

参赛队号:

2021 年 (第七届) 全国大学生统计建模大赛

参赛学校: 华东理工大学

论文题目: 《基于 CES 与柯布-道格拉斯生产函数对数据动能的理论分析及五个制造业的相关数据分析》

参赛队员: 阎文彬、周明弘、方正

指导老师: 黄文亮, 郑祥祺

摘要

本文基于国内之前关于人力资本偏向性技术进步的动能研究工作,从 CES 生产函数出发,先对五个工业行业进行了数据分析,得到各个行业以 CES 生产函数为生产函数时的生产参数,并进一步得到了五个行业的资本增益效率和劳动力增益效率随时间的变化。

接着我们对这两个增益效率的数据内部结构尝试进行了理论分析,我们假设在一个行业对行业数据的使用以及对管理费用的使用都已较为成熟时,这两个增益效率函数是关于长期增益效率能力和短期增益效率能力的生产函数。并在假设该生产函数为柯布-道格拉斯生产函数的情况下,进行了理论上的模型讨论,通过对生产函数的内部参数的合理假设与公式的反解,得到了对长期增益效率能力的合理定义。最后利用该定义推导出了一个行业在满足三个模型假设情况下,该行业的数据动能定义式。

进一步,我们得到了从数据上判断是否满足那三个模型假设的具体判据,并对五个工业行业中满足该判据的两个行业——纺织业与纺织服装,服饰业,其增益效率的内部数据结构进行了数据上的分析,得到了这两个行业的长期增益效率能力、及其数据动能,从 2011 年 1 月到 2021 年 3 月间随时间的变化。

我们发现第一个行业的数据动能在 53.8%左右,第二个行业的数据动能在 34.5%左右。特别地,第一个行业的数据动能在逐年提升,并在 2017 年年底上升速度进一步提升;第二个行业在 2017 年之前数据动能在逐年提升,并在 2017 年年底开始下降。但是二者的数据动能在 2011 年 1 月至 2021 年 3 月期间并没有发生显著变化。

关键词: CES 生产函数 柯布-道格拉斯生产函数 长期增益效率能力 数据动能

目录

摘要.....	2
一、绪论.....	5
(一) 课题研究的背景.....	5
(二) 引入.....	5
1.理论的建立.....	5
2.对五个行业的具体数据分析.....	6
(三) 研究难点.....	6
二、符号说明.....	7
三、生产函数模型.....	7
(一) 生产函数.....	8
(二) 利润最大化模型.....	8
(三) CES 生产函数.....	9
(四) 估计方程.....	10
四、基于 CES 生产函数定义数据动能.....	11
(一) 对长期增幅能力 D_{jt} 的定义.....	11
(二) 对数据动能的定义.....	13
五、对五个行业的具体数据分析.....	14
(一) 对五个行业 CES 生产函数参数的测定和误差项的评估.....	15
1.五个行业主营成本和主营收入的时间序列.....	15
2.五个行业 $\ln(Q_{jt})$ 与 $\ln(M_{jt}), \ln(SM_{jt})$ 项的线性拟合.....	17
3.五个行业的 CES 生产函数生产参数具体数值.....	22
(二) 对资本增益效率与劳动力增益效率的数据结构分析.....	23
(三) 两行业数据动能随时间的变化.....	27
六、结论与建议.....	27
七、参考文献.....	29

附录.....30

表格和插图清单

表 1 符号说明.....7

表格 1.....20

表格 2.....22

表格 3.....25

表格 4.....26

一、绪论

（一）课题研究的背景

随着时代的进步与科技水平的不断发展，数据在逐步为我们人类创造出越来越多的价值。如今，世界各国也把推进经济数字化作为实现创新发展的重要动能，通过数据要素推动实体经济和数字经济融合发展，推动制造业加速向数字化、网络化、智能化发展。根据生产要素的重要性和时代性，数据生产要素成为了当代的关键要素。区别于以往生产要素的突出特点是，数据对其他要素资源具有乘数作用，可以放大劳动力、资本等生产要素在社会各行业价值链流转中产生的价值。

在过去的 10 年里，中国保持了很高的经济增长率，远远高于同期的其他国家。中国面临的一个重要问题是：这种增长在当代经济数字化的时代是可持续的吗？尤其是如今制造业加速向数字化、网络化、智能化发展的大背景下，制造业的增长是否能可持续化是前面这个问题中的一个很重要的组成部分。因此，确定中国制造业产值增长的数据动能及其变化就变得非常重要，这也是本文所主要探讨的问题。

（二）引入

1. 理论的建立

本文的理论部分基于参考文献[1]的工作，从 CES 生产函数出发，利用参考文献[1]中的方法，从可观测数据出发，求得了不可观测数据——资本增益效率和劳动力增益效率的公式，并对其中的误差项利用广义矩估计 (GMM) 进行估计。

进一步我们对这两个增益效率的数据内部结构尝试进行了分析。我们假设在一个行业对行业数据的收集和合理使用以及对管理费用的使用都已较为成熟时，这两个增益效率函数是关于长期增益效率能力和短期增益效率能力的生产函数。

接着，我们在假设该生产函数为柯布-道格拉斯生产函数的情况下，进行了理论上的模型讨论，其中最大的难点在于不可观测量过多，为解决这一难点，我们进行了如下处理。

由于该生产函数中的短期增益效率能力主要体现的是一个行业面对快速变化的市场时，决策层快速组织协调，以保证资金和劳动力尽可能地不被浪费的能力，我们采用类似于参考文献[2]中对资金投入，劳动力投入等利用资金具体金额，劳动力人数来进行代替的方法，我们利用一个行业的管理费用代替短期增益效率能力。

由于长期增益效率能力主要体现的，是该行业根据过去的大数据进行分析，进而对未来行情进行预测的能力，我们认为长期增益效率能力可近似认为是数据的收集与合理使用能力，但是这一能力很难通过观测量得到，因为目前国内尚确实可靠的对行业内数据收集与合理使用能力的评估方式，所以这实际上是一个不可观测量。这首先导致柯布-道格拉斯生产函数的三个生产参数中有两项都成为了无法具体计算得到的值，同时由于长期增益效率能力这一项本身无法具体求得，这使得资本增益效率与劳动力增益效率计算中的误差项也进一步不可解。

为了解决这一问题，我们首先发现柯布-道格拉斯生产函数剩下的两项无法确定的生产参数，仅仅只会对长期增益效率能力的数量级产生影响，并不会改变该数据除此以外别的数据结构，进而我们大胆通过对生产函数的内部参数进行合理假设，确定了这两个本来无法确定的参数。接着，我们将所有的生产参数代入柯布-道格拉斯生产函数公式，并进行反解，得到了长期增益效率能力，同时，再次通过广义矩估计 (GMM) 对资产增益效率能力与劳动力增益效率能力的误

差项进行估计。

在理论部分的最后，我们对行业产值对数的增长进行了分析，发现该增长中由五个子部分构成，资产投入增长项，中间投入增长项，劳动力投入增长项，短期增益效率能力增长项，以及长期增益效率能力增长项。自然地，我们利用最后一项与总增长的比值，定义了一个行业的数据动能定义式。该定义有以下前提假设：

- ① 其生产函数满足 CES 生产函数。
- ② 其对行业数据的收集和合理使用以及对管理费用的使用都已较为成熟。
- ③ 其资本增益效率和劳动力增益效率关于长期增益效率能力和短期增益效率能力满足柯布-道格拉斯生产函数。

2.对五个行业的具体数据分析

在理论建立完善后，我们对五个产业进行了具体的数据分析以验证我们的模型猜想，我们先后计算出了五个产业的生产函数参数，资本增益效率与劳动力增益效率，并对这五个产业的误差项利用广义矩估计（GMM）进行了检验，发现它们都很好地符合正态分布。

接着我们找到了一个合理的评判准则，来判断某一产业是否满足我们之前提到的最后两个前提假设，即其行业对行业数据的使用以及对管理费用的使用都已较为成熟，并满足其行业内部资本增益效率和劳动力增益效率关于长期增益效率能力和短期增益效率能力满足柯布-道格拉斯生产函数。事实上，这两点从数据的角度上是在描述同一件事情，即，真实数据并不很好地满足由柯布-道格拉斯生产函数所导出的线性条件（这一线性条件是建立于两个增益效率的对数与两个增益效率能力的对数之间的）。通过对目前已有数据的分析，我们初步判断该条件为实际数据和拟合数据差值的方差在 0.1 以内。

进一步，依据上述条件，我们找到了我们所分析的五个产业中，适合我们所有的三条模型假设的两个产业——纺织业与纺织服装，服饰业，并对他们的相关增益效率的内部数据结构进行了分析，最终得到了这两个产业的长期增益效率能力从 2011 年 1 月到 2021 年 3 月间随时间的变化。发现该数据相较于其数据的数量级随时间快速下降，其中，纺织服装，服饰业的下降速度相较纺织业更慢。

最后，我们根据理论部分所定义的数据动能定义式，分析了纺织业与纺织服装，服饰业的相关数据动能，得到其数据动能 2011 年 1 月到 2021 年 3 月期间的变化。发现第一个产业的数据动能在 53.8% 左右，第二个产业的数据动能在 34.5% 左右，同时第一个产业的数据动能在逐年微幅提升，并在 2017 年年底上升速度进一步提升，第二个产业在 2017 年之前数据动能在逐年微幅提升，在 2017 年年底开始微幅下降。但是总的而言，两个产业的数据动能在 2011 年 1 月至 2021 年 3 月期间并没有显著变化。

（三）研究难点

第一个难点在于在假设该生产函数为柯布-道格拉斯生产函数的情况下，进行了理论上的模型讨论时，不可观测量过多。为解决这一难点，我们做了如下处理。

由于该生产函数中的短期增益效率能力主要体现的是一个行业面对快速变化的市场时，决策层快速组织协调，以保证资金和劳动力尽可能地不被浪费的能力，我们采用类似于参考文献[2]中对资金投入，劳动力投入等利用资金具体金额，劳动力人数来进行代替的方法，因此，我们利用一个行业的管理费用代替短

期增益效率能力。而由于长期增益效率能力主要体现的，是该行业根据过去的大数据进行分析，进而对未来行情进行预测的能力，因此，我们认为长期增益效率能力可近似认为是数据的收集与合理使用能力，但是这一能力很难通过观测量得到，因为目前国内尚缺少可靠的对行业内数据收集与合理使用能力的评估方式，所以这实际上是一个不可观测量。这首先导致柯布-道格拉斯生产函数的三个生产参数中有两项都成为了无法具体计算得到的值，同时由于长期增益效率能力这一项本身无法具体求得，这使得资本增益效率与劳动力增益效率计算中的误差项也进一步不可解。

为了解决这一问题，我们首先发现柯布-道格拉斯生产函数剩下的两项无法确定的生产参数，仅仅只会对长期增益效率能力的数量级产生影响，并不会改变该数据除此以外别的数据结构，进而我们大胆通过对生产函数的内部参数进行合理假设，确定了这两个本来无法确定的参数。接着，我们将所有的生产参数代入柯布-道格拉斯生产函数公式，并进行反解，得到了长期增益效率能力，同时，再次通过广义矩估计 (GMM) 对资产增益效率能力与劳动力增益效率能力的误差项进行估计。

第二个难点在于在理论建立的最后部分，我们难以得到长期增益效率能力的具体数值。而我们通过对两个增益效率的生产函数的生产参数进行了一定的合理假设后得到了长期增益效率能力的具体数值。并且我们通过具体的公式推导证明了这样的假设是不会影响数据动能定义式的合理性，它所影响的仅仅只是长期增益效率能力的数量级，并不会影响到长期增益效率能力更重要的数据内部信息与结构，例如我们所要求的数据动能。即这样的假设也是合理的。

二、符号说明

表 1 符号说明

符号	意义
A_{jt}^k / A_{jt}	资本增益效率
A_{jt}^l / B_{jt}	劳动力增益效率
K_{jt}	资本投入
L_{jt}	劳动力投入
M_{jt}	中间投入
Y_{jt}	目标产出
Q_{jt}	实际产出（观测到的产出）
ε_{jt}	目标产出与实际产出之间的误差项
D_{it}	长期增益能力（数据的收集与合理使用能力）
O_{jt}	短期增益能力（管理投入）

三、生产函数模型

由于数据对其他要素资源具有乘数作用，可以放大劳动力、资本等生产要素在社会各行业价值链流转中产生的价值。所以为了研究数据对一个产业产量增长的贡献，正如参考文献[1],[2]中所述，对于一个行业的三个主要投入 资本投入，劳动投入而言，中间投入是最不容易产生增益效率的，真正容易产生增益效率的是资本投入和劳动力投入，而资本增益效率和劳动力增益效率对资本投入和劳动力投入是一种乘数作用，所以如果我们想要清楚认识数据动能，即数据对某一行业产量增速的贡献，我们就必须清楚认识数据对某一行业其他生产要素的乘数作

用。而对其他生产要素的乘数作用,我们通过参考文献[1],[2]可知主要是在考虑 A_{jt}, B_{jt} 。所以为了清楚认识数据动能,我们首先得清楚地知道应该采用哪种生产函数,某一行业资本增益效率和劳动力增益效率的表达式,我们所采用的生产函数有什么特征,以及我们所采用的生产函数中资本增益效率 A_{jt} 和劳动增益效率 B_{jt} 的具体表达式。并且我们认为某一行业应尽可能地使自己的产量最大化。

(一) 生产函数

首先,我们应该明确生产函数的特征,如果认为技术的变革会使得如资本,劳动力等相关要素朝着更好的方向发展,那么在这种情况下,要素效率的提高会以不同的方式改变不同要素的边际产量。由此,在允许技术以非常灵活的方式发生变革时,我们考虑发生资本增益和劳动力增益的技术变革时的生产函数。

假设 1(要素增益性): 技术的变革使得资本增加或者劳动力增加。

企业的生产函数被参数化为有限参数 θ 。在假设 1 下,生产函数取一般形式为:

$$Y_{jt} = F(A_{jt}K_{jt}, B_{jt}L_{jt}, M_{jt}; \theta)$$

其中 Y_{jt} 是企业的目标产出, A_{jt} 是资本增益效率, B_{jt} 为劳动力增益效率。资本效率-劳动力效率的比值为 $\frac{A_{jt}}{B_{jt}}$,它衡量技术偏向于资本或者劳动力的程度。 ω_{jt}, ν_{jt} 分别为资本增益效率和劳动力增益效率的对数,即 $\omega_{jt} = \ln A_{jt}, \nu_{jt} = \ln B_{jt}$ 。那么从现在开始,我们将引用 (ω_{jt}, ν_{jt}) 为资本效率和劳动力效率的对数, $\omega_{jt} - \nu_{jt}$ 为资本效率与劳动力效率比值的对数。再定义 K_{jt}, L_{jt} 和 M_{jt} 分别表示资本投入,劳动投入和中间投入。

进一步,我们可以假设生产函数满足以下条件:

