附 录

正文未报告部分

附录 1：企业静态投入选择

企业每期选择的材料（静态投入），最大化利润：

$$\max_{L_{it},M_{it}} \Pi_{it} = R_{it} - \left(P_{Kt}K_{it} + W_t L_{it} + P_{Mt}M_{it}\right)$$

P_{Kt} 为标准资本的使用成本。根据需求函数（3）和柯布-道格拉斯生产函数（4），企业期初计划销售收入为：

$$R_{it} = \Phi^{\frac{1}{\eta_{it}}}\left(K_{it}^{\beta_K}L_{it}^{\beta_L}M_{it}^{\beta_M}\right)^{1-\frac{1}{\eta_{it}}} e^{\left(1-\frac{1}{\eta_{it}}\right)\left[\omega_{it}-\alpha(\delta_{Qit})\right]} e^{\delta_{Qit}} e^{\frac{\delta_{Hit}}{\eta_{it}}}$$

从而根据 M_{it} 的一阶条件（对数）：

$$\left(1-\frac{1}{\eta_{it}}\right)\omega_{it} = p_{Mt} - \frac{1}{\eta_{it}}\varphi - \ln\beta_M - \ln\left(1-\frac{1}{\eta_{it}}\right) + \left(\frac{1}{\eta_{it}}-1\right)\beta_K k_{it} + \left(\frac{1}{\eta_{it}}-1\right)\beta_L l_{it}$$
$$+\left(\frac{1}{\eta_{it}}\beta_M - \beta_M + 1\right)m_{it} + \left(1-\frac{1}{\eta_{it}}\right)\alpha\left(\delta_{Qit}\right) - \delta_{Qit} - \frac{1}{\eta_{it}}\delta_{Hit}$$

此即文中（6）。

附录 2：质量选择问题

对于生产函数 $Y_{it} = K_{it}^{\beta_K}L_{it}^{\beta_L}M_{it}^{\beta_M}e^{\omega_{it}-\alpha(\delta_{Qit})}$，给定劳动、材料投入的价格 W_t 和 P_{Mt}，企业选择等价劳动和原材料投入最小化短期总成本，以实现目标的产出 Y_{it}：

$$\max_{L_{it},M_{it}} C_{it} = W_t L_{it} + P_{Mt}M_{it}$$
$$st: Y_{it} = K_{it}^{\beta_K}L_{it}^{\beta_L}M_{it}^{\beta_M}e^{\omega_{it}-\alpha(\delta_{Qit})}$$

求解这一问题可得到短期总成本函数：

$$C_{it} = vc\left(K_{it},W_t,P_{Mt}\right)Y_{it}^{\frac{1}{\beta_L+\beta_M}} e^{-\frac{\omega_{it}}{\beta_L+\beta_M}+\frac{\alpha\left(\delta_{Qit}\right)}{\beta_L+\beta_M}}$$

$vc\left(K_{it},W_t,P_{Mt}\right) \equiv \left(K_{it}\right)^{-\frac{\beta_K}{\beta_L+\beta_M}}\left(W_t\right)^{\frac{\beta_L}{\beta_L+\beta_M}}\left(P_{Mt}\right)^{\frac{\beta_M}{\beta_L+\beta_M}}$ 为不依赖于企业产品质量决策的外生变量或状态变量的函数。根据需求函数（3），企业的产出量和质量选择问题为：

$$\max_{\delta_{Qit},Y_{it}} \Pi = \Phi^{\frac{1}{\eta_{it}}} e^{\delta_{Qit}+\frac{1}{\eta_{it}}\delta_{Hit}} Y_{it}^{1-\frac{1}{\eta_{it}}} - vc\left(K_{it},W_t,P_{Mt}\right)Y_{it}^{\frac{1}{\beta_L+\beta_M}} e^{-\frac{\omega_{it}}{\beta_L+\beta_M}+\frac{\alpha\left(\delta_{Qit}\right)}{\beta_L+\beta_M}}$$

