

西北农林科技大学学术型硕士研究生课程考试试题（卷）

2022—2023 学年第 1 学期《计量经济学 II》课程 A 卷

考试课程：计量经济学 II

学年学期：2022-2023-1

试卷类型：A 卷

考试时间：2023-2-24

专业班级：

命题教师：

审题教师：

学生姓名：

学 号：

考试成绩：

一、计量经济学基础分析（共 4 小题，共 25 分）

得分：_____分

一项研究打算分析课程成绩与学生学积分及性别的关系。样本数据来自 856 名研究生的计量经济学课程考试情况。其中 $score_i$ 表示课程成绩（百分制）、 gpa_i 表示学积分（积分点数）、 $male_i$ 为性别虚拟变量（ $male_i = 1$ 表示男生， $male_i = 0$ 表示女生）。已知全体学生的平均学积分为 2.81。研究人员分别构建了如下 4 个计量模型并分别得到 OLS 估计结果（其中：第一行表示回归系数，第二行 s 表示标准误，第三行 t 统计量值，第四行 fit 表示拟合优度）：

- 【分析 1】成绩对学积分的线性模型及 OLS 估计：

$$\begin{array}{lcl} score_i = & +\alpha_1 + \alpha_2 gpa_i + u_{1i} & (1) \\ \widehat{score} = & +32.31 & +14.32 gpa_i \\ (s) & (2.0049) & (0.6997) \\ (t) & (+16.11) & (+20.46) \\ (fit) & R^2 = 0.3289 & \bar{R}^2 = 0.3281 \end{array}$$

- 【分析 2】成绩对性别和学积分的线性模型及 OLS 估计：

$$\begin{array}{lcll} score_i = & +\beta_1 + \beta_2 male_i + \beta_3 gpa_i + u_{2i} & (2) \\ \widehat{score} = & +29.66 & +3.83 male_i & +14.57 gpa_i \\ (s) & (2.0409) & (0.7418) & (0.6913) \\ (t) & (+14.53) & (+5.16) & (+21.08) \\ (fit) & R^2 = 0.3492 & \bar{R}^2 = 0.3477 \end{array}$$

- 【分析 3】成绩对性别、学积分、以及二者交互项的线性模型及 OLS 估计：

$$score_i = +\gamma_1 + \gamma_2 male_i + \gamma_3 gpa_i + \gamma_4 male * gpa_i + u_{3i} \quad (3)$$

$\widehat{score} =$	+30.36	+2.48male _i	+14.33gpa _i	+0.48male * gpa _i
(s)	(2.8578)	(3.9646)	(0.9845)	(1.3834)
(t)	(+10.62)	(+0.62)	(+14.56)	(+0.35)
(fit)	$R^2 = 0.3493$		$\bar{R}^2 = 0.3470$	

- 【分析 4】成绩对性别、学积分、以及性别和【去均值化后】学积分（也即 $gpa_dm_i = (gpa_i - 2.81)$ ）交互项的线性模型及 OLS 估计：

$$score_i = +\lambda_1 + \lambda_2 male_i + \lambda_3 gpa_i + \lambda_4 male * gpa_dm_i + u_{4i} \quad (4)$$

$\widehat{score} =$	+30.36	+3.82male _i	+14.33gpa _i	+0.48male * gpa _{_dm_i}
(s)	(2.8578)	(0.7422)	(0.9845)	(1.3834)
(t)	(+10.62)	(+5.15)	(+14.56)	(+0.35)
(fit)	$R^2 = 0.3493$		$\bar{R}^2 = 0.3470$	

根据上述信息，请依次按要求回答如下问题：

1. (6 分) 对比【分析 1】和【分析 2】的 OLS 估计结果，请你简单解释一下为什么后者的可决系数 R^2 大于前者？继续对比【分析 3】和【分析 4】的 OLS 估计结果，请你简单解释一下为什么二者的可决系数 R^2 会相等？

2. (7 分) 观察【分析 3】的 OLS 估计结果，关于性别 $male_i$ 估计系数的 t 检验是否显著？（给定理论查表值为 $t_{(0.975, 852)} = 1.96$ ）。如果给定某个女生的学积分为 $gpa_i = 2$ ，请计算出她的预期考试成绩是多少？（要求：写出主要理由；给出计算过程，结果保留两位小数。）