假设 2 (可微性): $F(\cdot, \cdot, \cdot)$ 在它的三个分量上都是二阶连续可微的。

假设 3 (正边际产量): $F_1, F_2, F_3 > 0$, 其中 F_1, F_2, F_3 分别为 F 关于三个分量的偏导数

假设 4 (边际产量递减): $F_{ii} < 0, F_{ij} > 0$ 对所有 $i, j = 1, 2, 3, i \neq j$ 成立。

考虑到估计生产力测度时,存在一个问题在于不同企业的投入和产出的质量可能不同。企业生产力和投入质量都会影响投入效率的估计。尤其估计的生产力测度将同时包含企业所用技术的生产力和投入质量差异的影响。而价格包含了关于商品质量的重要信息。因此,本文遵循文献得知,可以使用投入(或产出)价格作为投入(或产出)质量的衡量标准。也即:我们假设价格的分散反映了企业之间投入和产出的质量差异,还假设企业为质量调整后的产品支付相同的价格,这样就解决了企业使用不同单位来记录其投入和产出的问题。通过用质量调整后的投入与产出来代替生产函数中的 (K_{jt}, L_{jt}, M_{jt}) 和 Y_{jt} ,从而去除了投入质量差异。而 A_{jt} 和 B_{jt} 是资本效率和劳动力效率,所以 A_{jt} 和 B_{jt} 现在衡量的是企业技术带来的效率,而不是投入质量。

(二) 利润最大化模型

对于前文所述的生产函数模型,我们利用关于劳动力和材料选择的一阶条件来帮助识别这个模型。企业在投入市场和产出市场中都是价格接受者。它们可能面临不同的投入和产出价格,这反映了不同企业投入和产出的质量差异。因此,同一时期所有企业的质量调整价格是相同的。我们假设资本在每个时期的开始都

是固定的，并且投资是动态选择的，以使公司价值最大化。我们还假设企业在选择劳动力和材料之前就知道自己的生产力水平（即资本效率和劳动效率）。

假设 5（利润最大化）：企业在各时期开始时观察自己的资本效率、劳动效率和资本存量，静态地选择劳动力和材料，使自己的期间利润最大化。

将 P_t, W_t, P_t^m 表示为质量调整后的产品的产出价格、工资率和材料价格。企业对劳动力和材料的最优静态决策问题表达式为：

$$\max_{L_{jt}, M_{jt}} \{P_t F(\exp(\omega_{jt}) K_{jt}, \exp(v_{jt}) L_{jt}, M_{jt}) - W_t L_{jt} - P_t^m M_{jt}\}$$

对应的一阶条件为：

$$\begin{aligned} \exp(v_{jt}) P_t F_2(\cdot) &= W_t \\ P_t F_3(\cdot) &= P_t^m \end{aligned}$$

其中 F_2, F_3 分别是 F 关于其第二个和第三个分量的偏导数。将第一个方程的两边同时乘以 $\frac{L_{jt}}{P_t Y_{jt}}$ ，第二个方程的两边同时乘以 $\frac{M_{jt}}{P_t Y_{jt}}$ ，再整理得到

$$\begin{aligned} \exp(v_{jt}) L_{jt} \frac{F_2(\cdot)}{F(\cdot)} &= \frac{W_t L_{jt}}{P_t Y_{jt}} \\ M_{jt} \frac{F_3(\cdot)}{F(\cdot)} &= \frac{P_t^m M_{jt}}{P_t Y_{jt}} \end{aligned} \quad (1)$$

我们假设观测到的产出 Q_{jt} 受独立同分布（i.i.d）测量误差 ε_{jt} 的影响。即 $Q_{jt} = Y_{jt} \exp(\varepsilon_{jt})$ ，其中 Y_{jt} 是公司的目标产出。将 S_{Ljt}, S_{Mjt} 分别表示为数据中观测到的劳动力和材料的收入份额。 S_{Ljt}, S_{Mjt} 的表达式为：

$$\begin{aligned} S_{Ljt} &= \frac{W_t L_{jt}}{P_t Y_{jt} \exp(\varepsilon_{jt})} \\ S_{Mjt} &= \frac{P_t^m M_{jt}}{P_t Y_{jt} \exp(\varepsilon_{jt})} \end{aligned} \quad (2)$$

将 (1) 中 $\frac{P_L L_{jt}}{P_t Y_{jt}}, \frac{P_M M_{jt}}{P_t Y_{jt}}$ 的表达式代入 (2) 中的等式得到

$$\begin{aligned} \exp(v_{jt}) L_{jt} \frac{F_2(\exp(\omega_{jt}) K_{jt}, \exp(v_{jt}) L_{jt}, M_{jt})}{F(\exp(\omega_{jt}) K_{jt}, \exp(v_{jt}) L_{jt}, M_{jt})} &= S_{Ljt} \exp(\varepsilon_{jt}) \\ M_{jt} \frac{F_3(\exp(\omega_{jt}) K_{jt}, \exp(v_{jt}) L_{jt}, M_{jt})}{F(\exp(\omega_{jt}) K_{jt}, \exp(v_{jt}) L_{jt}, M_{jt})} &= S_{Mjt} \exp(\varepsilon_{jt}) \end{aligned} \quad (3)$$

这个双方程体系中有两个未观察到的结构误差 ω_{jt}, v_{jt} 。我们可以从这个双方程体系中求解出 ω_{jt}, v_{jt} ，它们表示为未观察到的数据、非结构误差 ε_{jt} 和生产参数的函数。一阶条件还建立了效率比和替代弹性（由生产参数暗示）之间的联系。

我们想根据投入和产出的数据来估计生产函数中的所有参数。基本思想是从一阶条件中恢复未观测到的生产力。如果我们可以从一阶条件中唯一地恢复未观测到的真实生产力，则此限制可以帮助我们识别模型。在一阶条件下，有两个独立的方程和两个未知数，一般来说，我们可以求解出未知数。而参考文献[1]中的命题 1（可逆条件）则建立了我们可以反解一阶条件从而恢复未观察到的生产力的条件。接下来，我们将会介绍本文采用的 CES 生产函数，并说明在 CES 生产函数中恢复得到的 (ω_{jt}, ν_{jt}) 是唯一的。

（三）CES 生产函数

我们接下来将详细介绍 CES 生产函数以及资本增益效率与劳动力增益效率的表达式。

CES 生产函数由下式给出：

$$Y_{jt} = C[(A_{jt}K_{jt})^\gamma + (B_{jt}L_{jt})^\gamma + (M_{jt})^\gamma]^{\frac{s}{\gamma}}$$

其中 $A_{jt} = \exp(\omega_{jt})$, $B_{jt} = \exp(v_{jt})$, $s > 0$, 从而来衡量生产过程中的规模经济。我们假设企业在投入和产出市场都是价格接受者。它们面临不同的投入和产出价格, 这衡量投入和产出的质量。CES 的一阶条件是：

$$\frac{s \cdot L^\gamma B_{jt}^\gamma}{[(A_{jt}K_{jt})^\gamma + (B_{jt}L_{jt})^\gamma + (M_{jt})^\gamma]} = S_{Ljt} \exp(\varepsilon_{jt})$$

$$\frac{s \cdot M_{jt}^\gamma}{[(A_{jt}K_{jt})^\gamma + (B_{jt}L_{jt})^\gamma + (M_{jt})^\gamma]} = S_{Mjt} \exp(\varepsilon_{jt})$$

在 $s > 0, \gamma \neq 0$ 的限制下, 我们可以求解出资本效率和劳动效率的封闭解。

$$A_{jt} = \left(\frac{s - S_{Ljt} \exp(\varepsilon_{jt}) - S_{Mjt} \exp(\varepsilon_{jt})}{S_{Mjt} \exp(\varepsilon_{jt})} \right)^{\frac{1}{\gamma}} \frac{M_{jt}}{K_{jt}}$$

$$B_{jt} = \left(\frac{S_{Ljt}}{S_{Mjt}} \right)^{\frac{1}{\gamma}} \frac{M_{jt}}{L_{jt}}$$

将它们代入生产函数中, 取对数并将独立同分布 (i.i.d) 测量误差 ε_{jt} 添加到输出中, 得到估计方程：

$$\ln Q_{jt} = \left(\ln C + \frac{s}{\gamma} \ln s \right) - \frac{s}{\gamma} \ln S_{Mjt} + s \ln M_{jt} + \left(1 - \frac{s}{\gamma} \right) \varepsilon_{jt} \quad (4)$$

(四) 估计方程

在假设 1 下, 我们可以从关于劳动力和材料的一阶条件中恢复未观察到的多维生产率, 作为资本、劳动力、材料、劳动力份额、材料份额和生产参数的函数。将解表示为：

$$\omega_{jt}^* = \omega(K_{jt}, L_{jt}, M_{jt}, S_{Ljt}, S_{Mjt}, \varepsilon_{jt}; \theta)$$

$$v_{jt}^* = v(K_{jt}, L_{jt}, M_{jt}, S_{Ljt}, S_{Mjt}, \varepsilon_{jt}; \theta) \quad (5)$$

将生产函数中的 ω_{jt} 与 ν_{jt} 替换为等式 (5) 得到估计方程：

$$Y_{jt} = F(\exp(\omega_{jt}^*) K_{jt}, \exp(v_{jt}^*) L_{jt}, M_{jt})$$

其中 $\omega_{jt}^* = \omega(K_{jt}, L_{jt}, M_{jt}, S_{Ljt}, S_{Mjt}, \varepsilon_{jt}; \theta)$ 且 $v_{jt}^* = v(K_{jt}, L_{jt}, M_{jt}, S_{Ljt}, S_{Mjt}, \varepsilon_{jt}; \theta)$ 。 Y_{jt} 是公司的目标产出, 他没有被计量经济学家观察到。用带有测量误差的观测输出 (Q_{jt}) 替换它, 得出估计方程：

$$\ln Q_{jt} = \ln F(\exp(\omega_{jt}^*) K_{jt}, \exp(v_{jt}^*) L_{jt}, M_{jt}) + \varepsilon_{jt} \quad (6)$$

这是一个参数方程, 它在误差项 ε_{jt} 中是不可分的, 因为 ω_{jt}^*, v_{jt}^* 也包含 ε_{jt} 。识别条件要求对于任何数据点 ($Q_{jt}, K_{jt}, L_{jt}, M_{jt}, S_{Ljt}, S_{Mjt}$) 的任何生产参数 θ , ε_{jt} 由等式 (6) 唯一确定。即：

$$\varepsilon_{jt} = \varepsilon(Q_{jt}, K_{jt}, L_{jt}, M_{jt}, S_{Ljt}, S_{Mjt}; \theta).$$

在与命题 1 相同的条件下, 很容易证明这样一个函数的存在性。上面的例子表明, 在 CES 中, ε_{jt} 在生产函数中是唯一确定的。鉴于 ε_{jt} 可以由方程 (6) 唯一确定为 $\varepsilon_{jt} = \varepsilon(Q_{jt}, K_{jt}, L_{jt}, M_{jt}, S_{Ljt}, S_{Mjt}; \theta)$, 我们可以使用非线性最小二乘法 (NLLS) 或广义矩方法 (GMM)。矩条件如下：

$$E[Z'_{jt} \varepsilon_{jt}] = 0$$

其中 ε_{jt} 是输出中的测量误差, Z_{jt} 表示一组仪器变量。在估计中,我们选择 Z_{jt} 来包括 $\ln K_{jt}, \ln L_{jt}, \ln M_{jt}$ 的所有一阶和二阶项,包括所有交叉项。除此之外,我们还可以利用 $S_{Ljt} \exp(\varepsilon_{jt}) = \frac{W_t L_{jt}}{P_{jt} Y_{jt}}, S_{Mjt} \exp(\varepsilon_{jt}) = \frac{P_t^m M_{jt}}{P_{jt} Y_{jt}}, S_{Mjt} \exp(\varepsilon_{jt}) = \frac{P_t^m M_{jt}}{P_{jt} Y_{jt}}$ 均与 ε_{jt} 正交的事实来形成额外的矩。

四、基于 CES 生产函数定义数据动能

我们现在已经大致清楚了 A_{jt}, B_{jt} 在基于 CES 生产函数下的表达式,为了获得数据动能的合理定义,我们接下来需要先对 A_{jt}, B_{jt} 的内部数据结构进行适当猜想,并由此出发,定义数据的收集与合理使用能力的表达式,进而最终通过分析 Y_{jt} 的对数随时间的导数中对数据的收集与合理使用能力的贡献,得到最终的数据动能定义式。我们的讨论主要分为以下几个部分,先讨论我们模型的假设,再对 A_{jt}, B_{jt} 的内部数据结构的相关表达式进行分析,最后通过前两部分内容得到数据动能定义式。

首先,我们假定 A_{jt}, B_{jt} 来自一个对行业数据的使用以及对管理费用的使用都已较为成熟的行业,并且假定 A_{jt} 与 B_{jt} 都分别是关于短期增益效率能力 O_{jt} 与长期增益效率能力 D_{jt} 的函数,则 A_{jt}, B_{jt} 关于 O_{jt} 与 D_{jt} 的函数都分别满足生产函数条件。这是因为短期增益效率能力 O_{jt} 与长期增益效率能力 D_{jt} 的提升都会使得 A_{jt} 与 B_{jt} 增大,但增大速度分别随这两种能力的提升而减小,同时这两种能力相互配合提升时会使得 A_{jt} 与 B_{jt} 的增大速度增大,那么当 A_{jt}, B_{jt} 关于 O_{jt} 与 D_{jt} 都有一定光滑性时,如上原因分析则说明了 A_{jt}, B_{jt} 关于 O_{jt} 与 D_{jt} 的函数满足了生产函数的条件(即假设 A1~A4)。由此,我们可以提出假设 A6。

假设 A6: 当一个行业对行业数据的使用以及对管理费用的使用都已较为成熟时,则其长期增益效率能力主要由其对数据的收集与合理使用的能力决定,其短期增益效率能力主要由对管理投入的合理使用的能力决定,且 A_{jt}, B_{jt} 都是关于短期增益效率能力 O_{jt} 与长期增幅能力 D_{jt} 的生产函数。

基于假设 A6,我们之后将就三点展开讨论。

1. 如何评价一个行业对数据的使用是否已较为成熟
2. 如何得到 A_{jt} 与 B_{jt} 两个生产函数的生产参数并得到长期增益效率能力 D_{jt} (或者说数据的收集与合理使用能力)的表达式?
3. 如何得到数据的收集与合理使用能力对行业(企业)的产量变化的贡献。

其中第一点我们将在之后的具体数据分析中提出一个判断方法。而后两点我们将在接下来的理论模型提出中进行讨论。

(一) 对长期增幅能力 D_{jt} 的定义

对于 A_{jt}, B_{jt} 生产函数的建立以及关于短期增益效率能力 O_{jt} 与长期增幅能力 D_{jt} 的定义,我们接下来将利用柯布-道格拉斯生产函数作为 A_{jt} 与 B_{jt} 的生产函数。所以设

$$\begin{aligned} A_{jt} &= A \cdot D_{jt}^{\alpha_A} \cdot O_{jt}^{\beta_A} \\ B_{jt} &= B \cdot D_{jt}^{\alpha_B} \cdot O_{jt}^{\beta_B} \end{aligned}$$

