

连享会 · 2025 暑期班 · 高级班

B4. 面板函数系数模型

连玉君 (中山大学)

arlionn@163.com

C. Functional Coefficient Estimation of MLTECH

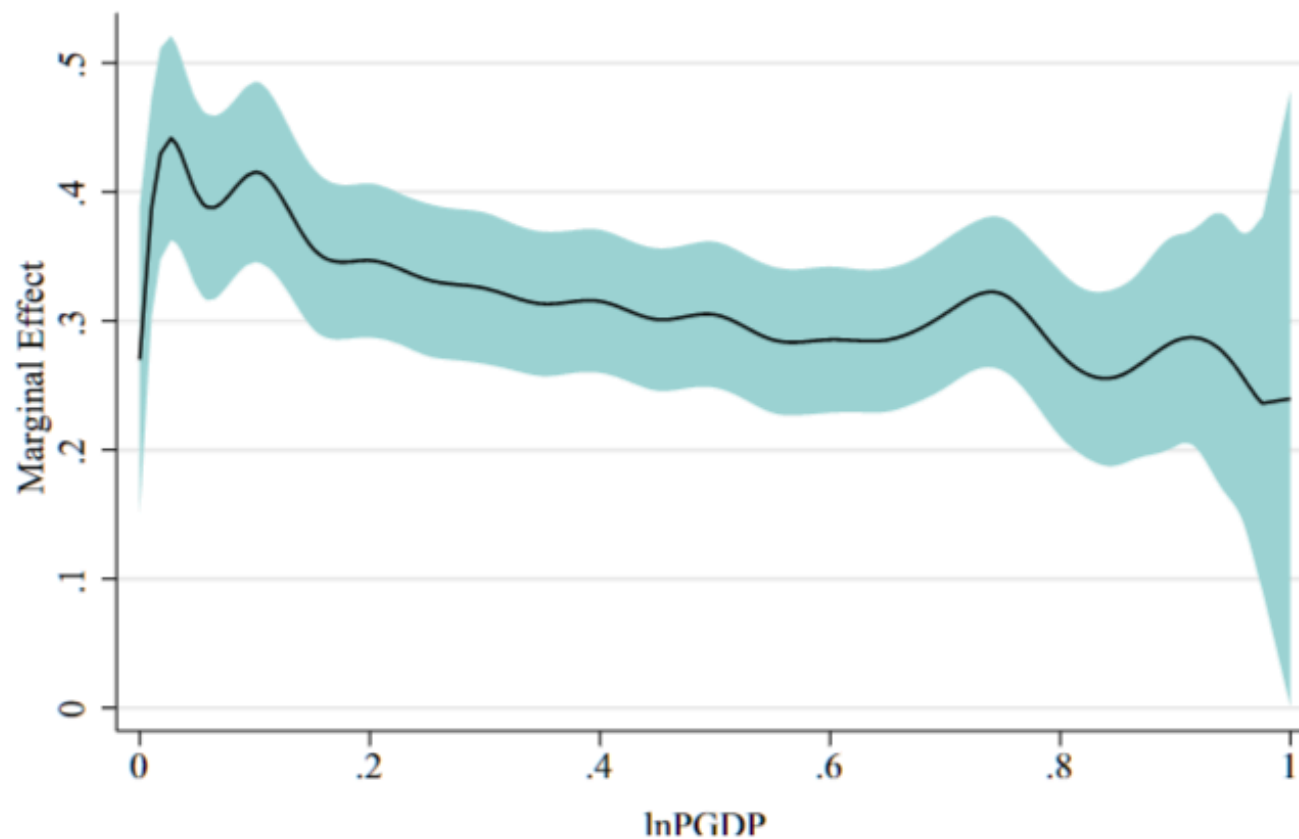


Fig. 5. Functional coefficients estimates.

Partially Linear Functional-Coefficient Models (PLFC)

- Du, K., Cheng, Y., & Yao, X. (2021). Environmental regulation, green technology innovation, and industrial structure upgrading: The road to the green transformation of Chinese cities. **Energy Economics**, 98, 105247. [Link](#) (rep), [PDF](#), [-Replication-](#), [Google](#).
 - 🍎 课件：【..\Du_2021_EE】
- Du, C., Cao, Y., Ling, Y., Jin, Z., Wang, S., & Wang, D. (2024). Does manufacturing agglomeration promote green productivity growth in China? Fresh evidence from partially linear functional-coefficient models. **Energy Economics**, 131, 107352. [Link](#) (rep), [PDF](#), [Google](#). [-Replication-](#)
 - 🍎 课件：【..\Du_2024_EE】
- Stata 命令：
 - `xtp1fc`：用于估计 PLFC 模型的 Stata 命令
 - `ivxtp1fc`：用于估计 PLFC 模型的工具变量版本
 - `xtdp1fc`：用于估计双向固定效应的 PLFC 模型

1. 问题背景

- 系数异质性
- 交乘项
- 分组回归

系数异质性 (异质性边际效应)

$$y_{it} = \alpha_i + x_{it}\beta + \varepsilon_{it} \quad \frac{\partial y}{\partial x} = \beta \text{ (constant)}$$

$$y_{it} = \alpha_i + x_{it}\beta_g + \varepsilon_{it} \quad g = \{FC, NFC\}, \{Male, Female\}$$

更一般化

- $y_{it} = \alpha_i + x_{it}\beta_i + \varepsilon_{it}$
- $y_{it} = \alpha_i + x_{it}\beta_t + \varepsilon_{it}$
- $y_{it} = \alpha_i + x_{it}\beta_{it} + \varepsilon_{it}$

Q: $\beta_{it} = ?$

- $FDI \rightarrow GDP$: β_i , 国家层面的制度、文化差异 (α_i 或 `i.country`), FDI 影响 GDP 的程度不同
- $FDI \rightarrow GDP$: β_t , 投资时机差异 (λ_t 或 `i.year`), FDI 影响 GDP 的程度不同
- $FDI \rightarrow GDP$: β_{it} , 资本存量 (K_{it}) 差异, FDI 影响 GDP 的程度不同

系数异质性：交乘项

$$y_{it} = \alpha_i + x_{it}\beta_{it} + \varepsilon_{it} \quad \frac{\partial y}{\partial x} = \beta_{it}$$

$$\beta_{it} = g(q_{it}) = q_{it}\theta$$

- 单变量： $\beta_{it} = \beta_0 + \theta q_{it}$ (Note: q_{it} 也可以是 x_{it} 自己)

- $y_{it} = x_{it}\beta_0 + (x_{it} \times q_{it})\theta + \varepsilon_{it}$

```
reghdfe y x c.x#c.q controls, absorb(id year)
```

- 双变量： $\beta_{it} = \beta_0 + \theta_1 q_{it} + \theta_2 w_{it}$

- $y_{it} = x_{it}\beta_0 + (x_{it} \times q_{it})\theta_1 + (x_{it} \times w_{it})\theta_2 + \varepsilon_{it}$

