arXiv:2302.01236v1 [stat.AP] 2 Feb 2023

# 一种衡量气候适应的机器学习方法

Max Vilgalys*

**摘要**

我通过比较有害天气的短期变化与长期变化所对应的弹性，来衡量对气候变化的适应。我提出了一种去偏机器学习方法，以便在面板数据情境下灵活地衡量这些弹性。在一项模拟实验中，我表明，相较于标准机器学习或普通最小二乘法，去偏机器学习具有显著优势，尤其是在高维情境下。随后，我衡量了美国玉米和大豆生产对有害高温暴露的适应。利用丰富的温度和降水变异数据，我发现，有害高温的短期影响在长期内会得到显著抵消。我表明，这是因为热暴露长期变化的影响并不遵循与热暴露短期冲击相同的函数形式。

**关键词：** 气候变化、方向导数、机器学习、面板数据

**JEL分类：** C14、C33、C55、Q51、Q54

## 1 引言

衡量对近期气候变化的适应，对于制定气候政策和预测未来气候变化造成的损害至关重要。近期适应的程度能够表明何时需要政策干预，也有助于了解我们能够在多大程度上适应预计的变化。研究人员可以通过建模天气冲击对经济结果的影响，并考察影响如何随着天气冲击暴露程度的增加（例如，Burke 和 Emerick（2016））或时间推移（例如，Barreca 等，2016年）而变化，来开展这项研究。这些研究依赖于对天气与经济结果之间关系的准确建模。

*max.vilgalys@erg.com Eastern Research Group, Inc.

当研究人员拥有丰富的数据，却缺乏专家指导来确定函数形式时，机器学习（机器学习）可以帮助建模这些关系。如果领域专家提供了天气冲击如何影响经济结果的模型，研究人员便可以据此准确衡量和比较影响。否则，研究人员就必须从数据中学习这种关系。经济学家通常使用经典统计工具，在不施加强函数形式假设的情况下对天气—经济关系进行建模（Hsiang，2016年）。这类工具适用于低维天气变化，但随着协变量维度增加，可能导致方差较高或估计结果不一致。即使模型只包含温度和降水，如果研究人员灵活地对每个变量及其相互作用进行建模，也可能变成高维模型。机器学习非常适合此类情境下的灵活建模（Mullainathan 和 Spiess，2017年），并能帮助研究人员充分利用现代面板数据集中的高维天气变异。

本文提出一种机器学习方法，用于研究美国（U.S.）玉米和大豆生产如何适应有害高温暴露。高温通常会损害作物生长，尽管长期适应可能抵消部分损害。Schlenker 和 Michael J. Roberts（2006）提出了一个简约模型，说明热暴露年度冲击如何影响作物产量。他们将作物产量建模为热暴露的分段线性函数，并以作物特定的温度阈值为界：低于阈值的热暴露有益，高于阈值的热暴露则有害。通过利用不同的变异来源估计有害高温的参数，Schlenker 和 Michael J Roberts（2009）、Burke 和 Emerick（2016）以及 Lemoine（2018）认为，观察到的对有害高温暴露的适应程度不足以抵消预期的气候损害。

我先根据数据学习作物产量与天气之间的关系，再衡量适应程度。我比较有害高温暴露在短期和长期变化下的影响。对于低维线性模型，这等同于Burke和Emerick（2016）提出的方法。该方法允许我采用更灵活的模型，并纳入高维天气变异，因此适用于天气变异丰富、且缺乏关于这种变异如何影响经济结果的专家指导的应用场景。我使用机器学习对作物产量与天气之间的关系建模，具体采用最小绝对收缩和选择算子（Lasso）和神经网络（NNet）。我的机器学习模型纳入加性固定效应，同时灵活地对温度、降水及二者之间的交互作用建模。

为估计适应程度，我估计作物产量相对于有害高温暴露的弹性。该弹性概括了有害高温暴露对作物产量的影响，并且可以针对每个模型计算。与Burke和Emerick（2016）的方法不同，我无法在不同模型之间比较某个单一参数。相反，我通过对数作物产量关于温度和降水的回归函数来拟合并计算沿有害高温暴露边际增加方向的平均方向导数，从而求得该弹性。我使用普通最小二乘法（OLS）和机器学习估计这些回归函数。

我采用去偏程序来处理标准机器学习模型产生的偏差。机器学习方法可能因过拟合或正则化而引入偏差，但在估计因果参数或基于回归函数的统计量时，有一些方法可以减少这种偏差（CCDDHNR，2018年；Chernozhukov、Newey 和 Singh，2022a）。我对 Chernozhukov、Newey 和 Singh（2022a）的估计量进行了调整。该方法采用双重机器学习（DML），利用第二种机器学习算法校正标准机器学习估计值。我调整了第二种机器学习算法，使其适用于面板数据情境。随后，我应用这一双重机器学习程序，对弹性估计值进行去偏。

我通过比较弹性，估计有害高温暴露造成的短期影响在长期内被抵消的程度。我根据1990—2019年的数据，构建了长期和短期作物产量及天气变量变异的面板数据集。随后，我分别在每个数据集中估计作物产量相对于有害高温暴露的弹性。我使用普通最小二乘法、双重机器学习（DML）以及未经偏差校正的机器学习来计算这一弹性。通过比较这些估计值，我发现有害高温暴露造成的短期影响在长期内被抵消的程度。

在将这一方法应用于数据之前，我先开展一项模拟实验，评估双重机器学习对弹性的估计值，并将其与普通最小二乘法和标准机器学习方法进行比较。每次模拟试验都采用美国县温度和降水的经验分布，但模拟结果变量。我考察双重机器学习程序相对于普通最小二乘法以及未经偏差校正的机器学习方法的表现，比较恢复真实弹性时的偏差和方差。我使用三种可能的天气变量集合：一组简单且常用的年气温和降水变量；一组包含更丰富年气温变异的变量；以及一组包含月度降水观测值和丰富温度变异的变量。

这项模拟实验清晰地凸显了在高维情境下使用去偏机器学习的优势。在高维情境下，神经网络和 Lasso 都比普通最小二乘法更具优势，

尽管神经网络的偏差更低。利用丰富的年气温变化数据，双重机器学习估计值的方差显著低于普通最小二乘法，偏差也低于标准机器学习估计值。利用月份温度和降水数据集，双重机器学习估计值的偏差和方差均显著低于普通最小二乘法，偏差也低于标准机器学习。

随后，我将这一方法应用于美国作物产量数据集，发现有证据表明，适应正在抵消有害高温暴露带来的影响。我在1990年至2019年玉米和大豆产量的县级数据集上，采用 Lasso、神经网络和普通最小二乘法估计量。通过短期和长期变化，我比较了弹性的估计值。按照模拟实验中的做法，我使用三组天气变化，对估计弹性进行了500次自助法试验。采用 Schlenker 和 Michael J Roberts（2009年）以及 Burke和Emerick（2016年）所用的简单年度天气变量集后，我发现，气候变化对玉米或大豆生产几乎没有产生显著的适应。这证实了 Burke和Emerick（2016年）的研究结果。

采用更灵活的天气变化组合后，我发现，有害高温暴露造成的短期影响很大一部分在长期内得到了抵消。根据短期变异，我发现有害高温暴露边际增加会导致产量出现统计上和经济上显著的下降。然而，采用长期变化时，我没有发现此类下降的证据。玉米和大豆的结果均如此。主要区别在于使用了更丰富的天气变量集。使用更灵活的天气变化组合进行普通最小二乘法估计时，我也得出了相同结论，不过双重机器学习方法得出的置信区间更窄。这表明，有害高温暴露造成的短期影响有相当大一部分在长期内得到了抵消，说明人们对这种热暴露进行了显著适应。

这一结果与 Burke和Emerick（2016）以及其他分析（Schlenker 和 Michael J Roberts，2009年；Lemoine，2018年）的结论大相径庭。这很可能是由于模型对长期热暴露变化对作物产量的影响设定错误所致。我表明，在包含长期变化的面板数据中，Schlenker 和 Michael J. Roberts（2006年）提出的简单模型无法充分概括温度变量的灵活作用。尽管有害高温暴露与作物产量下降相关，但其他温度变化更能解释这些下降。这表明，人们对气候变化的某些有害特征的适应程度有限，但并未适应有害高温暴露边际增加这一特定特征。

本文是　　与几项文献相关。首先是关于估计适应程度的文献　　f

气候变化适应。衡量对气候变化的适应，需要了解天气如何影响经济结果。有关衡量天气经济影响的经济学文献综述，请参见 Dell、Jones 和 Olken（2014）。Hsiang（2016）概述了衡量这些影响的计量经济学方法。这类文献大多聚焦于农业，因为该部门直接暴露于天气之下，因此尤其容易受到气候变化的影响（Shukla 等，2019）。研究气候变化影响的首种方法是李嘉图方法，研究者通过横截面比较农业用地的价值。Mendelsohn、Nordhaus 和 Shaw（1994）通过对美国县横截面数据中的平均气温和农业地产价值进行回归，预测气候变化的影响。这种方法容易受到遗漏变量偏误的影响，后续研究则致力于处理特定的遗漏变量，例如农民技术的内生变化（Kurukulasuriya、Kala 和 Mendelsohn，2011）或非农收入（Ortiz-Bobea，2020）。其他用于估计农业适应潜力的方法包括全经济模拟（Costinot、Donaldson 和 Smith，2016）、历史迁徙期间的生产变化（Sutch，2011；Olmstead 和 Rhode，2011）、自然实验（Hornbeck，2012；Hagerty，2021）或面板方法。

面板方法通过识别面板数据集中年度或长期变异所体现的适应，来解决遗漏变量偏误。Schlenker 和 Michael J Roberts（2009年）利用面板数据估算作物产量对极端高温暴露的弹性，并得出结论：由于美国南部和北部的损害相近，尽管两地气候存在差异，但适应气候变化的潜力有限。Barreca等（2016年）使用灵活的温度暴露模型，记录了极端高温对死亡率的影响在 $20^{\text{th}}$ 世纪以来如何减弱。Burke和Emerick（2016年）通过使用不同的变异来源来识别天气冲击和平均气温变化的影响，表明可以利用面板变异估算对近期气候变化的适应程度。Lemoine（2018年）提出了另一种方法，通过考察事前适应和事后适应在应对高温暴露冲击中的作用，对可能的适应程度进行部分识别。

我的论文与Burke和Emerick（2016年）最为密切相关。与他们的论文一样，我估计短期天气变化对玉米和大豆产量造成的损害在更长期暴露下得到抵消的程度。我还使用了来自美国农业的作物和天气数据。不同之处在于，我考虑了更丰富的天气变量集合，并使用双重机器学习来建模和学习这些数据与作物产量之间的关系

我得出结论认为，对于有害高温暴露，适应程度有所提高。

其次，越来越多的文献开始探讨在经济学中应用机器学习方法。Kleinberg 等人（2015年）讨论了预测性机器学习在经济学中的应用，Varian（2014年）和 Mullainathan 和 Spiess（2017年）则提供了实用的算法指南。近期有几篇论文利用机器学习来衡量环境经济学中的重要结果。Crane-Droesch（2018年）提出了一种半参数神经网络，并用其研究气候变化对玉米产量的影响。Deryugina 等人（2019年）采用机器学习方法来衡量空气污染的成本。Burlig 等人（2020年）利用机器学习来改进能源效率提升效果的估计。Stetter、Mennig 和 Sauer（2022年）采用双重机器学习方法来衡量一项农业干预措施的有效性；Klosin 和 Vilgalys（2022年）则提出了一种双重机器学习方法，用于衡量面板数据情境下的弹性。农业领域也有许多相关应用；Liakos 等人（2018年）对此进行了综述。

本文与 Crane-Droesch（2018）和 Klosin 和 Vilgalys（2022）的研究最为相关。Crane-Droesch（2018）估计了一个能够控制不可观测县级固定效应的神经网络，并利用该模型预测反事实气候变化情景下的产量。我采用相同的技术处理县级固定效应，但修改了算法，以便从网络中求出导数，并确保网络可微。Crane-Droesch 研究未来气候变化的影响，而我则利用这一工具来理解近期气候变化下的适应。与 Klosin 和 Vilgalys（2022）类似，我估计了作物产量相对于有害高温暴露增加的弹性。我采用更灵活的方式表示温度变化，并将该估计量用于衡量对近期气候变化的适应。

在机器学习文献中，本文应用了新兴的双重机器学习（DML）领域的研究成果。CC DDHNR（2018）证明，在某些情境下，样本分割并构造 Neyman 正交矩条件，可以得到近似去偏的机器学习估计。Semenova 和 Chernozhukov（2021）将 Neyman 正交矩条件方法扩展至其他若干统计目标，包括结构导数。另一种方法是，Chernozhukov、Newey 和 Singh（2022a）以及 Chernozhukov、Newey 和 Singh（2022b）基于里斯表示定理，为更一般的一类线性泛函提出了近似去偏估计量。Klosin 和 Vilgalys（2022）在面板设定中采用了这一方法，并证明了平均导数估计量的渐近正态性。与 Klosin 和 Vilgalys（2022）类似，

我的方法在面板设定中应用了 Chernozhukov、Newey 和 Singh（2022a）的结果；我采用不同的方法处理固定效应，并考虑对温度变化进行更灵活的表示。

本文其余部分安排如下。第 2 节介绍本项目使用的数据。我将展示气候变化的程度，并描述生成生长度日所需的变换。第 3 节介绍所用方法，包括如何计算平均导数的详细说明，以及对去偏机器学习估计方法的解释。第 4 节详细介绍模拟实验及其结果。第 5 节展示并讨论使用该估计程序衡量气候变化适应程度所得的结果。第 6 节为结论。

## 2 数据

在实证应用中，我使用1990—2019年美国玉米和大豆生产的天气与作物数据。我考察西经 100° 子午线以东的县，这些县构成了玉米和大豆种植较为集中的农业区。1990—2019年，该地区生产的玉米占全国总产量的93%以上，大豆占全国总产量的99%以上。作物数据包括美国农业部《农业调查》中的玉米和大豆年度产量（蒲式耳/英亩）。这些数据还包括各县的种植面积（英亩）。

天气数据由 Schlenker 和 Michael J Roberts 慷慨提供，他们提供了一个网格数据集，其中包含来自一组持续报告数据的气象站的日温度和降水数据。与 Schlenker 和 Michael J Roberts（2009年）以及 Ortiz-Bobea（2013年）一样，我考察3月至8月生长季期间的天气。我将网格数据集汇总为县级数据集，其中包括日最高气温、最低气温和降水。随后，我将日温度暴露转换为按月份和生长季计算的生长度日（GDD）。图1展示了从日温度观测值转换为生长度日的过程。

本文始终考察每个县的三组天气变量：生长季中高于和低于29°C的总热暴露，加上生长季总降水量（年度线性）；生长季中每个1°C温度区间的总热暴露，加上生长季总降水量（年度灵活）；以及生长季每个月各1°C温度区间的热暴露和每月降水量（月度灵活）。图1展示了温度变量的转换方式，

