

计量经济学实例分析

目录

1 多元回归分析	2
1.1 儿童死亡率与人均 GDP	2
1.2 儿童死亡率与妇女识字率	4
1.3 儿童死亡率与人均 GNP 和妇女识字率	5
1.4 标准化三变量回归分析	6
2 多重共线性	7
3 对数回归	9
4 虚拟变量分段线性回归	10
5 多项式拟合	11
5.1 一阶多项式拟合	12
5.2 二阶多项式拟合	12
5.3 三阶多项式拟合	13
5.4 四阶多项式拟合	14
6 时间序列	14
源码	16

1 多元回归分析

我们以儿童死亡率数据进行多元回归分析。我们取前五进行展示。

	CM	FLR	PGNP	TFR
0	128	37	1870	6.66
1	204	22	130	6.15
2	202	16	310	7.00
3	197	65	570	6.25
4	96	76	2050	3.81

图 1 数据集前五

其中，数据解释：

CM：每千人中不足五岁便死亡的人数。

PGNP:人均 GDP。

FLR：妇女识字率。

接下来，进行数据描述，对数据集有一个基本的认识。

	CM	FLR	PGNP	TFR
count	64.000000	64.000000	64.000000	64.000000
mean	141.500000	51.187500	1401.250000	5.549687
std	75.978067	26.007859	2725.695775	1.508993
min	12.000000	9.000000	120.000000	1.690000
25%	82.000000	29.000000	300.000000	4.607500
50%	138.500000	48.000000	620.000000	6.040000
75%	192.500000	77.250000	1317.500000	6.615000
max	312.000000	95.000000	19830.000000	8.490000

图 2 数据描述

1.1 儿童死亡率与人均 GDP

我们首先进行单变量的线性回归。观察儿童死亡率与人均 GDP 的关系。首先将该二维数据画图表示。

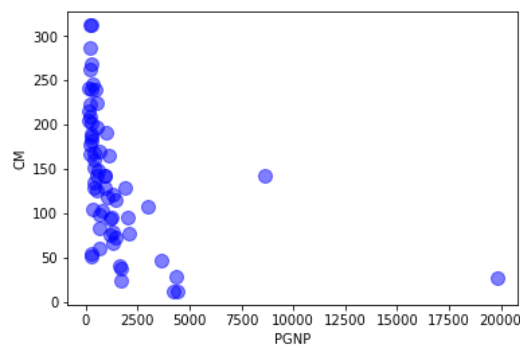


图 3 二维图表示

可以观察到，该数据有两个异常值。不过总体呈现出有规律的趋势。接下来，我们进行最小二乘回归分析。得到结果：

OLS Regression Results						
=====						
Dep. Variable:	CM	R-squared:	0.166			
Model:	OLS	Adj. R-squared:	0.153			
Method:	Least Squares	F-statistic:	12.36			
Date:	Wed, 28 Mar 2018	Prob (F-statistic):	0.000826			
Time:	21:48:05	Log-Likelihood:	-361.64			
No. Observations:	64	AIC:	727.3			
Df Residuals:	62	BIC:	731.6			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	157.4244	9.846	15.989	0.000	137.743	177.105
PGNP	-0.0114	0.003	-3.516	0.001	-0.018	-0.005
=====						
Omnibus:	3.321	Durbin-Watson:	1.931			
Prob(Omnibus):	0.190	Jarque-Bera (JB):	2.545			
Skew:	0.345	Prob(JB):	0.280			
Kurtosis:	2.309	Cond. No.	3.43e+03			
=====						

图 4 最小二乘回归结果

可以看到，回归方程为：
 $CM=157.4244 - 0.0114 \times PGNP$
PGNP 的系数为-0.0114，与预期一致，对 CM 有负影响，人均 GNP 每提高 1 单位，CM 将减少 0.0114 单位。
 $t=-3.516$ ， $p=0.001$ ，均显著，表明 PGNP 对 CM 有影响，应该包含在模型当中。
 R^2 为 0.166，说明儿童死亡率的变异中，有 16.6%可由 PGNP 来解释，这个 R^2 并不高，说明模型拟合效果一般。
雅克-贝拉检验表明：JB 统计量 2.545，在正态性假定下，得到这样一个统计量的概率为 28%。不拒绝残差项是正态分布的假设。
F-statistic:12.36, Prob (F-statistic):0.000826，拒绝人均 GNP 对儿童死亡率没有影响的假设。
接下来，我们将回归方程展现在图上。