- `reghdfe y x c.x#c.(q w) controls, absorb(id year)`

- 固定效应： $\beta_{it} = \beta_0 + \theta_i D_i$

- $y_{it} = x_{it}\beta_0 + (x_{it} \times D_i)\theta_i + \varepsilon_{it}$

```
reg y x i.id i.id#c.x controls
```

- 时间效应： $\beta_{it} = \beta_0 + \theta_t D_t$

- $y_{it} = x_{it}\beta_0 + (x_{it} \times D_t)\theta_t + \varepsilon_{it}$

```
reg y x i.id i.year#c.x controls
```

可能的挑战

$$y_{it} = \alpha_i + x_{it}\beta_{it} + \varepsilon_{it} \quad \frac{\partial y}{\partial x} = \beta_{it}$$

挑战 1: 模型中包含的参数个数超过了样本数 $N \times T$ ，是无法识别的。

- **应对思路:** 设定一些约束条件，以便减少模型中待估参数的个数 (降维)。
 - **合并。** 在行业 (地区) 层面上考虑异质性，而不是公司层面上
 - **简化。** 用时间趋势代替时间虚拟变量。傻傻分不清：时间趋势项与时间虚拟变量
 - **组合。** 主成分分析 - 将宏观冲击归结为几个主要成分；异质性用因子载荷来反映。本质上是降维。
 - **regife:** 面板交互固定效应模型-Interactive Fixed Effect

挑战 2: $\beta_{it} = g(z_{it}, c, \gamma)$

- 机制分析：选谁做 z_{it} ?
- $g(z_{it}, c, \gamma)$ 是线性还是非线性?

2. PLFC 模型

- 模型设定
- 估计方法
- 应用实例

2.1 模型设定

$$Y_{it} = \delta_i + \beta_{it}Z_{it} + \mu_{it} \quad (1)$$

$$Y_{it} = \gamma(U_{it})Z_{it} + \beta_0X_{it} + \delta_i + \mu_{it} \quad (2)$$

在这种设定下， Z 对 Y 的边际影响可以表示为：

$$M_{it}^{YZ} = \frac{\partial(Y_{it})}{\partial(Z_{it})} = \gamma(U_{it}) \quad (3)$$

- Li et al. (2002) 采用非参数估计方法，以避免模型误设偏误。
- An et al. (2016) 进一步将 Li et al. (2002) 的模型从截面数据情形扩展到包含固定效应的面板数据情形下，称之为「**部分线性变系数面板模型**」。
- Du et al. (2020) 编写了相应的 Stata 命令 `xtp1fc` , `ivxtp1fc` 和 `xtdp1fc`
- Du et al. (2021) 应用该方法研究中国城市层面的环境规制与绿色创新之间的关系。

2.2 估计方法 (1)

第一步，进行一阶差分，消除固定效应 δ_i 。

$$\Delta Y_{it} = \Delta \gamma(U_{it})Z_{it} + \beta_0 \Delta X_{it} + \Delta \mu_{it} \quad (4)$$

第二步，函数系数的近似。用 k 个基函数的线性组合近似变系数函数 $\gamma(U_{it})$ ：

$$\gamma(U) \approx h(U)^T \theta = \sum_{j=1}^p \theta_j h_j(U)$$

其中， $h(U_{it})$ 是 $k \times 1$ 的基函数向量， θ 是 $k \times 1$ 的未知参数向量。当 p 增大时，存在 $h_j(U_{it})$ 的线性组合能够很好地近似任何光滑函数 $\gamma(U_{it})$ ，并使近似均方误差（MSE）尽可能小。

为了便于理解，考虑如下四个简单的基函数：

- $h_1(U) = 1, h_2(U) = U, h_3(U) = U^2, h_4(U) = U^3$

直觉上来讲，相当于设定

$$\gamma(U) \approx \theta_0 + \theta_1 U + \theta_2 U^2 + \theta_3 U^3 \quad (5)$$

当然，实际估计中，为了尽可能地拟合数据，还会酌情加入更复杂的基函数。

2.2 估计方法 (2)

将 (5) 代入差分模型 (4) 后, 可得:

$$\Delta Y_{it} = \Delta Z_{it} h(U_{it})^T \theta + \beta_0 \Delta X_{it} + \Delta \varepsilon_{it} \quad (6)$$

其中, $\Delta \varepsilon_{it}$ 表示序列近似误差: $\Delta \varepsilon_{it} = \Delta \mu_{it} + \Delta(\gamma(U_{it})Z_{it}) - \Delta(Z_{it}h(U_{it})^T)\theta$

第三步, 最小二乘估计

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\theta} \end{bmatrix} = (\Delta X^T \Delta X)^{-1} \Delta X^T \Delta Y \quad (7)$$

其中,

- $\Delta Y = [\Delta Y_{12}, \dots, \Delta Y_{NT}]^T$;
- $\Delta X = [\Delta X_{11}, \dots, \Delta Z_{NT-1} p(U_{NT-1})^T]^T$ 。

2.3 边际效应及其置信区间估计

目的： 检验在不同 U 水平下边际效应的统计显著性

1. 估计边际效应函数：

$$\hat{\gamma}(U) = h(U)^T \hat{\theta} \quad (8)$$

2. 估计残差项：

$$\hat{\varepsilon}_{it} = Y_{it} - \hat{\gamma}(U_{it})Z_{it} - \hat{\beta}^T X_{it} \quad (9)$$

3. 构造协方差矩阵估计 $\hat{\Sigma}$ ：若 H 是所有 H_{it} 的堆叠矩阵，则：

$$\hat{\Sigma} = (H^T H)^{-1} H^T \text{diag}(\hat{\varepsilon}^2) H (H^T H)^{-1} \quad (10)$$

4. 估计 $\gamma(U)$ 在 U_0 处的标准误：

$$\widehat{se}[\gamma(U_0)] = \sqrt{h(U_0)^T \hat{\Sigma} h(U_0)} \quad (11)$$

5. 构建 $\gamma(U)$ 的 95% 的置信区间：

$$[\hat{\gamma}(U_0) \pm 1.96 \cdot \widehat{se}[\gamma(U_0)]] \quad (12)$$

2.4 总结：PLFC 模型的优势

- 刻画连续非线性效应：边际效应随 PGDP 光滑变化
- 避免人为误设门槛值或模型误设问题
- 具备良好的解释力与可视化能力
- 适用于具有明显异质性、样本量大的面板数据结构

3. Stata 实现

3.1 命令语法

$$Y_{it} = \gamma(U_{it})Z_{it} + \beta_0 X_{it} + \delta_i + \mu_{it} \quad (2)$$

xtpfnc 命令的语法格式如下：

```
xtpfnc Y varlist, zvars(varlist) uvars(varname) generate(string) [options]
```

必填项：

- `varlist`：控制变量
- `zvars(varlist)`：指定具有函数系数的变量列表 Z_{it} (核心解释变量)。
- `uvars(varlist)`：指定（连续）变量 U_{it} ，这些变量以交互形式进入函数系数。
- `generate(prefix)`：指定一个前缀，用于存储函数系数拟合值 $\hat{\gamma}(U_{it})$ ，参见 (10) 式。

