

当代 DID 方法：内生性问题

从理论到实践应用

许文立, wxu@cityu.edu.mo

澳门城市大学金融学院

2025 年 3 月 10 日



澳門城市大學
Universidade da Cidade de Macau
City University of Macau

① 因果推断中的内生性

② DID 中的协变量

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

⑥ 参考文献

① 因果推断中的内生性

② DID 中的协变量

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

⑥ 参考文献

1 因果推断中的内生性

2 DID 中的协变量

3 DID 中的机器学习

一个例子：天气热了，买冰淇淋吗？

FWL 定理

参数 DDML

非参数 DDML

因果机器学习的发展与 DID 中的应用

DID+ML: 两个应用

4 内生的处理配置

5 DID-IV 与 IV 型 DID

1 因果推断中的内生性

2 DID 中的协变量

3 DID 中的机器学习

一个例子：天气热了，买冰淇淋吗？

FWL 定理

参数 DDML

非参数 DDML

因果机器学习的发展与 DID 中的应用

DID+ML: 两个应用

4 内生的处理配置

5 DID-IV 与 IV 型 DID

1 因果推断中的内生性

2 DID 中的协变量

3 DID 中的机器学习

一个例子：天气热了，买冰淇淋吗？

FWL 定理

参数 DDML

非参数 DDML

因果机器学习的发展与 DID 中的应用

DID+ML: 两个应用

4 内生的处理配置

5 DID-IV 与 IV 型 DID

1 因果推断中的内生性

2 DID 中的协变量

3 DID 中的机器学习

一个例子：天气热了，买冰淇淋吗？

FWL 定理

参数 DDML

非参数 DDML

因果机器学习的发展与 DID 中的应用

DID+ML: 两个应用

4 内生的处理配置

5 DID-IV 与 IV 型 DID

- 回忆一下，DDML 将数据建模如下：

$$Y_i = \hat{M}_y(X_i) + \tau(X_i)(T_i - \hat{M}_t(X)) + \hat{\epsilon}_i$$

- 回忆一下，DDML 将数据建模如下：

$$Y_i = \hat{M}_y(X_i) + \tau(X_i)(T_i - \hat{M}_t(X)) + \hat{\epsilon}_i$$

- 重排上式

$$\hat{\epsilon}_i = (Y_i - \hat{M}_y(X_i)) - \tau(X_i)(T_i - \hat{M}_t(X))$$

- 回忆一下，DDML 将数据建模如下：

$$Y_i = \hat{M}_y(X_i) + \tau(X_i)(T_i - \hat{M}_t(X)) + \hat{\epsilon}_i$$

- 重排上式

$$\hat{\epsilon}_i = (Y_i - \hat{M}_y(X_i)) - \tau(X_i)(T_i - \hat{M}_t(X))$$

- 因果损失函数

$$\hat{L}_n(\tau(x)) = \frac{1}{n} \sum_{i=1}^n \left((Y_i - \hat{M}_y(X_i)) - \tau(X_i)(T_i - \hat{M}_t(X)) \right)^2$$

- 回忆一下，DDML 将数据建模如下：

$$Y_i = \hat{M}_y(X_i) + \tau(X_i)(T_i - \hat{M}_t(X)) + \hat{\epsilon}_i$$

- 重排上式

$$\hat{\epsilon}_i = (Y_i - \hat{M}_y(X_i)) - \tau(X_i)(T_i - \hat{M}_t(X))$$

- 因果损失函数

$$\hat{L}_n(\tau(x)) = \frac{1}{n} \sum_{i=1}^n \left((Y_i - \hat{M}_y(X_i)) - \tau(X_i)(T_i - \hat{M}_t(X)) \right)^2$$

- 用余值版本

$$\hat{L}_n(\tau(x)) = \frac{1}{n} \sum_{i=1}^n \left(\tilde{Y}_i - \tau(X_i) \tilde{T}_i \right)^2$$

$$\hat{L}_n(\tau(x)) = \frac{1}{n} \sum_{i=1}^n \tilde{T}_i^2 \left(\frac{\tilde{Y}_i}{\tilde{T}_i} - \tau(X_i) \right)^2$$

