

《智能制造背景下的生产函数估计偏误》附录

附录 1 蒙特卡洛模拟

为了评估本文提出的新的生产函数估计方法的表现，在此进行了蒙特卡洛模拟，以呈现出当企业进行内生技术选择时，该方法如何修正生产函数的估计偏差。附录首先描述了数据的生成过程，然后基于数据的可获得性，分别使用四种不同的方法估计生产函数，并对对比分析估计结果。

（一）数据生成过程

本节概述了蒙特卡洛模拟的数据生成过程。具体而言，给定一组感兴趣的真实参数（ θ 、 α 、 η 和 $\frac{b_K}{b_L}$ ），我们生成了 N 套包含 J 个企业和 T 个时期的模拟数据集。在每套数据中，我们为每个企业 i 在每个时期 t 均模拟生成包含生产率（ ω_{it} ）、异质性要素价格（ w_i 、 r_i ）和总产出（ Q_{it} ）的序列。具体生成过程描述如下。

假定企业 i 的（对数）生产率 ω_{it} 的变化服从一阶马尔科夫过程（即 AR（1）过程）：

$$\omega_{it+1} = g_0 + g_1\omega_{it} + \epsilon_{it+1}^\omega$$

其中， ϵ_{it+1}^ω 为服从正态分布的误差项，其均值为零、标准差为 $sd(\epsilon)$ ，满足独立同分布假设。企业 i 的初始（对数）生产率（ ω_{i0} ）从均值为 ω_0 、标准差为 $sd(\omega_0)$ 的正态分布中抽取。

存在企业异质性的劳动和资本投入价格（ w_{it} 和 r_{it} ）由如下步骤生成：¹

$$\begin{aligned} w_{it} &= w_t e^{\epsilon_{it}^L} \\ r_{it} &= r_t e^{\epsilon_{it}^K} \end{aligned}$$

其中，冲击 ϵ_{it}^L 和 ϵ_{it}^K 是对行业层面投入价格 w_t 和 r_t 的偏离。此外，本文假定冲击 ϵ_{it}^L 和 ϵ_{it}^K 服从均值为零、标准差为 $sd(\epsilon^L)$ 和 $sd(\epsilon^K)$ 的正态分布，满足独立同分布假设。

企业 i 的初始（对数）产出（ q_{i0} ）从均值为 q_0 、标准差为 $sd(q_0)$ 的正态分布中抽取。本文假定企业 i 的（对数）产出 q_{it} 的变化服从 AR（1）过程：

$$q_{it+1} = h_0 + h_1 q_{it} + \epsilon_{it+1}^q$$

其中， ϵ_{it+1}^q 为 $t+1$ 期实现的冲击。

为每个企业 i 在每个时期 t 模拟出 $\{\omega_{it}, Q_{it}, w_{it}, r_{it}, P_{Mit}\}$ 后，根据与企业静态成本最小化问题相关的一阶条件，可以得到每个企业 i 在时期 t 的最优的劳动、资本和中间品投入选择（ L_{it} 、 K_{it} 和 M_{it} ）。具体地，根据企业成本最小化问题的一阶条件，可以求得：

$$K_{it} = \left(\frac{1 - t^* r_i}{t^* w_i} \right)^{\frac{1}{t^*-1}} L_{it}^{\frac{t^*}{1-t^*}} \quad (A1)$$

$$M_{it} = \left((1 - t^*) \frac{P_{Mit}}{w_i} \right)^{\frac{1}{\theta-1}} \left(\frac{1 - t^* r_i}{t^* w_i} \right)^{\frac{t^*}{t^*-1}} L_{it}^{\frac{1-2t^*+2t^{*2}}{1-t^*}} L_{it}^{\frac{t^*}{1-\theta}} \quad (A2)$$