可选项：

- `te`：指定是否包含时间固定效应。
- `power(numlist)`：指定样条的幂（默认值为 3）。
- `nknots(numlist)`：指定用于样条插值的结点数量（默认值为 2）。
- `quantile`：指定基于经验分位数创建结点（默认情况下，结点通过等距规则生成）。
- `maxnknots(numlist)`：指定用于最小二乘交叉验证（LSCV）的最大结点数量。
- `minnknots(numlist)`：指定用于 LSCV 的最小结点数量（默认值为 2）。
- `brep(#)`：指定 bootstrap 复制次数（默认值为 200，建议根据实际需求调整）。
- `wild`：指定使用 wild bootstrap（默认采用 `cluster(panelvar)` 的残差 bootstrap）。
- `predict(prspec)`：用指定变量名存储条件均值和固定效应的预测值。可接受变量列表或前缀，第一个变量名为条件均值，第二个为固定效应。
- `level(#)`：设置置信水平（默认值为 95）。
- `fast`：使用 Mata 函数加速计算。
- `tenfoldcv`：使用十折交叉验证替代 LSCV。

4. 应用

应用 1: Du et al. (2021, EE)

- Du, K., Cheng, Y., & Yao, X. (2021). Environmental regulation, green technology innovation, and industrial structure upgrading: The road to the green transformation of Chinese cities. **Energy Economics**, 98, 105247.
[Link](#) (rep), [PDF](#), [-Replication-](#), [Google](#).
 - 🍎 课件: 【..\Du_2021_EE】

问题背景

- 环境管制 (Z) 可能会促进 绿色技术创新 (Y_1)、产业结构升级 (Y_2)，但作用效果会受 经济发展水平 (调节变量 U) 的影响。
- 政策启示：在经济欠发达地区或经济发展水平较低的阶段，不宜过度事实环境管制。

模型设定

$$Y_{it} = \gamma (U_{it-1})Z_{it-1} + \beta' X_{it-1} + \delta_i + \mu_{it} \quad (13)$$

- Y_{it} : 下列两个变量之一:
 - **lnGTI**: 绿色技术创新 或 **lnIS**: 产业结构
- Z_{it-1} : 第 i 个城市在 $t - 1$ 时刻的环境管制水平 (ER)
- U_{it-1} : 经济发展水平, 用人均 GDP 的对数衡量 (lnGDP)
- X_{it-1} : 控制变量, 包括
 - **lnRD**: 科技与教育经费投入
 - **lnPOP**: 城市规模
 - **lnHC**: 人力资本
 - **lnINV**: 固定资产投资
 - **lnFDI**: 经济开放度
- δ_i : 不可观测的个体固定效应

Stata 实操: 🍎 课件: [\[..\Du_2021_EE\]](#)

```
use "ERdata.dta", clear // ---- Table 4, col (1)

local controls "lnrd lnpop lnhc lninv lnfdi" // 控制变量

xtplfc lngti,          ///
    zvars(er)          /// 核心解释变量 (Z)
    uvar(lngdp)         /// 调节变量 (U)
    gen(fcoe_gti)       /// g(U) 的拟合值, 用于绘图
    maxnknots(20)       /// 设置最多 20 个结点进行平滑处理
    brep(1000)          // bootstrap 1000 次获取 SE
```

*----- Figure 2

```

gen lb_gti = fcoe_gti_1 - 1.96*fcoe_gti_1_sd // 95% CI lower bound
gen ub_gti = fcoe_gti_1 + 1.96*fcoe_gti_1_sd // 95% CI upper bound

local plot1 "line fcoe_gti_1 lngdp, color(black) sort" // 边际效应
local plot2 "rarea lb_gti ub_gti lngdp, color(gs12) sort" // 95% CI

twoway (`plot2') (`plot1'), ///
      legend(label(1 "95% CI") ///
              label(2 "Functional coefficients") ///
              ring(0) pos(5)) ///
      xtitle(, margin(t+2)) ///
      ytitle("Marginal effects: {&gamma} (lngdp)")

graph export "$out/Du2021_EE_Figure02.png", width(700) replace

```

Y1: 绿色技术创新 (GTI)