根据 δ_{Qit} 的一阶条件

$$\Phi^{\frac{1}{\eta_{it}}} e^{\delta_{Qit}+\frac{1}{\eta_{it}}\delta_{Hit}} Y_{it}^{1-\frac{1}{\eta_{it}}} = \frac{1}{\beta_L+\beta_M} vc\left(K_{it},W_t,P_{Mt}\right)Y_{it}^{\frac{1}{\beta_L+\beta_M}} e^{-\frac{\omega_{it}}{\beta_L+\beta_M}+\frac{\alpha\left(\delta_{Qit}\right)}{\beta_L+\beta_M}}\alpha_{it}'$$

和 Y_{it} 的一阶条件：

$$\left(1-\frac{1}{\eta_{it}}\right)\Phi^{\frac{1}{\eta_{it}}} e^{\delta_{Qit}+\frac{1}{\eta_{it}}\delta_{Hit}} Y_{it}^{-\frac{1}{\eta_{it}}} = \frac{1}{\beta_L+\beta_M} vc\left(K_{it},W_t,P_{Mt}\right)Y_{it}^{\frac{1}{\beta_L+\beta_M}-1} e^{-\frac{\omega_{it}}{\beta_L+\beta_M}+\frac{\alpha\left(\delta_{Qit}\right)}{\beta_L+\beta_M}}$$

从而：

$$\alpha'\left(\delta_{Qit}\right)=\frac{\eta_{it}}{\eta_{it}-1}$$

附录 3：企业退出的样本选择问题

由于只能观测到持续经营的样本，退出企业生产信息的疏漏可能导致企业生产率估计的偏误。具体地，只可观测到持续经营企业的收入方程，即生产率演化过程（2）只刻画持续经营的企业

$$\omega_{it}=\gamma_t+g\left(\omega_{it-1},\theta_t=1\right)+\xi_{it}$$

借鉴 Olley 和 Pakes（1996）的思路，[1]设企业退出决策取决于其综合异质性水平 ω_{it} 。设企业当期在不低于门槛水平时（$\omega_{it}\geq\underline{\omega}_{it}$）继续经营（用示性函数表示为 $\theta_{it}=1$），否则退出。从而

$$g\left(\omega_{it-1},\theta_t=1\right)=E_t\left[\omega_{it}\left|\omega_{it-1},\theta_t=1\right.\right]=E_t\left[\omega_{it}\left|\omega_{it-1},\omega_{it}\geq\underline{\omega}_{it}\right.\right]=\phi(\omega_{it-1},\underline{\omega}_{it})$$

又由于企业 i 在第 t 期持续经营的概率可以表示为

$$\Pr_{it|t-1}=\Pr(\theta_t=1|\omega_{it-1},I_{t-1})=\Pr(\omega_{it}\geq\underline{\omega}_{it}|\omega_{it-1},I_{t-1})=\Psi_{t-1}(\underline{\omega}_{it},\omega_{it-1})$$

其中，I_{t-1} 为 t-1 期除了 ω_{it-1} 之外的所有信息集。由此企业门槛生产率水平 $\underline{\omega}_{it}=\Psi_{t-1}^{-1}(\Pr_{it|t-1},\omega_{it-1})$，将其带入到生产率的条件期望

$$\phi(\omega_{it-1},\underline{\omega}_{it})=\phi\left(\omega_{it-1},\Psi_{t-1}^{-1}(\Pr_{it|t-1},\omega_{it-1})\right)=f(\omega_{it-1},\Pr_{it|t-1})$$

从而，只需要在生产率的一阶马尔科夫过程中加入在 t-1 期对企业在下一期继续存在的概率预测，即可以解决因企业退出引起的样本选择问题。具体地，用如下模型估计企业的持续经营概率

$$\Pr_{it|t-1}=\Gamma(k_{it-1},l_{it-1},m_{it-1},h_{it-1})$$

然后用 $f(\omega_{it-1},\Pr_{it|t-1})$ 代替中（5）的期望生产率 $g\left(\omega_{it-1}\right)$。

附录 4：跨年匹配和面板构建

面板数据构建具体思路分为三步。

一、利用企业代码在任意两个年份间的进行尽量精确匹配。

在每年删除很少的企业代码重复现象后，企业共用代码的可能性很小了。但为了进一步控制可能的共用代码的错误，尽量保证代码相同的企业不论是纵向（时间）还是横向（截面）确实是同一企业，我们还要求代码相同的企业在企业名称、法人代表姓名、地址、电话号码、邮政编码、行业代码、开业年份等 7 项信息中至少一项完全相同。

