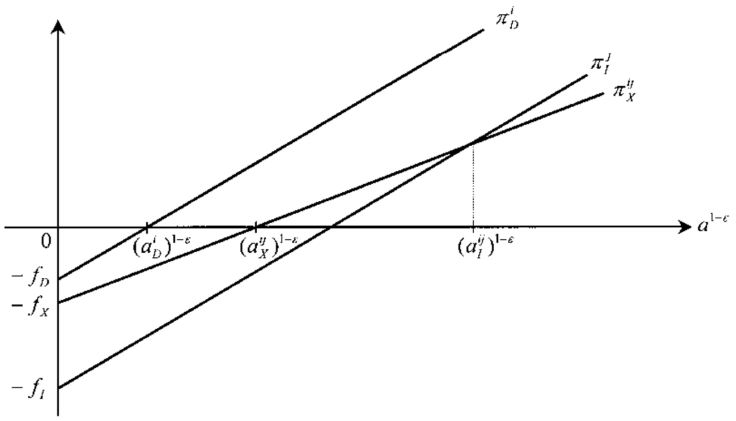
Stata Exercises

Due Nov. 18, 2026

第二题：面板二项选择（选做）

HMY (2004) 在 Melitz (2003) 异质性企业模型基础上引入 FDI 选项，其核心思想是：企业在生产率上存在异质性，企业可以选择只服务国内市场、出口，或通过对外直接投资在国外生产并服务当地市场；出口需要支付可变贸易成本 $\tau \geq 1$ 和固定出口成本 f_X，FDI 需要支付更高的固定成本 f_I 但可以节省贸易成本。因此，只有生产率最高的企业才会选择 FDI，生产率居中的企业选择出口，生产率较低的企业只服务国内市场：

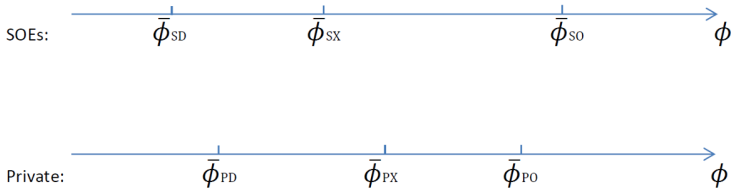
$$\phi_D^* < \phi_X^* < \phi_I^*$$



Chen et al. (2019) 在 HMY 基础上做了两个关键扩展：第一，在生产函数中加入资本（或土地）投入；第二，引入跨国的非对称制度扭曲。具体而言，私有企业在中国国内生产时面临更高的资本租金和土地价格，设其国内投入价格是国有企业的 $c > 1$ 倍，但企业到海外生产时，面对的是东道国要素市场，不再受中国国内信贷和土地市场歧视，于是，私有企业通过对外直接投资可以“制度套利”（institutional arbitrage），绕开国内投入品扭曲。其结果是：私有企业在国内生产的利润被压低，国内生存和出口门槛更高；但 FDI 相对于出口的额外收益对私有企业更大，因为出口仍然使用受扭曲的国内投入，而 FDI 使用国外投入。因此，私有企业进入 FDI 的临界值可能低于国有企业：

$$\bar{\phi}_{SO} > \bar{\phi}_{PO}$$

即出现“选择反转”（selection reversal）：私有跨国公司平均生产率低于国有跨国公司，但私有企业成为跨国公司的比例反而更高；同时，控制生产率、规模、出口等企业特征后，国有企业进行对外直接投资的概率更低。



Econometric Model Specification

原文并未给出计量模型的具体表达式，因为被解释变量 FDI_{it} 是一个二值变量，分别采用线性概率模型（LPM）、Logit 模型和 Complementary Log-Log 模型进行估计。

1. Linear Probability Model

$$FDI_{it} = \alpha + \beta SOE_i + \gamma' X_{it} + \delta_j + \lambda_t + \varepsilon_{it}$$

其中：

- FDI_{it}：企业 i 在年份 t 是否已经开展对外直接投资；
- SOE_i：企业是否为国有企业，基本不随时间变化；
- X_{it}：企业层面的控制变量，包括全要素生产率、企业规模和出口情况；
- δ_j：行业固定效应；
- λ_t：年份固定效应；
- ε_{it}：随机误差项。

2. Logit Model

企业开展对外直接投资的条件概率为：

$$P(FDI_{it} = 1 \mid SOE_i, X_{it}) = p_{it}$$

Logit 模型的概率函数为：

$$p_{it} = \Lambda\left(\alpha + \beta SOE_i + \gamma' X_{it} + \delta_j + \lambda_t\right)$$