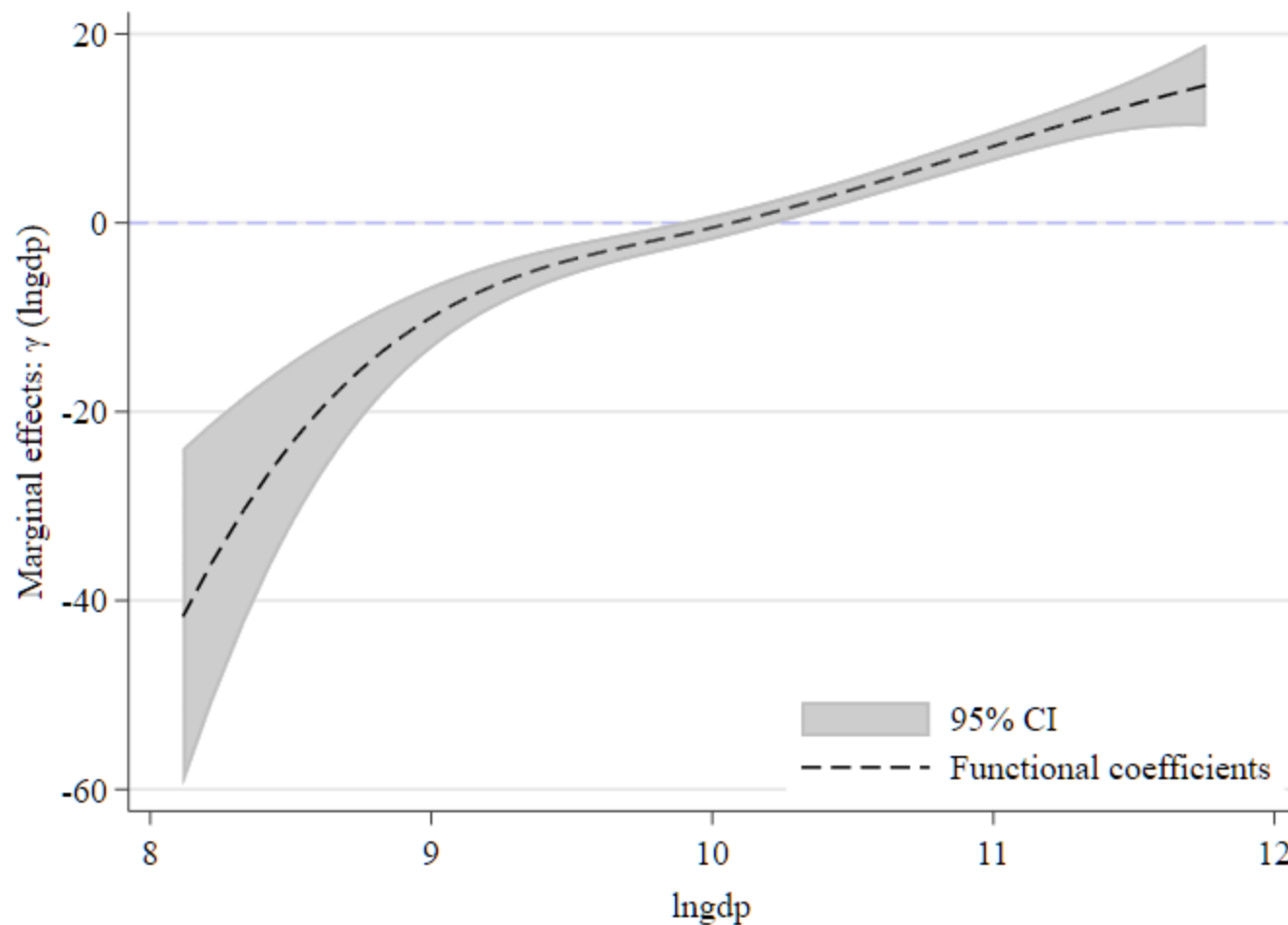
Fig. 2. ER 对 $\ln(\text{GTI})$ 的边际影响

$$Y_{it} = \ln(\text{gti}), Z_{it} = \text{ER};$$

$$U_{it-1} = \ln(\text{gdp})$$

主要结论：

- 在经济较低发展水平下，环境规制对绿色技术创新有抑制作用；
- 当 $\ln\text{GDP}$ 超过 10 后，环境规制开始显著促进绿色技术创新。



Y2: 产业结构优化升级 (IS)

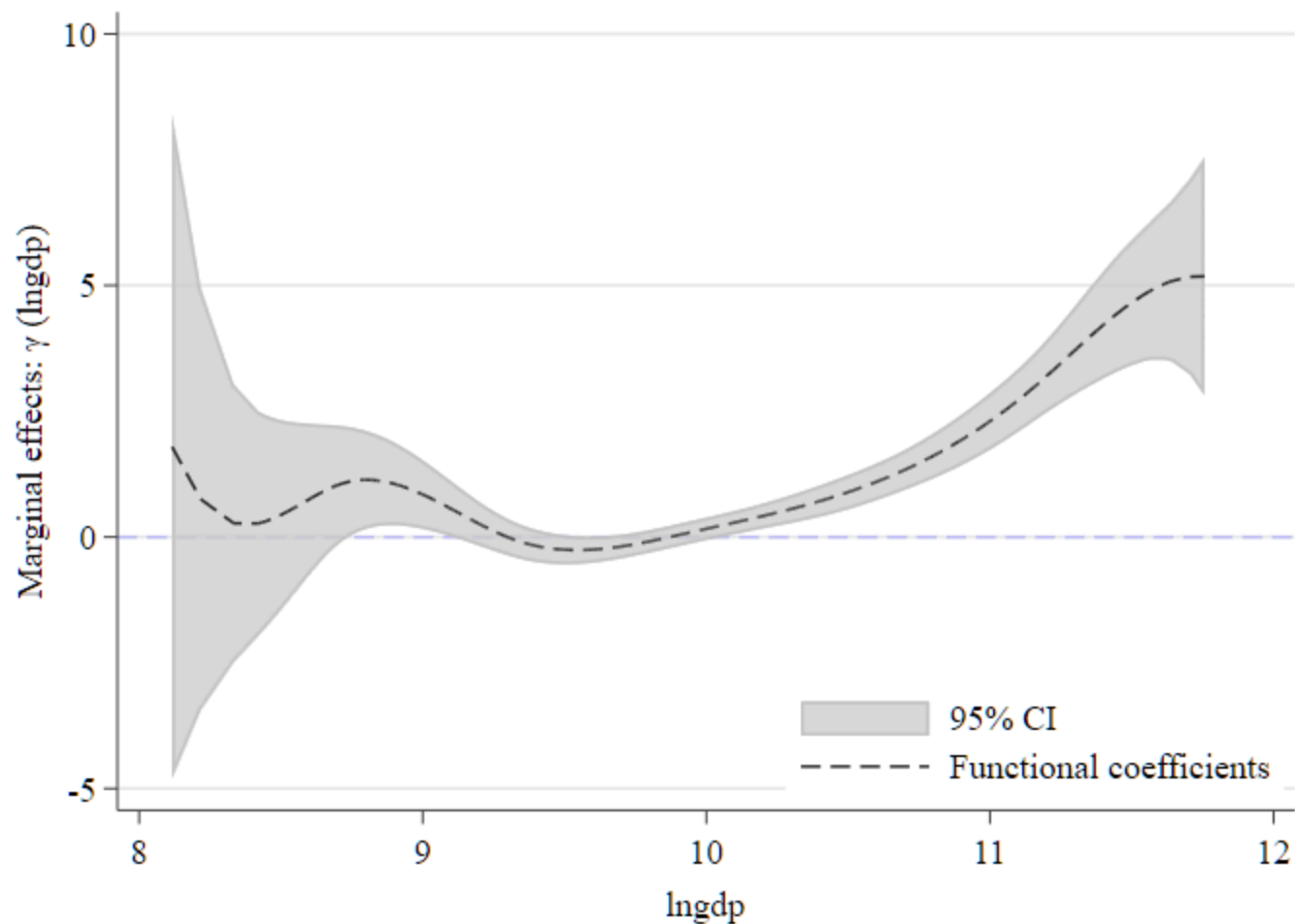
Fig. 3. ER 对 IS 的边际影响

$$Y_{it} = IS, Z_{it} = ER;$$

$$U_{it-1} = \ln(gdp)$$

主要结论：

- 在经济欠发达城市，环境规制对产业结构升级作用不显著；
- 在经济较发达城市，环境规制能有效促进产业结构优化升级。
- 门槛值大约在 10.5 左右。



边际效应的时序差异

Figure 4 (b) ER 的 ME (2005)

- □: $\ln(\text{GDP})$ - 较低;
- △: $\ln(\text{GDP})$ - 中等;
- ○: $\ln(\text{GDP})$ - 较高。
- 红色: $\text{ER} _ (\ln\text{GTI}, \ln\text{IS})$;
- 黄色: $\text{ER} \rightarrow \ln\text{IS}$, 但 $\text{ER} _ \ln\text{GTI}$;
- 蓝色: $\text{ER} _ \ln\text{IS}$, 但 $\text{ER} \rightarrow \ln\text{GTI}$;
- 绿色: $\text{ER} \rightarrow (\ln\text{GTI}, \ln\text{IS})$ 。