给定如正文式（1）所示的生产函数形式，分别用式（A1）和式（A2）替代资本 K_{it} 和中间品投入 M_{it} ，进而可以求得劳动 L_{it} 。之后，再将劳动 L_{it} 代入（A1）式和（A2）式便可得到资本 K_{it} 和中间品投入 M_{it} 。

通过上述方式，便为每个企业 i 在每个时期 t 生成了一套数据集 $\{\omega_{it}, Q_{it}, K_{it}, L_{it}, M_{it}, w_{it}, r_{it}, P_{Mit}, P_t\}$ 。附表 1 呈现了用于生成上述数据集的参数。

¹ 为便于分析，本文设定 $w_t = r_t = 0.1$ 。

附表 1 蒙特卡洛模拟参数设定

参数	参数描述	参数值
α	可替代范围	0.2
η	资本生产率的参数	15
b_K/b_L	资本的相对生产率	1.8
θ	生产函数参数	-0.4
g_0	生产率演化参数	0.2
g_1	生产率演化参数	0.95
h_0	产出演化参数	0.2
h_1	产出演化参数	0.95
$sd(q_0)$	初始产出的标准差（对数形式）	1
$sd(\omega_0)$	初始生产率的标准差（对数形式）	0.05
$sd(\varepsilon^\omega)$	生产率冲击的标准差	0.01
$sd(\varepsilon^q)$	产出冲击的标准差	0.01
w	工资的平均值	0.1
r	资本价格的平均值	0.1
P_M	中间品投入价格的平均值	0.1
$sd(w)$	工资的标准差	0.2
$sd(r)$	资本价格的标准差	0.2
$sd(P_M)$	中间品投入价格的标准差	0.2
$sd(P)$	产出价格的标准差	0.2
T	时间跨度	10
J	企业数量	100
N	蒙特卡洛模拟次数	1000

注：劳动、资本和中间品投入价格均为对数正态分布（log-normally distributed）。

（二）方法简介

假定研究人员可以观测到每个企业 i 在每个时期 t 的 $\{Q_{it}, M_{it}, K_{it}, L_{it}, w_{it}, r_{it}\}$ 。我们首先使用本文提出的新方法估计模型，如上所述，使用一阶条件来求解劳动 L_{it} 、资本 K_{it} 和中间投入 M_{it} 。之后，我们将利用本文方法得到的估计值与真实参数进行对比，以此评估新方法的表现。此外，本文还使用 Gandhi et al.（2020）提出的非参方法和 Levinsohn and Petrin（2003）与 Akerberg et al.（2015）提出的传统方法估计了要素投入份额，并将之与新方法的估计结果相比较，以此来揭示本文估计方法的合理性。

本文将 Gandhi et al.（2020）提出的非参方法作为主要比较对象，该方法沿用了已有研究中普遍采用的代理变量方法（proxy variable methods），同时对企业的一阶条件进行变换来识别总产出形式的生产函数（gross output production function）。本文将这种方法称为 GNR 法。此外，本文还将新方法与 Levinsohn and Petrin（2003）和 Akerberg et al.（2015）提出的传统方法进行比较¹。具体而言，传统方法使用控制函数法（control function approach），利用中间品投入作为代理变量来控制不可观测的生产率，以此解决生产函数估计中存在的内生性问题。本文将上述两种传统方法分别称为 LP 法和 ACF 法。

1 由于 Olley and Pakes（1996）使用投资作为代理变量来控制不可观测的生产率的内生性问题，而本文在蒙特卡洛模拟中并未生成投资数据，因此本文未使用 OP 法估计生产函数。