- 得到非参数因果损失函数的步骤

- 得到非参数因果损失函数的步骤

- 1 创建权重 \tilde{T}_i^2

- 得到非参数因果损失函数的步骤

- 1 创建权重 \tilde{T}_i^2

- 2 创建目标 $\frac{\tilde{Y}_i}{\tilde{T}_i}$

- 得到非参数因果损失函数的步骤

- 1 创建权重 \tilde{T}_i^2

- 2 创建目标 $\frac{\tilde{Y}_i}{\tilde{T}_i}$

- 3 用预测方法来得到目标 2，并使用权重 1

1 因果推断中的内生性

2 DID 中的协变量

3 DID 中的机器学习

一个例子：天气热了，买冰淇淋吗？

FWL 定理

参数 DDML

非参数 DDML

因果机器学习的发展与 DID 中的应用

DID+ML: 两个应用

4 内生的处理配置

5 DID-IV 与 IV 型 DID

Frame Title

- 在过去 15 年，因果机器学习取得了许多发展，是最引人注目的因果推断方法
 - Belloni, Chernozhukov and Hansen (2014)
 - Farrell (2015)
 - Belloni, Chernozhukov, Fernández-Val and Hansen (2017)
 - Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey and Robins (2018)
 - Athey and Wager (2018)
 - Athey, Tibshirani and Wager (2019)
 - Chernozhukov, Demirer, Duflo and Fernández-Val (2022)
- 这些机器学习方法已经在最近的 DID 理论中有所应用：Chang (2020); Callaway, Drukker, Liu and Sant'Anna (2023); Julia Hatamyar, Noemi Kreif, Rudi Rocha, Martin Huber(2024); Zhang(2024); Martin Huber, Eva-Maria OeSS(2024)

1 因果推断中的内生性

2 DID 中的协变量

3 DID 中的机器学习

一个例子：天气热了，买冰淇淋吗？

FWL 定理

参数 DDML

非参数 DDML

因果机器学习的发展与 DID 中的应用

DID+ML: 两个应用

4 内生的处理配置

5 DID-IV 与 IV 型 DID

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 初步参考：【香樟推文 2546】不完美的市场和不完美的监管——以美国电力市场为例

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 初步参考：【香樟推文 2546】不完美的市场和不完美的监管——以美国电力市场为例
- 识别策略：DID

$$y_{pt} = \tau D_{pt} + \gamma_p + \delta_{tr} + \epsilon_{pt}$$

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 初步参考：【香樟推文 2546】不完美的市场和不完美的监管——以美国电力市场为例
- 识别策略：DID

$$y_{pt} = \tau D_{pt} + \gamma_p + \delta_{tr} + \epsilon_{pt}$$

- Cicala (2022) 指出，即使与没有市场化改革的地区相比，DID 并不足以估计电力市场化后结果的变化

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 初步参考：【香樟推文 2546】不完美的市场和不完美的监管——以美国电力市场为例
- 识别策略：DID

$$y_{pt} = \tau D_{pt} + \gamma_p + \delta_{tr} + \epsilon_{pt}$$

- Cicala (2022) 指出，即使与没有市场化改革的地区相比，DID 并不足以估计电力市场化后结果的变化
- 在 PCA，结果会被时变的原料价格混淆。区域间的当期差异会被存在装机容量差异的 PCA 所混淆，因此，共同的时变冲击会有差异化的影响

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 初步参考：【香樟推文 2546】不完美的市场和不完美的监管——以美国电力市场为例
- 识别策略：DID

$$y_{pt} = \tau D_{pt} + \gamma_p + \delta_{tr} + \epsilon_{pt}$$

- Cicala (2022) 指出，即使与没有市场化改革的地区相比，DID 并不足以估计电力市场化后结果的变化
- 在 PCA，结果会被时变的原料价格混淆。区域间的当期差异会被存在装机容量差异的 PCA 所混淆，因此，共同的时变冲击会有差异化的影响
- 改进方法：政策函数法——给定一些变量，用电厂生产的历史模式来估计预测的电力生产配置。这些预测值应用于观测电厂成本来估计如果没有发生处理，生产成本是多少

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 在三个假设下，可以将上述 DID 方程扩展为

$$Y_{pt} - F(\psi_{p,y-1}^0, \mathbf{X}_{pt}) = \tau D_{pt} + \delta_{tr} + \gamma_p + v_{pt}$$