3. (6 分) 在【分析 3】中, 有人认为因为自变量 $male_i$ 和交互项 $male_i * gpi_i$ 之间有线性相关关系 (同样也可认为自变量 gpa_i 与上述交互项有线性相关关系), 因此导致模型出现多重共线性问题。你认为是否需要担心这个问题? 请简要说明你的理由。

4. (6 分) 对比【分析 3】和【分析 4】的 OLS 估计结果, 请你说明二者的估计结果有何异同? 根据上述比较, 你最终会更倾向于选择采用模型(3)还是采用模型(4)? 请简要说明你的理由。

二、时间序列分析（共 3 小题，共 25 分）

得分：_____分

考虑某资产 1983-2008 年每股季度收益率序列 r_t ，其时间序列图和 ACF 图如下所示（见图 1）：

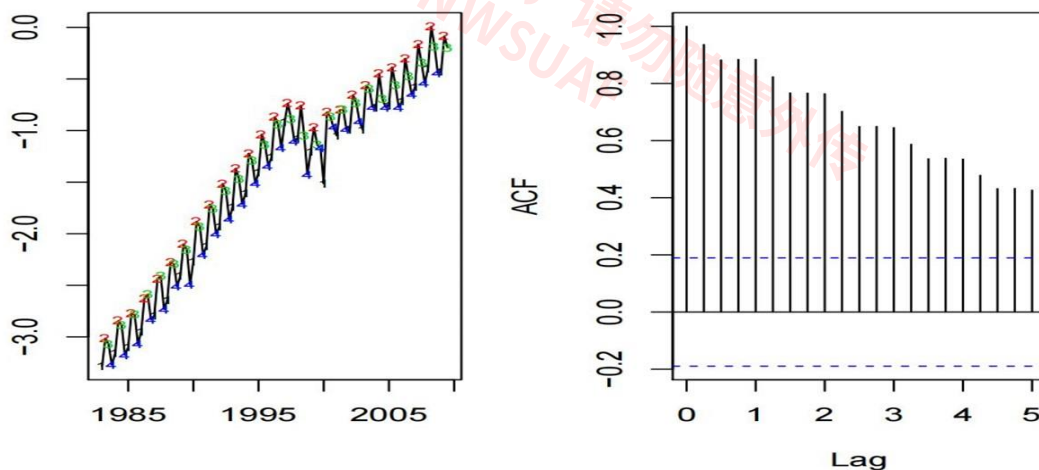


图 1: r_t 时序图和 ACF 图

x_t 序列为对 r_t 进行某种处理后得到的新序列，其时序图和 ACF 图如下（见图 2）：

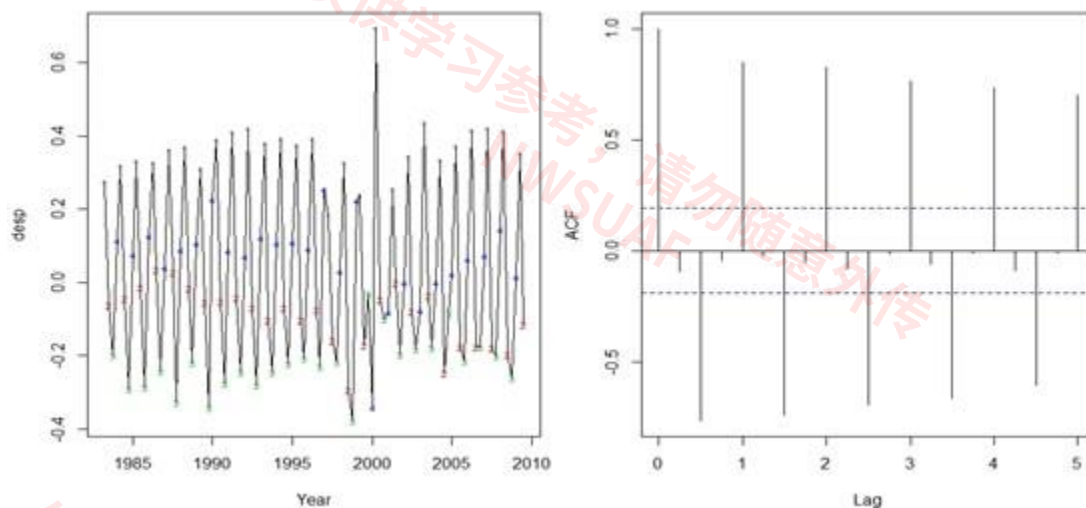


图 2: x_t 时序图和 ACF 图

z_t 序列为 x_t 进行某种处理后得到的新序列，其时序图和 ACF 图如下（见图 3）：

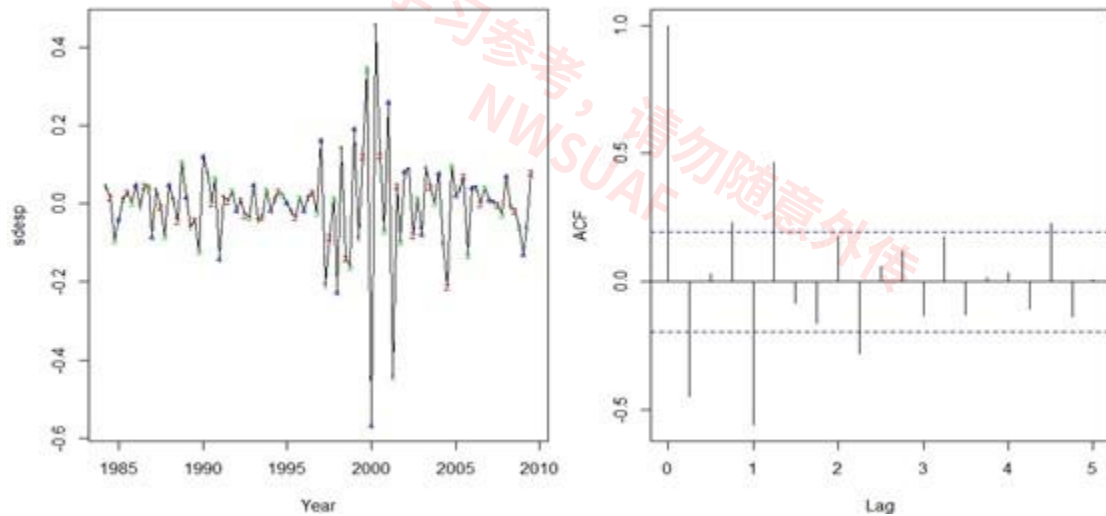


图 3: z_t 时序图和 ACF 图

1. (6 分) 根据图 1，请你判断 r_t 序列的平稳性，并说明原因。

2. (9 分) 结合前述图 1、图 2 和图 3，你认为序列 r_t 表现出什么样的特征？试写出适合 r_t 的时间序列模型。

3. (10 分) 若对于某资产的收益率序列 r_t 构建如下模型 (见式(5))。其中： a_t 为高斯白噪音序列， B 为滞后算子。

