计量经济学II
(Econometrics II)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2021-09-26

模块1：计量经济学基础

第04章 扩展方法

4.1 模型函数形式

4.2 虚拟变量模型

4.1 模型函数形式

4.1.1 过原点回归

过原点回归的模型形式

过原点回归（regression through the origin）：没有截距项的线性模型

在实践中，双变量PRM过原点回归采取如下的形式：

$$Y_i = \beta_2 X_i + u_i$$

适用于这种模型的例子：

- 弗里德曼的持久收入假说(permanent income hypothesis)；
- 资本资产定价模型(the capital Asset Pricing Model, CAPM)等。

资本资产定价模型（CAPM）

资本资产定价模型(the capital Asset Pricing Model, CAPM):

$$\left(ER_i - r_f\right) = \beta_i \left(ER_m - r_f\right)$$

其中:

- ER_i证券 i的期望回报率;
- ER_m市场证券组合的期望回报率（如标准普尔S&P500综合股票指数）;
- r_f为无风险回报率（90天国债回报率）。
- β_i为系数，表明第 i种证券回报率与市场互动程度的度量。(注:不要把这个 β_i和双变量回归的斜率系数 β_2混同起来。)

一个大于1的 β_i意味着证券 i是一种易波动或进攻型证券；一个小于1的 β_i意味着证券 i是一种防御型证券。

资本资产定价模型（CAPM）

$$R_i - r_f = \beta_i \left(R_m - r_f\right) + u_i$$
$$R_i - r_f = \alpha_i + \beta_i \left(R_m - r_f\right) + u_i$$

- 如果CAPM成立，则预期 α_i 为0。
- 这样的模型如何估计呢？

这类模型的SRM可以写成：

$$Y_i = \hat{\beta}_2 X_i + \mathrm{e}_i$$

OLS方法下求解回归系数：

$$\sum \mathrm{e}_i^2 = \sum \left(Y_i - \hat{\beta}_2 X_i\right)^2$$

$$\begin{aligned}
\frac{\partial \sum \mathrm{e}_i^2}{\partial \hat{\beta}_2} &= 2\sum \left(Y_i - \hat{\beta}_2 X_i\right)(-X_i) = 0 \\
\hat{\beta}_2 &= \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i (\beta_2 X_i + u_i)}{\sum X_i^2} = \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2} \\
E\left(\hat{\beta}_2\right) &= \beta_2
\end{aligned}$$

OLS方法下求解得到的方差：

$$
\begin{aligned}
Var\left(\hat{\beta}_2\right) &= E\left(\hat{\beta}_2 - \beta_2\right)^2 \\
&= E\left[\frac{\sum X_i u_i}{\sum X_i^2}\right]^2 \\
&= \frac{\sigma^2}{\sum X_i^2}
\end{aligned}
$$

$$
\hat{\sigma}^2 = \frac{\sum e_i^2}{n-1}; \quad \mathrm{E}\left(\hat{\sigma}^2\right) = \sigma^2
$$

OLS估计量对比：无截距和有截距的差异：

$$Y_i = \hat{\beta}_2 X_i + \mathrm{e}_i$$

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$$

$$Var\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum X_i^2}$$

$$\hat{\sigma}^2 = \frac{\sum \mathrm{e}_i^2}{n-1}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \mathrm{e}_i$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$Var\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2}$$

$$\hat{\sigma}^2 = \frac{\sum \mathrm{e}_i^2}{n-2}$$

- 第一，对有截距项的模型来说，总有 $\sum e_i = 0$；对无截距项的模型来说，$\sum e_i = 0$ 不一定成立，只有 $\sum e_i X_i = 0$成立。
- 第二，对有截距项的模型，判定系数 $r^2 \geq 0$；但是，对无截距模型来说，r^2时可能出现负值。

资本资产定价模型（CAPM）

过原点回归的判定系数 r^2的计算公式如下：

$$TSS=\sum y_i^2=\sum Y_i^2-n\overline{Y}^2$$
$$RSS=\sum e_i^2=\sum Y_i^2-\hat{\beta}_2^2\sum X_i^2$$
$$r^2=1-\frac{RSS}{TSS}>0;\quad r^2=1-\frac{RSS}{TSS}<0;$$

因此，对于无截距模型，我们给出拟合优度指标为**毛判定系数**（Raw r^2）：

$$Raw\quad r^2=\frac{\sum(X_iY_i)^2}{\sum X_i^2\sum Y_i^2}$$

- 第一，尽管模型含有截距项，但若该项的出现是统计上不显著的(即统计上等于零)，则从任何实际方面考虑，都可认为这个结果是一个过原点回归模型。
- 第二，如果在模型中确实有截距，而我们却执意拟合一个过原点回归，我们就犯了**设定错误**。

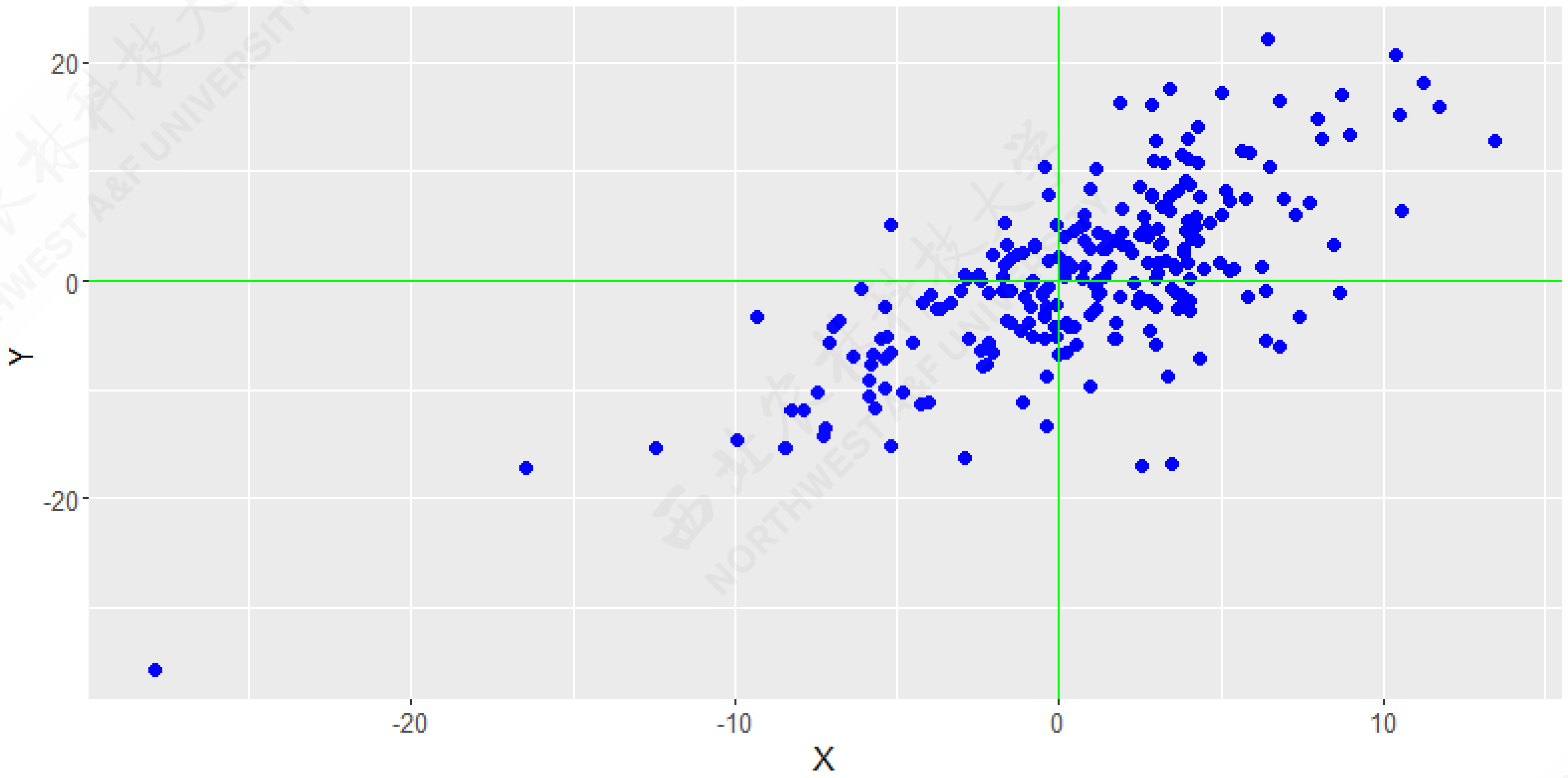
1980.01-1999.12年间104 种股票构成的一个指数的超额回报率 Y_t(%)和英国总体股票指数的超额回报率 X_t(%)的月度数据共n=240个月观测。其中超额回报率指的是超过无风险资产回报率的部分。

year	month	X	Y
1980	1	7.2634	6.0802
1980	2	6.3399	-0.9242
1980	3	-9.2852	-3.2862
1980	4	0.7933	5.2120
1980	5	-2.9024	-16.1642
1980	6	8.6132	-1.0547
1980	7	3.9821	11.1724

Showing 1 to 7 of 240 entries

Previous 1 2 3 4 5 … 35 Next

下面先直接给出二者的散点图：



两类模型回归结果对比：

无截距模型：

$$Y_i = \hat{\beta}_2 X_i + e_i$$

$$\begin{array}{ll}
\widehat{Y} = & +1.16X \\
(\mathrm{t}) & (15.5320) \\
(\mathrm{se}) & (0.0744) \\
(\mathrm{fitness}) & R^2 = 0.5023; \bar{R}^2 = 0.5003 \\
 & F^* = 241.24; p = 0.0000
\end{array}$$

