变量回归模型的自然共轭先验', *Journal of Econometrics***138**(1), 63–103.

Kato, K. (2013), '非参数工具变量模型的准贝叶斯分析', 统计学年刊**41**(5), 2359–2390.

Kleibergen, F. and van Dijk, H. K. (1998), '使用降秩结构的贝叶斯联立方程分析', 计量经济学理论**14**(6),701–743.

Kleibergen, F. and Zivot, E. (2003), '工具变量回归的贝叶斯与经典方法', *Journal of Econometrics***114**(1),29–72.

Kleijn, B. and van der Vaart, A. (2012), '误设条件下的伯恩斯坦-冯·米塞斯定理', 统计学电子期刊**6**,354–381.

Lehmann, E. L. 和 Casella, G.（1998），点估计理论（施普林格统计学丛书），第2版，施普林格。

Liao, Y. 和 Jiang, W.（2011），「非参数条件矩约束模型的后验一致性」，统计年鉴**39**(6)，第3003–3031页。

Newey, W. K. 和 McFadden, D.（1994），第36章：大样本估计与假设检验，计量经济学手册第4卷，爱思唯尔，第2111–2245页。

Schennach, S. M.（2005），「贝叶斯指数倾斜经验似然」，生物统计学**92**(1)，第31–46页。

Schennach, S. M.（2007），《使用指数倾斜经验似然进行点估计》，统计年鉴 **35**(2)，634–672。

Shin, M.（2014），贝叶斯广义矩估计，技术报告，宾夕法尼亚大学。

Sin, C. 和 White, H.（1996年），《选择可能设定错误的参数模型的信息准则》，*Journal of Econometrics* **71**(1-2)，207–225。

Sueishi, N.（2013），《设定错误下指数倾斜估计量的识别问题》，《经济学快报》**118**(3)，509–511。

van der Vaart, A.（2002），半参数统计学，第1781号载于《数学讲义》，施普林格，第331–457页。MR1915446。

Van der Vaart, A. W.（1998），渐近统计学，剑桥统计与概率数学丛书。

Van der Vaart, A. W. 和 Wellner, J. A.（1996年），弱收敛与经验过程，施普林格统计学丛书。

Wooldridge, J. M.（2010），横截面数据与面板数据的计量经济分析，麻省理工学院出版社。

在线附录，用于
内生性检验：基于矩的贝叶斯方法

B 与基于广义矩估计的准则的比较

有一些频率学派方法可用于我们所处情境下的模型（或矩）选择。Andrews（1999）提出了一类矩选择准则（MSC）。以下是属于这一类的一些常用准则：

$$\begin{aligned} \text{GMM-BIC} &= J_n(c) - (|c| - p)\ln n \\ \text{GMM-AIC} &= J_n(c) - 2(|c| - p) \\ \text{GMM-HQIC} &= J_n(c) - 2.01(|c| - p)\ln\ln n, \end{aligned} \tag{B.1}$$

其中， c 是矩选择向量， $|c|$ 是 c所选矩条件的数量， p是待估参数的数量，而 $J_n(c)$是使用 c 和最优加权矩阵构造的过度识别约束 J 检验统计量。与传统的贝叶斯信息准则类似，这些准则根据参数数量和施加的约束数量对模型复杂度进行惩罚。参数数量增加或约束数量减少时，模型复杂度会提高。Hong等（2003）将这一思想拓展至广义经验似然估计。

我们重新考察了正文中最初呈现的模拟实验（表3），并报告了其他频率学派方法的结果：GMM-BIC、GMM-AIC 和 GMM-HQIC。根据该表，我们得出以下观察结果。首先，所有方法都表现出模型选择一致性，即随着观测值数量增加，选择真实模型的概率趋近于1。其次，与 GMM-BIC 相比，当 ρ 接近零时，我们的方法具有更强的判别能力。第三，当 ρ 不为零（不存在内生性）时，GMM-AIC 和 GMM-HQIC 更常选出正确模型。然而，在

不存在内生性的情况下，它们似乎会过度选择具有内生性的模型。总之，在本例所考虑的数据生成过程中，基于 BETEL 的模型选择优于其他替代方法，尤其是在有限样本中。

表3：采用其他频率学派方法的表1。模型选择频率来自按第3.4节设计模拟的数据的100次重复。对于 n 与Cov(ε, u) $= \rho$的每种组合，表中条目给出在数据的100次重复中，$\mathcal{M}_e$ 的对数边际似然超过$\mathcal{M}_b$的对数边际似然的次数。BETEL的数值与正文中报告的略有不同，因为它们基于不同的模拟数据集合，即随机数种子不同。

BETEL	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
250	99	96	82	48	12	2	18	54	93	100	100
500	100	100	98	76	17	1	29	87	99	100	100
1000	100	100	100	96	46	1	46	100	100	100	100
2000	100	100	100	100	80	1	70	100	100	100	100
GMM-BIC	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
250	100	97	77	35	7	3	11	40	84	99	100
500	100	100	96	72	8	1	16	74	99	100	100
1000	100	100	100	92	29	1	25	99	100	100	100
2000	100	100	100	99	63	1	47	100	100	100	100
GMM-AIC	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
250	100	100	96	74	28	15	37	79	98	100	100
500	100	100	100	94	46	11	60	98	100	100	100
1000	100	100	100	99	71	11	76	100	100	100	100
2000	100	100	100	100	95	12	94	100	100	100	100
GMM-HQIC	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
250	100	99	85	54	17	5	20	62	94	99	100
500	100	100	100	82	23	3	31	89	100	100	100
1000	100	100	100	98	54	2	56	100	100	100	100
2000	100	100	100	100	79	1	73	100	100	100	100

C 补充蒙特卡洛示例

C.1 Monte Carlo示例1：两个潜在内生回归变量

我们现在用一个包含两个潜在内生回归变量的蒙特卡洛设计来说明上述逻辑，其中在数据生成过程中只有 x_1 是内生的。数据的生成方式如下：

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_0 + \gamma_1 z_{1i} + \varepsilon_i,$$

其中（β_1，β_2，γ_0，γ_1）=（1.0，0.8，1.0，0.6），样本量为 $n = 500$，真实内生性参数为（v_1，v_2）=（协方差（x_{1i}，ε_i），协方差（x_{2i}，ε_i））=（0.5，0）。

我们考虑四个模型，其差异在于将 x_1 与 x_2 中的哪些视为外生或内生；表 4列出了这些模型及其对数边际似然。条目 v_i 对应于矩条件$\mathbf{E}[\varepsilon$（θ）$x_i] = 0$：设定 $v_i = 0$即施加该矩（因此 x_i 为外生），而使 v_i 自由则放松该矩（因此 x_i 为内生）。四个模型都使用基于 Z_i =（x_{1i}，x_{2i}，1，z_{1i}，z_{2ai}，z_{2bi}）′的同一矩向量，约束仅通过将 v的哪些分量设为零来施加。

更明确地，令 θ =(β_1，β_2，γ_0，γ_1)，并定义回归残差

$$\varepsilon_i(\theta) = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \gamma_0 - \gamma_1 z_{1i}.$$

矩向量为

$$\mathbf{E}[\varepsilon_i(\theta) Z_i] = \begin{pmatrix} \mathbf{E}[\varepsilon_i(\theta) x_{1i}] \\ \mathbf{E}[\varepsilon_i(\theta) x_{2i}] \\ \mathbf{E}[\varepsilon_i(\theta)] \\ \mathbf{E}[\varepsilon_i(\theta) z_{1i}] \\ \mathbf{E}[\varepsilon_i(\theta) z_{2ai}] \\ \mathbf{E}[\varepsilon_i(\theta) z_{2bi}] \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

模型$\mathcal{M}_b$ 施加了 $v_1 = v_2 = 0$ （两个正交条件$\mathbf{E}[\varepsilon_i$(θ)$x_{1i}] = 0$ 和$\mathbf{E}[\varepsilon_i$(θ)$x_{2i}] =$

0），模型$\mathcal{M}_{e_2}$仅施加了 $v_2 = 0$ （放宽了 x_1 矩条件），模型$\mathcal{M}_{e_1}$仅施加了 $v_1 = 0$（放宽了 x_2 矩条件），模型$\mathcal{M}_e$则不对（v_1, v_2）作限制（放宽了两个矩条件）。

表 4：蒙特卡罗 1：对数边际似然

模型	v 设为零	内生性状态	$\log p(y \mid \mathcal{M})$
$\mathcal{M}_b$	$v_1 = 0, v_2 = 0$	x_1 外生, x_2 外生	-3147.5009
$\mathcal{M}_{e_2}$	$v_2 = 0$	x_1 内生, x_2 外生	-3121.9751
$\mathcal{M}_{e_1}$	$v_1 = 0$	x_1 外生, x_2 内生	-3147.6291
$\mathcal{M}_e$	none	x_1 内生, x_2 内生	-3122.7344

边际似然最大的是$\mathcal{M}_{e_2}$，这与真实的数据生成过程（x_1 内生， x_2 外生）相符。模型$\mathcal{M}_e$ 排名第二，但略低于前者，而$\mathcal{M}_b$ 和$\mathcal{M}_{e_1}$ 得到的支持要弱得多。排名$\mathcal{M}_{e_2} \succ \mathcal{M}_e \gg \{\mathcal{M}_b, \mathcal{M}_{e_1}\}$ </style>表明，有力证据支持将 x_1视为内生变量，而将 x_2 视为内生变量的支持较弱。

C.2 矩　蒙特卡洛示例 2：内生性与模型设定

接下来，我们考虑一种蒙特卡洛设计，以联合检验内生性和函数形式。数据生成方式如下

$$y_i = \gamma_0 + \beta_1 x_i + \beta_2 x_i^2 + \gamma_1 z_{1i} + \varepsilon_i,$$

其中 $(\gamma_0, \beta_1, \beta_2, \gamma_1) = (0。5, 1。0, 1。0, 0。8)$， $n = 500$，且 $\mathrm{Cov}(x_i, \varepsilon_i) = 0.5$。我们考虑四个模型，它们在函数形式（线性与二次）以及是否施加或放宽涉及 x 和 x^2 的正交条件方面有所不同；表 5 列出了这些模型及其对数边际似然。线性设定施加 $\beta_2 = 0$；二次设定则让 β_2 保持自由。所有模型都使用相同的工具变量向量 $Z_i = (x_i, x_i^2, 1, z_{1i}, z_{2i}, z_{2i}^2)'$，因此边际似然的差异仅反映了（i）线性与二次设定的区别（即 x^2 是否进入回归），以及（ii）施加了哪些正交约束。

令 $\varepsilon_i^L(\theta_L) = y_i - \gamma_0 - \beta_1 x_i - \gamma_1 z_{1i}$ 表示线性设定残差， $\varepsilon_i^Q(\theta_Q) =$

$y_i - \gamma_0 - \beta_1 x_i - \beta_2 x_i^2 - \gamma_1 z_{1i}$ 表示二次设定残差。相应的矩向量始终由 Z_i构成。在外生模型中，所有矩都设为零：$\mathbf{E}[\varepsilon_i(\theta) Z_i] = 0$。在内生模型中，涉及 x 和 x^2 的矩不受约束（无论回归是否包含 x^2）：

$$\mathbf{E}[\varepsilon_i(\theta) Z_i] = \begin{pmatrix} v_x \\ v_{x^2} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

其中，$v_x = 0$（分别为 $v_{x^2} = 0$）施加矩条件$\mathbf{E}[\varepsilon_i(\theta) x_i] = 0$（分别为$\mathbf{E}[\varepsilon_i(\theta) x_i^2] = 0$）。因此，$\mathcal{M}_{L,e}$ 和$\mathcal{M}_{Q,e}$ 均放宽了 x 和 x^2 正交条件（允许 $v_x \neq 0$和 $v_{x^2} \neq 0$）。

表 5：蒙特卡洛 2：对数边际似然

模型	放宽的矩条件	β_2 设为零	设定	对数 $p(y \mid \mathcal{M})$
$\mathcal{M}_{L,b}$	none	yes	线性、外生	-7426.6056
$\mathcal{M}_{Q,b}$	none	no	二次、外生	-3146.1535
$\mathcal{M}_{L,e}$	x, x^2	yes	线性、内生	-3138.3612
$\mathcal{M}_{Q,e}$	x, x^2	no	二次、内生	-3123.3361

同样，最高的边际似然被赋予真实模型 $\mathcal{M}_{Q,\ e}$。次优模型是 $\mathcal{M}_{L,\ e}$，其后是 $\mathcal{M}_{Q,\ b}$，而 $\mathcal{M}_{L,\ b}$ 则远远落后。该排名表明，在两种设定中，考虑内生性都至关重要；此外，允许包含二次项还能进一步改善拟合效果。

D ETEL后验与边际似然的计算

hood

对于基准模型 $\mathcal{M}_b$，参数为 $\theta \in \Theta$。对于扩展模型 $\mathcal{M}_e$，参数为 $\psi := (\theta, \upsilon) \in \Psi$。令 H_n 表示当前模型的 ETEL可行集。基准模型的 ETEL后验为

$$\pi^n(\theta \mid w_{1:n}, \mathcal{M}_b) \propto \pi_b(\theta)\, \widehat{q}(w_{1:n} \mid \theta, \mathcal{M}_b)\, I[\theta \in H_n], \tag{D.1}$$

其中 $\int \pi_b(\theta) d\theta = 1$。扩展模型的后验分布 $\pi^n(\psi \mid w_{1:n}, \mathcal{M}_e)$ 的定义类似，其中 $\psi = (\theta, \upsilon)$。

D.1 定制单块Metropolis–Hastings

我们使用定制的单块 Metropolis–Hastings（M-H）算法从 (D.1) 中抽样：

1. 通过数值优化计算后验众数 $\widehat{\theta}$

$$\log \pi_b(\theta) + \log \widehat{q}(w_{1:n} \mid \theta, \mathcal{M}_b),$$

满足约束 $\theta \in H_n$。

2. 计算众数处的逆海森矩阵，

$$\widehat{V} := \left[-\nabla^2_\theta \log \pi^n(\theta \mid w_{1:n}, \mathcal{M}_b)\right]^{-1}_{\theta=\widehat{\theta}},$$

并将其用作提议分布的尺度矩阵。

3. 使用多元 Student-t 密度的独立提议

$$q(\theta') = t_\kappa\left(\theta'; \widehat{\theta}, \widehat{V}\right).$$

4. 如果 $\theta' \notin H_n$，立即拒绝。

5. 若 $\theta' \in H_n$，则以概率接受

$$\alpha(\theta, \theta') = \min\left\{1,\ \frac{\widehat{q}(w_{1:n} \mid \theta', \mathcal{M}_b)\, \pi_b(\theta')\, q(\theta)}{\widehat{q}(w_{1:n} \mid \theta, \mathcal{M}_b)\, \pi_b(\theta)\, q(\theta')}\right\}.$$

(D.1) 中的可行性指示函数通过直接拒绝 H_n 之外的提议来实现。等价地，可将 ETEL 似然定义为在 H_n 之外取零，从而隐式实现步骤 4，使得当 $\theta' \notin H_n$ 时接受概率为零。 $\psi = (\theta,\ \upsilon)$ 情形的定义方式相同，只需将 θ 替换为 ψ、将 π_b 替换为 π_e，并将 $\mathcal{M}_b$ 替换为 $\mathcal{M}_e$。

D.2 对数边际似然的计算

基准模型的 ETEL 边际似然为

$$m(w_{1:n} \mid \mathcal{M}_b) = \int_{H_n} \widehat{q}(w_{1:n} \mid \theta, \mathcal{M}_b)\, \pi_b(\theta)\, d\theta = \int_{\Theta} \widehat{q}(w_{1:n} \mid \theta, \mathcal{M}_b)\, \pi_b(\theta)\, I[\theta \in H_n]\, d\theta. \quad \text{(D.2)}$$

对于任意固定的 $\theta^* \in H_n$，Chib恒等式给出

$$\log m(w_{1:n} \mid \mathcal{M}_b) = \log \widehat{q}(w_{1:n} \mid \theta^*, \mathcal{M}_b) + \log \pi_b(\theta^*) - \log \pi^n(\theta^* \mid w_{1:n}, \mathcal{M}_b). \quad \text{(D.3)}$$

我们利用 Metropolis–Hastings采样器的输出和后验众数作为 θ^*，根据该恒等式计算边际似然（Chib 和 Jeliazkov，2001年）。

D.3 替代先验设定

与正文不同，有些研究者可能更倾向于直接使用限制在可行集上的先验。通过将可行性示性函数附加到先验上，即可实现这种设定。尽管这种方法明确了先验，因为我们只在似然有定义的地方赋予 θ 先验质量，但必须接受，先验的支撑集 θ

会通过 H_n依赖于样本。

对于这一先验设定，MCMC算法基本保持不变，只有边际似然的计算有所修改，具体如下。定义可行集上的先验质量及归一化先验为

$$\begin{aligned} p_{H_n}(\mathcal{M}_b) &:= \int_{\Theta} \pi_b(\theta)\, I[\theta \in H_n]\, d\theta \\ \pi_{b,H_n}(\theta) &:= \frac{\pi_b(\theta)\, I[\theta \in H_n]}{p_{H_n}(\mathcal{M}_b)}. \end{aligned} \tag{D.4}$$

在这一先验设定下，边际似然为

$$m_{H_n}(w_{1:n} \mid \mathcal{M}_b) := \int_{H_n} \widehat{q}(w_{1:n} \mid \theta, \mathcal{M}_b)\, \pi_{b,H_n}(\theta)\, d\theta,$$

因此 $m(w_{1:n} \mid \mathcal{M}_b) = p_{H_n}(\mathcal{M}_b)\, m_{H_n}(w_{1:n} \mid \mathcal{M}_b)$ 且

$$\log m_{H_n}(w_{1:n} \mid \mathcal{M}_b) = \log m(w_{1:n} \mid \mathcal{M}_b) - \log p_H(\mathcal{M}_b). \tag{D.5}$$

因此，只有当先验在 H_n 上归一化时，才会出现 $p_{H_n}(\mathcal{M}_b)$；相对于（D.3）中无约束先验的定义，它只是一个常数偏移。

我们可以通过先验模拟来估计 $p_{H_n}(\mathcal{M}_b)$：从 $\pi_b(\theta)$ 中独立同分布地抽取$\{\tilde{\theta}^{(b)}\}_{b=1}^{B}$，并计算

$$\widehat{p}_{H_n}(\mathcal{M}_b) := \frac{1}{B} \sum_{b=1}^{B} I[\tilde{\theta}^{(b)} \in H_n].$$

E 主要结果的证明

本节证明中将使用以下记号。附加记号将在 F.1 节中引入，并用于该节的证明。省略第 i 个观测值向量中的 y_i 时，记作 $\widetilde{w}_i := (x_i', z_i')'$；若进一步省略 $z_{2,\,i}$，则记作 $\widetilde{w}_{1,\,i} := (x_i', z_{1,\,i}')'$。我们用 $\mathbf{E}_n[\cdot] := \frac{1}{n}\sum_{i=1}^{n}[\cdot]$ 表示经验均值。对于概率 Q，我们用记号 $\mathbf{E}^{Q}[\cdot]$ 表示关于 Q 的期望以及

$\mathbb{V}ar_Q$关于 Q 的方差。对于真实分布 P: $\mathbf{E}[\cdot] := \mathbf{E}^P[\cdot]$。我们采用经验过程理论中的标准记号： $\mathbb{P}_n := \mathbf{E}_n[\delta_{w_i}]$ ，其中 δ_x 是位于 x 处的狄拉克测度；并且对每个函数 f，有 $\mathbb{G}_n g := \sqrt{n}(\mathbb{P}_n f - \mathbf{E} f)$。

对于 θ的函数 $\lambda(\theta)$，定义 $\tau_i(\lambda,\theta) := \frac{e^{\lambda(\theta)' g_i(\theta)}}{\mathbf{E}_n[e^{\lambda(\theta)' g_j(\theta)}]}$, $\tau\dagger_i(\lambda,\theta) := \frac{e^{\lambda(\theta)' g_i(\theta)}}{\mathbf{E}[e^{\lambda(\theta)' g_j(\theta)}]}$, $\tau_i^{\diamond}(\lambda,\theta) := e^{\lambda(\theta)' g_i(\theta)}$，使得 $\tau_i(\widehat{\lambda},\theta) = n\widehat{p}_i(\theta)$ 且 $\tau\dagger_i(\lambda_*,\theta) = dQ^*(\theta)/dP$。我们还使用如下记号： $\Omega_*^{\diamond}(\lambda,\theta) := \mathbf{E}[\tau_i^{\diamond}(\lambda,\theta)\varepsilon_i(\theta)\,\widetilde{w}_i\widetilde{w}_i']$， $\Omega_*^{\diamond}(\theta) := \Omega_*^{\diamond}(\lambda_*,\theta)$， $\Omega_*^{\dagger}(\theta) := \mathbf{E}[\tau_i^{\dagger}(\lambda_*,\theta)\varepsilon_i(\theta)\,\widetilde{w}_i\widetilde{w}_i'] = \mathbf{E}^{Q^*}(\theta)[\varepsilon_i(\theta)\,\widetilde{w}_i\widetilde{w}_i']$， 以及$\Omega_\circ := \mathbf{E}[\varepsilon_i(\theta_\circ)^2\widetilde{w}_i\widetilde{w}_i']$。此外， $\Omega_* \equiv \Omega_*^{\diamond}(\theta_\circ)$。

F 有用辅助结果的证明

F.1 符号说明

下文中，我们省略矩函数中的下标“b”和“e”，对于基准模型和扩展模型，均简记为$g(w,\theta)$。我们用 $g_i(\theta) := g(w_i,\theta)$表示在 w_i 处取值的矩$\widehat{\quad}$ $\widehat{\quad}$函数，并用 $\lambda(\theta) := \lambda(w_{1:n},\theta)$表示倾斜参数。此外， $\varepsilon_i := \varepsilon_i(\theta_\circ)$表示在真实参数处取值的模型误差项。当从第 i个观测值的向量中省略 y_i 时，我们记作 $\widetilde{w}_i := (x_i', z_i')'$；若还省略$z_{2,i}$ ，则记作$\widetilde{w}_{1,i} := (x_i', z_{1,i}')'$。我们用 $\widetilde{v} := (v', 0)$表示由 d_x-辅助参数和一个 d_z维零向量组成的向量。对于每个 $\theta \in \mathbb{R}^p$，令 $h := \sqrt{n}(\theta - \theta_*)$，并用 $\pi_h^n(\cdot|w_{1:n})$表示其后验分布。

我们用记号 $\mathbf{E}_n[\cdot] := \frac{1}{n}\sum_{i=1}^n[\cdot]$ 表示经验均值。此外， $\widehat{g}(\theta) := \mathbf{E}_n[g_i(\theta)]$,$dg_i(\theta)/d\theta' = -\widetilde{w}_i\widetilde{w}_{1,i}'$。对于概率 Q，我们用记号 $\mathbf{E}^Q[\cdot]$ 表示相对于 Q的期望，并用 $\mathbb{V}ar_Q$表示相对于 Q的方差。对于真实分布 P： $\mathbf{E}[\cdot] := \mathbf{E}^P[\cdot]$。单个观测值 w_i 的对数似然函数记为$\ell_{n,\theta}(w_i)$：

$$\ell_{n,\theta}(w_i) := \log \widehat{q}_i(\theta) = \log \frac{e^{\widehat{\lambda}(\theta)' g(w_i,\theta)}}{\sum_{k=1}^n e^{\widehat{\lambda}(\theta)' g_k(\theta)}} = -\log n + \log \frac{e^{\widehat{\lambda}(\theta)' g(w_i,\theta)}}{\frac{1}{n}\sum_{k=1}^n e^{\widehat{\lambda}(\theta)' g_k(\theta)}}$$

因此，对数 ETEL 函数为 $\ell_{n,\theta}(w_{1:n}) = \sum_{i=1}^n \ell_{n,\theta}(w_i) = \log \prod_{i=1}^n \widehat{q}_i(\theta) = \log\ \widehat{q}(w_{1:n}|\theta)$.