其中， $F(\psi_{p,y-1}^0, \mathbf{X}_{pt})$ 表示电力生产相关的政策函数，上标 0 表示未处理的反事实

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 在三个假设下，可以将上述 DID 方程扩展为

$$Y_{pt} - F(\psi_{p,y-1}^0, \mathbf{X}_{pt}) = \tau D_{pt} + \delta_{tr} + \gamma_p + v_{pt}$$

其中， $F(\psi_{p,y-1}^0, \mathbf{X}_{pt})$ 表示电力生产相关的政策函数，上标 0 表示未处理的事实

- 政策函数法消除了反事实结果估计中的燃料价格变动的作用

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 在三个假设下，可以将上述 DID 方程扩展为

$$Y_{pt} - F(\psi_{p,y-1}^0, \mathbf{X}_{pt}) = \tau D_{pt} + \delta_{tr} + \gamma_p + v_{pt}$$

其中， $F(\psi_{p,y-1}^0, \mathbf{X}_{pt})$ 表示电力生产相关的政策函数，上标 0 表示未处理的事实

- 政策函数法消除了反事实结果估计中的燃料价格变动的作用
- 估计政策函数要求平衡灵活性和过度拟合风险

Steven Cicala(2017,NBER): 美国电力市场化改革的效应

- 在三个假设下，可以将上述 DID 方程扩展为

$$Y_{pt} - F(\psi_{p,y-1}^0, \mathbf{X}_{pt}) = \tau D_{pt} + \delta_{tr} + \gamma_p + v_{pt}$$

其中， $F(\psi_{p,y-1}^0, \mathbf{X}_{pt})$ 表示电力生产相关的政策函数，上标 0 表示未处理的反事实

- 政策函数法消除了反事实结果估计中的燃料价格变动的作用
- 估计政策函数要求平衡灵活性和过度拟合风险
- Cicala 认为机器学习方法非常适合估计/预测政策函数

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 面板数据模型：用丰富的固定效应和控制变量来分参数化分离能源效率升级的因果效应和其它混淆因子

$$Y_{ith} = \beta D_{it} + \alpha_{ith} + \varepsilon_{ith}$$

由于数据很丰富，可以包括许多维度的固定效应，例如，
 school-by-hour-of-day 固定效应，
 school-by-hour-by-month-of-year 固定效应

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 面板数据模型：用丰富的固定效应和控制变量来分参数化分离能源效率升级的因果效应和其它混淆因子

$$Y_{ith} = \beta D_{it} + \alpha_{ith} + \varepsilon_{ith}$$

由于数据很丰富，可以包括许多维度的固定效应，例如，
school-by-hour-of-day 固定效应，
school-by-hour-by-month-of-year 固定效应

- 还利用下列方程评估事前学校能源节约率

$$Y_{ith} = -\beta \bar{S}_i \times D_{it} + \alpha_{ith} + \varepsilon_{ith}$$

其中， \bar{S}_i 表示学校各异的平均能源节约率，没有采用能源节约措施的学校为 0

Burlig(2020,JAERE): 美国学校能源效率升级的效应

Table 2: Panel fixed effects results

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Average program estimates</i>						
Realization rate	0.68	0.81	0.52	0.31	0.43	0.39
Point estimate	-2.88	-3.47	-2.15	-1.26	-1.74	-1.60
	(0.44)	(0.44)	(0.46)	(0.45)	(0.47)	(0.45)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360	57,480,360
<i>Panel B: Average school-specific estimates</i>						
Realization rate	0.53	0.59	0.51	0.41	0.46	0.45
	(0.12)	(0.13)	(0.13)	(0.12)	(0.12)	(0.13)
Observations	55,818,652	55,817,256	55,817,256	55,818,652	55,817,256	57,480,360
School-Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes	Yes
Time trend	No	No	Yes	No	No	No
Month of Sample FE	No	No	No	Yes	Yes	Yes
Temp Ctrl	No	No	No	No	No	Yes

Notes: Panel A in this table reports results from estimating Equation (3.1), with hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses. Realization rates are calculated by dividing the regression results on a complementary regression of *ex ante* engineering energy savings where expected (and zero otherwise) on our treatment variable, where we include the same set of controls and fixed effects. Panel B reports results from estimating Equation (3.2), in which the independent variable equals (the negative of) average expected savings for treated schools after their first upgrade, and 0 otherwise.