其中 $\alpha_A, \alpha_B, \beta_A, \beta_B$ 为生产参数。取对数则得到

$$\ln A_{jt} = \alpha_A \cdot \ln D_{jt} + \beta_A \cdot \ln O_{jt} + \ln A$$

$$\ln B_{jt} = \alpha_B \cdot \ln D_{jt} + \beta_B \cdot \ln O_{jt} + \ln B$$

从而得到

$$\frac{d \ln A_{jt}}{d \ln O_{jt}} = \beta_A$$

$$\frac{d \ln B_{jt}}{d \ln O_{jt}} = \beta_B$$

进一步, 令

$$\begin{aligned} DA_{jt} &:= \ln A_{jt} - \beta_A \cdot \ln O_{jt} \\ DB_{jt} &:= \ln B_{jt} - \beta_B \cdot \ln O_{jt} \end{aligned} \quad (7)$$

从而得到

$$DA_{jt} = \ln A_{jt} - \beta_A \cdot \ln O_{jt} = \alpha_A \cdot \ln D_{jt} + \ln A$$

$$DB_{jt} = \ln B_{jt} - \beta_B \cdot \ln O_{jt} = \alpha_B \cdot \ln D_{jt} + \ln B$$

则得到

$$\ln D_{jt} = \frac{\ln A_{jt} - \beta_A \ln O_{jt}}{\alpha_A} = \frac{\ln B_{jt} - \beta_B \ln O_{jt}}{\alpha_B} \quad (8)$$

$$\frac{d(DB_{jt} - \ln B)}{d(DA_{jt} - \ln A)} = \frac{dDB_{jt}/d \ln D_{jt}}{dDA_{jt}/d \ln D_{jt}} = \frac{\alpha_B}{\alpha_A} \quad (9)$$

那么我们在假设 $\frac{\alpha_B}{\beta_B} = \frac{\beta_A}{\alpha_A}, A = 1$ 的条件下 (在之后的推导过程中我们会说明这样的假设是合理的), 代入上式得到

$$\frac{\alpha_B}{\alpha_A} \cdot \frac{d(DA_{jt} - \ln A)}{d(DB_{jt} - \ln B)} = \frac{\alpha_B}{\alpha_A} \cdot \frac{\alpha_A}{\alpha_B} = 1$$

但注意到在真实情况中, $\frac{\alpha_B}{\alpha_A} \frac{dDA_{jt}}{d(DB_{jt} - \ln B)}$ 的值可能存在一定误差。因此, 我们通过类似于之前在生产函数中对产出 Q_{jt} 受*i.i.d*测量误差 ε_{jt} 的影响, 即 $Q_{jt} = Y_{jt} \exp(\varepsilon_{jt})$ 的考虑, 同理不妨设

$$\frac{\alpha_B}{\alpha_A} \frac{dDA_{jt}}{d(DB_{jt} - \ln B)} = \exp(\eta_{jt})$$

并且类似于 ε_{jt} 的讨论, 我们认为 $E[\eta_{jt}] = 0$ 且 η_{jt} 为正态分布。而由等式 (6)

$$\ln D_{jt} = \frac{\ln A_{jt} - \beta_A \ln O_{jt}}{\alpha_A} = \frac{\ln B_{jt} - \beta_B \ln O_{jt}}{\alpha_B}$$

由于误差项 η_{jt} 的存在, 在误差允许范围内, 我们不妨取

$$\ln D_{jt} = \frac{1}{2} \left(\frac{\ln A_{jt} - \beta_A \ln O_{jt}}{\alpha_A} + \frac{\ln B_{jt} - \beta_B \ln O_{jt}}{\alpha_B} \right)$$

则得到长期增益效率能力 D_{jt} 的表达式为:

$$\begin{aligned} D_{jt} &= \exp \left(\frac{\ln A_{jt} - \beta_A \ln O_{jt}}{2\alpha_A} + \frac{\ln B_{jt} - \beta_B \ln O_{jt}}{2\alpha_B} \right) \\ (10) \quad &\approx \exp \left(\frac{\ln A_{jt} - \beta_A \ln O_{jt}}{\alpha_A} \right) \approx \exp \left(\frac{\log B_{jt} - \beta_B \ln O_{jt}}{\alpha_B} \right) \end{aligned}$$

特别地, 当 $\eta_{jt} \equiv 0$ 时, 它满足 (6) 式。

接下来，我们将说明假设 $\frac{\alpha_B}{\beta_B} = \frac{\beta_A}{\alpha_A}, A=1$ 的条件，这是合理的，主要有以下两个原因：

1. D_{jt} 本身就是一个不可观测的量，它的量级基于 A 与 α_A 。事实上，通过结合前面得到的关系式 (5)、(6)、(8) 考虑， A 的增减与 α_A 的增减同样会使得相应 B 与 α_B 的增减，最后只会影响到 D_{jt} 的量级而不会影响其数据结构（如大小关系等）。

2. 实际上，我们对 D_{jt} 的定义最终是想对数据动能进行合理的定义，而 A 与 α_A 的增减并不会影响我们随后得到的数据动能定义式所确定的数值，这一点在之后我们得到的数据动能定义式中可以被证明。

(二). 对数据动能的定义

根据对 CES 生产函数的介绍以及二中的讨论可得

$$\ln Y_{jt} = \frac{s}{\gamma} \ln ((A_{jt} K_{jt})^\gamma + (B_{jt} L_{jt})^\gamma + (M_{jt})^\gamma) + \ln C$$

$$\frac{d \ln Y_{jt}}{dt} = \frac{s}{\gamma} \cdot \frac{\gamma (A_{jt} K_{jt})^{\gamma-1} \cdot \frac{d(A_{jt} K_{jt})}{dt} + \gamma (B_{jt} L_{jt})^{\gamma-1} \cdot \frac{d(B_{jt} L_{jt})}{dt} + \gamma (M_{jt})^{\gamma-1} \frac{dM_{jt}}{dt}}{(A_{jt} K_{jt})^\gamma + (B_{jt} L_{jt})^\gamma + (M_{jt})^\gamma}$$

令

$$S_{jt}^k := \frac{s \cdot (A_{jt} \cdot K_{jt})^\gamma}{(A_{jt} K_{jt})^\gamma + (B_{jt} L_{jt})^\gamma + (M_{jt})^\gamma}$$

$$S_{jt}^l := \frac{s \cdot (B_{jt} L_{jt})^\gamma}{(A_{jt} K_{jt})^\gamma + (B_{jt} L_{jt})^\gamma + (M_{jt})^\gamma}$$

$$S_{jt}^m := s - S_{jt}^k - S_{jt}^l$$

于是得到

$$\frac{d \ln Y_{jt}}{dt} = S_{jt}^k \frac{d \ln (A_{jt} \cdot K_{jt})}{dt} + S_{jt}^l \frac{d \ln (B_{jt} L_{jt})}{dt} + S_{jt}^m \frac{d \ln M_{jt}}{dt}$$

而由 A_{jt} 与 B_{jt} 的生产函数可得

$$\frac{d \ln A_{jt}}{dt} = \alpha_A \frac{d \ln D_{jt}}{dt} + \beta_B \frac{d \ln O_{jt}}{dt}$$

$$\frac{d \ln B_{jt}}{dt} = \alpha_B \frac{d \ln D_{jt}}{dt} + \beta_B \frac{d \ln D_{jt}}{dt}$$

进一步得到

$$\begin{aligned} \frac{d \ln Y_{jt}}{dt} &= S_{jt}^k \frac{d \ln K_{jt}}{dt} + S_{jt}^l \cdot \frac{d \ln L_{jt}}{dt} + S_{jt}^m \cdot \frac{d \ln M_{jt}}{dt} + \\ &\quad (\alpha_A S_{jt}^k + \alpha_B \cdot S_{jt}^l) \frac{d \ln D_{jt}}{dt} + (\beta_A S_{jt}^k + \beta_B S_{jt}^l) \frac{d \ln O_{jt}}{dt} \end{aligned}$$

其中 $\frac{d \ln Y_{jt}}{dt}$ 即产出的增长可以由五个来源来解释。前三个来源是由于物理投入

（资本、劳动力和中间投入）的增长，这对应于等式中的前三项，而等式的后两项说明最后两个来源对应于长期增益效率能力与短期增益效率能力的增长。

我们以长期增益效率能力的增长与产值的增长的比值定义数据动能，表示长期增益效率能力对产值的变化所起到的贡献，即为下式：

$$\frac{\frac{d \ln D_{jt}}{dt} (\alpha_A S_{jt}^k + \alpha_B S_{jt}^l)}{\frac{d \ln Y_{jt}}{dt}}$$

可以观察到该定义式，它由长期增益效率能力 D_{jt} 与计划产值 Y_{jt} 的对数线性相关系数乘以资本的收入份额 S_{jt}^k 和劳动力的收入份额 S_{jt}^l 的线性组合来决定，其中长期增益效率能力 D_{jt} 与计划产值 Y_{jt} 的对数线性相关系数表示了一个长期范围内数据动能的基本值，而资本的收入份额 S_{jt}^k 和劳动力的收入份额 S_{jt}^l 的线性组合一方面表示了数据动能的短期周期性波动，另外一方面则表示了数据对资本和劳动力的作用能力（主要由 α_A, α_B 所表示）

最后，我们通过这个数据动能定义式证明（二）中的假设 $\frac{\alpha_B}{\beta_B} = \frac{\beta_A}{\alpha_A}, A=1$ 的合理性。将 $A=1$ 改为一般情形 $A = \tilde{A}$ 时，则得到改变后的新的长期增益能力 \tilde{D}_{jt} 满足

$$\frac{d \ln \tilde{D}_{jt}}{dt} = \frac{\frac{d \tilde{D}_{jt}}{dt}}{\tilde{D}_{jt}} = \frac{\tilde{A}^{-\frac{1}{\alpha_A}} \cdot \frac{d D_{jt}}{dt}}{\tilde{A}^{-\frac{1}{\alpha_A}} \cdot D_{jt}} = \frac{\frac{d D_{jt}}{dt}}{D_{jt}} = \frac{d \ln D_{jt}}{dt}$$

即更改前后 $\frac{d \ln D_{jt}}{dt}$ 保持不变。将 $\frac{\alpha_B}{\beta_B} = \frac{\beta_A}{\alpha_A}$ 改为一般情形 $\frac{a \cdot \tilde{\alpha}_B}{\beta_B} = \frac{\beta_A}{\tilde{\alpha}_A} (a > 0)$ 时得到改变后的新的长期增益能力 \tilde{D}_{jt} 满足：

$$\begin{aligned} \frac{d \ln \tilde{D}_{jt}}{dt} (\tilde{\alpha}_A S_{jt}^k + \tilde{\alpha}_B S_{jt}^l) &= \frac{1}{\sqrt{a}} \frac{d \ln D_{jt}^{\sqrt{a}}}{dt} (\alpha_A S_{jt}^k + \alpha_B S_{jt}^l) \\ &= \frac{d \ln D_{jt}}{dt} (\alpha_A S_{jt}^k + \alpha_B S_{jt}^l) \end{aligned}$$

即更改前后 $\frac{d \ln D_{jt}}{dt} (\alpha_A S_{jt}^k + \alpha_B S_{jt}^l)$ 保持不变。综合可知，我们做出

$\frac{\alpha_B}{\beta_B} = \frac{\beta_A}{\alpha_A}, A=1$ 的假设并不会改变得到的数据动能定义式所确定的数值。

五、对五个行业的具体数据分析

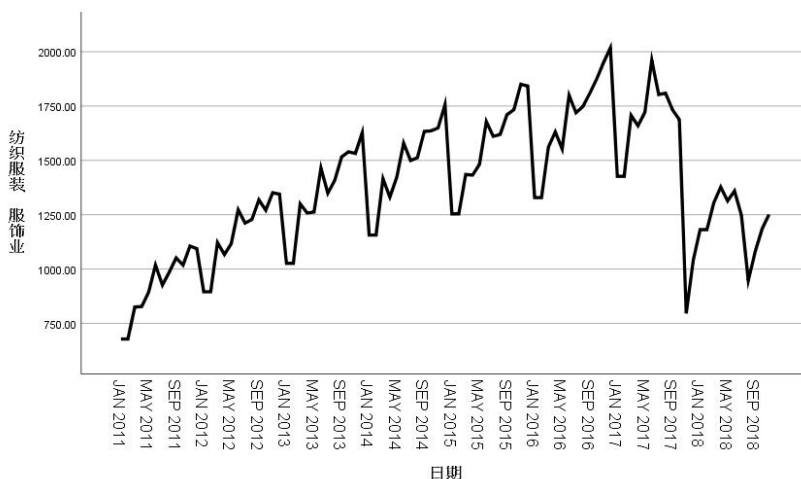
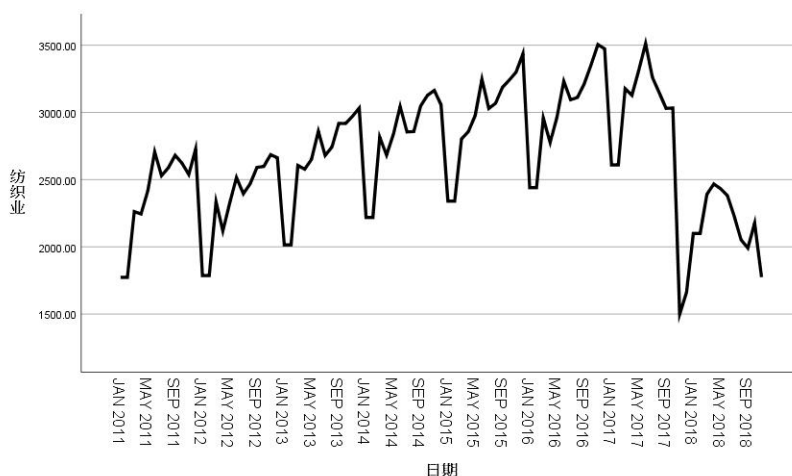
接下来，我们将基于上文的理论分析建立对五个产业的 CES 生产函数的生产参数测定，并对生产函数的误差项是否满足正态分布进行评估。进一步，根据反解出的资本增益效率和劳动力增益效率，利用理论部分对两个增益效率的内部数据结构分析，对五个行业的真实数据满足内部数据结构所导出的线性条件的好坏程度进行评价，依此找到一个合理的判据来判断某一产业是否满足两个前提假设，即其行业对行业数据的使用以及对管理费用的使用都已较为成熟，并满足其行业内部资本增益效率和劳动力增益效率关于长期增益效率能力和短期增益效率能力满足柯布-道格拉斯生产函数。进一步，通过对满足这个判据的两个行业的增益效率的生产函数的生产参数进行测定，得到了这两个行业的数据的收集和使用能力的具体数值和数据动能的具体数值。

