arXiv:2303.05888v1 [econ.TH] 10 Mar 2023

分布鲁棒随机效用模型

Emerson Melo[§] David M¨uller[*] Ruben Schlotter[‡]2023年3月13日

摘要

本文提出了分布鲁棒随机效用模型（DRO-RUM），该模型允许偏好冲击（未观测异质性）的分布存在设定错误或未知的情况。我们运用鲁棒优化文献中的工具，作出了三项贡献。首先，通过利用分布鲁棒社会剩余函数这一概念，我们证明，DRO-RUM 会内生地产生一种冲击分布，其中不同备选项的效用之间存在相关性。其次，我们证明，分布鲁棒社会剩余函数的梯度给出了选择概率向量。这一结果将著名的 William-Daly-Zachary 定理推广至冲击分布未知的情形。第三，我们说明了 DRO-RUM 如何利用选择市场数据，以非参数方式识别相应的平均效用向量。这一结果将需求反演方法扩展至冲击分布未知或设定错误的情形。我们进行了多项数值实验，将 DRO-RUM 的表现与传统的多项 Logit 模型和 Probit 模型进行比较。

关键词：离散选择、随机效用、凸分析、分布鲁棒优化。

JEL classification: C35, C61, D90.

[§]印第安纳大学经济学系，美国印第安纳州布卢明顿，邮编 47408。电子邮箱：emelo@iu.edu。

[*]开姆尼茨工业大学数学系，德国开姆尼茨，邮编 09126。电子邮箱：david.mueller@mathematik.tu-chemnitz.de

[‡]汉诺威再保险，Karl-Wiechert-Allee 50，德国汉诺威，邮编 30625。电子邮箱：ruben.schlotter@hannover-re.com

1 引言

由 Marschak（1959年）、Block 和 Marschak（1959年）以及 Becker 等人（1963）提出的随机效用模型（RUM），已成为解决随机选择问题的标准方法。McFadden（1978a,c, 1981）的奠基性研究，使随机效用模型成为一种便于实证应用的方法，适用于应用微观计量经济学的诸多领域，包括劳动力市场、产业组织、健康经济学、交通运输和运营管理。尤其是，McFadden 提供了经济学基础和计量经济学框架，将可观测行为与随机选择行为联系起来。后一特点使随机效用模型适用于处理复杂的选择环境和开展福利分析（McFadden（2001）和 Train（2009年））。

在随机效用模型（RUM）中，决策者（DM）面对由多个备选项组成的离散选择集，其中每个选项都关联着一种随机效用。随后，决策者以一定概率选择某个特定选项，该概率等于该备选项在所有可用备选项中效用最高的事件发生的概率。大多数应用文献将与每个备选项相关联的随机效用建模为可观测且确定性的部分与随机偏好冲击之和。在这种加性设定下，对随机偏好冲击作出不同的分布假设会产生不同的随机选择规则。因此，所有努力都在于为偏好冲击的分布提供条件，使选择概率符合随机效用最大化假说（McFadden（1981））。

更重要的是，假设分析者已知冲击分布，我们便可以估计描述与每个备选项相关联的确定性效用的参数，开展反事实福利分析，并预测未来的选择行为。从建模角度来看，这一假设意味着分析者能够正确地指定描述决策者行为中未观测到的异质性的冲击分布。

本文提出一个随机效用模型框架，允许分析者（或决策者）并不知道真实的冲击分布。为此，我们提出一个分布鲁棒框架，放宽冲击分布预先已知这一假设。具体而言，我们提出一个随机效用模型框架，允许冲击分布存在模型设定错误。通过对真实分布的不确定性进行建模，我们沿用分布鲁棒优化文献的做法，考虑这样一种情形：分析者可以获得一个参考分布 F。该分布对应于生成偏好冲击实现值的真实统计规律的一种近似。我们将 F 称为名义分布。因此，我们通过一个不确定集对分布不确定性进行建模，该集合包含所有与 F接近 的概率分布。我们借助统计 散度的概念来衡量概率分布之间的距离。更确切地说，我们采用 ϕ-散度（Csiszar（1967年）；Liese和Vajda（1987年））。 ϕ-散度的例子包括Kullback-Leibler散度、Rényi距离和Cressie-Read距离等。因此，不确定

集包含名义 F 以及根据 ϕ-散度衡量、位于特定半径范围内的所有可行分布。

基于不确定集，我们引入稳健社会剩余函数，它对应于所有可行分布下可实现的最大社会剩余。与传统随机效用模型一样，稳健社会剩余是一个凸函数，包含了研究和理解分布鲁棒随机效用模型（DRO-RUM）所需的全部相关信息。

1.1 贡献

我们有三项贡献。首先，我们证明，对 DRO-RUM 的分析等价于研究一个严格凸的有限维随机优化规划的性质。这一刻画直接表明，与 DRO-RUM 相关的内生稳健分布会在偏好冲击之间引入相关性，即使名义 F 可能假设这些冲击相互独立。

其次，我们证明稳健社会剩余函数的梯度给出了选择概率向量。后一结果将著名的威廉姆斯–戴利–扎卡里定理（WDZ）推广到了真实冲击分布未知的环境，这一推广并非显而易见。此外，我们证明 DRO-RUM 保留了传统随机效用模型的凸结构。具体而言，我们建立了一个稳健芬歇尔对偶框架，将稳健社会剩余与其凸共轭联系起来。

作为第三项贡献，我们刻画了分布鲁棒随机效用模型的经验内容。具体而言，我们证明，对于一个观测选择概率向量，存在一个唯一的平均效用向量，能够通过分布鲁棒随机效用模型对观测数据进行合理解释。特别是，我们证明该平均效用向量对应于稳健社会剩余函数的凸共轭的梯度。我们的结果在经济学上的意义在于，分布鲁棒随机效用模型能够用一个唯一的平均效用向量来合理解释观测到的行为（即一个选择概率向量）；该向量对应于一个严格凸随机规划问题的唯一解。

最后，为总结我们的理论贡献，我们进行了若干数值模拟，探讨了该框架的性质。具体而言，我们比较了分布鲁棒随机效用模型与多项Logit模型和多项Probit模型的选择行为。我们主要关注所谓的稳健性参数所产生的影响，该参数决定可行集的大小，进而影响选择概率和剩余函数。

1.2 相关文献

本文与多个研究领域的文献相关。首先，本文与随机效用模型和凸分析相关。与本文最接近的文献包括 Chiong 等（2016）、Galichon 和 Salani´e（2021）以及 Fosgerau 等（2021）的研究。与本文类似，这些研究利用随机效用模型的凸结构来

在可获得市场汇总数据（即观测到的选择概率）时，研究平均效用向量的非参数识别。本文及其结果与这些研究有显著不同，因为本文允许采用关于分布假设的更灵活框架。

其次，本文与半参数选择模型（SCM）文献相关。Natarajan 等人（2009年）的研究提出了 SCM，该模型适用于真实联合分布未知、但分析者可以获得与每个备选项相关的边缘分布的情形。SCM 的这一具体实例称为边际分布模型（MDM）。Mishra 等人（2014）研究了 MDM 方法的理论和实证表现。Mishra 等人（2012）研究了 SCM 的第二种实例，该实例利用交叉矩约束。具体而言，他们假设真实分布未知，但分析者可以获得真实的方差-协方差矩阵，该矩阵刻画了离散备选项集合中的相关结构。

乍看之下，我们的方法与 SCM 相似。正如 Feng 等人（2017年）所述，后者通常定义为对一组分布取上确界。我们通过引入 DRO-RUM 来适应这一设定：真实分布未知，因此分析者会考虑不确定集中的所有分布。

尽管一般的半参数选择模型（SCM）与我们的方法相似，但这两种框架存在重要差异。首先，我们的方法无需对边际分布或方差-协方差矩阵作出任何假设。相反，我们的模型只要求已知一个名义分布。其次，利用 ϕ-散度的概念，研究者能够纳入稳健性，并通过选择稳健性参数来控制冲击分布的不确定性。因此，可行集并非通过固定某些矩或边缘分布来明确确定，而是通过选择名义分布和稳健性参数的大小来隐式构造。此外，我们的方法允许选择多种 ϕ-散度函数和不同的名义分布，从而能够生成不同的模型。第三，我们证明，分布鲁棒随机效用模型（DRO-RUM）保留了传统随机效用模型（RUM）方法的凸结构（及其对偶性）。具体而言，我们推广了 WDZ 定理，并对稳健 Fenchel 对偶进行了分析。第四，我们利用稳健凸对偶结果，以非参数方式识别平均效用向量。这一结果使我们能够用分布鲁棒随机效用模型来解释市场汇总数据（选择概率）。具体而言，我们的识别结果对应于一种稳健需求反演方法。我们的论文也与稳健性相关文献有关——

宏观经济学中的稳健性（Hansen 和 Sargent（2001, 2008年））。然而，这一文献聚焦于使用 Kullback-Leibler 距离的递归问题。Christensen 和 Connault（2023年）近期发表的论文引入了稳健性思想，以分析反事实对结构模型中潜变量分布的参数假设有多敏感。他们关注的问题与本文研究的问题不同。最后，本文与以下领域的研究密切相关：

分布鲁棒优化。Shapiro（2017a）和 Kuhn 等人（2019）对这一主题进行了最新综述。相关应用涵盖从库存管理到机器学习中的正则化。[1]然而，据我们所知，这一领域的文献尚未研究随机效用模型（RUM）的分布鲁棒性问题。

本文其余部分安排如下。第2节回顾传统的随机效用模型方法，并介绍稳健性问题。第3节介绍分布鲁棒随机效用模型，并讨论其主要性质。第4节讨论 DRO-RUM 方法的经验内容。第5节包含若干数值实验，将DRO-RUM的结果与多项Logit模型和多项Probit模型进行比较。第6节通过概述可能的扩展来总结全文。

符号。本文通篇使用以下符号和定义。令 $\mathbb{R}=\mathbb{R}\cup\{-\infty,+\infty\}$ 表示，并考虑扩展实值函数

$$f:\mathbb{V}\to\bar{\mathbb{R}},$$

其中， $\mathbb{V}$ 是一个有限维实向量空间。因此，我们用$\mathbb{V}^*$ 表示由所有线性泛函构成的对偶空间。特别地，我们经常研究 $\mathbb{R}^n$的子空间。由下式定义的集合

$$\mathrm{dom} f=\{x\in\mathbb{V}:f(x)<+\infty\}$$

称为 f的（有效）定义域 。如果一个函数处处不取值 $-\infty$ ，且 $\mathrm{dom} f\neq\emptyset$，则称其为正常函数 。对于正常函数 $f:\mathbb{V}\to\mathbb{R}$ ，集合$\partial f(x)$表示其在 $x\in\mathrm{dom} f$处的次微分 ，即

$$\partial f(x)=\{g\in\mathbb{V}^*:f(y)\geq f(x)+\langle g,y-x\rangle,\quad \text{for all}\quad y\in\mathbb{R}^n\},$$