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 对奇异值较为敏感

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 对奇异值较为敏感
- 采用“武汉面窝”声明，即去掉处理前后紧邻的一期来解释可能的处理日期的测量误差问题，还采用了连续时变处理声明，结果也很敏感

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 对奇异值较为敏感
- 采用“武汉面窝”声明，即去掉处理前后紧邻的一期来解释可能的处理日期的测量误差问题，还采用了连续时变处理声明，结果也很敏感
- 选择“正确”控制变量很难，尤其是带有控制变量与个体/时间固定效应的交乘项（见上一期的协变量内容）

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 对奇异值较为敏感
- 采用“武汉面窝”声明，即去掉处理前后紧邻的一期来解释可能的处理日期的测量误差问题，还采用了连续时变处理声明，结果也很敏感
- 选择“正确”控制变量很难，尤其是带有控制变量与个体/时间固定效应的交乘项（见上一期的协变量内容）
- 丰富的数据使得研究者更加难以选择可信的回归模型

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 为了克服这些挑战，用机器学习方法

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 为了克服这些挑战，用机器学习方法
- 用高频数据和机器学习方法来选择可能的协变量集合

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 为了克服这些挑战，用机器学习方法
- 用高频数据和机器学习方法来选择可能的协变量集合
- 两步：第一步，用机器学习工具创建个体异质的结果模型：

$$Y_{it} = \beta D_{it} + \gamma_i X_{it} + \alpha_{ith} + \epsilon_{it}$$

即用机器学习来选择 X_{it} ，然后用处理前的观测值来获得 γ_i 。

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 为了克服这些挑战，用机器学习方法
- 用高频数据和机器学习方法来选择可能的协变量集合
- 两步：第一步，用机器学习工具创建个体异质的结果模型：

$$Y_{it} = \beta D_{it} + \gamma_i X_{it} + \alpha_{ith} + \epsilon_{it}$$

即用机器学习来选择 X_{it} ，然后用处理前的观测值来获得 γ_i 。

- 这允许我们从处理效应估计中分离出学校各异的系数选择 γ_i

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 为了克服这些挑战，用机器学习方法
- 用高频数据和机器学习方法来选择可能的协变量集合
- 两步：第一步，用机器学习工具创建个体异质的结果模型：

$$Y_{it} = \beta D_{it} + \gamma_i X_{it} + \alpha_{ith} + \epsilon_{it}$$

即用机器学习来选择 X_{it} ，然后用处理前的观测值来获得 γ_i 。

- 这允许我们从处理效应估计中分离出学校各异的系数选择 γ_i
- 使用这些模型来预测处理后的结果

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 为了克服这些挑战，用机器学习方法
- 用高频数据和机器学习方法来选择可能的协变量集合
- 两步：第一步，用机器学习工具创建个体异质的结果模型：

$$Y_{it} = \beta D_{it} + \gamma_i X_{it} + \alpha_{ith} + \epsilon_{it}$$

即用机器学习来选择 X_{it} ，然后用处理前的观测值来获得 γ_i 。

- 这允许我们从处理效应估计中分离出学校各异的系数选择 γ_i
- 使用这些模型来预测处理后的结果
- 处理后实际结果和预测结果之间的差异就是处理效应的一个粗略估计

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 为了克服这些挑战，用机器学习方法
- 用高频数据和机器学习方法来选择可能的协变量集合
- 两步：第一步，用机器学习工具创建个体异质的结果模型：

$$Y_{it} = \beta D_{it} + \gamma_i X_{it} + \alpha_{ith} + \epsilon_{it}$$

即用机器学习来选择 X_{it} ，然后用处理前的观测值来获得 γ_i 。

- 这允许我们从处理效应估计中分离出学校各异的系数选择 γ_i
- 使用这些模型来预测处理后的结果
- 处理后实际结果和预测结果之间的差异就是处理效应的一个粗略估计
- 然后，第一步并不能恰当的控制时间趋势和其它的混淆因子（如 DID 中控制的那些）