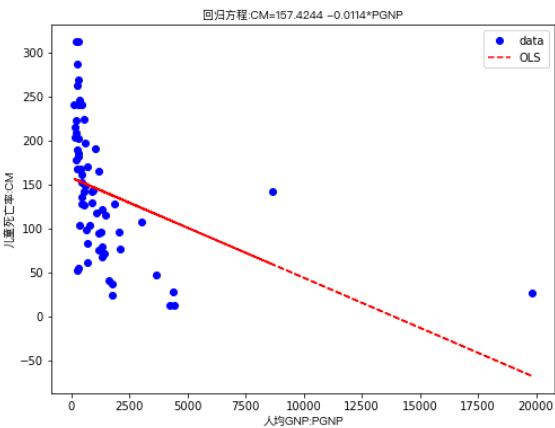


图 5 回归方程

从上图可以看出异常值对模型的影响非常大，因此使用普通最小二乘法建模时，先要清洗一下异常值。接下来进行去掉两个异常值之后的回归。
去掉异常点后，图形如下：

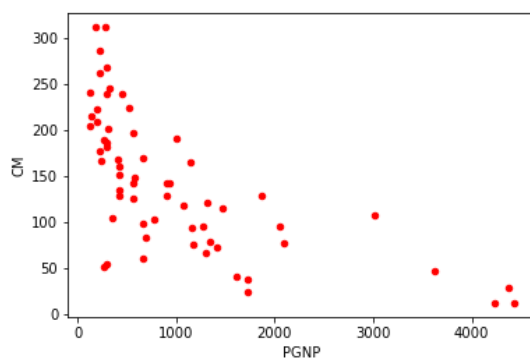


图6 去掉异常点后的数据点

同样，进行回归分析。

OLS Regression Results						
=====						
Dep. Variable:	CM	R-squared:	0.437			
Model:	OLS	Adj. R-squared:	0.428			
Method:	Least Squares	F-statistic:	46.62			
Date:	Wed, 28 Mar 2018	Prob (F-statistic):	4.91e-09			
Time:	21:48:06	Log-Likelihood:	-337.98			
No. Observations:	62	AIC:	680.0			
Df Residuals:	60	BIC:	684.2			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	191.1177	10.098	18.927	0.000	170.919	211.316
PGNP	-0.0484	0.007	-6.828	0.000	-0.063	-0.034

Omnibus:	0.136	Durbin-Watson:	1.905			
Prob(Omnibus):	0.934	Jarque-Bera (JB):	0.184			
Skew:	0.103	Prob(JB):	0.912			
Kurtosis:	2.832	Cond. No.	1.98e+03			
=====						

图7 去掉异常点后的回归分析

常数项系数：157.4244 增加到 191.1177。PGNP:-0.0114 增加到 -0.0484。
r 方由 0.166 增加到 0.437。说明普通最小二乘法对异常值异常敏感。

1.2 儿童死亡率与妇女识字率

首先作出其二维图观察其关系。

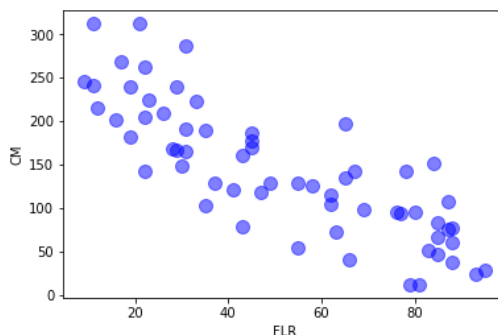


图8 数据关系

可以发现两个变量之间关系显著，接下来进行回归分析。

OLS Regression Results						
=====						
Dep. Variable:	CM	R-squared:	0.679			
Model:	OLS	Adj. R-squared:	0.673			
Method:	Least Squares	F-statistic:	126.8			
Date:	Wed, 28 Mar 2018	Prob (F-statistic):	1.99e-16			
Time:	21:48:06	Log-Likelihood:	-320.60			
No. Observations:	62	AIC:	645.2			
Df Residuals:	60	BIC:	649.5			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	264.0537	12.051	21.911	0.000	239.948	288.159
FLR	-2.3662	0.210	-11.258	0.000	-2.787	-1.946
=====						
Omnibus:	0.683	Durbin-Watson:	2.259			
Prob(Omnibus):	0.711	Jarque-Bera (JB):	0.610			
Skew:	0.234	Prob(JB):	0.737			
Kurtosis:	2.866	Cond. No.	126.			
=====						