其中 $g\in\mathbb{V}^*$ 称为次梯度。如果次微分集合是单元素集，即次梯度 g唯一，则用 $\nabla f(x)$表示函数 f 在 $x\in\mathrm{int}\,(\mathrm{dom} f)$处的梯度 。正常函数 $f:\mathbb{V}\to\mathbb{R}$ 的凸共轭 为

$$f^*(g)=\sup_{x\in\mathbb{V}}\{\langle x,g\rangle-f(x)\},\quad g\in\mathbb{V}^*.$$

$\mathbb{E}_F(\cdot)$表示相对于分布 F的期望算子。

2 随机效用模型

考虑一位决策者（DM），在备选项 $j\in\mathcal{J}=\{0,1,\ldots,J\}$中进行效用最大化的离散选择。选项 j的效用为

$$\tilde{u}_j=u_j+\varepsilon_j,\tag{1}$$

其中 $u=(u_0,u_1,\ldots,u_J)^T$是确定性的，而 $\varepsilon=(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_J)^T$是随机效用冲击向量。备选项 0 表示外部选项。按照离散选择文献的惯例，我们设定 $u_0=0$。

[1]在经济学中，最早研究鲁棒优化问题的论文之一是 Scarf（1958年）

遵循 McFadden (1978a, 1981) 的做法，上述描述对应于经典的加性随机效用模型（RUM）。本文对随机效用模型框架的介绍将着重强调其凸分析性质。

假设 1 随机向量 ε 服从分布 F，该分布绝对连续、具有有限均值，且与 u独立，并在 $\mathbb{R}^{J+1}$上具有完全支撑。

假设 1 未具体规定 ε 的分布，因此允许存在远远超出常用 Logit模型范围的多种选择概率系统。该假设允许 ε_j之间存在任意相关性，这在实际应用中可能很重要。作为假设 1 的直接推论，决策者的选择概率对应于：

$$p_j(u) \equiv \mathbb{P}\left(u_j + \varepsilon_j = \max_{j' \in \mathcal{J}} \{u_{j'} + \varepsilon_{j'}\}\right), \quad j = 0, 1, \ldots, J.$$

随机效用模型框架中的一个重要对象是离散选择模型的剩余函数 （由 McFadden（1981）命名）。其定义为

$$W(u) = \mathbb{E}_F\left[\max_{j \in \mathcal{J}} \{u_j + \varepsilon_j\}\right]. \tag{2}$$

在假设 1 下， W 是凸且可微的，并且选择概率向量 $p(u)$ 与 W[2]的梯度重合：

$$\frac{\partial}{\partial u_k} W(u) = p_k(u) \text{ for } k = 0, 1, \ldots, J$$

或者，使用向量记法， $p(u) = \nabla W(u)$。上述结果即著名的 Williams–Daly–Zachary 定理（以下简称 WDZ），在离散选择文献中广为人知（McFadden（1978a，1981））。

应用最广泛的随机效用模型之一是多项Logit模型，它假设 $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_J)^T$中的各项均服从尺度参数为 η 的独立同分布冈贝尔分布。根据这一假设，我们可以将社会剩余函数写成闭式形式：

$$W(u) = \eta \log\left(\sum_{j=0}^{J} e^{u_j/\eta}\right) + \eta\gamma, \tag{3}$$

其中， γ 是欧拉-马歇罗尼常数。由 (3) 可知，WDZ 定理意味着 $p_j(u)$ 可表示为：

$$\frac{\partial W(u)}{\partial u_j} = \frac{e^{u_j/\eta}}{\sum_{l=0}^{J} e^{u_l/\eta}} \quad \text{for } j \in \mathcal{J}. \tag{4}$$

MNL 模型属于一类更广泛的随机效用模型，称为广义极值（GEV）模型，由 McFadden（1978b）提出。这类模型通过生成函数 G 定义： $\mathbb{R}_+^{J+1} \to \mathbb{R}_+$，该函数必须满足以下性质：

[2] W 的凸性源于最大值函数的凸性。可微性则源于 ε的绝对连续性。

（G1） G 是 $\frac{1}{\eta} > 0$次齐次的。 –

(G2) $G(x_0, x_1, \ldots, x_j, \ldots, x_J) \to \infty$ as $x_j \to \infty$, $j = 0, 1, \ldots, J$.

(G3) 对于关于 k 个不同变量的 G 偏导数，有:

$$\frac{\partial^{k+1} G(x_0, \ldots, x_J)}{\partial x_{j_0}, \partial x_{j_1} \cdots \partial x_{j_k}} \geq 0 \text{ if } k+1 \text{ is odd}, \quad \frac{\partial^{k+1} G(x_0, \ldots, x_J)}{\partial x_{j_0}, \partial x_{j_1} \cdots \partial x_{j_k}} \leq 0 \text{ if } k+1 \text{ is even}.$$

McFadden（1978b，1981）表明，满足条件（G1）—（G3）的函数 G 意味着随机向量 ε 的联合分布对应于以下概率密度函数:

$$f_\epsilon(y_0, y_1, \ldots, y_J) = \frac{\partial^{J+1} \exp\left(-G\left(e^{-y_0}, \ldots, e^{-y_J}\right)\right)}{\partial y_0 \cdots \partial y_J},$$

GEV 类的一个重要性质是，其社会剩余函数对应于（McFadden，1978b）

$$W(u) = \eta \ln G(e^u) + \eta\gamma,$$

其中 γ 是欧拉-马歇罗尼常数。根据 WDZ 定理可知，第 j 个备选项的选择概率为:

$$p_j(u) = \frac{\partial W(u)}{\partial u_j} = \eta \frac{\partial G(e^u)}{\partial e^{u_j}} \cdot \frac{e^{u_j}}{G(e^u)} \quad \forall j \in \mathcal{J}.$$

不难看出，生成函数

$$G(e^u) = \sum_{j=0}^{J} e^{u_j/\eta} = 1 + \sum_{j=1}^{J} e^{u_j/\eta}$$

导出 MNL 模型。

GEV 类的主要优势在于能够灵活地捕捉随机变量 ε_j 之间复杂的相关性模式。例如，嵌套 Logit (NL)、配对组合 Logit (PCL)、有序 GEV (OGEV) 和广义嵌套 Logit (GNL) 模型，都是 GEV 家族的特例。

2.1 随机效用模型的稳健框架

随机效用模型的一个基本假设是，研究者（以及决策者）知道冲击分布。这意味着，ε 的分布设定是正确的。本文的主要目标是放宽这一条件，允许 ε 的分布未知。相反，ε 的分布是一个优化问题中的决策变量，该问题对应于社会剩余函数的定义。我们通过将表达式 (2) 替换为稳健社会剩余 函数来形式化这一想法:

$$W^{RO}(u) = \sup_{G \in \mathcal{M}(F)} \mathbb{E}_G \left[\max_{j \in \mathcal{J}} \{u_j + \varepsilon_j\} \right], \tag{5}$$

其中， $\mathcal{M}(F)$ 是一组与预先设定的分布 F 接近的概率分布，而该分布满足假设 1。分布 F 可视为分析者对误差项联合分布的最佳估计或先验知识。我们将 F 称为名义分布。为了更好地应对模型设定错误，分析者会考虑所有与名义分布接近的可能分布。我们的方法的一个关键方面与集合 $\mathcal{M}(F)$ 的结构有关。在第3节中，我们用 ϕ-散度函数来刻画 $\mathcal{M}(F)$，从而能够运用概率分布之间统计散度的概念（Csiszar（1967年）、Liese 和 Vajda（1987年）以及 Pardo（2005年））。因此，我们将其称为分布鲁棒-随机效用模型（ DRO-RUM）。正如我们将看到的，这样做使我们能够通过一个凸有限维优化规划来刻画由此得到的 DRO-RUM 剩余函数。这一刻画对于研究 DRO-RUM 方法的性质至关重要。

令 $G^\star$ 表示在 (5) 中达到最优值的分布（或一列分布的极限）。在此模型下，备选项 j 的选择概率为（前提是其定义良好）：

$$p_j^{RO}(u) = \mathbb{P}_{G^\star}\left(j = \arg\max_{j' \in \mathcal{J}} \{u_{j'} + \varepsilon_{j'}\}\right) \tag{6}$$

从经济学角度看，我们可以用两种不同的方式来解释规划 (5)。首先，稳健-随机效用模型考虑的是这样一种情形：决策者面临偏好冲击，但对于产生这些误差的分布有一定的灵活性。

其次，分析者可能不确定随机向量 ε 的分布，而是考虑一组可能的分布。因此，分析者假设决策者是理性的，并且产生社会剩余的冲击分布对应于 $\mathcal{M}$ 中的一个元素，这是合理的。

2.2 与半参数选择模型的联系

值得指出的是，RO-RUM 的定义与半参数选择模型 （SCM）相似。SCM 最近由运筹学文献引入（Natarajan 等人（2009））。在这两类模型中，剩余函数均定义为对各分布取上确界。通过这种方式，SCM 能够捕捉选择集 $\mathcal{J}$中不同备选项之间复杂的替代模式和相关性。Feng 等人（2017年）详细概述了若干离散选择模型，并将 SCM 描述为对一个一般分布集合取上确界。因此，鲁棒随机效用模型可以视为半参数选择模型的一个实例。文献中已有一些 SCM 的实例。在原始论文中，Natarajan 等人（2009年）将可行集限定为边缘分布信息已知的联合分布。这个特定的 SCM 实例称为

边际分布模型 （MDM）。[3] 第二类 SCM 利用交叉矩约束。具体而言，Mishra 等人（2012）研究了交叉矩模型 （CMM），其中将集合 $\mathcal{M}$ 视为与已知 方差-协方差矩阵相一致的分布集合。[4]

尽管表面上十分相似，我们的方法在若干方面不同于现有的集合约束模型。正如本文后文将要看到的，我们的框架与集合约束模型在分布集合的设定上有所不同。具体而言，在现有的集合约束模型中，分析者需要显式构造可行集，例如固定边缘分布。相比之下，在我们的稳健方法中，分析者通过确定标称分布 F，并对其他分布与 F 之间的距离设定上界，来隐式指定可行集。因此，DRO-RUM 方法不需要了解任何一个方差-协方差矩阵的边际分布。在第3节中，我们将看到，在分布鲁棒随机效用模型中，研究者通过选择稳健性参数的大小来控制距离。因此，我们的方法遵循的原则与现有的集合约束模型截然不同。

此外，我们证明，分布鲁棒随机效用模型对应于一个凸有限维优化问题的解。这一结果使我们能够将 WDZ 定理推广到冲击分布设定有误的情形。最后，第4节展示了分布鲁棒随机效用模型如何帮助我们恢复平均效用向量 u。

3 分布鲁棒随机效用模型

本节将正式介绍 DRO-RUM 方法。按照分布鲁棒优化领域的文献，我们考虑这样一种情形：研究者（或决策者）可以获得参考分布 F，它可能是支配 ε 实现值的真实统计规律的近似（或估计）。我们将 F 称为名义分布。接下来，我们定义一个概率分布集合，其中的分布与 F足够接近。我们借助统计距离来形式化概率分布之间的距离概念。