ER $_$ $\ln\text{GTI}$: 表示 ER 对 $\ln\text{GTI}$ 的影响不显著

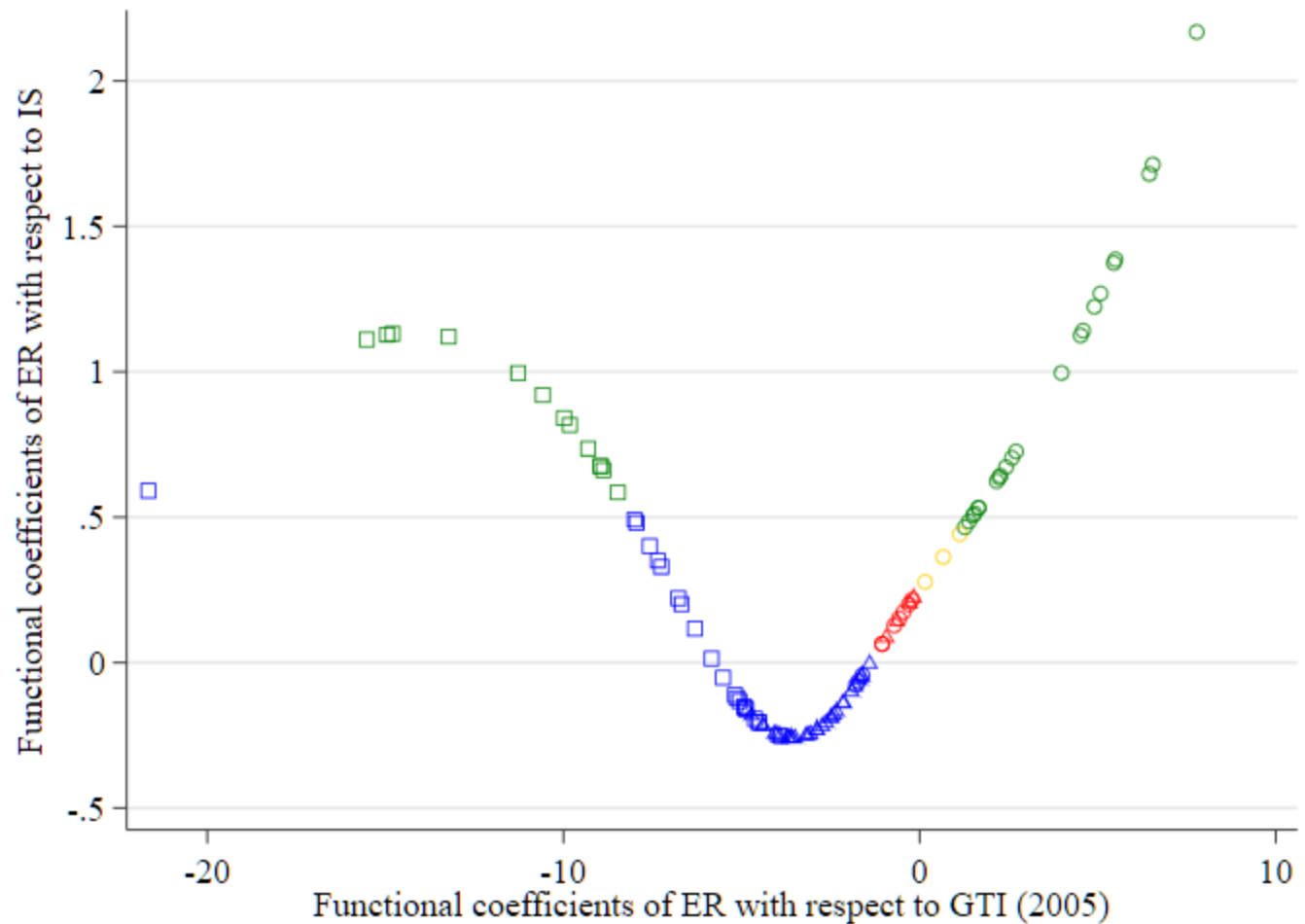


Figure 4 (b) ER 的 ME (2010)

- \square : $\ln(\text{GDP})$ - 较低;
- \triangle : $\ln(\text{GDP})$ - 中等;
- \bigcirc : $\ln(\text{GDP})$ - 较高。
- 红色: $\text{ER} \rightarrow (\ln \text{GTI}, \ln \text{IS})$;
- 黄色: $\text{ER} \rightarrow \ln \text{IS}$, 但 $\text{ER} \nrightarrow \ln \text{GTI}$;
- 蓝色: $\text{ER} \nrightarrow \ln \text{IS}$, 但 $\text{ER} \rightarrow \ln \text{GTI}$;
- 绿色: $\text{ER} \rightarrow (\ln \text{GTI}, \ln \text{IS})$ 。