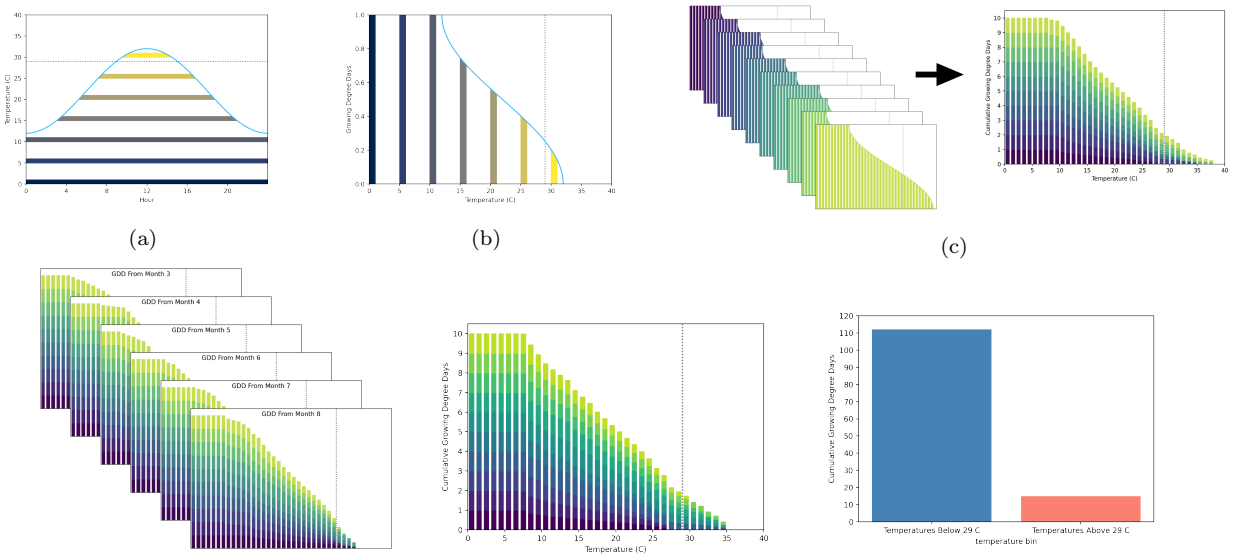
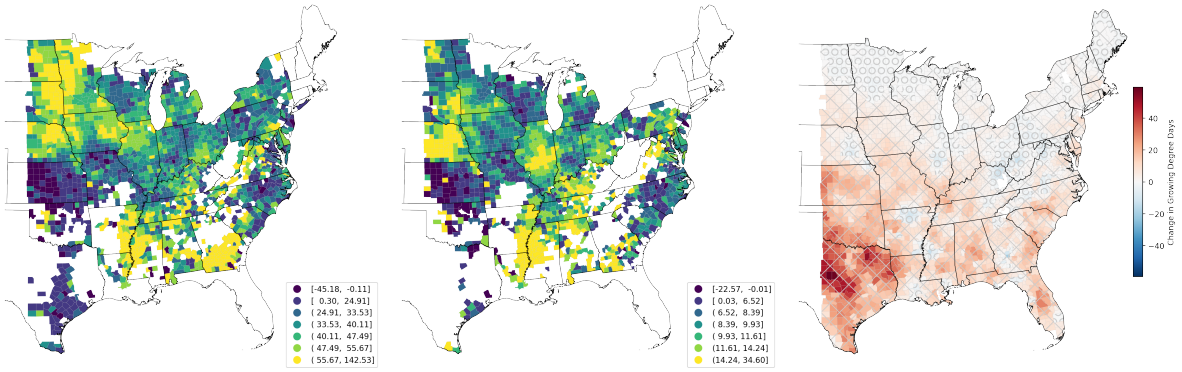



图1：将每日最低和最高气温记录转换为分析中使用的天气变量。(a) 显示了一天内的日温度。(b) 中，该温度被转换为生长度日（GDD）。(c) 展示了如何将每日 GDD 观测值汇总为累积 GDD 暴露量。(d) 为月度灵活变量，即将每日温度按月份和温度区间汇总为生长度日。(e) 中，这些数据进一步汇总为各温度区间的整个生长季热暴露量。(f) 中，这些数据进一步汇总为虚线 29°C 以上和以下的整个生长季热暴露量。

这些转换方式如图1所示。图1f展示的年度线性转换在经济学分析中被广泛采用。Schlenker 和 Michael J Roberts（2009）指出，在面板数据情境下，基于高于和低于作物特定的致害阈值的热暴露进行的回归，对产量变异的解释程度与更灵活的模型相当。Burke和Emerick（2016）确定，长期来看，29°C是玉米和大豆遭受有害高温暴露的阈值。本文在整个分析中采用29°C这一阈值，并将高于29°C的热暴露称为有害高温暴露。

我们利用各县气候变化程度的差异，估算样本中的长期弹性。图2c展示了1990—1999年与2010—2019年间，美国各县超过29°C的生长度日（GDD）暴露量的差异。样本中80%的县热暴露有所增加，但即使在同一州内，增温程度也存在显著差异。许多相邻县中，一个县的有害高温暴露有所增加，而另一个县则有所减少。这种变异可能与土壤质量或州固定效应等不可观测因素不相关。详见Burke




(a) 玉米产量　　(b) 大豆产量　　(c) 高于29°C的生长度日

图2：1990—1999年平均值与2010—2019年平均值之间，玉米产量、大豆产量以及生长季高于29°C的生长度日的县级数值差异。地图包含西经100度子午线以东的县。在(a)和(b)中，第一个颜色区间涵盖所有产量下降的县（约占县总数的4%），其余分布则在剩余区间中均匀划分。在(c)中，“X”（“O”）阴影线表示高于29°C的热暴露增加（减少）的县。

和Emerick（2016），其中详细论述了这种变异如何用于识别气候变化的影响。

尽管许多县的有害高温暴露有所增加或减少，但只有少数县在这两个时间段之间的平均产量下降。图2a和图2b显示了这一时间段内玉米和大豆平均产量的变化。在1990—1999年和2010—2019年两个时期都种植玉米或大豆的县中，只有4.03%的县玉米产量下降，只有3.96%的县大豆产量下降。这表明，农业技术的进步提高了样本中各县的产量，而且这些进步带来的提升超过了有害高温暴露增加所造成的损害。

## 3 方法

本节讨论用于估计适应程度的方法。首先，我从平均方向导数的角度定义适应，并介绍如何测量这些平均方向导数。随后，我将介绍普通最小二乘法（OLS）、Lasso 和神经网络（NNet）程序，包括如何在各模型中处理固定效应项，以及交叉折叠训练程序的具体细节。我还将介绍用于调整标准机器学习（ML）估计中偏差的双重机器学习（DML）方法。

### 3.1 适应

我将适应定义为：长期内抵消的有害高温暴露对产量造成的短期影响。这个定义涵盖了主体为应对气候变化而对生产技术、管理实践或各类作物的品种选择所采取的所有适应行为

。但它未涵盖其他一些重要的适应途径，例如作物转换或退出农业。Burke和Emerick（2016年）提供的证据表明，这些适应途径的作用有限。这个定义也未涵盖那些不依赖于个体气候变化暴露程度、却会改变有害高温影响的变化，例如Barreca等（2016年）研究的热相关死亡率下降。首先，我将定义由Burke和Emerick（2016年）提出的估计目标。然后，我将以弹性的形式对该定义加以推广，并讨论如何使用灵活函数形式和／或更丰富的天气变异来估计这些弹性。

遵循Burke和Emerick（2016）的做法，我通过比较短期和长期变异下作物产量对极端高温的弹性来衡量适应。短期变异来自年度面板数据中天气观测值和作物产量对数的逐年变化。为捕捉长期变化，我首先对较长时期内的天气和作物产量数据取平均值（我采用10年为一个时期），然后构建两期面板数据。利用短期变异计算出的弹性，反映单一年份的极端高温天气冲击对产量的影响程度；利用长期变异计算出的弹性，则反映长期暴露于升高的热暴露对平均作物产量的影响程度。使用长期变异得出的估计值，或许能够反映气候变化造成的损害程度，因为在相对较长的时期内，平均热暴露发生了变化，农民也有时间适应这些变化。图2展示了样本中的长期变化，比较了我所研究的数十年间长期平均作物产量和有害温度暴露的差异。比较基于这些变异来源得出的影响，我可以判断短期损害是否会在长期内得到抵消。

在一个产量对整个生长季热暴露的对数呈线性关系的模型中，直接估计这些弹性是很简单的。Burke和Emerick（2016年）采用了这样一个模型，下面我将对其进行调整：

$$y_{it} = a_i + \beta_1 lower_{it} + \beta_2 higher_{it} + g(prec_{it}) + \varepsilon_{it} \tag{1}$$

其中， $higher_{it}$ （$lower_{it}$）表示生长季内高于（低于）有害温度阈值的总热暴露， $g$ 是降水量的某种函数， $a_i$ 是每个县的加性固定效应项， $\varepsilon_{it}$ 是加性误差项。Burke和Emerick（2016年）使用

普通最小二乘法（OLS）估计该方程，并通过组内变换去除固定效应项[1]。这里的关键参数是 $\beta_2$，它反映了有害热暴露每边际增加时，作物产量对数的变化幅度。这等同于作物产量相对于有害热暴露的弹性。请注意， $\beta_2$ 在短期内预计为负，因为这一区间的温度冲击会损害作物生长（Schlenker 和 Michael J. Roberts，2006年；Schlenker 和 Michael J Roberts，2009年）。因此，长期内抵消的短期损害占比估计值为（$\hat{}\beta_2^{SR} - \hat{}\beta_2^{LR}$）/ $\hat{}\beta_2^{SR} = 1 - \hat{}\beta_2^{LR}/\hat{}\beta_2^{SR}$，其中$\hat{}\beta_2^{SR}$（$\hat{}\beta_2^{LR}$）是使用短期（长期）变异得出的估计值。Burke和Emerick（2016年）对这一比率进行自助抽样，但未能拒绝该比率不同于0的零假设。

为使用更灵活的函数形式重现这一比率，我用平均方向导数替代参数估计值。考虑一个更一般的函数形式：

$$y_{it} = a_i + \gamma(X_{it}) + \varepsilon_{it} \tag{2}$$

这里， $X_{it}$ 是一组天气变量， $\gamma$ 是一个一般函数， $a_i$ 是一个加性固定效应项， $\varepsilon_{it}$ 是一个加性误差项。我对 $X_{it}$使用三组不同的天气变量，具体如图1所示。由于 $y_{it}$ 是作物产量的对数，因此作物产量相对于某个变量的弹性，等于 $\gamma$ 相对于该变量的平均方向导数。

因此，方程（1）中的 $\beta_2$ 对应项是 $\gamma$ 相对于 $higher_{it}$的平均方向导数。令 $\theta^{SR}$ （$\theta^{LR}$）表示使用短期（长期）变异计算的这一平均方向导数。随后，我对比率 $1 - \hat{}\theta_2^{LR}/\hat{}\theta_2^{SR}$进行自助抽样，以检验短期热暴露造成的损害是否会在较长期内得到抵消。

使用ˆ年度线性天气变量组时，这一平均方向导数等同于平均偏导数。具体而言，取 $\theta = \mathbb{E}[\partial\hat{}\gamma(X_{it})/\partial higher_{it}]$。使用线性设定时，这一平均导数等同于普通最小二乘法的参数估计值。使用神经网络时，可以在网络训练完成后得到该导数（见第3.3.b节）。当设定包含基函数时（例如，使用多项式基函数的普通最小二乘法或 Lasso），我通过计算基函数梯度与估计系数点积的平均值来得到这一导数。我将在第3.2节介绍这一过程。

当我使用一个 另一组天气变化，同时也有必要考虑 t

[1]Burke和Emerick（2016）建议使用一阶差分来消除长期变化面板数据中的固定效应；我采用组内变换方法，因为在两期面板数据中，这在数值上等同于计算一阶差分。

每个天气变量对生长季热暴露总量的贡献程度。我使用链式法则来计算：

$$\hat{\theta} = \sum_{X_{it} \in X_{it}^{higher}} \frac{\partial \hat{\gamma}(X_{it})}{\partial X_{it}} \frac{\partial X_{it}}{\partial higher_{it}} \tag{3}$$

其中， $X_{it}^{higher}$ 是天气变量的集合，将这些变量加总后可得到超过温度阈值的生长季总热暴露量。我设定 $\partial X_{it}/\partial higher_{it} = X_{it}/higher_{it}$；这体现了一个假设：额外的边际热量将按照经验热暴露分布成比例分配。

这使我能够利用近期的面板变异来衡量对气候变化的适应，同时采用灵活模型刻画高维天气变异。在以下各节中，我将介绍如何使用各种估计技术估计回归函数和平均方向导数。

### 3.2 普通最小二乘法方法

普通最小二乘法（OLS）是估计这种弹性的经典统计方法。我采用线性函数形式的方法（线性普通最小二乘法），以及对多项式函数和交互作用进行基函数变换后采用的方法（多项式普通最小二乘法）。在这两种情况下，我都使用组内变换去除县级固定效应项，然后通过虚拟变量纳入年度固定效应项。

我使用的基函数包含原始数据的交互作用和灵活函数形式。我为所有项指定多项式基函数，并纳入这些多项式展开之间的交互作用。为构建一个便于处理的模型，我将可能的交互作用范围限制为同一时间段内热暴露与降水变量之间的交互作用。例如，在月度灵活设定中，我考虑七月28至29°C温度区间的累积生长度日与七月降水之间的交互作用，以及各项的平方值和这些多项式展开之间的交互作用，但不考虑该变量与其他温度区间的累积生长度日之间的交互作用，也不考虑其与其他月份降水之间的交互作用。年度线性和年度灵活变量集采用三阶多项式，月度灵活变量集采用二阶多项式。随后，我对每个灵活基函数进行缩放，使其均值为零、方差为1。

对于线性普通最小二乘法，我使用恒等基函数集；也就是说， $b(X_{it}) = X_{it}$。对于年度

线性， $X_{it}$ 具有 3 个协变量；对于年度灵活， $X_{it}$ 具有 41 个；对于月度灵活， $X_{it}$ 具有 246 个。对于多项式普通最小二乘法，我使用上述所述的基函数集。对于年度线性， $b(X_{it})$具有 30 个协变量；对于年度灵活， $b(X_{it})$具有 486 个；对于月度灵活， $b(X_{it})$具有 1464 个。

普通最小二乘法假设以下模型能够真实描述这种关系：

$$y_{it} = a_i + b(X_{it})\beta_0 + \varepsilon_{it} \tag{4}$$

与经济学中的常见做法一样，我假设我们拥有相对较短的面板数据，因此无法通过纳入虚拟变量来一致地估计 $a_i$ 。因此，我使用组内变换来消除县级固定效应项：

$$\ddot{y}_{it} = \ddot{b}(X_{it})\beta_0 + \varepsilon_{it} \tag{5}$$

¨其中，双点表示组内变换，即 $\ddot{y}_{it} := y_{it} - \text{mean}(\{y_{it} \forall t\})$ 和 $b(X_{it}) :=¨\text{mean}(\{b(X_{it} \forall t\})$。请注意，为了构造 $b(X_{it})$，在应用基函数变换后，要减去该面板单位中所有观测值的均值。这确保了不同模型中的 $\beta_0$ 是相同的ˆ 参数向量。我使用普通最小二乘法来估计 $\beta$。