3.1 ϕ-散度

我们用所谓的ϕ-散度来衡量两个概率分布之间的距离。

设 $\phi:\mathbb{R}\to(-\infty,+\infty]$ 是一个适当闭凸函数，且 dom ϕ 是端点为 $\alpha<\beta$的区间，因此，int (dom ϕ) $=(\alpha,\beta)$。由于 ϕ 是闭的，若 α 有限，则有 $\lim_{t\to\alpha_+}\phi(t)=\phi(\alpha)$；若 β 有限，则有 $\lim_{t\to\beta_-}\phi(t)=\phi(\beta)$。

[3]在 MDM 中，随机向量 ε 的边缘分布是固定的。形式上，我们写作$\varepsilon_i\sim F_i$，其中 F_i 是第 i 个误差的边缘分布函数，$i=1$、.....、J。在这种情况下，我们定义 $\mathcal{M}\triangleq \text{MAR}=\{F:\ \varepsilon_i\sim F_i\quad \forall i\in\mathcal{J}\}$。[4]形式上，CMM 考虑分布集合 $\mathcal{M}(0,\ \Sigma)\ =\ \{G:\ \ \mathbb{E}_F(\varepsilon)\ =0,\ \ \ \mathbb{E}_G(\varepsilon\varepsilon^\top)\ =\ \Sigma\}$。在 $\mathcal{M}(0,\ \Sigma)$ 的定义中，假定方差-协方差矩阵 Σ 已知。

本文始终假设 ϕ 非负，且在点 $1 \in$ int (dom ϕ)处取得其最小值，即 ϕ (1) $= 0$。这类函数记为 Φ。

定义 1 给定 $\phi \in \Phi$，概率测度 G 相对于 F 的 ϕ-散度为

$$D_\phi(G\|F) = \begin{cases} \int_{\mathbb{R}^{J+1}} \phi\left(\frac{g(\varepsilon)}{f(\varepsilon)}\right) f(\varepsilon) d\varepsilon & if \ \mathrm{G} \ll \mathrm{F} \\ +\infty & otherwise \end{cases} \tag{7}$$

其中，f 和 g 分别是 F 和 G 对应的密度。

为避免病态情形，本文始终作如下假设:

$$\phi(0) < \infty, 0 \cdot \phi\left(\frac{0}{0}\right) \equiv 0, 0 \cdot \phi\left(\frac{s}{0}\right) = \lim_{\varepsilon \to 0} \varepsilon \cdot \phi\left(\frac{s}{\varepsilon}\right) = s \lim_{t \to \infty} \frac{\phi(t)}{t}, \quad s > 0. \tag{8}$$

若测度 G 关于 F 绝对连续，即 $G \ll F$，则 ϕ-散度可方便地写为:

$$D_\phi(G\|F) = \mathbb{E}_F\left(\phi(L(\varepsilon))\right), \tag{9}$$

其中，$L(\varepsilon) \triangleq g(\varepsilon)/f(\varepsilon)$ 是密度 g 和 f 之间的似然比，也称为这两个测度的拉东-尼科迪姆导数。利用表达式 (9) 并结合 ϕ 的凸性，根据詹森不等式可得

$$D_\phi(G\|F) \geq \phi\left(\mathbb{E}_F\left(L(\varepsilon)\right)\right) = \phi(1) = 0 \tag{10}$$

当 $G = F$ 时等号成立，因此 $D_\phi(G\|F)$ 是衡量 G 与 F 之间距离的量。[5] 此外，ϕ-散度泛函关于其两个自变量均为凸函数。下面的命题总结了这些关键性质。

命题 1 ϕ-散度泛函 *(7)* 定义良好且非负。当且仅当 $f_1(t) = f_2(t)$ 几乎处处成立时，其值为零。此外，D_ϕ 对其每个自变量都是凸的。

证明。 $D_\phi(G\|F)$ 定义良好且非负，这由 (Ben-Tal 和 Teboulle，1987年，命题 1) 可得。 D_ϕ 的凸性由 (Ben-Tal 和 Teboulle，1987年，命题 2) 可得。□

在我们的分析中，ϕ的凸共轭是一个关键要素。对于 $\phi \in \Phi$，其凸共轭记为 ϕ^*：

$$\phi^*(s) = \sup_{t \in \mathbb{R}}\{st - \phi(t)\} = \sup_{t \in \operatorname{dom}\phi}\{st - \phi(t)\} = \sup_{t \in \operatorname{int}\operatorname{dom}\phi}\{st - \phi(t)\}, \tag{11}$$

其中最后一个等式由 (Rockafellar，1970年，推论 12.2.2) 得出。凸共轭 ϕ^* 是一个闭真凸函数，其 int dom $\phi^* = (a, b)$，其中

$$a = \lim_{t \to -\infty} t^{-1}\phi(t) \in [-\infty, +\infty); b = \lim_{t \to +\infty} t^{-1}\phi(t) \in (-\infty, +\infty].$$

[5]我们回顾一下，$1 \in$ int dom ϕ 是 ϕ 取到其最小值 0 的点。

此外，由于 ϕ 是凸且闭的，我们有其双共轭 ϕ^{**} =ϕ（Rockafellar（1970年））。值得注意的是，利用 1 是 ϕ 的极小化点且位于其定义域的内部这一事实，可得 $\phi'(1) = 0$ 成立。此外，利用 ϕ 是凸且闭的这一性质，根据 Fenchel 等式，我们有 $y = \phi'(x)$ 当且仅当 $x = \phi^{*'}(y)$。将这一观察应用于 $x = 1$ 和$y = 0$ ，可得 $\phi^{*\prime}(0) = 1$。

3.2 分布鲁棒随机效用最大化框架

核心思想是考虑这样一种环境：分析者（或决策者）不知道决定冲击向量ε实现值的真实分布。在这种环境中， F 的作用是作为对“真实”未知分布的一种近似或最佳猜测。考虑到这种歧义或分布 F可能存在的模型设定错误，我们使用 ϕ-散度来定义不确定集 $\mathcal{M}_\phi(F)$:

$$\mathcal{M}_\phi(F) = \{G \ll F : D_\phi(G||F) \leq \rho\}, \tag{12}$$

形式上， $\mathcal{M}_\phi(F)$是所有满足以下条件的概率测度 G 构成的集合：相对于 F 绝对连续，且按照 ϕ-散度衡量，与 F的距离至多为 ρ。超参数 ρ 是 $\mathcal{M}_\phi(F)$的半径，它反映了研究者（或决策者）对 F 是否正确的可信程度有多不确定。下面进一步阐释这一解释。遵循Hansen 和 Sargent（2001、2008）、Shapiro（2017b）以及Kuhn等人（2019）的做法，我们将集合（12）解释为这样一种环境：分析者（或决策者）对真实未知概率分布有一个最佳猜测 F ，但并不完全信任它。例如，研究者可能认为名义分布 F对应冈贝尔分布。在这种情况下， $\mathcal{M}_\phi(F)$允许许多其他概率分布 G 纳入可行范围，其中 ρ 决定可行集的大小。

给定集合 $\mathcal{M}_\phi(F)$，我们可以修改表达式 (5)，得到分布鲁棒剩余函数。因此，DRO- RUM 的剩余函数对应以下优化问题：

$$W^{DRO}(u) = \sup_{G \in \mathcal{M}_\phi(\mathcal{F})} \left\{ \mathbb{E}_G \left[\max_{j \in \mathcal{J}} \{u_j + \varepsilon_j\} \right] \right\} \tag{13}$$

这里有几点需要说明。首先，规划 (13) 的一个基本方面是参数 ρ 的作用，它控制 $\mathcal{M}_\phi$（F）的大小。因此，我们可以将 ρ 解释为鲁棒性指数。更具体地说，当$\rho = 0$ 时，我们得到 $\mathcal{M}_\phi$（F） $= \{F\}$，这意味着我们恢复了分布为 F 时的随机效用模型（RUM）。[6]另一方面，当 $\rho \longrightarrow \infty$ 时，不确定集 $\mathcal{M}_\phi$（F）包含大得多的一组可能分布，其中包括那些可能不

[6]我们注意到 $\rho = 0$ 意味着 $D_\phi(G\|F) = 0$。根据命题 1，我们知道，当且仅当 $F = G$ 时，后一个等式成立。

满足假设 1。[7] 分布鲁棒随机效用模型（DRO-RUM）旨在设置 ρ，以反映研究者（或决策者）对 ε 的分布假设所感知到的不确定性。

下面的引理确立了 $W^{DRO}(u)$ 的一些基本性质。

引理 1 对于分布鲁棒随机效用模型，剩余函数 $W^{DRO}(u)$ 满足：

(i) $W^{DRO}(u + c \cdot e) = W^{DRO}(u) + c$ *for all* $c \in \mathbb{R}, u \in \mathbb{R}^J$.

(ii) $W^{DRO}(u) \geq W^{DRO}(v)$ 对所有满足 $u \geq v$ 的 u、$v \in \mathbb{R}^J$ 均成立。

(iii) $W^{DRO}(u) \geq \max_{j \in \mathcal{J}} u_i + \min_{j \in \mathcal{J}} \mathbb{E}_F[\varepsilon_i]$。

(iv) $W^{DRO}(u)$ 关于 u 是凸的。

证明。

(i) 根据定义可得

$$W^{DRO}(u + c \cdot e) = \sup_{G \in \mathcal{M}_\phi(F)} \left\{ \mathbb{E}_G \left[\max_{j \in \mathcal{J}} \{u_j + \varepsilon_j + c\} \right] \right\}$$

由于期望的线性性，有

$$c + \sup_{G \in \mathcal{M}_\phi(F)} \mathbb{E}_G \left[\max_{j \in \mathcal{J}} \{u_j + \varepsilon_j\} \right] = c + W^{DRO}(u).$$

(ii) 任取 $u, v \in \mathbb{R}^J$，其中 $u \geq v$。首先注意，对于任意可行分布 $G \in \mathcal{M}_\phi(F)$，有

$$W^{DRO}(u) \geq \mathbb{E}_G \left[\max_{j \in \mathcal{J}} \{u_j + \varepsilon_j\} \right] \overset{(*)}{\geq} \mathbb{E}_G \left[\max_{j \in \mathcal{J}} \{v_j + \varepsilon_j\} \right],$$

其中 $(*)$ 由期望的单调性成立。对右侧取上确界，可得 $W^{DRO}(u) \geq W^{DRO}(v)$。

(iii) 可推得，对于任意 $i \in \mathcal{J}$

$$W^{DRO}(u) \geq \mathbb{E}_F \left[\max_{j \in \mathcal{J}} \{u_j + \varepsilon_j\} \right] \geq \mathbb{E}_F [u_i + \varepsilon_i] \geq u_i + \min_{j \in \mathcal{J}} \mathbb{E}_F [\varepsilon_j],$$