$$(1 - \varphi_1 B - \varphi_2 B^2)r_t = \varphi_0 + (1 - \theta_1 B - \theta_2 B^2)a_t \quad (5)$$

(1) (3 分) 该模型是什么模型? (要求: 写出名称和阶数。)

(2) (3 分) 若要基于该模型进行预测, 以 h 时刻为预测原点分别写出向前 1 步、2 步和 3 步预测值 $\hat{r}_h(1)$ 、 $\hat{r}_h(2)$ 和 $\hat{r}_h(3)$ 。

(3) (4 分) 若要判断该模型样本外预测能力, 可选用的指标有哪些? 请写出相应计算公式。

三、离散变量回归模型（共 4 小题，共 25 分）

得分：_____分

为研究某国女性参与劳动力市场情况，研究者随机在该国抽取 2000 名女性问她们参与了劳动力市场（ $work_i = 1$ ）还是未参与劳动力市场（ $work_i = 0$ ）；同时又收集了她们的年龄（ age_i ，用岁表示），婚姻状况（已婚则 $married_i = 1$ ，未婚则 $married_i = 0$ ），育有的子女数量（ $children_i$ ，用个表示），受教育程度（ $education_i$ ，用年表示）。请回答以下问题：

1.（5 分）若使用 $work_i$ 变量对其他所有变量回归并分别建立 probit 模型和线性概率模型（LPM），请问在此例中哪个模型更为合适？请说明理由。

2.（8 分）若问题 1 中建立的 probit 模型，使用 STATA 软件输出回归结果见下面图 4，请问女性是否参与劳动力市场依赖于哪些因素？并对该模型从整体的显著性和拟合优度等进行整体评价。