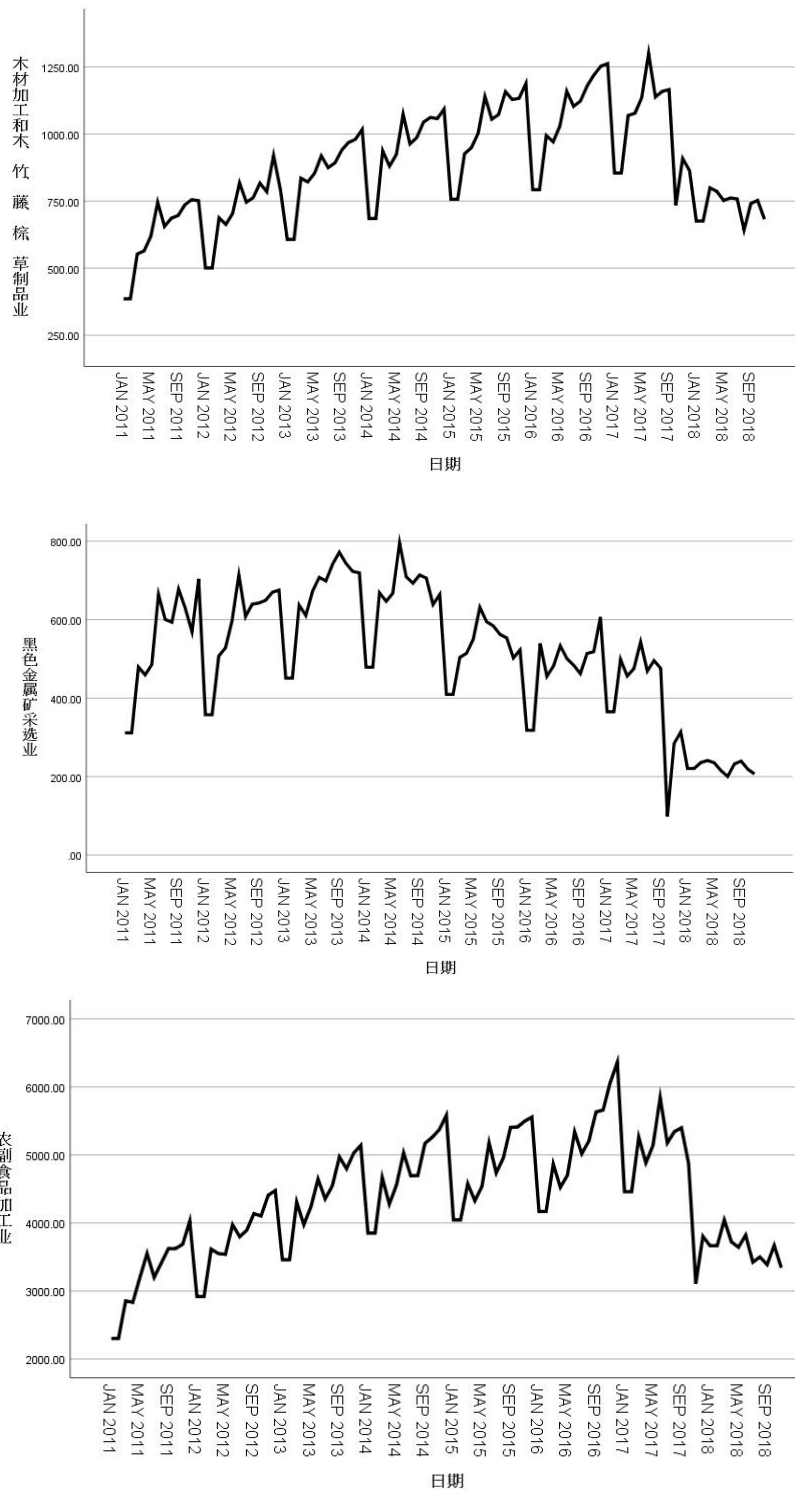
（一）对五个行业 CES 生产函数参数的测定和误差项的评估

1.五个行业主营成本和主营收入的时间序列

在程序设计中，我们主要通过对两个变量进行线性拟合、将拟合结果可视化并通过判断点的离散程度、和拟合直线的吻合程度等等得出两个变量的相关性及其数据的好坏。在以下所有数据处理中，我们经常采用取对数的形式来简化计算和美化可视化结果。通过查找国家统计局公布的数据，我们选取 2011 年 1 月到 2021 年 2 月作为本文研究的时间范围。注意到主营业务收入和主营业务成本这两个变量在国家统计局中搜索不到自 2019 年以来的数据，故我们首先利用 SPSS 的时间序列预测功能给出 2019 年 1 月到 2021 年 2 月这两个数据的大致估计。以下为五个行业数据的时间序列图。

主营业务成本：





经过回判，我们发现五个行业中的前三个的预测模型效果不佳，故本文对于这三个行业的数据仅采用能直接在国家统计局上下载的数据，而后两个行业，即纺织业和纺织服装、服饰业的数据包括能在国家统计局上下载的数据以及通过时间预测模型所推断的 2019 年到 2021 年 2 月的数据，预测数据如下图所示。

主营业务成本的预测数据：

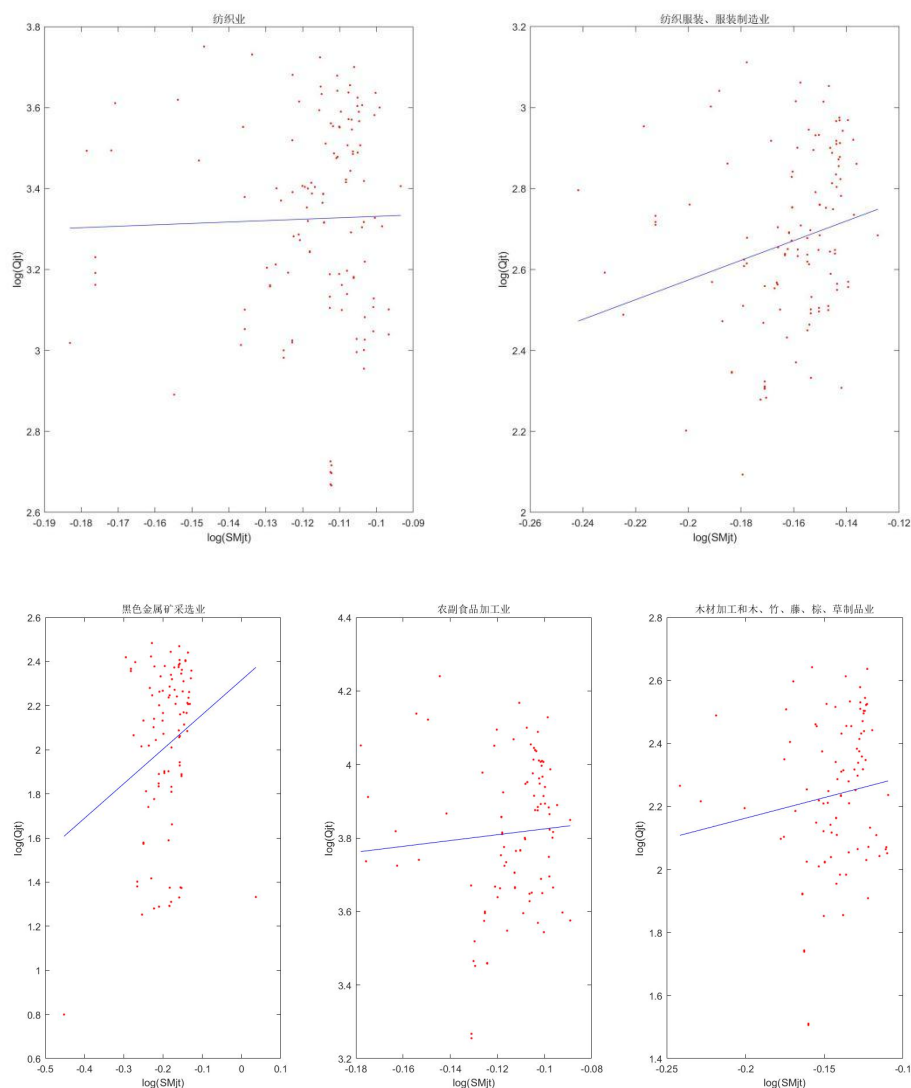
年	月	黑色金属矿采选业主营业务成本(亿元)	农副食品加工业主营业务成本(亿元)	纺织业主营业务成本(亿元)	纺织服装、服饰业主营业务成本(亿元)	木材加工和木、竹、藤、棕、草制品业主营业务成本(亿元)
2018	12	293.62	3735	1962.53	1333.3	741.39
2019	1	56.75	2392.45	1269.56	920.26	407.21
2019	2	56.75	2378.92	1267.39	920.26	406.51
2019	3	201.54	3043.3	1775.19	1135.9	599.24
2019	4	181.74	2780.63	1713.59	1124.93	587.98
2019	5	213.42	2964.08	1846.3	1147.39	626.68
2019	6	293.32	3442.57	2044.3	1318.66	738.39
2019	7	240.27	3069.66	1865.94	1222.75	660.54
2019	8	250.65	3213.34	1861.35	1209.07	664.49
2019	9	261.03	3486.02	1939.63	1283.27	716.72
2019	10	206.79	3443.32	1991.22	1294.84	672.57
2019	11	206.7	3342.8	1791.91	1237.73	710.63
2019	12	293.62	3735	1962.53	1333.3	741.39
2020	1	56.75	2392.45	1269.56	920.26	407.21
2020	2	56.75	2378.92	1267.39	920.26	406.51
2020	3	201.54	3043.3	1775.19	1135.9	599.24
2020	4	181.74	2780.63	1713.59	1124.93	587.98
2020	5	213.42	2964.08	1846.3	1147.39	626.68
2020	6	293.32	3442.57	2044.3	1318.66	738.39
2020	7	240.27	3069.66	1865.94	1222.75	660.54
2020	8	250.65	3213.34	1861.35	1209.07	664.49
2020	9	261.03	3486.02	1939.63	1283.27	716.72
2020	10	206.79	3443.32	1991.22	1294.84	672.57
2020	11	206.7	3342.8	1791.91	1237.73	710.63
2020	12	293.62	3735	1962.53	1333.3	741.39
2021	1	56.75	2392.45	1269.56	920.26	407.21
2021	2	56.75	2378.92	1267.39	920.26	406.51

主营业务收入的预测数据:

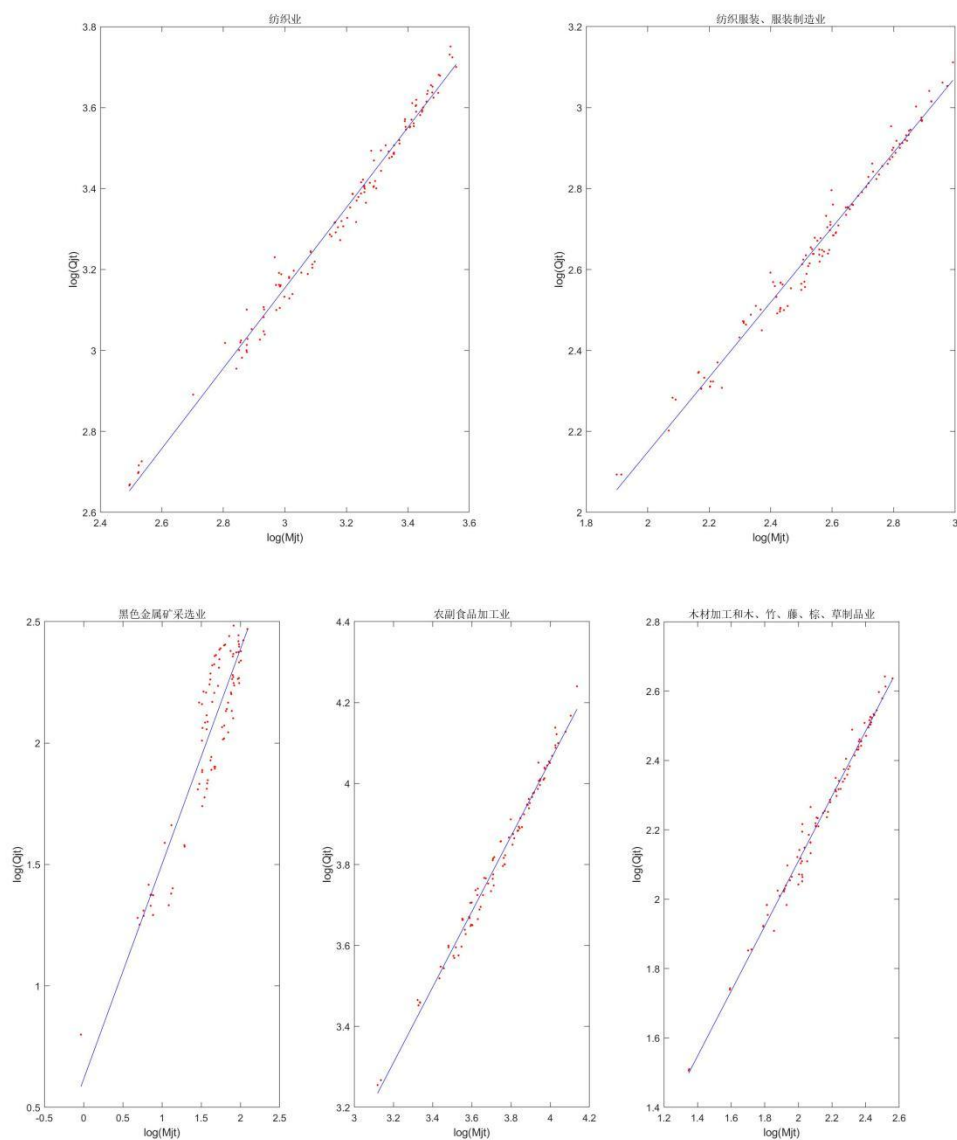
年	月	黑色金属矿采选业主营业务收入	农副食品加工业主营业务收入	纺织业主营业务收入	纺织服装、服饰业主营业务收入	木材加工和木、竹、藤、棕、草制品业主营业务收入
2018	12	283.22	4452.37	2340.65	1648.93	883.71
2019	1	160.59	2665.24	1420.69	1091.94	451.47
2019	2	152.6	2663.83	1417.91	1091.93	449.94
2019	3	200.31	3357.74	1972.51	1320.31	657.29
2019	4	194.65	3056.43	1900.26	1311.64	643.74
2019	5	206.38	3258.19	2046.99	1328.98	689.71
2019	6	235.81	3787.28	2277.44	1523.48	816.27
2019	7	220.54	3357.21	2055.37	1405.62	723.25
2019	8	228.09	3522.59	2058.9	1395.76	730.04
2019	9	228.11	3854.82	2163.87	1498.2	796.33
2019	10	187.93	3817.66	2228.66	1517.23	752.05
2019	11	207.54	3790.2	2052.35	1480.19	813.59
2019	12	213.78	4452.37	2340.65	1648.93	883.71
2020	1	120.39	2665.24	1420.69	1091.94	451.47
2020	2	113.59	2663.83	1417.91	1091.93	449.94
2020	3	147.99	3357.74	1972.51	1320.31	657.29
2020	4	142.67	3056.43	1900.26	1311.64	643.74
2020	5	150.01	3258.19	2046.99	1328.98	689.71
2020	6	169.91	3787.28	2277.44	1523.48	816.27
2020	7	157.44	3357.21	2055.37	1405.62	723.25
2020	8	161.23	3522.59	2058.9	1395.76	730.04
2020	9	159.57	3854.82	2163.87	1498.2	796.33
2020	10	130.02	3817.66	2228.66	1517.23	752.05
2020	11	141.89	3790.2	2052.35	1480.19	813.59
2020	12	144.33	4452.37	2340.65	1648.93	883.71
2021	1	80.19	2665.24	1420.69	1091.94	451.47
2021	2	74.57	2663.83	1417.91	1091.93	449.94
2021	3	95.66	3357.74	1972.51	1320.31	657.29