有截距模型：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

$$\begin{array}{llll}
\widehat{Y} = & -0.45 & & +1.17X \\
(\mathrm{t}) & (-1.2329) & & (15.5350) \\
(\mathrm{se}) & (0.3629) & & (0.0754) \\
(\mathrm{fitness}) & R^2 = 0.5035; \bar{R}^2 = 0.5014 & & \\
 & F^* = 241.34; p = 0.0000 & &
\end{array}$$

4.1.2 尺度与测量单位

案例数据

在回归分析中，因变量Y和解释变量X的测量单位的不同会造成回归结果的差异吗？

Year	GPDIB	GPDIM	GDPB	GDPM	GPDIB_std	GDPB_std
1990	886.6	886600	7112.5	7112500	-1.2942	-1.3459
1991	829.1	829100	7100.5	7100500	-1.4624	-1.3550
1992	878.3	878300	7336.6	7336600	-1.3185	-1.1768
1993	953.5	953500	7532.7	7532700	-1.0986	-1.0287
1994	1042.3	1042300	7835.5	7835500	-0.8389	-0.8002

Showing 1 to 5 of 16 entries

Previous 1 2 3 4 Next

- GPDIB = 以2000年**10亿**(Billions)美元计国内私人总投资； GPDIM = 以2000年**百万**(millions)美元计国内私人总投资；

尺度变换

把某一测量单位下的回归模型，变换为另一测量单位的回归模型：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \mathrm{e}_i$$
$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i + \mathrm{e}_i^*$$

尺度因子： $\omega_1;\omega_2$分别表示为Y和X的尺度因子。

$$Y_i^* = \omega_1 Y_i \quad X_i^* = \omega_2 X_i$$

如果 $(Y_i;X_i)$都是以**10亿**(billion)美元计量的，我们把它们改为用**百万**(million)美元去度量，就会有：

$$Y_i^* = 1000Y_i; \quad X_i^* = 1000X_i; \quad \omega_1 = \omega_2 = 1000$$

OLS估计

进行数据转换，新模型的OLS估计量如下：

$$
\begin{aligned}
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \mathrm{e}_i \\
Y_i^* &= \hat{\beta}_1^* + \hat{\beta}_2^* X_i + \mathrm{e}_i^* \\
Y_i^* &= \omega_1 Y_i; \quad X_i^* = \omega_2 X_i; \quad e_i^* = \omega_1 e_i
\end{aligned}
$$

$$
\begin{aligned}
\hat{\beta}_2* &= \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} && \Leftarrow \operatorname{var}\left(\hat{\beta}_2^*\right) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \\
\hat{\beta}_1^* &= \overline{Y}^* - \hat{\beta}_2^* \overline{X}^* && \Leftarrow \operatorname{var}\left(\hat{\beta}_1^*\right) = \frac{\sum X_i^{*2}}{n \sum x_i^{*2}} \cdot \sigma^{*2} \\
\hat{\sigma}^{*2} &= \frac{\sum e_i^{*2}}{n-2}
\end{aligned}
$$

OLS估计

进行数据转换，两个模型下OLS估计量有如下关系：

$$
\begin{aligned}
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \mathrm{e}_i \\
Y_i^* &= \hat{\beta}_1^* + \hat{\beta}_2^* X_i + \mathrm{e}_i^* \\
Y_i^* &= \omega_1 Y_i; \quad X_i^* = \omega_2 X_i; \quad e_i^* = \omega_1 e_i
\end{aligned}
$$

$$
\begin{aligned}
\hat{\beta}_2^* &= \left(\frac{\omega_1}{\omega_2}\right) \hat{\beta}_2 && \Leftarrow \operatorname{var}\left(\hat{\beta}_2^*\right) = \left(\frac{\omega_1}{\omega_2}\right)^2 \operatorname{var}\left(\hat{\beta}_2\right) \\
\operatorname{var}\left(\hat{\beta}_2^*\right) &= \left(\frac{\omega_1}{\omega_2}\right)^2 \operatorname{var}\left(\hat{\beta}_2\right) && \Leftarrow \operatorname{var}\left(\hat{\beta}_1^*\right) = \omega_1^2 \operatorname{var}\left(\hat{\beta}_1\right) \\
\hat{\sigma}^{*2} &= \omega_1^2 \hat{\sigma}^2 \\
r_{xy}^2 &= r_{x*y^*}^2
\end{aligned}
$$

相关结论

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \mathrm{e}_i$$
$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i + \mathrm{e}_i^*$$
$$Y_i^* = \omega_1 Y_i; \quad X_i^* = \omega_2 X_i; \quad e_i^* = \omega_1 e_i$$

模型对比，得出如下主要结论：

- $\omega_1 = \omega_2$，即尺度因子相等时，**斜率系数及其标准误**不受尺度从（Y_i, X_i）到（Y_i^*, X_i^*）的影响。**截距及其标准误**却放大或缩小至 ω_1 倍。
- X_i尺度不变 $\omega_2 = 1$，Y_i尺度因子 ω_1变化，那么，斜率和截距系数以及它们各自的标准误都要乘以同样的因子 ω_1。
- Y_i尺度不变 $\omega_1 = 1$，而 X_i尺度因子 ω_2变化，那么，斜率系数及其标准误都要乘以因子 $1/\omega_1$，而截距系数及其标准误不变。

案例分析结果

GPDI和GDP都以十亿美元计算：

$$
\begin{array}{lll}
\widehat{GPDIB} = & -926.09 & +0.25GDPB \\
(\mathrm{t}) & (-7.9590) & (19.5824) \\
(\mathrm{se}) & (116.3577) & (0.0129) \\
(\mathrm{fitness}) & R^2 = 0.9648; \bar{R}^2 = 0.9623 & \\
 & F^* = 383.47; p = 0.0000 &
\end{array}
$$

GPDI和GDP都以百万美元计算：

$$
\begin{array}{lll}
\widehat{GPDIM} = & -926090.39 & +0.25GDPM \\
(\mathrm{t}) & (-7.9590) & (19.5824) \\
(\mathrm{se}) & (116357.6965) & (0.0129) \\
(\mathrm{fitness}) & R^2 = 0.9648; \ \bar{R}^2 = 0.9623 & \\
 & F^* = 383.47; \ p = 0.0000 &
\end{array}
$$

GPDI以十亿美元计算而GDP以百万美元计算：

$$
\begin{array}{lll}
\widehat{GPDIB} = & -926.09 & +0.00GDPM \\
(\mathrm{t}) & (-7.9590) & (19.5824) \\
(\mathrm{se}) & (116.3577) & (0.0000) \\
(\mathrm{fitness}) & R^2 = 0.9648; \bar{R}^2 = 0.9623 & \\
 & F^* = 383.47; p = 0.0000 &
\end{array}
$$

GPDl以百万美元计算而GDP 以十亿美元计算：

4.1.3 标准化变量回归

标准化变量回归

假设如下双变量回归：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \mathrm{e}_i$$

对Y和X作如下标准化变换，得到相应的**标准化变量**：

$$Y_i^* = \frac{Y_i - \overline{Y}}{S_Y}; \quad X_i^* = \frac{X_i - \overline{X}}{S_X}$$

得到如下新的双变量回归模型:

$$\begin{aligned} Y_i^* &= \hat{\beta}_1^* + \hat{\beta}_2^* X_i + e_i^* \\ &= \hat{\beta}_2^* X_i + e_i^* \end{aligned}$$

- 标准化变量的特征是：其均值总是0 和标准差总是1。
- 对标准化的回归子和回归元做回归，截距项总是零!
- 实际上变成了过原点回归模型!

标准化变量回归

模型比较与结论：

- 第一，由于标准化回归本质上是一个过原点回归，而我们在已经指出通常过原点回归的不能使用 r^2，所以我们就没有给出其 r^2值。
- 第二，传统模型的系数与这里的系数之间存在一种有趣的关系。在双变量情形中，这种关系如下（证明过程略：自学练习题！）：

$$\hat{\beta}_2^* = \frac{\mathrm{S}_X}{\mathrm{S}_\mathrm{Y}} \hat{\beta}_2$$

- 第三，在多元回归中，变量标准化可以去除多个自变量之间数量尺度(量纲)的差别，因而具有一定的优点！

下面我们对以十亿美元计的GPDIB和GDPB进行标准化数据变换:

Year	GPDIB	GDPB	GPDIB_std	GDPB_std
1990	886.6	7112.5	-1.2942	-1.3459
1991	829.1	7100.5	-1.4624	-1.3550
1992	878.3	7336.6	-1.3185	-1.1768
1993	953.5	7532.7	-1.0986	-1.0287
1994	1042.3	7835.5	-0.8389	-0.8002
1995	1109.6	8031.7	-0.6421	-0.6521
1996	1209.2	8328.9	-0.3508	-0.4277
1997	1320.6	8703.5	-0.0250	-0.1450

Showing 1 to 8 of 16 entries

Previous 1 2 Next

GPDI和GDP都以十亿美元计算：

$$
\begin{aligned}
&\widehat{GPDIB} = && -926.09 && +0.25GDPB \\
&\text{(t)} && (-7.9590) && (19.5824) \\
&\text{(se)} && (116.3577) && (0.0129) \\
&\text{(fitness)} && R^2 = 0.9648; \bar{R}^2 = 0.9623 \\
& && F^* = 383.47; p = 0.0000
\end{aligned}
$$

