

西北农林科技大学学术型硕士研究生课程考试试题（卷）

参考答案与评分标准

考试课程：计量经济学

学年学期：2022-2023-1

试卷类型：A 卷

考试时间：2023-02-24

专业年级：

任课老师：

一、计量经济学基础分析（共 4 小题，共 25 分）

1. (6 分) 答：（1）（3 分）【分析 2】比【分析 1】增加了一个新的变量 $male_i$ ，从理论上来看这个增加的自变量（性别）将会对因变量（成绩）产生影响，因此对因变量具有一定解释力，而更大的可决系数 R^2 也印证了这一点。（2）（3 分）【分析 3】和【分析 4】对比来看，并没有增加新的变量，只是做了一次变量变换，也即将原变量 gpa_i ，去均值变换为 $(gpa_i - 2.81)$ ，这样做并不会导致对因变量解释力的增加或减低，因此模型的可决系数也不会发生改变。

2. (7 分) 答：（1）（3 分）【分析 3】的 OLS 估计结果表明 $t_{\gamma_2}^* = 0.62$ ，明显小于理论 t 查表值 $t_{(0.975, 852)} = 1.96$ ，因此 t 检验不显著。（2）（4 分）给定学积分 $gpa_i = 2$ ，则其预期考试成绩为 $E(score | (gpa = 2, male = 0)) = \gamma_1 + \gamma_3 gpa_i = 30.36 + 2.48 * 2 = 35.32$ 。

3. (6 分) 答：（1）（2 分）首先自变量 $male_i$ 和交互项 $male_i * gpi_i$ 之间有线性相关关系确实可能会带来多重共线性问题。（2）（4 分）主要理由：【分析 3】在【分析 2】基础上增加交互项之后， $male_i$ 变量 t 检验不显著，而且其系数的估计精度大大下降（其标准误差从 0.7418 增大 3.9646），因此可以认为多重共线性确实给模型估计带来了干扰。其根源在于给定一个男性，其交互项中的学积分变量 gpa_i 基本不大可能为 0，从而导致 $male_i$ 和交互项 $male_i * gpi_i$ 高度线性相关。

4. (6分) 答: (1) 【分析3】和【分析4】的结果可以看出, 只有 $male_i$ 的系数估计、标准误和t统计量发生了改变, 其他变量的这些估计量都不变。(2) 结合问题3的相关考虑, 我们更应该倾向选择【分析4】设定的计量模型, 因为变量的去均值变换是一种有效的缓解多重共线性问题的手段之一。

二、时间序列分析 (共3小题, 共25分)

1. (6分) 答: r_t 序列为非平稳性序列 (3分)。由图2-1 (左) 可知其具有明显时间趋势, 随时间推移序列实现值 (1分); 图2-1 (右) 为 r_t 自回归函数图, 由图可知其前20阶自相关函数没有出现截尾现象 (显著不等于0), 衰减过程非常缓慢; 综合图2-1所提供信息可以判断 r_t 为非平稳序列 (1分)。

2. (9分) 答: 由图2-1可以看出序列 r_t 表现出较明显的时间趋势, 首先对其做1阶差分处理消除时间趋势 (图2-2)。结合图2-1与图2-2可以看出序列 r_t 表现出较为明显的周期为4的季节性特征, 对其做周期为4的季节差分处理调整季节效应 (图2-3) (5分, 答出时间趋势或季节效应即可得分)。

- Step1: 一阶差分

$$\Delta r_t = r_t - r_{t-1} = (1 - B)r_t$$

- Step2: 季节差分

$$\Delta_4(\Delta r_t) = \Delta r_t - \Delta r_{t-4} = (1 - B^4)\Delta r_t = (1 - B^4)(1 - B)r_t$$

- Step3: 图2-3 (右) ACF图信息判断季节差分后的序列在5阶后截尾, 则可考虑AR(5)模型 (4分, 最高阶数为5即可得分)

$$(1 - B^4)(1 - B)r_t = \theta_0 + (1 - \theta_1 - \theta_2 - \theta_3 - \theta_4 - \theta_5)a_t$$