2.五个行业 $\ln(Q_{jt})$ 与 $\ln(M_{jt}), \ln(SM_{jt})$ 项的线性拟合

然后，我们对 SM_{jt} 的对数和 Q_{jt} 的对数进行线性拟合，所选取的五个行业的结果如下图所示：



我们看到，对这五个行业来说， Q_{jt} 关于 SM_{jt} 的拟合程度很差，这是因为 $\ln(SM_{jt})$ 的斜率很小，导致它可能在一个维度上呈现出较强的线性程度，但图中从另一个维度观测，它可能线性程度大幅减弱，呈现出较大的离散性。而在对 $\ln(M_{jt})$ 和 $\ln(Q_{jt})$ 的线性拟合中，结果如下图所示：



我们看到，对于这五个行业来说， $\ln(Q_{jt})$ 关于 $\ln(M_{jt})$ 都呈现出很强的线性关系，数据值几乎都紧贴在拟合直线的两侧。

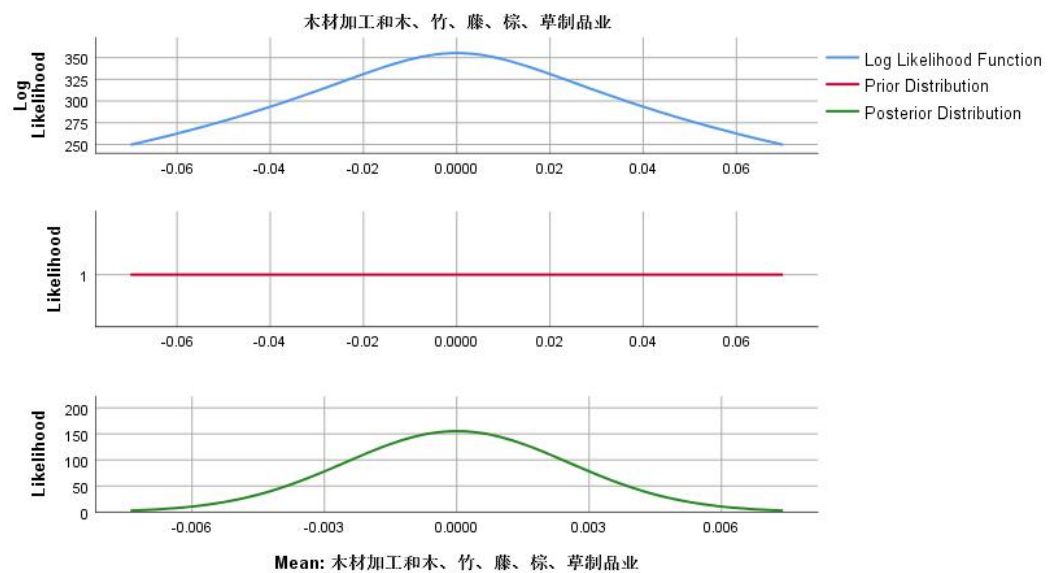
3.对五个行业关于产量的误差项分析

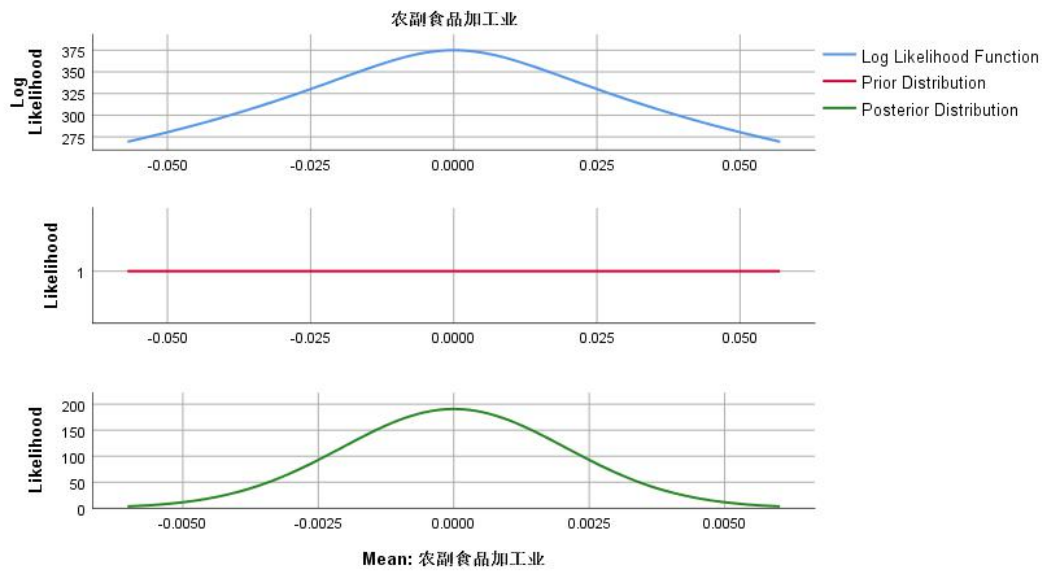
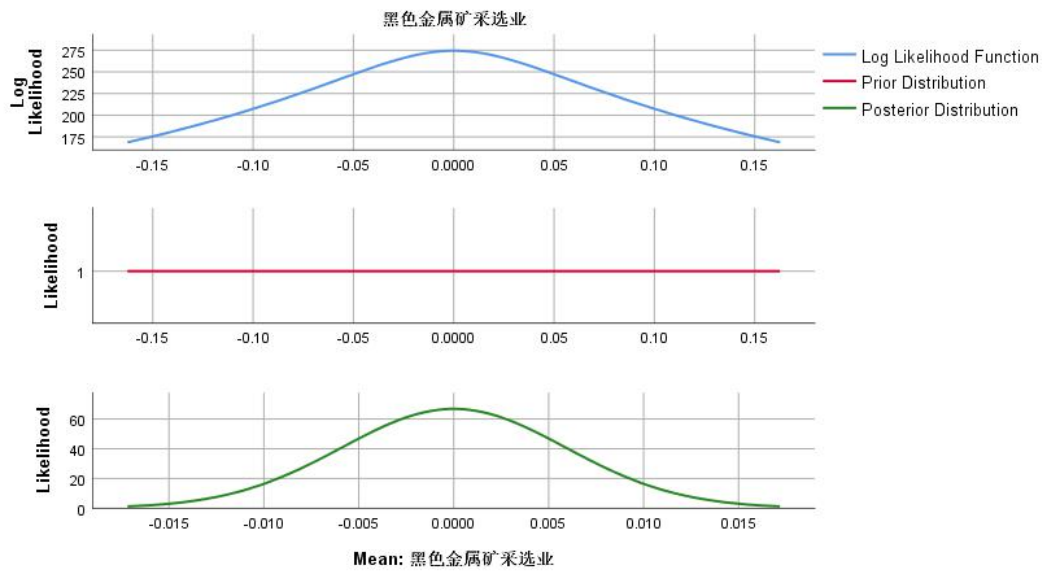
之后我们计算关于 Q_{jt} 和 Y_{jt} 的误差项 ε ，并对该误差项进行单样本正态分布检验，SPSS 输出结果如下：

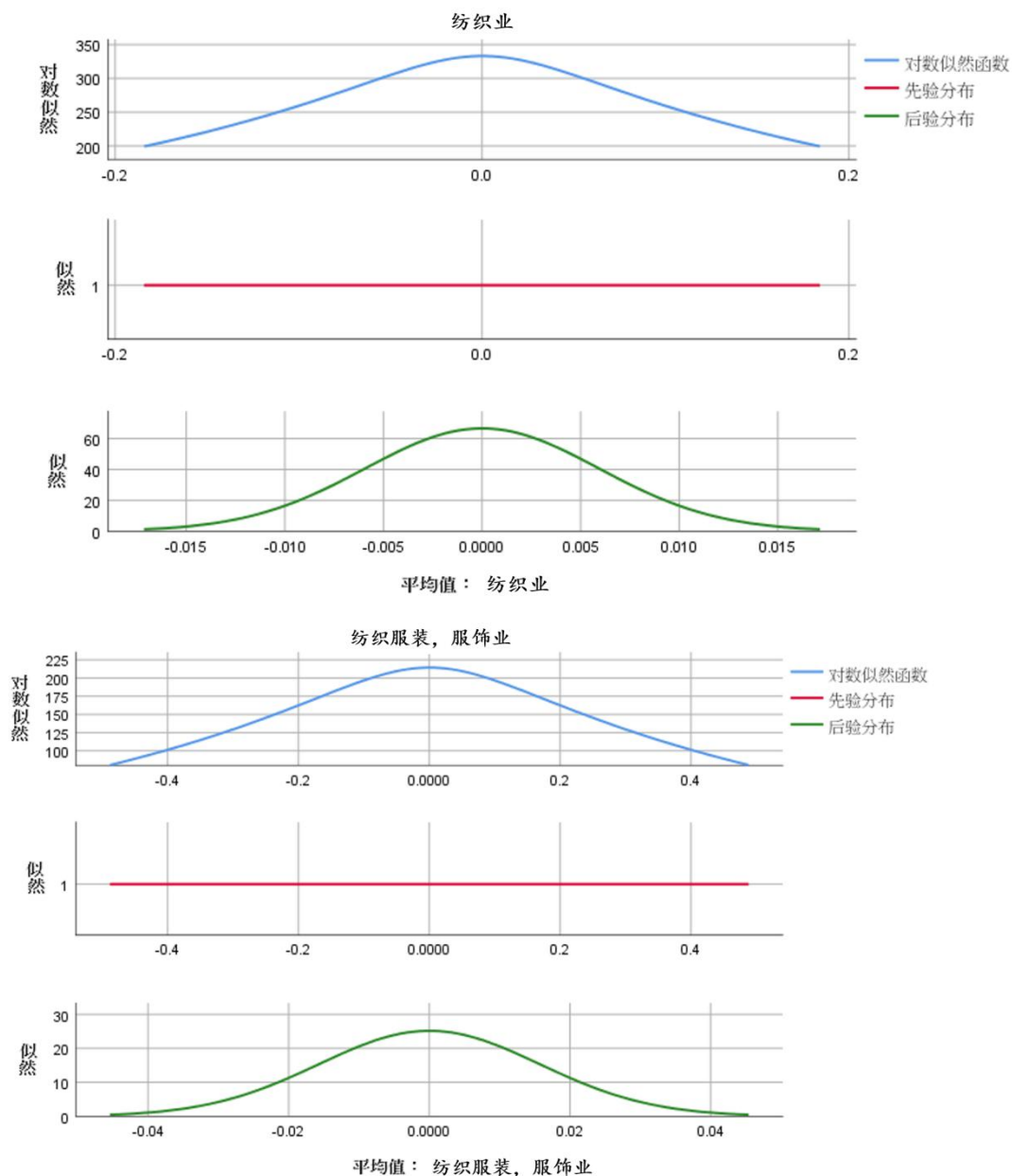
表格 1

对 ε_{jt} 的单样本平均值的后验分布特征

	个案数	后验			95% 置信区间	
		众数	平均值	方差	下限	上限
黑色金属矿采选业	96	-.0000000000093	-.0000000000093	.000	-.0118224870451	.01182248702638
		75	75		31	1
农副食品加工业	96	-.0000000000250	-.0000000000250	.000	-.0041430734655	.00414307341554
		00	00		44	4
木材加工和木、藤、棕、草制品业	96	.0000000000000	.0000000000000	.000	-.0050881397338	.00508813973386
		00	0		62	2
纺织业	122	.0000000000000	.0000000000000	.000	-.011822056	.011822056
		00	0			
纺织服装，服饰业	122	.0000000000000	.0000000000000	.000	-.031320184	.031320184
		00	0			







我们看到输出结果中，误差项 ε 的对数似然函数和后验分布都具有良好的正态性，而先验分布的似然函数值与直线 $y=1$ 十分贴合。

3.五个行业的 CES 生产函数生产参数具体数值

通过上述线性拟合与误差分析，得到五个行业各自的生产参数：

表格 2

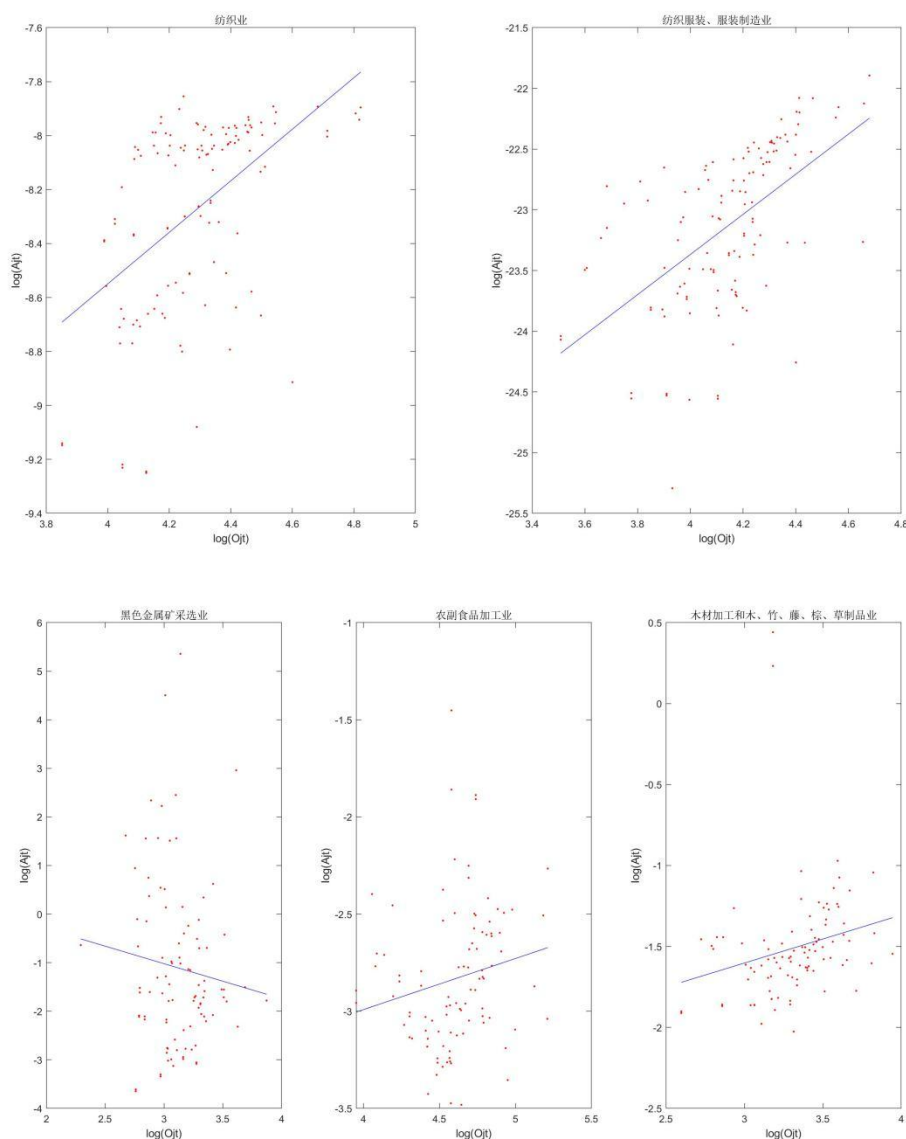
参数名	纺织业	纺织服装、服饰业	黑色金属矿采选业	农副食品加工业	木材加工和木、竹、藤、棕、草制品业
γ	-2.83789	-0.38065	-0.56453	-1.17892	-0.71627
s	0.99167	0.92551	0.88563	0.93298	0.93636
C	1.24324	1.65537	2.06249	1.43358	1.40274