$\widehat{}$将 $\lambda(\theta)$替换为其真值 $\lambda_*(\theta)$，我们定义：

$$\ell_{*,\theta}(w) := \log \frac{e^{\lambda_*(\theta)' g(w,\theta)}}{\sum_{k=1}^n e^{\lambda_*(\theta)' g_k(\theta)}} =: \log p_w^*(\theta) \qquad \text{and} \qquad \ell_{*,\theta}(w_{1:n}) := \sum_{i=1}^n \ell_{*,\theta}(w_i).$$

$\dot{}\theta \mapsto \ell_{n,\theta}(w_{1:n})\,\theta_1$ 处的一阶（分别为二阶）导数记作$\ell_{n,\theta_1}(w_{1:n})$（分别为$\ddot{\ell}_{n,\theta_1}(w_{1:n})$）。此外，对于关于 θ的函数 $\lambda(\theta)$，定义 $\tau_i(\lambda,\theta) := \frac{e^{\lambda(\theta)' g_i(\theta)}}{\mathrm{E}_n[e^{\lambda(\theta)' g_j(\theta)}]}$, $\tau\dagger_i(\lambda,\theta) := \frac{e^{\lambda(\theta)' g_i(\theta)}}{\mathrm{E}_{[e^{\lambda(\theta)' g_j(\theta)}]}}$, $\tau_i^\diamond(\lambda,\theta) := e^{\lambda(\theta)' g_i(\theta)}$。因此，$\tau_i(\widehat{\lambda},\theta) = n\widehat{p}_i(\theta)$，且 $\tau\dagger_i(\lambda_*,\theta) = dQ^*(\theta)/dP$。

我们还使用以下记号：$\check{\Omega}(\lambda,\theta) := \mathrm{E}_n[\tau_i(\lambda,\theta)\varepsilon_i(\theta)^2\widetilde{w}_i\widetilde{w}_i']$，$\check{\Omega}^\diamond(\lambda,\theta) := \mathrm{E}_n[\tau_i^\diamond(\lambda,\theta)\varepsilon_i(\theta)^2\widetilde{w}_i\widetilde{w}_i']$，$\check{\Omega}^\dagger(\lambda,\theta) := \mathrm{E}_n[\tau_i^\dagger(\lambda,\theta)\varepsilon_i(\theta)^2\widetilde{w}_i\widetilde{w}_i']$，$\Omega_*^\diamond(\lambda,\theta) := \mathrm{E}[\tau_i^\diamond(\lambda,\theta)\varepsilon_i(\theta)^2\widetilde{w}_i\widetilde{w}_i']$，$\Omega_*^\diamond(\theta) := \Omega_*^\diamond(\lambda_*,\theta)$，以及$\Omega_*^\dagger(\theta) := \mathrm{E}[\tau_i^\dagger(\lambda_*,\theta)\varepsilon_i(\theta)^2\widetilde{w}_i\widetilde{w}_i'] = \mathrm{E}^{Q^*(\theta)}[\varepsilon_i(\theta)^2\widetilde{w}_i\widetilde{w}_i']$。此外，$\Omega_* \equiv \Omega_*^\diamond(\theta_\circ)$且$\Omega_\circ := \mathrm{E}[\varepsilon_i(\theta_\circ)^2\widetilde{w}_i\widetilde{w}_i']$。对于扩展模型，我们使用记号$\Omega_{\psi_0} := \mathrm{E}[g_e(w_i,\psi_\circ)g_e(w_i,\psi_\circ)'] = \mathrm{E}[\varepsilon_i^2\widetilde{w}_i\widetilde{w}_i'] - \widetilde{v}_\circ\widetilde{v}_\circ'$。

令 $\|\cdot\|_2$表示欧几里得范数，$\|\cdot\|_{op}$表示算子范数。我们用记号 C表示一个任意的正数常量。我们将以向量 a 为中心、半径为 r_n的闭球记为 $B_{r_n}(a)$：$B_{r_n}(a) := \{b;\ \|b-a\|_2 \le Cr_n\}$。我们将 Θ_n 记为：$= \{\|\theta-\theta_*\| \le M_n/\sqrt{n}\}$。我们用 $L_2(P)$ 表示关于 P 平方可积函数的空间，用 $\|\cdot\|_{P,2}$ 表示 $L_2(P)$ 空间中的范数。全变差范数记为 $\|\cdot\|_{TV}$。最后，令 CS、T、J、MVT、CMT 和 DCT 分别表示柯西-施瓦茨不等式、三角不等式、詹森不等式、中值定理、连续映射定理和控制收敛定理。

对于集合 $\mathcal{A} \subset \mathbb{R}^m$，我们将其相对于 $\mathbb{R}^m$的内部记为 $int(\mathcal{A})$。我们采用经验过程理论中的标准记号：$\mathbb{P}_n := \mathbf{E}_n[\delta_{w_i}]$，其中 δ_x 是位于 x处的狄拉克测度，并且对每个函数 f，$\mathbb{G}_n g := \sqrt{n}(\mathbb{P}_n f - \mathbf{E}f)$。需要时还会引入其他记号。

此外，我们还使用了以下恒等式，这些恒等式是在下文的一些假设下由引理 F.1 和 F.2 推导得出的：$\forall\theta \in B_{*,\,n}$，

$$\frac{d\widehat{\lambda}(\theta)'}{d\theta} = -\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\theta)\frac{dg_i(\theta)'}{d\theta}(I + \widehat{\lambda}(\theta)g_i(\theta)')\right]\check{\Omega}(\widehat{\lambda},\theta)^{-1}$$
$$= \mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda},\theta)\widetilde{w}_{1,i}\widetilde{w}_i'(I + \widehat{\lambda}(\theta)g_i(\theta)')\right]\check{\Omega}^\diamond(\widehat{\lambda},\theta)^{-1}, \quad \text{(F.1)}$$

and

$$\frac{d\lambda_*(\theta)'}{d\theta} = \mathbf{E}^{Q^*(\theta)}\left[\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta)g_i(\theta)')\right]\Omega_*^\dagger(\theta)^{-1}$$
$$= \mathbf{E}\left[e^{\lambda_*(\theta)'g_i(\theta)}\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta)g_i(\theta)')\right]\left(\mathbf{E}[e^{\lambda_*(\theta)'g_i(\theta)}\varepsilon_i(\theta)^2\widetilde{w}_i\widetilde{w}_i']\right)^{-1}, \quad \text{(F.2)}$$

分别为 $\frac{dg_i(\theta')}{d\theta} = \widetilde{w}_{1,i}\widetilde{w}_i'$。

最后，在假设1下，θ_* 的一阶条件为

$$\frac{d\lambda_*(\theta_*)'}{d\theta}\mathbf{E}[g_i(\theta_*)] - \mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\lambda_*(\theta_*) + \mathbf{E}[\tau_i^\dagger(\lambda_*,\theta_*)\widetilde{w}_{1i}\widetilde{w}_i']\lambda_*(\theta_*) = 0, \quad \text{(F.3)}$$

并且$\mathbf{E}[\tau_i(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i] = 0$，因为这是 λ_*的一阶条件。

在假设 1 和假设 3 下，$\ell_{*,\theta}(w_{1:n})$关于 θ的一阶导数，记为$\dot{\ell}_{*,\theta_*}(w_{1:n})$，由下式给出：$\forall\theta \in B_{*,n}$:

$$\frac{1}{n}\dot{\ell}_{*,\theta}(w_{1:n}) = \frac{d\lambda_*(\theta)'}{d\theta}\mathbf{E}_n[(1-\tau_i(\lambda_*,\theta))g_i(\theta)] + \mathbf{E}_n\left[(1-\tau_i(\lambda_*,\theta))\widetilde{w}_{1,i}\widetilde{w}_i'\right]'\lambda_*(\theta)$$

其概率随着 $n \to \infty$趋近于1。

F.2 基础模型和扩展模型的随机局部渐近正态性（LAN）

现在我们给出三个定理，以证明基础模型和扩展模型具有随机局部渐近正态性。每个定理之后给出相应的证明。这些证明是全新的，与 Chib 等人（2018）中类似结果的证明有实质性不同。

定理 F.1（基准模型中的随机局部渐近正态性。） 设 V_{θ_*} 为一个正定矩阵，其表达式见定理 *F.6* 的陈述。假设 *1–6* 成立。对于每个 $\theta \in \mathbb{R}^p$，令 $h := \sqrt{n}(\theta - \theta_*)$。那么，对于每个以零为中心、半径为 $M_n \to \infty$ 的闭球 $K_n \subset \mathbb{R}^p$

使得 $M_n = o\left(\sqrt{n}\right)^{-}$,

$$\sup_{h\in K_n}\left|\log\frac{\widehat{q}(w_{1:n}|\theta_*+h/\sqrt{n})}{\widehat{q}(w_{1:n}|\theta_*)}-h'V_{\theta_*}^{-1}\Delta_{n,\theta_*}+\frac{1}{2}h'V_{\theta_*}^{-1}h\right|\xrightarrow{p}0\qquad as\ n\to\infty, \tag{F.4}$$

其中， θ_* 如(3.4)所定义，$h'V_{\theta_*}^{-1}\Delta_{n,\ \theta_*}:=\frac{h'}{\sqrt{n}}\dot{\ell}_{n,\ \theta_*}(w_{1:n})\ \xrightarrow{d}\boldsymbol{N}(0,\ h'H_*h)$ 依概率有界，而 H_* 是下文引理 *F.5* 中定义的正定矩阵。

证明。 首先，紧集 $K_n:=\{h\in\mathbb{R}^p;\|h\|\le M_n\}$ 使得相应的 $\theta:=\theta_*+h/\sqrt{n}$ 属于$B_{*,\ n}$ 因此，根据假设 3，存在一个 $N\ge 1$，使得对于每个$n>N$，对数 ETEL 函数$\ell_{n,\ \theta}(w_{1:n})$在$B_{*,\ n}$上均有定义。

我们对$\theta\mapsto\ell_{n,\ \theta}(w_{1:n})$在 θ_*附近应用二阶中值定理展开，利用 $\widehat{\lambda}(\theta)$的一阶条件，即$\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta)'g_i(\theta)}g_i(\theta)\right]=0$，以及引理 F.1（该引理保证 $\theta\mapsto\widehat{\lambda}(\theta)$在$B_{*,\ n}$上为$\mathcal{C}^2$），得到：
$\forall\theta\in B_{*,\ n}$，其成立概率趋近于 1，

$$\begin{aligned}
&\ell_{n,\theta}(w_{1:n})-\ell_{n,\theta_*}(w_{1:n})=(\theta-\theta_*)'\dot{\ell}_{n,\theta_*}(w_{1:n})+\frac{1}{2}(\theta-\theta_*)'\ddot{\ell}_{n,\widetilde{\theta}}(w_{1:n})(\theta-\theta_*)\\
&=(\theta-\theta_*)'\frac{d\widehat{\lambda}(\theta_*)'}{d\theta}n\widehat{g}(\theta_*)+(\theta-\theta_*)'\frac{d\widehat{g}(\theta_*)'}{d\theta}\widehat{\lambda}(\theta_*)n+\frac{n}{2}(\theta-\theta_*)'\frac{d^2[\widehat{\lambda}(\widetilde{\theta})'\widehat{g}(\widetilde{\theta})]}{d\theta d\theta'}(\theta-\theta_*)\\
&-n(\theta-\theta_*)'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\theta_*)\frac{dg_i(\theta_*)'}{d\theta}\right]\widehat{\lambda}(\theta)-\frac{n}{2}(\theta-\theta_*)'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\frac{dg_i(\widetilde{\theta})'}{d\theta}\right]\frac{d\widehat{\lambda}(\theta)}{d\theta'}(\theta-\theta_*)\\
&-\frac{n}{2}(\theta-\theta_*)'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\frac{dg_i(\widetilde{\theta})'}{d\theta}\widehat{\lambda}(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\frac{dg_i(\widetilde{\theta})}{d\theta'}\right](\theta-\theta_*)\\
&-\frac{n}{2}(\theta-\theta_*)'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\theta_*)'\frac{dg_i(\widetilde{\theta})}{d\theta'}\right](\theta-\theta_*)\\
&+\frac{n}{2}(\theta-\theta_*)'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\frac{dg_i(\widetilde{\theta})'}{d\theta}\right]\widehat{\lambda}(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\frac{dg_i(\widetilde{\theta})}{d\theta'}\right](\theta-\theta_*)
\end{aligned}$$

$\widetilde{\ }\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad-\qquad\widetilde{\ }$对于 $\theta=\tau\theta+(1-\tau)\theta_*$以及某些 $\tau\in[0,1]$。将 θ替换为 $\theta_*+h/\sqrt{n}$，使得 $\theta=\theta_*+\tau h/\sqrt{n}$，并利用 $g_i(\theta)$ 的表达式及其关于 θ 的一阶导数，前述表达式可简化为： $\forall h\in K_n$,

$$\ell_{n,\theta_*+h/\sqrt{n}}(w_{1:n})-\ell_{n,\theta_*}(w_{1:n})=h'\frac{d\widehat{\lambda}(\theta_*)'}{d\theta}\sqrt{n}\widehat{g}(\theta_*)-h'\sqrt{n}\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']\widehat{\lambda}(\theta_*)+\frac{1}{2}h'\frac{d^2[\widehat{\lambda}(\widetilde{\theta})'\widehat{g}(\widetilde{\theta})]}{d\theta d\theta'}h$$

$$+\sqrt{n}h'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*)+\frac{1}{2}h'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\widetilde{w}_{1,i}\widetilde{w}_i'\right]\frac{d\widehat{\lambda}(\widetilde{\theta})}{d\theta'}h$$
$$-\frac{1}{2}\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]+\frac{1}{2}h'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h$$
$$+\frac{1}{2}h'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\widetilde{w}_i\widetilde{w}_{1,i}'\right]h. \quad \text{(F.5)}$$

首先考虑 (F.5) 中的一阶项，并在此基础上加上再减去 (F.3) 中给出的关于 θ_* 的一阶条件：$\forall h\in K_n$,

$$h'\frac{d\widehat{\lambda}(\theta_*)'}{d\theta}\sqrt{n}\widehat{g}(\theta_*)-h'\sqrt{n}\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']\widehat{\lambda}(\theta_*)+\sqrt{n}h'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta)=$$
$$h'\sqrt{n}\left(\frac{d\widehat{\lambda}(\theta_*)'}{d\theta}\widehat{g}(\theta_*)-\frac{d\lambda_*(\theta_*)'}{d\theta}\mathbf{E}[g_i(\theta_*)]\right)-h'\sqrt{n}\left(\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']\widehat{\lambda}(\theta_*)-\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\lambda_*(\theta_*)\right)$$
$$+\sqrt{n}h'\left(\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}_*(\theta)-\mathbf{E}[\tau_i^\dagger(\lambda_*,\theta_*)\widetilde{w}_{1i}\widetilde{w}_i']\lambda_*(\theta_*)\right)=:h'V_{\theta_*}^{-1}\Delta_{n,\theta_*}.$$

根据下文的引理 F.5，$h'V_{\theta_*}^{-1}\Delta_{n,\ \theta_*}$ 是渐近正态的，均值为零，方差等于非奇异矩阵 $h'H_*h$，其表达式在引理 F.5 的陈述中给出。

现在，让我们考虑（F.5）中的二阶项。首先，由引理 F.12，$h'\frac{d^2[\widehat{\lambda}(\widetilde{\theta})'\widehat{g}(\widetilde{\theta})]}{d\theta d\theta'}h$ 在 $h\in K_n$上一致地依概率有界。因为 $h/\sqrt{n}\to 0$，由于 $\|h\|\leq M_n$ 以及 $M_n=o(\sqrt{n})$，于是当 $n\to\infty$时，$\theta\to\theta_*$ 在 $h\in K_n$ 上一致成立。此外，我们使用当 $n\to\infty$时的如下极限：（1）由 $\theta\mapsto\lambda_*(\theta)$的连续性（根据补充材料中的引理 F.2），以及 $\theta\mapsto g_i(\theta)$的连续性，我们有：$\lambda_*(\theta)\to\lambda_*(\theta_*)$，以及 $g_i(\theta)\to g_i(\theta_*)$。（2）由引理 F.9，则 $\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\widetilde{w}_{1,i}\widetilde{w}_i'\right]$ 当 $n\to\infty$时在 $h\in K_n$上一致地依概率收敛于 $\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]$。（3）由引理 F.10，则 $\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]$ 当 $n\to\infty$时在 $h\in K_n$上一致地依概率收敛于 $\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)(h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*))^2\right]$。（4）由引理 F.11，则

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\varepsilon_i(\widetilde{\theta})'\widetilde{w}_i'\right]\xrightarrow{p}\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\varepsilon_i(\theta_*)'\widetilde{w}_i'\right]$$

$\widehat{\ }\ \widetilde{}$在（$\lambda$（$\theta$） h）$\in B_{1/\sqrt{n}}$（λ_*（θ_*））$\times K$中一致。结合上述（2）和（4）中的收敛，

~~引理 F.6，以及 $\frac{d\lambda}{d\theta}(\theta')$ 的表达式，我们有 $\forall \theta \in B_{*,n}$

$$h' \frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta} \xrightarrow{p} h' \mathbf{E}^{Q^*(\theta_*)}[\widetilde{w}_{1,i}\widetilde{w}_i'(I + \lambda_*(\theta_*)g_i(\theta_*)')]\Omega_*^{\diamond}(\theta_*)^{-1}$$

~~在 $(\lambda(\theta), h) \in B_{1/\sqrt{n}}(\lambda_*(\theta_*)) \times K$ 上一致收敛。因此，利用这些极限，(F.5) 中的二阶项等于：

$$\begin{aligned}
&\frac{1}{2}h' \frac{d^2\lambda_*(\theta_*)}{d\theta d\theta'} \mathbf{E}[g_i(\theta_*)]h - h' \frac{d\lambda_*(\theta_*)'}{d\theta} \mathbf{E}\left[\widetilde{w}_i \widetilde{w}_{1,i}'\right] h + \frac{1}{2}h' \mathbf{E}\left[\tau_i^\dagger(\lambda_*, \theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'\right] \frac{d\lambda_*(\theta_*)}{d\theta'}h \\
&\quad - \frac{1}{2}\mathbf{E}\left[\tau_i^\dagger(\lambda_*, \theta_*)\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\right)^2\right] + \frac{h'}{2}\frac{d\lambda_*(\theta_*)'}{d\theta}\mathbf{E}\left[\tau_i^\dagger(\lambda_*, \theta_*)g_i(\theta_*)\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'\right] h \\
&\quad + \frac{1}{2}h'\mathbf{E}\left[\tau_i^\dagger(\lambda_*, \theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\lambda_*(\theta_*)'\mathbf{E}\left[\tau_i^\dagger(\lambda_*, \theta_*)\widetilde{w}_i\widetilde{w}_{1,i}'\right] h + o_p(1). \quad \text{(F.6)}
\end{aligned}$$

其中 $o_p(1)$ 在 $(\widehat{\lambda}(\widetilde{\theta}), h) \in B_{1/\sqrt{n}}(\lambda_*(\theta_*)) \times K$ 上一致收敛。注意到 $\tau_i^\dagger(\lambda_*, \theta_*) = dQ^*(\theta_*)/dP$，且对每个 $l > d_x$，$\mathbf{E}[g_{i,l}(\theta_*)] = 0$，前面的表达式可简化为

$$\begin{aligned}
&\frac{1}{2}h' \frac{d^2\lambda_*(\theta_*)}{d\theta d\theta'} \mathbf{E}[g_i(\theta_*)]h - h' \frac{d\lambda_*(\theta_*)'}{d\theta} \mathbf{E}\left[\widetilde{w}_i \widetilde{w}_{1,i}'\right] h - \frac{1}{2}h' \mathbb{V}ar_{Q^*(\theta_*)}\left[\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\right] h \\
&\quad + \frac{1}{2}h'\mathbf{E}^{Q^*(\theta_*)}\left[\widetilde{w}_{1,i}\widetilde{w}_i'\left(I + \lambda_*(\theta_*)g_i(\theta_*)'\right)\right]\frac{d\lambda_*(\theta_*)}{d\theta'}h + o_p(1) =: -\frac{1}{2}h'V_{\theta_*}^{-1}h + o_p(1).
\end{aligned}$$

将这些要素综合起来，可得：

$$\ell_{n,\theta_*+h/\sqrt{n}}(w_{1:n}) - \ell_{n,\theta_*}(w_{1:n}) = h'V_{\theta_*}^{-1}\Delta_{n,\theta_*} - \frac{1}{2}h'V_{\theta_*}^{-1}h + o_p(1),$$

其中 $h'V_{\theta_*}^{-1}\Delta_{n,\theta_*} \xrightarrow{d}$ 正态分布$(0, h'H_*h)$，$\frac{h'}{\sqrt{n}}\dot{\ell}_{n,\theta_*} = h'V_{\theta_*}^{-1}\Delta_{n,\theta_*}$ 且 $V_{\theta_*}^{-1} = \text{plim}\ddot{\ell}_{n,\theta_*}/n + o_p(1)$。因此，我们得到了定理的结论。

□

定理 F.2（外生性下基准模型中的随机局部渐近正态性）假设假设 *2-6* 成立，且对于每个 $\theta \in \mathbb{R}^p$，令 $h := \sqrt{n}(\theta - \theta_\circ)$。那么，对于每个以零为中心的闭球 $K_n \subset \mathbb{R}^p$