然后，我将对基函数偏导数的估计投影到 $\beta_0$ 上，以计算平均方向导数。将基函数的梯度定义为 $b_{higher}$；这是一个 1 乘 $p$ 的字典，列出了每个基函数相对于 $higher_{it}$的导数。真实的平均ˆ ˆ 方向导数为 $\theta_0 = \mathbb{E}[b_{higher}(X_{it})\beta_0]$ ，其估计值为 $\theta = \mathbb{E}[b_{higher}(X_{it})\beta]$。

在主要分析的模型设定中，我将逐年固定效应作为虚拟变量纳入模型。我假设每个时间段都有足够多的观测值，从而能够分别一致地估计这些变量。每个逐年固定效应的导数均为零，因此纳入这些项不会改变我对平均方向导数的估计方式。

### 3.3 机器学习方法

机器学习（ML）方法包括一系列估计技术，可在高维情境下实现一致的函数逼近，即协变量数量相对于观测值数量较大时。这类情境给普通最小二乘法、分箱或核回归等经典统计方法带来了挑战。我使用 Lasso 和神经网络（NNets），不过同样的程序也可用于随机森林、支持向量机或其他方法等算法。我重点关注这两种机器学习算法，因为它们都能让研究者纳入线性固定效应，并在无需数值

微分的情况下评估导数。标准机器学习方法可能会在回归分析中引入偏差；我通过采用 Chernozhukov、Newey 和 Singh（2022a）提出的程序来克服这一偏差。

平均导数是通过对输出变量关于天气输入变量进行机器学习回归计算得出的。用 $\gamma(\cdot;\lambda)$ 表示灵活的机器学习函数，以强调其对超参数 $\lambda$的依赖。超参数是由研究者指定、会影响模型行为的取值，例如 Lasso 中的正则化惩罚或神经网络（NNet）中的网络宽度。机器学习器被估计为回归函数： $\mathbb{E}[\ddot{y}_{it}|\ddot{X}_{it}] = \hat{\ddot{\gamma}}(X_{it};\lambda)$。第 3.3.a 节和第3.3.b节分别详细介绍了各估计程序。令 $m(\gamma, X_{it};\lambda)$ 表示在观测值 $X_{it}$ 上计算的 $\gamma(\cdot;\lambda)$ 的方向导数。

我采用 Chernozhukov、Newey 和 Singh（2022a）提出的双重机器学习（DML）程序，求得平均方向导数的近似去偏估计。我会在第 3.3.c 节中更详细地讨论这一程序。简言之，我估计第二个机器学习器，并据此构造平均方向导数的估计量，使其对第一个或第二个机器学习器估计中的误差均具有稳健性。令 $\alpha(X;\kappa)$ 表示这个第二个机器学习器，以强调其对超参数 $\kappa$ 的依赖。Chernozhukov、Newey 和 Singh（2022a）表明，表达式 $\mathbb{E}[m(\hat{\gamma}, X_{it};\lambda)+\hat{\alpha}(X_{it};\kappa)\hat{\varepsilon}_{it}]$ 是真实平均方向导数的近似无偏估计，其中 $\hat{\varepsilon}_{it}$ 是估计 $y_{it}$ 时得到的残差。我对 Chernozhukov、Newey 和 Singh（2022a）中的双重稳健估计量形式作了修改，以考虑数据的面板结构；详情见第 3.3.c 节。

我使用数据驱动过程来确定超参数的取值。首先，将每个面板单位（一个美国县）划分到用于交叉验证的 $k$ 折之一。将来自同一面板单位的观测值分到同一折，可以降低训练数据与测试数据之间的相关性，因为各县具有不可观测特征，而这些特征可能会影响天气和作物产量的分布。令 $\mathcal{I}_\ell$ 表示第 $\ell$ 折中的索引集合，其中 $\ell \in \{1, 2, \ldots, k\}$。对于 $k$ 折中的每一折，我使用不在第 $\ell$ 折中的数据训练机器学习和双重机器学习算法，并仅在第 $\ell$ 折中的索引上评估算法。令 $\hat{\gamma}_\ell$ 和 $\hat{\alpha}_\ell$ 分别表示使用不在第 $\ell$ 折中的索引集合训练得到的机器学习估计量和双重机器学习估计量。然后，通过在一组候选值组成的网格上搜索，选择使损失函数最小的超参数值。令 $\mathcal{L}_\gamma(\gamma_\ell, \mathcal{I}_\ell;\lambda)$ 表示函数 $\gamma_\ell$ 在超参数 $\lambda$ 下、基于 $\mathcal{I}_\ell$ 中数据计算的均方误差。令 $\mathcal{L}_\alpha(\alpha_\ell, \mathcal{I}_\ell;\kappa)$ 表示函数 $\alpha_\ell$ 在超参数 $\kappa$ 下、基于 $\mathcal{I}_\ell$中数据计算的损失函数；我在附录 A.3 中介绍该损失函数。

选定超参数后，我使用上述定义的相同折，在测试集上评估（去偏）得分。去偏得分的完整估计程序如下。若要使用此程序但不进行去偏校正，请在每个步骤中省略偏差校正项 $\hat{}\alpha_\ell(X_{it}; \hat{}\kappa)(\ddot{y}_{it} - \hat{}\ddot{\gamma}_\ell(X_{it}; \hat{}\lambda))$。

1. 选择 $\hat{}$ t 超参数 $\lambda$，使回归的测试集均方误差最小 :

$$\hat{\lambda} = \arg\min_{\lambda} \sum_{\ell=1}^{k} \mathcal{L}_\gamma(\hat{\gamma}_\ell, \mathcal{I}_\ell; \lambda)$$

2. 选 择超参数 $\hat{}\kappa$，使双重机器学习的测试集损失最小 r:

$$\hat{\kappa} = \arg\min_{\kappa} \sum_{\ell=1}^{k} \mathcal{L}_\alpha(\hat{\alpha}_\ell, \mathcal{I}_\ell; \kappa)$$

3. 使用这些超参数计算去偏得分:

$$\hat{\theta} = \frac{1}{N} \sum_{\ell=1}^{k} \sum_{it \in \mathcal{I}_\ell} m(\hat{\gamma}_\ell, X_{it}; \hat{\lambda}) + \hat{\alpha}_\ell(X_{it}; \hat{\kappa})(\ddot{y}_{it} - \hat{\ddot{\gamma}}_\ell(X_{it}; \hat{\lambda}))$$

4. 在调整面板单元内相关性后，求估计量的渐近方差。令 $\hat{}\ \theta_{\ell;it} := m(\hat{}\gamma_\ell,\ X_{it}; \hat{}\lambda) + \hat{}\alpha_\ell(X_{it}; \hat{}\kappa)(\ddot{y}_{it} - \hat{}\ddot{\gamma}_\ell(X_{it}; \hat{}\lambda))$ 且 $\hat{}\ \theta_{\ell;i} := 1/T \sum_{t=1}^{T} \hat{}\theta_{\ell;it}$。那么

渐近方差为:

$$\hat{V} = \frac{1}{N} \sum_{\ell=1}^{k} \sum_{i \in \mathcal{I}_\ell} \left\{ \sum_{t=1}^{T} (\hat{\theta}_{\ell;it} - \hat{\theta})^2 + 2 \sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} (\hat{\theta}_{\ell;it} - \hat{\theta}_{\ell;i})(\hat{\theta}_{\ell;it'} - \hat{\theta}_{\ell;i}) \right\}$$

在以下小节中，我将介绍如何训练和评估 Lasso 和神经网络估计量，并介绍双重机器学习程序。

### 3.3.a Lasso

最小绝对收缩和选择算子（Lasso）是一种回归方法，它从研究者指定的一组基函数中选出一个稀疏线性模型。该方法通过最小化估计的平方误差，同时利用正则化惩罚更复杂的模型，从而找到这一稀疏线性组合。该方法与多项式普通最小二乘法类似，但由于这种惩罚，估计结果可能存在很大差异。超参数 $\lambda$ 是正则化项的大小，我通过上述交叉拟合程序确定该参数。

与普通最小二乘法一样，我假设存在一个形式如方程（4）所示的真实线性模型，并进行组内变换以消除县级固定效应，从而得到方程（5）。我采用与多项式普通最小二乘法中相同的灵活基函数集合，该集合定义见第3.2节。与多项式普通最小二乘法不同，我假设真实参数向量是稀疏的，并通过求解正则化优化问题来得到估计值。

对于每个十字 ˆ s 折交叉验证中，我通过以下最小化问题得到估计值 $\beta_\ell$ m:

$$\hat{\beta}_\ell = \arg\min_{\beta} \sum_{i \in \mathcal{I}_\ell} \sum_{t=1}^{T} (\ddot{y}_{it} - \ddot{b}(X_{it})\beta)^2 + \lambda|\beta|_1 \tag{6}$$

其中， $|\beta|_1$ 是 $\ell 1$ 惩罚项，即 $\beta$各个分量绝对值之和。为纳入年度固定效应，我加入虚拟变量，并不对这些虚拟变量对应的系数施加正则化惩罚。这反映了这样一个假设：尽管 $\gamma$ 中的系数是稀疏的，但固定效应项中的系数并非如此（Belloni 等人，2016年）。关于 Lasso 程序的更多细节，请参见附录 A.1。

ˆ 估计出 $\beta_\ell$后，我通过将 $\beta_0$的估计值投影到基函数的偏导数上，计算方向导数。这与使用普通最小二乘法计算导数的过程相同，详见第3.2节。

### 3.3.b 神经网络

神经网络（NNet）通过迭代过程学习输入与输出之间的关系：从输入特征所有潜在交互作用的灵活变换所构成的巨大空间中，选出拟合效果最佳的模型。神经网络对我的应用有几个优势。首先，计算整个网络输出相对于每个输入特征的梯度非常直接。其次，正如 Crane-Droesch（2018年）所示，神经网络可以在面板模型中纳入固定效应。第三，估计神经网络不要求研究者指定基函数。

我通过提取网络训练过程中计算出的梯度，来计算神经网络的导数。神经网络是由激活函数（用户指定的变换）加权组合而成，并作用于输入向量。内部参数随机初始化后，模型会经过多次迭代：预测输出变量、根据训练数据计算损失，并更新参数值。我使用均方误差作为损失函数。在训练或评估网络时，算法会计算输出相对于每个输入的导数。这通常用于在训练过程中调整参数值，这一过程称为反向传播。我还使用这些自动计算的导数，来计算预测值相对于每个天气输入的偏导数。

我沿用 Crane-Droesch（2018）的方法，估计对天气输入变量进行非参数化处理并包含加性线性固定效应的模型。训练该网络涉及一个迭代过程，

这个过程可理解为从一个庞大的候选函数空间中选择基函数。完成训练后，我使用标准计量经济学方法处理线性项，将输入的非线性变换视为线性模型中的一个特征。具体而言，我使用组内变换去除个体固定效应，并使用普通最小二乘法求得其他线性项，例如年度固定效应。迭代过程在训练数据集上进行，普通最小二乘法步骤则在测试集上进行。这要求模型架构的顶层对于所有固定效应分量以及非线性网络变换的输出均为线性。有关此类网络及其在估计农业产量时相较于线性模型或完全非参数模型的表现，更多详情请参阅 Crane-Droesch（2018）。附录 A.2 中提供了神经网络架构和训练过程的更多详情。

### 3.3.c 双重机器学习

双重机器学习（DML）是一种用于消除标准机器学习算法偏差的方法。我采用了 Chernozhukov、Newey 和 Singh（2022a）提出的方法，该方法依赖于 Riesz 表示元的统计理论。

在估计平均导数时，存在两个相互关联但彼此正交的问题。第一个是回归函数，第二个是作用于回归函数的导数算子。Chernozhukov、Newey 和 Singh（2022a）展示了如何在不估计回归函数的情况下完成第二步，并表明可以将这两种方法结合起来，构造出原始目标导数的近似去偏估计。

第二个函数可以利用里斯表示定理从数据中估计得到。我将这个函数记为 $\alpha_0$，其估计值记为 $\hat{\alpha}$。里斯表示意味着 $\mathbb{E}[\alpha_0(X_{i,t})\ddot{\gamma}_0(X_{i,t})] = m(\ddot{\gamma}_0, X_{i,t})$。Chernozhukov 、 Newey 和 Singh （ 2022a）展示了如何利用这一性质从数据中估计 $\alpha_0$ 。在面板数据情境中，我使用经过组内变换的函数 $\ddot{\gamma}$；我使用 $m(\ddot{\gamma}_0, X_{i,t}) = m(\gamma_0, X_{i,t})$，因为泛函 $m$ 是方向导数，而均值的导数为 0。我假设 $\alpha_0$ 是基函数集合 $b$的线性函数。因此，真实的 Riesz 表示元为 $\alpha_0(X_{it}) := b(X_{it})\rho_0$ ，其估计值为 $\hat{\alpha}(X_{it}) := b(X_{it})\hat{\rho}$。我根据 Chernozhukov、Newey 和 Singh（2022a）确定的矩条件，使用优化软件包估计 $\hat{\rho}$ 。Klosin 和 Vilgalys（2022）证明，这一估计程序在面板设定中有效。附录 A.3 进一步说明了这一 Riesz 表示元的动机以及

估计程序。

随后使用这一估计值构造以下双重稳健得分：

$$\hat{\theta} = \frac{1}{N}\sum_{\ell=1}^{k}\sum_{it\in\mathcal{I}_\ell} m(\hat{\gamma}_\ell, X_{it}; \hat{\lambda}) + (\ddot{y}_{it} - \hat{\ddot{\gamma}}_\ell(X_{it}))\hat{\alpha}_\ell(X_{it}; \hat{\kappa})$$

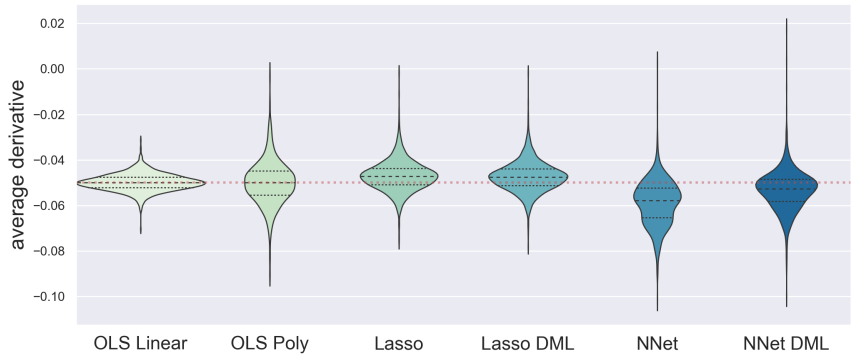
**4 模拟实验**

我开展了一项模拟实验，在改变所用天气变量集合的同时，将该估计程序的表现与普通最小二乘法进行比较。为尽可能贴近实际情境，我在所有模拟试验中都使用天气协变量的经验分布。随后，我采用 Schlenker 和 Michael J Roberts（2009）提出的分段线性函数形式，构造一个模拟生产函数。