（三）蒙特卡洛模拟结果

附表 2 汇报了蒙特卡洛模拟的结果。对于每一种方法，表中呈现的结果均是 1000 次蒙特卡洛模拟估计值的平均值，括号内为标准误。由附表 2 中 Panel A 的结果可知，本文新方法很好地揭示了真实参数，误差几乎可以忽略不计，估计参数的标准差也非常小，模型可以很好地拟合数据。Panel B 中的结果表明，GNR 法对要素投入弹性的估计存在较大偏差，其对于劳动投入弹性的估计值偏高，同时也严重低估了中间品投入的弹性。这说明，尽管 Gandhi et al. (2020) 将生产函数设定为非参形式，即其对生产函数的形式并不敏感，但估计结果不尽如人意。同时，Panel C 和 Panel D 的结果显示 LP 法和 ACF 法两种传统方法的估计结果更为糟糕：LP 法估计得到的资本弹性为负值，而 ACF 法则更甚——劳动弹性和中间品投入弹性均为负值，这与现实大相径庭。出现这一结果的原因在于：Gandhi et al. (2020) 通过理论推导与蒙特卡洛模拟证实了 LP 法和 ACF 法在估计生产函数时需要足够大的企业投入价格的变化来源，其模拟结果表明当企业数量足够多且样本期足够长时 ($J = 500$ 且 $T = 50$)，二者的估计结果才更接近于真实值。基于本文的数据生成过程所设定的企业数量与样本时长，除了未考虑企业内生技术选择外，二者的估计偏差在一定程度上也源于此。

根据上述蒙特卡洛模拟结果可知，当企业内生地选择其生产技术时，传统方法和非参方法在生产函数估计时会产生较大的偏差，即使 GNR 法并不依赖于生产函数的特殊设定。由于本文从理论层面考虑了企业的内生技术选择，因此一定程度上避免了上述方法存在的问题，从而能准确地揭示真实参数。这为后续研究结果提供了可靠保证。

附表 2 蒙特卡洛模拟：不同方法的生产函数估计结果

Panel A	WW 法			
	α	η	b_K/b_L	θ
真实值	0.2	15	1.5	0.5
估计值	0.2	15	1.5	0.5
标准误 (s.e.)	(.0000)	(.0000)	(.0000)	(.0000)
Panel B	GNR 法			
	劳动	资本	中间品投入	加总
估计值	0.7775	0.1079	0.0483	0.9337
标准误 (s.e.)	(.0285)	(.0481)	(.0012)	(.0500)
Panel C	LP 法			
	劳动	资本	中间品投入	加总
估计值	0.3724	-0.0595	0.5287	0.9606
标准误 (s.e.)	(.0657)	(.0156)	(.0598)	—
Panel D	ACF 法			
	劳动	资本	中间品投入	加总
估计值	-0.0763	0.0205	-0.1401	-0.1959
标准误 (s.e.)	(.0524)	(.0159)	(.1047)	—

注：表 A.2 汇报了 N 次模拟的平均值。括号内为标准误。本文新方法在结果展示中简称为 WW 法。

附录 2 生产率分解方法

本小节量化了在位企业、进入企业和退出企业对生产率变化的相对贡献。参考 Melitz and Polanec (2015) 的思路，本文将生产率分解为四个部分：企业内部生产率的变化、企业内部市场份额的变化、进入企业生产率的变化和退出企业生产率的变化。

Olley and Pakes (1996) 将时期 t 的生产率分解为如下形式：

$$\Omega_t = \bar{\omega}_t + \sum_i (s_{it} - \bar{s}_t)(\omega_{it} - \bar{\omega}_t) = \bar{\omega}_t + cov(s_{it}, \omega_{it})$$

其中， $\bar{\omega}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \omega_{it}$ 为未加权的企业生产率平均值， $\bar{s}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} s_{it}$ 为市场份额的平均值。Melitz and Polanec (2015) 在此基础上加入了企业的进入和退出，以考察生产率的动态变化。令 $S_{Gt} = \sum_{i \in G} s_{it}$ 表示类别 G 的总市场份额， $\Omega_{Gt} = \sum_{i \in G} (s_{it}/S_{Gt}) \omega_{it}$ 表示类别 G 的总生产率。我们可以将每个时期的总生产率写成三类企业（在位企业 S 、进入企业 E 和退出企业 X ）各自的总份额和总生产率的函数：

$$\Omega_1 = s_{S1} \Omega_{S1} + s_{X1} \Omega_{X1} = \Omega_{S1} + s_{X1} (\Omega_{X1} - \Omega_{S1})$$

$$\Omega_2 = s_{S2} \Omega_{S2} + s_{E2} \Omega_{E2} = \Omega_{S2} + s_{E2} (\Omega_{E2} - \Omega_{S2})$$