半径为 $M_n \to \infty$，使得 $M_n = o(\sqrt{n})$，

$$\sup_{h \in K_n} \left| \log \frac{\widehat{q}(w_{1:n}|\theta_\circ + h/\sqrt{n})}{\widehat{q}(w_{1:n}|\theta_\circ)} - h' V_{\theta_\circ}^{-1} \Delta_{n,\theta_\circ} + \frac{1}{2} h' V_{\theta_\circ}^{-1} h \right| \xrightarrow{p} 0 \qquad as\ n \to \infty \tag{F.7}$$

其中 $\theta_\circ$ 是 θ 的真值，$V_{\theta_\circ}^{-1} := \mathbf{E}\left[\widetilde{w}_{1,i}\widetilde{w}_i'\right]\left(\mathbf{E}[\varepsilon_i^2 \widetilde{w}_i \widetilde{w}_i']\right)^{-1} \mathbf{E}\left[\widetilde{w}_i \widetilde{w}_{1,i}'\right]$ 假设其为非奇异的，并且 $V_{\theta_\circ}^{-1}\Delta_{n,\theta_\circ} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{E}\left[\widetilde{w}_{1,i}\widetilde{w}_i'\right]\left(\mathbf{E}[\varepsilon_i^2 \widetilde{w}_i \widetilde{w}_i']\right)^{-1} \varepsilon_i \widetilde{w}_i$ 依概率有界。

证明。 首先，紧集 $K_n := \{h \in \mathbb{R}^p; \|h\| \le M_n\}$ 满足相应的 $\theta := \theta_\circ + h/\sqrt{n}$ 属于 $B_{\circ,n}$ 因此，根据假设 3，存在一个 $N \ge 1$，使得对于每个$n > N$，对数 ETEL 函数 $\ell_{n,\theta}(w_{1:n})$ 在 $B_{\circ,n}$ 上均有定义。

证明过程与上文定理 F.1 的证明相同，但现在在真实的 $\theta_\circ$ 处取值的倾斜参数为零：$\lambda_*(\theta_\circ) = 0_\circ$ 此外，$\mathbf{E}[g_i(\theta_\circ)] = 0$, $\tau_i(\lambda_*, \theta_\circ) = 1$ 且 $\frac{d\lambda_*(\theta_\circ)}{d\theta'} = \mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i'](\mathbf{E}[\varepsilon_i^2 \widetilde{w}_i \widetilde{w}_i'])^{-1}$。因此，定理 F.1 证明中的一阶项得以简化。为说明这一点，我们将各项分别处理。先看这一项：$h' \frac{d\lambda(\theta_\circ)}{d\theta'}$。利用 (F.1) 中 $\frac{d\lambda_{(\theta_\circ)}}{d\theta}$ 的表达式，可得如下分解：

$$\begin{aligned} h' \frac{d\widehat{\lambda}(\theta_\circ)'}{d\theta} \breve{\Omega}^\diamond(\widehat{\lambda}, \theta_\circ) - h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i'] &= h'\mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda}, \theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'(I + \widehat{\lambda}(\theta_\circ) g_i(\theta_\circ)')\right] - h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i'] \\ &= h'\mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda}, \theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'\right] - h'\mathbf{E}[\tau_i^\diamond(\widehat{\lambda}, \theta_\circ)\widetilde{w}_{1i}\widetilde{w}_i'] \\ &\quad + h'\mathbf{E}[(\tau_i^\diamond(\widehat{\lambda}, \theta_\circ) - 1)\widetilde{w}_{1i}\widetilde{w}_i'] + h'\mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda}, \theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\theta_\circ) g_i(\theta_\circ)'\right]. \end{aligned} \tag{F.8}$$

根据引理 F.3，对于每个 $\eta > 0$，都存在有限的 C 和有限的 $N(C, \eta)$，使得对于每个 $n > N(C, \eta)$，$\|\lambda(\theta_\circ) - \lambda_*(\theta_\circ)\|_2 < C/\sqrt{n}$ 的概率至少为 $\eta_\circ$ 因此，(F.8) 右边前两项满足以下上界：

$$\begin{aligned} &\sup_{h \in K_n} \left\| h'\mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda}, \theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i\right] - h'\mathbf{E}[\tau_i^\diamond(\widehat{\lambda}, \theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i] \right\| \\ &\qquad \le \sup_{h \in K_n} \sup_{\lambda \in B_{1/\sqrt{n}}(0)} \left\| h'\mathbf{E}_n\left[e^{\lambda' g_i(\theta_\circ)}\widetilde{w}_{1,i}\widetilde{w}_i\right] - h'\mathbf{E}[e^{\lambda' g_i(\theta_\circ)}\widetilde{w}_{1i}\widetilde{w}_i] \right\| \end{aligned}$$

根据紧性，该表达式在假设 6 *(c)* 下收敛于零（其中 $(j, \ell, \ell') = (1, 1, 1)$），

$\widetilde{\ }B_{1/\sqrt{n}}(0)$，并由 Newey 和 McFadden（1994，引理 2.4）得出。接下来，我们控制 (F.8) 中的项 $h'\mathrm{E}[(\tau_i^{\diamond}(\lambda,\theta_\circ)-\widehat{\ }1)\widetilde{w}_{1i}\widetilde{w}_i]$。根据连续映射定理和引理 F.3：对每个 $i=1$，……，n，$\sup_{h\in K_n}\|e^{\lambda(\theta_\circ)'g_i(\theta_\circ)}h'\widetilde{w}_{1,i}\widetilde{w}_i-h'\widetilde{w}_{1,i}\widetilde{w}_i\|\xrightarrow{p}0$。由此以及控制收敛定理可知，在假设 6 (c) 下（其中 $(j,\ell,\ell')=(1,1,1)$），$\mathrm{E}[(\tau_i^{\diamond}(\widehat{\lambda},\theta_\circ)-1)h'\widetilde{w}_{1i}\widetilde{w}_i']\xrightarrow{p}0$ 关于 $h\in K_n$ 一致成立。(F.8) 右边的最后一项可通过使用上界以类似方式处理：

$$\sup_{h\in K_n}\left\|h'\mathbf{E}_n\left[\tau_i^{\diamond}(\widehat{\lambda},\theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\theta_\circ)g_{i,}(\theta_\circ)\right]\right\|$$
$$\leq\sup_{h\in K_n}\sup_{\lambda\in B_{1/\sqrt{n}}(0)}\|h'\mathbf{E}_n\left[\tau_i^{\diamond}(\lambda,\theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'\lambda g_i(\theta_\circ)\right]-h'\mathbf{E}\left[\tau_i^{\diamond}(\lambda,\theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'\lambda g_i(\theta_\circ)\right]\|$$
$$+\sup_{h\in K_n}\left\|h'\mathbf{E}\left[\tau_i^{\diamond}(\widehat{\lambda},\theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\theta_\circ)g_i(\theta_\circ)\right]\right\|.$$

在假设 6*(f)*、Newey 和 McFadden（1994，引理 2.4）、引理 F.3、连续映射定理、支配收敛定理以及 $\lambda_*(\theta_\circ)=0$ 这一事实的条件下，前一表达式右边的两项依概率收敛于零。综合以上各项，并利用引理 F.6 控制表达式 $\frac{d\lambda}{d\theta}(\theta_\circ)'$ 中的 $\widehat{\ }\check{\ }\Omega^{\diamond}(\widehat{\lambda},\theta_\circ)$，可得 $h'\frac{d\lambda(\theta_\circ)}{d\theta}'\sqrt{n}\widehat{g}(\theta_\circ)$ 等于 $h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\mathbf{E}[\varepsilon_i^2\widetilde{w}_i\widetilde{w}_i']^{-1}\mathbb{G}_n[g(w_i,\theta_\circ)]+o_p(1)$，其中 $o_p(1)$ 项关于 $h\in K_n$ 一致。

接下来，我们分析定理 F.1 证明中其他的一阶项，即 $-h'\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']\sqrt{n}\lambda(\theta_\circ)+h'\mathbf{E}_n[\tau_i(\lambda,\theta_\circ)\widetilde{w}_{1i}\widetilde{w}_i']\sqrt{n}\lambda(\theta_\circ)$。乘在 $\sqrt{n}\lambda(\theta_\circ)$ 前面的因子可分解如下：

$$-h'\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']+h'\mathbf{E}_n[\tau_i(\widehat{\lambda},\theta_\circ)\widetilde{w}_{1i}\widetilde{w}_i']=$$
$$=h'\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'-\mathbf{E}\left(\tau_i(\widehat{\lambda},\theta_\circ)\widetilde{w}_{1,i}\widetilde{w}_i'\right)\right]+\mathbf{E}\left[\left(\tau_i(\widehat{\lambda},\theta_\circ)-1\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]$$
$$-\mathbf{E}_n\left[h'\widetilde{w}_{1,i}\widetilde{w}_i'-\mathbf{E}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\right)\right].\quad\text{(F.9)}$$

前一个表达式右边的前两项与 F.8 中已证明一致收敛于零的前三项中的前两项相同。在假设 7 *(c)* 下，根据一致大数定律，最后一项收敛于零。接下来，由于根据引理 F.4，$\sqrt{n}\widehat{\lambda}(\theta_\circ)=-\mathbf{E}[\varepsilon_i^2\widetilde{w}_i\widetilde{w}_i']^{-1}\mathbb{G}_n[\varepsilon_i\widetilde{w}_i]+o_p(1)$ 且 $\mathbb{G}_n[\varepsilon_i\widetilde{w}_i]\xrightarrow{d}$ 正态分布(0，$\mathbf{E}[\varepsilon_i^2\widetilde{w}_i\widetilde{w}_i']$)，

那么 $\sqrt{n}\widehat{\lambda}(\theta_\circ) = O_p(1)$，因此，在假设 6 *(c)* 下，$h'\mathbf{E}_n[(\tau_i(\widehat{\lambda}, \theta_\circ) - 1)\widetilde{w}_{1i}\widetilde{w}_i']\sqrt{n}\widehat{\lambda}(\theta_\circ) \xrightarrow{p} 0$，其中 $(j, \ell, \ell') = (1, 1, 1)$。

综合以上所有要素，我们得到定理 F.1 的证明中的一阶项等于以下简化表达式：$\forall h \in K_n$,

$$h'\frac{d\widehat{\lambda}(\theta_\circ)'}{d\theta}\sqrt{n}\widehat{g}(\theta_\circ) - h'\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']\sqrt{n}\widehat{\lambda}(\theta_\circ) + h'\mathbf{E}_n[\tau_i(\widehat{\lambda}, \theta_\circ)\widetilde{w}_{1i}\widetilde{w}_i']\sqrt{n}\widehat{\lambda}(\theta_\circ)$$
$$= h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\mathbf{E}[\varepsilon_i^2\widetilde{w}_i\widetilde{w}_i']^{-1}\mathbb{G}_n[g(w_i, \theta_\circ)] + o_p(1)$$

其中，$o_p(1)$ 项关于 $h \in K_n$ 是一致的。

最后，由于 $\mathbb{G}_n[g(w_i, \theta_\circ)] \xrightarrow{d}$ 正态分布$(0, \mathbf{E}[\varepsilon_i(\theta_\circ)^2\widetilde{w}_i\widetilde{w}_i'])$，根据假设 4 *(a)* 的林德伯格-列维中心极限定理，我们得出前一项对每个 $h \in K_n$ 都收敛于正态分布$(0, h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\mathbf{E}[\varepsilon_i(\theta_\circ)^2\widetilde{w}_i\widetilde{w}_i']^{-1}\mathbf{E}[\widetilde{w}_i\widetilde{w}_{1i}']h)$ 分布。

现在，我们来考察 (F.5) 中的二阶项，其中 $\theta := \theta_\circ + \tau h/\sqrt{n} \in B_{\circ,\, n}$ 对于某个 $\tau \in [0, 1]$ 和 $h \in K_n$：

$$\frac{1}{2}h'\sum_{l=1}^{d}\frac{d^2\widehat{\lambda}_l(\widetilde{\theta})}{d\theta d\theta'}\widehat{g}_l(\widetilde{\theta})h - h'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\mathbf{E}_n\left[\widetilde{w}_i\widetilde{w}_{1,i}'\right]h + \frac{1}{2}h'\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \widetilde{\theta})\widetilde{w}_{1,i}\widetilde{w}_i'\right]\frac{d\widehat{\lambda}(\widetilde{\theta})}{d\theta'}h$$
$$- \frac{1}{2}\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right] + \frac{1}{2}h'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h$$
$$+ \frac{1}{2}h'\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \widetilde{\theta})\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \widetilde{\theta})\widetilde{w}_i\widetilde{w}_{1,i}'\right]h, \quad \text{(F.10)}$$

其中，$\widehat{\lambda}_l(\widetilde{\theta})$（分别为 $\widehat{g}_l(\widetilde{\theta})$）表示向量 $\widehat{\lambda}(\widetilde{\theta})$（分别为 $\widehat{g}(\widetilde{\theta})$）的第 l 个元素；根据引理 F.1，它是 C^2 by。由于 h 有界，因此序列 $\theta \to \theta_\circ$ 在 $h \in K_n$ 上随 $n \to \infty$ 一致收敛。为处理项 (F.10)，我们使用以下通过类似前述方法的论证得到的极限。(1) 根据 Newey 和 McFadden（1994，引理 2.4）、连续映射定理和控制收敛定理，在假设 7 (a) 下，$\widehat{g}(\widetilde{\theta}) \xrightarrow{p} 0$在 $h \in K_n$ 上一致收敛。(2) 根据引理 F.14，$h'\mathbf{E}_n[\tau_i(\widehat{\lambda}, \widetilde{\theta})\widetilde{w}_{1,\,i}\widetilde{w}_i] \xrightarrow{p} h'\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i]$在 $h \in K_n$ 上一致收敛。(3) 根据引理 F.15，$\mathbf{E}_n[\tau_i(\widehat{\lambda}, \widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2] \xrightarrow{p} 0$在 $h \in K_n$ 上一致收敛。(4) 根据引理 F.16，$\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,\,i}'\right]h \xrightarrow{p} 0$ 在 $h \in K_n$ 上一致收敛。(5) 结合 (2)、(4) 和引理 F.6，并以 $\theta_\circ$ 替换 θ_*，可得

$h' \frac{d\lambda(\theta)'}{d\theta} \xrightarrow{p} h'\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\Omega_*^\diamond(\theta_\circ)^{-1}$ 在 $h \in K_n$ 上一致收敛。最后，根据引理 F.3 和 F.2，$\widehat{\lambda}(\widetilde{\theta}) \xrightarrow{p} 0$，且 $\theta \to \theta_\circ$ 在 $h \in K$ 上一致收敛。将这些极限代入 (F.10)，可得 (F.10) 等于：$-\frac{1}{2}h'\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i'](\mathbf{E}[\varepsilon_i^2\widetilde{w}_{1,i}\widetilde{w}_i])^{-1}\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i'] + o_p(1) =: -\frac{1}{2}h'V_{\theta_\circ}^{-1}h + o_p(1)$ 在 $h \in K_n$ 上一致收敛，且在假设 4 *(c)* 下，$V_{\theta_\circ}^{-1}$ 定义良好，其中 $\lambda = \lambda_*(\theta_*) = 0$，并满足假设 5。

□

定理 F.3（扩展模型中的随机 LAN。） 假设假设 *2–6* 成立，并且对每个 $\psi \in \Psi$，令 $h := \sqrt{n}(\psi - \psi_\circ)$。那么，对每个以零为中心、半径为 $M_n \to \infty$ 的闭球 $K_n \subset \mathbb{R}^{p+d_x}$，只要 $M_n = o(\sqrt{n})$，

$$\sup_{h \in K_n} \left| \log \frac{\widehat{q}(w_{1:n}|\psi_\circ + h/\sqrt{n})}{\widehat{q}(w_{1:n}|\psi_\circ)} - h'V_{\psi_\circ}^{-1}\Delta_{n,\psi_\circ} + \frac{1}{2}h'V_{\psi_\circ}^{-1}h \right| \xrightarrow{p} 0 \qquad \text{as } n \to \infty \tag{F.11}$$

其中，$\psi_\circ$ 是 ψ 的真值，$V_{\psi_\circ}^{-1} := V_{\psi_\circ}^{-1} := \mathbf{E}\left[\frac{dg_e(w_i\,\psi_\circ)'}{d\psi}\right]\Omega_{\psi_\circ}^{-1}\mathbf{E}\left[\frac{dg_e(w_i\,\psi_\circ)}{d\psi'}\right]$ 假定为非奇异，且 $V_{\psi_\circ}^{-1}\Delta_{n,\psi_\circ} := \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{E}\left[\frac{dg_e(w_i\,\psi_\circ)'}{d\psi}\right]\Omega_{\psi_\circ}^{-1}(\varepsilon_i\widetilde{w}_i - \widetilde{v}_\circ)$ 依概率有界。

证明。 首先，紧集 $K_n := \{h \in \mathbb{R}^p; \|h\| \le M_n\}$ 满足相应的 $\theta := \theta_\circ + h/\sqrt{n}$ 属于 $B_{\circ,n}$ 因此，根据假设 3，存在一个 $N \ge 1$，使得对于每个$n > N$，对数 ETEL 函数 $\ell_{n,\theta}(w_{1:n})$ 在 $B_{\circ,n}$ 上有定义。

证明过程与定理 F.2 的证明相同，只需将 $\lambda_*(\theta_\circ) = 0$替换为 $\lambda_*(\psi_\circ) = 0$，将 $\lambda(\theta)$替换为 $\widehat{\lambda}(\psi)$，将 $\mathrm{E}[g_i(\theta_\circ)] = 0$ 替换为 $\mathrm{E}[g_i(\theta_\circ)] = v_{0\circ}$ 令 $g_{e,i}(\psi) := g_e(w_i, \psi)$, $\tau_{e,i}^\diamond(\lambda, \psi) := e^{\lambda(\psi)'g_{e,i}(\psi)}$, $\check{\Omega}_\psi^\diamond(\lambda, \psi) := \mathrm{E}_n[\tau_{e,i}^\diamond(\lambda, \psi)g_{e,i}(\psi)g_{e,i}(\psi)'] = \Omega^\diamond(\lambda, \theta) - \widetilde{v}\widetilde{v}'$。此外，注意到

$$h'\frac{d\widehat{\lambda}(\psi)'}{d\psi}\check{\Omega}_\psi^\diamond(\widehat{\lambda}, \psi) = h_\theta'\mathbf{E}_n[\tau_{e,i}^\diamond(\widehat{\lambda}, \psi)\widetilde{w}_{1,i}\widetilde{w}_i']I_d + h_\theta'\mathbf{E}_n[\tau_{e,i}^\diamond(\widehat{\lambda}, \psi)\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\theta)\varepsilon_i(\theta)\widetilde{w}_i'] + (h_v'\mathbf{E}_n[\tau_{e,i}^\diamond(\widehat{\lambda}, \psi)], 0')'$$

其中 $h := (h_\theta', h_v')' \in K_n := K_{n,\theta} \times K_{n,v}$ 且 0 具有匹配维度。以下分解成立：

$$h'\frac{d\widehat{\lambda}(\psi_\circ)'}{d\psi}\check{\Omega}(\widehat{\lambda},\psi_\circ)-h'_\theta\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}'_i]=(h'_v\mathbf{E}_n[\tau^\diamond_{e,i}(\widehat{\lambda},\psi)],0')'$$
$$+h'_\theta\mathbf{E}_n\left[\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}'_i\right]-h'_\theta\mathbf{E}[\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)\widetilde{w}_{1i}\widetilde{w}'_i]$$
$$+h'_\theta\mathbf{E}[(\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)-1)\widetilde{w}_{1i}\widetilde{w}'_i]+h'_\theta\mathbf{E}_n\left[\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}'_i\widehat{\lambda}(\psi_\circ)\varepsilon_i(\theta)\widetilde{w}'_i\right]. \quad \text{(F.12)}$$

根据引理 F.3，对于任意 $\eta>0$，存在有限的 $\delta>0$ 和有限的 $N(\delta,\ \eta)>0$，使得对于每个 $n>N(\delta,\ \eta)$，$\|\lambda(\psi_\circ)\|_2\leq C/\sqrt{n}$ 以大于 η 的概率成立。因此，

$$\sup_{h_\theta\in K_{n,\theta}}\left\|h'_\theta\mathbf{E}_n\left[\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}_i\right]-h'_\theta\mathbf{E}[\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}_i]\right\|$$
$$\leq\sup_{h_\theta\in K_{n,\theta}}\sup_{\lambda\in B_{1/\sqrt{n}}(0)}\left\|h'_\theta\mathbf{E}_n\left[e^{\lambda' g_{e,i}(\psi_\circ)}\widetilde{w}_{1,i}\widetilde{w}_i\right]-h'_\theta\mathbf{E}[e^{\lambda' g_{e,i}(\psi_\circ)}\widetilde{w}_{1i}\widetilde{w}_i]\right\|$$

根据紧性和 Newey 和 McFadden（1994，引理 2.4），在假设 6 (c) 下（其中 $(j,\ell,\ell')=(1,1,1)$），该项收敛于零。接下来，我们控制 (F.12) 中的项 $h'_\theta\mathbf{E}[(\tau^\diamond_{e,i}(\lambda,\psi_\circ)-1)\widetilde{w}_{1i}\widetilde{w}_i]$。根据连续映射定理和引理 F.3：$\|e^{\lambda(\psi_\circ)'g_{e,i}(\psi_\circ)}h'\widetilde{w}_{1,i}\widetilde{w}_i-h'_\theta\widetilde{w}_{1,i}\widetilde{w}_i\|\xrightarrow{p}0$ 对每个 $i=1,\ldots,n$，关于 $h\in K_{n,\theta}$ 一致成立。由此以及控制收敛定理（在假设 6 (c) 下，当 $(j,\ell,\ell')=(1,1,1)$ 时，该定理成立）可知，$\mathbf{E}[(\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)-1)h'\widetilde{w}_{1i}\widetilde{w}'_i]\xrightarrow{p}0$ 关于 $h\in K_{n,\theta}$ 一致成立。利用上界，可以类似地处理 (F.12) 右边的最后一项：

$$\sup_{h_\theta\in K_{n,\theta}}\left\|h'_\theta\mathbf{E}_n\left[\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}'_i\widehat{\lambda}(\psi_\circ)\varepsilon_i(\theta_\circ)\widetilde{w}_i\right]\right\|\leq\sup_{h_\theta\in K_{n,\theta}}\sup_{\lambda\in B_{1/\sqrt{n}}(0)}\left\|h'_\theta\mathbf{E}_n\left[\tau^\diamond_{e,i}(\lambda,\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}'_i\lambda(\psi_\circ)\varepsilon_i(\theta_\circ)\widetilde{w}_i\right]\right.$$
$$\left.-h'_\theta\mathbf{E}\left[\tau^\diamond_{e,i}(\lambda,\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}'_i\lambda(\psi_\circ)\varepsilon_i(\theta_\circ)\widetilde{w}_i\right]\right\|$$
$$+\sup_{h_\theta\in K_{n,\theta}}\sup_{\lambda\in B_{1/\sqrt{n}}(0)}\left\|h'_\theta\mathbf{E}\left[\tau^\diamond_{e,i}(\widehat{\lambda},\psi_\circ)\widetilde{w}_{1,i}\widetilde{w}'_i\widehat{\lambda}(\psi_\circ)\varepsilon_i(\theta_\circ)\widetilde{w}_i\right]\right\|$$