标准化变量后的模型估计：

$$
\begin{aligned}
&\widehat{GPDIB}_s td = +0.98GDPB_{std} \\
&\text{(t)} && (20.2697) \\
&\text{(se)} && (0.0485) \\
&\text{(fitness)} && R^2 = 0.9648; && \bar{R}^2 = 0.9624 \\
& && F^* = 410.86; && p = 0.0000
\end{aligned}
$$

回归模型的函数形式

- 对数线性模型
- 半对数模型
- 倒数模型

4.1.4 对数线性模型

对数线性模型的形式

指数回归模型（exponential regression model）

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$$

可化为：

$$\begin{aligned}
\ln Y_i &= \ln \beta_1 + \beta_2 \ln X_i + u_i \\
\ln Y_i &= \alpha + \beta_2 \ln X_i + u_i \qquad \Leftarrow \alpha = \ln \beta_1
\end{aligned}$$

这种模型被称为对数-对数(log-log)，双对数(double-log)或对数一线性(log-linear)模型。进而有：

$$Y_i^* = \alpha + \beta_2 X_i^* + u_i \qquad \Leftarrow \left[Y_i^* = \ln Y_i; \quad X_i^* = \ln X_i \right]$$

从而可用OLS方法可以得到BLUE估计量：

$$Y_i^* = \hat{\alpha} + \hat{\beta}_2 X_i^* + \mathrm{e}_i$$

双数线性模型：

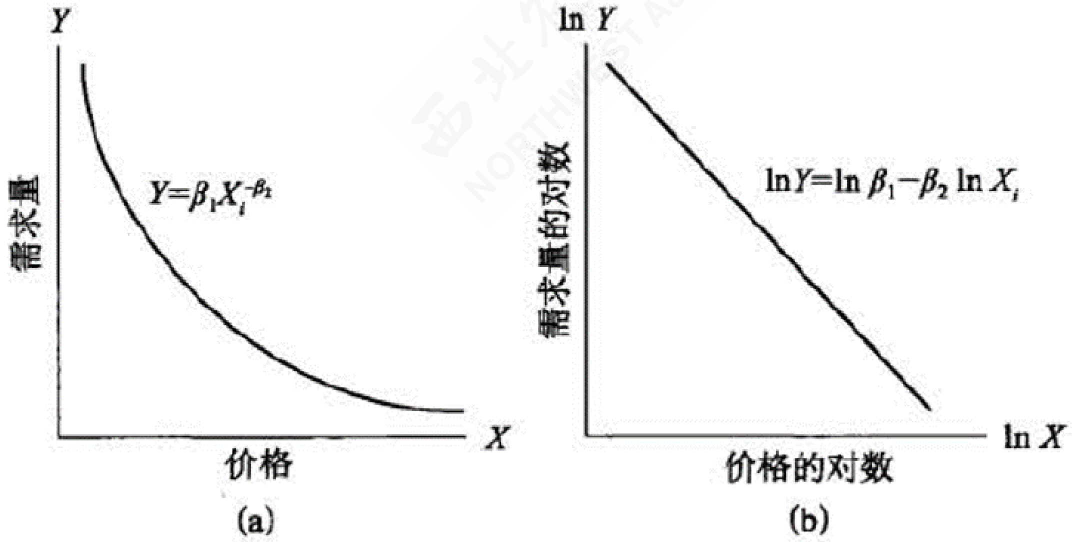
$$\begin{aligned} \ln Y_i &= \alpha + \beta_2 \ln X_i + u_i \\ Y_i^* &= \hat{\alpha} + \hat{\beta}_2 X_i^* + \mathrm{e}_i \qquad\qquad \Leftarrow \hat{\alpha} = \ln \hat{\beta}_1 \end{aligned}$$

$$\beta_2 = \frac{d(\ln Y)}{d(\ln X)} = \frac{\frac{1}{Y} dY}{\frac{1}{X} dX} = \frac{dY/Y}{dX/X}$$

斜率就是Y对X的弹性！如果Y代表商品需求量Q，X代表商品价格P，则就表示该商品的需求价格弹性。

双数线性模型有如下性质：

- Y对X的弹性在整个研究范围内是常数，一直为 β_2，因此这种模型也称为不变弹性模型(constant elasticity model)。
- 虽然 $\hat{\alpha}$和 $\hat{\beta}_2$是无偏估计量，但是进入原始模型的参数 β_1的估计值 $\hat{\beta}_1$却是有偏估计，而且 $\beta_1 = antilog\hat{\alpha}$。



耐用品消费案例

耐用品支出与个人消费总支出的关系：

obs	EXPDUR	PCEXP	ln_expdur	ln_pcexp
2003-I	971.4	7184.9	6.8787	8.8797
2003-II	1009.8	7249.3	6.9175	8.8887
2003-III	1049.6	7352.9	6.9562	8.9029
2003-IV	1051.4	7394.3	6.9579	8.9085
2004-I	1067	7479.8	6.9726	8.9200
2004-II	1071.4	7534.4	6.9767	8.9272
2004-III	1093.9	7607.1	6.9975	8.9368