Probit regression		Number of obs = 2000				
		LR chi2(4) = 478.32				
		Prob > chi2 = 0.0000				
Log likelihood = -1027.0616		Pseudo R2 = 0.1889				

work	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

age	.0347211	.0042293	8.21	0.000	.0264318	.0430105
married	.4308575	.074208	5.81	0.000	.2854125	.5763025
children	.4473249	.0287417	15.56	0.000	.3909922	.5036576
education	.0583645	.0109742	5.32	0.000	.0368555	.0798735
_cons	-2.467365	.1925635	-12.81	0.000	-2.844782	-2.089948

图 4: Probit 模型回归结果

3. (8 分) 若计算 Probit 模型的平均边际效应结果见下面图 5，请解释各解释变量的边际效应。

Marginal effects after probit
y = Pr(work) (predict)
= .71835948

variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
age	.011721	.00142	8.25	0.000	.008935	.014507		36.208
married*	.150478	.02641	5.70	0.000	.098716	.20224		.6705
children	.1510059	.00922	16.38	0.000	.132939	.169073		1.6445
educat~n	.0197024	.0037	5.32	0.000	.012442	.026963		13.084

(*) dy/dx is for discrete change of dummy variable from 0 to 1

图 5: Probit 模型的边际效应

4. (4 分) 若以上所有回归使用 Logit 模型重新估计，你预测 Logit 模型与 Probit 模型的估计结果是否有很大差异？相比 Probit 模型而言，Logit 模型是否有优势？请说明理由。

四、联立方程模型（共 2 小题，共 25 分）

得分：_____分

1.（小计 10 分）假定工资水平的“真实模型”为： $Wage_i = \beta_1 + \beta_2 Edu_i + \beta_3 Abl_i + \epsilon_i$ ，其中随机干扰项 $\epsilon_i \sim \text{i.i.d } N(0, \sigma^2)$ 。进一步假定个体受教育水平 edu_i 与其能力 abl_i 存在线性相关关系。

学生 A 基于某些考虑构建了一个可操作的计量分析模型 $Wage_i = \alpha_1 + \alpha_2 Edu_i + v_i$ ，其中 v_i 为随机干扰项。

学生 B 认为能力变量 Abl_i 不好测量，但是可以用智商分值 IQ_i 替代能力变量，因此构建了一个可操作的计量分析模型 $Wage_i = \gamma_1 + \gamma_2 Edu_i + \gamma_3 IQ_i + u_i$ ，其中 u_i 为随机干扰项。

根据上述信息，请回答下列问题：

（1）（4 分）学生 A 和学生 B 构建的计量分析模型分别会存在什么问题？引起这些问题的主要原因分别是什么？

（2）（6 分）请你简要证明：在这些情形下，学生 A 构建的模型中 $cov(Edu_i, v_i) \neq 0$ ；学生 B 构建的模型中 $cov(Edu_i, u_i) \neq 0$ 。

2. (小计 15 分) 松露是一种美味的食材，它们是生长在地下的食用菌。研究人员构建了如下的结构化联立方程模型：

$$\begin{cases} Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 PS_i + \alpha_4 DI_i + u_{1i} & (\text{eq1}) \\ Q_i = \beta_1 + \beta_2 P_i + \beta_3 PF_i + u_{2i} & (\text{eq2}) \end{cases}$$

其中： Q_i = 松露的交易量； P_i = 松露的市场价格； 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； 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）； u_{1i} 和 u_{2i} 表示随机干扰项。

根据上述信息，请回答下列问题：

(1) (4 分) 根据经济学理论，请你判断上述结构化联立方程中方程 (eq1) 和方程 (eq2) 中，哪个是供给方程，哪个是需求方程？

(2) (6 分) 给定变量 PS_i 、 DI_i 和 PF_i 都是外生变量，请你根据题干中的结构化联立方程，写出其对应的约简化联立方程。（要求：约简参数记号请分别使用 $\pi_1, \pi_2, \dots$ ，以及 $\lambda_1, \lambda_2, \dots$ 。随机干扰项用 v_{1i} 和 v_{2i} 表示。）

(3) (5 分) 在采用两阶段最小二乘法 (2SLS) 估计上述结构化联立方程之前，请你判断其中方程 (eq1) 和方程 (eq2) 是否可以识别（不可识别、恰好识别或过度识别）？说明你的判断理由。