（其中我们利用了 $\lambda_*(\psi_\circ)=0$ 这一事实）。根据 Newey 和 McFadden（1994，引理 2.4）、引理 F.3、连续映射定理和支配收敛定理，在假设 6 (f) 下，前一个表达式右边的两项依概率收敛于零。综合所有这些要素，并利用引理 F.6 控制表达式 $\frac{d\lambda}{d\psi}(\psi_\circ)'$ 中的 $\check{\Omega}^\diamond_\psi(\widehat{\lambda},\ \psi_\circ)$，我们得到 $h'\frac{d\lambda(\psi_\circ)}{d\psi}'\sqrt{n}\widehat{g}_e(\psi_\circ)$ 等于

$(h'_\theta \mathbf{E}[\widetilde{w}_{1i}\widetilde{w}'_i] + (h'_v,\ 0'))\Omega^{-1}_{\psi_\circ}\mathbb{G}_n[g_e(w_i,\ \psi_\circ)] + o_p(1)$，其中 $o_p(1)$ 项关于 $h \in K_n$ 是一致的。

对于 $\ell_{n,\psi_*+h/\sqrt{n}}(w_{1:n}) - \ell_{n,\psi_*}(w_{1:n})$ 的中值定理展开中的其他一阶项，我们采用与定理 F.2 的证明类似的方法，并使用与上文类似的论证，可知这些项关于 $h \in K_n$ 依概率一致收敛于零。因此，$\ell_{n,\psi_*+h/\sqrt{n}}(w_{1:n}) - \ell_{n,\psi_*}(w_{1:n})$ 的中值定理展开中的一阶项等于：$(h'_\theta \mathbf{E}[\widetilde{w}_{1i}\widetilde{w}'_i] + (h'_v, 0'))\Omega^{-1}_{\psi_\circ}\mathbb{G}_n[g_e(w_i, \psi_\circ)] + o_p(1)$，其中 $o_p(1)$ 关于 $h \in K_n$ 一致。最后，由于根据假设 4 (a) 的林德伯格-列维中心极限定理，$\mathbb{G}_n[g_e(w_i, \psi_\circ)] \xrightarrow{d}$ 服从 $N(0, \Omega_{\psi_\circ})$，我们得出 $\ell_{n,\psi_*+h/\sqrt{n}}(w_{1:n}) - \ell_{n,\psi_*}(w_{1:n})$ 的中值定理展开中的一阶项依概率有界。

现在，考虑（F.5）中的二阶项（针对扩展模型进行调整），其中 $\widetilde{\psi} := \psi_\circ + \tau h/\sqrt{n}$，且对于某个 $\tau \in [0, 1]$ 和 $h \in K_n$：

$$\begin{aligned}
&\frac{1}{2}h' \sum_{l=1}^{d} \frac{d^2\widehat{\lambda}_l(\widetilde{\psi})}{d\psi d\psi'}\widehat{g}_{e,l}(\widetilde{\psi})h - h'\frac{d\widehat{\lambda}(\widetilde{\psi})'}{d\psi}\mathbf{E}_n\left[\frac{dg_e(\widetilde{\psi})}{d\psi'}\right]h + \frac{1}{2}h'\mathbf{E}_n\left[\tau_{e,i}(\widehat{\lambda},\widetilde{\psi})\frac{dg_e(\widetilde{\psi})'}{d\psi}\right]\frac{d\widehat{\lambda}(\widetilde{\psi})}{d\psi'}h \\
&- \frac{1}{2}\mathbf{E}_n\left[\tau_{e,i}(\widehat{\lambda},\widetilde{\psi})\left(h'\frac{dg_e(\widetilde{\psi})'}{d\psi}\widehat{\lambda}(\widetilde{\psi})\right)^2\right] + \frac{1}{2}h'\frac{d\widehat{\lambda}(\widetilde{\psi})'}{d\psi}\mathbf{E}_n\left[\tau_{e,i}(\widehat{\lambda},\widetilde{\psi})g_{e,i}(\widetilde{\psi})\widehat{\lambda}(\widetilde{\psi})'\frac{dg_e(\widetilde{\psi})}{d\psi'}\right]h \\
&+ \frac{1}{2}h'\mathbf{E}_n\left[\tau_{e,i}(\widehat{\lambda},\widetilde{\psi})\frac{dg_e(\widetilde{\psi})'}{d\psi}\right]\widehat{\lambda}(\widetilde{\psi})\widehat{\lambda}(\widetilde{\psi})'\mathbf{E}_n\left[\tau_{e,i}(\widehat{\lambda},\widetilde{\psi})\frac{dg_e(\widetilde{\psi})}{d\psi'}\right]h, \quad \text{(F.13)}
\end{aligned}$$

其中，$\widehat{\lambda}_l(\widetilde{\psi})$（分别为 $\widehat{g}_{e,\,l}(\psi)$）表示向量 $\widehat{\lambda}(\widetilde{\psi})$（分别为 $\widehat{g}_e(\psi)$）的第 l 个元素，根据引理 F.1，该元素为 $\mathcal{C}^2$。由于 h 有界，因此当 $n \to \infty$ 时，序列 $\widetilde{\psi} \to \psi_\circ$ 关于 $h \in K_n$ 一致收敛。为处理项 (F.13)，我们使用以下极限，这些极限可通过与前述类似的论证得到。(1) 根据假设 7 (a)、$B_{1/\sqrt{n}}(\psi_\circ)$ 的紧致性、Newey 和 McFadden（1994，引理 2.4）、连续映射定理和控制收敛定理，$\widehat{g}_e(\widetilde{\psi}) \xrightarrow{p} 0$ 关于 $h \in K_n$ 一致收敛。(2) 根据引理 F.14，$h'\mathbf{E}_n[\tau_{e,i}(\widehat{\lambda}, \widetilde{\psi})\,\widetilde{w}_{1i}\widetilde{w}_i] \xrightarrow{p} h'\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i]$ 关于 $h \in K_n$ 一致收敛。(3) 根据引理 F.15，

$\mathbf{E}_n[\tau_{e,i}(\widehat{\lambda},\ \widetilde{\psi})\ \left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\ \widetilde{\psi})\right)^2] \xrightarrow{p} 0$ 关于 $h \in K_n$ 一致收敛。(4) 根据引理 F.16,

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\psi})g_{e,i}(\widetilde{\psi})\widehat{\lambda}(\widetilde{\psi})'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h \xrightarrow{p} 0$$

在 $h \in K_n$ 上一致。(5) 结合 (2)、(4) 以及将 θ_* 替换为 $\psi_\circ$ 的引理 F.6，我们有$\widehat{\ }\ \widetilde{\ }$，即 $h'^{d\lambda}(\ \xrightarrow[d\psi]{\psi\ '\ p}\ )\ (h_\theta'\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i'] + (h_v', 0'))\Omega_{\psi_\circ}^{-1}$在 $h \in K_n$ 上一致。最后，由引理 F.3、F.2 和 $\psi \to \psi_\circ$，$\widehat{\lambda}(\widetilde{\psi}) \xrightarrow{p} 0$在 $h \in K_n$ 上一致$\widetilde{\ }$。将这些极限代入 (F.13)，可得 (F.13) 等于：$-\frac{1}{2}h'\mathbf{E}\left[{}^{dg_e}(\ \underset{d\psi}{\widetilde{\psi}\ '}\ )\right]\Omega_{\psi_\circ}^{-1}\mathbf{E}_n[\ {}^{dg_e}(\widetilde{\psi})_{\ d\psi'}\ ]h + o_p(1) =: -\frac{1}{2}h'V_{\psi_\circ}^{-1}h + o_p(1)$ 在 $h \in K_n$ 上一致，并且在取 $\lambda = \bar{\lambda}_*(\psi_\circ) = 0$ 的假设 4 (c) 以及假设 5 下，$V_{\psi_\circ}^{-1}$是良定义的。

□

F.3 后验一致性

定理 F.4（后验一致性——基准模型）假设假设 *1–8* 成立。设关于 θ 的先验是连续概率测度，其相对于勒贝格测度存在密度，且在 θ_* 的某个邻域上为正。那么，对于任意 $M_n \to \infty$,

$$\pi^n\left(\theta \in H_n;\ \sqrt{n}\|\theta - \theta_*\| > M_n \middle|\ w_{1:n}\right) \xrightarrow{p} 0\ ,\ \textit{as}\ n \to \infty\ . \tag{F.14}$$

证明。定义事件 $A_{n,\ 1} := \left\{\sup_{\theta \in H_n \cap \Theta_n^c} \frac{1}{n}\sum_{i=1}^n (\ell_{n,\ \theta}(w_i) - \ell_{n,\ \theta_*}(w_i))\ \leq -CM_n^2/n\right\}$ 以及

$$A_{n,2} := \left\{\int_{\Theta \cap H_n} \frac{\widehat{p}(w_{1:n}|\theta)}{\widehat{p}(w_{1:n}|\theta_*)}\pi(\theta)d(\theta) \geq e^{-CM_n^2/2}\right\}.$$

根据假设 8，$P(A_{n,\ 1}^c) \to 0$；根据 Chib 等人（2018，引理 E.3），$P(A_{n,\ 2}^c) \to 0$。因此，根据全期望定律

$$\mathbf{E}\left[\pi^n\ (H_n \cap \Theta_n^c |\ w_{1:n})\ \right] \leq$$

$$\mathbf{E}\left[\pi\left(\theta \in H_n; \sqrt{n}\|\theta - \theta_*\| > M_n \middle|\ w_{1:n}\right) \middle|\ A_{n,1} \cap A_{n,2}\right]\ P(A_{n,1} \cap A_{n,2}) + o(1)$$

$$= \mathbf{E}\left[\frac{\int_{H_n\cap\Theta_n^c} e^{\sum_{i=1}^n(\ell_{n,\theta}(w_i)-\ell_{n,\theta_*}(w_i))}\pi(\theta)d\theta}{\int_{\Theta\cap H_n} e^{\sum_{i=1}^n(\ell_{n,\theta}(w_i)-\ell_{n,\theta_*}(w_i))}\pi(\theta)d\theta}\middle| A_{n,1}\cap A_{n,2}\right] P(A_{n,1}\cap A_{n,2}) + o(1)$$

$$\leq e^{-CM_n^2}\pi(\Theta_n^c)\mathbf{E}\left[\left(\int_{\Theta\cap H_n}\frac{\widehat{p}(w_{1:n}|\theta)}{\widehat{p}(w_{1:n}|\theta_*)}\pi(\theta)d\theta\right)^{-1}\middle| A_{n,1}\cap A_{n,2}\right] + o(1)$$

$$\leq e^{-CM_n^2}e^{CM_n^2/2}\pi(\Theta_n^c) + o(1) = o(1), \quad \text{(F.15)}$$

其中，趋于零是由 $M_n \to \infty$ 得出的。这证明了该定理的结论。

□

定理 F.5（后验一致性——扩展模型）假设假设 *1–6* 和假设 *9* 成立。设 ψ 上的先验是一个连续概率测度，它相对于勒贝格测度具有密度，且在 $\psi_\circ$ 的某个邻域上为正。那么，对任意 $M_n \to \infty$，

$$\pi^n\left(\psi \in H_n \times \mathcal{V}; \sqrt{n}\|\psi - \psi_\circ\| > M_n \middle| w_{1:n}\right) \xrightarrow{p} 0 \text{ , } as\ n \to \infty \text{ .} \quad \text{(F.16)}$$

证明。证明过程与定理 F.4 的证明相同，因此省略。

□

F.4 Bern基准模型和扩展模型的斯坦-冯·米塞斯定理 s

定理 F.6（基准模型中的伯恩斯坦-冯·米塞斯定理）假设定理 *F.1* 和 *F.4* 的条件成立。那么， $h := \sqrt{n}(\theta - \theta_*)$ 的后验分布序列在全变差意义下收敛到一个正态分布，即

$$\left\|\pi_h^n(\sqrt{n}(\theta-\theta_*)|w_{1:n}) - \mathcal{N}_{\Delta_{n,\theta_*},V_{\theta_*}}\right\|_{TV} \xrightarrow{p} 0, \quad \text{(F.17)}$$

其中 $\Delta_{n,\ \theta_*} := \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{\theta_*}\dot{\ell}_{n,\ \theta_*}(w_i) + o_p(1) = O_p(1)$，且 V_{θ_*} 是一个正定矩阵，等于以下矩阵的逆：

$$V_{\theta_*}^{-1} = \mathbf{E}^{Q^*(\theta_*)}\left[\tilde{w}_{1,i}\tilde{w}_i'(I + \lambda_*(\theta_*)\varepsilon_i\tilde{w}_i')\right]\left(\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_i^2\tilde{w}_i\tilde{w}_i'\right]\right)^{-1}$$
$$\times\left(2\mathbf{E}[\tilde{w}_i\tilde{w}_{1,i}'] - \mathbf{E}^{Q^*(\theta_*)}\left[\tilde{w}_{1,i}\tilde{w}_i'(I + \lambda_*(\theta_*)\varepsilon_i\tilde{w}_i')\right]\right)$$
$$-\sum_{j=1}^{d_x}\ddot{\lambda}_{*,j}(\theta_*)\mathbf{E}[\varepsilon_i x_{i,j}] + \mathbb{V}ar_{Q^*(\theta_*)}[\tilde{w}_{1,i}\tilde{w}_i'\lambda_*(\theta_*)],$$

其中 $\mathbb{V}ar_{Q^*(\theta_*)}$ 表示相对于分布 $Q^*(\theta_*)$ 所取的方差。

证明。 本定理的证明沿用 Chib 等人（2018，定理 2.2）的证明。它依赖于两个中间结果：定理 F.4 的后验一致性结果，以及下文定理 F.1 所建立的随机局部渐近正态展开（F.4）。

□

推论 F.1（外生性下基准模型中的伯恩斯坦-冯·米塞斯） 令 $\theta_\circ$ 表示 θ 的真值，并假设 $\mathbf{E}[\varepsilon_i(\theta_\circ)x_i] = 0$。设假设 *2–6* 成立，其中 θ_* 替换为 $\theta_\circ$，$\lambda_*(\theta_\circ)$替换为零，且假设 *4*（*c*）中的矩阵替换为矩阵 $\mathbf{E}[\varepsilon_i(\theta_\circ)^2\widetilde{w}_i\widetilde{w}_i']$。设假设 *8* 成立，且 θ 上的先验是一个连续概率测度，该测度关于勒贝格测度具有密度，并在$\theta_\circ$的某个邻域上为正。则 $h := \sqrt{n}(\theta - \theta_*)$ 的后验分布序列依全变差收敛于正态分布，即，

$$\left\|\pi_h^n(\sqrt{n}(\theta - \theta_\circ)|w_{1:n}) - \mathcal{N}_{\Delta_{n,\theta_\circ},V_{\theta_\circ}}\right\|_{TV} \xrightarrow{p} 0, \tag{F.18}$$

其中 $\Delta_{n,\,\theta_\circ} := \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{\theta_\circ}\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\left(\mathbf{E}[\varepsilon_i^2\widetilde{w}_i\widetilde{w}_i]\right)^{-1}\varepsilon_i\widetilde{w}_i$，且 $V_{\theta_\circ}$ 是 $V_{\theta_\circ}^{-1} := \mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\left(\mathbf{E}[\varepsilon_i^2\widetilde{w}_i\widetilde{w}_i']\right)^{-1}\mathbf{E}\left[\widetilde{w}_i\widetilde{w}_{1,\,i}'\right]$的逆。

证明。 该推论的证明过程与定理 F.6 的证明相同。唯一的区别在于：在后验一致性结果中，我们必须将伪真值 θ_* 替换为真值 $\theta_\circ$；此外，对于设定正确的情形，我们在上面的定理 F.2 中给出了随机 LAN 展开。

□

定理 F.7（扩展模型中的伯恩斯坦-冯·米塞斯） 令 $\psi_\circ := (\theta'_\circ \; v'_\circ)'$ 表示 ψ 的真值，并令假设 *2–6* 成立，其中将 θ_* 替换为 $\psi_\circ$，将 $\lambda_*(\psi_\circ)$ 替换为零，并将假设 *4 (c)* 中的矩阵替换为矩阵 $\mathbf{E}[\varepsilon_i(\psi_\circ)^2 \widetilde{w}_i \widetilde{w}_i']$。假设 *9* 成立。令 ψ 上的先验为一个相对于勒贝格测度具有密度、且在 $\psi_\circ$的某个邻域上为正的连续概率测度。则后验分布序列在全变差意义下收敛于正态分布，即

$$\left\| \pi^n(\sqrt{n}(\psi - \psi_\circ)|w_{1:n}) - \mathcal{N}_{\Delta_{n,\psi_\circ}, V_{\psi_\circ}} \right\|_{TV} \xrightarrow{p} 0, \tag{F.19}$$

其中 $\Delta_{n,\ \psi_\circ} = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_{\psi_\circ}$**期望**$\left[\frac{dg_e(w_i\ \psi_\circ)'}{d\psi} \right], \Omega_{\psi_\circ}^{-1}(\varepsilon_i \widetilde{w}_i - \widetilde{v}_\circ)$，其中 $\widetilde{v}_\circ := (v'_\circ, 0', 0')'$ 且 $v_\circ :=$**期望**$[\varepsilon_i x_i]$，并且 $V_{\psi_\circ}$ 是.....的逆

$$V_{\psi_\circ}^{-1} = \mathbf{E}\left[\frac{dg_e(w_i, \psi_\circ)'}{d\psi} \right] \Omega_{\psi_\circ}^{-1} \mathbf{E}\left[\frac{dg_e(w_i, \psi_\circ)}{d\psi'} \right]$$

其中 $\Omega_{\psi_\circ} :=$**期望**$[g_e(w_i, \psi_\circ) g_e(w_i, \psi_\circ)'] =$**期望**$[\varepsilon_i^2 \widetilde{w}_i \widetilde{w}_i'] - \widetilde{v}_\circ \widetilde{v}_\circ'$ 且 **期望**$\left[\frac{dg_e(w_i\ \psi_\circ)'}{d\psi} \right] = -\begin{pmatrix} , \text{期望} \\ [\widetilde{w}_{1i} x_i'] \quad [\widetilde{w}_{1i} z_i'] \end{pmatrix}$**期望**。

证明。 该定理的证明与推论 F.1 的证明过程相同，只需将 $\theta_\circ$ 替换为 $\psi_\circ$，并使用上述定理 F.3 中的随机 LAN 展开。

□

F.5 基础模型的技术结果

本节简化记号，记 $g_i(\theta) := g_b(w_i,\ \theta)$。F.1 的结果在渐近意义下成立，即对于 $n \geq N$，其中 N 是一个足够大的正整数。

引理 F.1 假设假设 *1*、*3* 和 *4 (c)* 成立。设 λ_* 是方程 $F_n(\lambda, \theta_*) = 0$ 在 λ 中的一个解，其中 $F_n(\lambda, \theta) := -\sum_{i=1}^n \frac{e^{\lambda' \widetilde{w}_i \varepsilon_i(\theta)}}{\sum_{j=1}^n e^{\lambda' \widetilde{w}_j \varepsilon_j(\theta)}} \varepsilon_i(\theta) \widetilde{w}_i$。那么，存在一个邻域

$U \subset \mathbb{R}^d$，其中心为 $\widehat{\lambda}_*$，以及一个 $N \geq 1$，使得 $\forall n \geq N$ 存在唯一的连续映射 $\widehat{\lambda}$，将 $B_{*,n}$ 映入 U，并满足：

(i) $\widehat{\lambda}(\theta_*) = \lambda_*$；

(ii) $F_n(\widehat{\lambda}(\theta), \theta) = 0$ 对每个 $\theta \in B_{*,n}$；

(iii) $\widehat{\lambda} \in C^2$ on $B_{*,n}$ 和 $\forall \theta \in B_{*,n}$，

$$\frac{\partial \widehat{\lambda}(\theta)}{\partial \theta'} = \check{\Omega}^{\diamond}(\widehat{\lambda}, \theta)^{-1} \mathbf{E}_n \left[\tau_i^{\diamond}(\widehat{\lambda}, \theta) w_{1,i} \widetilde{w}_i' (I + \widehat{\lambda}(\theta) g_i(\theta)') \right]. \tag{F.20}$$

证明。 我们打算应用隐函数定理（例如 Dieudonné（1969，定理 10.2.3）），为此只需检查其条件是否成立。首先，函数 $(\lambda, \theta) \mapsto F(\lambda, \theta)$ 在 $\mathbb{R}^{d+p}$ 上至少二阶连续可微。其次，在假设 1 和 3 下，存在一个 λ，使得对每个 $\theta \in B_{*,n}$，$F(\lambda, \theta) = 0$ 以趋近于 1 的概率成立。此外，在先前的 (θ, λ) 处计算的一阶偏导数 $\partial F(\lambda, \theta)/\partial \lambda$，以趋近于 1 的概率为：

$$\frac{\partial F(\widehat{\lambda}, \theta)}{\partial \widehat{\lambda}} = -\sum_{i=1}^{n} \frac{e^{\widehat{\lambda}' \widetilde{w}_i \varepsilon_i(\theta)}}{\sum_{j=1}^{n} [e^{\widehat{\lambda}' \widetilde{w}_j \varepsilon_j(\theta)}]} \varepsilon_i(\theta)^2 \widetilde{w}_i \widetilde{w}_i'. \tag{F.21}$$

这里我们用到了这一事实：$P(F(\widehat{\lambda}, \theta) = 0) \to 1$。根据假设 4 *(c)*，矩阵 $\frac{\partial F(\lambda, \theta)}{\partial \widehat{\lambda}}$ 不为零，因此可逆。

然后，根据隐函数定理，存在 $\lambda(\theta_*)$的一个开邻域$\mathcal{U}$ 以及一个$N \geq 1$，使得 $\forall n \geq N$：