$\text{ER} \nrightarrow \ln \text{GTI}$: 表示 ER 对 $\ln \text{GTI}$ 的影响不显著

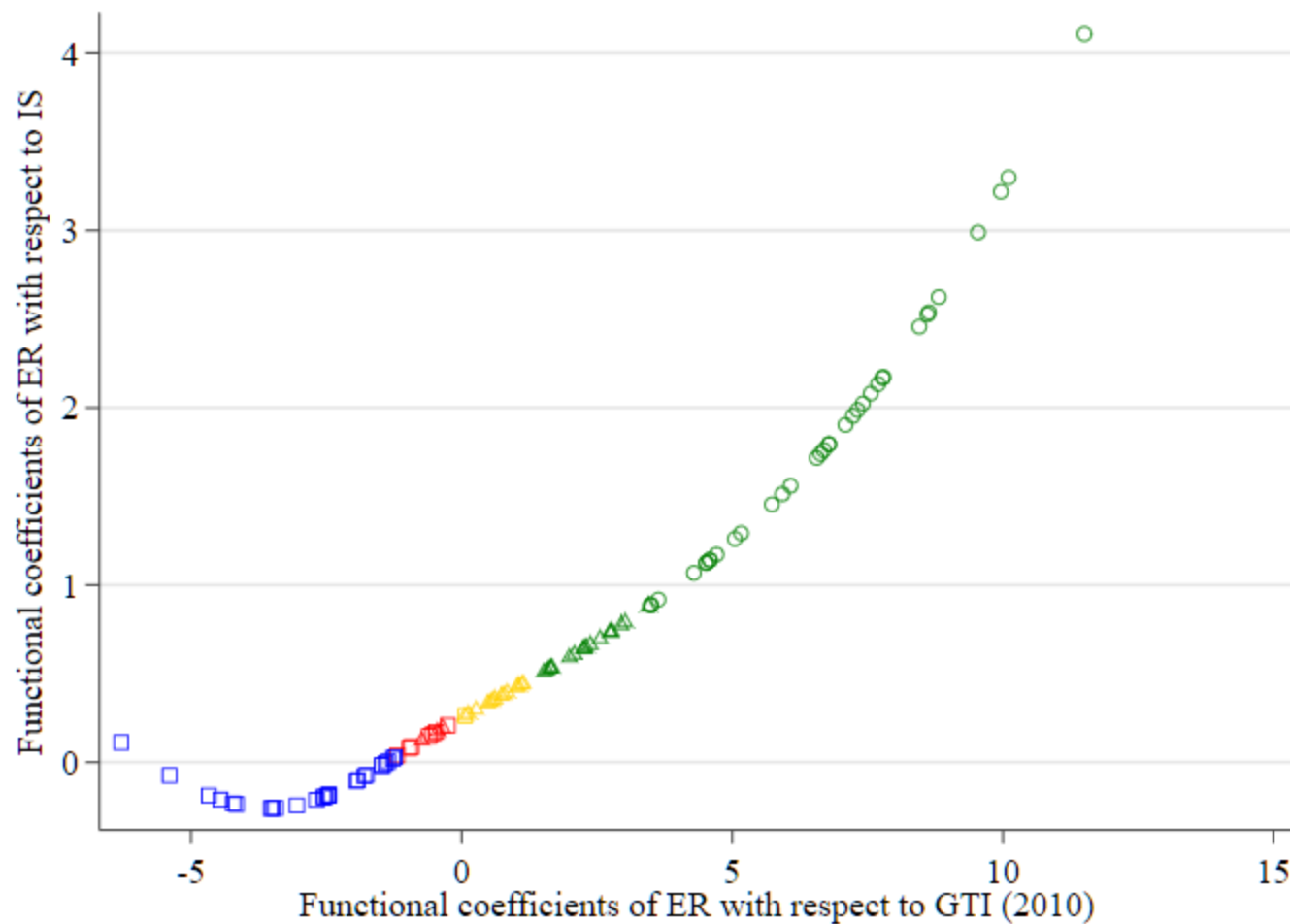
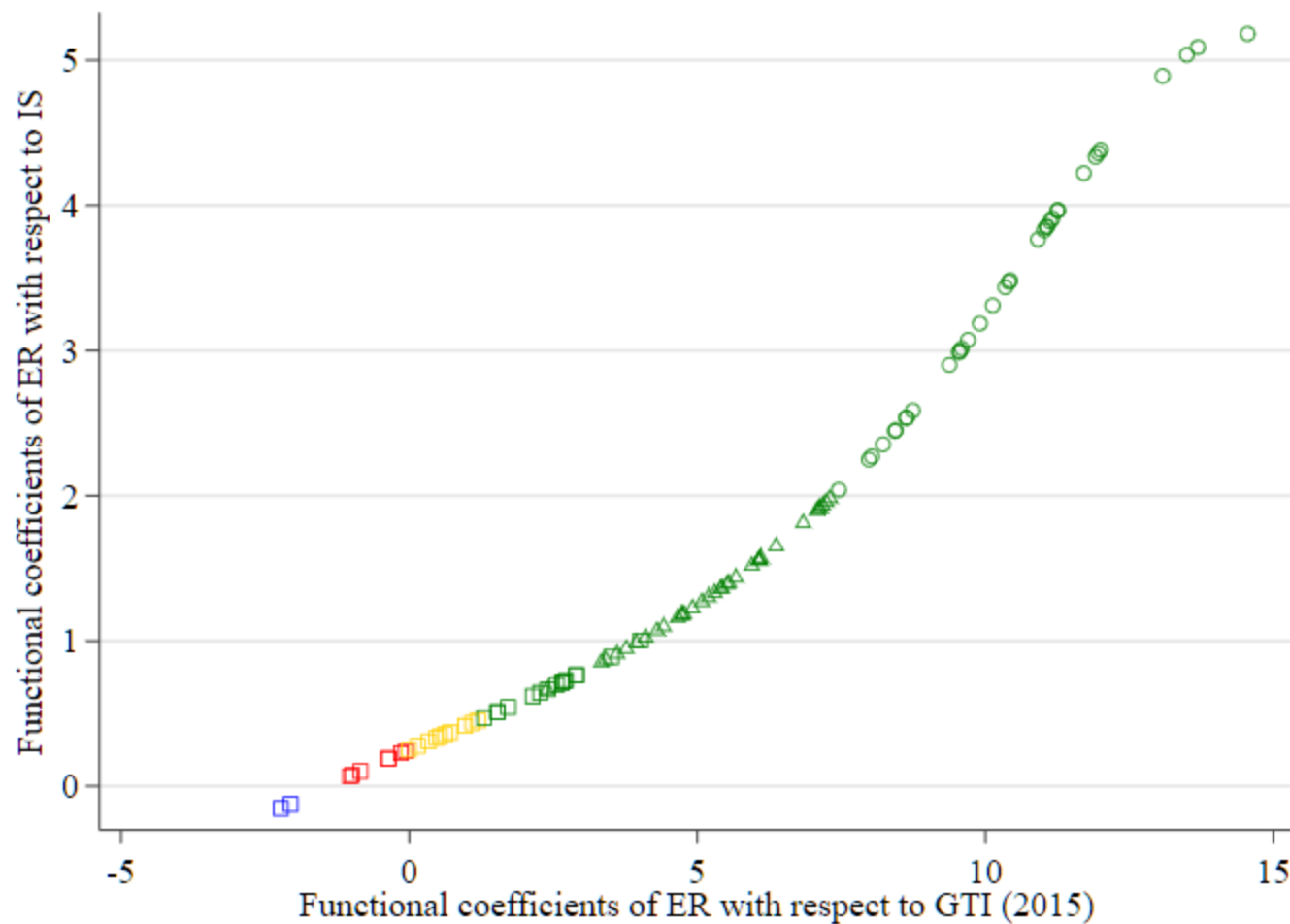


Figure 4 (c) ER 的 ME (2015)

- □: $\ln(\text{GDP})$ - 较低;
- △: $\ln(\text{GDP})$ - 中等;
- ○: $\ln(\text{GDP})$ - 较高。
- 红色: $\text{ER} _ (\ln\text{GTI}, \ln\text{IS})$;
- 黄色: $\text{ER} \rightarrow \ln\text{IS}$, 但 $\text{ER} _ \ln\text{GTI}$;
- 蓝色: $\text{ER} _ \ln\text{IS}$, 但 $\text{ER} \rightarrow \ln\text{GTI}$;
- 绿色: $\text{ER} \rightarrow (\ln\text{GTI}, \ln\text{IS})$ 。

ER $_ \ln\text{GTI}$: 表示 ER 对 $\ln\text{GTI}$ 的影响不显著



应用 2：Du et al. (2024, EE)

- Du, C., et al. (2024). Does manufacturing agglomeration promote green productivity growth in China? **Energy Economics**. [Link](#), [PDF](#), [Replication](#)
 - 🍎 课件：【..\Du_2024_EE】

研究背景与动机

- 中国制造业集聚（MA）显著增强，带来规模经济、知识溢出等外部性，提升了经济效率。但同时，环境污染和资源消耗问题日益突出。在“碳达峰、碳中和”目标下，如何平衡制造业集聚与绿色发展，成为重要议题。
- 绿色全要素生产率（GTFP）衡量经济增长与环境效率的协同，兼顾传统投入与污染排放，更突出技术进步和资源利用效率。
- 尽管理论上，制造业集聚（MA）可通过多种机制提升绿色全要素生产率（GTFP），但已有研究发现，集聚程度过高也可能产生**拥挤效应**（congestion effect），如环境容量紧张、企业过度竞争、基础设施超负荷，导致 GTFP 边际效应递减甚至为负。
- 因此，**制造业集聚对 GTFP 的真实影响是否为正，是否存在“适度集聚”的最优区间**，是一个亟需实证检验的重要议题。

核心问题

1. **非线性问题**: MA 对 GTFP 的影响是否随着经济发展水平的不同而发生变化? 例如, 在低 PGDP 地区可能具有促进作用, 而在高 PGDP 地区则可能引发拥挤效应, 导致边际效应递减甚至变负;
2. **异质性问题**: MA 对 GTFP 的影响是否因城市地理区位或人口规模的不同而异? 这涉及到政策制定时“因地制宜”的重要性;
3. **机制分解问题**: MA 对 GTFP 的正面影响是主要通过技术进步 (MLTECH) 传导, 还是通过管理效率提升 (MLEFFCH) 传导?