根据假设 1，该值有限。

[7] 为说明这一点，我们注意到，当 $\rho \longrightarrow \infty$ 时，ϕ-散度无界。这一事实意味着，集合 $\mathcal{M}_\phi(F)$ 由所有相对于 F 绝对连续的分布构成。由于 F 具有完全支撑，这仅意味着 $\mathcal{M}_\phi(F)$ 中的分布必须是连续的，但当然不一定在 $\mathbb{R}^{J+1}$ 上具有完全支撑。事实上，$\mathcal{M}_\phi(F)$ 可能由相对于勒贝格测度绝对连续但没有有限均值的分布构成。例如，形状参数为 $\alpha = 1$ 的帕累托分布是绝对连续的，但其均值并非有限。

(iv) 设 $\alpha \in [0,1]$，并令 u 和 v 为两个确定性效用向量。对于固定分布 $G \in \mathcal{M}_\phi(F)$，由于 max$\{\cdot\}$ 算子的凸性，

$$W^{DRO}(\alpha u+(1-\alpha)v) \leq \alpha \mathbb{E}_G\left(\max_{j\in\mathcal{J}}\{u_j+\varepsilon_j\}\right)+(1-\alpha)\mathbb{E}_G\left(\max_{j\in\mathcal{J}}\{v_j+\varepsilon_j\}\right).$$

在右侧，对 G over$\mathcal{M}_\phi(F)$ 取上确界，可得

$$W^{DRO}(\alpha u+(1-\alpha)v) \leq \alpha W^{DRO}(u)+(1-\alpha)W^{DRO}(v).$$

由此可得 $W^{DRO}(u)$ 的凸性。

以下结果刻画了 $W^{DRO}(u)$。

命题 2 设假设 *1* 成立，并定义随机变量 $H(u,\varepsilon) \triangleq \max_{j\in\mathcal{J}}\{u_j+\varepsilon_j\}$。那么，问题 (13) 等价于求解以下有限维凸规划：

$$W^{DRO}(u)=\inf_{\lambda\geq 0,\mu\in\mathbb{R}}\left\{\lambda\rho+\mu+\lambda\mathbb{E}_F\left[\phi^*\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right]\right\}, \tag{14}$$

其中，λ 是与不确定集 $\mathcal{M}_\phi(F)$ 相关联的拉格朗日乘子，而μ 是与 G 为概率测度这一条件相关联的乘子。此外，规划 *(14)* 关于 μ 和 λ 是凸的。

证明。该结果可由直接应用（Ruszczy´nski 和 Shapiro，2021年，命题 7.9）得到。为完整起见，我们给出论证的细节。首先注意，对于固定的效用向量 u，并利用似然比$L(\varepsilon) \triangleq dG(\varepsilon)/dF(\varepsilon)$，式 (13) 中的分布鲁棒随机效用模型可表示为：

$$\begin{aligned} W^{DRO}(u) &= \sup_G\{\mathbb{E}_G(H(u,\varepsilon)) : G\in\mathcal{M}_\phi(F)\} \\ &= \sup_{L\geq 0}\{\mathbb{E}_F[L(\varepsilon)H(u;\varepsilon)] \mid \mathbb{E}_F[\phi(L(\varepsilon))]\leq\rho, \mathbb{E}_F[L(\varepsilon)]=1\} \end{aligned} \tag{15}$$

其中上确界取自一组可测函数。

问题 (15) 的拉格朗日函数为：

$$\mathcal{L}(L,\lambda,\mu)=\int_{\mathbb{R}^{J+1}}[L(\varepsilon)H(u,\varepsilon)-\lambda\phi(L(\varepsilon))-\mu L(\varepsilon)]dF(\varepsilon)+\lambda\rho+\mu. \tag{16}$$

问题 (16) 的拉格朗日对偶问题为

$$\inf_{\lambda\geq 0,\mu\in\mathbb{R}}\sup_{L\geq 0}\mathcal{L}(L,\lambda,\mu) \tag{17}$$

由于问题 (16) 满足 Slater 条件[8]，，因此问题 (16) 与其对偶问题 (17) 之间不存在对偶间隙。此外，对偶问题的最优解集合非空且有界。

[8]例如，我们可以取 $L(\varepsilon)=1$ 对所有 $\varepsilon\in\mathbb{R}^{J+1}$。

根据可交换性原理（(Rockafellar, 1976, 定理 3A)），(17) 中的最大值可以移入积分号内，即

$$\begin{aligned}&\sup_{L\geq 0}\int_{\mathbb{R}^{J+1}}[L(\varepsilon)H(u,\varepsilon)-\mu L(\varepsilon)-\lambda\phi(L(\varepsilon))]dF(\varepsilon)\\&=\int_{\mathbb{R}^{J+1}}\sup_{t\geq 0}\{t(H(u,\varepsilon)-\mu)-\lambda\phi(t)\}dF(\varepsilon),\end{aligned}$$

注意到 $(\lambda\phi)^*(H(u,\varepsilon)-\mu)=\sup_{t\geq 0}\{t(H(u,\varepsilon)-\mu)-\lambda\phi(t)\}$，于是可得

$$W^{DRO}(u)=\inf_{\lambda\geq 0,\mu\in\mathbb{R}}\left\{\lambda\rho+\mu+\mathbb{E}_F\left[(\lambda\phi)^*(H(u,\varepsilon)-\mu)\right]\right\}. \tag{18}$$

为证明关于 λ 和 μ 的凸性，我们注意到，在 (17) 和 (18) 中，只需对 $\lambda>0$ 而非 $\lambda\geq 0$取 inf，并且对于 $\lambda>0$，$(\lambda\phi)^*(y)=\lambda\phi^*(y/\lambda)$。因此，$W^{DRO}(u)$由以下问题的最优值给出：

$$\inf_{\lambda>0,\mu\in\mathbb{R}}\left\{\lambda\rho+\mu+\lambda\mathbb{E}_F\left[\phi^*((H(u,\varepsilon)-\mu)/\lambda)\right]\right\} \tag{19}$$

注意到 $\phi^*(\cdot)$ 是凸的。因此，$\lambda\phi^*(y/\lambda)$ 关于 y 和 $\lambda>0$是联合凸的。由此可知，问题 (19) 的目标函数是关于 $\lambda>0$和 $\mu\in\mathbb{R}$ 的凸函数，其中 $y=H(u,\varepsilon)-\mu$。因此，(19) 是一个凸问题。□

命题 2 的一个重要推论是，我们可以将函数 $W^{DRO}(u)$刻画为一个有限维凸优化问题的解。求解规划 (14) 的效率在很大程度上取决于相对于名义分布 F 的期望以及凸共轭 ϕ^*的性质。

下一个推论形式化了在 $\rho=0$时，$W(u)$与 $W^{DRO}(u)$之间的联系。

推论 1 设假设 *1* 成立。则对于 $\rho=0$，我们有 $W^{DRO}(u)=W(u)$。

证明。 考察问题 (15)。若 $\rho=0$，由其中一个约束可得

$$\mathbb{E}_F[\phi(L(\varepsilon))]\leq 0.$$

根据 ϕ 的定义，这意味着 $L(\varepsilon)=1$。因此，拉格朗日函数得以简化，因为对密度取上确界变得显而易见。将 $L(\varepsilon)=1$ 代入方程（16）：

$$\mathcal{L}(L,\lambda,\mu)=\int_{\mathbb{R}^{J+1}}H(u,\varepsilon)-\lambda\cdot\underbrace{\phi(1)}_{=0}-\mu\cdot 1]dF(\varepsilon)+\lambda\cdot 0+\mu.$$

后者等价于

$$\mathbb{E}_F\left[H(u,\varepsilon)-\mu\right]+\mu=\mathbb{E}_F\left[H(u,\varepsilon)\right],$$

其中，最后一个等式由期望的线性性得出。对于任何满足假设 1 的分布，我们确实可以恢复 $W(u)$。□

3.3 一个稳健的 WDZ 定理

随机效用模型的一个基本特点是，在对 ε 作出特定分布假设时，可以刻画选择概率。形式上，作为假设 1 的推论，WDZ 定理指出， $W(u)$ 的梯度给出选择概率向量 $p(u)$。本节将证明，在分布鲁棒随机效用模型中也有类似结果。具体而言，我们将证明 $\nabla W^{DRO}(u) = p^\star(u)$，其中 $p^\star(u)$ 对应于通过式 (14) 方法的最优解生成的选择概率向量。为确立这一结果，我们需要以下假设。

假设 2 $\phi^*(s)$ 严格凸且可微，满足 $\phi^{*\prime}(s) \geq 0$ 对所有 s均成立。

我们指出，许多 ϕ-散度函数都满足假设2。表1概述了三种满足该假设的常用 ϕ-散度。

散度	$\phi(t)$	$\phi^*(s)$	定义域	$\phi^{*'}$	$\phi^{*''}$
*Kullback-Leibler*散度	$t\log\ t$	e^{s-1}	$\mathbb{R}$	e^{s-1}	e^{s-1}
反向*Kullback-Leibler*散度	$-\log(t)$	$-1-\log(-s)$	$\mathbb{R}_{--}$	$-\frac{1}{s}$	$\frac{1}{s^2}$
海林格距离	$(\sqrt{t}-1)^2$	$\frac{s}{1-s}$	$s<1$	$\frac{1}{(1-s)^2}$	$-\frac{2}{(s-1)^3}$

表1： ϕ-散度及其凸共轭、一阶和二阶导数。

作为假设2的直接推论，我们可以证明（14）的最优解严格凸且唯一。

引理2 设假设 *1* 和 *2* 成立。则规划(14) 是严格凸的，并且具有唯一的最优解 $\lambda^\star$ 和 $\mu^\star$。

证明。根据假设2，函数 ϕ^* 是严格凸的。按照 Dacorogna 和 Mar´echal（2008年）的类似步骤，可以推出 $\lambda\cdot\phi^*\left(\frac{s}{\lambda}\right)$， $\lambda>0$是严格凸的。此外，凸函数与严格凸函数之和是严格凸函数。这一事实立即意味着 λ 和 μ中的目标函数是严格凸的。鉴于 λ 和 μ中的严格凸性，可以推出规划（14）有唯一解。□

假设2的第二个重要推论是，可以刻画规划（14）最优解所对应的鲁棒密度。

引理 3 假设假设 *1* 和 *2* 成立。对于固定的 $u\in\mathcal{U}$，令 $\lambda^\star>0$ *and*$\mu^\star\in\mathbb{R}$ 为问题 *(14)* 的唯一最优解。则唯一的鲁棒密度 $g^\star(\varepsilon)$ 对应于：

$$g^\star(\varepsilon)=\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)f(\varepsilon)\quad\forall\varepsilon\in\mathbb{R}^{J+1}. \tag{20}$$