二、对剩余的样本利用企业代码之外的信息在任意两个年份间的尽量广泛地匹配。

根据企业代码之外其它信息的跨年识别企业，并据此在这两个年份间企业法人代码间建立对应。

对于第一步没有在两个年份间建立匹配关系的样本，进行如下处理。首先，根据某组识别信息分别对两年的样本进行分组，即识别信息唯一的样本组和识别信息重复的样本组。其次，在这两年识别信息唯一的样本组间进行匹配，识别信息完全一样的企业被认为是同一企业，在同一企业两年的企业代码间建立对应，这样得到根据该组信息识别为同一企业的样本集。然后，将各年该类信息重复的样本与根据该类信息不能识别的样本合并在一起，再用另一组识别信息重复以上过程。最后，将各类信息识别的结果（包括第一步）合并在一起。

[1] Collard-Wexler and De Loecker （2015）在分析技术变化对美国钢铁业生产率以及再配置效应影响时，也沿用了这一方法处理门槛生产率。

关键是识别信息的选取。我们可能利用的企业代码之外的基本识别信息包括企业名称、法人代表姓名、企业所在省市县 6 位代码、企业地址、邮编、固定电话号码、行业代码、主要产品名称、开业年份等。然而在样本期间这些基本信息都可能发生变化。例如企业可能改变名称、法人代表可能更换、固定电话号码可能改变。而且，企业在填报这些基本信息时表现出一定的随意性，如企业名称有时填“XX 公司”、有时填“XX 有限责任公司”，企业有些年份填写从省市开始的详细地址，而有时只填写“XX 镇（街）XX 号”。单独使用某项基本信息识别效果很差。为此我们对基本信息进行合并、重组，运用检测程序找到识别结果完善的信息组合。

我们用于匹配的识别信息组合包括 25 个组合：企业名称+省市县 6 位代码的前 4 位（即同一省、地（市））；企业名称+开业年份的后两位数字（企业一般填写 4 位，如 1980，也有只填后两位的，如 80）；企业名称+邮编；企业名称+行业代码；企业名称+法人姓名；企业名称的后 9 个字符+开业年份+企业地址；企业名称的后 9 个字符+法人姓名+省市县 6 位代码的前 4 位；企业名称的后 9 个字符+省市县 6 位代码的前 4 位+固定电话号码；企业名称的倒数第 14-9 个字符+开业年份；企业名称的倒数第 12-5 个字符+法人姓名+省市县 6 位代码的前 4 位；企业名称的倒数第 12-5 个字符+固定电话号码；企业名称的倒数第 13-9 个字符+行业代码+邮编；企业名称的倒数第 12-5 个字符+行业代码+邮政编码；企业名称前 8 个字符+开业年份+邮编；企业名称前 8 个字符+法人姓名；企业名称前 8 个字符+固定电话号码；企业名称前 13 个字符+行业代码+邮编；法人姓名+固定电话号码；法人姓名+行业代码+开业年份+省市县 6 位代码；法人姓名+邮编+开业年份；法人姓名+邮编+行业代码+省市县 6 位代码；固定电话号码+行业代码+省市县 6 位代码+开业年份；固定电话号码+行业代码+邮编+开业年份；企业地址+行业代码+开业年份+省市县 6 位代码；企业代码的前 6 位+法人姓名+行业代码。

经过以上信息组合识别后，我们对两个年份间未匹配上的样本进行了程序检测和手工检测，发现根据已有信息基本上不能进行有效匹配。

三、构建面板，同时检验、修正利用企业代码之外的信息时可能发生的错配。

利用以上两年间匹配结果，可以构建任意时段间的不平衡面板数据。构建面板还可以进一步增加两两匹配效果，这是因为更充分地利用了两两匹配的信息。例如若某企业由于在样本期间发生了企业代码变更，进行两两匹配时，一些年份匹配上了，而一些年份没有匹配上。构建面板数据时不仅综合了两两匹配的直接信息，也利用了间接匹配信息，这就有可能更多的年份匹配。