其中，$\Lambda(z)$ 为 Logistic 累积分布函数：

$$\Lambda(z) = \frac{e^z}{1 + e^z}$$

进一步地，Logit 模型可以表示为：

$$\ln\left(\frac{p_{it}}{1 - p_{it}}\right) = \alpha + \beta SOE_i + \gamma' X_{it} + \delta_j + \lambda_t$$

左侧表示事件发生的对数优势（log odds）。

3. Complementary Log-Log Model

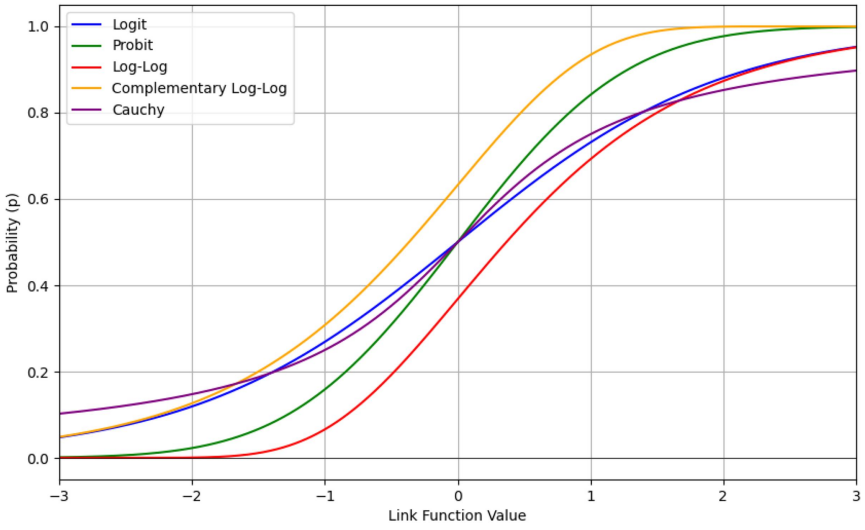
Complementary Log-Log（简称 Cloglog，互补对数-对数模型）也是一种用于二值因变量的非线性概率模型，与 Logit 和 Probit 模型一样，可以用于分析某一事件发生的概率。

Cloglog 模型的概率函数为：

$$\boxed{p = 1 - \exp\left\{-\exp\left(z\right)\right\}}$$

对概率函数进行变换，可以得到 Cloglog 模型的连接函数（link function）：

$$\boxed{\ln[-\ln(1-p)] = z}$$



Stata Programming Requirements

原文没有提供原始数据。请根据以下变量定义编写 Stata 代码，完成实证分析。无需实际运行代码，但代码应当符合 Stata 编程规范，并准确反映上述模型设定。

变量名	含义	用途
`fdi_indicator`	企业是否进行对外直接投资（1= 是，0= 否）	被解释变量
`soe_narrow`	企业是否为国有企业（1= 国有，0= 民营）	核心解释变量
`soe_broad`	广义国企定义：国有股比例 $\geq 50\%$	列 (6) 替代 `soe_narrow`
`tfp`	企业全要素生产率	控制变量，纳入 `$controls`
`ln_labor`	企业员工人数对数	控制变量，纳入 `$controls`
`export_indicator`	企业是否出口（1= 是，0= 否）	控制变量，纳入 `$controls`
`foreign_firm`	企业是否为外商投资企业（1= 是，0= 否）	样本筛选：列 (1)(2) 保留；列 (3) 起使用 `if foreign_firm == 0`
`tax_haven`	是否流向避税天堂目的地（1= 是，0= 否）	列 (7)：`if tax_haven == 0`
`distr_fdi`	是否为 distribution-oriented FDI（1= 是，0= 否）	列 (8)：`if distr_fdi == 0`
`soe_switcher`	是否为 SOE 身份发生切换的企业（1= 是，0= 否）	列 (9)(10)：`if soe_switcher == 0`
`ma_deals`	是否为并购型 FDI 的企业（1= 是，0= 否）	列 (10)：`if ma_deals == 0`
`ind`	2 位行业代码	行业固定效应
`year`	年份	年份固定效应
`firmid`	企业识别码	聚类层级

Questions:

线性概率模型

(a) 使用 Stata 的 `reghdfe` 命令估计 LPM。

(b) 解释表 6 中 SOE 的估计系数。

(c) LPM 有哪些缺点?

(d) 论文为什么加入行业和年份固定效应而没有使用企业固定效应?

Logit 模型与边际效应

(a) 使用 Stata 的 `xtlogit` 命令估计 Logit 模型。

(b) 与 LPM 相比，Logit 模型有哪些优点？如果换用 Probit 模型，Logit 和 Probit 模型的系数符号和边际效应符号是否一致?