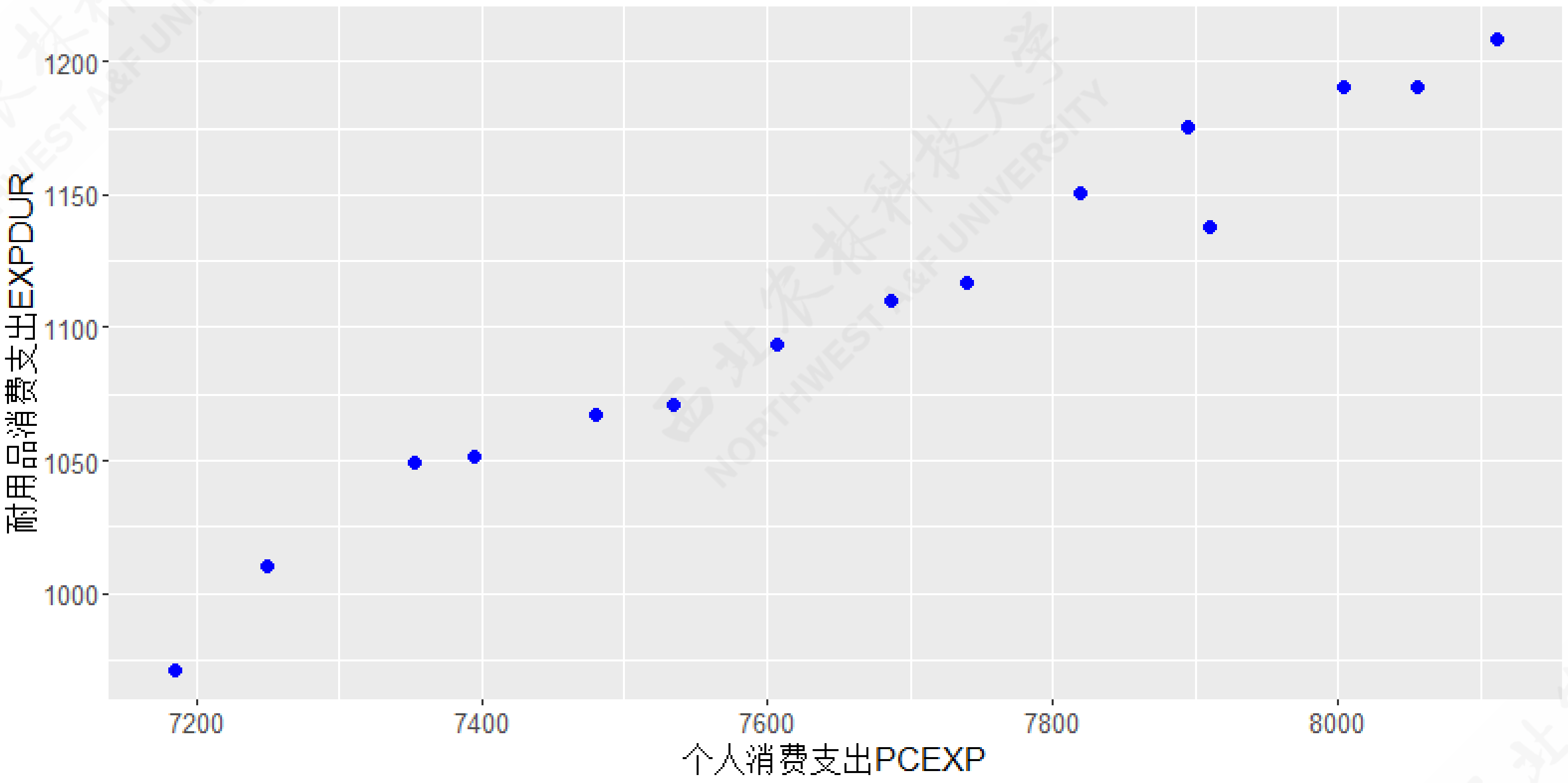
Showing 1 to 7 of 15 entries

Previous 1 2 3 Next

其中：PCEXP=个人消费支出, EXPDUR=耐用品消费支出， 单位10亿美元（按2000年价格计）

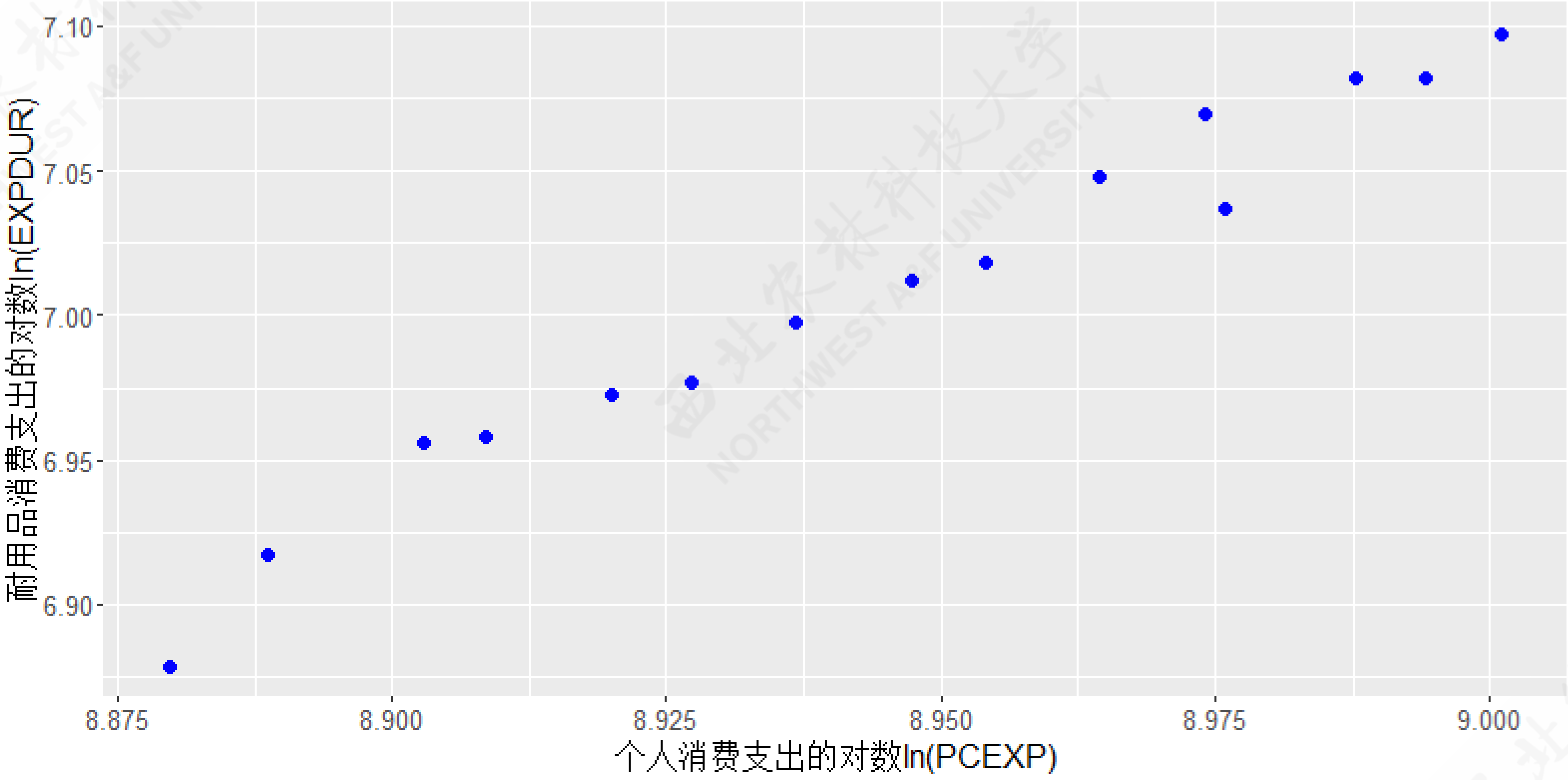
假设我们想求出耐用品支出对个人消费总支出的**斜率**。

将耐用品支出相对于个人消费总支出做散点图：



假设我们想求出耐用品支出对个人消费总支出的**弹性**。

将耐用品支出的**对数**相对于个人消费总支出的**对数**做散点图：



耐用品消费案例中，我们可以实证得到如下的**双对数模型**：

$$
\begin{aligned}
&\widehat{log(EXPDUR)} = -7.54 && +1.63log(PCEXP)\\
&\text{(t)} \quad (-10.5309) && (20.3152)\\
&\text{(se)} \quad (0.7161) && (0.0801)\\
&\text{(fitness)} \quad R^2 = 0.9695; \bar{R}^2 = 0.9671\\
&\qquad F^* = 412.71; p = 0.0000
\end{aligned}
$$

4.1.5 半对数模型

线性到对数模型

怎样测量增长率？经济学家、企业人员与政府常常对于求出某些经济变量的增长率感兴趣，如人口、GNP、货币供给、就业、生产力、贸易赤字等。

$$Y_t = Y_0(1+r)^t$$

Y_t＝时期t的劳务实际支出； Y_0＝劳务实际支出的初始值(为2002年第四季度末的值)；r是Y的复合增长率。

线性到对数模型

劳务支出数据

obs	t	EXPSERVICES	ln_expservices
2003-I	1	4143.3	8.3292
2003-II	2	4161.3	8.3336
2003-III	3	4190.7	8.3406
2003-IV	4	4220.2	8.3476
2004-I	5	4268.2	8.3589
2004-II	6	4308.4	8.3683
2004-III	7	4341.5	8.3760
2004-IV	8	4377.4	8.3842

Showing 1 to 8 of 15 entries

Previous 1 2 Next

半对数模型(semilog models)：

- 线性到对数模型(log-lin model)：只有回归子Y取对数
- 对数到线性模型(lin-log model)：只有回归元X取对数

线性到对数模型

半对数模型的形式：

$$\begin{aligned} Y_t &= Y_0(1+r)^t \\ \ln Y_t &= \ln Y_0 + t\ln(1+r) \\ \ln Y_t &= \beta_1 + \beta_2 t && \Leftarrow [\beta_1 = \ln Y_0; \quad \beta_2 = \ln(1+r)] \\ \ln Y_t &= \beta_1 + \beta_2 t + u_t \end{aligned}$$