3. (小计 10 分)

(1) (3 分) 答: 该模型为 ARMA 模型 (或自回归移动平均模型) (2 分)。其中 AR 阶数为 2, MA 阶数为 2 (或者自回归阶数为 2, 移动平均阶数为 2) (1 分)。

(2) (3 分) 答:

- 若以 h 时刻为预测原点, 向前 1 步预测值 $\hat{r}_t(1)$ 为 (1 分):

$$\hat{r}_h(1) = E(r_{h+1} | F_h) = \varphi_0 + \varphi_1 r_h + \varphi_2 r_{h-1} - \theta_1 a_h - \theta_2 a_{h-1}$$

- 若以 h 时刻为预测原点, 向前 2 步预测值 $\hat{r}_t(2)$ 为 (1 分):

$$\hat{r}_h(2) = E(r_{h+2} | F_h) = \varphi_0 + \varphi_1 \hat{r}_h(1) + \varphi_2 r_h - \theta_2 a_h$$

- 若以 h 时刻为预测原点, 向前 3 步预测值 $\hat{r}_t(3)$ 为 (1 分):

$$\hat{r}_h(3) = E(r_{h+3} | F_h) = \varphi_0 + \varphi_1 \hat{r}_h(2) + \varphi_2 \hat{r}_h(1)$$

(3) (4 分) 答: 若要判断该模型样本外预测能力, 可选用的指标有 mean square of forecast errors (MSFE)、mean absolute forecast errors (MAFE) 或 Bias (2 分, 任写出一个即得分)。

- 对模型若以 h 时刻为预测原点:

$$\text{MSFE}(m) = \frac{\sum_{j=h}^{T-1} [e_j(1)]^2}{T-h}; \quad \text{MAFE}(m) = \frac{\sum_{j=h}^{T-1} |e_j(1)|}{T-h}; \quad \text{Bias}(m) = \frac{\sum_{j=h}^{T-1} e_j(1)}{T-h}$$

其中 $e_j(1)$ 为 j 时刻向前 1 步预测误差 $e_j(1) = r_{j+1} - \hat{r}_j(1)$ (2 分, 任写出一个即得分)。

三、离散变量回归模型（共 4 小题，共 25 分）

1. (5 分) 答：Probit 模型更为合适。因为被解释变量 $work_i$ 为二值变量，如果用线性概率模型估计有以下缺点：缺点之一，虽然 y 的取值非 0 即 1，但根据线性概率模型所作的预测值却可能出现 $\hat{y} > 1$ 或 $\hat{y} < 0$ 的不现实情形，因为概率必须介于 0-1 之间（2 分）。即使我们可以使得某些方法将其限制在 0-1 之间，但是误差项 u 无法满足同方差的假定（1 分）；缺点之二，概率不可能与自变量所有的可能值线性地相关（2 分）。

2. (8 分) 答：女性参与劳动力市场依赖于女性的年龄、女性的婚姻状况、受教育程度和子女数量（4 分）。该模型的准 R^2 为 0.19，检验整个方程显著性的 LR 通过统计量为 478.32， p 值为 0.0000，整个方程高度显著（4 分）。

3. (8 分) 答：女性的年龄、女性的婚姻状况、受教育程度和子女数量都对女性参与劳动力市场有显著影响，此结果为所有解释变量在样本均值处的边际效应。保持其他因素不变，女性年龄若比平均年龄增加 1 岁则其参与劳动力市场的概率会提高 1.2%（2 分）。若女性从未婚变成已婚则参与劳动力市场的概率会提高 15%（2 分）。女性拥有的子女的数量在平均数量上增加 1 个，则其参与劳动力市场的概率会提高 15%（2 分）。女性受教育年限比平均年限增加 1 年，则其参与劳动力市场的概率会提高 2%（2 分）。