贡献

- **方法创新：**首次将 PLFC 用于研究 MA 与 GTFP 的关系。相比传统门槛模型，该方法具有如下优势：
 - 不需要预设门槛个数与具体数值；
 - 支持连续、非线性、非参数估计；
 - 可识别 MA 边际效应如何随 PGDP 连续变化而变化。
- **机制分析：**通过 GTFP 分解，区分 MA 对技术进步（MLTECH）和效率改进（MLEFFCH）的作用路径，有助于进一步明确政策应侧重于“扶持技术扩散”还是“优化管理效率”。
- **异质性探索：**在模型中引入交互项、门槛变量与函数系数，识别城市地理位置（东中西部）、人口规模（大中小城市）对 MA-GTFP 关系的调节效应。

模型四：部分线性函数系数面板模型（PLFC）

Note: 作者先用传统的交乘项设定、静态面板门槛模型和动态面板门槛模型，然后才开始使用 PLFC 模型。

$$y_{it} = g(u_{it})x_{it} + \beta'z_{it} + \alpha_i + \varepsilon_{it} \quad (1)$$

- y_{it} : 因变量，本研究为 $\ln GTFP_{it}$
- x_{it} : 核心变量，为 $\ln MA_{it-1}$
- z_{it} : 控制变量向量，包括 IS、STI、GC、HC、SE 等
- u_{it} : 调节变量，本研究中为归一化的 $\ln PGDP_{it-1}$
- $g(u)$: 未知光滑函数，表示 x_{it} 对 y_{it} 的边际效应随 u_{it} 变化的函数
- α_i : 城市固定效应； ε_{it} : 误差项

u_{it} 归一化的目的是将 u_{it} 映射至 $[0, 1]$ 区间，方便后续非参数估计，公式为：

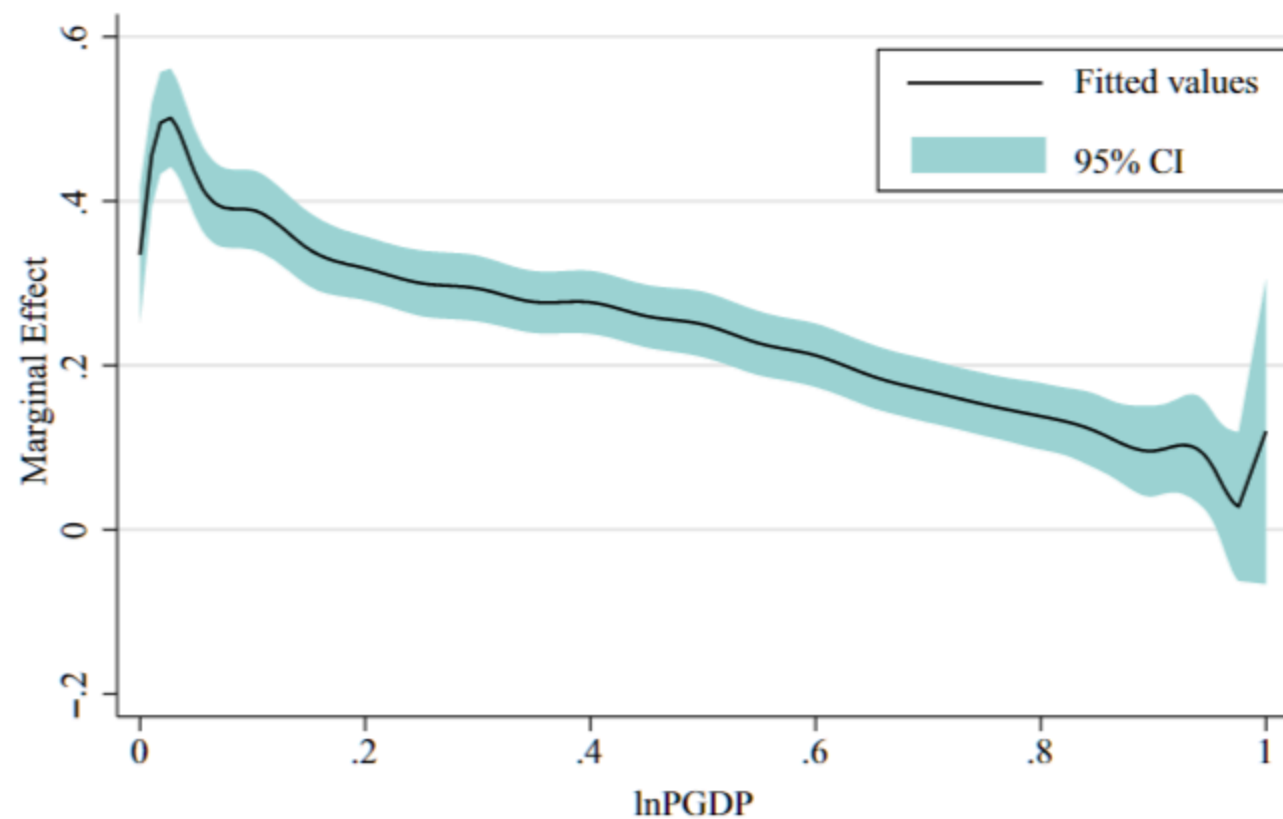
$$u_{it} = \frac{nPGDP_{it-1} - \min_i nPGDP_{it-1}}{\max_i nPGDP_{it-1} - \min_i nPGDP_{it-1}} \quad (2)$$