构建面板过程还具有另外一个重要的作用，即纠正错配。这种错配源于在第二阶段两两直接匹配过程中运用非企业代码匹配。[2]虽然在第二阶段我们将错配的可能性降低到了最小，然而由于样本量巨大、企业个别情形复杂，错配还是有可能发生的。例如同一自然人可能拥有名称相似的多个企业，运用企业名称或其它信息匹配时就有可能出错。在构建面板时，如果某年两个不同的企业出现了相同的代码，就表明出现了错配。[3]具体地：

第一步，从相邻三年开始，在相邻的子区间内运用前向法和后向法形成核心变量，使之成为在截面维度和时间维度上都能识别企业的标志，然后根据核心变量分别构建相邻三年子区间内的非平衡面板。然后把所有相邻三年的子区间内非平衡面板的匹配信息合并到前一步

[2] 第一步已经尽量控制了根据企业代码匹配可能发生的错配。

[3] 一个错配的例子。某企业 1998 年企业代码为 101103682，1999 年其代码没有发生变化，所以用为企业代码能够与 1999 年的样本匹配上。根据法人代表姓名和固定电话号码，1998 年它也与 2000 年企业代码为 101135983 的企业匹配上了（但 1999 年与 2000 年没有匹配上）。然而在 1999 年、2000 年都有另外一个企业代码也为 101135983。这样在合并面板时在 2000 年就出现了两个企业的代码都是 101135983。本文仔细核对发现两个企业法人代表姓名和固定电话号码都相同，但确实不是同一个企业。

形成的整个研究区间内初步的非平衡面板中。

运用前（后）向法形成核心变量：首先，从每个企业出现的最近（早）一年开始，将当年的企业代码赋值给该变量；然后，利用这一年与上（下）一年企业代码间的对应关系，将能够匹配上的企业的该变量替代为上（下）一年的企业代码；重复这一过程，一直到样本区内的最早（近）年份。这样，我们把每个企业的核心变量都设定为其第一次（最后一次）出现时的企业代码。利用这个两个核心变量，构建相邻三年的不平衡面板，并检测每年是否出现企业代码重复，否则取消匹配信息。

第二步，逐步扩大子区间的长度，重复第一步，直至形成覆盖整个样本区间的面板。

第三步，利用两两直接匹配信息，对上一步形成的整个研究区间内的非平衡面板进行更新、补充。上一步形成的非平衡面板可能遗失部分直接匹配的信息，这一步将这些遗失的匹配信息重新找回。基础思路是先利用两两直接匹配信息构建排除了错配的整个样本区间面板，然后将其合并到第二步形成的面板中，最后再进一次错配检测。这样，我们就把所有直接和间接匹配的信息都整合到了整个研究区间内的非平衡面板中，而且尽最大可能减少了错配。

附录 5：行业定义

行业 1	13 农副食品加工业；14 食品制造业；15 酒、饮料和精制茶制造业；16 烟草制品业
行业 2	17 纺织业；18 纺织服装、服饰业；19 皮革、毛皮、羽毛及其制品和制鞋业
行业 3	20 木材加工和木、竹、藤、棕、草制品业；21 家具制造业
行业 4	22 造纸及纸制品业；23 印刷业和记录媒介的复制
行业 5	26 化学原料及化学制品制造业；27 医药制造业；28 化学纤维制造业；29 橡胶和塑料制品业
行业 6	30 非金属矿物制品业
行业 7	31 黑色金属冶炼及压延加工业；32 有色金属冶炼及压延加工业；33 金属制品业
行业 8	34 通用设备制造业；35 专用设备制造业
行业 9	36 汽车制造业；37 铁路、船舶、航空航天和其他运输设备制造业
行业 10	38 电气机械和器材制造业；39 计算机、通信和其他电子设备制造业；40 仪器仪表制造业