(i) 对每个 $\theta \in B_{*,n}$ 都有唯一的 $\lambda \in \mathcal{U}$ 与之对应，使得 $F(\lambda, \theta) = 0$；(ii) 该 λ 可写为 $\lambda = \lambda(\theta)$，并且满足 $F(\lambda(\theta), \theta) = 0$，其中 $\theta \in B_{*,n}$；(iii) $\widehat{\lambda} \in \mathcal{C}^2$ 上 $B_{*,n}$ 且有：

$$\frac{\partial \widehat{\lambda}(\theta)}{\partial \theta'} = \left(\frac{\partial F(\widehat{\lambda}, \theta)}{\partial \lambda'} \right)^{-1} \frac{\partial F(\widehat{\lambda}, \theta)}{\partial \theta'} \equiv \check{\Omega}^{\diamond}(\widehat{\lambda}, \theta)^{-1} \mathbf{E}_n \left[\tau_i^{\diamond}(\widehat{\lambda}, \theta) w_{1,i} \widetilde{w}_i' (I + \widehat{\lambda}(\theta) g_i(\theta)') \right].$$

□

引理 F.2（连续性）令 $\lambda_* := \operatorname{argmin}\mathbf{E}\left[e^{\lambda' g_{i(\theta_*)}}\right]$，并假设假设 *3* 成立。则存在 λ_* 的一个邻域 $U \subset \mathbb{R}^d$ 以及一个 $N \geq 1$，使得 $\forall n \geq N$ 存在唯一的连续映射 λ_*，将 $B_{*,\,n}$ 映入 U，并满足：

(i) $\lambda_*(\theta_*) = \lambda_*$;

(ii) $\mathbf{E}[e^{\lambda_*(\theta)'\widetilde{w}_i \varepsilon_i(\theta)} \varepsilon_i(\theta)\, \widetilde{w}_i] = 0$ 对每个 $\theta \in B_{*,n}$;

(iii) $\lambda_* \in C^2$ *on* $B_{*,n}$ 和 $\forall \theta \in B_{*,n}$,

$$\frac{\partial \lambda_*(\theta)}{\partial \theta'} = \Omega_*^{\diamond}(\theta_*)^{-1}\mathbf{E}_n\left[\tau_i^{\diamond}(\lambda_*, \theta) w_{1,i} \widetilde{w}_i'(I + \lambda_*(\theta) g_i(\theta)')\right]. \tag{F.22}$$

(iv) $\theta \mapsto \mathbf{E}\left[e^{\lambda_*(\theta)' g(w_i, \theta)}\right]$ 在 $B_{*,n}$ 上连续。

证明。 (i)-(iii) 的证明与引理 F.1 的证明类似，通过应用隐函数定理即可，因此略去。结果 (iv) 可由 $\theta \mapsto \lambda_*(\theta)$ 在 $B_{*,n}$ 上的连续性，以及函数 $\theta \mapsto \mathbf{E}\left[e^{\lambda' g_{i(\theta)}} g_i(\theta)\right]$ 的连续性得出。

□

引理 F.3 假设　若假设 *1*、*3*、*4(a)* 和 *6(d)-(e)* 成立。那么，对于每个给定的 $\theta \in B_{*,\,n}$,

$$\sqrt{n}\|\widehat{\lambda}(\theta) - \lambda_*(\theta)\|_2 = \mathcal{O}_p(1)$$

或者，等价地，对于任意 $\eta > 0$，存在有限的 $\delta > 0$ 和有限的 $N(\delta、\ \eta) > 0$，使得：$\forall n > N(\delta、\ \eta)$,

$$P\left(\sqrt{n}\|\widehat{\lambda}(\theta) - \lambda_*(\theta)\|_2 < \delta\right) > \eta.$$

证明。 为了证明该引理，我们打算将 Van der Vaart and Wellner（1996，推论 3.2.6）应用于 $\mathbb{M}_n = -\mathbf{E}_n\left[e^{\lambda' g_{i(\theta)}}\right]$。此处，$\theta \in B_{*,n}$ 保持固定，我们将 $\mathbb{M}_n$ 视为数据 w_i 和 λ 的函数。对于 $\lambda_* = \arg\max_{\lambda \in \mathbb{R}^d} \mathbb{M}$，其中 $\mathbb{M} := -\mathbf{E}\left[e^{\lambda' g_{i(\theta)}}\right]$，定义函数集

$\mathcal{M}_\delta(\theta) := \left\{ w_i \mapsto e^{\lambda_*' g_i(\theta)} - e^{\lambda' g_i(\theta)}; \lambda \in B_\delta(\lambda_*) \right\}$ 由 $\lambda \in B_\delta(\lambda_*)$ 索引。由于 $\mathcal{M}_\delta(\theta)$ 中的函数关于索引参数 $\lambda \in B_\delta(\lambda_*)$ 满足利普希茨条件，因此根据中值定理和柯西-施瓦茨不等式，存在一个 $\tau \in (0,1)$ 使得 $\lambda_{12} := \lambda_1 + \tau(\lambda_2 - \lambda_1)$，其中 $\lambda_1, \lambda_2 \in B_\delta(\lambda_*)$ 满足

$$|e^{\lambda_2' g_i(\theta)} - e^{\lambda_1' g_i(\theta)}| \leq 2\delta \|g_i(\theta) e^{\widetilde{\lambda}_{12}' g_i(\theta)}\|_2$$

对于每个 w_i 以及每个固定的 $\theta \in B_{*,n}$。因此，$\mathcal{M}_\delta(\theta)$ 的一个包络函数，记为 $M_\delta(\theta)$，可表示为

$$M_\delta(\theta) = 2\delta \|g_i(\theta)\|_2 \sup_{\lambda_1, \lambda_2 \in B_\delta(\lambda_*)} e^{\widetilde{\lambda}_{12}' g_i(\theta)}$$

（参见 Van der Vaart and Wellner（1996，定理 2.7.11）之前关于包络函数的讨论）。此外，根据 Van der Vaart（1998，示例 19.7），存在一个常数 K，使得 $\mathrm{M}_\delta(\theta)$ 的括号覆盖数 $N_{[]}(\varepsilon \|M_\delta(\theta)\|_P, 2, \mathrm{M}_\delta(\theta)\ \|\cdot\|_P, 2)$ 被 $K(2\delta)^d/\varepsilon^d$ 上界所界定；对于每个 $0 < \varepsilon < 2\delta$，只要 $\|M_\delta(\theta)\|_{P,2} < C/\sqrt{n}$，则该结论成立，而在假设 6 (d) 下，对于 $\delta = 1/\sqrt{n}$，这一条件得到保证。因此，括号积分 $J_{[]}(1, \mathrm{M}_\delta(\theta)\ \|\cdot\|_P, 2) := \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon \|M_\delta(\theta)\|_P, 2, \mathrm{M}_\delta(\theta)\ \|\cdot\|_P, 2)} d\varepsilon$ 是有界的，并且函数类 $\mathrm{M}_\delta(\theta)$ 是 P-Glivenko-Cantelli 类（参见例如 Van der Vaart and Wellner（1996，定理 2.5.6）以及 Van der Vaart and Wellner（1996，第 243–244 页）中讨论的简化版本）。这确保了对于每个 $\theta \in B_{*,n}$，$|\mathbb{M}_n - \mathbb{M}| \to 0$ 在 P-概率意义下关于 $\mathrm{M}_\delta(\theta)$ 一致成立。

因此，根据 Van der Vaart and Wellner（1996，推论 3.2.3），对于每个固定的 $\theta \in B_{*,n}$，$\widehat{\lambda}(\theta) \xrightarrow{p} \lambda_*(\theta)$。最后，为了应用 Van der Vaart and Wellner（1996，推论 3.2.6），我们需要给 $\mathbf{E}(\sup_{f \in \mathcal{M}_{1/\sqrt{n}}} |\sqrt{n}\mathbf{E}_n[f - \mathbf{E}(f)]|)$ 建立上界。根据 Van der Vaart and Wellner（1996，推论 3.2.6）之后的讨论（并且由于前面的论证已表明 $J_{[]}(1, \mathcal{M}_\delta, \|\cdot\|_{P,2})$ 有界），后者的上界为 $\|M_{1/\sqrt{n}}\|_{P,2}$ 的常数倍。由于 $\|M_{1/\sqrt{n}}\|_{P,2} \leq C/\sqrt{n}$，因此 $r_n^4 \|M_{1/r_n}\|_{P,2}^2 \leq n$ 得出 $r_n \asymp \sqrt{n}$。所以，Van der Vaart and Wellner（1996，推论 3.2.6）的结果在 $r_n \leq \sqrt{n}$ 下成立。这就证明了引理的结论。

□

引理 F.4 设假设 *1*、*3*、*4(a)*、*6(e)*成立。则，

$$\sqrt{n}(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*)) = -\Omega_*^{\dagger}(\theta_*)^{-1}\mathbb{G}_n[\tau_i^{\dagger}(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i] + o_p(1).$$

证明。令 $\lambda_* \equiv \lambda(\theta_*)$ 和 $\lambda_* \equiv \lambda_*(\theta_*)$。对函数 $\lambda \mapsto \mathbb{E}_n[e^{\lambda'\widetilde{w}_i\varepsilon_i(\theta_*)}\varepsilon_i(\theta_*)\,\widetilde{w}_i]$ 在 λ_* 附近、于 $\widehat{\lambda}$ 处取值进行二阶中值定理展开，得到：$\widehat{\lambda}_* - \lambda_* = -\Omega_*^{\diamond}(\theta_*)^{-1}\mathbb{E}_n[\tau_i^{\diamond}(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\,\widetilde{w}_i] + O_p(1/n)$，其中 $\Omega_*^{\diamond}(\lambda_*, \theta_*)$ 依概率收敛于 $\Omega_*^{\diamond}(\theta_*)$，这是由大数定律得出的；$O_p(1/n)$ 项则由引理 F.3 和假设 6 (e)（其中 $(\ell, \ell', i) = (1, 3, 1)$）推出，它们使得我们能够控制中值定理展开中的二次项。此外，$\mathbb{E}[\tau_i^{\diamond}(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\,\widetilde{w}_i] = 0$。因此，

$$\begin{aligned}\sqrt{n}\left(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*)\right) &= -\Omega_*^{\diamond}(\theta_*)^{-1}\mathbb{G}_n[\tau_i^{\diamond}(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i] + o_p(1) \\ &= -\Omega_*^{\dagger}(\theta_*)^{-1}\mathbb{G}_n[\tau_i^{\dagger}(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i] + o_p(1).\end{aligned}$$

□

引理 F.5 设假设 *1–6* 成立。则，对每一个属于以零为中心、半径为 $M_n \to \infty$ 的闭球 $K_n \subset \mathbb{R}^p$ 的 h，

$$\begin{aligned}h'\sqrt{n}\left(\frac{d\widehat{\lambda}(\theta_*)'}{d\theta}\widehat{g}(\theta_*) - \frac{d\lambda_*(\theta_*)'}{d\theta}\mathbf{E}[g_i(\theta_*)]\right) - h'\sqrt{n}\left(\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']\widehat{\lambda}(\theta_*) - \mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\lambda_*(\theta_*)\right) \\ + \sqrt{n}h'\left(\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}_*(\theta) - \mathbf{E}[\tau_i^{\dagger}(\lambda_*, \theta_*)\widetilde{w}_{1i}\widetilde{w}_i']\lambda_*(\theta_*)\right)\end{aligned}$$

渐近正态，均值为零，方差 $h'H_*h$定义为

$$\begin{aligned}h'H_*h := h'A\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^{\dagger}(\lambda_*, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]A'h \\ + Var\Big(\tau_i^{\dagger}(\lambda_*, \theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I + \lambda_*(\theta_*)g_i(\theta_*)')B + h'a'(1 - \tau_i^{\dagger}(\lambda_*, \theta_*))\varepsilon_{i*}(\theta_*)\widetilde{w}_i \\ + h'a'\tau_i^{\dagger}(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i'\Omega_*^{\dagger}(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}^3\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right]B \\ - h'\tau_i^{\dagger}(\lambda_*, \theta_*)\varepsilon_{i*}\widetilde{w}_i'\Omega_*^{\dagger}(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[g_i(\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'(I + \lambda_*(\theta_*)g_i(\theta_*)')\right]B\end{aligned}$$

$$
\begin{aligned}
&- h'a'\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'B - \tau_i^\dagger(\lambda_*,\theta_*)\mathbf{E}^{Q_*(\theta_*)}\left[h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\\
&\qquad - \left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\Big)\\
&+ 2h'A\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')B\widetilde{w}_{1,i}'h\right]\\
&+ 2h'A\mathbf{E}^{Q^*(\theta_*)}\left[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]ah\\
&+ 2h'A\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}(\theta_*)^3\widetilde{w}_i\widetilde{w}_i'B\widetilde{w}_i'\right]ah\\
&- 2h'A\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[g_i(\theta_*)\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')B\widetilde{w}_{1,i}'\right]h\\
&- 2h'A\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^3\widetilde{w}_i\widetilde{w}_i'B\widetilde{w}_i'\right]ah\\
&- 2h'A\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i\right]\lambda_*(\theta_*)'\mathbf{E}^{Q^*(\theta_*)}[\widetilde{w}_i\widetilde{w}_{1,i}'h]\\
&\qquad - 2h'A\mathbf{E}^{Q^*(\theta_*)}\left[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\varepsilon_{i*}\widetilde{w}_i\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h,
\end{aligned}
$$

其中，H_* 是一个正定矩阵，$A := E[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}$ 是一个 $(p\times d)$ 矩阵，$B := \Omega_*^\dagger(\theta_*)^{-1}E[g_i(\theta_*)]$是一个 d 维向量，而 $a := \Omega_*^\dagger(\theta)^{-1}E^{Q^*(\theta)}\left[(I+g_i(\theta)\lambda_*(\theta)')\,\widetilde{w}_i\widetilde{w}_{1,i}'\right]$。

证明。 在本证明中，我们用 $\varepsilon_{i*} := \varepsilon_i(\theta_*)$ 来简化表达式。我们首先如下定义项$\mathcal{A}_1$,$\mathcal{A}_2$ 和 $\mathcal{A}_3$：

$$
\begin{aligned}
&h'\sqrt{n}\left(\frac{d\widehat{\lambda}(\theta_*)'}{d\theta}\widehat{g}(\theta_*) - \frac{d\lambda_*(\theta_*)'}{d\theta}\mathbf{E}[g_i(\theta_*)]\right) - h'\sqrt{n}\left(\mathbf{E}_n[\widetilde{w}_{1,i}\widetilde{w}_i']\widehat{\lambda}(\theta_*) - \mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\lambda_*(\theta_*)\right)\\
&\quad + \sqrt{n}h'\left(\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*) - \mathbf{E}[\tau_i^\dagger(\lambda_*,\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i']\lambda_*(\theta_*)\right) =: \mathcal{A}_1+\mathcal{A}_2+\mathcal{A}_3.
\end{aligned}
$$

首先，我们分析$\mathcal{A}_1$:

$$
\mathcal{A}_1 = h'\sqrt{n}\left(\frac{d\widehat{\lambda}(\theta_*)'}{d\theta} - \frac{d\lambda_*(\theta_*)'}{d\theta}\right)\widehat{g}(\theta_*) + h'\frac{d\lambda_*(\theta_*)'}{d\theta}\sqrt{n}\mathbf{E}_n\left[g_i(\theta_*) - \mathbf{E}[g_i(\theta_*)]\right]
$$

并由 u 利用引理 F.1 和 F.2 中给出的$\frac{d\lambda_{(\theta_*')}}{d\theta}$ 和 $\frac{d\lambda_{*(\theta_*')}}{d\theta}$ 的表达式，我们得到 ：

$$
\sqrt{n}h'\left(\frac{d\widehat{\lambda}(\theta_*)'}{d\theta} - \frac{d\lambda_*(\theta_*)'}{d\theta}\right) = \sqrt{n}\Big(\mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda},\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\theta_*)g_i(\theta_*)')\right]\breve{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}
$$

$$- \mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\left(\mathbf{E}[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i']\right)^{-1}\Big).$$

对 $\lambda \mapsto \tau_i^\diamond(\lambda,\theta_*)$ 在 $\lambda_*(\theta_*)$ 处作一阶中值定理展开，并在 $\lambda(\theta_*)$ 处取值，存在一个 $\tau \in [0,1]$ 使得：$\tau_i^\diamond(\lambda,\theta_*) = \tau_i^\diamond(\lambda_*,\theta_*) + \tau_i^\diamond(\lambda_*,\theta_*)g_i(\theta_*)'(\lambda(\theta_*) - \lambda_*(\theta_*))$，其中 $\lambda_* := \tau\lambda(\theta_*) + (1-\tau)\lambda_*(\theta_*)$。将其代入前一个表达式，我们得到：

$$\begin{aligned}\sqrt{n}h'\left(\frac{d\widehat{\lambda}(\theta_*)'}{d\theta} - \frac{d\lambda_*(\theta_*)'}{d\psi}\right) &= \mathbb{G}_n\left[\tau_i^\diamond(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\\ &+ \sqrt{n}\mathbf{E}_n\left[\tau_i^\diamond(\widetilde{\lambda}_*,\theta_*)(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))'g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\theta_*)g_i(\theta_*)')\right]\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\\ &+ \sqrt{n}\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\left(\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1} - \Omega_*^\diamond(\theta_*)^{-1}\right). \quad \text{(F.23)}\end{aligned}$$

将 (F.23) 的第二项右乘 $\widehat{g}(\theta_*)$，结果等于（利用上述引理 F.4 以及 LLN）

$$\begin{aligned}&\sqrt{n}\mathbf{E}_n\left[\tau_i^\diamond(\widetilde{\lambda}_*,\theta_*)(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))'g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\theta_*)g_i(\theta_*)')\right]\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*)\\ &= \sqrt{n}(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))'\mathbf{E}_n\left[\tau_i^\dagger(\widetilde{\lambda}_*,\theta_*)g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\theta_*)g_i(\theta_*)')\right]\check{\Omega}^\dagger(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*)\\ &= -\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)].\end{aligned}$$

接下来，我们分析 (F.23) 的第三项右乘 $\widehat{g}(\theta_*)$ 后的结果：

$$\begin{aligned}&\sqrt{n}\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\left(\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1} - \Omega_*^\diamond(\theta_*)^{-1}\right)\widehat{g}(\theta_*)\\ &= h'\sqrt{n}\mathbf{E}^{Q^*(\theta_*)}\left[\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\Omega_*^\dagger(\theta_*)^{-1}\left(\Omega_*^\diamond(\theta_*) - \check{\Omega}^\diamond(\widehat{\lambda},\theta_*)\right)\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*)\\ &=: h'a'\sqrt{n}\left(\Omega_*^\diamond(\theta_*) - \check{\Omega}^\diamond(\widehat{\lambda},\theta_*)\right)\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*), \quad \text{(F.24)}\end{aligned}$$

我们打算利用 van der Vaart（2002 年，定理 6.15）对其加以控制。因此，我们验证该定理的条件：

(1) 因为根据引理 F.3 以及连续映射定理，当 $n\to\infty$ 时，$\widehat{\lambda}(\theta_*) \xrightarrow{p} \lambda_*(\theta_*)$，$|h'a'\tau_i^\diamond(\widehat{\lambda},\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{i,k} - h'a'\tau_i^\diamond(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{i,k}| \xrightarrow{p} 0$ 对每个 $k=1,\dots,d$ 都成立。此外，对于

每个 $k = 1, \ldots, d$，对于任意 $\delta > 0$ 和 $\eta > 0$，都存在一个数 $N(\delta, \eta)$，使得对于每个 $n > N(\delta, \eta)$，

$$|a'\tau_i^{\diamond}(\widehat{\lambda}, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{i,k} - a'\tau_i^{\diamond}(\lambda_*, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{i,k}|^2 \leq 4 \sup_{\lambda \in B_\delta(\lambda_*(\theta_*))} e^{2\lambda' g_i(\theta_*)}\varepsilon_{i*}^4 h'a'\widetilde{w}_i\widetilde{w}_i' ah\widetilde{w}_{ik}^2$$

概率为 $\geq \eta$。根据这一点以及假设 6 *(e)*（其中 $(\ell, \ell', i) = (2, 4, 2)$），前述上界具有可积性，因此我们可以应用控制收敛定理，从而得到：对于每个 $k = 1, \ldots, d$，$\mathbf{E}[(h'a'\tau_i^{\diamond}(\widehat{\lambda}, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{i,k} - h'a'\tau_i^{\diamond}(\lambda_*, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{i,k}$。后者是 van der Vaart（2002 年，定理 6.15）的条件之一。

(2) 根据假设 6 *(e)*（其中 $(\ell, \ell', i) = (2, 4, 2)$），对于每个 $k = 1, \ldots, d$，$\mathbf{E}\left[(h'a'\tau_i^{\diamond}(\lambda_*, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{ik})^2\right]$ 都有界。

(3) 对于每个 $k = 1$、...、d 以及任意 $\delta > 0$，定义函数类 $\mathrm{F}_{k,\delta} := \{h'a'e^{\lambda' g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{ik}; \lambda \in B_\delta(\lambda_*(\theta_*))\}$。在假设 6 (e)（其中 $(\ell, \ell', i) = (2, 4, 2)$）下，该函数类是 Donsker类，适用于某个 $\delta > 0$（依据 Van der Vaart和Wellner（1996年，定理 2.7.11和推论2.7.10））[1]。根据引理 F.3 的结果，对于任意 $\delta > 0$ 和任意 $\eta > 0$，存在一个数 $N(\delta, \eta)$，使得对于每个 $n > N(\delta, \eta)$，$h'a'\tau_i^{\diamond}(\widehat{\lambda}, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{ik} \in \mathrm{F}_{k,\delta}$ 以概率 $\geq \eta$ 对每个 $k = 1$、...、d 均成立。

因此，van der Vaart（2002 年，定理 6.15）保证：

$$\sqrt{n}h'a'\left(\check{\Omega}^{\diamond}(\widehat{\lambda}, \theta_*) - \check{\Omega}^{\diamond}(\lambda_*, \theta_*) - \left[\Omega_*^{\diamond}(\widehat{\lambda}, \theta_*) - \Omega_*^{\diamond}(\theta_*)\right]\right) \xrightarrow{p} 0$$

对于每个非随机矩阵 $a \in \mathbb{R}^{d \times p}$。为了利用这一结果，我们必须在（F.24）中加上并减去 $\check{\Omega}^{\diamond}(\lambda_*, \theta_*) + \Omega_*^{\diamond}(\lambda, \theta_*) - \Omega_*^{\diamond}(\theta_*)$，从而得到：

$$\sqrt{n}\mathbf{E}\left[e^{\lambda_*(\theta_*)' g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i'(I + \lambda_*(\theta_*)g_i(\theta_*)')\right]\left(\check{\Omega}^{\diamond}(\widehat{\lambda}, \theta_*)^{-1} - \Omega_*^{\diamond}(\theta_*)^{-1}\right)\widehat{g}(\theta_*)$$