其中， $\ln PGDP_{it-1}$ 为人均 GDP 的自然对数值。

实证结果（PLFC 模型）

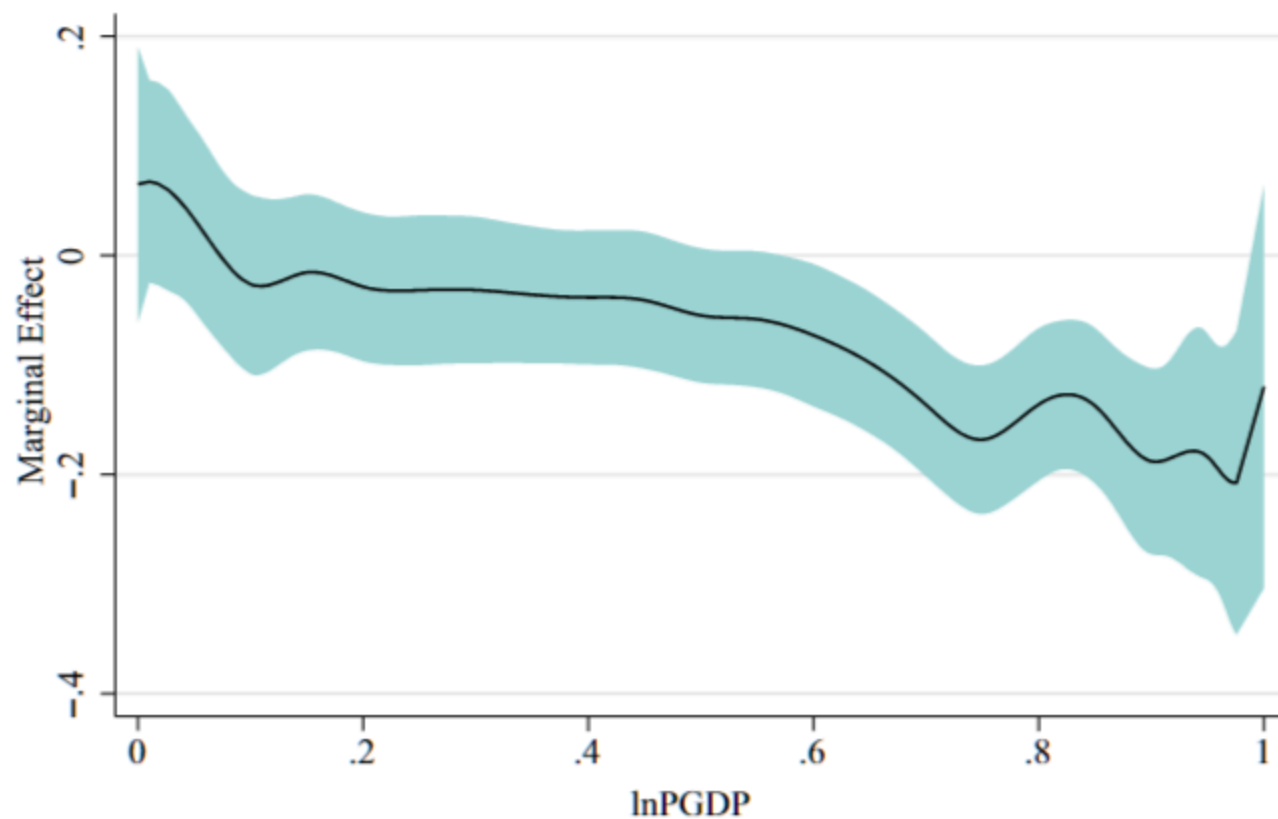
- MA 对 GTFP 的边际效应在低 PGDP 时为 0.4 左右，随着 PGDP 增加逐步下降，最终逼近 0；
 - 说明在经济发展尚处初期阶段，集聚效应显著，但随着经济增长，拥挤、竞争、资源瓶颈等负面因素逐步显现，抵消正效应；
- 将 GTFP 分解为 MLTECH（技术进步）与 MLEFFCH（效率提升）后发现：
 - MA 对 MLEFFCH 的影响在高 PGDP 区间为负，表明高集聚可能引发管理效率下降。
 - MA 对 MLTECH 的正向作用稳定；

A Functional Coefficient Estimation of GTFP



MA 对 MLEFFCH 的影响在高 PGDP 区间为负，表明高集聚可能引发管理效率下降。

B. Functional Coefficient Estimation of MLEFFCH



MA 对 MLTECH 的正向作用稳定

C. Functional Coefficient Estimation of MLTECH

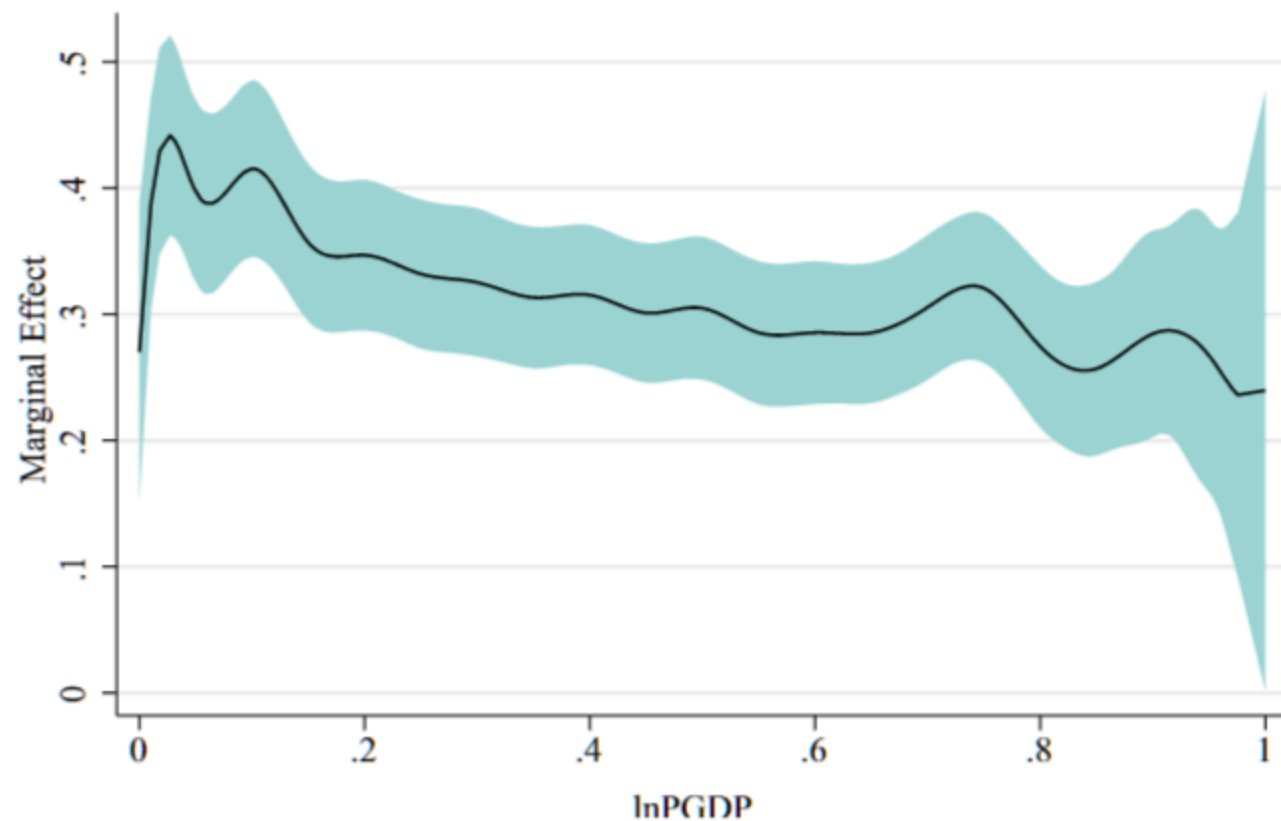


Fig. 5. Functional coefficients estimates.

5. 小结

- 总体而言，PLFC 模型是最灵活的，也最具解释性，更新后也支持「非平行面板」
- 上述模型本质上都可以用传统的交乘项设定来代替
- 故事主角： $x \rightarrow y, U$
 - y 可以是多个 (从不同角度衡量 Outcome)
 - U 也可以是多个 (可以作为论文机制分析的一个重要手段)
- AI 助手：你要怎么问？

Thanks

lianxh.cn