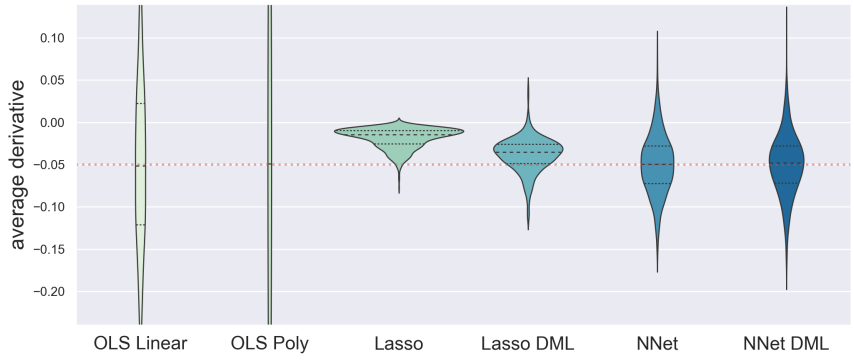
这项模拟实验重点比较在高维变异回归情境下，机器学习、双重机器学习和普通最小二乘法的表现。在所有模拟试验中，普通最小二乘法的设定均正确。如果普通最小二乘法的设定不正确，机器学习和双重机器学习可能会表现得更好，因为它们能够表示更丰富、更灵活的函数形式。CCDDHNR（2018）的模拟实验展示了这一特性。

我对这一估计程序进行了 1,000 次蒙特卡洛模拟试验。在每次试验中，我首先随机抽取一个包含 1,000 个县的县级样本，再从这些县中随机选取一个两年面板。我使用该样本中的天气观测值，从而确保天气变量之间存在符合实际情况的相关性。接着，我根据以下函数生成一个 $y$ 变量：$y_{it} = a_i + \beta_1 lower_{it} + \beta_2 higher_{it} + \beta_3 prec_{it} + \varepsilon_{it}$，其中 $lower$($higher$) 表示生长季中低于（高于）29°C 的累计生长度日总量，$prec$表示生长季的总降水量。这一函数形式与 Schlenker 和 Michael J Roberts（2009）提出的简约形式非常接近。请注意，在这种函数形式下，对于我所考虑的所有天气变量集合，线性普通最小二乘法的设定都是正确的。我设定 $\beta_1 = 0.02$；$\beta_2 = -0.05$；$\beta_3 = 0.001$，并取 $a_i \sim N(1, 1)$ 和 $\varepsilon_{it} \sim N(0, 1)$。

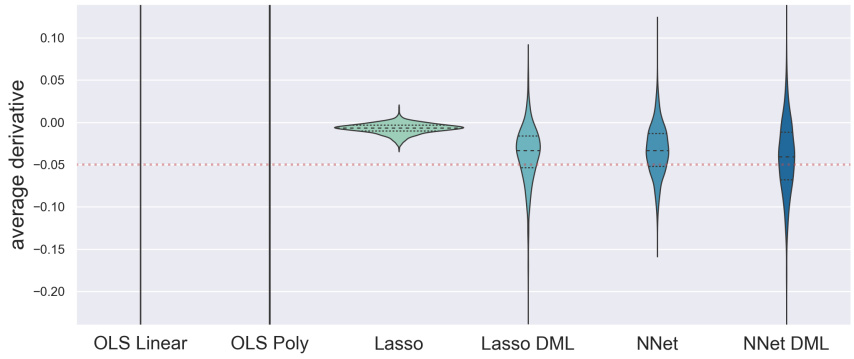
这项模拟实验的结果以可视化形式汇总于图3，并以数值形式汇总于表1。我使用普通最小二乘法、机器学习和双重机器学习，针对年度线性、年度灵活和月度灵活天气变量集合（如图1所示）估计关注参数 $\beta_2$。对于普通最小二乘法，我采用第3.2节所述的线性普通最小二乘法和多项式普通最小二乘法。对于机器学习，我考虑第 3.3.a 节和第3.3.b节所述的 Lasso 和神经网络估计量。对于双重机器学习，我使用第 3.3.c 节中的双重机器学习程序对每项机器学习结果进行校正。我用套索双重机器学习和神经网络双重机器学习来描述经过双重机器学习校正的 Lasso 和神经网络估计值。我采用交叉折叠和样本分割程序




(a) 年度线性




(b) 年度灵活




(c) 月度灵活

图3：基于1,000次蒙特卡罗模拟所得参数估计值分布的小提琴图。各子图标明所使用的天气变量组合，标注方式沿用图1。每个子图中，每种方法均对应一个单独的小提琴图。图中显示了基于模拟试验得到的参数在纵轴各取值处的密度。每个图内的点线表示该分布的中位数和上四分位数/下四分位数。贯穿各图的水平点线表示参数的真实值，即-0.05。

| 天气 | 方法 | 线性普通最小二乘法 | 多项式普通最小二乘法 | Lasso | 套索双重机器学习 | NNet | 神经网络双重机器学习 |
|---|---|---|---|---|---|---|---|
| 年度线性 | $\theta$ | -0.0498 | -0.0499 | -0.047 | -0.0473 | -0.0588 | -0.0533 |
| | | (0.003756) | (0.009589) | (0.006619) | (0.006594) | (0.009762) | (0.008802) |
| | MSE | 0.995 | 0.968 | 1.01 | 1.01 | 1.08 | 1.08 |
| 年度灵活 | $\theta$ | -0.0507 | -0.0469 | -0.0184 | -0.0383 | -0.0498 | -0.0498 |
| | | (0.108) | (0.386) | (0.01192) | (0.02) | (0.03537) | (0.03649) |
| | MSE | 0.954 | 0.511 | 1.03 | 1.03 | 1.1 | 1.1 |
| 月度灵活 | $\theta$ | 2.36 | -0.0317 | -0.00716 | -0.0367 | -0.0327 | -0.0413 |
| | | (58.22) | (0.9779) | (0.006234) | (0.03336) | (0.02993) | (0.04387) |
| | MSE | 0.752 | 7.32e-09 | 1.08 | 1.08 | 1.13 | 1.13 |

表1：估计程序进行1000次蒙特卡洛模拟抽样的结果汇总。所有试验中，平均导数的真实值均为-0.05； $\theta$ 是所有试验中估计得到的平均导数的平均值。每一列代表一种用于回归函数的方法。括号内为标准误，依据自助法数值的分布计算得出。

第 3 节中所述。

使用年度线性天气输入时，普通最小二乘法在所有模型中表现最佳。多项式普通最小二乘法和线性普通最小二乘法的偏差最低，而线性普通最小二乘法的方差最低。机器学习模型的表现尚可，尤其是在进行双重机器学习校正后，不过其方差和偏差均高于普通最小二乘法的结果。两种双重机器学习方法得出的中心估计值与真实弹性相差均在 7% 以内，且 95%置信区间包含真实弹性。

使用年度灵活天气变量时，机器学习估计值的方差远低于普通最小二乘法。线性普通最小二乘法的估计值标准差约为双重机器学习估计值的三倍，而多项式普通最小二乘法估计值的标准差约为其十倍。套索估计值存在显著偏差，不过双重机器学习校正大幅降低了这种偏差。神经网络和神经网络双重机器学习估计值的偏差都很小；两者的中心估计值与真实值相差均在 0.4% 以内，且偏差小于线性普通最小二乘法。两种双重机器学习程序得出的 95%置信区间均包含真实弹性，而未使用双重机器学习的 Lasso 所得 95%置信区间则不包含。

使用月度灵活变量时，采用双重机器学习的优势更加显著。线性普通最小二乘法和多项式普通最小二乘法的方差极高，且两者的偏差都大于双重机器学习方法。高方差尤其构成限制，因为采用这些方法得到的置信区间无法提供有用信息。如图3c所示，在纵轴范围为真实目标弹性绝对值 8 倍的图中，自助法估计值的四分位距几乎无法看见。线性普通最小二乘法估计的均值为 2.359，而真实值为 -0.05。多项式普通最小二乘法的偏差为 .018，

高于套索双重机器学习（0.012）或神经网络双重机器学习（0.0087）的偏差。

请注意，虽然均方误差（MSE）有助于判断哪种模型更可取，但直接比较均方误差并不能选出偏差最低的模型。机器学习模型报告的均方误差高于普通最小二乘法，是因为机器学习报告的是模型在测试集上的均方误差。当机器学习模型的均方误差较低时，说明机器学习拟合效果更好，因为该模型对未见数据的外推表现良好。当普通最小二乘法的均方误差较低时，则可能意味着过拟合。在所有模拟试验中，多项式普通最小二乘法的均方误差都最低，但在年度灵活和月度灵活两种情况下，它对数据过拟合。在数据生成过程中，加性误差项的方差为 1。多项式普通最小二乘法的均方误差远低于数据生成过程中的真实平方误差，这表明多项式普通最小二乘法拟合的是噪声，而非数据中所需的模式。采用月度灵活天气项的线性普通最小二乘法，其均方误差显著低于真实平方误差，这表明在这种设置下，线性普通最小二乘法可能对数据过拟合。这说明，不应仅凭均方误差的比较来选择更可取的模型；研究者还可以考察测试集均方误差以及模拟试验中的表现。

这项模拟实验表明，当研究者希望利用高维天气变量集来衡量某种关系时，双重机器学习能够提供比普通最小二乘法偏差更低、方差更小的结果。对于年度灵活或月度灵活天气变量，神经网络双重机器学习在所有模型中偏差最低，且双重机器学习方法的方差显著更低。普通最小二乘法在线性年度情形下表现最佳，这是预期之中的，因为普通最小二乘法的设定正确，且天气变化是低维的。

## 5 结果

我采用上述估计程序，研究美国玉米和大豆生产对有害高温暴露的适应。在使用年度线性天气变量集时（如 Burke 和 Emerick（2016）所采用的），我几乎没有发现适应的证据。然而，在使用更丰富的天气变化集时，我发现证据表明，随着对这些温度的暴露增加，极端高温造成的短期损害中有相当大的份额被抵消。为帮助解释这一差异，我将采用年度灵活温度变化时的普通最小二乘法结果进行可视化，以表明简单的线性模型可能无法准确描述极端高温的长期作用。

我进行了 500 次自助法抽样 第 3 节所述程序的自助法试验，用于衡量 f

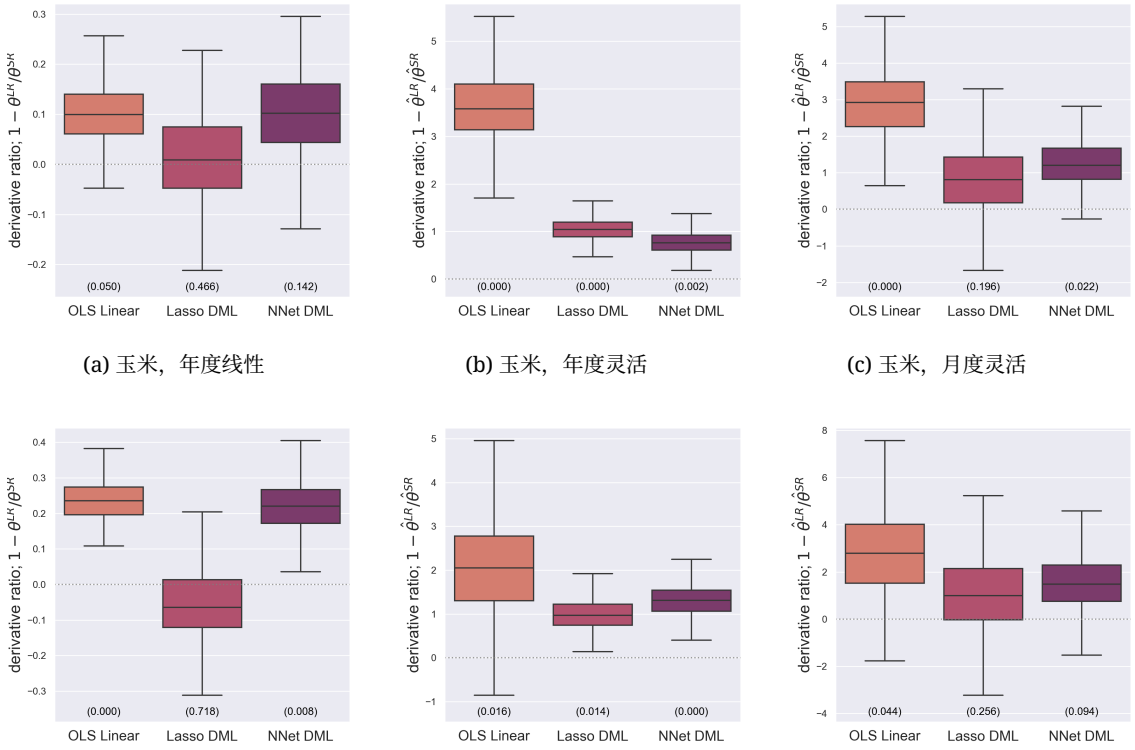
作物产量对极端高温的弹性。如第 2 节所述，我研究了西经100度子午线以东各县1990—2019年的玉米和大豆生产。我使用完整数据构成的面板数据来捕捉短期变异，并通过比较1990—1999年与2010—2019年的平均产量和天气来捕捉长期变化。每次自助法试验中，我随机抽取80%的县作为子样本。正如Kapetanios（2008）所建议的，这种重抽样方案能够处理时间相关性，同时通过在训练数据和测试数据中纳入相同的观测值，避免机器学习方法产生样本内偏差的风险。

我使用图1中描述的三组天气变化，以及第 3 节中描述的所有估计方法来估计弹性。每种方法都使用组内变换去除个体固定效应，并纳入加性的年度固定效应。对于使用一组基函数的估计方法，我对所有温度和降水变量使用多项式展开，并将每个降水变量的多项式展开与每个温度变量的多项式展开进行交互。对于年度线性和年度灵活天气变量，我使用三阶多项式展开；对于月度灵活天气变量，则使用二阶多项式展开。机器学习模型采用五折交叉验证程序。

本节仅报告线性普通最小二乘法、套索双重机器学习和神经网络双重机器学习的结果。模拟结果表明，去偏程序能有效降低朴素机器学习方法造成的偏差，而使用多项式普通最小二乘法得到的估计可能具有极高的方差。未采用去偏程序的机器学习模型结果，以及使用多项式基函数的普通最小二乘法结果，均列于附录B。总体而言，未经偏差校正的机器学习估计与相应的去偏估计相近；而在使用更丰富的天气变化集合时，多项式普通最小二乘法估计的方差高得无法接受。

令人欣慰的是，使用年度线性天气变化得出的结果与 Burke 和 Emerick（2016）的结果相似。表2汇总了自助法试验的结果。长期和短期数据集均发现，额外的极端高温会降低作物产量，对玉米种植和大豆种植的影响幅度相近。这些弹性的幅度在经济意义上显著——估计值为 -0.005 意味着，作物每多经历一天高于 29°C 的温度，作物产量就会下降 0.5%。研究发现与 Burke 和 Emerick（2016）报告的结果相似。他们发现，玉米产量的弹性范围为 $-0.0037$至 $-0.0062$。