（二）对资本增益效率与劳动力增益效率的数据结构分析

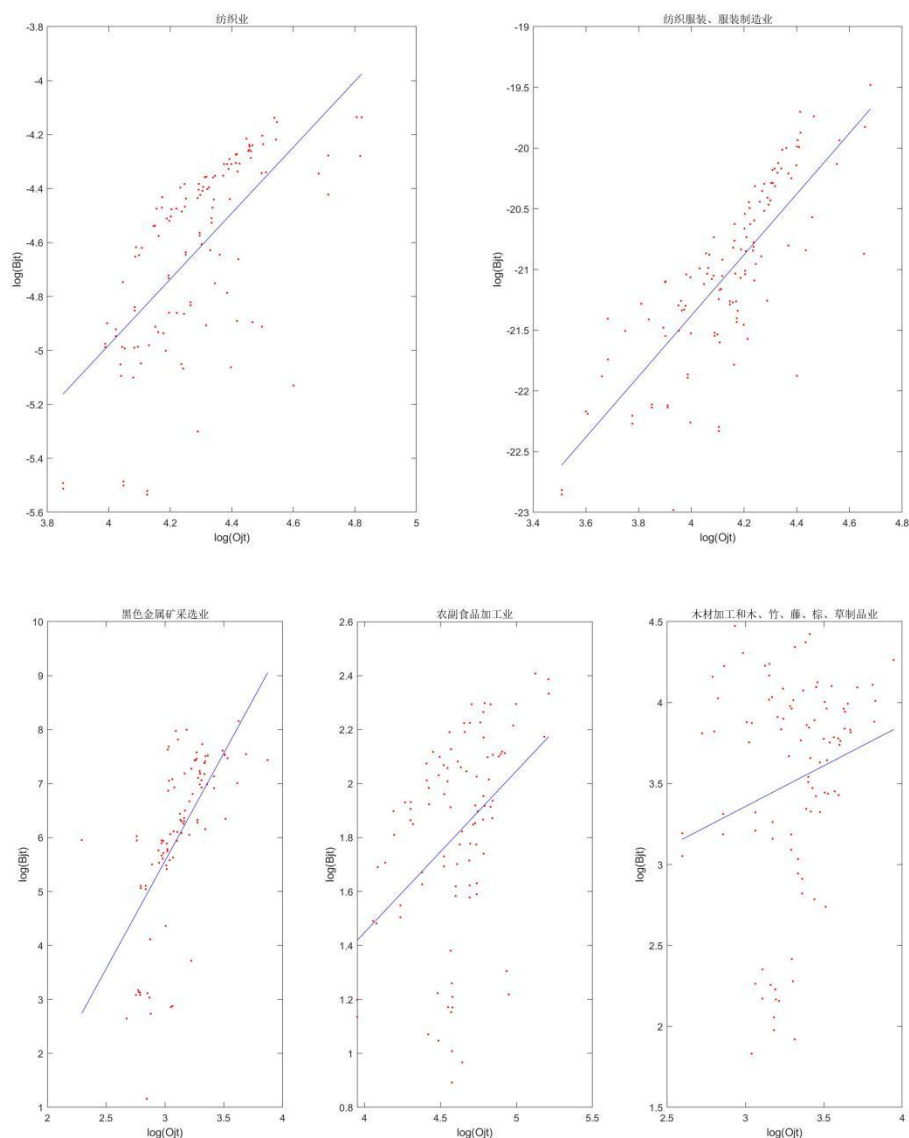
1. 对 $\ln(A_{jt})$ ， $\ln(B_{jt})$ 与 $\ln(O_{jt})$ 的线性拟合以及 $\ln(DA_{jt})$ 与 $\ln(DB_{jt}) - \ln(B_{jt})$ 的线性拟合

在得出误差项 ε 后，我们计算资本增益效率 A_{jt} 和劳动力增益效率 B_{jt} ，并将它们关于管理费用 O_{jt} 进行线性拟合，拟合结果如下图所示。

A_{jt} 的对数关于 O_{jt} 的对数：



我们看到， $\ln(A_{jt})$ 的关于 $\ln(O_{jt})$ 的数据值的分布大多比较分散，不能够呈现出较好的线性关系，所以我们可以得出 A_{jt} 与 O_{jt} 之间的相关程度不大，即认为 A_{jt} （资本增益效率）的内部结构中 O_{jt} 所占的比重不大，这是符合我们的直观认知的。



而从 B_{jt} 的对数关于 O_{jt} 的对数的拟合图中，可以看出纺织业、纺织服装、服饰业以及黑色金属矿采选业中 B_{jt} 和 O_{jt} 有良好的相关关系。

2.对模型假设是否成立的判据

我们为了研究资本增益效率和劳动力增益效率内部的数据结构，在上文中，试图构造出两种不同的增益效率关于长期增益效率能力和短期增益效率能力的生产函数，但是这样的生产函数能够被建立是基于各个行业对数据的收集和有效使用的能力是否成熟以及对管理费用的使用是否恰当。正如前文所说，我们所收集的数据来自于其行业中的大中型企业，所以我们基本可以认为具有相当的参考性以及对管理费用的使用方案已经成熟。但就“大数据”而言，事实上，从它引起人们的广泛重视至今只有十余年，我们不能认为所有行业对“数据”有强大的运用能力。同时我们应该注意到，柯布-道格拉斯生产函数具有其应用范围的局限性，故我们所建立的模型只能应用在对数据使用成熟且满足柯布-道格拉斯生产函数的行业中。为判断所选的行业是否符合上述条件，我们定义了两个统计量 DA_{jt} 和 DB_{jt} ，为了便于计算，我们先取它们的对数，然后作出它们之间线性拟

合的直线，并计算出 DA_{jt} 的数值与其拟合直线上同月的数值之差的方差，提前设定一个预估好的评判界限（ <0.1 ），对于方差小于该界限的对应行业，我们即认为它们是符合前面所述要求的。

下表是我们计算所选取的五个行业的相关方差的程序输出。

描述统计				
	案例数	均值	标准 偏差	方差
黑色金属矿采选业	96	.000000000	1.0505236245	1.104
农副食品加工业	96	.000000000	.3344044462	.112
木材加工和木、竹、藤、棕、草制品业	96	.000000000	.6478896527	.420
纺织业	122	.000000000	.0653983243	.004
纺织服装、服饰业	122	.000000000	.1732598442	.030

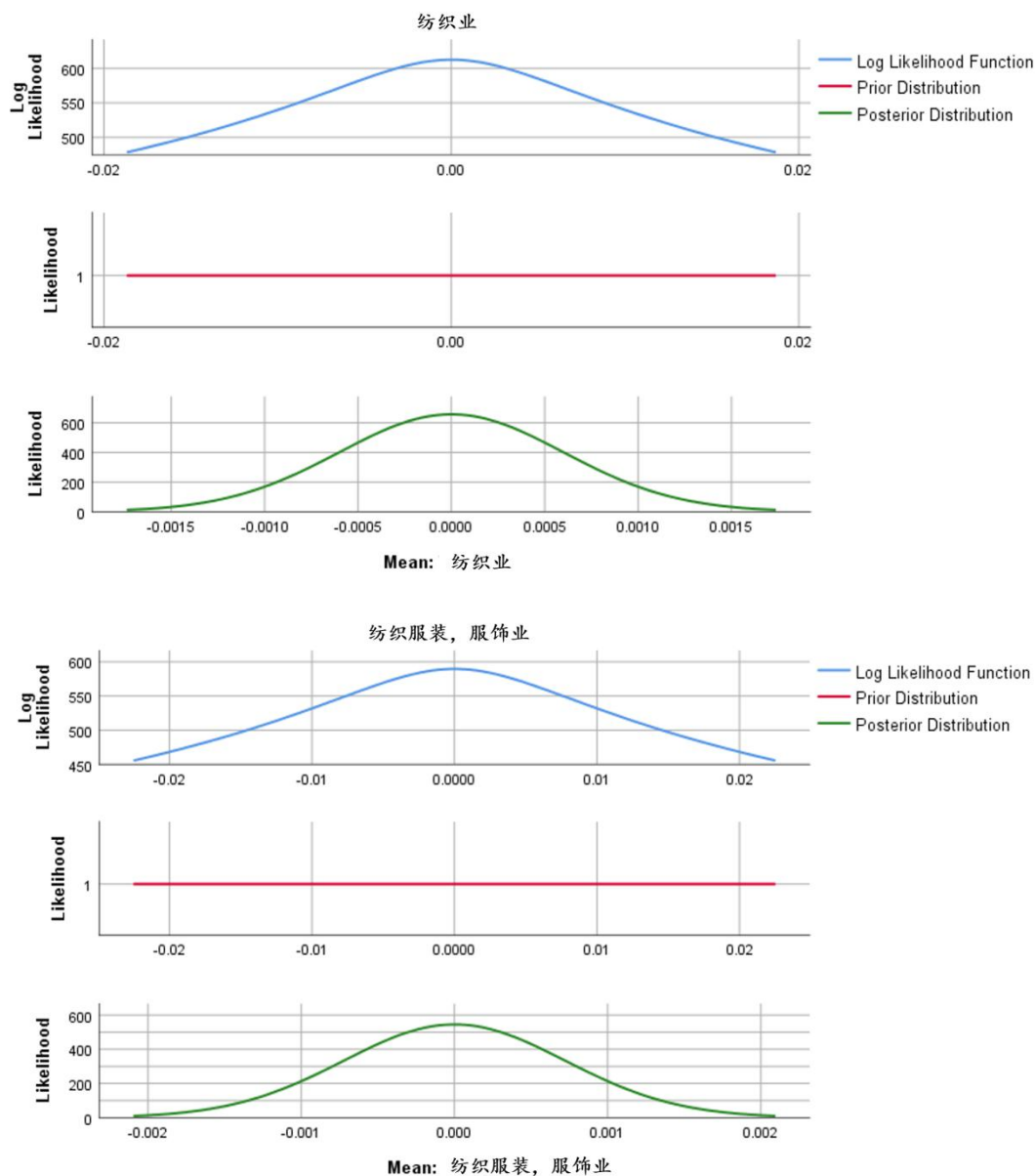
从表中可以看出，纺织业和纺织服装、服饰业的方差小于 0.1，符合评判标准，而黑色金属矿采选业、农副食品加工业、木材加工和木、竹、藤、棕、草制品业的方差均大于 0.1，不符合所设要求。故在接下来模型中的数据处理与计算中，我们将五个行业中的前三个和后两个区分开，在后续的分析中，我们将看到，我们所设立的评判标准是合理的，分开后的两类行业的相关数据将在计算中走向两种不同的情况。

3.对两个行业增益效率误差项的分析

然后，用类同计算误差项 ε 的方式，同样能够算出误差项 η 的值，我们发现，前三个行业的 η 值解得虚部不为 0，这与我们提前设想的结果是矛盾的，在本问题中复变量会给计算带来相当大的困难，于是我们放弃对前三个行业进行进一步的数据分析。而对于后两个行业，与 ε 误差项相同，我们也对它进行单样本正态分布检验，SPSS 的软件输出如下图：

表格 3

对 η_{jt} 的单样本平均值的后验分布特征						
	个案数	众数	后验 平均值	方差	95% 置信区间	
					下限	上限
纺织业	122	.000000000000000000	.000000000000000000	.000	-.001198593995170	.001198593995170
纺织服装、服饰业	122	.000000000000000000	.000000000000000000	.000	-.001445132588993	.001445132588993



与 ε 误差项相似, 我们看到 η 误差项的对数似然函数和后验分布都具有极好的正态性, 而先验分布几乎与直线 $y=1$ 重合。

4.对 A_{jt} 与 B_{jt} 柯布-道格拉斯生产函数的生产参数

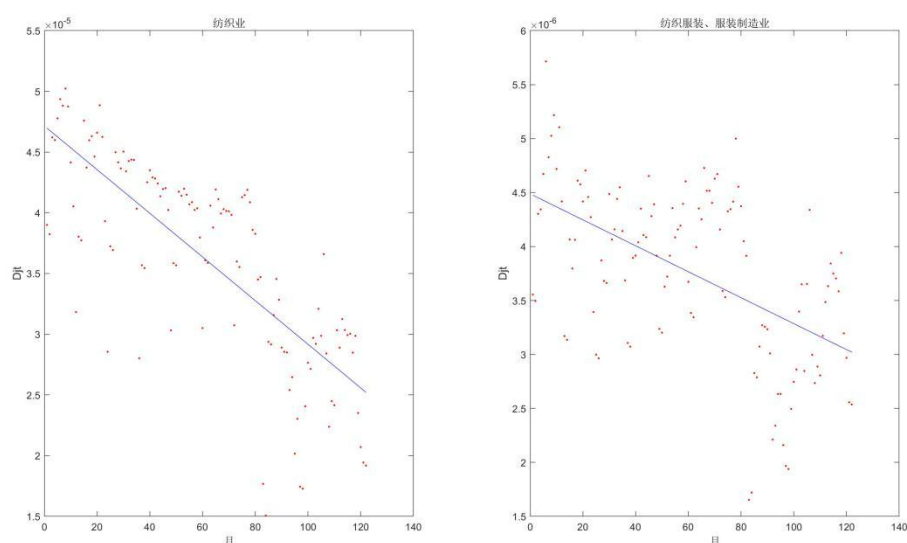
最终获得两个行业的 A_{jt} 与 B_{jt} 生产函数参数

表格 4

参数名	纺织业	纺织服装、服饰业
α_A	1. 205391664	2. 394283883
α_B	0. 967714696	1. 72826819
β_A	0. 954698921	1. 652561628
β_B	1. 221825228	2. 50396996
B	1. 094174477	0. 000059717963