证明。定义 $\Psi(\lambda,\mu) := \lambda\rho + \mu + \lambda\mathbb{E}_F\left(\phi^*\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right)$。针对 λ and μ 优化 $\Psi(\lambda,\mu)$，结合假设2的一阶条件表明，最优解 $\lambda^\star$ and $\mu^\star$ 必须满足

$$\begin{aligned}\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)\right) &= 1\\ \int_{\mathbb{R}^{J+1}}\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)f(\varepsilon)d\varepsilon &= 1\end{aligned}$$

定义 $g^\star(\varepsilon) \triangleq \phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)f(\varepsilon)$。由此可得 $\int_{\mathbb{R}^{J+1}} g^\star(\varepsilon)d\varepsilon = 1$。此外，根据假设2，可得对于所有 $\varepsilon \in \mathbb{R}^{J+1}$，$g^\star(\varepsilon) \geq 0$。因此，我们得出结论，$g^\star(\varepsilon)$ 确实是一个概率密度，我们称之为问题 (14) 对应的鲁棒密度。□

这里有几点说明。首先，鲁棒密度 $g^\star$ 取决于通过其共轭 ϕ^* 所体现的 ϕ-散度的选择。此外，鲁棒密度通过 $H(u,\varepsilon)$ 依赖于确定性效用向量，尽管根据假设 1，名义分布 F 并不依赖于 u。此外，$g^\star$ 使我们能够定义鲁棒分布函数 $G^\star$，正如我们将看到的，它在给出 $W^{DRO}(u)$ 的显式形式时发挥关键作用。其次，引理 3 证明，鲁棒密度 $g^\star(\varepsilon)$ 通过因子 $\phi^{*\prime}((H(u,\varepsilon) - \mu^\star)/\lambda^\star)$ 纳入了随机向量 ε 各元素之间的相关性。因此，即使名义分布 F 可能假设 $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_J$ 相互独立，DRO-RUM 方法仍会引入这些项之间的相关性。

例 1 *[KL-散度]现在我们考虑 Kullback-Leibler 散度的情形。为此，我们定义 ϕ 如下：*

$$\phi(t) \triangleq t\log t,\ t \geq 0 \tag{21}$$

注意，在前一个表达式中，$0\log 0 = 0$。这里

$$\int_{\mathbb{R}^J}\phi(L(\varepsilon))dF(\varepsilon) \tag{22}$$

定义了 *Kullback-Leibler* 散度，记作 $D_{KL}(G\|F)$。对于 $\lambda > 0$，$\lambda\phi$ 的共轭为 $(\lambda\phi)^*(y) = \lambda\left(e^{y/\lambda} - 1\right)$。由命题 *2*可知，

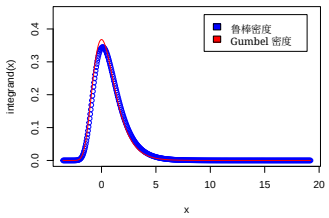
$$W^{DRO}(u) = \inf_{\lambda\geq 0,\mu}\left\{\lambda\rho + \mu + \lambda e^{-\mu/\lambda}\mathbb{E}_F\left[e^{H(u,\varepsilon)/\lambda}\right] - \lambda\right\} \tag{23}$$

在 *(23)* 中，对 μ 求最小值得到 $\mu^\star = \lambda\ln\mathbb{E}_F\left[e^{H(u,\varepsilon)/\lambda}\right]$。将$\mu^\star$ 代入 *(23)*，我们得到 $\lambda^\star$ 作为以下问题的解：

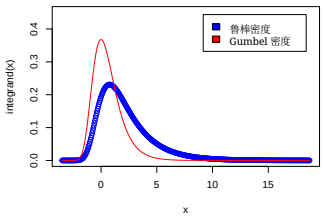
$$W^{DRO}(u) = \inf_{\lambda>0}\left\{\lambda\rho + \lambda\ln\mathbb{E}_F\left[e^{H(u,\varepsilon)/\lambda}\right]\right\}. \tag{24}$$

众所周知，在 *KL*散度的情形下（例如 *Hu* 和 *Hong*（*2012*）以及 *Hansen* 和 *Sargent*（*2001*）），「稳健」密度为：

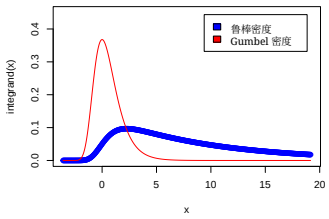
$$f^{DRO}(\varepsilon) = f(\varepsilon)\frac{e^{H(u,\varepsilon)/\lambda^\star}}{\mathbb{E}_F(e^{H(u,\varepsilon)/\lambda^\star})} \tag{25}$$



(a) $\rho = 0.01$



(b) $\rho = 0.5$



(c) $\rho = 3.5$

图 1： ρ 的三个实例

其中， $f(\varepsilon)$ 是与名义分布相关联的密度， $H(u, \varepsilon) = \max_{j \in \mathcal{J}}\{u_j + \varepsilon_j\}$ ，而 $\lambda^\star$ 是（*24*）的唯一最优解。图 *1* 展示了最优密度(25) 与名义分布为冈贝尔分布时的情形相比有何不同。

引理 3 的结果使我们能够以类似于著名的 WDZ 定理的方式刻画选择概率向量 $p^\star(u)$。

定理 1 设假设 *1* 和 *2* 成立。设 $\lambda^\star$ 和 $\mu^\star$ 为规划 *(14)* 的唯一最优解，它们分别导出作为最优密度和分布函数的 $g^\star$ 和 $G^\star$ 。则以下陈述成立：

(i) 稳健社会剩余对应如下：

$$W^{DRO}(u) = \mathbb{E}_{G^\star}\left(\max_{j \in \mathcal{J}}\{u_j + \varepsilon_j\}\right)$$

(ii) 选择概率向量 $p^\star(u)$ 为

$$\nabla W^{DRO}(u) = p^\star(u).$$

证明。(i) 为证明第一部分，我们定义函数 $\Psi(\cdot)$ 如下$\Psi(\lambda, \mu) \triangleq \lambda\rho + \mu + \lambda\mathbb{E}_F\left(\phi^*\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right)$。关于 $\Psi(\lambda, \mu)$ 进行优化

λ 和 μ 我们得到

$$\begin{aligned}\frac{\partial\Psi(\lambda,\mu)}{\partial\lambda} &= \rho+\mathbb{E}_F\left(\phi^*\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right) \\ &+ \lambda\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\left(\frac{H(u,\varepsilon)-\mu}{-\lambda^2}\right)\right)=0 \\ \frac{\partial\Psi(\lambda,\mu)}{\partial\mu} &= 1-\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right)=0\end{aligned} \tag{26}$$

重新整理第一个方程，得到:

$$\begin{aligned}&\lambda\rho+\lambda\mathbb{E}_F\left(\phi^*\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right)+\mu\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right) \\ =&\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)H(u,\varepsilon)\right).\end{aligned}$$

同理，对于第二个方程，我们有:

$$\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu}{\lambda}\right)\right)=1$$

结合这两个表达式可知，最优的 $\lambda^\star$ 和 $\mu^\star$ 必须满足:

$$\lambda^\star\rho+\lambda^\star\mathbb{E}_F\left(\phi^*\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)\right)+\mu^\star=\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)H(u,\varepsilon)\right).$$

利用引理 3 中的表达式 (20)，得到:

$$\Psi(\lambda^\star,\mu^\star)=\mathbb{E}_F\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)H(u,\varepsilon)\right)=\mathbb{E}_{G^\star}\left(\max_{j\in\mathcal{J}}\{u_j+\varepsilon_j\}\right).$$

因此，我们得出

$$W^{DRO}(u)=\mathbb{E}_{G^\star}\left(\max_{j\in J}\{u_j+\varepsilon_j\}\right).$$

(ii) 为了证明 $\nabla W^{DRO}(u)=p^\star(u)$，注意到使用优化后的值 $\Psi(\lambda^\star,\ \mu^\star)=\lambda^\star\rho+\mu^\star+\lambda^\star\mathbb{E}_F\left(\phi^*\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)\right)$，我们得到:

$$\begin{aligned}\frac{\partial W^{DRO}(u)}{\partial u_j} &= \frac{\partial\Psi(\lambda^\star,\mu^\star)}{\partial u_j} \\ &= \int_{\varepsilon\in\mathbb{R}^{J+1}}\left(\phi^{*\prime}\left(\frac{H(u,\varepsilon)-\mu^\star}{\lambda^\star}\right)\frac{\partial H(u,\varepsilon)}{\partial u_j}\right)f(\varepsilon)d\varepsilon \\ &= \mathbb{E}_{G^\star}\left(\frac{\partial H(u,\varepsilon)}{\partial u_j}\right) \\ &= p_j^\star(u).\end{aligned}$$

由于前述结果对所有 $j\in\mathcal{J}$ 均成立，我们得出 $\nabla W^{DRO}(u)=p^\star(u)$。

□

定理的第 (i) 部分表明，给定最优解 $\lambda^\star$ 和$\mu^\star$，剩余函数 $W^{DRO}(u)$ 具有熟悉的期望最大值形式

该形式刻画了随机效用模型（参见式 (2)）。第 (i) 部分的刻画与随机效用模型中的剩余函数之间，主要区别在于表达式 (1) 是关于稳健分布 $G^\star$ 的期望。第 (ii) 部分表明，$W^{DRO}(u)$ 的梯度给出选择概率向量 $p^\star(u)$。后一结果将 Williams–Daly–Zachary 定理推广到了名义分布 F 可能存在误设或不准确的情形。换言之，定理 1 表明，分布鲁棒随机效用模型保留了广受欢迎的随机效用模型的期望最大值形式和梯度结构。

4 分布鲁棒随机效用模型的经验内容

本节讨论分布鲁棒随机效用模型的经验内容。具体而言，我们将展示我们的方法如何适用于恢复平均效用向量，同时允许生成 ε 的真实分布存在不确定性。

为获得一些直观认识，考虑这样一种情形：观测到来自市场数据的选择概率向量 p 。此时，分析者的目标是找到一个向量 u ，使观测到的 p 得到合理解释。根据 Berry (1994)，这一问题称为需求反演。具体而言，Berry (1994) 表明，在多项Logit模型 u 的情形下，满足以下条件

$$p_j = \frac{e^{u_j}}{1 + \sum_{j'=1}^{J} e^{u_{j'}}} \quad \text{for } j = 1, \ldots, J.$$

and

$$p_0 = \frac{1}{1 + \sum_{j'=1}^{J} e^{u_{j'}}}$$

然后，利用前述表达式，我们可以将平均效用向量 u 表示为 p 的函数：

$$\log(p_j / p_0) = u_j \quad \text{for } j = 1, \ldots, J.$$

换言之，我们可以根据观测选择概率向量 p 表示 u 。

我们可以用类似的论证，在嵌套 Logit、随机系数多项 Logit 模型（Berry (1994)；Berry 等人 (1995)）以及 Fosgerau 等人 (2022) 提出的逆向产品差异化 Logit 模型中求得向量 u 。对于多项Logit模型及其变体以外的一般随机效用模型，Galichon 和 Salani´e (2021) 提出了一种基于凸对偶和质量传输技术的通用方法。他们证明，对于 ε 的任意固定 分布，均可根据观测选择概率 p识别平均效用向量 u 。