与 Burke 和 d Emerick（2016年），采用年度线性天气变化时，我几乎没有发现证据 e




(a) 玉米，年度线性　(b) 玉米，年度灵活　(c) 玉米，月度灵活

(d) 大豆，年度线性　(e) 大豆，年度灵活　(f) 大豆，月度灵活

图4：玉米和大豆生产中比率 $1-\hat{\theta}^{LR}/\hat{\theta}^{SR}$ 的 500 个自助抽样样本箱线图。子图标示了所使用的天气变量组，采用图1中的记法。每个箱线图下方括号内的数字是针对比率大于 0 的单侧检验的 $p$ 值。每个子图都为所用的每种方法分别绘制了一个箱线图。每个箱体内部的线表示中位数，箱体边缘表示分布的上四分位数和下四分位数，从箱体延伸出的须线长度为四分位距的 1.5 倍。

表明短期损害会在长期得到抵消。图4展示了对比率 $1-\hat{\theta}^{LR}/\hat{\theta}^{SR}$ 进行自助抽样所得的结果。如第3.1节所述，$\hat{\theta}^{SR}$ ($\hat{\theta}^{LR}$) 是根据短期（长期）变异估计的弹性。对于玉米种植，我无法拒绝该比率不等于 0 的零假设，无论采用哪种估计方法。我在 $p=0.05$ 水平进行检验，并通过邦费罗尼校正来考虑进行三项假设检验的情况。对于大豆种植，我得到的结果不一：采用线性普通最小二乘法和神经网络双重机器学习时，我在该水平上拒绝零假设；采用套索双重机器学习时，则无法拒绝零假设。对于这些大豆估计值，采用线性普通最小二乘法时，95%置信区间为 $[0.1278, 0.3488]$；采用神经网络双重机器学习时，95%置信区间为 $[0.07153, 0.3649]$。置信区间根据自助法试验中样本均值和标准差计算得出。

使用年度灵活天气变量，我发现了适应的显著证据。采用短期变异的面板数据显示，极端高温暴露的边际

| | | | 长期 | | | 短期 | | |
|---|---|---|---|---|---|---|---|---|
| | | 方法 | 线性普通最小二乘法 | Lasso双重机器学习 | 神经网络双重机器学习 | 普通最小二乘线性回归 | 套索双重机器学习 | 神经网络双重机器学习 |
| Crop | 天气 | | | | | | | |
| Corn | 年度线性 | $\theta$ | -0.00481*** | -0.007*** | -0.00597*** | -0.00535*** | -0.00709*** | -0.00663*** |
| | | | (0.000326) | (0.000568) | (0.000464) | (5.68e-05) | (8.05e-05) | (0.000337) |
| | | MSE | 0.00652 | 0.00648 | 0.00673 | 0.0387 | 0.0363 | 0.0362 |
| | | N | 2850 | 2850 | 2850 | 38633 | 38633 | 38633 |
| | 年度灵活 | $\theta$ | 0.0322*** | 0.000429 | -0.00249 | -0.0125*** | -0.0106*** | -0.0104*** |
| | | | (0.00826) | (0.00234) | (0.00247) | (0.00106) | (0.00115) | (0.000603) |
| | | MSE | 0.00531 | 0.00582 | 0.00603 | 0.0376 | 0.0343 | 0.0352 |
| | | N | 2850 | 2850 | 2850 | 38633 | 38633 | 38633 |
| | 月度灵活 | $\theta$ | 0.0213* | -0.000793 | 0.00148 | -0.0114*** | -0.00505*** | -0.00674*** |
| | | | (0.00874) | (0.00533) | (0.00404) | (0.00101) | (0.000612) | (0.000718) |
| | | MSE | 0.00263 | 0.00544 | 0.00461 | 0.0334 | 0.0303 | 0.0306 |
| | | N | 2850 | 2850 | 2850 | 38633 | 38633 | 38633 |
| Soy | 年度线性 | $\theta$ | -0.00409*** | -0.0055*** | -0.00409*** | -0.00537*** | -0.00525*** | -0.00525*** |
| | | | (0.000307) | (0.000574) | (0.000331) | (5.2e-05) | (8.07e-05) | (0.000252) |
| | | MSE | 0.00579 | 0.00562 | 0.00578 | 0.0308 | 0.029 | 0.0291 |
| | | N | 2422 | 2422 | 2422 | 33799 | 33799 | 33799 |
| | 年度灵活 | $\theta$ | 0.00805 | -7.48e-05 | 0.00239 | -0.00788*** | -0.00698*** | -0.00764*** |
| | | | (0.0079) | (0.00276) | (0.0026) | (0.00109) | (0.000843) | (0.000648) |
| | | MSE | 0.00432 | 0.0056 | 0.00519 | 0.03 | 0.0276 | 0.0282 |
| | | N | 2422 | 2422 | 2422 | 33799 | 33799 | 33799 |
| | 月度灵活 | $\theta$ | 0.00955 | 0.000192 | 0.00247 | -0.00565*** | -0.00362*** | -0.00465*** |
| | | | (0.00914) | (0.0069) | (0.00572) | (0.00104) | (0.000717) | (0.000611) |
| | | MSE | 0.00204 | 0.00451 | 0.00421 | 0.0253 | 0.0236 | 0.0234 |
| | | N | 2422 | 2422 | 2422 | 33799 | 33799 | 33799 |

表2：估计程序的500次自助法抽样结果汇总。 $\theta$ 是自助法试验中平均导数的均值。标准误以括号标出，根据自助法数值的分布计算得出。样本量是每次模拟试验的平均样本数。星号表示在 $p = 0.05$(*) 0.01(**)，以及 0.001(***) 水平上具有显著性，依据的是自助法试验均值和标准误计算得出的Z分数。

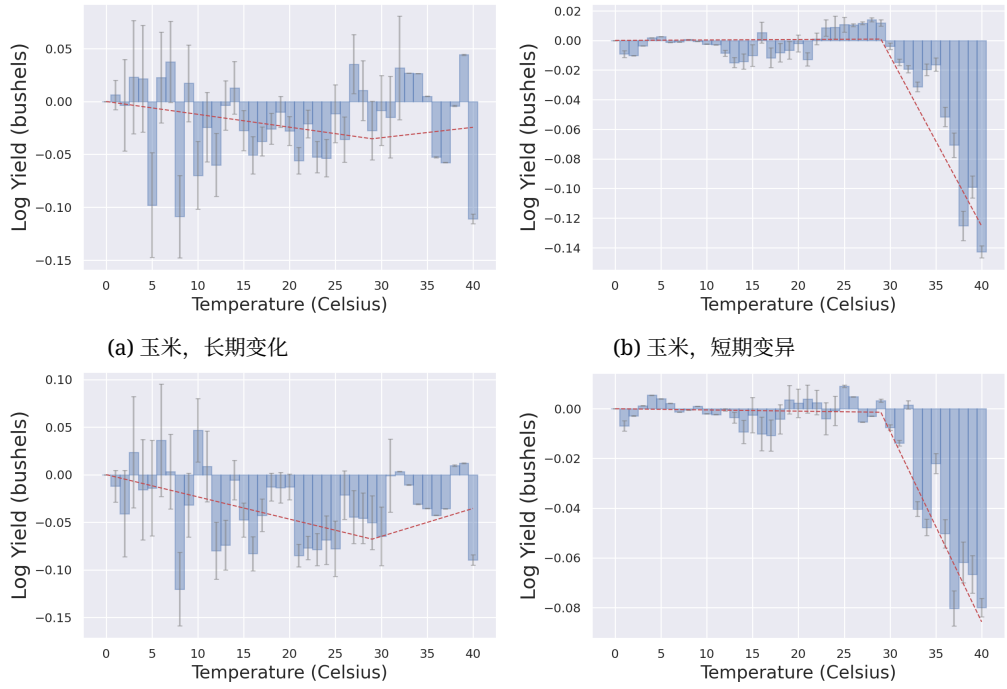
增加会造成在统计和经济意义上均显著的损害。玉米的中心估计值范围为 $-0.01057$ 至 $-0.01248$ ，大豆为 $-0.006983$ 至 $-0.00788$ 。然而，对于两种作物，在所有设定下，我都无法拒绝长期弹性不为零这一零假设。图4证实了这一发现。对于两种作物，使用所有估计方法时，我都拒绝了短期损害中被长期抵消的比例等于零这一零假设。

使用这组天气变量时，我偏好的估计量是神经网络双重机器学习。模拟结果表明，使用年度灵活天气变量时，该估计量的表现最佳。均方误差也表明，神经网络双重方法在此情形下表现良好。对于玉米和大豆，神经网络双重机器学习在长期回归中的均方误差均有所下降。由于神经网络双重机器学习的均方误差是测试集均方误差，相较于使用年度线性天气变量时的均方误差有所改善，说明模型对真实函数形式的拟合有所改进，而非出现过拟合。使用这一偏好的估计结果，我发现该比率的95%置信区间为：玉米为 $[0.3023, 1.222]$ ，大豆为 $[0.6331, 2.001]$ 。

使用月度灵活天气变量组得到的结果与使用年度灵活天气变量组的结果相似，但方差更大。方差较大并不意外，因为模拟结果表明，使用这组天气变量时，双重机器学习和普通最小二乘法的偏差和方差都更高。使用这组天气变量考察短期变异时，热暴露边际增加会导致产量出现具有经济意义且统计显著的下降。考察长期变化时，没有证据表明热暴露边际增加会导致产量显著下降。令人欣慰的是，这一证据与使用年度灵活天气变化得到的结果一致。鉴于这些估计值的方差较大，除使用双重机器学习方法外，我无法拒绝“这些弹性的比率不同于零”这一零假设。使用线性普通最小二乘法时，我发现热暴露边际增加会显著提高作物产量。

这些结果表明，在灵活地对温度与作物产量关系进行建模时，长期有害高温暴露的边际变化并不会导致产量出现显著的平均下降。我没有估计出长期暴露于29°C以上温度会造成显著减产，因为其他变异更能解释长期平均产量的差异。短期变异的情况则不同：灵活建模证实，暴露于29°C以上温度的边际增加与作物产量显著下降相关。这一发现具有新颖性，并表明人们已显著适应有害高温暴露。

为更好地解释这一结果，我考察了年度灵活天气变量各个天气区间的温度系数估计值。我使用神经网络双重机器学习程序估计这些图。附录C列出了使用其他估计程序的结果。由于使用双重机器学习程序估计每个平均导数的计算成本较高，我没有针对每个温度系数重复运行双重机器学习程序；结果相似。图5展示了这些回归结果。图5b和图5d展示了使用短期变异进行估计的结果。这些结果与 Schlenker 和 Michael J Roberts（2009年）的结果相似；他们指出，分段线性拟合（图中的虚线）可以解释灵活建模方法（图中的柱状图）所揭示的大部分变异。在这里，分段线性模型似乎拟合良好：在约29°C之前，额外的热暴露通常与产量增加相关；超过这一温度后，额外的热暴露则与产量下降相关。




(a) 玉米，长期变化

(b) 玉米，短期变异

(c) 大豆，长期变化

(d) 大豆，短期变异

图5：比较温度与产量之间的关系在短期变异和长期变化下的差异。每个图均汇总了使用神经网络双重机器学习、基于年度灵活天气变量集得到的系数估计值。柱状图显示在给定温度水平下暴露于一个度日所对应的产量预期变化。误差棒表示各系数估计值上下一个标准差的范围。折线表示对柱状图进行的最佳分段线性拟合，其中一段为0–29°C，另一段为29–40°C。

长期变化的结果并未呈现这一模式。图5a和图5c展示了使用长期变化进行估计的结果。系数没有呈现明显的分段线性模式。这表明，在长期模型中，描述温度暴露与作物产量之间关系的分段线性模型可能并不适用。在附录C中，我按不同时间段重新绘制了这张图，起始时间段为1950—1979年。这些图证实，与长期变化相比，分段线性模型在使用短期变异时拟合得更好。年度线性分析的结果显示，长期有害高温暴露与作物产量下降相关；而这些结果表明，其他温度变化更能解释作物产量的长期变化。这说明，灵活建模对于理解有害高温暴露边际增加的作用十分重要。

## 6 讨论

我介绍了 aDML 程序，该程序能够在高维情境下估计平均方向导数，并通过一项模拟

实验展示了该程序相较于普通最小二乘估计的优势。结果表明，与普通最小二乘法相比，该程序能够以更低的偏差和方差估计平均方向导数。模拟结果还表明，与能够获得底层函数真实且简约的模型相比，我的方法可靠性较低。在研究者面对高维天气变异，且对真实函数形式没有强先验认知的情境下，可以使用这种双重机器学习方法来估计弹性以及对某一天气特征的适应程度。

将这一估计量应用于包含丰富温度变化数据的美国玉米和大豆产量面板数据后，我得出结论：人们已经显著适应了有害高温暴露。使用这种灵活的方法对短期损害进行面板分析发现，温度超过29°C后，热暴露的边际增加会对玉米和大豆产量造成损害。然而，通过构建一个记录平均变化的数据集来捕捉长期变化，我无法拒绝以下零假设：长期有害高温暴露的边际增加与长期作物产量无关。使用一组年度灵活的天气变量进行普通最小二乘法估计，也支持这一发现。我证明，长期暴露于29°C以上的温度并不像短期暴露那样与产量下降存在明确关联。这意味着在这种情境下，人们已经显著适应了气候变化，因为短期损害在长期内得到了显著抵消。

这种方法无法揭示这种适应可能采取何种形式。农民有许多可能的适应机制，例如投资灌溉、购买改良种子或调整种植时间。为了研究这些机制的长期有效性，或了解能否将其应用于其他农业情境，有必要明确哪些机制可能负责在长期内抵消损害。还需要开展更多研究，以理解这些损害是如何得到抵消的。

这一结果与先前文献之间看似矛盾，这表明，尽管人们已适应有害高温暴露的边际增加，但对气候变化的其他某些有害特征的适应有限。Burke和Emerick（2016）等先前研究并未发现有力证据表明人们已显著适应有害高温暴露。我的分析表明，尽管有害高温暴露的长期增加与作物产量下降相关，但其他温度水平下热暴露的变异更能解释这一模式。这表明，仅使用有益和有害热暴露进行长期估计会产生遗漏变量偏误。尽管先前的结果表明，人们对温度分布的某些变化适应有限