资料来源：作者整理。

附录 6：变量定义

资本存量 K_{it}。定义为标准质量资本。企业报告的固定资产原价合计是每一年名义投资的累加，一方面缺少反映投资价格逐年变化的信息，另一方面没有考虑折旧。本文借鉴 Brandtet al.（2012）的思路，采用永续盘存法估计企业实际资本存量。具体地，本文根据企业成立的时间不同分两种情况对资本和投资进行平滑，以得到企业实际资本存量。1998 年之前成立的企业，假设名义资产以稳定速度增长，计算出名义资本增长率。结合 1998 年（或之后任意一年）的名义资本存量反推出企业建立时的名义资本存量，通过固定资产投资价格指数，可以得到企业建立年的实际资本。最后利用永续盘存法，逐年计算企业的实际投资和资本存量。名义投资额在 1998 年之前根据名义资本增长率计算，1998 年之后由两年固定资产原价合计的差额进行计算。如果存在数据缺失情况，则用 1998 年之前数据的估算方法得到名义投资。这个过程中假设企业的折旧率为 9%。1998 年（含 1998）年之后成立的企业，建立年的固定资产原价合计即为企业建立年的名义资本存量，根据固定资产投资价格指数将其转化为实际资本之后，同样利用永续盘存法得到每年的实际资本存量。其中每年新增投资由两年的固定资产原价合计的差值得到，如果有数据缺失情况，计算平均增长率进行进一步

估算。投资价格平减指数包括两个来源：2006 年之前使用 Brandt-Rawski 投资价格指数；，2006 年之后用《中国统计年鉴》的全国固定资产投资价格指数。

材料支出 $P_{Mt}M_{it}$。与本文模型相关的材料支出强调企业生产阶段的可变成本，即生产过程的中间投入（直接材料+生产耗用的间接材料）。本文采用统一的方法重新估计样本期内的材料支出。具体地，按照会计处理规则，产品销售成本项目核算的是企业当年产品销售收入对应的直接材料、直接人工和制造费用。因此，材料支出等于产品销售成本扣除不属于材料投入的部分，即制造费用中包含的劳动者报酬（工资、福利费等）和折旧费。2005-2007 年有根据国家统计局定义的中间投入分项数据，包括制造费用中的中间投入、管理费用中的中间投入和营业费用中的中间投入。本文设企业劳动者报酬和折旧费在制造费用、管理费用和营业费用中分摊的比例是稳定的，利用 2005-2007 年数据中相应比例，估算产品销售成本中需要扣除的不属于材料投入的部分。

标准质量材料价格 P_{Mt}。来源于《中国统计年鉴》（2014）中《工业生产者购进价格指数》的总指数。这个指数以上一年为 100 给出，换算成以 1998 年定基的价格指数。本文考虑了 1998 年基数的省际差异，设定北京市为 100，其它省根据 1998 年各省 CPI 水平进行换算。

工资福利支出 W_tL_{it}。包括工资总额、职工福利费、劳动保险费、待业保险费、养老保险和医疗保险费、住房公积金和住房补贴。

标准质量工资 W_t。行业总工资除以行业全部从业人数（平均数）。

产出价格 P_t。来源于各年的《中国统计年鉴》中《按工业行业分工业生产者出厂价格指数》。由于年鉴中两位数行业出厂价格指数的行业定义从 2015 年才开始调整，本文按照《国民经济行业分类》（GB/T4754-2011）与《国民经济行业分类》（GB/T4754-2002）的对应将该指数调整为新的行业定义。

营销努力 sc_{it}。营业费用（产品销售费用）除以销售收入（**企业**报告的销售收入加上增值税销项税额）。

国有企业虚拟变量 soe_{it}。实收资本中国有股份大于 0，则 $soe_{it}=1$。本文还尝试了按登记注册类型定义国有企业，即注册登记类型为国有企业（110）、集体企业（120）、国有联营企业（141）、集体联营企业（142）、国有与集体联营企业（143）或国有独资公司（151），结果维持不变。

产出税率 τ_{it} **和材料税率** τ_{Mit}。根据增值税的会计处理，数据中企业报告的销售收入（主营业务收入）中不包含增值税销项税额，但包含营业税金及附加（包括消费税、营业税、城市维护建设税、资源税和教育费附加等）。增值税进项税也没有进入材料成本。注意到需求函数（1）中价格是消费者面临的价格，即包括增值税的价格，从而（4）中销售收入也应该包括消费者承担的增值税销项税额。因此，本文定义企业销售收入 R_{it} 为企业报告的销售收入加上增值税销项税额。又由于增值税通过产品销售和材料成本两个渠道影响企业的短期利润，本文定义[456]：