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 两步：第二步，用类似的面板固定效应模型，但是用来自第一步的结果变量预测误差作为因变量：

$$Y_{ith} - \hat{Y}_{ith} = \beta D_{it} + \alpha_{ith} + \gamma \text{posttrain}_{ith} + \varepsilon_{ith}$$

其中， posttrain_{ith} 是样本外预测时期的虚拟变量

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 两步：第二步，用类似的面板固定效应模型，但是用来自第一步的结果变量预测误差作为因变量：

$$Y_{ith} - \hat{Y}_{ith} = \beta D_{it} + \alpha_{ith} + \gamma \text{posttrain}_{ith} + \varepsilon_{ith}$$

其中， posttrain_{ith} 是样本外预测时期的虚拟变量

- 在上述方程声明中，可以控制时间趋势和其它混淆因子

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 两步：第二步，用类似的面板固定效应模型，但是用来自第一步的结果变量预测误差作为因变量：

$$Y_{ith} - \hat{Y}_{ith} = \beta D_{it} + \alpha_{ith} + \gamma \text{posttrain}_{ith} + \varepsilon_{ith}$$

其中， posttrain_{ith} 是样本外预测时期的虚拟变量

- 在上述方程声明中，可以控制时间趋势和其它混淆因子
- 其它的声明：

$$Y_{ith} - \hat{Y}_{ith} = -\beta \bar{S}_i D_{it} + \alpha_{ith} + \gamma \text{posttrain}_{ith} + \varepsilon_{ith}$$

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 两步：第二步，用类似的面板固定效应模型，但是用来自第一步的结果变量预测误差作为因变量：

$$Y_{ith} - \hat{Y}_{ith} = \beta D_{it} + \alpha_{ith} + \gamma \text{posttrain}_{ith} + \varepsilon_{ith}$$

其中， posttrain_{ith} 是样本外预测时期的虚拟变量

- 在上述方程声明中，可以控制时间趋势和其它混淆因子
- 其它的声明：

$$Y_{ith} - \hat{Y}_{ith} = -\beta \bar{S}_i D_{it} + \alpha_{ith} + \gamma \text{posttrain}_{ith} + \varepsilon_{ith}$$

- 识别：要求处理学校和未处理学校的预测误差（不是能源消费）有类似的变化趋势

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 注意，上述机器学习方法并不能避免源自于能源效率升级直接相关的能源消费变化的偏误

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 注意，上述机器学习方法并不能避免源自于能源效率升级直接相关的能源消费变化的偏误
- 1 如果一所学校采取了其他的能源节约措施，会高估处理效应

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 注意，上述机器学习方法并不能避免源自于能源效率升级直接相关的能源消费变化的偏误
 - 1 如果一所学校采取了其他的能源节约措施，会高估处理效应
 - 2 任何正向选择进入处理都会导致估计量偏离 0

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 注意，上述机器学习方法并不能避免源自于能源效率升级直接相关的能源消费变化的偏误
 - 1 如果一所学校采取了其他的能源节约措施，会高估处理效应
 - 2 任何正向选择进入处理都会导致估计量偏离 0
 - 3 对于一个导致结果趋向 0 的混淆因子，一所学校可能会随着省级而增加能源使用

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 第一步：预测反事实

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 第一步：预测反事实

1 对于处理学校，将升级发生前的时期定义为处理前时期

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 第一步：预测反事实

- 1 对于处理学校，将升级发生前的时期定义为处理前时期

- 2 对于未处理学校，随机配置处理日期，以便可以定义处理前时期

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 第一步：预测反事实

- 1 对于处理学校，将升级发生前的时期定义为处理前时期
- 2 对于未处理学校，随机配置处理日期，以便可以定义处理前时期
- 3 仅仅用处理前时期数据来训练预测模型