[1]根据 Van der Vaart and Wellner（1996年，定理 2.7.11）中的上述讨论，我们将包络函数取为

$$\sup_{\|\lambda - \lambda_*\|_2 \leq \delta} \left|h'a'\tau_i^{\diamond}(\lambda, \theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_{ik}\right|.$$

$$= h'a'\sqrt{n}\left(\Omega_*^\diamond(\theta_*) - \check{\Omega}^\diamond(\lambda_*,\theta_*) - \Omega_*^\diamond(\widehat{\lambda},\theta_*) + \Omega_*^\diamond(\theta_*)\right)\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*) + o_p(1). \quad \text{(F.25)}$$

我们首先分析（F.25）中方括号内的前两项。由于根据引理 F.6，$\check{\Omega}^\diamond(\widehat{\lambda},\ \theta_*) \xrightarrow{p} \Omega_*^\diamond(\theta_*)$，并且根据大数定律，$\widehat{g}(\theta_*) \xrightarrow{p} \mathbf{E}[g_i(\theta_*)]$，因此我们得到：

$$\begin{aligned}
&- h'a'\sqrt{n}\left(\check{\Omega}^\diamond(\lambda_*,\theta_*) - \Omega_*^\diamond(\theta_*)\right)\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*) \\
&\qquad = -h'a'\mathbb{G}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\diamond(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] + o_p(1),
\end{aligned}$$

其中，正如下文将证明的，项 $\mathbb{G}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]$ 依概率有界。接下来，我们分析 (F.25) 中括号内的最后两项。根据引理 F.7，结合引理 F.6 得出的 $\check{\Omega}^\diamond(\widehat{\lambda},\theta_*) \xrightarrow{p} \Omega_*^\diamond(\theta_*)$、由大数定律得出的 $\widehat{g}(\theta_*) \xrightarrow{p} \mathbf{E}[g_i(\theta_*)]$，以及引理 F.4，我们得到：

$$\begin{aligned}
&\sqrt{n}h'a'\left[-\Omega_*^\diamond(\widehat{\lambda},\theta_*) + \Omega_*^\diamond(\lambda_*)\right]\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*) \\
&\qquad = -\sqrt{n}(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))'\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}^3 h'a'\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right]\check{\Omega}^\dagger(\widehat{\lambda},\theta_*)^{-1}\widehat{g}(\theta_*) + o_p(1) \\
&\qquad = \mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}^3 h'a'\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] + o_p(1)
\end{aligned}$$

其中，正如下文将要证明的，项 $\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\ \theta_*)\varepsilon_{i*}\widetilde{w}_i']$ 依概率有界。将所有这些要素综合起来，项 $\mathcal{A}_1$ 等于：

$$\begin{aligned}
\mathcal{A}_1 = {} & \mathbb{G}_n\left[\tau_i^\diamond(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I + \lambda_*(\theta_*)g_i(\theta_*)')\right]\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] \\
& + h'\frac{d\lambda_*(\theta_*)'}{d\theta}\sqrt{n}\mathbf{E}_n\left[g_i(\theta_*) - \mathbf{E}[g_i(\theta_*)]\right] \\
& - \mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}\widetilde{w}_i h'\widetilde{w}_{1,i}\widetilde{w}_i'(I + \lambda_*(\theta_*)g_i(\theta_*)')\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] \\
& - h'a'\mathbb{G}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\diamond(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] \\
& + h'a'\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}^3\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] + o_p(1).
\end{aligned}$$

接下来，我们分析$\mathcal{A}_2+\mathcal{A}_3$，它可以写作

$$\begin{aligned}\mathcal{A}_2+\mathcal{A}_3 = h'\sqrt{n}\mathbf{E}_n\left[\left(\tau_i(\widehat{\lambda},\theta_*)-\tau_i^\dagger(\lambda_*,\theta_*)\right)\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*)\\ -h'\mathbf{E}_n[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\widetilde{w}_{1,i}\widetilde{w}_i']\sqrt{n}\left(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*)\right)\\ -\mathbb{G}_n[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)].\end{aligned}$$

我们利用引理 F.4 和 F.8，以及根据大数定律有 $h'\mathbf{E}_n[\left(1-\tau_i^\dagger(\lambda_*,\ \theta_*)\right)\widetilde{w}_{1,\ i}\widetilde{w}_i] \xrightarrow{p} h'\mathbf{E}[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\widetilde{w}_{1,\ i}\widetilde{w}_i]$这一事实，得到：

$$\begin{aligned}\mathcal{A}_2+\mathcal{A}_3 &= -\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\\ &\quad -\mathbb{G}_n\left(\tau_i^\dagger(\lambda_*,\theta_*)\right)\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\\ &\quad +h'\mathbf{E}[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\widetilde{w}_{1,i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i]\\ &\quad -\mathbb{G}_n[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)]+o_p(1)\\ &= -\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\frac{d\lambda_*(\theta_*)}{d\theta'}-\mathbb{G}_n\left(\tau_i^\dagger(\lambda_*,\theta_*)\right)\mathbf{E}^{Q_*(\theta_*)}\left[h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\\ &\quad +h'\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i]-\mathbb{G}_n[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)]+o_p(1)\end{aligned}$$

最后，我们将$\mathcal{A}_1$,$\mathcal{A}_2$ 和$\mathcal{A}_3$ 组合起来，得到：

$$\begin{aligned}&\mathcal{A}_1+\mathcal{A}_2+\mathcal{A}_3 =\\ &\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]+h'\frac{d\lambda_*(\theta_*)'}{d\theta}\mathbb{G}_n\left[(1-\tau_i^\dagger(\lambda_*,\theta_*))\varepsilon_{i*}\widetilde{w}_i\right]\\ &\quad +h'a'\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}^3\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]\\ &-h'\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[g_i(\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]\\ &-h'a'\mathbb{G}_n\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]-\mathbb{G}_n\left(\tau_i^\dagger(\lambda_*,\theta_*)\right)\mathbf{E}^{Q_*(\theta_*)}\left[h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\\ &\quad +h'\mathbf{E}[\widetilde{w}_{1,i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i]-h'\mathbb{G}_n[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)]+o_p(1).\end{aligned}$$

根据林德伯格-列维中心极限定理，$\mathcal{A}_1+\mathcal{A}_2+\mathcal{A}_3$ 渐近正态，均值为零

均值和方差为

$$
\begin{aligned}
&h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[\widetilde{w}_i\widetilde{w}_{1i}']h \\
+Var\Big(&\tau_i^\dagger(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]+h'\frac{d\lambda_*(\theta_*)'}{d\theta}(1-\tau_i^\dagger(\lambda_*,\theta_*))\varepsilon_{i*}(\theta_*)\widetilde{w}_i \\
&+h'a'\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i'\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\varepsilon_{i*}^3\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] \\
&-h'\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i'\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[g_i(\theta_*)\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)] \\
&-h'a'\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]-\tau_i^\dagger(\lambda_*,\theta_*)\mathbf{E}^{Q_*(\theta_*)}\left[h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*) \\
&-\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\Big) \\
&+2h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*}(\theta_*)\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]\widetilde{w}_{1,i}'h\right]+2h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*) \\
&^{-1}\mathbf{E}^{Q^*}(\theta_*)\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*}(\theta_*)\left[\varepsilon_{i*}(\theta_*)^3\widetilde{w}_i\widetilde{w}_i'\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]\widetilde{w}_i'\right]ah-2h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*}(\theta_*) \\
&+2h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\frac{d\lambda_*(\theta_*)}{d\theta'}h \\
&\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^2\widetilde{w}_i\widetilde{w}_i'\right]\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*}(\theta_*)\left[g_i(\theta_*)\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\right.
\end{aligned}
$$

θ_*

$$
\begin{aligned}
&-2h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}^3\widetilde{w}_i\widetilde{w}_i'\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}[g_i(\theta_*)]\widetilde{w}_i'\right]ah \\
&-2h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_{i*}\widetilde{w}_i\right]\lambda_*(\theta_*)'\mathbf{E}^{Q^*(\theta_*)}[\widetilde{w}_i\widetilde{w}_{1,i}'h] \\
&-2h'\mathbf{E}[\widetilde{w}_{1i}\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}^{Q^*(\theta_*)}\left[\left(1-\tau_i^\dagger(\lambda_*,\theta_*)\right)\varepsilon_{i*}\widetilde{w}_i\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h=:h'H_*h.
\end{aligned}
$$

在假设 5 *(b)* 和 4 *(c)*下，矩阵 H_* 非奇异。

□

引理 F.6 设假设 *1*、*4 (a)*、*6 (e)-(f)*成立，则，

$$\left\|\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)-\Omega_*^\diamond(\theta_*)\right\|\xrightarrow{p}0$$

当 $n\to\infty$时。

证明。 记 $\varepsilon_{i*}:=\varepsilon_i(\theta_*)$。由引理 F.3， $\|\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*)\|_2\xrightarrow{p}0$，因此对于任意 $\delta>0\hat{}$ 和 $\eta>0$，存在一个数 $N(\delta,\ \eta)$，使得对于每个 $n>N(\delta,\ \eta)$， $\|\lambda(\theta_*)-\lambda_*(\theta_*)\|_2\le\delta$

$\widehat{\ }$，其概率为 $\geq \eta$。因此，对于每个 $n > N(\delta,\ \eta)$，$\lambda(\theta_*) \in B_\delta(\lambda_*(\theta_*))$。根据假设 6*(e)*，结合 $(\ell,\ \ell',\ i) = (1,\ 2,\ 0)$、$B_\delta(\lambda_*(\theta_*))$ 的紧性以及 Newey 和 McFadden（1994，引理 2.4），我们有，对于每个 $n > N(\delta,\ \eta)$，

$$\left|\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right] - \mathbf{E}\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right]\right| \leq \sup_{\lambda\in B_\delta(\lambda_*(\theta_*))}\left|\mathbf{E}_n\left[e^{\lambda'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right] - \mathbf{E}\left[e^{\lambda'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right]\right| \xrightarrow{p} 0 \quad \text{(F.26)}$$

永远 $\widehat{\ }$满足引理 F.3 结论的随机变量序列 $\lambda(\theta_*)$。

$\widetilde{\ }$此外，根据连续映射定理和引理 F.3，$e^{\lambda(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,\ j}\widetilde{w}_{i,\ k} \xrightarrow{p} e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,\ j}\widetilde{w}_{i,\ k}$对每个 $j,\ k = 1,\ \ldots,\ d$均成立。接下来，根据控制收敛定理（假设 6 *(e)* 适用于 $(\ell,\ \ell',\ i) = (1,\ 2,\ 0)$），可得

$$\left|\mathbf{E}\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right] - \mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right]\right| \xrightarrow{p} 0 \quad \text{(F.27)}$$

对每个 $j,\ k = 1,\ \ldots,\ d$均成立。将 F.26 和 (F.27) 合并，并利用 T，我们有：

$$\begin{aligned}\left\|\check{\Omega}^\diamond(\widehat{\lambda},\theta_*) - \Omega_*^\diamond(\theta_*)\right\| &\leq d\max_{j,k\in\{1,\ldots,d\}}\left|\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right] - \mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right]\right| \\ &\leq d\max_{j,k\in\{1,\ldots,d\}}\left|\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right] - \mathbf{E}\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right]\right| \\ &\quad + d\max_{j,k\in\{1,\ldots,d\}}\left|\mathbf{E}\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right] - \mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_{i*}^2\widetilde{w}_{i,j}\widetilde{w}_{i,k}\right]\right| = o_p(1).\end{aligned}$$

□

引理 F.7 设假设 *1*、*3*、*4 (a)*、*6 (d)-(e)*成立。那么，当 $n \to \infty$时，

$$\left[\Omega_*^\diamond(\widehat{\lambda},\theta_*) - \Omega_*^\diamond(\theta_*)\right] = (\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))'\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_i(\theta_*)^3\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right] + o_p(1). \quad \text{(F.28)}$$

$\widetilde{\ }$**证明**。根据一阶中值定理展开，存在一个 $\tau \in [0,1]$，使得 $\lambda_* := \tau\lambda_*(\theta_*) + (1-\tau)\widehat{\lambda}(\theta_*)$ 满足 $e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)} = e^{\lambda_*(\theta_*)'g_i(\theta_*)} + e^{\widetilde{\lambda}_*'g_i(\theta_*)}g_i(\theta_*)'(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))$。将其代入

$\widehat{\Omega}_*^\diamond(\lambda,\ \theta_*)$ 的表达式，我们得到

$$\left[\Omega_*^\diamond(\widehat{\lambda},\theta_*)-\Omega_*^\diamond(\theta_*)\right]=(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))'\mathbf{E}\left[e^{\widetilde{\lambda}_*'g_i(\theta_*)}\varepsilon_i(\theta_*)^3\widetilde{w}_i\widetilde{w}_i\widetilde{w}_i'\right]. \tag{F.29}$$

引理 F.3 和连续映射定理表明，对每个 $k=1$、.....、 d,

$$e^{\widetilde{\lambda}_*'g_i(\theta_*)}\varepsilon_i(\theta_*)^3\widetilde{w}_{i,k}\widetilde{w}_i\widetilde{w}_i'\xrightarrow{p}e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_i(\theta_*)^3\widetilde{w}_{i,k}\widetilde{w}_i\widetilde{w}_i'.$$

利用假设 6*(e)*（其中 $(\ell,\ \ell',\ i)=(1,\ 3,\ 1)$），我们可以应用控制收敛定理，从而保证：

$$\left\|\mathbf{E}\left[e^{\widetilde{\lambda}_*'g_i(\theta_*)}\varepsilon_i(\theta_*)^3\widetilde{w}_{i,k}\widetilde{w}_i\widetilde{w}_i'\right]-\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\varepsilon_i(\theta_*)^3\widetilde{w}_{i,k}\widetilde{w}_i\widetilde{w}_i'\right]\right\|\to 0.$$

证明至此完成。

□

引理 F.8 假设假设 *1*、*3*、*4 (a)*、*6 (a)-(c)*、*(e)* 成立。那么，当 $n\to\infty$ 时，

$$\begin{aligned}\sqrt{n}\mathbf{E}_n&\left[\left(\tau_i(\widehat{\lambda},\theta_*)-\tau_i^\dagger(\lambda_*,\theta_*)\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*)\\&=-\mathbb{G}_n[\tau_i^\dagger(\lambda_*,\theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\\&\quad-\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\mathbb{G}_n\left(\tau_i^\dagger(\lambda_*,\theta_*)\right)+o_p(1),\end{aligned}$$

其中， $o_p(1)$ 对任意紧集 $K\subset\mathbb{R}^p$，在 $h\in K$ 上是一致的。

证明。通过对函数 $\lambda\mapsto e^{\lambda'g_i}(\theta_*)$ 在 $\lambda_*(\theta_*)$ 附近作一阶中值定理展开，并在$\widehat{\lambda}(\theta_*)$处求值，存在一个 $\tau\in[0,1]$ 使得： $e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}=e^{\lambda_*(\theta_*)'g_i(\theta_*)}+e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)'(\widehat{\lambda}(\theta_*)-\lambda_*(\theta_*))$，其中 $\lambda_*(\theta_*):=\tau\lambda(\theta_*)+(1-\tau)\lambda_*(\theta_*)$。我们利用这一结果得到下面的第二个等式：

$$\sqrt{n}\mathbf{E}_n\left[\left(\tau_i(\widehat{\lambda},\theta_*)-\tau_i^\dagger(\lambda_*,\theta_*)\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*)$$

$$= \sqrt{n}\mathbf{E}_n\left[\left(e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)} - e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*)\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right]}$$

$$+ \sqrt{n}\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*)\left(\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right]} - \frac{1}{\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]}\right)$$

$$= \sqrt{n}(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))'\mathbf{E}_n\left[e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*)\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right]} +$$

$$\sqrt{n}\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\frac{\widehat{\lambda}(\theta_*)}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right]}\left(\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right] - \mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right]\right)\frac{1}{\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]}. \tag{F.30}$$

我们考虑 (F.30) 右边第二项中的因子 $\sqrt{n}\left(\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right] - \mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right]\right)$：

$$\begin{aligned}
&\sqrt{n}\left(\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right] - \mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right]\right) \\
&= -\sqrt{n}\mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)} - \mathbf{E}[e^{\lambda_*(\theta_*)'g_i(\theta_*)}]\right] - \sqrt{n}\left(\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right] - \mathbf{E}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]\right) \\
&= -\mathbb{G}_n\left(e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right) - \sqrt{n}\mathbf{E}_n\left[e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)'\right]\left(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*)\right), \quad \text{(F.31)}
\end{aligned}$$

其中我们再次使用了函数 $\lambda \mapsto e^{\lambda' g_i(\theta_*)}$ 在 $\lambda_*(\theta_*)$ 附近、于 $\lambda(\theta_*)$ 处取值的一阶中值定理展开。根据引理 F.3，对于任意 $\eta > 0$，存在有限的 $\delta > 0$ 和有限的 $N(\delta, \eta)$，使得对于每个 $n > N(\delta, \eta)$，$\lambda_*(\theta_*) \in B_\delta(\lambda_*(\theta_*))$ 的概率为 $\geq \eta$。由此，且由于 $\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)\right] = 0$，根据 T：$\forall n > N(\delta, \eta)$,

$$\begin{aligned}
\left\|\mathbf{E}_n\left[e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)\right]\right\|_2 &= \left\|\mathbf{E}_n\left[e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)\right] - \mathbf{E}\left[e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)\right]\right\|_2 \\
&\quad + \left\|\mathbf{E}\left[\left(e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)} - e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right)g_i(\theta_*)\right]\right\|_2 \\
&\leq \sup_{\lambda \in B_\delta(\lambda_*(\theta_*))}\left\|\mathbf{E}_n\left[e^{\lambda' g_i(\theta_*)}g_i(\theta_*)\right] - \mathbf{E}\left[e^{\lambda' g_i(\theta_*)}g_{i,k}(\theta_*)\right]\right\|_2 \\
&\quad + \left\|\mathbf{E}\left[\left(e^{\widetilde{\lambda}_*(\theta_*)'g_i(\theta_*)} - e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right)g_i(\theta_*)\right]\right\|_2
\end{aligned}$$

以概率 $\geq \eta$。根据假设 6*(a)*、$B_\delta(\lambda_*(\theta_*))$ 的紧性以及 Newey 和 McFadden

（1994，引理 2.4），前一个表达式右边的第一项依概率收敛于零。为处理第二项，我们使用连续映射定理和引理 F.3，它们保证对每个 $k = 1, \ldots, d$，$\left| \left(e^{\widetilde{\lambda}_*(\theta_*)' g_i(\theta_*)} - e^{\lambda_*(\theta_*)' g_i(\theta_*)} \right) g_{i,k}(\theta_*) \right| \xrightarrow{p} 0$。根据假设 6 *(a)* 和控制收敛定理，我们得出

$$\left\| \mathbf{E} \left[\left(e^{\widetilde{\lambda}_*(\theta_*)' g_i(\theta_*)} - e^{\lambda_*(\theta_*)' g_i(\theta_*)} \right) g_i(\theta_*)' \right] \right\| \xrightarrow{p} 0.$$

于是我们得出结论：$\left\| \mathbf{E}_n \left[e^{\widetilde{\lambda}_*(\theta_*)' g_i(\theta_*)} g_i(\theta_*) \right] \right\|_2 = o_p(1)$；又由于根据引理 F.3，$\sqrt{n} \left(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*) \right) = \mathcal{O}_p(1)$，因此 (F.31) 中的项 $\sqrt{n} \mathbf{E}_n \left[e^{\widetilde{\lambda}_*(\theta_*)' g_i(\theta_*)} g_i(\theta_*)' \right] \left(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*) \right)$ 也依概率收敛于零。

我们现在分析 (F.30) 右边的第一项。根据引理 F.4：

$$\sqrt{n}(\widehat{\lambda}(\theta_*) - \lambda_*(\theta_*))' = -\mathbb{G}_n[\tau_i^\dagger(\lambda_*, \theta_*) \varepsilon_i(\theta_*) \widetilde{w}_i'] \Omega_*^\dagger(\theta_*)^{-1} + o_p(1),$$

根据在假设 6 (c) 下成立的一致大数定律（其中 $(j, \ell, \ell') = (1, 1, 1)$），我们有 $\mathrm{E}_n \left[e^{\lambda_*(\theta_*)' g_i(\theta_*)} h' \widetilde{w}_{1,i} \widetilde{w}_i' \right] \xrightarrow{p} \mathrm{E} \left[e^{\lambda_*(\theta_*)' g_i(\theta_*)} h' \widetilde{w}_{1,i} \widetilde{w}_i' \right]$ 关于 $h \in K$ 一致成立，并且根据引理 F.3：$\widehat{\lambda}(\theta_*) \xrightarrow{p} \lambda_*(\theta_*)$。接下来还需考虑项 $\mathrm{E}_n \left[e^{\widetilde{\lambda}_*(\theta_*)' g_i(\theta_*)} g_i(\theta_*) h' \widetilde{w}_{1,i} \widetilde{w}_i' \right]$以及 $\mathrm{E}_n \left[e^{\widehat{\lambda}(\theta_*)' g_i(\theta_*)} \right]$。我们从第一项开始，它是一个 $(d \times d)$ 矩阵，我们分析该矩阵的每一个元素：根据 T，对每一个 $j, k = 1, \ldots, d$，以及 $\forall n > N(\delta, \eta)$,

$$\begin{aligned}
&\sup_{h \in K} \left| \mathbf{E}_n \left[e^{\widetilde{\lambda}_*(\theta_*)' g_i(\theta_*)} g_{i,j}(\theta_*) h' \widetilde{w}_{1,i} \widetilde{w}_{i,k} \right] - \mathbf{E} \left[e^{\lambda_*(\theta_*)' g_i(\theta_*)} g_{i,j}(\theta_*) h' \widetilde{w}_{1,i} \widetilde{w}_{i,k} \right] \right| \\
&\quad \leq \sup_{h \in K} \sup_{\lambda \in B_\delta(\lambda_*(\theta_*))} \left| \mathbf{E}_n \left[e^{\lambda' g_i(\theta_*)} g_{i,j}(\theta_*) h' \widetilde{w}_{1,i} \widetilde{w}_{i,k} \right] - \mathbf{E} \left[e^{\lambda' g_i(\theta_*)} g_{i,j}(\theta_*) h' \widetilde{w}_{1,i} \widetilde{w}_{i,k} \right] \right| \\
&\qquad\qquad + \sup_{h \in K} \left| \mathbf{E} \left[\left(e^{\widetilde{\lambda}_*(\theta_*)' g_i(\theta_*)} - e^{\lambda_*(\theta_*)' g_i(\theta_*)} \right) g_{i,j}(\theta_*) h' \widetilde{w}_{1,i} \widetilde{w}_{i,k} \right] \right|
\end{aligned}$$