5.两行业长期增益效率能力项随时间的变化

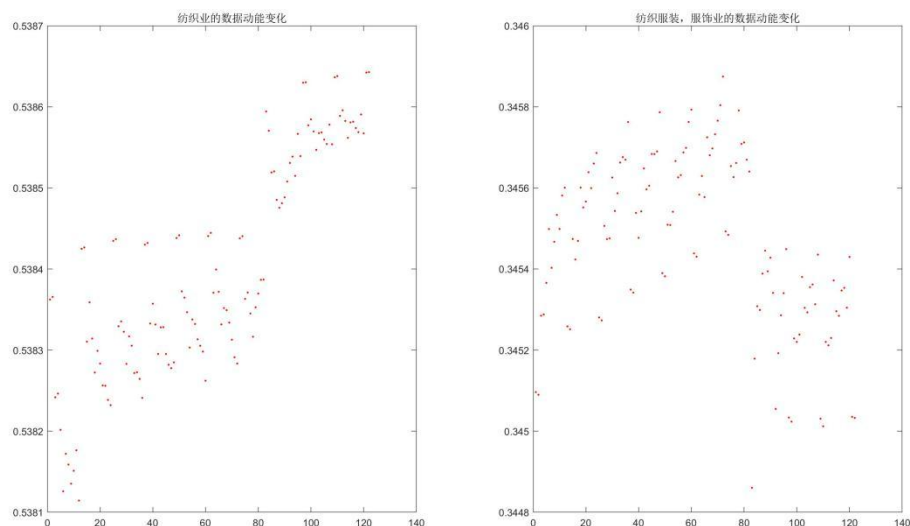
这样我们就可以计算长期增益效率能力 D_{jt} ，然后对月份数 t 和 D_{jt} 进行线性拟合，也就是观察 D_{jt} 随着时间的变化趋势，如下图所示：



可以看到， D_{jt} 整体在保持大于 0 的情况下，整体随时间呈下降趋势，说明纺织业和纺织服装、服饰业的长期增益效率能力较差，对其行业发展有阻碍作用。

（三）两行业数据动能随时间的变化

最后，我们能够得出本文中最关键的部分，即数据动能 WD_{jt} ，然后我们画出数据动能关于时间的变化，如下图所示。



我们可以看到，纺织业的数据动能随着时间的变化整体呈上升趋势，而纺织服装、服饰业的数据动能整体先是有上升趋势，而在 80 个月后就大约 2017 年下半年突然出现断崖式下降，可以推断出纺织服装、服饰业在当时受到了严重打击。

六、结论与建议

在数据动能的计算中，我们发现黑色金属矿采选业、农副食品加工业、木材加工和木、竹、藤、棕、草制品业无法算出数据动能，所以我们推测在这三个行业中，目前并不存在数据动能或者说数据动能几乎为 0，而通过计算纺织业和纺织服装、服饰业的数据动能，我们看到纺织业的数据动能随着时间变化整体呈上升趋势，可以判断出该行业对于数据的收集和应用能力近年来有所提升，而纺织服装、服饰业的数据动能在 2017 年以前和纺织业一样整体上升，但其数值低于纺织业的数据动能，且在 2017 年下半年出现断崖式下降，由此可以推测该行业在当时可能遇到了突发情况，导致某个原始数据在 2017 年发生了突变，从而影响之后的计算。在此结论上，我们给出如下建议：

1. 对于本文中所选取的纺织业、纺织服装、服饰业和农副食品加工业等传统行业，在如今的大数据时代下，应加强行业变革，注重对数据的收集和合理、充分运用，加强对数据的管理。
2. 对于经典的柯布-道格拉斯函数，我们应深度挖掘其中资本增益效率和劳动力增益效率的内部数据结构，分析其成分及其作用，充分发挥数据动能的作用，让数据要素成为经济发展新动能。
3. 对于数据动能的研究，由于该变量的不可测性和各个行业的特殊性，需要综合考虑各种情况的发生，得到的结论未必具有广泛的适用性，所以需要加大研究行业的范围，探索典型案例，这也是本文所需要进一步改进的地方。

七、参考文献

- [1] Zhang, H. Biased Technology and Contribution of Technological Change to Economic Growth: Firm-Level Evidence[R]. Working paper, University of Hong Kong, Hong Kong, 2014.
- [2] 郑江淮, 宋建, 张玉昌, 等. 中国经济增长新旧动能转换的进展评估[J]. 中国工业经济, 2018, 000(006):24-42.
- [3] 张兴祥, 范明宗. 技能偏向性技术进步与要素的收入分配——基于 CES 生产函数的建模与理论解释 [J]. 福建论坛 (人文社会科学版), 2020, No. 335(04):145-153.
- [4] Kongsamut P., S. Rebelo, and D. Xie. Beyond Balanced Growth [J]. Review of Economic Studies, 2001, 68 (4): 869-82.
- [5] 柴士改,李金昌.中国经济增长新动能统计测度研究[J].统计与信息论坛,2021,36(01):47-58.

附录

1. 主营业务成本时间序列模型摘要

模型拟合度

拟合统计	平均值	标准误差	最小值	最大值	百分位数		
					5	10	25
平稳 R 方	.504	.083	.386	.589	.386	.386	.423
R 方	.861	.050	.787	.910	.787	.787	.810
RMSE	149.482	99.693	56.024	285.034	56.024	56.024	58.846
MAPE	5.983	2.876	4.086	11.024	4.086	4.086	4.277
MaxAPE	128.233	118.307	54.378	336.063	54.378	54.378	54.628
MAE	82.968	55.295	36.050	166.481	36.050	36.050	36.211
MaxAE	923.933	591.410	332.367	1690.616	332.367	332.367	367.778
正 态 化 BIC	9.708	1.457	8.147	11.401	8.147	8.147	8.243

模型统计

模型	预测变量数	模型拟合度统计			
		平稳 R 方	杨-博克斯 Q(18)统计	DF	显著性
黑色金属矿采选业-模型_1	0	.573	9.281	16	.901
农副食品加工业-模型_20	0	.460	23.069	16	.112
纺织业-模型_3	0	.386	17.689	16	.342
纺织服装、服饰业-模型_4	0	.509	21.327	16	.166
木材加工和木、竹、藤、棕、草制品业-模型_5	0	.589	23.510	16	.101

2. 主营业务收入时间序列模型摘要

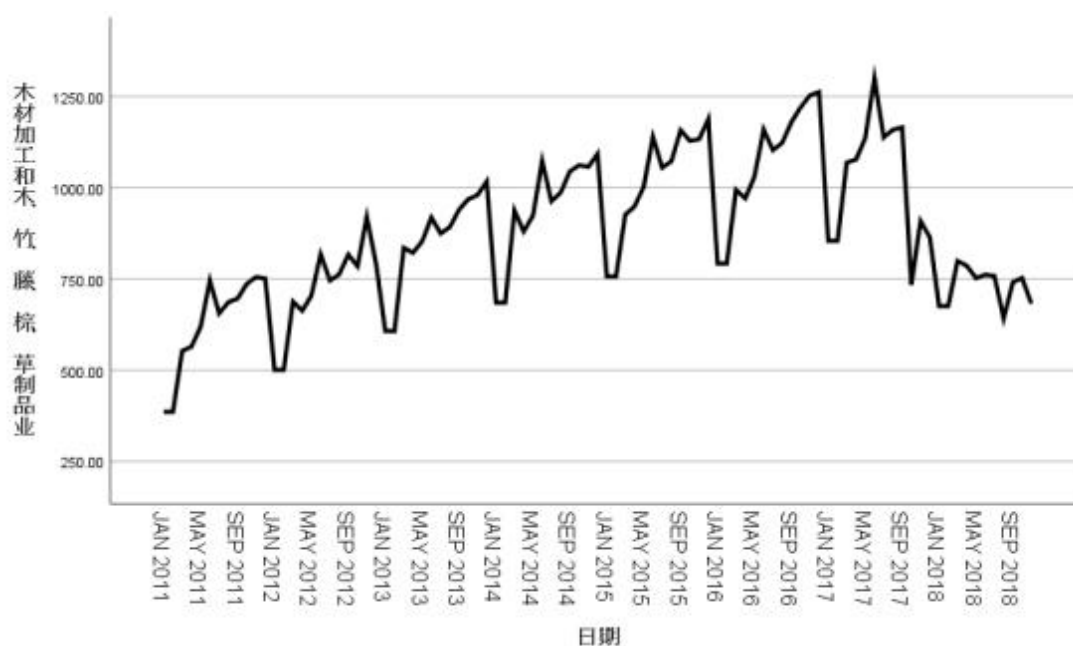
模型拟合度

拟合统计	平均值	标准误差	最小值	最大值	百分位数		
					5	10	25
平稳 R 方	.489	.081	.378	.571	.378	.378	.409
R 方	.867	.053	.790	.911	.790	.790	.812
RMSE	167.186	110.628	62.187	318.411	62.187	62.187	65.953
MAPE	5.479	2.083	3.925	9.072	3.925	3.925	4.145
MaxAPE	108.698	83.203	52.634	253.472	52.634	52.634	53.794
MAE	91.842	59.413	40.166	180.401	40.166	40.166	40.624
MaxAE	1029.795	638.940	393.896	1862.927	393.896	393.896	426.575
正 态 化 BIC	9.946	1.438	8.404	11.623	8.404	8.404	8.494

模型统计

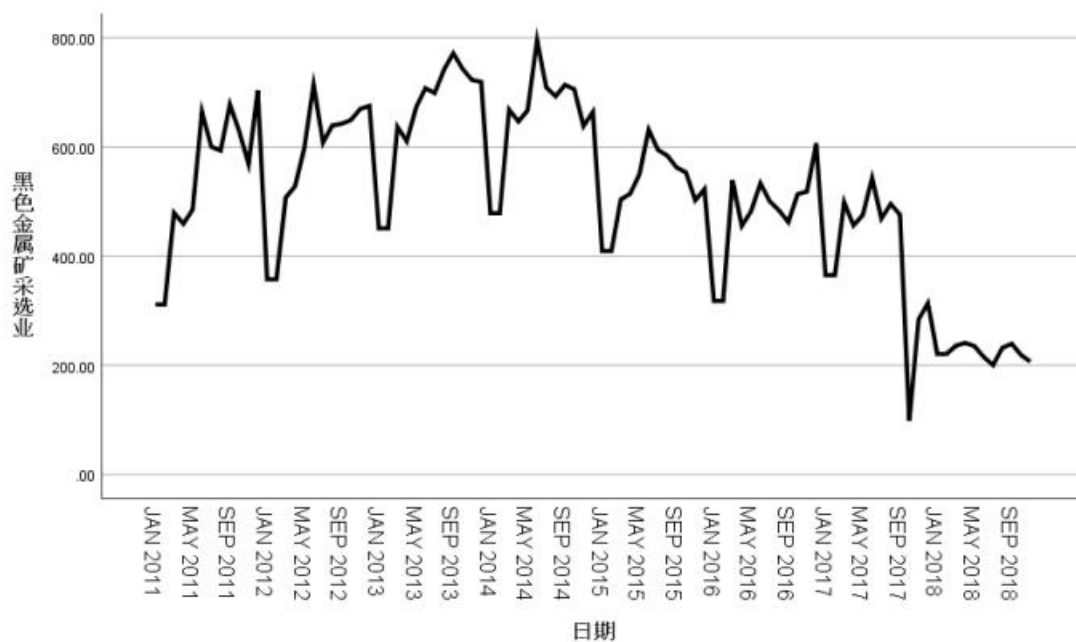
模型	预测变量数	模型拟合度统计			
		平稳	R 方	杨-博克斯统计	Q(18) 显著性
黑色金属矿采选业-模型_1	0	.557		11.842	15 .691
农副食品加工业-模型_2	0	.440		20.829	16 .185
纺织业-模型_3	0	.378		18.703	16 .284
纺织服装、服饰业-模型_4	0	.499		22.235	16 .136
木材加工和木、竹、藤、棕、草制品业-模型_5	0	.571		23.748	16 .095

3.