(c) 解释表 6 中 SOE 的估计系数。

(d) 计算所有解释变量的平均边际效应（APE），并报告模型几率比（Odds Ratio）。

稀有事件 Logit[1]

King & Zeng (2001) 指出，当因变量中取值为 1 的事件非常少时，传统 Logit 模型可能存在稀有事件估计偏差。而在大多数年份，开展 FDI 的制造业企业比例不到 1%，因此，FDI 是一个典型的稀有事件 (rare events)。为此，作者使用 Rare Events Logit 对传统 Logit 估计进行修正。

(a) Rare Events Logit 的基本思想是什么？它与普通 Logit 模型有什么区别?

(b) 使用 Stata 的 `relogit` 命令估计 Rare Events Logit 模型。

(c) 比较普通 Logit 和 Rare Events Logit 中 SOE 的系数。为什么 Rare Events Logit 的 SOE 系数绝对值可能更大?

(d) Rare Events Logit 的系数是否可以直接解释为概率变化？能否计算边际效应?

[1]Rare Events Logit 处理的是结果变量是 0/1，但 1 很少 (FDI 企业占比不足 1%)，普通 Logit 会低估事件概率、把系数向 0 压缩。可以类比零膨胀泊松/负二项模型 (ZIP/ZINB) 处理的是结果变量是计数 (0,1,2,⋯)，但 0 特别多，且零可能来自两个不同过程：结构零 (企业从不对外投资)、抽样零 (企业可能投但恰好没投)。

Complementary Log-Log 模型

除了 Rare Events Logit，作者还使用 Cloglog 模型处理稀有事件。

(a) 写出 Logit 和 Cloglog 的模型形式，说明二者在连接函数上的区别。

(b) 为什么 Cloglog 模型适合处理事件发生概率较低、分布具有非对称性的情形?

(c) 使用 Stata 的 `cloglog` 命令估计 Cloglog 模型。

(d) 比较 Logit、Rare Events Logit 和 Cloglog 的估计结果。

Table 6. *Private Firms Are More Likely to Undertake FDI.*

Regressand:	LPM	Logit	Logit	Rare event Logit		Complementary log-log				
FDI indicator										
Year coverage:	2000–8								2004–8	
SOE defined:	*Narrow*	*Narrow*	*Narrow*	*Narrow*	*Narrow*	*Broad*	*Narrow*	*Narrow*	*Narrow*	*Narrow*
Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
SOE indicator	−0.002**	−0.454*	−0.757***	−1.306***	−0.693***	-0.682***	−1.179***	−0.532*	−0.703***	−0.662**
	(−2.41)	(−1.85)	(−2.88)	(−12.63)	(−2.81)	(−2.88)	(−3.26)	(−1.73)	(−2.68)	(−2.56)
Firm TFP	0.009***	1.333*	1.716**	4.237***	1.838**	1.843**	1.552*	1.603	2.360***	2.500***
	(4.14)	(1.80)	(2.00)	(18.50)	(2.25)	(2.26)	(1.66)	(1.28)	(4.25)	(4.96)
Log firm labour	0.003***	0.589***	0.623***	0.588***	0.587***	0.587***	0.565***	0.734***	0.574***	0.567***
	(6.55)	(10.85)	(9.78)	(38.49)	(8.86)	(8.86)	(7.86)	(8.21)	(9.37)	(11.03)
Export indicator	0.004***	0.900***	1.142***	1.102***	1.145***	1.145***	1.167***	0.736***	1.145***	1.174***
	(7.42)	(4.45)	(6.03)	(26.01)	(6.03)	(6.03)	(5.43)	(3.81)	(5.82)	(6.49)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign firms dropped	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tax haven dropped	No	No	No	No	No	No	Yes	No	No	No
Distr. FDI dropped	No	No	No	No	No	No	No	Yes	No	No
Switching SOE Dropped	No	No	No	No	No	No	No	No	Yes	Yes
M&A deals dropped	No	No	No	No	No	No	No	No	No	Yes
Observations	1,136,603	1,135,467	895,209	896,314	895,209	895,210	894,815	893,754	701,277	701,204

Notes: The regressand is the FDI indicator. All columns include industry-specific fixed effects and year-specific fixed effects. The numbers in parentheses are *t*-values clustered at the firm level. *** (**) denotes significance at the 1% (5%) level. Columns (1)–(2) include foreign-invested firms whereas all other columns drop those firms. Columns (1)–(8) cover data over the period of 2000–8 whereas columns (9)–(10) cover data over the period of 2004–8. Column (6) uses broadly defined SOE. Column (7) drops outward FDI to tax haven destinations. Column (8) drops distribution-oriented FDI (i.e., Distr. FDI). Column (9) drops the switching SOEs (i.e., switching from SOEs to private firms). Column (10) drops both switching SOEs and merger and acquisition deals. In all columns,TFP is measured by augmented Olley–Pakes controlling for input price distortion and SOE status.