本节旨在说明，在分析者不知道 ε 的真实分布时，分布鲁棒随机效用模型可用于研究需求反演问题。因此，我们的方法可以在支配 ε 实现值的分布存在模型设定错误的情况下，识别 u 。

4.1 稳健需求反演

我们的主要结果采用了适用于离散选择模型的分布鲁棒版本 Fenchel 等式。为建立这一结果，我们定义 $\mathcal{U} \triangleq \{u \in \mathbb{R}^{J+1} : \ u_0 = 0\}$。换言之， $\mathcal{U}$ 是满足外部选项归一化条件 $u_0 = 0$ 的平均效用向量集合。我们首先考察 $W^{DRO}(u)$ 的凸共轭的性质：

$$W^{(DRO)^*}(p) = \sup_{u \in \mathcal{U}} \left\{ \langle u, p \rangle - W^{DRO}(u) \right\}. \tag{27}$$

特别是，我们感兴趣的是了解 $W^{(DRO)^*}(p)$ 在其有效定义域上的行为：

$$\operatorname{dom} W^{(DRO)^*} = \left\{ p \in \mathbb{R}^{J+1} \,|\, W^{DRO^*}(p) < \infty \right\}.$$

以下引理在我们的分析中起着关键作用。

引理 4 假设假设 *1* 和 *2* 成立，则 $W^{DRO}(u)$ 关于u严格凸。

证明。 给定 λ和 μ，对于 u_1, u_2和 $\alpha \in (0,1)$，有

$$\begin{aligned} &\lambda \mathbb{E}_F \left(\phi^* \left(\frac{H(\alpha u_1 + (1-\alpha) u_2, \varepsilon) - \mu}{\lambda} \right) \right) \\ &\overset{(\star)}{\leq} \lambda \mathbb{E}_F \left(\phi^* \left(\frac{\alpha H(u_1, \varepsilon) + (1-\alpha) H(u_2, \varepsilon) - \mu}{\lambda} \right) \right), \end{aligned}$$

其中，$(\star)$ 的成立源于 H 的凸性，以及由假设2可知 ϕ^* 的单调性。进一步利用 ϕ^* 的严格凸性以及期望算子的线性和单调性，可得：

$$\begin{aligned} &\lambda \mathbb{E}_F \left(\phi^* \left(\frac{\alpha H(u_1, \varepsilon) + (1-\alpha) H(u_2, \varepsilon) - \mu}{\lambda} \right) \right) \\ &< \alpha \lambda \mathbb{E}_F \left(\phi^* \left(\frac{H(u_1, \varepsilon) - \mu}{\lambda} \right) \right) + (1-\alpha) \lambda \mathbb{E}_F \left(\phi^* \left(\frac{H(u_2, \varepsilon) - \mu}{\lambda} \right) \right). \end{aligned}$$

因此可知， $W^{DRO}(u)$关于 u严格凸。 □

以下定理证明了 $W^{(DRO)^*}$的连续性和光滑性。

定理 2 设假设 *1* 和 *2* 成立。凸共轭 $W^{(DRO)^*}$在其定义域 *dom* $W^{(DRO)^*}$ 上连续，该定义域与概率单纯形 Δ_{J+1}一致。此外， $W^{(DRO)^*}$ 在 *int dom* $W^{(DRO)^*}$上连续可微。

证明。 我们首先证明 dom $W^{(DRO)^*} \subseteq \Delta_{J+1}$。固定一个效用向量 $\bar{u}$，并任取满足 $\langle p, e \rangle \neq 1$的 $p \in \mathbb{R}^{J+1}$。然后，利用引理 1(iii)，可得

$$\begin{aligned} W^{(DRO)^*}(p) &\geq \sup_{\gamma \in \mathbb{R}} \langle p, \bar{u} + \gamma \cdot e \rangle - W^{DRO}(\bar{u} + \gamma \cdot e) \\ &\overset{(iii)}{=} \langle p, \bar{u} \rangle - W^{DRO}(\bar{u}) + \sup_{\gamma \in \mathbb{R}} \gamma \left(\langle e, p \rangle - 1 \right) = \infty. \end{aligned}$$

接下来，任取一个向量 $p \in \mathbb{R}^{J+1}$，使得对某个 $i \in \{0,1,\ldots,J\}$ 有 $p_i < 0$。根据引理 1 (ii)，可得

$$W^{(DRO)^*}(p) \geq \sup_{\gamma<0}\langle p, \gamma \cdot e_i\rangle - W^{DRO}(e_i) \overset{(ii)}{\geq} \sup_{\gamma<0} \gamma \cdot p_i, -W^{DRO}(0) = \infty.$$

因此，只需证明反向蕴含，即 $\Delta_{J+1} \subseteq \text{dom } W^{(DRO)^*}$。为此，我们推导凸共轭在单纯形上的一个上界：

$$\begin{aligned}\sup_{p\in\Delta_{J+1}} W^{(DRO)^*}(p) &= \sup_{p\in\Delta_{J+1}} \left(\sup_{u\in\mathcal{U}}\langle p,u\rangle - W^{DRO}(u)\right) \\ &= \sup_{u\in\mathcal{U}} \left(\sup_{p\in\Delta_{J+1}} \langle p,u\rangle - W^{DRO}(u)\right).\end{aligned}$$

我们应用引理 1 中的 (iii)，得到

$$\sup_{u\in\mathcal{U}} \left(\sup_{p\in\Delta_{J+1}} \langle p,u\rangle - W^{DRO}(u)\right) = \sup_{u\in\mathcal{U}} \left(\max_{i\in\mathcal{J}} u_i - W^{DRO}(u)\right) \leq -\min_{i\in\mathcal{J}} \mathbb{E}_F\left[\varepsilon_i\right].$$

因此，定义域与单纯形重合。关于连续性，我们首先注意到 $W^{(DRO)^*}$ 是凸的，因此它在其定义域的相对内部上连续。若定义域是多面体，则 Gale-Klee-Rockafellar 定理保证$W^{(DRO)^*}$ 上半连续；而此处的定义域确为多面体（Rockafellar，1970年）。此外，凸共轭总是下半连续的，因此可得连续性。为证明$W^{(DRO)^*}$ 连续可微，我们注意到引理 4 表明，$W^{(DRO)^*}$ 在 u中严格凸。于是，根据 Hiriart-Urruty 和 Lemarechal（1993年，定理 4.1.1），我们知道，W^{DRO}(u)的严格凸性意味着$W^{(DRO)^*}$(p)在 int $\left(\text{dom } W^{(DRO)^*}\right)$ 上连续可微。□

前述结果对于识别平均效用这一目标至关重要。为说明这一点，我们注意到，根据定理 1，对于备选项 $j \in \mathcal{J}$，我们知道：

$$p_j = \frac{\partial W^{DRO}(u)}{\partial u_j}$$

此外，根据定理 2，我们得到：

$$u_j = \frac{\partial W^{(DRO)^*}(p)}{\partial p_j}$$

其中 u 在（27）中取得最大值。根据 Fenchel 对偶定理，我们知道这两个条件等价。因此，给定鲁棒分布 $G^\star$，我们得出结论：u 可由 p识别。换言之，我们可以找到一个向量 u，使观测选择概率向量 p得到合理解释。

以下结果确立了分布鲁棒随机效用模型的经验内容。

定理 3 设假设 *1* 和 *2* 成立。则以下陈述等价：

(i) 选择概率向量 $p \in \Delta_{J+1}$ 满足：

$$p = \nabla W^{DRO}(u). \tag{28}$$

(ii) 确定性效用向量 $u \in \mathcal{U}$ 满足：

$$u = \nabla W^{(DRO)^*}(p). \tag{29}$$

(iii) $(u^\star, \lambda^\star, \mu^\star)$ 是严格凸优化问题的唯一解：

$$-W^{(DRO)^*}(p) = \inf_{u\in\mathcal{U},\lambda\in\mathbb{R}_+,\mu\in\mathbb{R}} \left\{ \lambda\rho + \mu + \lambda \mathbb{E}_F \left(\phi^* \left(\frac{H(u,\varepsilon) - \mu}{\lambda} \right) \right) - \langle p, u \rangle \right\}. \tag{30}$$

证明。 (i) 和 (ii) 等价，这由定理 2 得出；由此我们可以应用 Fenchel 等式得出结论。为证明 (iii)，我们来看一下 $W^{(DRO)^*}(p)$。根据定义，我们知道

$$W^{(DRO)^*}(p) = \sup_{u\in\mathcal{U}} \{\langle p, u\rangle - W^{DRO}(u)\}.$$

命题 2 表明，前一个表达式对应于

$$W^{(DRO)^*}(p) = \sup_{u\in\mathcal{U}} \left\{ \langle p, u\rangle - \inf_{\lambda>0,\mu\in\mathbb{R}} \left\{ \lambda\rho + \mu + \lambda \mathbb{E}_F \left(\phi^* \left(\frac{H(u,\varepsilon) - \mu}{\lambda} \right) \right) \right\} \right\}.$$

等价地，我们有：

$$W^{(DRO)^*}(p) = - \inf_{u\in\mathcal{U},\lambda>0,\mu\in\mathbb{R}} \left\{ \lambda\rho + \mu + \lambda \mathbb{E}_F \left(\phi^* \left(\frac{H(u,\varepsilon) - \mu}{\lambda} \right) \right) - \langle p, u\rangle \right\}.$$

因此，我们得到：

$$-W^{(DRO)^*}(p) = \inf_{u\in\mathcal{U},\lambda>0,\mu\in\mathbb{R}} \left\{ \lambda\rho + \mu + \lambda \mathbb{E}_F \left(\phi^* \left(\frac{H(u,\varepsilon) - \mu}{\lambda} \right) \right) - \langle p, u\rangle \right\}.$$

结合引理 2 和引理 4，我们得到，(30) 关于 $u, \lambda,$ 和μ严格凸。因此，问题 (30) 存在唯一解。□

正如本节引言中所述，对于给定的 ε 分布，(i) 和 (ii) 两部分已由 Galichon 和 Salani´e（2021年）证明。我们的结果与他们的结果存在一个根本区别：我们实现了对平均效用向量 u的识别，同时放宽了已知 ε 分布这一假设。换言之，即使冲击分布可能存在模型设定错误，我们的结果仍能实现对 u 的非参数识别。同样，我们的结果也与动态离散选择模型的 “ 反演 ” 方法相关。[9] 例如，Hotz 和 Miller（1993年）以及 Arcidiacono 和 Miller（2011年）的论文证明，平均效用向量 u 可以表示为 $\nabla^{-1}W(p) = u$。