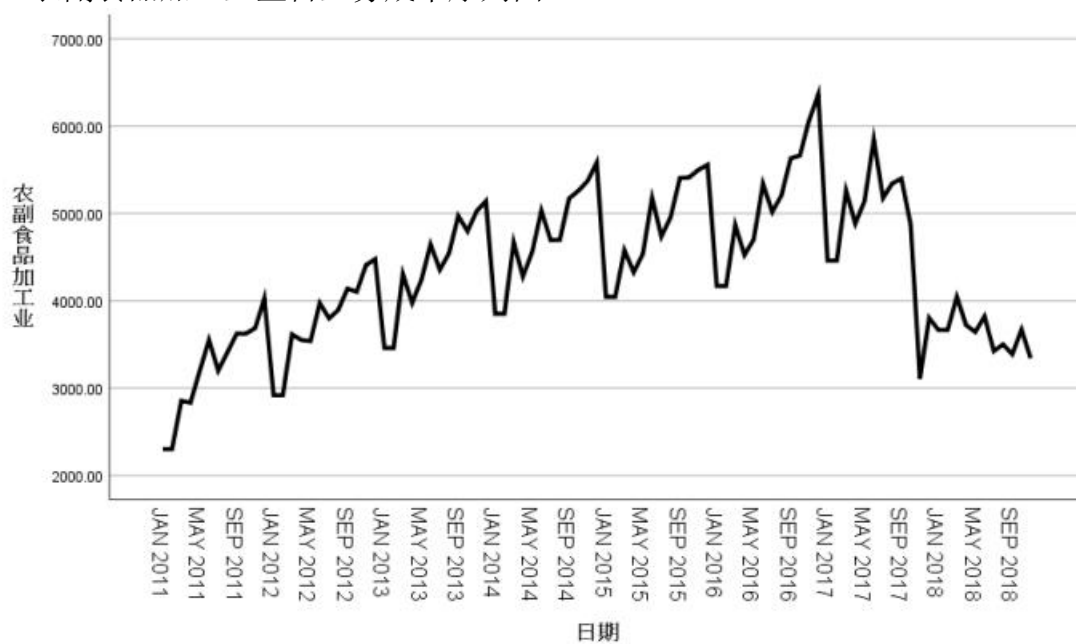


5. 木材加工和木、竹、藤、棕、草制品业主营业成本序列图

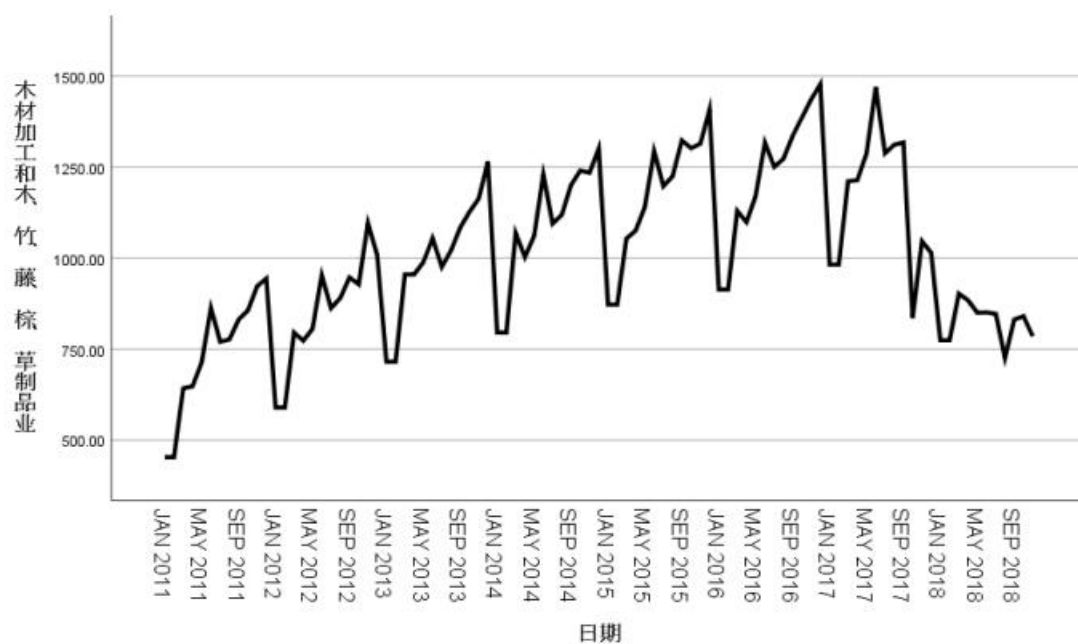
6. 黑色金属矿采选业主营业务成本序列图



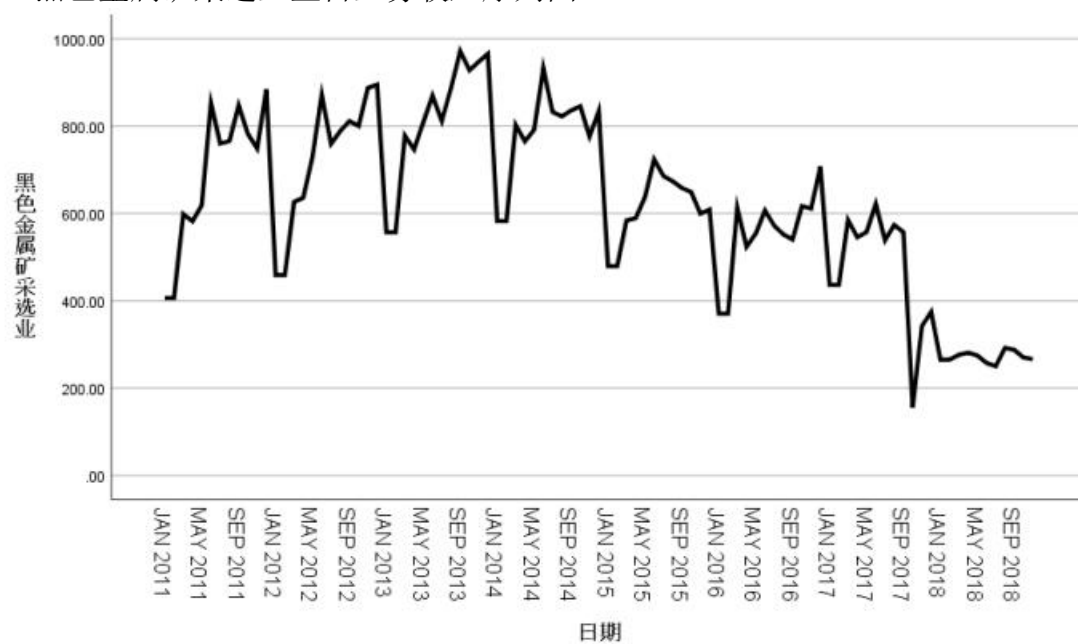
7. 农副食品加工业主营业务成本序列图



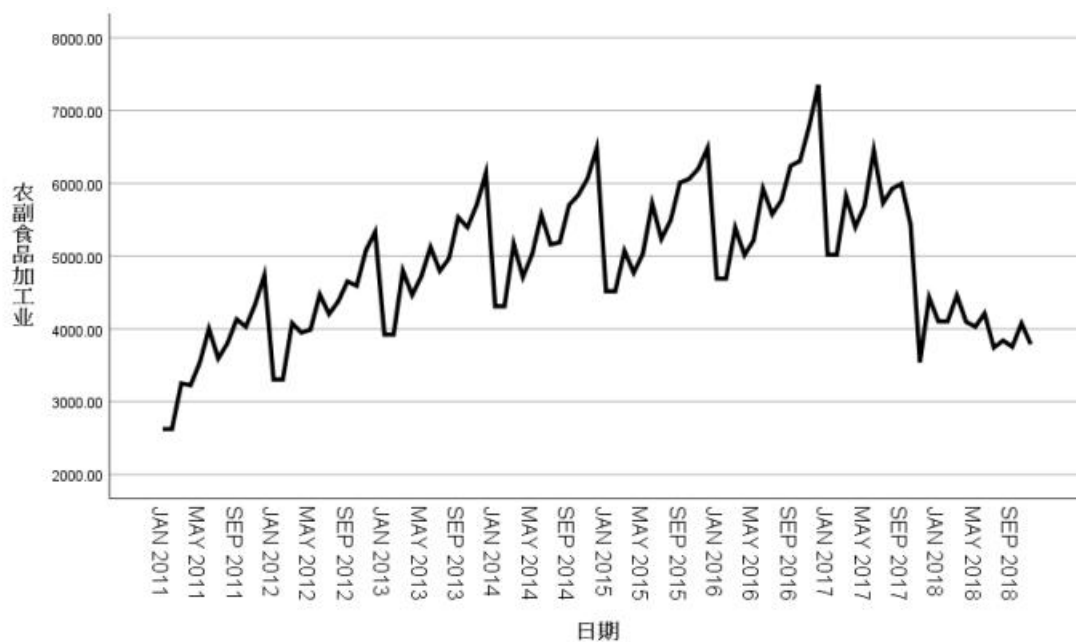
8. 木材加工和木、竹、藤、棕、草制品业主营业务收入序列图



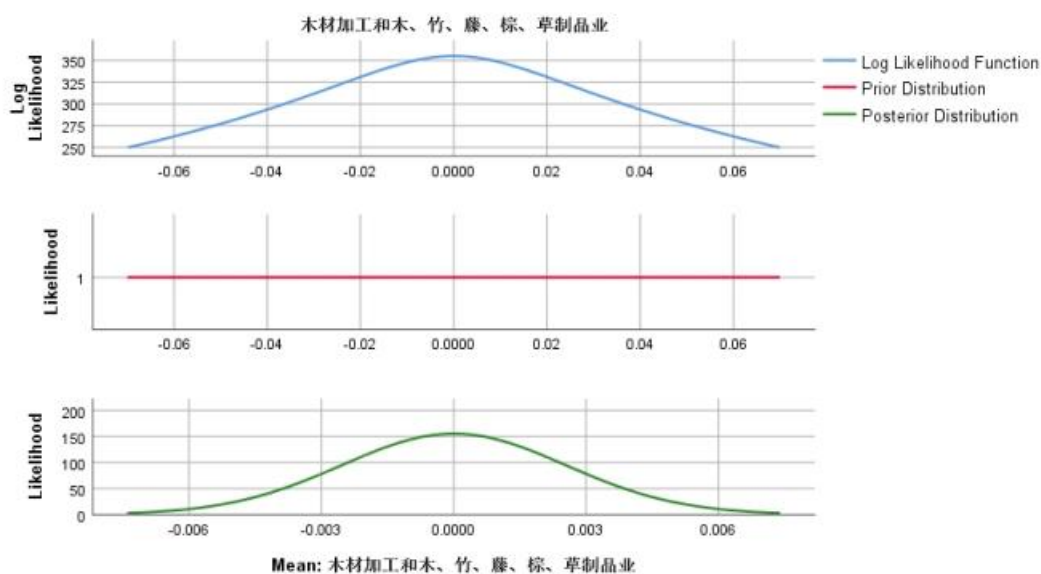
9. 黑色金属矿采选业主营业务收入序列图



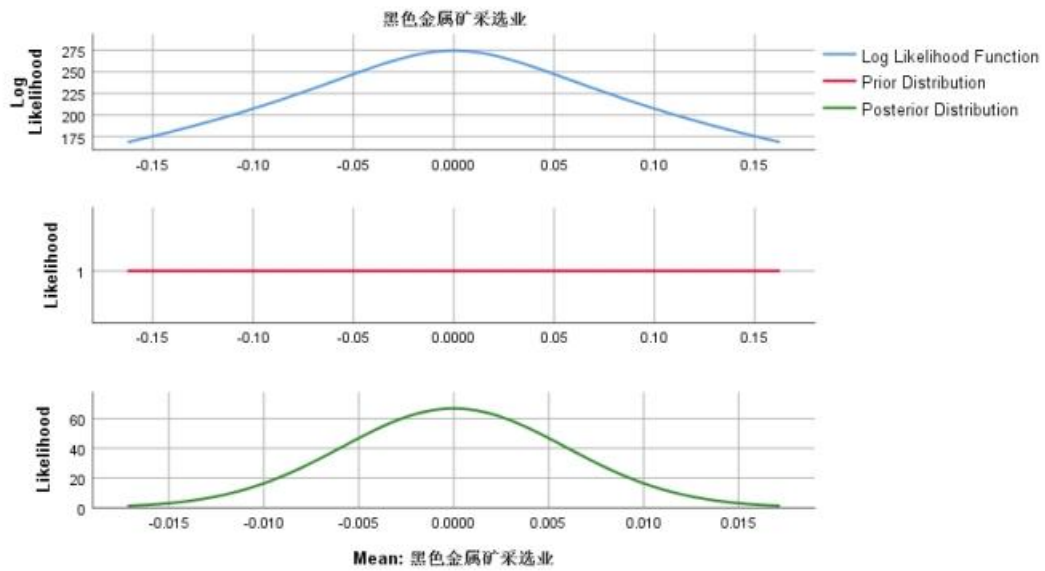
10. 农副食品加工业主营业务收入序列图



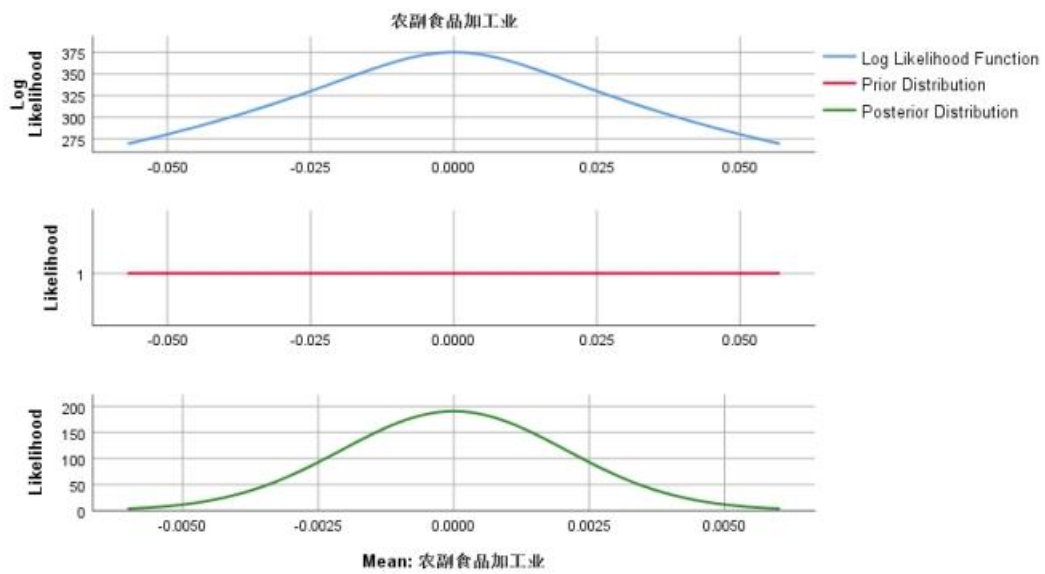
11.木材加工和木、竹、藤、棕、草制品业 ε 误差项单样本正态分布检验



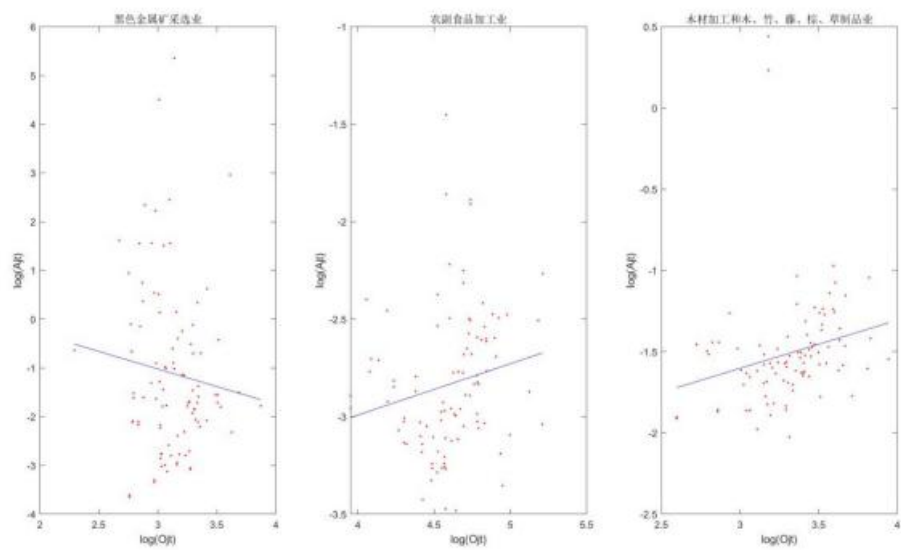
12.黑色金属矿采选业 ε 误差项单样本正态分布检验



13. 农副食品加工业 ε 误差项单样本正态分布检验



14. $\lg(O_{jt})$ 和 $\lg(A_{jt})$ 的拟合图 2



15. $\lg(O_{jt})$ 和 $\lg(B_{jt})$ 的拟合图 2