图 9 回归分析

由回归分析结果也可以看出，儿童死亡率与妇女识字率相关性很强。

1.3 儿童死亡率与人均 GNP 和妇女识字率

接下来进行三变量回归，其回归结果如下：

OLS Regression Results						
=====						
Dep. Variable:	CM	R-squared:	0.725			
Model:	OLS	Adj. R-squared:	0.716			
Method:	Least Squares	F-statistic:	77.88			
Date:	Wed, 28 Mar 2018	Prob (F-statistic):	2.80e-17			
Time:	21:48:06	Log-Likelihood:	-315.75			
No. Observations:	62	AIC:	637.5			
Df Residuals:	59	BIC:	643.9			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	260.1982	11.304	23.019	0.000	237.580	282.816
PGNP	-0.0196	0.006	-3.162	0.002	-0.032	-0.007
FLR	-1.9116	0.243	-7.865	0.000	-2.398	-1.425
=====						
Omnibus:	1.250	Durbin-Watson:	2.209			
Prob(Omnibus):	0.535	Jarque-Bera (JB):	0.792			
Skew:	0.269	Prob(JB):	0.673			
Kurtosis:	3.129	Cond. No.	3.14e+03			

图 10 回归分析

回归方程： $CM=260-0.0196PGNP-1.9116FLR$

PGNP 的偏回归系数为-0.0196，与预期一致，对 CM 有负影响，在保持其它因素不变时，PGNP 每提高 1 单位，CM 将减少 0.0196 个单位。

与-0.0484 相比，PGNP 对 CM 的影响减小了，说明我们的双变量回归模型有设定误差。

$t=-3.162$ ， $p=0.002$ ，均显著，表明 PGNP 对 CM 有影响，应该包含在模型当中。如果虚拟假设正确，得到一个小于 3.162 的 t 值得概率为 0.002。

FLR 的偏回归系数为-1.916，与预期一致，对 CM 有负影响，妇女识字率越高，儿童死亡率越低。保持其它因素不变，FLR 每提高 1 单位，CM 将减少 1.916 单位。

$t=-7.865$ ， $p=0.000$ ，均显著，表明 PGNP 对 CM 有影响，应该包含在模型当中。调整 R^2 为 0.716，说明儿童死亡率的变异中，有 71.6%可由模型来解释，这个

R2 很高，说明模型拟合效果很好。
雅克-贝拉检验表明：JB 统计量 0.673，在正态性假定下，得到这样一个统计量的概率为 79.2%。不拒绝残差项是正态分布的假设。
显著性检验。F-statistic:73.83, 在零假设下，得到这样一个值得概率, Prob (F-statistic):5.12e-17, 认为妇女识字率和人均 GNP 对儿童死亡率有显著影响。
残差项的分布图：