[9]值得指出的是，Chiong 等（2016）将 Galichon 和 Salani´e（2021年）的方法应用于广泛使用的动态离散选择模型中。

他们的方法仅适用于多项Logit模型和广义极值模型。Fosgerau 等人（2021年）运用凸优化技术，将 Hotz 和 Miller（1993年）以及 Arcidiacono 和 Miller（2011年）的反演方法扩展到了远超 GEV 类的模型。类似地，Li（2018年）采用凸最小化算法求解需求反演问题。他以 Berry 等人 (1995) 的随机系数 Logit 需求模型和 Berry 和 Pakes（2007年）的纯特征模型为例，展示了该方法。然而，Fosgerau 等人（2021年）和 Li（2018年）的结果仅在已知 ε 分布的假设下适用。相比之下，(iii) 部分证明，给定选择概率向量 p，我们可以将平均效用向量 u 识别为严格凸优化规划 (30) 的唯一解。这个特征刻画通过拉格朗日乘子 $\lambda^\star$ 和 $\mu^\star$ 的取值，体现了模型设定错误的作用。因此，定理 3 给出了一个分布鲁棒非参数识别结果。

4.2 稳健随机系数模型

为了解定理 3 如何应用，我们分析随机系数模型，并假设 ϕ-散度对应于 Kullback-Leibler 距离。遵循 Berry 等人 (1995) 以及 Galichon 和 Salani´e (2021)，我们考虑一个随机系数模型，其中 $\varepsilon = Ze + T\eta$， e 是取值于 $\mathbb{R}^k$、分布为 F_e 的随机向量， Z 是一个 $|\mathcal{J}| \times k$ 矩阵， $T > 0$ 是一个标量参数，而 η 是一个由 $|\mathcal{J}|$ 个冈贝尔随机变量组成的向量，其分布函数为 F_η。假设 e 和 η 统计独立。固定分布 F_e和 F_η后，利用迭代期望并结合 e 与 η 的独立性（Galichon 和 Salani´e，2021，公式 B.6-B.7），我们得到

$$\begin{aligned} W(u) &= \mathbb{E}_{F_e}\left(\mathbb{E}_\eta\left(\max_{j\in\mathcal{J}}\{u_j + (Ze)_j + T\eta_j\}\right)|e\right), \\ &= \mathbb{E}_{F_e}\left(W(u+Ze)\right), \end{aligned}$$

其中 $W(u+Ze) = \int_{\mathbb{R}^{J+1}} \max_{j\in\mathcal{J}}\{u_j + (Ze)_j + T\eta_j\} f_\eta d\eta$。利用$\eta$ 服从 Gumbel 分布这一事实，我们发现

$$W(u+Ze) = T\log\left(\sum_{j\in\mathcal{J}} e^{\frac{u_j+(Ze)_j}{T}}\right).$$

假设 F_e 近似于生成 e 的真实分布。那么，我们可以如下定义 $W^{DRO}(u)$:

$$W^{DRO}(u) = \sup_{G_e\in\mathcal{M}_\phi(F_e)} \mathbb{E}_{G_e}\left(T\log\left(\sum_{j\in\mathcal{J}} e^{\frac{u_j+(Ze)_j}{T}}\right)\right)$$

为应用定理 3，我们注意到 $H(u,\ \varepsilon) = T\log\left(\sum_{j\in\mathcal{J}} e^{\frac{u_j+(Ze)_j}{T}}\right)$。然后，利用 Kullback-Leibler 距离，对于可观测的选择概率向量 p，识别出的平均效用向量 $u^\star$ 对应于以下解

以下规划:

$$-W^{(DRO)^*}(p) = \inf_{u\in\mathcal{U},\lambda>0} \left\{\rho\lambda + \lambda \ln \mathbb{E}_{F_e} \|e^{u+Ze}\|_{T^{-1}} - \langle p, u\rangle\right\}, \qquad (31)$$

其中 $\|e^{u+Ze}\|_{T^{-1}} \triangleq \left(\sum_{j\in\mathcal{J}} e^{\frac{u_j + (Ze)_j}{T}}\right)^T$.

规划 (31) 使我们能够识别平均效用向量，同时允许 e 的分布存在一定程度的模型设定错误。值得指出的是，规划 (31) 相当易于求解，因此我们可以使用传统的随机规划算法求得其唯一解。

5 数值实验

本节讨论我们方法的数值模拟，并将分布鲁棒随机效用模型与多项Logit模型和多项Probit模型进行比较。[10]

我们的主要目标是分析稳健性指标 ρ 对选择概率的影响。我们考虑一个包含四个备选项的情景，其中 $\mathcal{J}=\{0,1,2,3\}$。效用向量 u 的第一种参数化设定为 $u=(0,1,2,2.1)^T$。基于这一设定，我们计算选择概率。在多项Logit模型中，选择概率通过式 (4) 计算，尺度参数等于 1（ $\eta=1$）。此外，我们假设每个冈贝尔误差的位置参数为零。对于多项Probit模型，我们考虑随机误差向量的方差-协方差矩阵的两种不同参数化设定； $\mathcal{N}(0,\Sigma_1)$ 和 $\mathcal{N}(0,\Sigma_2)$，其中

$$\Sigma_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} 2 & -0.5 & 0.5 & 1.3 \\ -0.5 & 2 & 0 & 0.15 \\ 0.5 & 0 & 2 & 1 \\ 1.3 & 0.15 & 1 & 2 \end{pmatrix}.$$

我们将后一个模型称为 MNP-dep，将前一个模型称为独立多项Probit模型，因为在前一个模型中，随机误差 $\varepsilon^{(i)}, i=1$、.....、4 相互独立。我们从误差向量中抽取 10,000,000 个样本，以稳定模拟结果并模拟选择概率。

对于分布鲁棒随机效用模型，我们选择例 1 中介绍的 Kullback-Leibler 散度情形。我们假设名义分布的误差项独立同分布，且服从位置参数为零、尺度参数为一的冈贝尔分布。这样便可考察分布鲁棒随机效用模型的行为和数值稳定性，以及 ρ 对选择概率的影响。

鲁棒选择概率的模拟方式与多项Probit模型类似。然而，对于分布鲁棒随机效用模型，我们必须从由密度 25 定义的分布中生成样本。首先，使用来自 4 个独立同分布的Gumbel分布的 50，000，000 次模拟，估计 (24) 中的最优 $\lambda^\star$ 以及 $\mathbb{E}_F\left(e^{H(u,\varepsilon)/\lambda^\star}\right)$。基于优化后的参数，一个更高维的

[10]我们回顾一下，多项Probit模型假设误差项服从具有特定方差-协方差矩阵的正态分布。

接受-拒绝算法提供了一种高效的采样方法。为提升性能，代码使用 Julia 编写。[11]

结果见表2

	备选方案1	备选方案2	备选方案3	备选方案4
MNL	5.1885%	14。1037%	38.3379%	42.3699%
独立多项Probit模型	1.6243%	10。2996%	41.443%	46.6331%
MNP-dep	1。7877%	19。359%	37.9993%	40。854%
$\rho = 0.1$	9.2391%	18.4783%	33。695%	38.587%
$\rho = 0.7$	13。2132%	19。7823%	31。6066%	35。3979%
$\rho = 1.3$	15。4725%	21.5501%	30。5269%	32。4505%
$\rho = 2.2$	18.1501%	21。46%	29。7536%	30。6363%
$\rho = 4.3$	21.5843%	23。2006%	27。1235%	28。0916%

表2：效用向量 $u =$(0，1，2，2。1)^T 的选择概率

.

在上表中，第一行显示了多项Logit模型的选择概率。第二行和第三行分别显示了 MNP-indep 和 MNP-dep 的选择概率。第四行显示了 $\rho = 0.1$ 时分布鲁棒随机效用模型的行为。对于这一参数化设定，分布鲁棒随机效用模型得出的选择概率与 MNP-dep 所示的概率相似（但并不相等）。在 $\rho = 0.7$ 的情况下也观察到类似行为。

第六至第八行显示了随着 ρ 增大，分布鲁棒随机效用模型的行为。正如预期，随着 ρ 增大，选择概率看起来会接近备选项之间的均匀选择。具体而言，在 ρ =4。3 的情况下，我们注意到分布鲁棒随机效用模型赋予的概率与均匀情形相似。直观来说，较大的 ρ 表示分析者对真实分布极不确定。因此，她的行为会过度谨慎，并考虑一大组可能（且可行）的分布。因此，当 $\rho \longrightarrow \infty$ 时，分析者的最佳选择是猜测均匀概率。

同样，从决策者的角度来看，ρ 较大的值表示对误差项采取谨慎且灵活的选择。因此，随机误差项可能服从一种分布，完全抵消确定性效用的影响，并保证每个备选项具有相同的整体随机效用。事实上，如图2所示，随着稳健性指标增大，鲁棒剩余函数 (23) 显著递增。图中绘制了该剩余函数在 $u =$ (0，1，2，2。1)^T 处取值随不同 ρ 变化的情况。

[11]代码可在 GitHub 上的 `https://github.com/rubsc/rejection DRO RUM`找到。

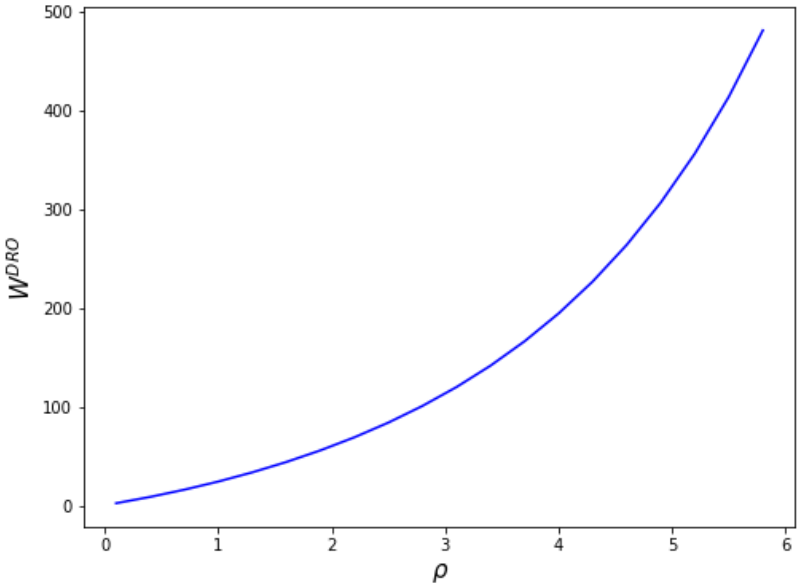


图2：效用向量 u 在不同稳健性指标 ρ取值下的稳健剩余 W^{DRO} 。

MNL 模型一个广为人知的缺陷是满足无关选项独立性性质。无关选项独立性性质指出，任意两个备选项的选择概率之比仅取决于这两个备选项效用之间的差异。由公式 (4) 可直接得到这一性质。由此直接可知，当一个备选项的确定性效用发生变化时，选择概率会按比例变化，因此备选项之间的概率比保持不变。相比之下，分布鲁棒随机效用模型将某种依赖结构纳入了 MNL。[12] 因此，模拟确定性效用向量略有变化时的选择概率颇有意义。表3汇总了效用向量为 $\tilde{u} = (0，1，2，2。2)^T$时各备选项的选择概率。