在假设 6 *(f)* 下，由 $B_\delta(\lambda_*(\theta_*))$ 的紧性依概率收敛于零,

Newey 和 McFadden（1994，引理 2.4）、连续映射定理和控制收敛定理。类似地， $\forall n > N(\delta, \eta)$,

$$\begin{aligned}
&\left|\mathbf{E}_n\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)}\right] - \mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]\right| \\
&\qquad \leq \sup_{\lambda \in B_\delta(\lambda_*(\theta_*))} \left|\mathbf{E}_n\left[e^{\lambda' g_i(\theta_*)}\right] - \mathbf{E}\left[e^{\lambda' g_i(\theta_*)}\right]\right| + \left|\mathbf{E}\left[e^{\widehat{\lambda}(\theta_*)'g_i(\theta_*)} - e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]\right|
\end{aligned}$$

在假设 6 *(b)*下依概率收敛于零。将所有这些元素代回 (F.30)，得到：

$$\begin{aligned}
&\sqrt{n}\mathbf{E}_n\left[\left(\tau_i(\widehat{\lambda}, \theta_*) - \tau_i^\dagger(\lambda_*, \theta_*)\right) h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\widehat{\lambda}(\theta_*) \\
&\quad = -\mathbb{G}_n[\tau_i^\dagger(\lambda_*, \theta_*)\varepsilon_i(\theta_*)\widetilde{w}_i']\Omega_*^\dagger(\theta_*)^{-1}\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\lambda_*(\theta_*)\frac{1}{\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]} \\
&\qquad - \mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\frac{\lambda_*(\theta_*)}{\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]}\mathbb{G}_n\left(e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right)\frac{1}{\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]} + o_p(1),
\end{aligned}$$

其中， o_p**(1)** 关于 $h \in K$ 一致。

□

引理 F.9 设假设 *1–3*、*4 (a)–(b)*、*6 (b)–(c)*成立，并且 $\theta := \theta_* + \frac{\tau}{\sqrt{n}}h$ 对某个 $\tau \in [0,1]$ 以及紧集 $K \subset \mathbb{R}^p$ 中任意 h 均成立。那么，当 $n \to \infty$ 时，

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda}, \widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i\right] \xrightarrow{p} \mathbf{E}\left[\tau_i^\dagger(\lambda_*, \theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i\right]$$

关于 $h \in K$ 一致。

证明。 注意到$\theta \in B_{*,\ n}$。下面的分解成立：

$$\begin{aligned}
&\mathbf{E}_n\left(\tau_i(\widehat{\lambda}, \widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i\right) - \mathbf{E}\left[\tau_i^\dagger(\lambda_*, \theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i\right] \\
&\qquad = \mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i - \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i\right]\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]} \\
&\qquad\quad + \mathbf{E}\left[\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})} - e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right)h'\widetilde{w}_{1,i}\widetilde{w}_i\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}
\end{aligned}$$

$$+\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}h'\widetilde{w}_{1,i}\widetilde{w}_i\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}\left(\mathbf{E}[e^{\lambda_*(\theta_*)'g_i(\theta_*)}]-\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]\right)\frac{1}{\mathbf{E}[e^{\lambda_*(\theta_*)'g_i(\theta_*)}]}. \tag{F.32}$$

由引理 F.3 可知，$\widehat{\lambda}(\theta)\in B_{1/\sqrt{n}}(\lambda_*(\theta_*))$ 对于 n 较大时成立；再结合$B_{1/\sqrt{n}}(\lambda_*(\theta_*))$ 和 $B_{1/\sqrt{n}}(\theta_*)$ 的紧性，以及 Newey 和 McFadden （1994，引理 2.4）（该结论在假设 6 *(c)* 且 $j=\ell=\ell'=1$ 的条件下成立），可得

$$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i'-\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\right]\xrightarrow{p}0$$

在 $(\lambda(\theta),h)\in B_{1/\sqrt{n}}(\lambda_*(\theta_*))\times K$ 中一致成立。由引理 F.3，$\theta\to\theta_*$，根据连续映射定理和控制收敛定理（在满足假设 6 (c) 且 $j=\ell=\ell'=1$ 时，该定理适用）我们有 $\mathrm{E}\left[\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}-e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\xrightarrow{p}0$ 在 $h\in K$ 中一致成立。类似地，在假设 6 (b) 下，$\mathrm{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]\xrightarrow{p}\mathrm{E}[e^{\lambda_*(\theta_*)'g_i(\theta_*)}]$ 在 $h\in K$ 中一致成立。因此，(F.32) 在 $h\in K$ 中依概率一致收敛于零。

□

引理 F.10 设假设 *1–3*、*4 (a)–(b)*、*6 (b)–(c)* 成立，并且 $\theta:=\theta_*+\frac{\tau}{\sqrt{n}}h$ 对于某个 $\tau\in[0,1]$ 以及紧集 $K\subset\mathbb{R}^p$ 中的任意 h。那么，当 $n\to\infty$ 时，

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]\xrightarrow{p}\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\right)^2\right]$$

关于 $h\in K$ 一致收敛。

证明。注意到 $\theta\in B_{*,n}$。以下分解成立：

$$\begin{aligned}
&\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]-\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\right)^2\right]\\
&\quad=\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2-\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}\\
&\quad+\left(\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]-\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\right)^2\right]\right)\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}
\end{aligned}$$

$$+\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\right)^2\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}\left(\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]-\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]\right)\frac{1}{\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]}. \tag{F.33}$$

根据引理 F.3（该引理保证当 n足够大时，$\lambda(\theta) \in B_{1/\sqrt{n}}(0)$），结合 $B_{1/\sqrt{n}}(0)$和 $B_{1/\sqrt{n}}(\theta_\circ)$的紧性，以及 Newey 和 McFadden（1994，引理 2.4）（该引理在假设 6 *(c)* 满足 $j=1$和 $\ell=\ell'=2$时成立），可得

$$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2-\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]\right]\xrightarrow{p}0$$

在 $(\lambda(\theta),h) \in B_{1/\sqrt{n}}(\lambda_*(\theta_*)) \times K$ 上一致成立。根据引理 F.3，由于 $\theta \to \theta_*$，并利用连续映射定理和控制收敛定理（在 $j=1$ 和 $\ell=\ell'=2$ 条件下，假设 6 (c) 可保证其成立），我们有 $\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]\xrightarrow{p}\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\lambda_*(\theta_*)\right)^2\right]$ 在 $h\in K$ 上一致成立。类似地，在假设 6 (b) 下，$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]\xrightarrow{p}\mathbf{E}[e^{\lambda_*(\theta_*)'g_i(\theta_*)}]$ 在 $h\in K$ 上一致成立。因此，(F.33) 关于 $h\in K$ 一致依概率收敛于零。

□

引理 F.11 设假设 *1–3*、*4 (a)–(b)*、*6 (b)*、*(g)*成立，且 $\theta := \theta_* + \frac{\tau}{\sqrt{n}}h$对某个 $\tau\in[0,1]$ 以及紧集 $K\subset\mathbb{R}^p$中的任意 h 均成立。那么，当 $n\to\infty$时，

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h\xrightarrow{p}\mathbf{E}\left[\tau_i^\dagger(\lambda_*,\theta_*)g_i(\theta_*)\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h$$

关于 $h\in K$一致。

证明。注意到$\theta\in B_{*,\ n\circ}$ 以下分解成立：

$$\begin{aligned}
&\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h-\mathbf{E}\left[\tau_i(\lambda_*,\theta_*)g_i(\theta_*)\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h\\
&\quad=\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h-\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}\\
&\quad+\left(\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]-\mathbf{E}\left[\tau_i^\diamond(\lambda_*,\theta_*)g_i(\theta_*)\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h\right)\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}
\end{aligned}$$

$$+\mathbf{E}\left[\tau_i^{\diamond}(\lambda_*,\theta_*)g_i(\theta_*)\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}\frac{\left(\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]-\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]\right)}{\mathbf{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]}. \tag{F.34}$$

根据引理 F.3（其保证当 n 充分大时，$\lambda(\theta) \in B_{1/\sqrt{n}}(\lambda_*(\theta_*))$），由$B_{1/\sqrt{n}}(\lambda_*(\theta_*))$的紧性以及 $B_{1/\sqrt{n}}(\theta_*)$的紧性，并由Newey 和 McFadden（1994，引理 2.4）（该结论在假设 6 *(f)*下成立），我们有

$$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]-\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\right]\xrightarrow{p}0$$

在 $(\lambda(\theta),h)\in B_{1/\sqrt{n}}(\lambda_*(\theta_*))\times K$ 中一致成立。根据引理 F.3，由 $\theta\to\theta_*$ 在 $h\in K$ 中一致成立，由连续映射定理和控制收敛定理（在假设 6 (f) 下成立），我们有 $\mathrm{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\xrightarrow{p}\mathrm{E}\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}g_i(\theta_*)\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]$ 在 $h\in K$ 中一致成立。类似地，在假设 6 (b) 下，$\mathrm{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]\xrightarrow{p}\mathrm{E}_n\left[e^{\lambda_*(\theta_*)'g_i(\theta_*)}\right]$在 $h\in K$ 中一致成立。因此，(F.34) 关于 $h\in K$ 一致依概率收敛至零。

□

引理 F.12 设假设 *1–3*、*4 (a)–(b)*、*6 (b)–(g)* 成立，并且 $\theta:=\theta_*+\frac{\tau}{\sqrt{n}}h$对某个 $\tau\in[0,1]$ 以及紧集 $K\subset\mathbb{R}^p$中的任意 h 成立。那么，当 $n\to\infty$时，

$$h'\frac{d^2[\widehat{\lambda}(\widetilde{\theta})'\widehat{g}(\widetilde{\theta})]}{d\theta d\theta'}h=\mathcal{O}_p(1)$$

关于 $h\in K$一致成立。

证明。注意到 $\widetilde{\theta}\in B_{*,n}$。首先，由于 $d^2\widehat{g}(\theta)/d\theta d\theta'=0$，因此 $h'\frac{d^2[\widehat{\lambda}(\widetilde{\theta})'\widehat{g}(\widetilde{\theta})]}{d\theta d\theta'}h=h'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\frac{d\widehat{g}(\widetilde{\theta})}{d\theta'}h\asymp 2h'\frac{d\lambda(\theta)'}{d\theta}\mathrm{E}_n\left[\widetilde{w}_i\widetilde{w}_{1,i}'\right]h$。根据假设 7 (c)、假设 6 (c)（其中 $j=\ell=\ell'=1$，$\ell''=0$）和 (g)，我们应用一致大数定律和引理 F.6 得到：$\asymp\mathrm{E}_n\left[\widetilde{w}_i\widetilde{w}_{1,i}'\right]h\xrightarrow{p}\mathrm{E}\left[\widetilde{w}_i\widetilde{w}_{1,i}'\right]h$ 以及 $h'\frac{d\lambda(\theta')}{d\theta}\xrightarrow{p}h'\frac{d\lambda(\theta_*')}{d\theta}$ 关于 $h\in K_n$一致成立。

接下来，我们考虑右边的第一项：

$$h'\frac{d^2\widehat{\lambda}(\widetilde{\theta})}{d\theta d\theta'}\widehat{g}(\widetilde{\theta})h = \frac{d\mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\widetilde{\theta})g_i(\widetilde{\theta})')\right]\check{\Omega}^\diamond(\widehat{\lambda},\widetilde{\theta})^{-1}}{d\theta'}\widehat{g}(\widetilde{\theta})h$$
$$= h'\mathbf{E}_n\left(\frac{d}{d\theta'}\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right)h\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\widetilde{\theta})g_i(\widetilde{\theta})') + \tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\widetilde{w}_{1,i}\widetilde{w}_i'\frac{d}{d\theta'}\left(\widehat{\lambda}(\widetilde{\theta})g(\widetilde{\theta})'\right)h\right)\check{\Omega}^\diamond(\widehat{\lambda},\widetilde{\theta})^{-1}\widehat{g}(\widetilde{\theta})$$
$$-\mathbf{E}_n\left[\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\widetilde{\theta})g_i(\widetilde{\theta})')\right]\check{\Omega}^\diamond(\widehat{\lambda},\widetilde{\theta})^{-1}$$
$$\times\mathbf{E}_n\left(\frac{d}{d\theta}\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right)\varepsilon_i(\widetilde{\theta})^2\widetilde{w}_i\widetilde{w}_i' + 2\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\varepsilon_i(\widetilde{\theta})\widetilde{w}_{1,i}'h\widetilde{w}_i\widetilde{w}_i'\right)\check{\Omega}^\diamond(\widehat{\lambda},\widetilde{\theta})^{-1}\widehat{g}(\widetilde{\theta}). \quad \text{(F.35)}$$

现在我们分析这些项各自的渐近行为。首先，根据引理 F.6：$\left\|\check{\Omega}^\diamond(\widehat{\lambda},\theta_*) - \Omega_*^\diamond(\theta_*)\right\| \xrightarrow{p} 0$，且当 $n\to\infty$ 时；根据连续映射定理：$\check{\Omega}^\diamond(\widehat{\lambda},\theta_*)^{-1} \xrightarrow{p} \Omega_*^\diamond(\theta_*)^{-1}$。此外，

$$\frac{d}{d\theta'}\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right)h = \tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\left[g_i(\widetilde{\theta})'\frac{d\widehat{\lambda}(\widetilde{\theta})}{d\theta'}h + \widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right].$$

我们将该表达式代入（F.35）。因此，第一个因子变为：

$$h'\mathbf{E}_n\left(\frac{d}{d\theta'}\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right)h\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\widetilde{\theta})g_i(\widetilde{\theta})') + \tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\widetilde{w}_{1,i}\widetilde{w}_i'\frac{d}{d\theta'}\left(\lambda_*(\widetilde{\theta})g(\widetilde{\theta})'\right)h\right)$$
$$= h'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\mathbf{E}_n\left(\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\left[g_i(\widetilde{\theta})+\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\widehat{\lambda}(\widetilde{\theta})g_i(\widetilde{\theta})')\right)$$
$$+\mathbf{E}_n\left(\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i'\left[\frac{d}{d\theta'}\widehat{\lambda}(\widetilde{\theta})hg(\widetilde{\theta})' + \widehat{\lambda}(\widetilde{\theta})\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\right)$$

根据假设 6 （c），当 $j=2$，$\ell=\ell'=\ell''=1$ 且当 $j=1$，$\ell=\ell''=1$，$\ell'=2$时；根据假设 6 （g），当（j，ℓ）=（1，2）以及（j，ℓ）=（2，1）时；（c）由此可以使用一致大数定律和控制收敛定理，前面的表达式依概率收敛于：

$$h'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\mathbf{E}\left(\tau_i^\diamond(\lambda_*,\theta_*)\left[g_i(\theta_*)+\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]h'\widetilde{w}_{1,i}\widetilde{w}_i'(I+\lambda_*(\theta_*)g_i(\theta_*)')\right)$$
$$+\mathbf{E}\left(\tau_i^\diamond(\lambda_*,\theta_*)h'\widetilde{w}_{1,i}\widetilde{w}_i'\left[\frac{d}{d\theta'}\widehat{\lambda}(\theta_*)hg(\theta_*)' + \lambda_*(\theta_*)\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\right)$$
$$= h'\frac{d\widehat{\lambda}(\widetilde{\theta})'}{d\theta}\mathbf{E}\left(\tau_i^\diamond(\lambda_*,\theta_*)\left[g_i(\theta_*)+\lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]h'\widetilde{w}_{1,i}\widetilde{w}_i'(2I+\lambda_*(\theta_*)g_i(\theta_*)')\right) \quad \text{(F.36)}$$

~~关于 $h\in K_n$ 一致成立。最后，$h'\frac{d\lambda(\theta)'}{d\theta}$ 依概率收敛至 $\mathbf{E}\left[e^{\lambda_*(\theta)'g_i(\theta)}\widetilde{w}_{1,\,i}\widetilde{w}_i'(I+\lambda_*(\theta)g_i(\theta)')\right]$

在假设 6 *(c)* 下，关于 $h \in K_n$ 一致成立，其中 $j = 2$，$\ell = \ell' = 1$，$\ell'' = 0$，且满足假设 6 *(f)*。现在我们分析 (F.35) 中的第二项：

$$\begin{aligned}
&\mathbf{E}_n\left(\frac{d}{d\theta}\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right)\varepsilon_i(\widetilde{\theta})^2\widetilde{w}_i\widetilde{w}_i' + 2\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\varepsilon_i(\widetilde{\theta})\widetilde{w}_{1,i}'h\widetilde{w}_i\widetilde{w}_i'\right)\\
&= \mathbf{E}_n\left(\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\left[g_i(\widetilde{\theta})'\frac{d\widehat{\lambda}(\widetilde{\theta})}{d\theta'}h + \widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\varepsilon_i(\widetilde{\theta})^2\widetilde{w}_i\widetilde{w}_i' + 2\tau_i^\diamond(\widehat{\lambda},\widetilde{\theta})\varepsilon_i(\widetilde{\theta})\widetilde{w}_{1,i}'h\widetilde{w}_i\widetilde{w}_i'\right) \quad \text{(F.37)}
\end{aligned}$$

在假设 6 *(e)* 满足 $i = \ell = 1, \ell' = 3$，且假设 6 *(c)* 满足 $(j, \ell, \ell', \ell'') = (2, 1, 1, 1)$ 的条件下，它依概率收敛于

$$\mathbf{E}\left(\tau_i^\diamond(\lambda_*,\theta_*)\left[g_i(\theta_*)'\frac{d\lambda_*(\theta_*)}{d\theta'}h + \lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\varepsilon_i(\theta_*)^2\widetilde{w}_i\widetilde{w}_i' + 2\tau_i^\diamond(\lambda_*,\theta_*)\varepsilon_i(\theta_*)\widetilde{w}_{1,i}'h\widetilde{w}_i\widetilde{w}_i'\right) \quad \text{(F.38)}$$

关于 $h \in K_n$ 一致收敛。

综合以上结果，我们得到

$$\begin{aligned}
&h'\frac{d^2[\widehat{\lambda}(\widetilde{\theta})'\widehat{g}(\widetilde{\theta})]}{d\theta d\theta'}h \xrightarrow{p}\\
&h'\frac{d\lambda_*(\theta_*)'}{d\theta}\Omega_*^\diamond(\theta_*)\mathbf{E}\left(\tau_i^\diamond(\lambda_*,\theta_*)\left[g_i(\theta_*) + \lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]h'\widetilde{w}_{1,i}\widetilde{w}_i'(2I + \lambda_*(\theta_*)g_i(\theta_*)')\right)\Omega_*^\diamond(\theta_*)\mathbf{E}[g_i(\theta_*)]\\
&\quad - h'\frac{d\lambda_*(\theta_*)'}{d\theta}\Omega_*^\diamond(\theta_*)\mathbf{E}\Big(\tau_i^\diamond(\lambda_*,\theta_*)\left[g_i(\theta_*)'\frac{d\lambda_*(\theta_*)}{d\theta'}h + \lambda_*(\theta_*)'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\varepsilon_i(\theta_*)^2\widetilde{w}_i\widetilde{w}_i'\\
&\quad + 2\tau_i^\diamond(\lambda_*,\theta_*)\varepsilon_i(\theta_*)\widetilde{w}_{1,i}'h\widetilde{w}_i\widetilde{w}_i'\Big)\Omega_*^\diamond(\theta_*)\mathbf{E}[g_i(\theta_*)] - h'\frac{d\lambda_*(\theta_*)'}{d\theta}\mathbf{E}\left[\widetilde{w}_i\widetilde{w}_{1,i}'\right]h\\
&\quad = h'\frac{d^2\lambda_*(\theta_*)}{d\theta d\theta'}\mathbf{E}[g_i(\theta_*)]h - h'\frac{d\lambda_*(\theta_*)'}{d\theta}\mathbf{E}\left[\widetilde{w}_i\widetilde{w}_{1,i}'\right]h
\end{aligned}$$

关于 $h \in K_n$ 一致。由此得出，$h'\frac{d^2[\widehat{\lambda}(\widetilde{\theta})'\widehat{g}(\widetilde{\theta})]}{d\theta d\theta'}h$ 关于 $h \in K_n$ 一致地依概率有界。

□

F.5.1 基准模型的技术结果（外生性）

引理 F.13 假设假设 *3*、*4(a)* 和 *6(d)-(e)* 成立。那么，对于任意给定的 $\theta \in B_{\circ,n}$，

$$\sqrt{n}\|\widehat{\lambda}(\theta)\|_2 = \mathcal{O}_p(1)$$

或者，等价地，对于任意 $\eta > 0$，存在有限的 $\delta > 0$ 和有限的 N（δ、η），使得：$\forall n > N$（δ、η），

$$P\left(\sqrt{n}\|\widehat{\lambda}(\theta)\|_2 < \delta\right) > \eta.$$

证明。 证明过程与引理 F.3 的证明相同，因此略去。

□

引理 F.14 设假设 *2-3*、*4(a)-(b)*、*6(b)-(c)*成立，且 $\theta := \theta_\circ + \frac{\tau}{\sqrt{n}}h$ 对于某个 $\tau \in [0,1]$ 以及紧集 $K \subset \mathbb{R}^p$ 中的任意 h。那么，随着 $n \to \infty$，

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i\right] \xrightarrow{p} \mathbf{E}\left[h'\widetilde{w}_{1,i}\widetilde{w}_i\right]$$

关于 $h \in K$ 一致地。

证明。 注意到 $\widetilde{\theta} \in B_{\circ,\ n\circ}$ 由于 $\lambda_*(\theta_\circ) = 0$，因此 $\tau_i^\dagger(\lambda_*,\ \theta_\circ) = 1$，从而有如下分解：

$$\begin{aligned}
&\mathbf{E}_n\left(\tau_i(\widehat{\lambda},\widetilde{\theta})h'\widetilde{w}_{1,i}\widetilde{w}_i\right) - \mathbf{E}\left[h'\widetilde{w}_{1,i}\widetilde{w}_i\right] \\
&\qquad = \mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i - \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i\right]\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]} \\
&\qquad\quad + \mathbf{E}\left[\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})} - 1\right)h'\widetilde{w}_{1,i}\widetilde{w}_i\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]} \\
&\qquad\quad + \mathbf{E}\left[h'\widetilde{w}_{1,i}\widetilde{w}_i\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}\left(1 - \mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]\right). \quad \text{(F.39)}
\end{aligned}$$