4. (4 分) 答：对于同样的模型或数据，probit 和 logit 模型的估计系数会有显著不同，由于 probit 回归使用的是标准正态分布，其标准差为 1，logit 回归使用的 logistic 分布，其标准差为约为 1.814，所以 logit 回归报告的系数是 probit 的系数大约 1.814 倍，但二者的边际效应基本相同。（2 分）Logit 模型的优势在于，逻辑分布的累积分布函数有解析表达式（标准正态没有），故计算 logit 更为方便；而且 logit 的回归系数更易解释其经济意义。（2 分）

四、联立方程模型（共 2 小题，共 25 分）

1.（小计 10 分）

（1）（4 分）答：学生 A 和 B 构建的模型都会导致出现内生自变量问题（2 分）。A 同学的模型问题源自于遗漏了重要变量（ Abl_i ）；B 同学的模型问题源自于变量出现了测量误差，简单使用 IQ_i 作为 Abl_i 的代理变量，而忽略了二者之间存在一定测量误差（2 分）。

（2）（6 分）答：根据题意，我们容易证明 A 和 B 同学的模型构建都会使得自变量 Edu_i 为内生自变量，从而导致模型内生性问题，我们只需要证明 Edu_i 与模型随机干扰项相关即可。分别证明如下：

- 对于 A 同学的模型：易知 $v_i = \beta_3 Abl_i$ ，又因为已知能力与教育水平相关，也即 $cov(Edu_i, Abl_i) \neq 0$ ，因此有 $cov(Edu_i, v_i) \neq 0$ 。问题得证。
- 对于 B 同学的模型：因为存在变量测量误差，实际上 $Abl_i = IQ_i + Other_Abl_i$ （能力=智商+能力的其他方面）。因此有 $u_i = Other_Abl_i + \epsilon_i$ ，又因为已知能力的其他方面（ $Other_Abl_i$ ）也会与教育水平相关，也即 $cov(Edu_i, Other_Abl_i) \neq 0$ ，因此有 $cov(Edu_i, u_i) \neq 0$ 。问题得证。

2.（小计 15 分）

（1）（4 分）答：a) 根据经济学理论， PS_i = 松露替代品的市场价格， DI_i = 当地居民每月人均可支配收入，二者是商品消费的重要决定变量，因此方程(eq1)表达的是需求函数。b) 同理，因为 PF_i = 生产要素的价格（本例中为使用小猪来帮忙采掘松露的每小时小猪租借费用），主要表现为生产产量的重要决定因素，因此方程(eq)表达的是供给函数。

（2）（6 分）答：已知变量 PS_i 、 DI_i 和 PF_i 都是外生变量，而 P_i 和 Q_i 为内生变量，根据题目的记号要求，对应的约简化联立方程可以表达为：

$$\begin{cases} Q_i = \pi_1 + \pi_2 PF_i + \pi_3 PS_i + \pi_4 DI_i + v_{1i} \\ P_i = \lambda_1 + \lambda_2 PF_i + \lambda_3 PS_i + \lambda_4 DI_i + v_{2i} \end{cases}$$

说明： Q_i 和 P_i 顺序调换不影响作答正确性；随机干扰项最好表达为 v_i ，这样可以与结构化联立方程的随机干扰项相区别。

(3) (5分) 答：我们可以使用阶条件规则来进行判断。根据题干，已知整个结构化联立方程的内生变量个数总共有 $M = 2$ 个，前定变量个数总共有 $K = 3$ 个，因此两个方程的具体判断规则分别如下：

- 对于方程(eq1)，易知内生变量个数 $m = 2$ ，前定变量个数 $k = 2$ 。因此阶条件规则有 $\{K - k = 1\}$ 等于 $\{m - 1 = 1\}$ ，因此方程(eq1)是恰好识别的。
- 对于方程(eq2)，易知内生变量个数 $m = 2$ ，前定变量个数 $k = 1$ 。因此阶条件规则有 $\{K - k = 2\}$ 大于 $\{m - 1 = 1\}$ ，因此方程(eq2)是过度识别的。

说明：当然我们这里没有进一步讨论识别的秩条件。