但我的详细分析发现，人们已经显著适应了有害高温暴露。

**参考文献**

ApS, MOSEK（2021）。*MOSEK Python 优化器 API 9.3.6*。Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone 和 Joseph S. Shapiro（2016）。“适应气候变化：20 世纪美国温度与死亡率关系的显著减弱”。载于：《政治经济学杂志》124.1，第 105–159 页。Barron, Jonathan T.（2017）。“连续可微指数线性单元”。载于：*arXiv* 预印本。arXiv: 1704.07483。Belloni, Alexandre, Victor Chernozhukov, Christian Hansen 和 Damian Kozbur（2016）。“高维面板模型中的推断及其在枪支管制中的应用”。载于：《商业与经济统计杂志》34.4，第 590–605 页。Burke, Marshall 和 Kyle Emerick（2016）。“适应气候变化：来自美国农业的证据”。载于：《美国经济杂志：经济政策》8.3，第 106–140 页。Burlig, Fiona, Christopher Knittel, David Rapson, Mar Reguant 和 Catherine Wolfram（2020）。“从学校获取能源效率方面的机器学习信息”。载于：《环境与资源经济学家协会杂志》7.6，第 1181–1217 页。Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney K Newey 和 James Robins（2018）。“针对处理效应和结构参数的双重/去偏机器学习”。载于：《计量经济学杂志》21.1，第 C1–C68 页。Chernozhukov, Victor, Whitney K Newey 和 Rahul Singh（2022a）。“因果效应和结构效应的自动去偏机器学习”。载于：《计量经济学》90.3，第 967–1027 页。—（2022b）。“利用正则化里斯表示元对全局和局部参数进行去偏机器学习”。载于：《计量经济学杂志》。Costinot, Arnaud, Dave Donaldson 和 Cory Smith （ 2016 ）。“ 比较优势的演变与气候变化对农产品市场的影响：来自全球 170 万块田地的证据”。载于：《政治经济学杂志》124.1。Crane-Droesch, Andrew（2018）。“用于农业作物产量预测和气候变化影响评估的机器学习方法”。载于：环境研究快报13，第 114003 页。Dell, Melissa, Benjamin F. Jones 和 Benjamin A. Olken（2014）。“我们从天气中学到了什么？新的气候经济学文献”。载于：《经济文献杂志》52.3，第 740–798 页。Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor 和 Julian Reif（2019）。“空气污染造成的死亡率与医疗费用：来自风向变化的证据”。载于：《美国经济评论》109.12，第 4178–4219 页。Diamond, Steven 和 Stephen Boyd（2016）。“CVXPY：一种嵌入 Python 的凸优化建模语言”。载于：《机器学习研究杂志》17.83，第 1–5 页。Hagerty, Nick（2021）。“灌溉农业对地表水短缺的适应”。载于：工作论文。Hornbeck, Richard（2012）。“美国沙尘暴的持久影响：对环境灾难的短期和长期调整”。载于：《美国经济评论》102.4，第 1477–1507 页。Hornik, Kurt, Maxwell Stinchcombe 和 Halbert White（1990）。“利用多层前馈神经网络对未知映射及其导数进行通用逼近”。载于：神经网络3.5，第 551–560 页。Hsiang, Solomon（2016）。“气候计量经济学”。载于：《资源经济学年度评论》8.1，第 43–75 页。Kapetanios, George（2008）。“针对包含大量截面单位的面板数据集的自助法程序”。载于：《计量经济学杂志》11.2，第 377–395 页。

Kingma, Diederik P. 和 Jimmy Lei Ba（2015）。“Adam：一种随机优化方法”。载于：第三届国际学习表征会议（*ICLR 2015*）会议论文集。国际学习表征会议，ICLR。arXiv: 1412.6980。Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan 和 Ziad Obermeyer（2015）。“预测政策问题”。载于：《美国经济评论》。第 105 卷第 5 期，第 491–495 页。Klosin, Sylvia 和 Max Vilgalys（2022）。“利用机器学习估计面板数据中的连续处理效应：一项农业应用”。载于：*arXiv* 预印本 *arXiv:2207.08789*。Kurukulasuriya, Pradeep, Namrata Kala 和 Robert Mendelsohn（2011）。“适应与气候变化影响：非洲灌溉和农场收入的结构性李嘉图模型”。载于：《气候变化经济学》2.02，第 149–174 页。Lemoine, Derek（2018）。“根据天气变异估计气候变化的后果”

载于：手稿。美国国家经济研究局 w25008。Liakos, Konstantinos G., Patrizia Busato, Dimitrios Moshou, Simon Pearson 和 Dionysis Bochtis（2018）。“农业中的机器学习：综述”。载于：18.8，第 2674 页。Mendelsohn, Robert, William D. Nordhaus 和 Daigee Shaw（1994）。“全球变暖对农业的影响：李嘉图分析”。载于 《美国经济评论》84，第 753–771 页。Mullainathan, Sendhil 和 Jann Spiess（2017）。“机器学习：一种应用计量经济学方法”。载于《经济展望杂志》。第 31 卷。第 2 期，第 87–106 页。Olmstead, Alan L 和 Paul W Rhode（2011）。“使北美小麦生产适应气候挑战，1839—2009年。”载于 《美国国家科学院院刊》108.2，第 480–485 页。Ortiz-Bobea, Ariel（2013）。“天气在农业生产中真的具有可加性吗？对气候变化影响的启示”。载于 *SSRN*电子期刊 1，第 3–41 页。—（2020）。“非农影响因素在李嘉图气候变化影响估计中的作用：以美国农业为例”。载于 《美国农业经济学杂志》102.3，第 934–959 页。Park, J. 和 I. W. Sandberg（1991）。“使用径向基函数网络实现通用逼近”。载于 《神经计算》3.2，第 246–257 页。Schlenker, Wolfram 和 Michael J Roberts（2009）。“非线性温度效应表明气候变化将严重损害美国作物产量。”载于 《美国国家科学院院刊》106.37，第 15594–8 页。—（2006）。“天气对玉米产量的非线性影响*”。载于 《农业经济学评论》28.3，第 391–398 页。Semenova, Vira 和 Victor Chernozhukov（2021）。“条件平均处理效应及其他因果函数的去偏机器学习”。载于《计量经济学杂志》24.2，第 264–289 页。arXiv：1702.06240。Shukla, P R 等（2019）。“政府间气候变化专门委员会，2019年：《气候变化与土地》：政府间气候变化专门委员会关于气候变化、荒漠化、土地退化、可持续土地管理、粮食安全以及陆地生态系统温室气体通量的特别报告”。Stetter, Christian, Philipp Mennig 和 Johannes Sauer（2022）。“运用机器学习识别欧盟农业环境计划的异质性影响：案例研究”。载于：《欧洲农业经济学评论》。Sutch, Richard（2011）。“1936年玉米带干旱对美国农民采用杂交玉米的影响”。载于 《农业研究与创新经济学》。芝加哥大学出版社，第 195–223 页。Varian, Hal R（2014）。“大数据：计量经济学的新技巧”。载于 《经济展望杂志》28.2，第 3–28 页。

## 附录A 机器学习估计细节

本附录进一步详细介绍机器学习估计程序。

### 附录 A.1 Lasso

我在由 15 个值组成的网格上调整 Lasso 超参数（正则化惩罚），这些值在对数空间中从 $10^{-10}$ 到 $10^0$均匀分布。交叉折叠验证通常会选取网格内部的超参数，这表明该范围是合适的。

回顾一下优化问题的形式：

$$\hat{\beta}_\ell = \arg\min_\beta \sum_{i \in \mathcal{I}_\ell} \sum_{t=1}^{T} w_i (\ddot{y}_{it} - \ddot{b}(X_{it})\beta)^2 + \lambda |\beta|_1 \tag{7}$$

ˆ 由于这是一个（弱）凸优化问题，我使用优化软件包 CVXPY（Diamond and Boyd, 2016年）和优化器 Mosek（ApS，2021年）求解 $\beta_\ell$。在优化软件包的实现中，优化器偶尔会无法找到唯一最优解。在这种情况下，我会引入一个为 $10^{-20}$ 的 $\ell 2$，以找到唯一解。使用优化软件包可确保 Lasso 程序收敛，但与坐标下降相比，该方法需要更多内存，并且若无法收敛则会报错。我发现模型在大约 5% 的模拟试验中无法收敛。

在主要设定中，我纳入了年度固定效应项，但不对这些项进行正则化。我通过有选择地将正则化因子应用于 $\beta$ 不属于这些年度固定效应项的项来实现这一点。在测试集上评估估计量之前，我会重新调整额外的固定效应；这不会影响梯度估计，但会降低估计量的均方误差。

我使用优化软件包有两个原因。求解器未返回错误时，可以保证所得解是最优的；而随机梯度下降或坐标下降等迭代方法则无法保证这一点。有些软件包为迭代方法提供了这项功能，但据我所知，Python 中尚无此类实现。

### 附录 A.2 神经网络

我采用相对简单的网络配置，并在每次模拟试验中改变神经网络（NNet）的宽度。使用神经网络时，研究者有三个关键的自由度：深度（

层数）、宽度（每个隐藏层的节点数）以及激活函数（应用于每层输出的非线性函数）。我使用仅含一个隐藏层的网络，这是神经网络能够逼近任意函数所需的最小深度（Hornik、Stinchcombe 和 White，1990年；Park 和 Sandberg，1991年）。在每次模拟试验中，我使用交叉折验证来选择网络的宽度。为确保网络可微，我使用连续可微指数线性单元（Barron，2017年）作为网络的激活函数。

我从 2 到 256 的网格空间中选择超参数，各取值在对数尺度上均匀分布。交叉折叠程序通常会选择该区间内部的宽度，这表明这个范围是合适的。训练网络时，我使用批量归一化、Adam优化器（Kingma 和 Ba，2015）、0.01 的学习率以及 1000 个训练轮次。在测试集上评估之前，我会重新调整参数组件（县级和年度固定效应项）。这不会影响梯度估计，但会改善估计量的均方误差。

模拟结果表明，这种简单的网络配置表现良好，不过更复杂的网络配置也可能表现更佳。考虑到探索各种潜在网络配置的计算成本，我将此类探索留待未来研究。

**附录 A.3 自动双重机器学习**

双重机器学习（DML）程序的理论基础来自里斯表示定理。该定理指出，对于线性泛函[2] $m$，存在一个函数 $\alpha_0$，使得对于任意函数 $h$:

$$\mathbb{E}[m(h, X_i)] = \mathbb{E}[h(X_i)\alpha_0(X_i)] \tag{8}$$

由于这一关系对任意函数 $h$ 都成立，因此可以设定已知函数，使目标 $m(h,\ X_i)$ 已知，并根据 $X$ 的经验分布估计 $\hat{\alpha}$。其他方法则要求研究者根据函数 $m$ 求解 $\alpha$ 的函数形式，并且通常还要估计密度或密度导数。我采用这一方法，因为它不要求研究者指定 $\alpha$ 的形式，而是从数据中学习 $\hat{\alpha}$。

我遵循 Chernozhukov、Newey 和 Singh（2022a）的方法，使用 Lasso 和基函数$\alpha$ 来估计 $\hat{\alpha}$

[2] 泛函是对函数的标量概括。在这种情形下，泛函是方向导数。

用于估计回归函数的函数。目标是找到使 $\hat{}\alpha$ 与 $\alpha_0$之间的均方误差最小的估计量，其中 $\alpha_0$ 是真实的 Riesz 表示元：

$$\hat{\alpha} = \arg\min_{\alpha} \mathbb{E}[(\alpha_0(X_{i,t}) - \alpha(X_{i,t}))^2]$$

¨采用 Lasso 函数形式，我有 $\hat{}\alpha = b(X_{i,t})\hat{}\rho$。我通过求解以下正则化问题来找到 $\hat{}\rho$：

$$\hat{\rho} = \arg\min_{\rho} \mathbb{E}\left[(\alpha_0(X_{i,t}) - \ddot{b}(X_{i,t})\rho)^2\right] + \kappa|\rho|_1$$

展开多项式并应用 Riesz 表示元的定义后，该表达式可简化为：

$$\hat{\rho} = \arg\min_{\rho} -2\mathbb{E}[m(\ddot{b}, X_{i,t})]\rho + \rho'\mathbb{E}[\ddot{b}(X_{i,t})'\ddot{b}(X_{i,t})]\rho + \kappa|\rho|_1$$

Chernozhukov、Newey 和 Singh（2022a）还加入了一种迭代程序，按各分量方差的倒数缩放其正则化项；我未作任何修改，直接采用他们建议的程序。

由于这是一个（弱）凸优化问题，对于给定的正则化 $\kappa$值，我使用优化软件包 Mosek（ApS, 2021）求参数向量 $\hat{}\rho$ 的最优值。Klosin 和 Vilgalys（2022）展示了模拟结果，将这种优化方法与 Chernozhukov、Newey 和 Singh（2022a）的迭代方法进行了比较；他们发现，在估计真实平均导数时，优化方法得到的均方误差更低。

我通过第 3.3 节所述的交叉折叠法，最小化上述损失函数，以此选取超参数 $\kappa$ 的值。也就是说：

$$\mathcal{L}_{\alpha}(\hat{\alpha}_{\ell}, X_{i,t}; \kappa) = -2m(\ddot{b}, X_{i,t})\hat{\rho}_{\ell} + \hat{\rho}_{\ell}'\ddot{b}(X_{i,t})'\ddot{b}(X_{i,t})\hat{\rho}_{\ell}$$

我在 Chernozhukov、Newey 和 Singh（2022a）建议的取值网格上搜索超参数： $\kappa = c(N - N_{\ell})^{-1/2}\Phi^{-1}(1 - .05/p)$，其中 $\Phi^{-1}$ 是标准正态密度函数的倒数， $N_{\ell}$ 是折中观测值集合的大小，适用于 $c \in \{5/4, 1, 3/4, 5/8, 9/16, 1/2\}$

$\ell$。

**附录B 补充结果**

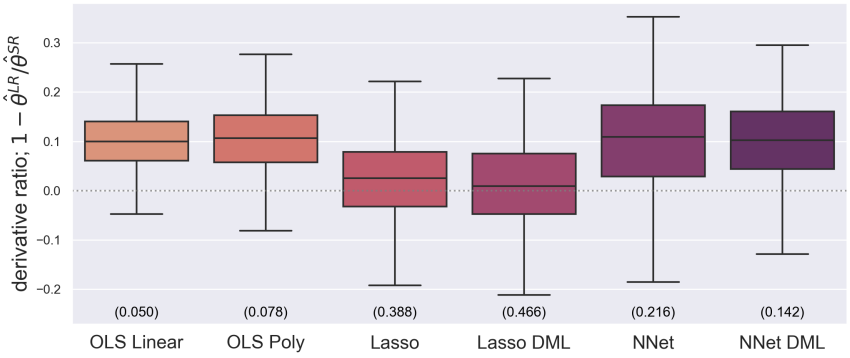
本附录包含第 5 节所述自助法试验的更多结果。表 3 展示了玉米产量估计的完整结果，表 4 展示了大豆产量估计的完整结果。正如预期，多项式普通最小二乘法的结果比其他方法具有更高的方差，尤其是在高维情形下。未进行偏差校正的机器学习方法方差很低，但相较于去偏估计值可能存在偏差。应用偏差校正后，方差通常会增加。图 6 展示了所有模型对玉米产量所得比率 $1 - \hat{\theta}^{LR}/\hat{\theta}^{SR}$ 的自助法取值，图 7 展示了该比率对大豆产量所得的自助法取值。请注意，多项式普通最小二乘法估计值的方差占主导地位。这些标准误很大且影响显著，尤其是在图 6c 和图 7c 中。