[12]我们回顾一下，假设名义分布为冈贝尔分布。

	备选方案1	备选方案2	备选方案3	备选方案4
MNL	4.9671%	13。5021%	36.7024%	44.8284%
独立多项Probit模型	1.4758%	9.5821%	39.2676%	49.6745%
MNP-dep	1。5544%	18.5851%	36.0749%	43.7856%
$\rho = 0.1$	6.8788%	15。0473%	36。5434%	41.5305%
$\rho = 0.7$	13。2076%	20.2437%	30.8569%	35。6918%
$\rho = 1.3$	15.2938%	21.5069%	30。7281%	32.4712%
$\rho = 2.2$	17.9704%	21。4406%	29.5254%	31。0636%
$\rho = 4.3$	21。5509%	23。264%	27。4421%	27.743%

表3：效用向量 $\tilde{u} = (0, 1, 2, 2.2)^T$ 对应的选择概率

图3展示了无关选项独立性的违反情况。请注意，在多项Logit模型中，选择备选方案4的概率下降后，选择备选项 $1 - 3$ 中某一个的概率会均匀上升，如虚线所示。与此同时，稳健模型的替代模式则灵活得多。

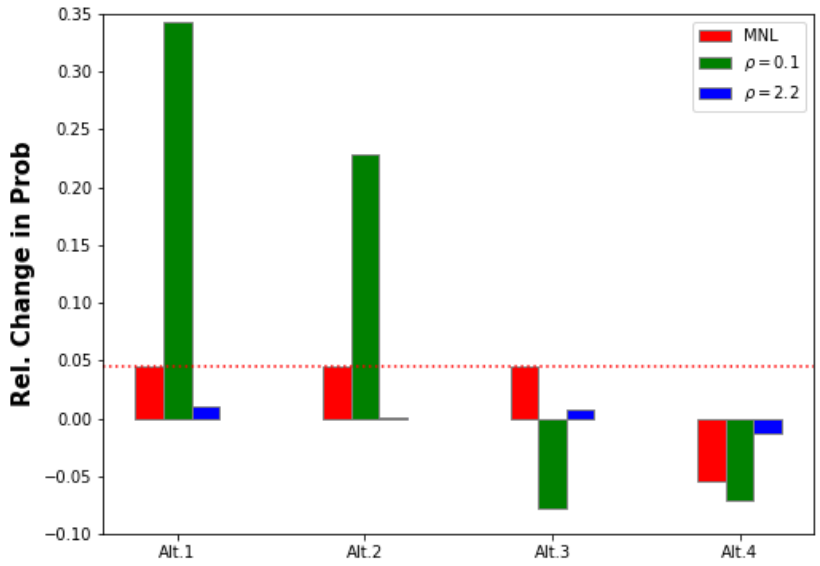


图3：确定性效用向量从 $\tilde{u}$ 变为 u 时，选择概率的相对变化。

6 总结

本文介绍了分布鲁棒随机效用模型（DRO-RUM），该模型允许冲击分布未知或设定有误。我们证明，DRO-RUM

保留了传统随机效用模型（RUM）的易处理性和凸结构。此外，我们刻画了DRO-RUM的经验内容，证明对于给定的观测选择概率向量，存在唯一的平均效用向量，能够从DRO-RUM的角度合理解释所观测到的行为。最后，我们展示了该方法的稳定性和数值特性。

还有几种可能的扩展。首先，一个自然的问题是，DRO-RUM 在市场数据上的经济表现如何。这一点尤其值得关注，因为我们的方法为分析者提供了一类丰富多样的模型。事实上，只需选择不同的ϕ 散度和/或名义分布，就能构建不同的模型。在此背景下，研究稳健性参数 ρ 的影响也很有意义。

其次，可以在稳健性框架下分析 RUM 的结果。例如，本文的结果有助于研究具有可转移效用的双边匹配市场。同样，我们的结果也有助于研究动态离散选择模型中的鲁棒识别。

可以进一步研究 DRO-RUM 的算法方面。例如，M¨uller 等人（2022年）近期提出了一类基于离散选择模型、定义在概率单纯形上的新近端函数。因此，探究能否从 DRO-RUM 中构造近端函数也很有意义。

此外，分布鲁棒方法在理论上也可以进一步拓展。一种自然的做法是采用不同的统计距离概念，例如 Wasserstein距离，并分析其可处理性。此外，还可以将其他此类稳健模型的性质与分布鲁棒随机效用模型进行比较。

参考文献

Peter Arcidiacono 和 Robert A. Miller。利用条件选择概率估计含未观测异质性的动态离散选择模型。《计量经济学》，79(6):1823–1867，2011年。Gordon M. Becker、Morris H. Degroot 和 Jacob Marschak。选择行为的随机模型。《行为科学》，8(1):41–55，1963年。A. Ben-Tal 和 M. Teboulle。通过 φ-散度泛函研究随机规划中的惩罚函数与对偶性。《运筹学数学》，12(2):224–240，1987年。Steven Berry 和 Ariel Pakes。纯特征需求模型。《国际经济评论》，48(4):1193–1225，2007年。Steven Berry、James Levinsohn 和 Ariel Pakes。市场均衡中的汽车价格。《计量经济学》，63(4):841–890，1995年。Steven T. Berry。产品差异化离散选择模型的估计。《兰德经济学杂志》，25(2):242–262，1994年。ISSN 07416261。

H.D. Block 和 Jacob Marschak。随机排序与随机响应理论。考尔斯基金会讨论论文 66，耶鲁大学考尔斯经济学研究基金会，1959年。Khai Xiang Chiong、Alfred Galichon 和 Matt Shum。动态离散选择模型中的对偶性。《数量经济学》，7(1):83–115，2016年。Timothy Christensen 和 Benjamin Connault。反事实敏感性与稳健性。《计量经济学》，91(1):263–298，2023年。I. Csiszar。概率分布差异的信息型度量与间接观测。《匈牙利数学科学研究》，2: 229–318，1967年。Bernard Dacorogna 和 Pierre Mar´echal。透视函数在凸性、多重凸性、秩一凸性和分别凸性中的作用。《凸分析杂志》，15(ARTICLE):271–284，2008年。Guiyun Feng、Xiaobo Li 和 Zizhuo Wang。技术说明——关于若干离散选择模型之间的关系。《运筹学》，65(6):1516–1525，2017年。Mogens Fosgerau、Emerson Melo、Matthew Shum 和 Jesper R.-V. Sørensen。关于动态模型基于条件选择概率的估计量的一些评论。《经济学快报》，204:109911，2021年。ISSN 0165-1765。Mogens Fosgerau、Julien Monardo 和 Andre de Palma。逆向产品差异化 Logit 模型。工作论文，2022年。Alfred Galichon 和 Bernard Salani´e。《丘比特的无形之手：匹配模型中的社会剩余与识别》。《经济研究评论》，89 (5):2600–2629，2021年12月。ISSN 0034-6527。Lars Peter Hansen 和 Thomas J. Sargent。稳健控制与模型不确定性。《美国经济评论》，91(2):60–66，2001。ISSN 00028282。Lars Peter Hansen 和 Thomas J. Sargent。稳健性。普林斯顿大学出版社，学生版，2008年。Jean-Baptiste Hiriart-Urruty 和 Claude Lemarechal。《凸分析与极小化算法 *II*》。Addison-Wesley Professional，第2版，1993年。V. Joseph Hotz 和 Robert A. Miller。条件选择概率与动态模型估计。《经济研究评论》，60(3): 497–529，1993年07月。ISSN 0034-6527。Zhaolin Hu 和 L. Jeff Hong。受 Kullback-Leibler 散度约束的分布鲁棒优化。工作论文，2012。

Daniel Kuhn、Peyman Mohajerin Esfahani、VietAnh Nguyen 和 Soroosh Shafieezadeh-Abadeh。《*Wasserstein*分布鲁棒优化：机器学习中的理论与应用》，第 6 章，第 130–166 页。2019 年。

Lixiong Li. 需求反演的一般方法，2018年。F. Liese 和 I. Vajda。凸统计距离。莱比锡：*Teubner*数学丛书，第*95*卷。，1987年。Jacob Marschak。随机效用指标的二元选择约束。技术报告，1959年。D. McFadden。住宅区位选择建模。载于 *A. Karlqvis*、*A.*、*Lundqvist*、*L.*、*Snickars*、*L.*、*Weibull*、*J.*（编），《空间相互作用理论与规划模型》（北荷兰出版社，阿姆斯特丹），第531–551页，1978a。D. McFadden。住宅区位选择建模。《交通运输研究记录》，（673）：72–77，1978b。D. McFadden。空间相互作用理论与住宅区位，章节：住宅区位选择建模，第75–96页。北荷兰出版社，阿姆斯特丹，1978c。D. McFadden。计量经济学应用中的离散数据结构分析，章节：概率选择的计量经济学模型，第198–272页。剑桥：麻省理工学院，1981。Daniel McFadden。经济选择。《美国经济评论》，91(3)：351–378，2001年6月。Vinit Kumar Mishra、Karthik Natarajan、Hua Tao 和 Chung-Piaw Teo。当效用相关时，利用半定优化进行选择预测。《*IEEE*自动控制汇刊》，57(10)：2450–2463，2012。Vinit Kumar Mishra、Karthik Natarajan、Dhanesh Padmanabhan、Chung-Piaw Teo 和 Xiaobo Li。边际分布选择模型的理论与实证方面。《管理科学》，60(6)：1511–1531，2014。David M¨uller、Yurii Nesterov 和 Vladimir Shikhman。单纯形上的离散选择近端函数。《运筹学数学》，47(1)：485–507，2022年。Karthik Natarajan、Miao Song 和 Chung-Piaw Teo。持久性模型及其在选择建模中的应用。《管理科学》，55(3)：453–469，2009年。L. Pardo。基于散度度量的统计推断。Chapman and Hall/CRC，第1版，2005年。

R. Tyrrell Rockafellar。积分泛函、正规被积函数与可测选择。载于 Jean Pierre Gossez、Enrique Jos´e Lami Dozo、Jean Mawhin 和 Lucien Waelbroeck 编，《非线性算子与变分法》，第157–207页，柏林、海德堡，1976。施普林格柏林海德堡。ISBN 978-3-540-38075-7。

T. R. Rockafellar。《凸分析》。1970年。Andrzej Ruszczy´nski 和 Alexander Shapiro。《随机规划讲义：建模与理论，第三版》。2021年。H. Scarf。库存问题的极小极大解。《库存与生产数学理论研究》，1958年。Alexander Shapiro。分布鲁棒随机规划。《*SIAM*优化杂志》，27(4):2258–2275，2017a。Alexander Shapiro。分布鲁棒随机规划。《*SIAM*优化杂志》，27(4):2258–2275，2017b。Kenneth E. Train。《离散选择模拟方法》。剑桥大学出版社，第2版，2009年。