[4] 2009 年和 2013 年增值税销项税额和进项税额数据缺失。为减少样本损耗，我们分别用企业上年的相应数据替代。

$$\tau_{it}=\frac{\text{销项税额}}{\text{企业报告的主营业务收入+增值税销项税额}}$$

$$\tau_{Mit}=\frac{\text{进项税额}}{\text{企业材料支出+进项税额}}$$

产销率τ_{it}：

$$\lambda_{it}=\frac{\text{当期销售量}}{\text{当期生产量}}=\frac{\text{当期产品销售成本}}{\text{当期产品销售成本+期末产成品金额-期初产成品金额}}$$

附录 7：样本清理标准

一般清理。

-将如下变量的缺失值和负值调整为 0：产品销售费用、产品销售税金及附加、管理费用中列支的税金、劳动保险费、待业保险费、养老保险和医疗保险费、住房公积金和住房补贴；

-由于 2009 和 2013 年增值税销项税额、增值税进项税额数据缺失，将其调整为上年值；

-将如下情形设为缺失值：实际资本存量$K_{it}\le 0$、全部从业人数$L_{it}\le 0$、材料支出$P_{Mt}M_{it}\le 0$、工资福利总额$W_{it}L_{it}\le 0$、企业报告的销售收入（主营业务收入）小于等于 0、企业报告的销售收入（主营业务收入）小于等于产品销售税金及附加；

-删除当期进项税或销项税为小于或等于 0 的样本；

-删除进项税率大于 1（原材料小于进项税）的样本；

-删除当期进项税大于当期销项税的样本；

-最后，删除进项税率和销项税率的两端各 1%的野值。

特别清理。

首先，在如下情形将关键变量（实际资本）设为缺失值：

-全部从业人数（年平均）小于 8 人；

-实际资本$K_{it}\le 30000$、材料支出$P_{Mt}M_{it}\le 30000$、企业报告的销售收入（主营业务收入）小于 3 万元；

-企业报告的销售收入（主营业务收入）小于出口交货值；企业报告的将销售收入（主营业务收入）小于工资福利总额；企业报告的将销售收入（主营业务收入）小于材料支出；实收资本小于国家资本金、集体资本金与外商资本金之和；企业报告的将销售收入（主营业务收入）小于或等于产品销售费用；开业时间早于 1950 年；注册登记类型为国有企业（110）、集体企业（120）、国有联营企业（141）、集体联营企业（142）、国有与集体联营企业（143）、国有独资公司（151）。

-产销率最低和最高的 1%。

然后，删除任何一个变量为缺失值的观测，选择 1998-2013 年间最长的连续序列（大于

---

[5] 在参数和生产率估计时τ_{it}的分子为（销项税额 + 产品销售税金及附加）。

[6] 从 2009 年开始全面实施增值税从生产型转向消费型的改革，进项税额中不仅包括原材料对应的抵扣，还包括固定资产的抵扣。本文所用的数据无法将其分开，但在样本清理时尝试从以下几个方面缓解这个问题：第一，删除当期进项税或销项税为小于或等于 0 的样本；第二，删除进项税率大于 1（原材料小于进项税）的样本；第三，删除当期进项税大于当期销项税的样本；最后，删除进项税率和销项税率的两端各 1%的野值。另一方面，由于每期产销可能是不平衡的，增值税进项税额不完全是当期销售产品中对应的材料抵扣额，即存在留抵税款问题，对此本文用产销率进行调整，数据清理标准也一定程度上缓解了这一问题。

或等于 2 期）企业（时期）

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明引文和下载附件出处。

引用示例：

参考文献引用范例：

[1] 朱军. 技术吸收、政府推动与中国全要素生产率提升[J].中国工业经济.2017,(1):5-24.

如果研究中使用了未在《中国工业经济》纸质版刊发、但在杂志网站上正式公开发表的数字内容（包括数据、程序、附录文件），请务必在研究成果正文中注明：

数据（及程序等附件）来自朱军（2017），参见在《中国工业经济》网站（http://www.ciejournal.org）附件下载。