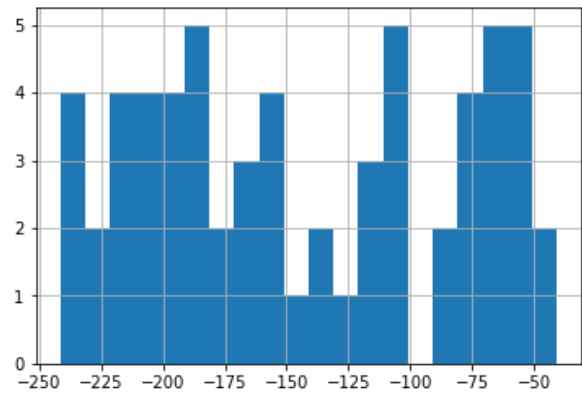


图 10 残差项

1.4 标准化三变量回归分析

刚刚的回归分析我们并没有对数据集进行标准化，这可能导致了一些误差。所以接下里，我们对数据集的每一列进行标准化。之后在进行回归分析

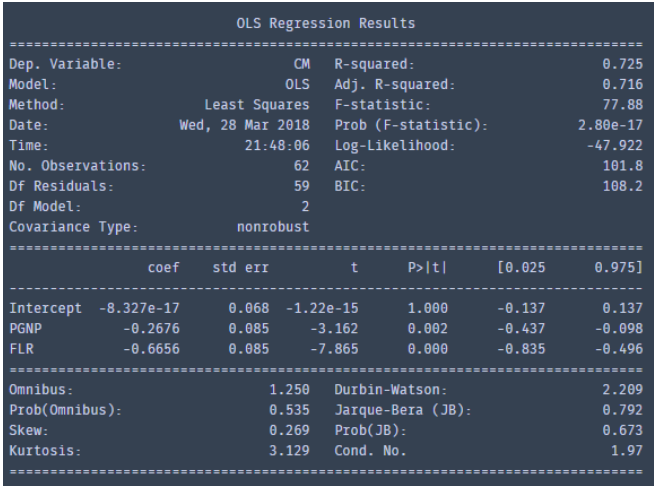


图 11 回归分析

保持其它变量不变，PGNP 每提高一个标准差，导致 CM 平均下降 0.2676 个标准差。
保持其它变量不变，FLR 提高一个标准差，导致 CM 平均下降 0.6656 个标准差相对而言，妇女识字率比人均 GNP 对儿童死亡率的影响更大。
标准化回归，可以看出哪个变量的影响更大，PGNP 和 FLR 同时提高 1 单位, 每 1000 名儿童中不足 5 岁便死亡的儿童数约降低 2.24 人。

2 多重共线性

在本节，我们探讨多重共线性的例子。数据集展示如下：

	Y	X2	X3
0	70	80	810
1	65	100	1009
2	90	120	1273
3	95	140	1425
4	110	160	1633

图 12 数据集展示

其中：

Y 消费支出

X2 收入

X3 财富

接着，我们进行数据集的简单统计分析：

	Y	X2	X3
count	10.000000	10.000000	10.000000
mean	111.000000	170.000000	1740.000000
std	31.428932	60.553007	617.731153
min	65.000000	80.000000	810.000000
25%	91.250000	125.000000	1311.000000
50%	112.500000	170.000000	1754.500000
75%	135.000000	215.000000	2163.750000
max	155.000000	260.000000	2686.000000

图 13 数据集分析

首先我们直接进行回归分析：

OLS Regression Results						
=====						
Dep. Variable:	Y	R-squared:	0.964			
Model:	OLS	Adj. R-squared:	0.953			
Method:	Least Squares	F-statistic:	92.40			
Date:	Tue, 03 Apr 2018	Prob (F-statistic):	9.29e-06			
Time:	16:22:53	Log-Likelihood:	-31.587			
No. Observations:	10	AIC:	69.17			
Df Residuals:	7	BIC:	70.08			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	24.7747	6.752	3.669	0.008	8.808	40.742
X2	0.9415	0.823	1.144	0.290	-1.004	2.887
X3	-0.0424	0.081	-0.526	0.615	-0.233	0.148
=====						
Omnibus:	0.869	Durbin-Watson:	2.891			
Prob(Omnibus):	0.648	Jarque-Bera (JB):	0.716			
Skew:	-0.511	Prob(JB):	0.699			
Kurtosis:	2.179	Cond. No.	5.79e+03			
=====						

图 14 回归分析

$$\hat{Y}_i = 24.7747 + 0.9415X_{2i} - 0.0424X_{3i}$$

财富回归系数为负值，越有钱消费越低，与预期不一致。

se= (6.752) (0.823) (0.081)
t= (3.669) (1.144) (-0.526)
p= (0.008) (0.290) (0.615) 回归系数均不显著 ，说明变量无影响。调整 $R^2=0.953$ 调整 R^2 非常高，说明模型拟合非常好。
所有这些看起来怪异的现象，是由于收入与财富的高度相关。接下来分别对 X2 和 X3 进行单变量回归分析。并且对 X2 与 X3 之间进行回归分析。