根据引理 F.13（其保证当 n 足够大时，$\widehat{\lambda}(\theta) \in B_{1/\sqrt{n}}(0)$），由 $B_{1/\sqrt{n}}(0)$ 及 $B_{1/\sqrt{n}}(\theta_\circ)$ 的紧性，以及在假设 6 *(c)* 与 $j = \ell = \ell' = 1$ 下成立的 Newey 和 McFadden（1994，引理 2.4），我们有

$$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i' - \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}h'\widetilde{w}_{1,i}\widetilde{w}_i'\right]\right] \xrightarrow{p} 0$$

在 $(\widehat{\lambda}(\theta),\ h)$ 中一致地 $\in B_{1/\sqrt{n}}(0) \times K$。根据引理 F.13，由 $\theta \to \theta_\circ$ 在$h \in K$上一致收敛、连续映射定理和控制收敛定理（在假设 6 *(c)* 且 $j = \ell = \ell' = 1$ 的条件下成立），可得 $\mathbf{E}\left[\left(e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})} - 1\right)h'\widetilde{w}_{1,i}\widetilde{w}_i'\right] \xrightarrow{p} 0$ 在 $h \in K$ 中一致地成立。类似地，在假设 6 *(b)* 下，$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right] \xrightarrow{p} 1$在 $h \in K$ 中一致地成立。因此，(F.39) 在 $h \in K$ 中一致地依概率收敛于零。

□

引理 F.15 设假设 *2*、*4 (a)-(b)*、*6 (b)-(c)* 成立，并设 $\theta := \theta_\circ + \frac{\tau}{\sqrt{n}}h$ 对某个 $\tau \in [0,1]$ 以及紧集 $K \subset \mathbb{R}^p$ 中任意 h 均成立。则当 $n \to \infty$时，

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right] \xrightarrow{p} 0$$

关于 $h \in K$一致地。

证明。 注意到$\theta \in B_{\circ,\ n\circ}$。以下分解成立：

$$\begin{aligned}
&\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right] \\
&\quad= \mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2 - \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]} \\
&\qquad\qquad + \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}. \quad \text{(F.40)}
\end{aligned}$$

根据引理 F.13，对于 n 足够大时可保证$\widehat{\lambda}(\theta) \in B_{1/\sqrt{n}}(0)$，再结合 $B_{1/\sqrt{n}}(0)$及 $B_{1/\sqrt{n}}(\theta_\circ)$的紧性，以及 Newey 和 McFadden（1994，引理 2.4）在假设条件

6 （*c*）下对 $j = 1$和 $\ell = \ell' = 2$成立，我们有

$$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2 - \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right]\right] \xrightarrow{p} 0$$

$\widehat{\ }$ $\widetilde{\ }$ $\bar{\ }$ $\widetilde{\ }$关于 $(\lambda(\theta)\ h), \in B_{1/\sqrt{n}}$ $(0) \times K$一致成立。根据引理 F.13，考虑到 $\theta \to \theta_\circ$在 $h \in K$上一致成立，根据连续映射定理和控制收敛定理（在假设 6 *(c)* 满足 $j = 1$和$\ell = \ell' = 2$的条件下，该定理适用），可得 $\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\left(h'\widetilde{w}_{1,i}\widetilde{w}_i'\widehat{\lambda}(\widetilde{\theta})\right)^2\right] \xrightarrow{p} 0$在 $h \in K$上一致成立。类似地，$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right] \xrightarrow{p} 1$在假设 6 *(b)* 下对 $h \in K$一致成立。因此，(F.40) 关于 $h \in K$一致依概率收敛于零。

□

$\widetilde{\ }$**引理 F.16** 设假设 *2*、*4 (a)-(b)*、*6 (b),(g)* 成立，且令 $\theta := \theta_\circ + \frac{\tau}{\sqrt{n}}h$对某个 $\tau \in [0,1]$以及紧集 $K \subset \mathbb{R}^p$中任意 h 均成立。那么，当 $n \to \infty$时，

$$\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h \xrightarrow{p} 0$$

对 $h \in K$一致。

证明。以下分解成立：

$$\begin{aligned}
&\mathbf{E}_n\left[\tau_i(\widehat{\lambda},\widetilde{\theta})g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'\right]h \\
&\quad= \mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h - \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]} \\
&\qquad\qquad + \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\frac{1}{\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}\right]}. \quad \text{(F.41)}
\end{aligned}$$

$\widehat{\ }$ $\widetilde{\ }$根据引理 F.13（它保证当 n 充分大时 $\lambda(\theta) \in B_{1/\sqrt{n}}(0)$），根据 $B_{1/\sqrt{n}}(0)$与 $B_{1/\sqrt{n}}(\theta_\circ)$的紧性，$\bar{\ }$以及在假设 6*(f)*下成立的Newey与McFadden（1994，引理 2.4），我们有

$$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h - \mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})'g_i(\widetilde{\theta})}g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})'\widetilde{w}_i\widetilde{w}_{1,i}'h\right]\right] \xrightarrow{p} 0$$

$\widehat{\ }$ $\widetilde{\ }$ $\bar{\ }$ $\widetilde{\ }$在$(\lambda(\theta)\ h) \in B_{1/\sqrt{n}}(0)$ $\times K$上一致成立。根据引理 F.13，根据 $\theta \to \theta_\circ$ 在 $h \in K$上一致成立这一事实，并根据连续映射定理与控制收敛定理（它们在假设 6 *(f)*下成立），我们有$\mathbf{E}\left[e^{\widehat{\lambda}(\widetilde{\theta})' g_i(\widetilde{\theta})} g_i(\widetilde{\theta})\widehat{\lambda}(\widetilde{\theta})' \widetilde{w}_i \widetilde{w}_{1,i}' h\right] \xrightarrow{p} 0$ 在 $h \in K$上一致成立。类似地，$\mathbf{E}_n\left[e^{\widehat{\lambda}(\widetilde{\theta})' g_i(\widetilde{\theta})}\right] \xrightarrow{p} 1$ 在假设 6 *(b)* 下于 $h \in K$ 上一致成立。因此，(F.41)在 $h \in K$上依概率一致收敛于零。

□

F.6 扩展模型的技术结果

引理 F.17 假设假设 *3*、*4(a)–(b)* 中的 Θ 被 Ψ 替换，且假设 *6(d)–(e)* 中的 θ_* 和 $\lambda_*(\theta_*)$ 分别被 $\theta_\circ$ 和 0 替换。则，

$$\sqrt{n}\widehat{\lambda}(\psi_\circ) = -\Omega_{\psi_\circ}^{-1}\mathbb{G}_n\left[\varepsilon_i(\theta_\circ)\widetilde{w}_i - \begin{pmatrix} \mathbf{E}[\varepsilon_i(\theta_\circ)x_i] \\ 0 \\ 0 \end{pmatrix}\right] + o_p(1),$$

其中 $\Omega_{\psi_\circ} := \mathbf{E}[g_e(w_i, \psi_\circ)g_e(w_i, \psi_\circ)']$。

证明。对函数 $\lambda \mapsto \mathrm{E}_n[e^{\widetilde{\lambda}' g_e(w_i,\psi_\circ)} g_e(w_i, \psi_\circ)]$ 在$\lambda_*(\psi_\circ) = 0$处进行二阶中值定理展开，并在 $\widehat{\lambda}(\psi_\circ)$ 处求值，可得：$\widehat{\lambda}(\psi_\circ) = -\mathrm{E}[g_e(w_i, \psi_\circ)g_e(w_i, \psi_\circ)']^{-1}\mathrm{E}_n[g_e(w_i, \psi_\circ)] + \mathrm{O}_p(1/n)$。其中，$\mathrm{O}_p(1/n)$ 来自大数定律（这意味着 $\mathrm{E}_n[g_e(w_i, \psi_\circ)g_e(w_i, \psi_\circ)']$ 依概率收敛于 $\mathrm{E}[g_e(w_i, \psi_\circ)g_e(w_i, \psi_\circ)']$），以及引理 F.18 和假设 6(e)（其中 $(\ell, \ell', i) = (1, 3, 1)$，且将 θ_* 替换为 $\theta_\circ$），这些条件允许控制中值定理展开中的二次项。此外，$\mathrm{E}[g_e(w_i, \psi_\circ)] = 0$。因此，

$$\sqrt{n}\widehat{\lambda}(\psi_\circ) = -\mathbf{E}[g_e(w_i, \psi_\circ)g_e(w_i, \psi_\circ)']^{-1}\mathbb{G}_n[g_e(w_i, \psi_\circ)] + o_p(1). \tag{F.42}$$

□

引理 F.18 假设假设 *3*、*4(a)–(b)*（将 Θ 替换为 Ψ）和假设 *6(e)*（分别将 θ_* 和$\lambda_*(\theta_*)$ 替换为 $\theta_\circ$和 0）成立。那么，对于任意给定的 $\psi \in \Psi$，

$$\sqrt{n}\|\widehat{\lambda}(\psi) - \lambda_*(\psi)\|_2 = \mathcal{O}_p(1).$$

证明。为证明该引理，我们打算将 Van der Vaart and Wellner（1996年，推论 3.2.6）应用于 $\mathbb{M}_n = -\mathbb{E}_n\left[e^{\lambda' g_{e,i}(\psi)}\right]$。我们固定 ψ，并将 $\mathbb{M}_n$ 视为数据 w_i 和 λ的函数。对于 $\lambda_* = \arg\max_{\lambda\in\mathbb{R}^d} \mathbb{M}$，其中 $\mathbb{M} := -\mathbb{E}\left[e^{\lambda' g_{e,i}(\psi)}\right]$，定义函数集 $\mathbb{M}_\delta := \left\{ w_i \mapsto e^{\lambda'_* g_{e,i}(\psi)} - e^{\lambda' g_{e,i}(\psi)}; \lambda \in B_\delta(\lambda_*) \right\}$。令 M_δ 表示 $\mathbb{M}_\delta$的一个包络函数。由于 $\mathbb{M}_\delta$ 中的函数关于索引参数 $\lambda \in B_\delta(\lambda_*)$ 满足利普希茨条件，并且

$$|e^{\lambda_2' g_{e,i}(\psi)} - e^{\lambda_1' g_{e,i}(\psi)}| \leq 2\delta \|g_{e,i}(\psi) e^{\widetilde{\lambda}_{12}' g_{e,i}(\psi)}\|_2$$

~对于每个 w_i 以及每个 λ_1，$\lambda_2 \in B_\delta(\lambda_*)$，由中值定理，对于给定的 $\tau \in (0, 1)$，存在某个值使得 $\lambda_{12} := \lambda_1 + \tau(\lambda_2 - \lambda_1)$；再由柯西-施瓦茨不等式，可得一个包络函数为

$$M_\delta = 2\delta \|g_{e,i}(\psi)\|_2 \sup_{\lambda_1,\lambda_2 \in B_\delta(\lambda_*)} e^{\widetilde{\lambda}_{12}' g_{e,i}(\psi)}$$

（参见 Van der Vaart and Wellner（1996，定理 2.7.11）之前的讨论）。此外，存在常数 K，使得对每个 $0 < \varepsilon < 2\delta$，若 $\|M_\delta\|_{P,2} < C/\sqrt{n}$，则 $\mathbb{M}_\delta$ 的括号覆盖数 $N_{[]}(\varepsilon\|M_\delta\|_P, \mathbb{M}_\delta, \|\cdot\|_P, 2)$ 根据 Van der Vaart（1998，例 19.7）可由 $K(2\delta)^d/\varepsilon^d$ 上界；在假设 6 (d) 和假设 4 (b) 下，将 Θ 替换为 Ψ，即可保证 $\delta = 1/\sqrt{n}$ 满足这一条件。因此，括号积分 $J_{[]}(1, \mathbb{M}_\delta, \|\cdot\|_P, 2) := \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon\|M_\delta\|_P, \mathbb{M}_\delta, \|\cdot\|_P, 2)}d\varepsilon$ 有界，且函数类 $\mathbb{M}_\delta$ 是 P-Glivenko-Cantelli 类（例如，参见 Van der Vaart and Wellner（1996，定理 2.5.6）及其简化版本；该版本的讨论见 Van der Vaart and Wellner（1996，第 243–244 页））。

因此，根据 Van der Vaart and Wellner（1996年，推论 3.2.3），对每个给定的 $\psi \in B(\circ\ n)$，$\widehat{\lambda}(\psi) \xrightarrow{p} \lambda_*(\psi)$。最后，为了应用 Van der Vaart and Wellner（1996年，推论 3.2.6），我们

需要对 $\mathbf{E}(\sup_{f\in\mathcal{M}_{1/\sqrt{n}}}|\sqrt{n}\mathbf{E}_n[f-\mathbf{E}(f)]|)$ 给出上界。根据 Van der Vaart and Wellner（1996年，推论 3.2.6）之后的讨论（并且由于根据前面的论证，$J_{[]}(1,\mathcal{M}_\delta,\|\cdot\|_{P,2})$ 有界），后者以上界为某个常数乘以 $\|M_{1/\sqrt{n}}\|_{P,2}$。由于 $\|M_{1/\sqrt{n}}\|_{P,2}\leq C/\sqrt{n}$，因此 $r_n^4\|M_{1/r_n}\|_{P,2}^2\leq n$ 得出 $r_n\asymp\sqrt{n}$。因此，Van der Vaart and Wellner（1996年，推论 3.2.6）的结果在 $r_n\leq\sqrt{n}$ 下成立。这证明了引理的结论。

□

G 工具变量设定下指数倾斜经验似然问题的对偶理论细节

在本节中，我们首先考虑一个假定已知的一般矩函数 $g(w,\theta)$：$\mathbb{R}^{d+1}\times\mathbb{R}^p\to\mathbb{R}^d$。然后，在必要时，关于 $g(w,\theta)$的条件将用关于 $g_b(w,\theta)$和$g_e(w,\theta)$的条件来表述。定义指数倾斜经验似然的原始优化问题$\mathcal{P}_s$ 由下式给出：

$$(\mathcal{P}_s)\qquad \min_{q_1,\ldots,q_n>0}\sum_{i=1}^n[q_i\log(nq_i)]$$
$$\text{subject to }\sum_{i=1}^n q_i=1,\qquad\text{and}\qquad\sum_{i=1}^n q_ig(w_i,\theta)=0,\tag{G.1}$$

它依赖于参数向量 $\theta\in\Theta$，或者更简单地

$$(\mathcal{P}_s)\qquad\min_{q_1,\ldots,q_n\in\mathcal{Q}_n(\theta)}\sum_{i=1}^n[q_i\log(nq_i)],\tag{G.2}$$

其中，$\mathcal{Q}_n(\theta)$ 是优化问题 $\mathcal{P}_s$ 的可行集，其定义为 s:

$$\mathcal{Q}_n(\theta):=\left\{\{q_i\}_{i=1}^n;\sum_{i=1}^n q_i=1,\ q_1>0,\ldots,q_n>0,\text{ and }\sum_{i=1}^n g(w_i,\theta)q_i=0\right\}.$$

我们注意到，$Q_n(\theta)$ 是 $\bigcup_{i=1}^n g(w_i,\theta)$ 的凸包。P_s 中的下标强调这是样本中的原问题。函数 $u \mapsto \phi(u) \equiv \log(u)u$ 是 P_s 第一行中的被加项，是定义在 $\mathbb{R}_+$ 上的严格凸函数。用 $\phi_+(\cdot)$ 表示 ϕ 到 $\mathbb{R}$ 的延拓，定义如下：若 $u>0$，则 $\phi_+(u)=\phi(u)$；若 $u=0$，则 $\phi_+(u)=0$；若 $u<0$，则 $\phi_+(u)=\infty$。函数 $u \mapsto \phi_+(u)$ 是严格凸且本质光滑的，并满足 $(0,\infty) \subset dom(\phi_+)$（参见 Borwein 和 Lewis（1991年，例5.6））。ϕ_+ 在 $\mathbb{R}_+$ 上的严格凸性以及 Borwein 和 Lewis（1991年，命题2.11）意味着 (P_s) 的任一最优解都是唯一的。

为将维数从 n 降至 d，使用对偶问题（记为 $\mathcal{D}_s$）会很方便。拉格朗日对偶问题为

$$(\mathcal{D}_s) \qquad \sup_{\gamma\in\mathbb{R},\lambda\in\mathbb{R}^d} \gamma - \frac{1}{n}\sum_{i=1}^n \phi_+^*(\gamma+\lambda' g(w_i,\theta)), \tag{G.3}$$

其中，$\phi_+^*(t)$ 是 ϕ_+ 的凸共轭：$\mathbb{R} \to (-\infty, \infty]$ 其定义为 $t \mapsto \phi_+^*(t) = e^{t-1}$。关于 γ 求解 (G.3) 中的优化问题，可得到对偶最优值 $\hat{\gamma}(\lambda, \theta) = 1 - \log(\mathbf{E}_n[\exp\{\lambda' g(w_i, \theta)\}])$；将其代回 (G.3) 后，所得对偶问题$\mathcal{D}_s$ 为：

$$(\mathcal{D}_s) \qquad \sup_{\lambda\in\mathbb{R}^d} -\log\left(\frac{1}{n}\sum_{i=1}^n e^{\lambda' g(w_i,\theta)}\right). \tag{G.4}$$

下一个引理建立了原始指数倾斜经验似然问题与对偶问题之间的关系，并回顾了保证对偶性结果的假设。

引理 *G.1* 假设对于所有 $j=1, \ldots\ldots, d$ 以及任意 $\theta \in \Theta$，$\sum_{i=1}^n |g_j(w_i,\theta)| < \infty$，$P$-*a.s.* 成立。那么，如果对于给定的 θ，存在一个关于 P_s 的可行解 $\{\hat{q}_i(\theta)\}_{i=1}^n$，即 $\{\hat{q}_i(\theta)\}_{i=1}^n \in Q_n(\theta)$ 且 $\{\hat{q}_i(\theta)\}_{i=1}^n = \arg\min_{q_1,\ldots,q_n\in Q_n(\theta)} \sum_{i=1}^n [q_i \log(nq_i)]$，，则 (P_s) 和 (D_s) 的值对于该特定的 θ 相等。

该引理表明，在假设 3 成立且函数 ϕ_+ 严格凸的情况下，

这两者共同保证存在可行解的概率趋近于 1，对每个$\theta \in \bigcap_{\{M_n;\ M_n\to\infty\}} B(\theta_*,\ M_n n^{-1/2})$ 都成立。此时，对于每个 $\theta \in \bigcap_{\{M_n;\ M_n\to\infty\}} B(\theta_*,\ M_n n^{-1/2})$，$(\mathcal{P}_s)$ 和 $(\mathcal{D}_s)$ 的取值相等的概率趋近于 1。

证明。 一旦我们证明该推论的条件得到满足，该定理的结果便可由 Borwein 和 Lewis（1991年，推论 2.6）得出。第一个条件要求积分域是一个有限测度空间。在我们的情形中，最小化问题中的积分是关于经验测度 $\frac{1}{n}\sum_{i=1}^{n}\delta_{w_i}$ 的，其中 δ_{w_i} 表示 w_i 上的狄拉克测度。于是，我们可以将积分域取为 $[\min_i w_i, \max_i w_i]$，这是一个有限测度空间。

Borwein 和 Lewis（1991年，推论 2.6）的第二个条件要求 ϕ_+ 是闭且凸的。凸性得到满足。关于闭性，只需注意到集合 $\{u \in \mathbb{R};\, \phi_+(u) \leq \alpha\}$ 对每个 $\alpha \in \mathbb{R}$ 都是闭集，因为当 $u \to 0^+$ 时，0 是 $\phi_+(u)$的极限点。最后，我们必须检验资格约束；由于 $(0,\infty) \subset dom(\phi_+)$，它等价于要求存在一个可行解 $\{\widehat{q}_i\}_{i=1}^n$，使得对每个 $i = 1,\ldots,n$ 都有 $\widehat{q}_i > 0$。后者得到满足，因为它是定义可行集 $\mathcal{Q}_n(\theta)$ 的条件之一。

□

引理 G.1 中对每个 $\theta \in \Theta$ 都成立的条件 $\sum_{i=1}^{n} |g(w_i,\theta)| < \infty$, P-几乎处处，变为对基础模型和扩展模型中的每个 $\theta \in \Theta$ 都成立的条件 $\sum_{i=1}^{n} |(y_i - \theta'\widetilde{w}_{1,i})\,\widetilde{w}_i| < \infty$, P-几乎处处。这是因为，正如 Borwein 和 Lewis（1991年）所示，辅助参数 v 不会出现在该条件中。

下一个引理根据对偶问题 $\mathcal{D}_s$ 的解刻画原问题 $\mathcal{P}_s$ 的解。

引理 G.2 假设函数 $w \mapsto g_j(w,\theta)$ 对所有 $j = 1$，……，d 以及每一个 $\theta \in \Theta$ 都是连续可微的，并假设引理 *G.1* 的条件得到满足。对每一个使 $Q_n(\theta)$ 非空的 θ，令 $(\widehat{\gamma}(\theta)\ \lambda(\theta))$ 为对偶最优解。则唯一的

原问题的最优解由下式给出： $\forall\theta$ ，使得 $Q_n(\theta) \neq \varnothing$,

$$\widehat{q}_i(\theta) = \frac{1}{n}((\phi_+)^*)'(\widehat{\gamma} + \widehat{\lambda}' g(w_i,\theta)), \tag{G.5}$$

$$= \frac{e^{\widehat{\lambda}' g(w_i,\theta)}}{n\mathbf{E}_n[e^{\widehat{\lambda}' g(w_j,\theta)}]} \qquad \textit{for } i = 1,\ldots,n. \tag{G.6}$$

证明。 在引理的假设下，根据 Borwein 和 Lewis（1991,

定理 5.5），结果可由 $D := \lim_{u\to\infty}(\phi(u)/u) = \infty$ 得出。

□

对于基础模型和扩展模型，连续可微性均成立。引理 G.2 的结果十分重要，这意味着，如果对偶问题存在一个可行解，那么对于每个 θ ，只要$\mathcal{Q}_n(\theta)$ 非空，我们就可以据此找到原问题的一个解。

即使函数 $\phi_+(\cdot)$ 是严格凸的，$\mathcal{P}_s$ 的解也不一定存在。事实上，对于某些 θ，可行集 $\mathcal{Q}_n(\theta)$ 可能为空。在假设1下，为了证明任何基于 ETEL 的估计程序合理，还需要以下假设。

假设 10 对于任意常数 $C > 0$ 以及任意 $\theta \in B(\theta_*, Cn^{-1/2})$:

$$P\left(\sum_{i=1}^{n} g(w_i,\theta)q_i = 0 \textit{ for at least one } q \in int\Delta_n\right) \to 1. \tag{G.7}$$

这一假设略弱于我们所要求的假设 3，因为该条件必须对稍小一些的集合 $B(\theta_*, Cn^{-1/2})$ 中的一个 θ 成立。