| 变异 | 天气 | 方法 | 线性普通最小二乘法 | 多项式普通最小二乘法 | Lasso | 套索双重机器学习 | NNet | 神经网络双重机器学习 |
|---|---|---|---|---|---|---|---|---|
| 长期 | 年度线性 | $\theta$ | -0.00481*** | -0.0063*** | -0.00692*** | -0.007*** | -0.00583*** | -0.00597*** |
| | | | (0.000326) | (0.00052) | (0.000547) | (0.000568) | (0.000485) | (0.000464) |
| | | MSE | 0.00652 | 0.00584 | 0.00648 | 0.00648 | 0.00673 | 0.00673 |
| | | N | 2850 | 2850 | 2850 | 2850 | 2850 | 2850 |
| | 年度灵活 | $\theta$ | 0.0322*** | 0.014 | -0.00135 | 0.000429 | 0.00147 | -0.00249 |
| | | | (0.00826) | (0.0294) | (0.000882) | (0.00234) | (0.00155) | (0.00247) |
| | | MSE | 0.00531 | 0.00159 | 0.00582 | 0.00582 | 0.00603 | 0.00603 |
| | | N | 2850 | 2850 | 2850 | 2850 | 2850 | 2850 |
| | 月度灵活 | $\theta$ | 0.0213* | -5.83e+04 | -0.00184*** | -0.000793 | 0.00356** | 0.00148 |
| | | | (0.00874) | (1.25e+06) | (0.00042) | (0.00533) | (0.00126) | (0.00404) |
| | | MSE | 0.00263 | 3.72e-06 | 0.00544 | 0.00544 | 0.00461 | 0.00461 |
| | | N | 2850 | 2850 | 2850 | 2850 | 2850 | 2850 |
| 短期 | 年度线性 | $\theta$ | -0.00535*** | -0.00704*** | -0.00708*** | -0.00709*** | -0.00645*** | -0.00663*** |
| | | | (5.68e-05) | (7.64e-05) | (7.96e-05) | (8.05e-05) | (0.000595) | (0.000337) |
| | | MSE | 0.0387 | 0.0363 | 0.0363 | 0.0363 | 0.0362 | 0.0362 |
| | | N | 38633 | 38633 | 38633 | 38633 | 38633 | 38633 |
| | 年度灵活 | $\theta$ | -0.0125*** | -0.0115*** | -0.0106*** | -0.0106*** | -0.00845*** | -0.0104*** |
| | | | (0.00106) | (0.000988) | (0.00112) | (0.00115) | (0.00055) | (0.000603) |
| | | MSE | 0.0376 | 0.0316 | 0.0343 | 0.0343 | 0.0352 | 0.0352 |
| | | N | 38633 | 38633 | 38633 | 38633 | 38633 | 38633 |
| | 月度灵活 | $\theta$ | -0.0114*** | -14 | -0.0038*** | -0.00505*** | -0.00562*** | -0.00674*** |
| | | | (0.00101) | (491) | (0.000589) | (0.000612) | (0.000677) | (0.000718) |
| | | MSE | 0.0334 | 0.0258 | 0.0303 | 0.0303 | 0.0306 | 0.0306 |
| | | N | 38633 | 38633 | 38633 | 38633 | 38633 | 38633 |

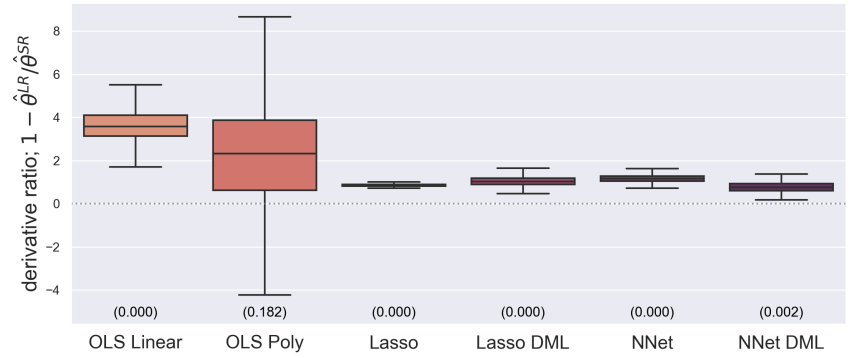
表 3：玉米生产估计程序的 500 次自助法抽样结果摘要。 $\theta$ 是各次自助法试验中平均导数的均值。标准误列于括号内，根据自助法数值的分布计算。样本量 N 是每次模拟试验的平均样本数。星号表示显著性水平为 $p = 0.05$(*) 0.01(**)，以及 0.001(***)，该显著性基于自助法试验均值和标准误计算的 Z 分数。

| 变异 | 天气 | 方法 | 线性普通最小二乘法 | 多项式普通最小二乘法 | Lasso | 套索双重机器学习 | NNet | 神经网络双重机器学习 |
|---|---|---|---|---|---|---|---|---|
| 长期 | 年度线性 | $\theta$ | -0.00409*** | -0.0042*** | -0.00538*** | -0.0055*** | -0.00404*** | -0.00409*** |
| | | | (0.000307) | (0.000485) | (0.000585) | (0.000574) | (0.000374) | (0.000331) |
| | | MSE | 0.00579 | 0.00479 | 0.00562 | 0.00562 | 0.00578 | 0.00578 |
| | | N | 2422 | 2422 | 2422 | 2422 | 2422 | 2422 |
| | 年度灵活 | $\theta$ | 0.00805 | -0.0554 | -0.00169* | -7.48e-05 | 0.00559** | 0.00239 |
| | | | (0.0079) | (0.0328) | (0.000677) | (0.00276) | (0.00204) | (0.0026) |
| | | MSE | 0.00432 | 0.000905 | 0.0056 | 0.0056 | 0.00519 | 0.00519 |
| | | N | 2422 | 2422 | 2422 | 2422 | 2422 | 2422 |
| | 月度灵活 | $\theta$ | 0.00955 | 0.00374 | -0.00182*** | 0.000192 | 0.00518*** | 0.00247 |
| | | | (0.00914) | (0.0779) | (0.000299) | (0.0069) | (0.00128) | (0.00572) |
| | | MSE | 0.00204 | 2.74e-23 | 0.00451 | 0.00451 | 0.00421 | 0.00421 |
| | | N | 2422 | 2422 | 2422 | 2422 | 2422 | 2422 |
| 短期 | 年度线性 | $\theta$ | -0.00537*** | -0.00522*** | -0.00525*** | -0.00525*** | -0.00509*** | -0.00525*** |
| | | | (5.2e-05) | (7.43e-05) | (8.01e-05) | (8.07e-05) | (0.000321) | (0.000252) |
| | | MSE | 0.0308 | 0.029 | 0.029 | 0.029 | 0.0291 | 0.0291 |
| | | N | 33799 | 33799 | 33799 | 33799 | 33799 | 33799 |
| | 年度灵活 | $\theta$ | -0.00788*** | -0.00651*** | -0.00686*** | -0.00698*** | -0.00649*** | -0.00764*** |
| | | | (0.00109) | (0.00102) | (0.000884) | (0.000843) | (0.000877) | (0.000648) |
| | | MSE | 0.03 | 0.0258 | 0.0276 | 0.0276 | 0.0282 | 0.0282 |
| | | N | 33799 | 33799 | 33799 | 33799 | 33799 | 33799 |
| | 月度灵活 | $\theta$ | -0.00565*** | -38.3 | -0.00396*** | -0.00362*** | -0.00444*** | -0.00465*** |
| | | | (0.00104) | (3.4e+03) | (0.000839) | (0.000717) | (0.000544) | (0.000611) |
| | | MSE | 0.0253 | 0.0193 | 0.0236 | 0.0236 | 0.0234 | 0.0234 |
| | | N | 33799 | 33799 | 33799 | 33799 | 33799 | 33799 |

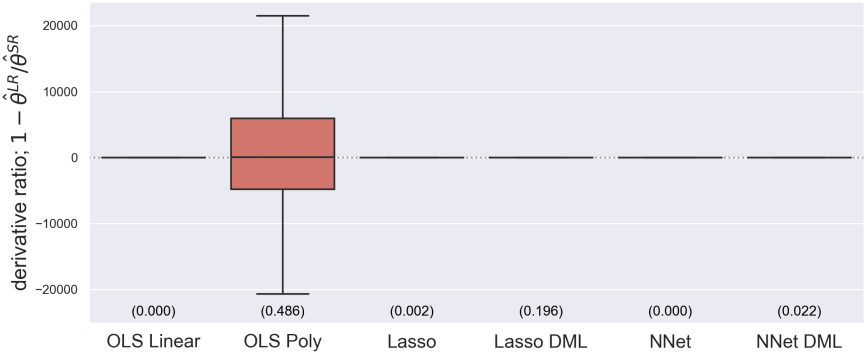
表 4：大豆生产估计程序 500 次自助法抽样的结果摘要。 $\theta$ 是自助法试验中平均导数的均值。标准误置于括号内，根据自助法数值的分布计算。样本量是每次模拟试验的平均样本数。星号表示显著性水平为 $p = 0.05$(*) 0.01(**)，和 0.001(***)，根据自助法试验的均值和标准误计算的 Z分数确定。




(a) 玉米，年度线性

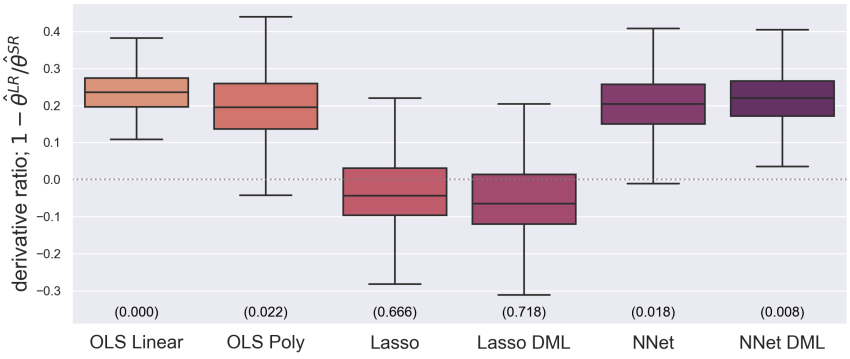



(b) 玉米，年度灵活

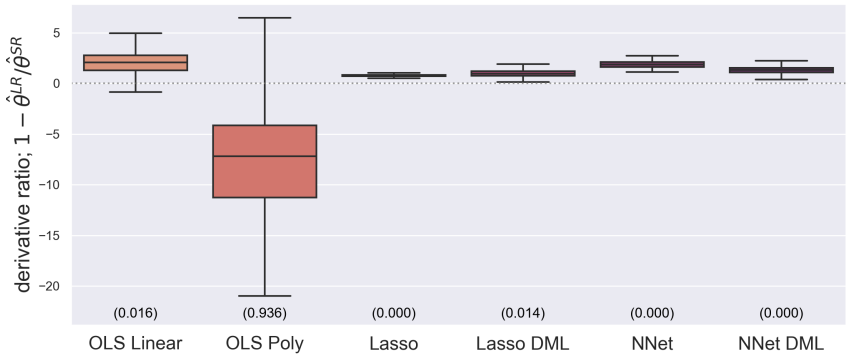



(c) 玉米，月度灵活

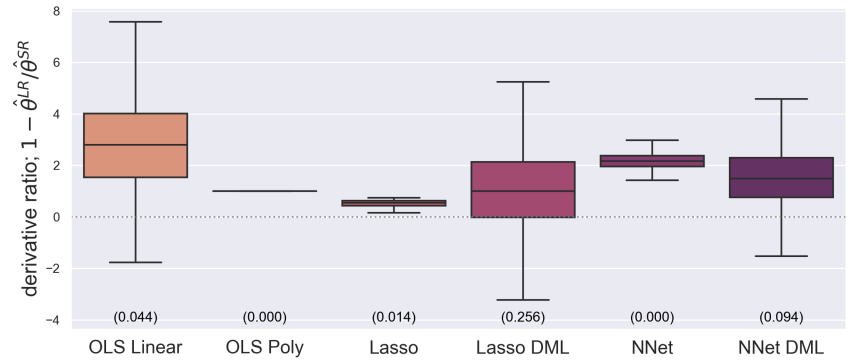
图 6：玉米生产中比率 $1-\hat{\theta}^{LR}/\hat{\theta}^{SR}$ 的 500 个自助抽样样本的箱线图。各子图表示所使用的天气变量集合，采用图1中的记号。底部括号中的数字是单侧检验该比率大于 0 时的 $p$ 值。每个子图都针对所使用的每种方法分别绘制了一个箱线图。每个箱体内的线表示中位数，箱体边缘表示分布的上四分位数和下四分位数，而从每个箱体延伸出的须线长度为四分位距的 1.5 倍。




(a) 大豆，年度线性




(b) 大豆，年度灵活




(c) 大豆，月度灵活

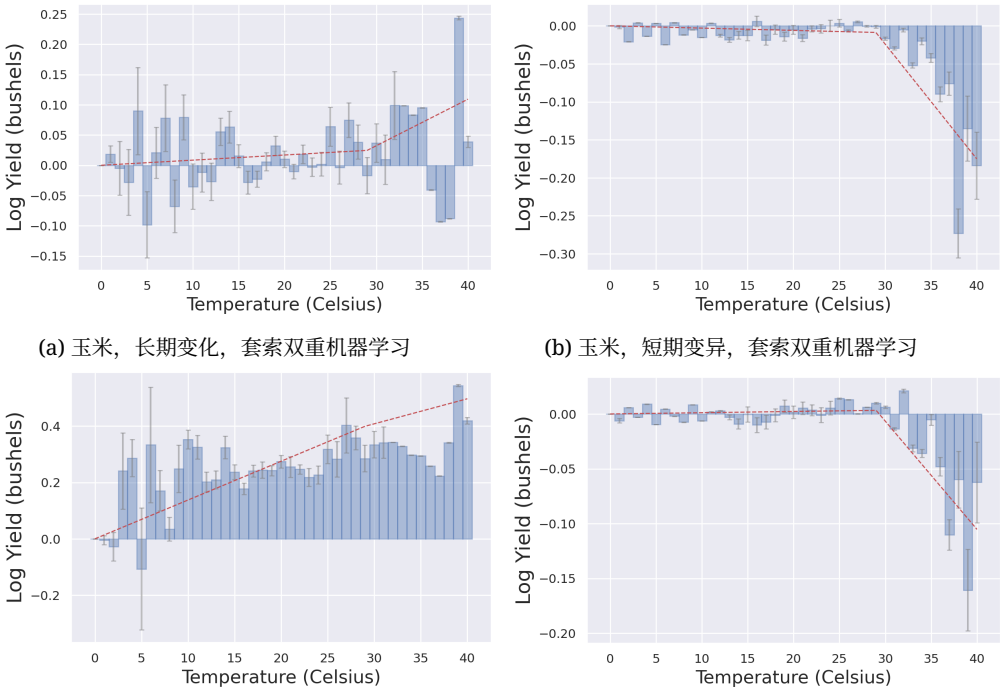
图 7：大豆生产中比率 $1-\hat{\theta}^{LR}/\hat{\theta}^{SR}$ 的 500 个自助抽样样本的箱线图。各子图标示了所使用的天气变量组，记法沿用图1。底部括号中的数字是单侧检验中比率大于 0 的 $p$ 值。每个子图都针对所用的每种方法分别绘制一个箱线图。每个箱体内的线表示中位数，箱体边缘表示分布的上四分位数和下四分位数，从箱体延伸出的须线长度为四分位距的 1.5 倍。

## 附录C 短期和长期年度系数估计值

我提供了更多结果，以展示年度系数区间的短期估计值与长期估计值之间的差异。这些结果证实，长期损害的函数形式并不符合已广为确立的短期损害模式。我还证明，这一结论在不同的建模选择和时间段下均稳健。