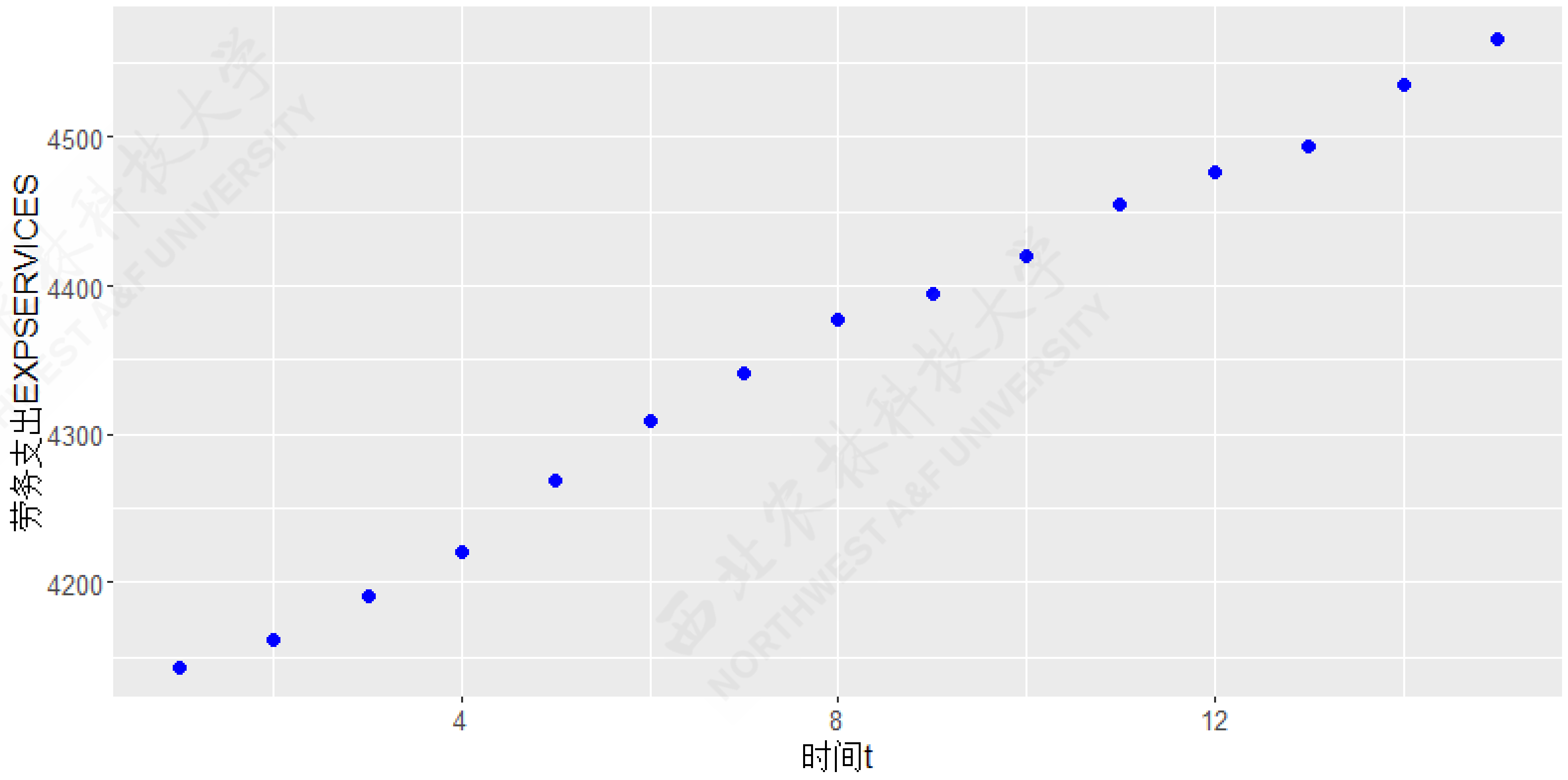
斜率 β_2的经济学含义：

$$\beta_2 = \frac{d\ln Y}{dt} = \frac{dY/Y}{dt}$$

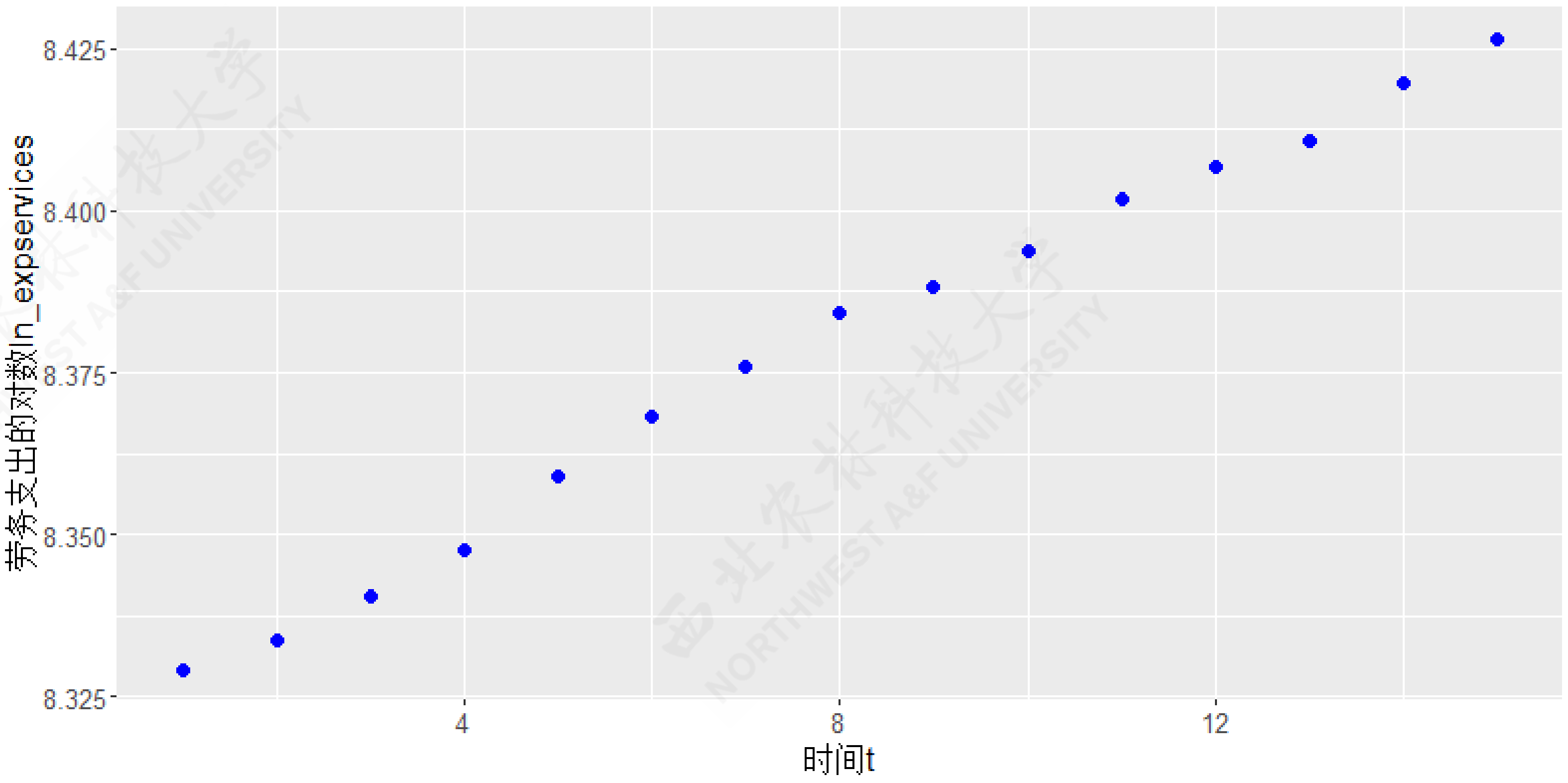
恒定相对增长率模型：上述模型描述了因变量Y的恒定相对增长率

- 恒定相对增长模型： $\beta_2 > 0$
- 恒定相对衰减模型： $\beta_2 < 0$

散点图I



散点图2



半对数模型：线性到对数模型

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \qquad \Leftarrow [\beta_1 = \ln Y_0; \quad \beta_2 = \ln(1+r)]$$

$$\begin{aligned}
log(\widehat{EXPSERVICES}) &= +8.32 && +0.01t \\
\text{(t)} & \quad (5186.2999) && (39.9648) \\
\text{(se)} & \quad (0.0016) && (0.0002) \\
\text{(fitness)} & \quad R^2 = 0.9919; && \bar{R}^2 = 0.9913 \\
& \quad F^* = 1597.18; p = 0.0000
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_2 &= \ln(1+r) = 0.00705 \\
r &= \text{antilog}\left(\hat{\beta}_2\right) - 1 \\
&= \text{antilog}(0.00705) - 1 \\
&= 0.00708
\end{aligned}$$

- $\hat{\beta}_2 = 0.00705$表示瞬时增长率
- $r = 0.00708$表示复合增长率

回归结果比较

下面做一个对比模型。线性趋势模型：Y直接对时间t回归：

$$Y_t = \beta_1 + \beta_2 t + u_t$$

$$
\begin{aligned}
\widehat{EXPSERVICES} &= +\,4111.54 && +\,30.67t \\
\text{(t)} & \quad (655.5628) && (44.4671) \\
\text{(se)} & \quad (6.2718) && (0.6898) \\
\text{(fitness)} & \quad R^2 = 0.9935; && \bar{R}^2 = 0.9930 \\
& \quad F^* = 1977.32; p = 0.0000 &&
\end{aligned}
$$

解释如下： 在2003年第1季度至2006年第3季度期间，劳务支出以每季度约300 亿美元的绝对速度(注意不是相对速度)增加，即劳务支出有上涨的趋势。

对数到线性模型(lin-log model)

如果我们的目的是测量X的一个百分比变化时，Y的绝对变化量，则要用**对数到线性模型**（lin-log model）。

$$Y_t = \beta_1 + \beta_2 \ln X_i + u_i$$

$$\beta_2 = \frac{dY}{d\ln X} = \frac{dY}{dX/X} = \frac{\Delta Y}{\Delta X/X}$$

$$\Delta Y = \beta_2 \frac{\Delta X}{X}$$

例如：恩格尔支出(Engel expenditure) 模型：

- “用于食物的总支出以算术级数增加，而总支出以几何级数增加。”

食物支出（`foodexp`）与家庭总支出（`totalexp`）的关系：

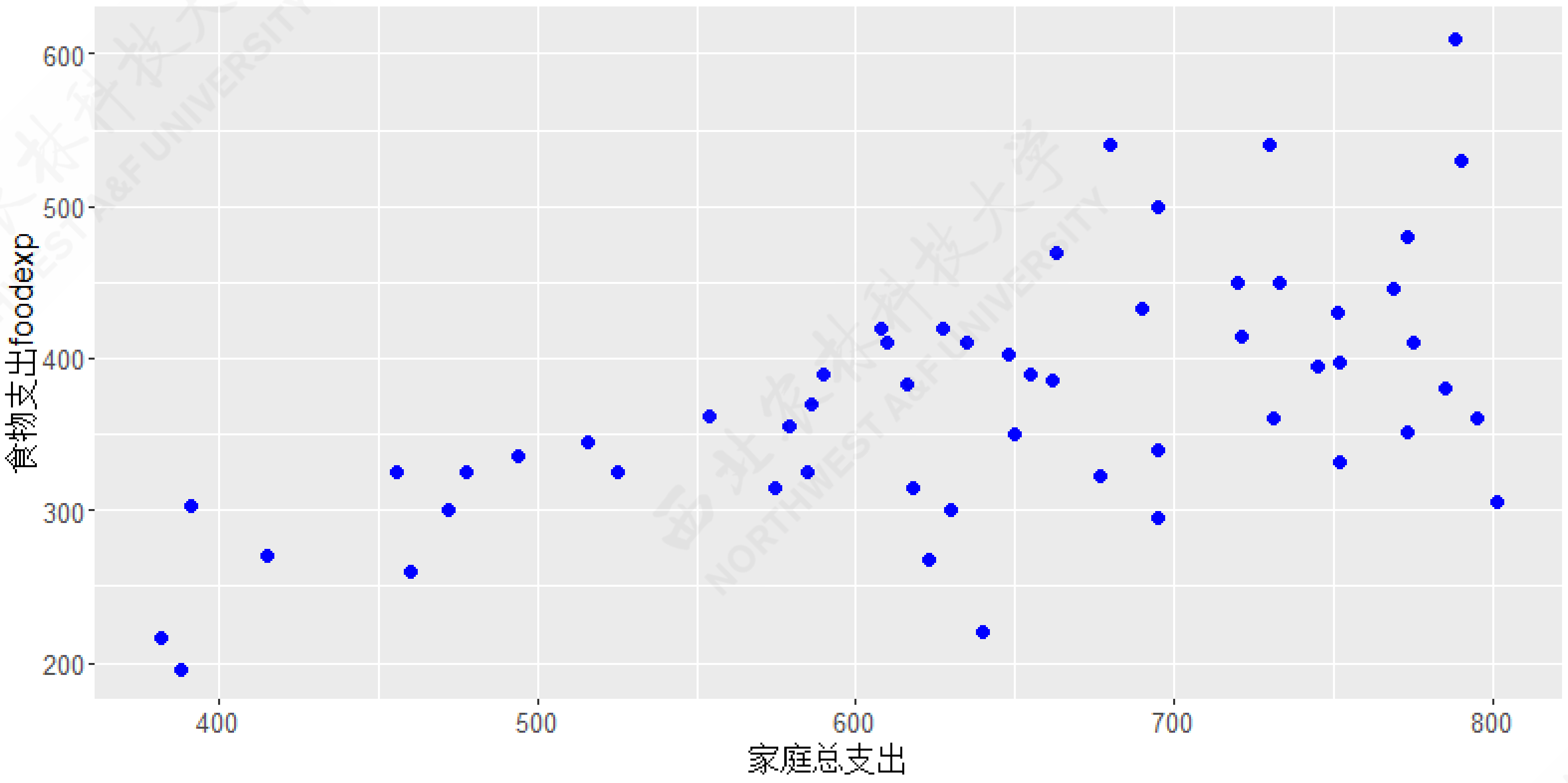
obs	foodexp	totalexp	ln_totalexp
1	217	382	5.9454
2	196	388	5.9610
3	303	391	5.9687
4	270	415	6.0283
5	325	456	6.1225
6	260	460	6.1312
7	300	472	6.1570
8	325	478	6.1696