OLS Regression Results						
=====						
Dep. Variable:	Y	R-squared:	0.962			
Model:	OLS	Adj. R-squared:	0.957			
Method:	Least Squares	F-statistic:	202.9			
Date:	Tue, 03 Apr 2018	Prob (F-statistic):	5.75e-07			
Time:	19:21:50	Log-Likelihood:	-31.781			
No. Observations:	10	AIC:	67.56			
Df Residuals:	8	BIC:	68.17			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	24.4545	6.414	3.813	0.005	9.664	39.245
X2	0.5091	0.036	14.243	0.000	0.427	0.592
=====						
Omnibus:	1.060	Durbin-Watson:	2.680			
Prob(Omnibus):	0.589	Jarque-Bera (JB):	0.777			
Skew:	-0.398	Prob(JB):	0.678			
Kurtosis:	1.891	Cond. No.	561.			
=====						

图 15 Y 与 X2 回归分析

OLS Regression Results						
=====						
Dep. Variable:	Y	R-squared:	0.957			
Model:	OLS	Adj. R-squared:	0.951			
Method:	Least Squares	F-statistic:	176.7			
Date:	Tue, 03 Apr 2018	Prob (F-statistic):	9.80e-07			
Time:	19:22:04	Log-Likelihood:	-32.444			
No. Observations:	10	AIC:	68.89			
Df Residuals:	8	BIC:	69.49			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	24.4110	6.874	3.551	0.007	8.559	40.263
X3	0.0498	0.004	13.292	0.000	0.041	0.058
=====						
Omnibus:	1.504	Durbin-Watson:	2.417			
Prob(Omnibus):	0.471	Jarque-Bera (JB):	0.767			
Skew:	-0.176	Prob(JB):	0.681			
Kurtosis:	1.689	Cond. No.	5.75e-03			
=====						

图 16 Y 与 X3 回归分析

OLS Regression Results						
=====						
Dep. Variable:	X2	R-squared:	0.998			
Model:	OLS	Adj. R-squared:	0.998			
Method:	Least Squares	F-statistic:	3849.			
Date:	Tue, 03 Apr 2018	Prob (F-statistic):	5.06e-12			
Time:	19:22:25	Log-Likelihood:	-23.807			
No. Observations:	10	AIC:	51.61			
Df Residuals:	8	BIC:	52.22			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	-0.3863	2.898	-0.133	0.897	-7.069	6.296
X3	0.0979	0.002	62.040	0.000	0.094	0.102
=====						
Omnibus:	0.003	Durbin-Watson:	2.069			
Prob(Omnibus):	0.999	Jarque-Bera (JB):	0.201			
Skew:	-0.028	Prob(JB):	0.904			
Kurtosis:	2.307	Cond. No.	5.75e-03			
=====						

图 17 X2 与 X3 回归分析

R^2 非常高，表明两者高度相关

3 对数回归

接下来，我们对美国消费函数进行对数回归。数据展示如下：

	Year	C	Yd	W	I
0	1947	976.4	1035.2	5166.815	-10.350940
1	1948	998.1	1090.0	5280.757	-4.719804
2	1949	1025.3	1095.6	5607.351	1.044063
3	1950	1090.9	1192.7	5759.515	0.407346
4	1951	1107.1	1227.0	6086.056	-5.283152

图 18 数据集展示

数据描述：

C 真实消费支出

Yd 真实个人可支配收入

W 真实财富

真实利率 I

建立模型

$$\ln C_i = \beta_1 + \beta_2 \ln Yd_i + \beta_3 \ln W_i + u_i$$

根据模型进行回归分析。

OLS Regression Results						
=====						
Dep. Variable:	np.log(C)	R-squared:	1.000			
Model:	OLS	Adj. R-squared:	1.000			
Method:	Least Squares	F-statistic:	3.783e-04			
Date:	Mon, 02 Apr 2018	Prob (F-statistic):	7.12e-84			
Time:	20:01:35	Log-Likelihood:	164.59			
No. Observations:	54	AIC:	-321.2			
Df Residuals:	50	BIC:	-313.2			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.4677	0.043	-10.933	0.000	-0.554	-0.382
np.log(Yd)	0.8049	0.017	45.998	0.000	0.770	0.840
np.log(W)	0.2013	0.018	11.441	0.000	0.166	0.237
I	-0.0027	0.001	-3.529	0.001	-0.004	-0.001
=====						
Omnibus:	1.995	Durbin-Watson:	1.289			
Prob(Omnibus):	0.369	Jarque-Bera (JB):	1.006			
Skew:	0.438	Prob(JB):	0.405			
Kurtosis:	2.811	Cond. No.	357.			
=====						