基于此，便可写出生产率变化的表达式：

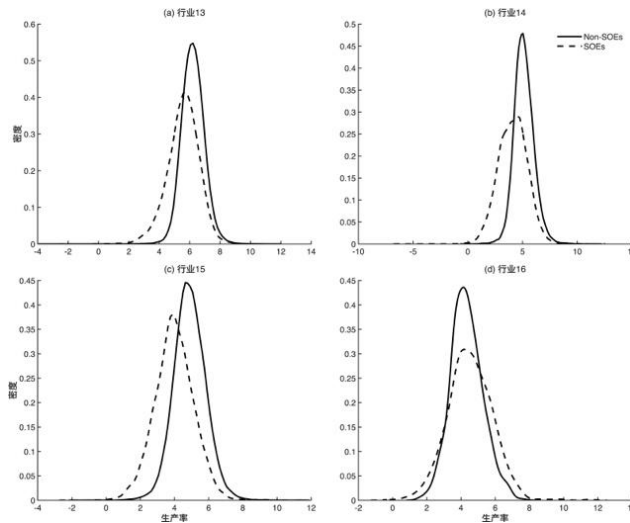
$$\Delta \Omega = (\Omega_{S2} - \Omega_{S1}) + s_{E2} (\Omega_{E2} - \Omega_{S2}) + s_{X1} (\Omega_{S1} - \Omega_{X1})$$

$$= \Delta \bar{\omega}_S + \Delta cov_S + s_{E2} (\Omega_{E2} - \Omega_{S2}) + s_{X1} (\Omega_{S1} - \Omega_{X1})$$

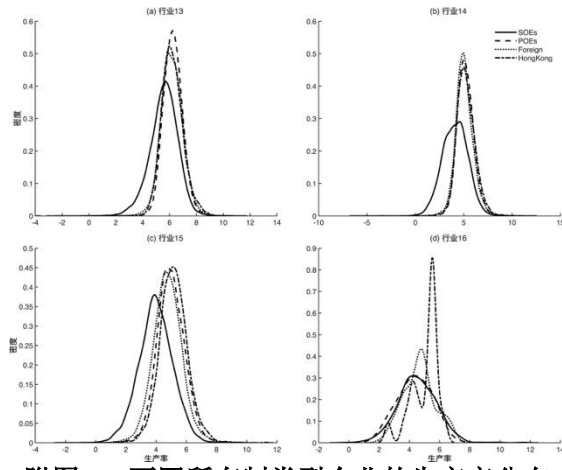
第一行将总生产率的变化分为三部分：在位企业、进入企业和退出企业。第二行应用了 Olley and Pakes (1996) 的分解方法，将在位企业的生产率变化分为两部分： $\Delta \bar{\omega}_S$ 为在位企业生产率均值的变化，刻画了企业内部生产率的变化（within component）； Δcov_S 为在位企业市场份额和生产率协方差的变化，刻画了企业内部市场份额的变化（reallocation component）。