Showing 1 to 8 of 55 entries

Previous 1 2 3 4 5 6 7 Next

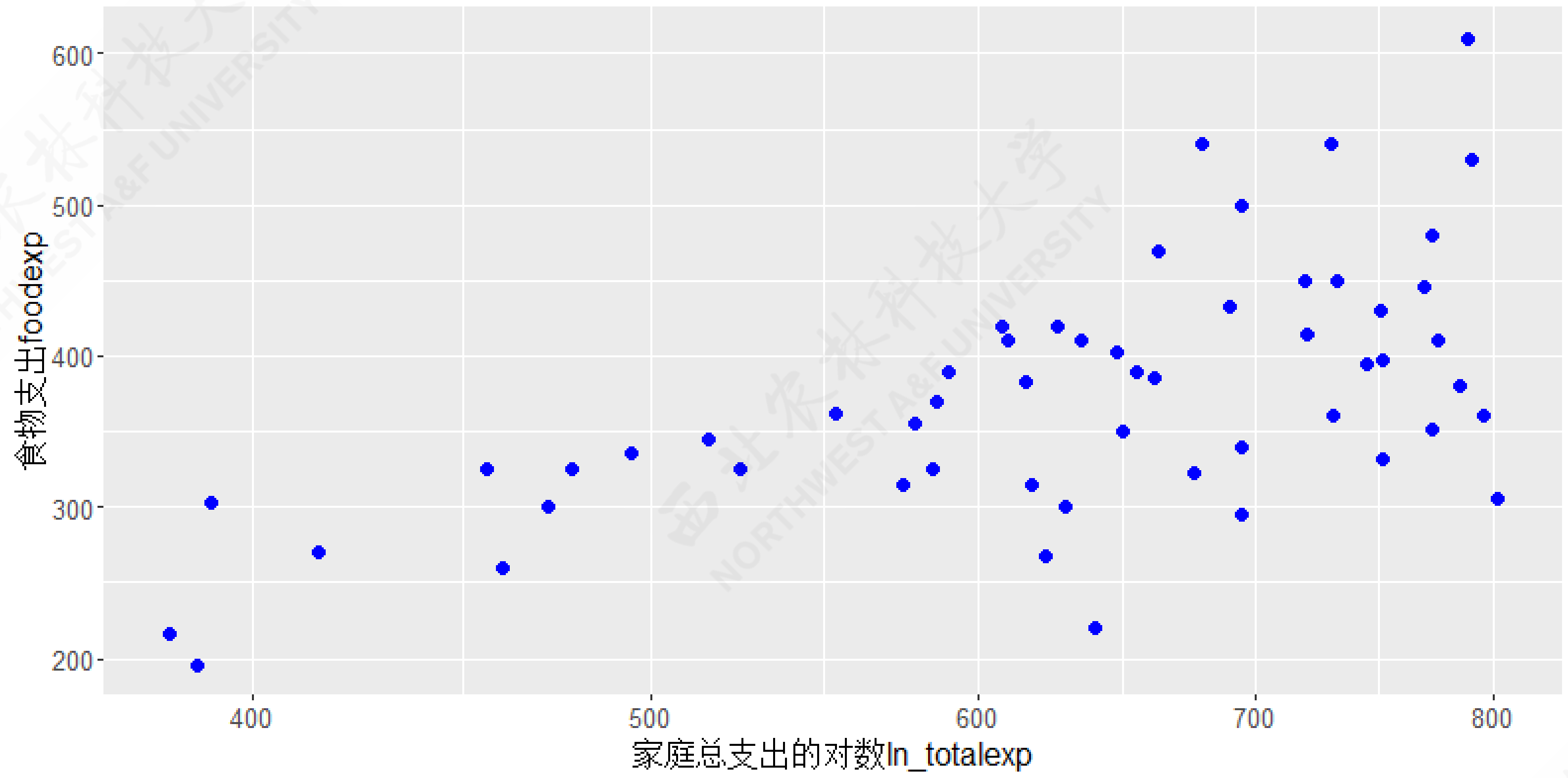
家庭食物支出案例

原始数据作散点图：



家庭食物支出案例

对家庭总支出取对数`ln(totalexp)`，再做散点图：



家庭食物支出案例

构建如下对数到线性模型：

$$Y_t = \beta_1 + \beta_2 \ln X_i + u_i$$

家庭食物支出案例的OLS估计结果如下：

$$\begin{array}{lll}
\widehat{foodexp} = & -1283.91 & +257.27log(totalexp) \\
\text{(t)} & (-4.3848) & (5.6625) \\
\text{(se)} & (292.8105) & (45.4341) \\
\text{(fitness)} & R^2 = 0.3769; \bar{R}^2 = 0.3652 & \\
 & F^* = 32.06;\ p = 0.0000 &
\end{array}$$

对比构建如下经典线性模型及其OLS估计结果：

$$Y_t = \beta_1 + \beta_2 X_i + u_i$$

$$\begin{array}{lll}
\widehat{foodexp} = & +94.21 & +0.44totalexp \\
\text{(t)} & (1.8524) & (5.5770) \\
\text{(se)} & (50.8563) & (0.0783) \\
\text{(fitness)} & R^2 = 0.3698; \bar{R}^2 = 0.3579 & \\
 & F^* = 31.10;\ p = 0.0000 &
\end{array}$$

4.1.6 倒数模型

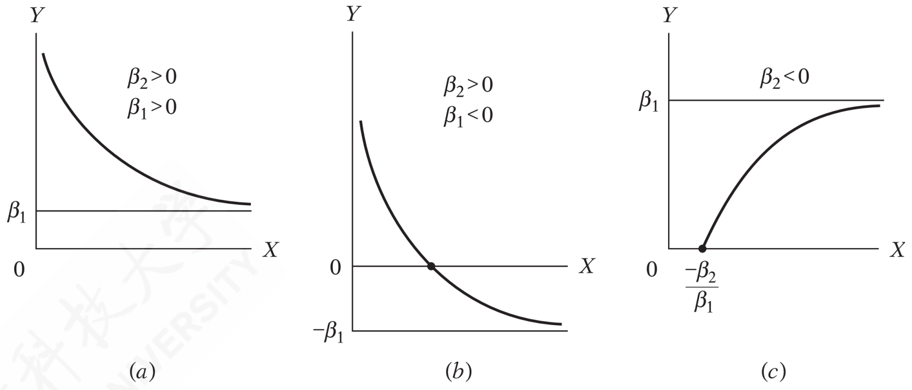
倒数模型

形式：

$$Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i} \right) + u_i$$

特征：总有一条内在的渐近线！

- a.平均固定成本(AFC)曲线
- b.菲利普斯曲线（Phillips curve）
- c.恩格尔曲线（the Engel expenditure curve）



$$X \to \infty; \quad \beta_2 \left(\frac{1}{X} \right) \to 0; \quad Y \to \beta_1$$