Burlig(2020,JAERE): 美国学校能源效率升级的效应

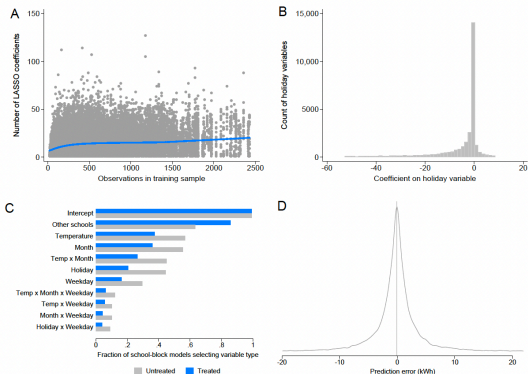
● 第一步：预测反事实

- 1 对于处理学校，将升级发生前的时期定义为处理前时期
- 2 对于未处理学校，随机配置处理日期，以便可以定义处理前时期
- 3 仅仅用处理前时期数据来训练预测模型
- 4 用一种机器学习方法（如 LASSO）来搜寻潜在协变量组合，包括许多协变量、协变量之间的交互项、协变量与固定效应之间的交互项等等，除了这些个体差异变量，还包括未处理学校的能源消费作为潜在的预测变量（类似于 SC）

Burlig(2020,JAERE): 美国学校能源效率升级的效应

● 第一步：预测绩效的四个有效性检验

Figure 3: Machine learning diagnostics



Notes: This figure presents three checks of our machine learning methodology. Panel A displays the relationship between the number of observations in the pre-treatment (“training”) dataset and the number of variables LASSO selects to include in the prediction model for each school in the sample. Panel B displays the marginal effect of holiday indicators in each school-specific prediction model. Panel C displays the categories of variables selected by our preferred LASSO method for untreated and treated schools. Finally, Panel D shows the distribution of average prediction errors out-of-sample for untreated schools (trimming the top and bottom 1 percent).

Burlig(2020,JAERE): 美国学校能源效率升级的效应

● 第二步：用预测误差跑面板回归

Table 4: Machine learning results

	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Average program estimates</i>					
Realization rate	0.86	0.92	0.75	0.53	0.60
Point estimate	-3.64	-3.92	-3.17	-2.10	-2.42
	(0.50)	(0.52)	(0.49)	(0.47)	(0.49)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360
<i>Panel B: Average school-specific estimates</i>					
Realization rate	0.57	0.58	0.55	0.50	0.50
	(0.13)	(0.14)	(0.14)	(0.13)	(0.13)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: Panel A in this table reports results from estimating Equation (3.3), with prediction errors in hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses. Realization rates are calculated by dividing the regression results on a complementary regression of ex-ante engineering energy savings where expected (and zero otherwise) on our treatment variable, also including the same set of controls. Panel B reports results from estimating Equation (3.4), in which the independent variable equals (the negative of) average expected savings for treated schools after their first upgrade, and 0 otherwise. All regressions include a control for being in the post-training period for the machine learning.

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 第二步：用预测误差跑面板回归

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 第二步：用预测误差跑面板回归
- 所有的面板固定效应模型稳健性都可以做

Burlig(2020,JAERE): 美国学校能源效率升级的效应

- 第二步：用预测误差跑面板回归
- 所有的面板固定效应模型稳健性都可以做
- 更换机器学习方法，例如使用 DDML(Chernozhukov et al. (2018))

① 因果推断中的内生性

② DID 中的协变量

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

⑥ 参考文献

① 因果推断中的内生性

② DID 中的协变量

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

⑥ 参考文献

① 因果推断中的内生性

② DID 中的协变量

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

⑥ 参考文献

- 1 Sant' Anna, DiD Resources, 2025
- 2 Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, James Robins, Double/debiased machine learning for treatment and structural parameters, The Econometrics Journal, Volume 21, Issue 1, 1 February 2018, Pages C1–C68
- 3 Alves, Matheus Facure. "Causal inference for the brave and true." São Paulo, Brazil. Available at <https://matheusfacure.github.io/python-causality-handbook/01-Introduction-To-Causality.html> (2022).
- 4 Goldsmith-Pinkham P, Hull P, Kolesár M. Contamination bias in linear regressions[J]. American Economic Review, 2024, 114(12): 4015-4051.

- 5 Burlig F, Knittel C, Rapson D, et al. Machine learning from schools about energy efficiency[J]. Journal of the Association of Environmental and Resource Economists, 2020, 7(6): 1181-1217.
- 6 Cicala S. Imperfect Markets versus Imperfect Regulation in US Electricity Generation[R]. National Bureau of Economic Research, 2017.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ 🔍 ↺ ↻