### 附录 C.1 使用替代估计量进行比较

以下图表展示了各年度系数区间的短期估计值与长期估计值的比较结果。所有图表均采用主要估计所用的样本，即1990—2019年的玉米和大豆产量。每张图概述了使用年度灵活天气变量集所得结果中的系数估计值，所用方法为套索双重机器学习、多项式普通最小二乘法或线性普通最小二乘法。柱状图显示，在给定温度下，暴露于一个度日所对应的产量预期变化。误差棒表示每个系数估计值上下一个标准差的范围。折线表示对柱状图进行的最佳分段线性拟合，其中一段覆盖0—29°C，另一段覆盖29—40°C。



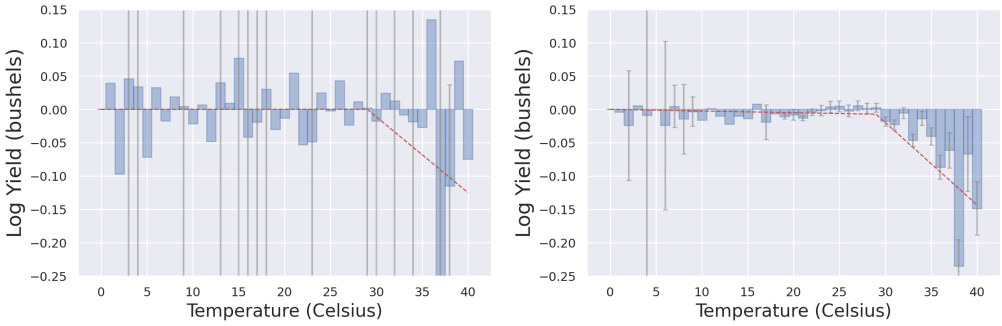

(a) 玉米，长期变化，套索双重机器学习

(b) 玉米，短期变异，套索双重机器学习

(c) 大豆，长期变化，套索双重机器学习

(d) 大豆，短期变异，套索双重机器学习

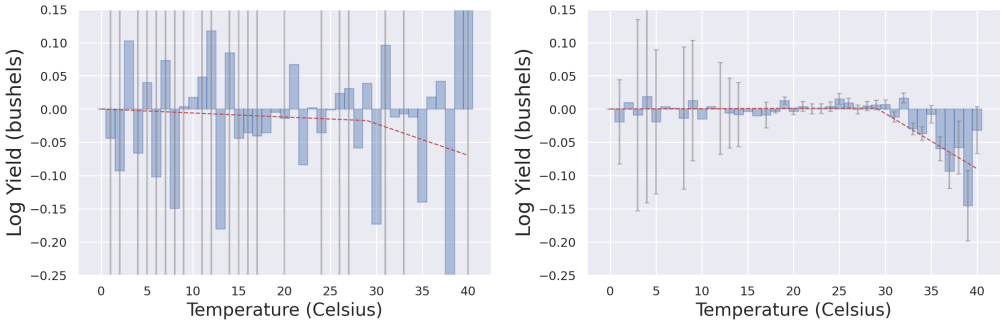
图 8：使用套索双重机器学习比较温度与产量之间在短期和长期变异下的关系。




（a）玉米，长期变化，多项式普通最小二乘法




（b）玉米，短期变异，套索双重机器学习

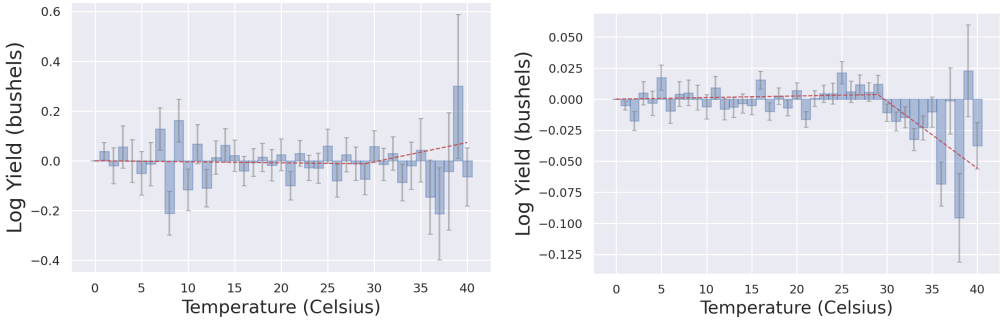



（c）大豆，长期变化，多项式普通最小二乘法




（d）大豆，短期变异，多项式普通最小二乘法

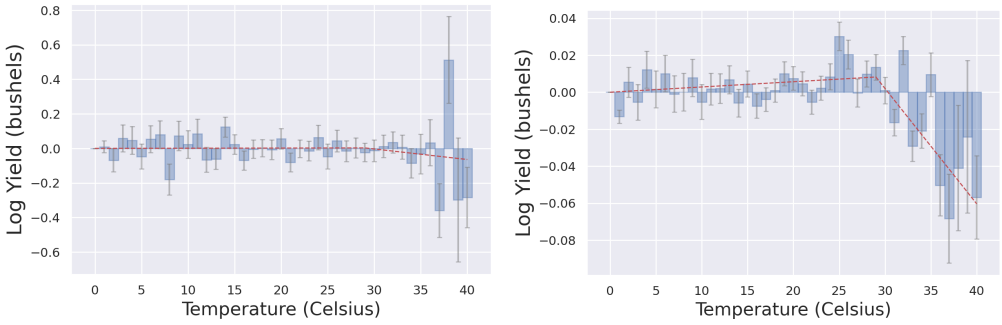
图9：使用多项式普通最小二乘法比较温度与产量之间在短期变异和长期变化下的关系。




（a）玉米，长期变化，线性普通最小二乘法




(b) 玉米、短期变异、套索双重机器学习




(c) 大豆、长期变化、线性普通最小二乘法




(d) 大豆、短期变异、线性普通最小二乘法

图10：使用线性普通最小二乘法比较温度与产量之间的关系，以及短期和长期变化。

### 附录C.2 跨时期比较

本节给出使用不同历史时期的短期和长期变化估计年度灵活天气变量的普通最小二乘法结果。每幅图中，我比较了短期估计值（使用涵盖这30年范围内所有年份的面板数据集）与长期估计值（比较该范围内第一个和最后一个十年的平均天气和作物产量值）。

每幅图包含四个独立的图表。每个图表概括了使用年度灵活天气变量组进行普通最小二乘法回归得到的系数估计值。柱状图显示在给定温度水平下，暴露于一个度日所导致的预期产量变化。误差棒表示每个估计系数上下一个标准差的范围。曲线表示对柱状图进行的最佳分段线性拟合，其中一段为0–29°C，另一段为29–40°C。回归是在进行组内变换后进行的，同时还包括年度固定效应项和降水；这些项未在图中显示。




（a）玉米，长期变化




(b) 玉米，短期变异

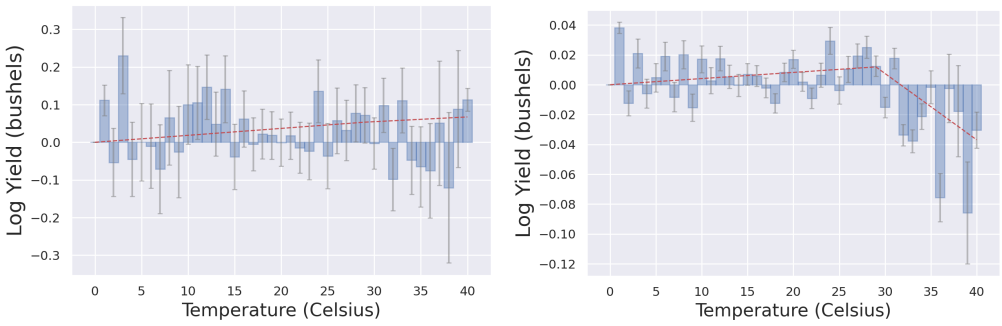



(c) 大豆，长期变化




(d) 大豆，短期变异

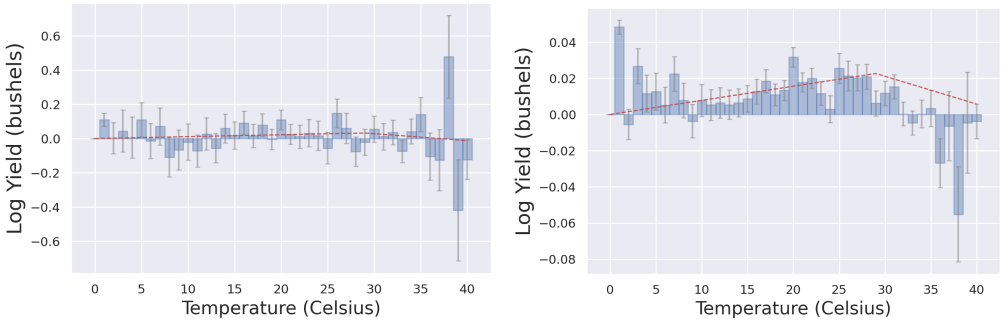
图11：玉米　　比较温度与产量之间的关系，以及短期和长期变化　　，
来自1950—1979年。




(a) 玉米，长期变化




(b) 玉米，短期变异

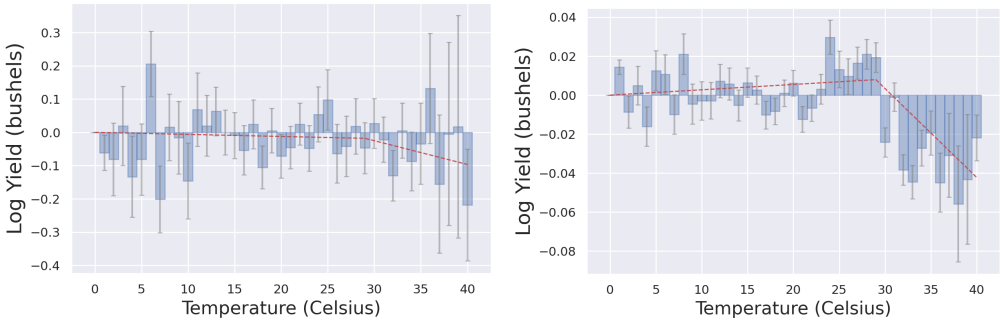



(c) 大豆，长期变化




(d) 大豆，短期变异

图12：1960—1989年温度与产量之间关系的短期和长期变异比较。




(a) 玉米，长期变化




(b) 玉米，短期变异

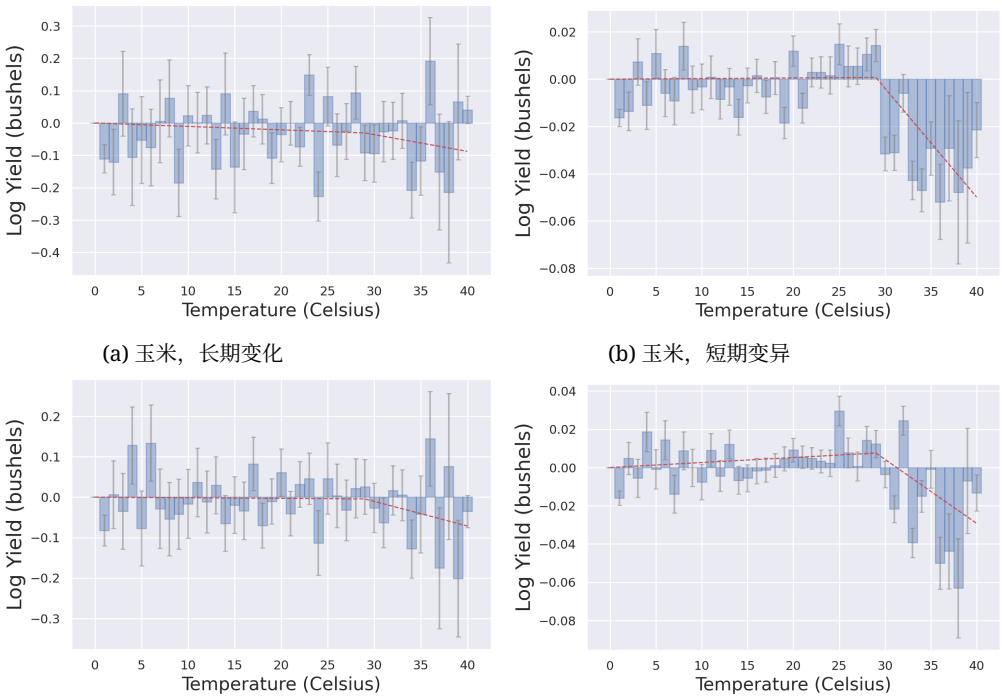



(c) 大豆，长期变异




(d) 大豆，短期变异

图13：1970—1999年温度与产量关系的短期和长期变化比较。



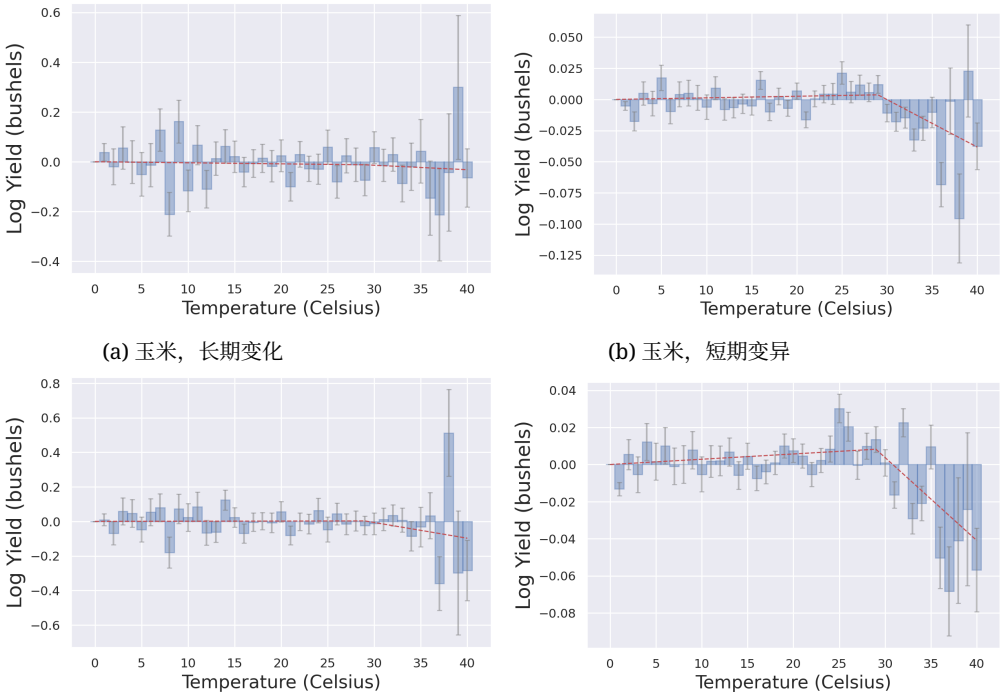

(a) 玉米，长期变化

(b) 玉米，短期变异

(c) 大豆，长期变化

(d) 大豆，短期变异

图14：1980—2009年温度与产量关系在短期变异和长期变化下的比较。




(a) 玉米，长期变化

(b) 玉米，短期变异

(c) 大豆，长期变化

(d) 大豆，短期变异

图15：1990—2019年温度与产量关系的短期和长期变化比较。