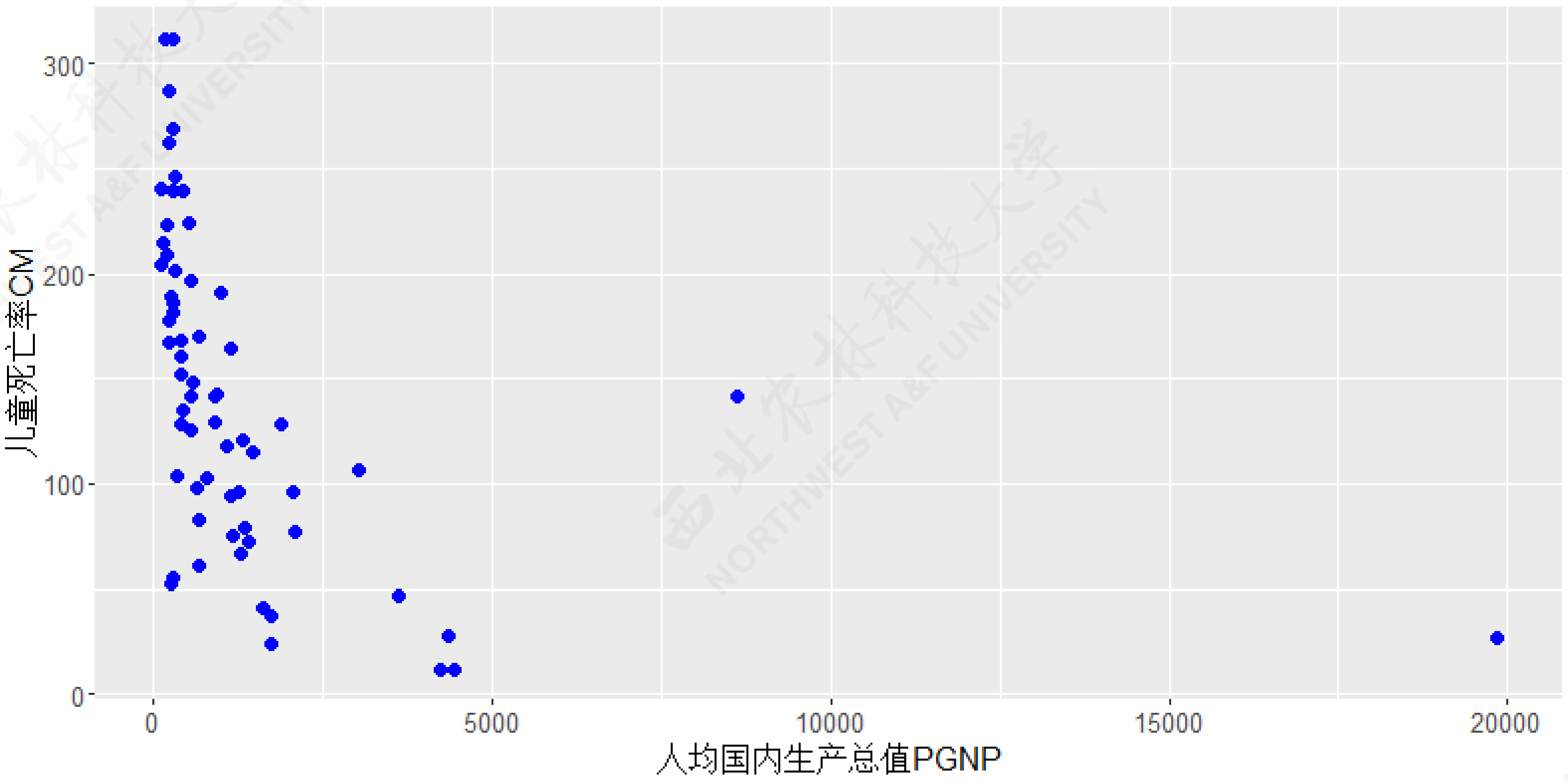
儿童死亡率（CM，千分数）与人均GNP（PGNP，1980年的人均GNP）的关系：

obs	CM	PGNP	rep_PGNP
1	128	1870	0.0005
2	204	130	0.0077
3	202	310	0.0032
4	197	570	0.0018
5	96	2050	0.0005
6	209	200	0.0050
7	170	670	0.0015
8	240	300	0.0033

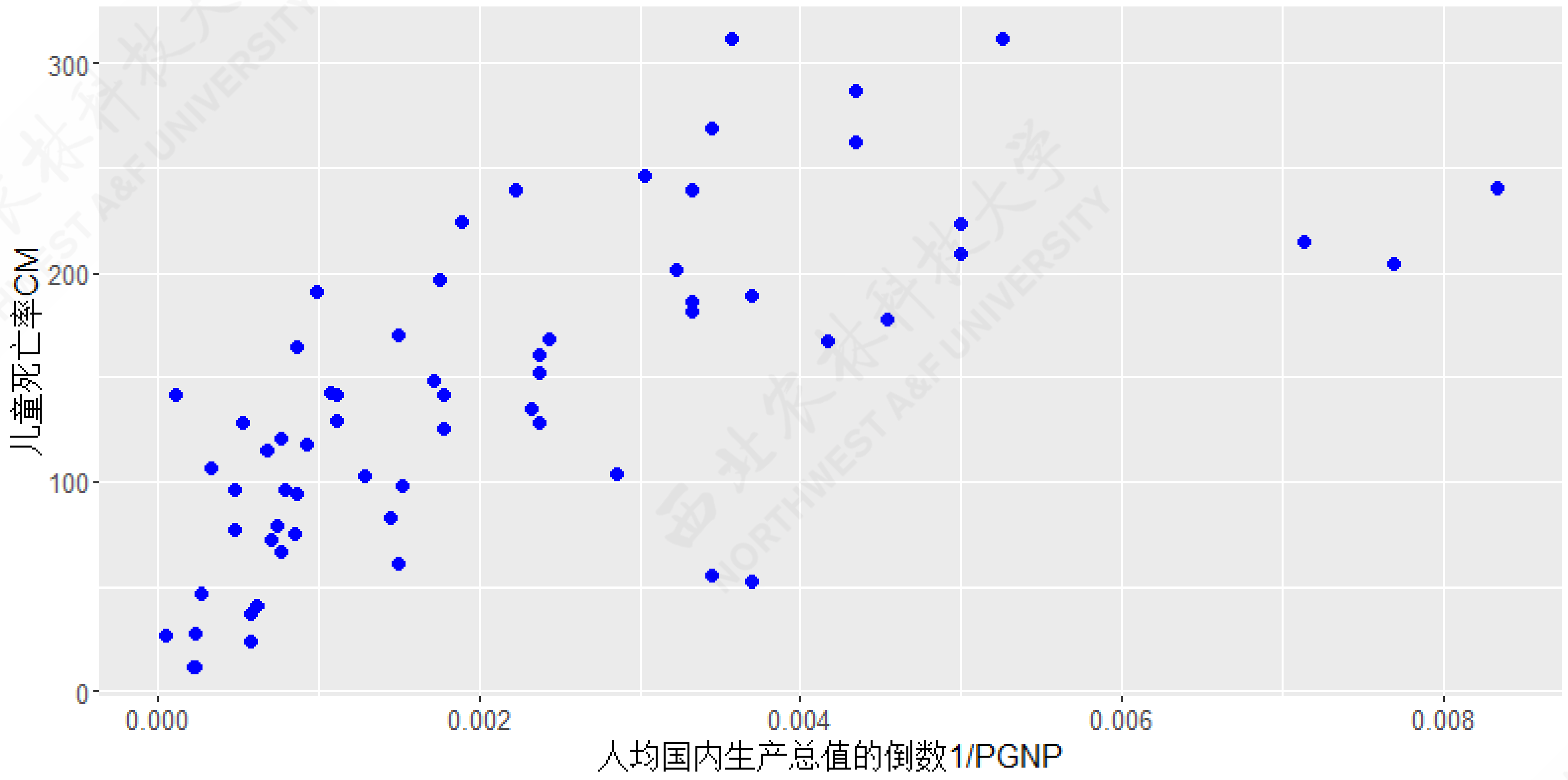
Showing 1 to 8 of 64 entries

Previous 1 2 3 4 5 … 8 Next

原始数据作散点图：



把PGNP取倒数 $1/PGNP$再作散点图：



儿童死亡率案例

构建如下倒数模型：

$$Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i$$

儿童死亡率案例倒数模型的OLS估计结果如下：

$$\begin{array}{llll}
\widehat{CM} = & +81.79 & +27273.17rep_{PGNP} \\
(t) & (7.5511) & (7.2535) \\
(se) & (10.8321) & (3759.9992) \\
(fitness) & R^2 = 0.4591; \bar{R}^2 = 0.4503 & \\
 & F^* = 52.61;\ p = 0.0000 &
\end{array}$$

对比构建如下经典线性模型：

$$Y_t = \beta_1 + \beta_2 X_i + u_i$$

其OLS估计结果如下：

$$\begin{array}{llll}
\widehat{CM} = & +157.42 & -0.01PGNP \\
(t) & (15.9893) & (-3.5157) \\
(se) & (9.8456) & (0.0032) \\
(fitness) & R^2 = 0.1662; \bar{R}^2 = 0.1528 & \\
 & F^* = 12.36;\ p = 0.0008 &
\end{array}$$