附录3 进一步分析结果

(一) 不同所有制类型

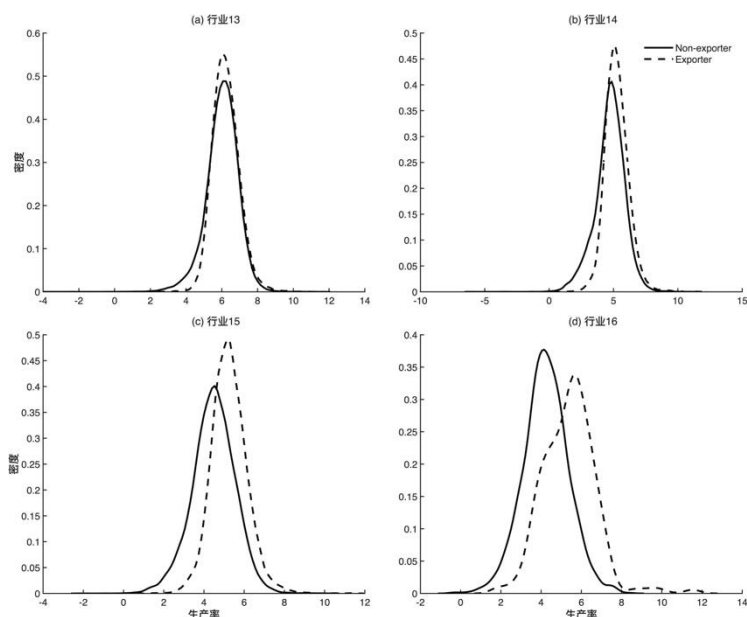


附图1 国有企业与非国有企业的生产率分布

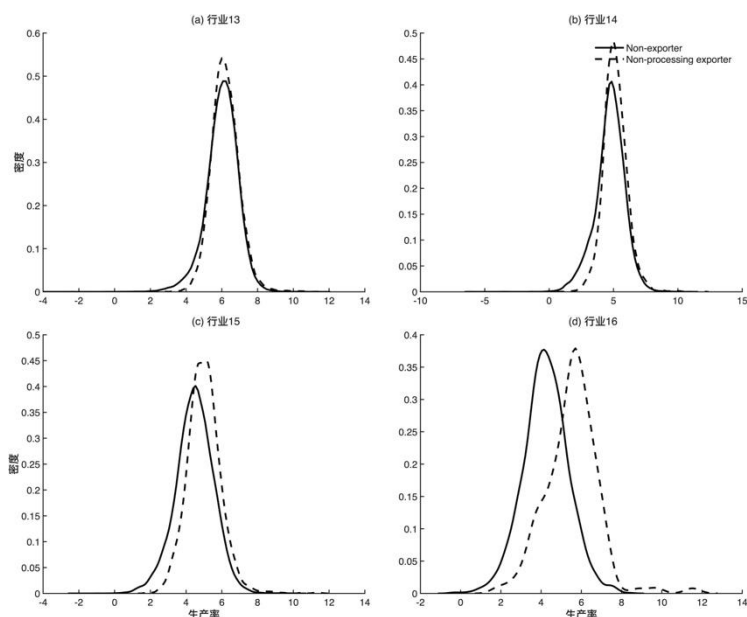


附图2 不同所有制类型企业的生产率分布

(二) 出口企业与非出口企业



附图3 出口企业与非出口企业的生产率分布（行业13-16）



附图4 非加工贸易出口企业与非出口企业的生产率分布

参考文献

- [1] Akerberg, D. A., K. Caves and G. Frazer, 2015, "Identification Properties of Recent Production Function Estimators", *Econometrica*, 83(6), pp.2411~2451.
- [2] Gandhi, A., S. Navarro and D. A. Rivers, 2020, "On the Identification of Gross Output Production Functions", *Journal of Political Economy*, 128(8), pp.2973~3016.
- [3] Levinsohn, J. and A. Petrin, 2003, "Estimating Production Functions Using Inputs to Control for Unobservables", *The Review of Economic Studies*, 70(2), pp.317~341.
- [4] Melitz, M. J. and S. Polanec, 2015, "Dynamic Olley-Pakes Productivity Decomposition with Entry and Exit", *The RAND Journal of Economics*, 46(2), pp.362~375.
- [5] Olley, G. S. and A. Pakes, 1996, "The Dynamics of Productivity in the Telecommunications Equipment", *Econometrica*, 64(6), pp.1263~1297.