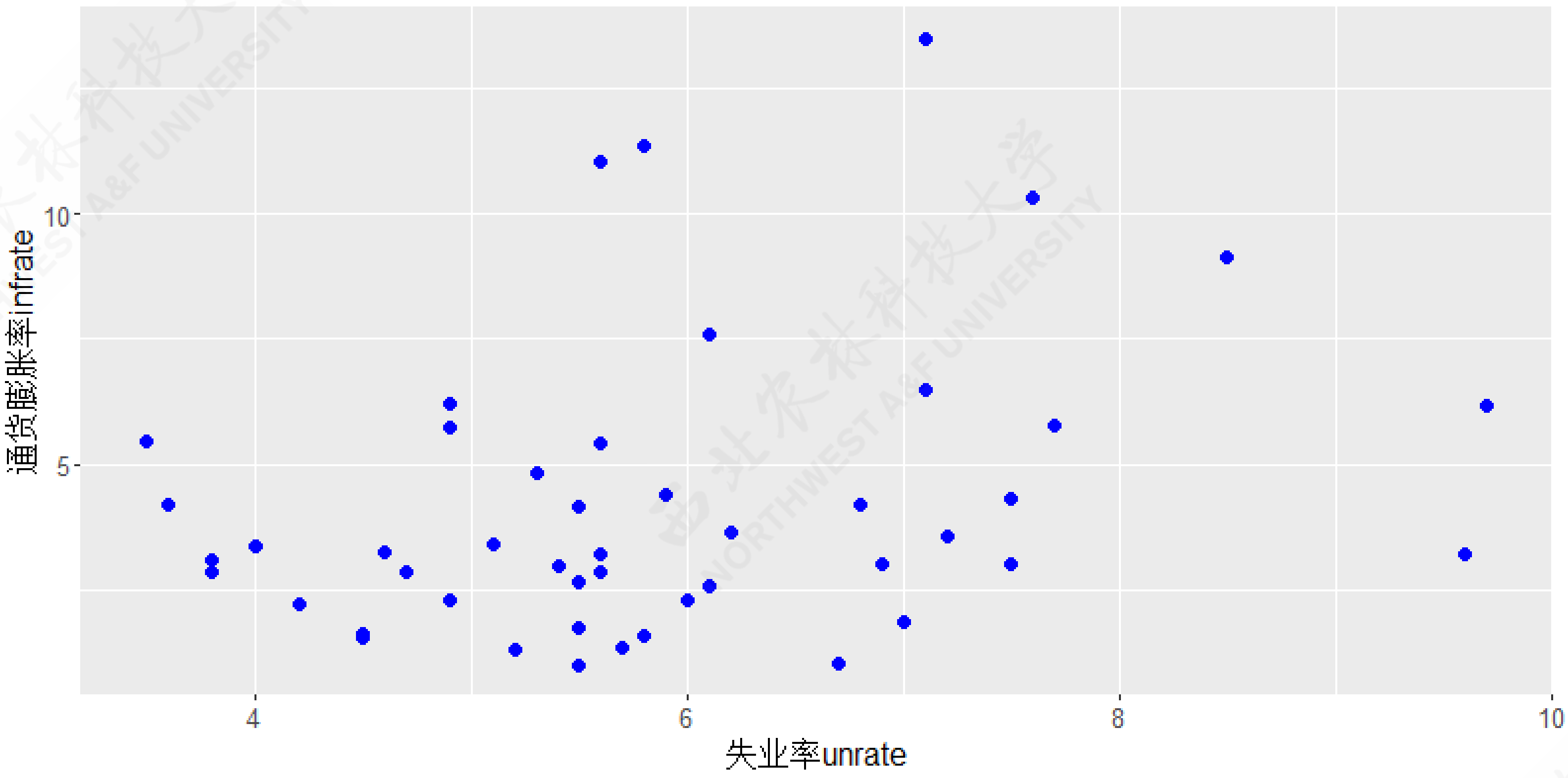
通货膨胀率（`infrate`，%）与失业率（`unrate`，%）的关系：

year	infrate	unrate	rep_unrate
1960	1.7182	5.5	0.1818
1961	1.0135	6.7	0.1493
1962	1.0033	5.5	0.1818
1963	1.3245	5.7	0.1754
1964	1.3072	5.2	0.1923
1965	1.6129	4.5	0.2222
1966	2.8571	3.8	0.2632
1967	3.0864	3.8	0.2632

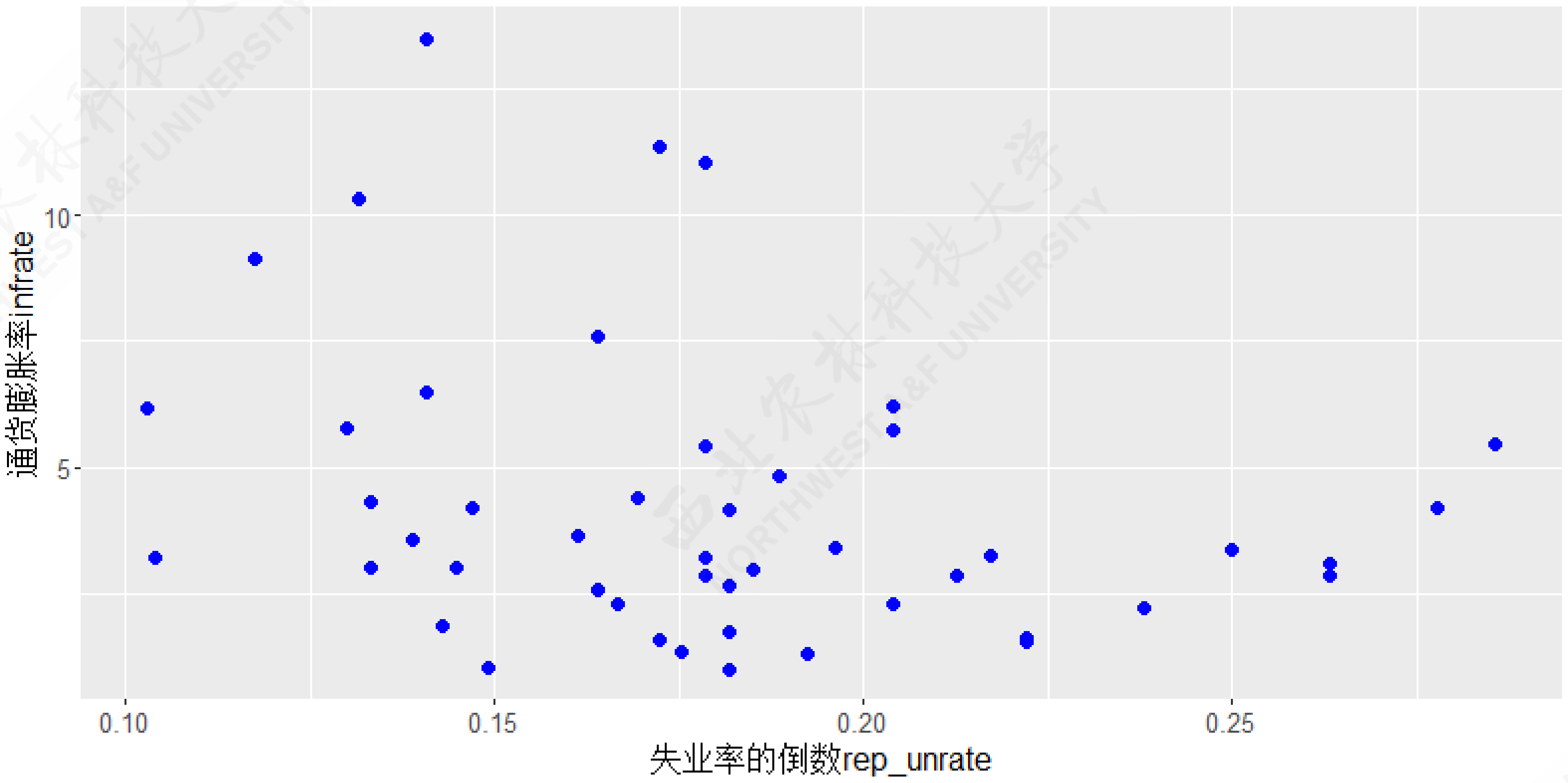
Showing 1 to 8 of 47 entries

Previous 1 2 3 4 5 6 Next

原始数据作散点图：



把失业率`unrate`取倒数 $1/unrate$ 再作散点图：



构建如下倒数模型：

$$Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i$$

菲利普斯曲线案例倒数模型的OLS估计结果如下：

$$\begin{array}{lll}
\widehat{inf rate} = & +7.37 & -17.37 rep_{unrate} \\
(\text{t}) & (4.1723) & (-1.8212) \\
(\text{se}) & (1.7670) & (9.5364) \\
(\text{fitness}) & R^2 = 0.0686; \bar{R}^2 = 0.0479 & \\
 & F^* = 3.32; \quad p = 0.0752 &
\end{array}$$

对比构建如下经典线性模型及其OLS估计结果：

$$Y_t = \beta_1 + \beta_2 X_i + u_i$$

$$\begin{array}{lll}
\widehat{inf rate} = & +0.81 & +0.59 unrate \\
(\text{t}) & (0.4642) & (2.0377) \\
(\text{se}) & (1.7347) & (0.2874) \\
(\text{fitness}) & R^2 = 0.0845; \bar{R}^2 = 0.0641 & \\
 & F^* = 4.15; \quad p = 0.0475 &
\end{array}$$

4.1.7 函数形式的选择

技巧和经验

选择适当模型时，需要一些技巧和经验：

1. 模型背后的理论(如菲利普斯曲线〉可能给出了一个特定的函数形式。
2. 最好能求出回归子相对回归元的变化率〈即斜率〉和回归子对回归元的弹性(见下页ppt)。
3. 所选模型的系数应该满足一定的先验预期。
4. 有时多个模型都能相当不错地拟合一个给定的数据集。
5. 通常不应该过分强调这个指标
6. 在有些情形中，确定一个特定的函数形式不是那么容易，此时我们或许可以使用所谓的博克斯-考克斯变换(Box-Cox transformations)

计算表一览

模型	方程	斜率	点弹性	平均弹性
models	eq	$\frac{dY}{dX}$	$\frac{dY}{dX}\cdot\frac{X_i}{Y_i}$	$\frac{dY}{dX}\cdot\frac{\bar{X}}{\bar{Y}}$
M_1线性模型	$Y_i=\beta_1+\beta_2X_i+u_i$	β_2	β_2X_i/Y_i	$\beta_2\bar{X}/\bar{Y}$
M_2过原点模型	$Y_i=\beta_2X_i+u_i$	β_2	β_2X_i/Y_i	$\beta_2\bar{X}/\bar{Y}$
M_3双对数模型	$ln(Y_i)=\beta_1+\beta_2ln(X_i)+u_i$	β_2Y_i/X_i	β_2	β_2
M_4线性到对数模型	$ln(Y_i)=\beta_1+\beta_2X_i+u_i$	β_2Y_i	β_2X_i	$\beta_2\bar{X}$
M_5对数到线性模型	$Y_i=\beta_1+\beta_2ln(X_i)+u_i$	β_2/X_i	β_2/Y_i	$\beta_2/\bar{Y}$
M_6倒数模型	$Y_i=\beta_1+\beta_2/X_i+u_i$	$-\beta_2/X_i^2$	$-\beta_2/(X_iY_i)$	$-\beta_2/(\bar{X}\bar{Y})$
M_7对数倒数模型	$ln(Y_i)=\beta_1-\beta_2/X_i+u_i$	β_2Y_i/X_i^2	β_2/X_i	$\beta_2/\bar{X}$

本章结束