图 19 回归分析

区间估计等于回归值加减两个标准误。所有的统计系数都是高度显著的，因为他们的 P 值都极小。

收入弹性约等于 0.80，即在其他变量保持不变的情况下，如果收入增加 1%，则消费支出平均约增加 0.8%。

财富弹性约等于 0.20，在保持其它变量不变的情况下，如果财富增加 1%，平均消费将增加约 0.2%。

如果利率上调 1 个百分点(注意它的 1 单位是 1 个百分点)，保持其它条件不变，消费支出将下降约 0.26%（半弹性）

所有回归元的符号都与先验预期一致，即收入与财富对消费有正影响，而利率对消费有负影响。

在该例子中不需要担心多重共线性的影响。因为所有的系数都具有正确的符号，每个系数又是个别统计显著的，F 值也统计显著，表明所有变量共同对消费有影响，调整 $R^2=1$ 非常高，表明模型拟合非常好。

4 虚拟变量分段线性回归

在本节，我们所要处理的数据是：

	Y	X
0	256	1000
1	414	2000
2	634	3000
3	778	4000
4	1003	5000

图 20 数据集

Y 表示总成本，X 表示总产出，研究 X 的变化对 Y 的影响。我们被告知，总成本在产出为 5500 个单位时，可能会改变斜率。由于只有两个变量，我们进行数据图像展示。

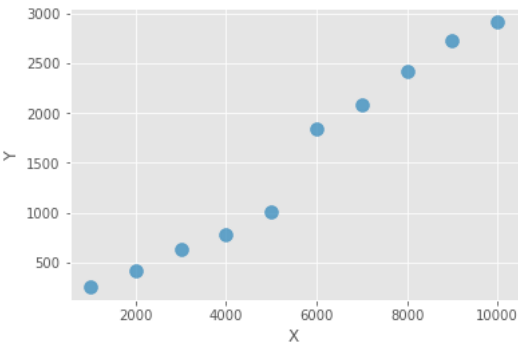


图 21 数据集展示

接下来，我们建立一系列虚拟变量来解决分段函数的问题。新的数据集为：

	Y	X	XK	D	X2	D2
0	256	1000	5500	0	-4500	0
1	414	2000	5500	0	-3500	0
2	634	3000	5500	0	-2500	0
3	778	4000	5500	0	-1500	0
4	1003	5000	5500	0	-500	0
5	1839	6000	5500	1	500	500
6	2081	7000	5500	1	1500	1500
7	2423	8000	5500	1	2500	2500
8	2734	9000	5500	1	3500	3500
9	2914	10000	5500	1	4500	4500

图 22 数据集展示

之后，我们进行回归分析：

OLS Regression Results						
=====						
Dep. Variable:	Y	R-squared:	0.974			
Model:	OLS	Adj. R-squared:	0.966			
Method:	Least Squares	F-statistic:	129.6			
Date:	Tue, 26 Sep 2017	Prob (F-statistic):	2.95e-06			
Time:	17:58:21	Log-Likelihood:	-64.588			
No. Observations:	10	AIC:	135.2			
Df Residuals:	7	BIC:	136.1			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	-145.7167	176.734	-0.824	0.437	-563.627	272.193
X	0.2791	0.046	6.067	0.001	0.170	0.388
D2	0.0945	0.083	1.145	0.290	-0.101	0.290
=====						
Omnibus:	0.417	Durbin-Watson:	1.504			
Prob(Omnibus):	0.812	Jarque-Bera (JB):	0.470			
Skew:	-0.082	Prob(JB):	0.791			
Kurtosis:	1.951	Cond. No.	1.96e-04			
=====						

图 22 回归分析

得到回归方程为：

$$\hat{Y}_i = -145.72 + 0.2791X_i + 0.0945(X_i - X_i^*)D_i$$
$$t = (-0.824) \quad (6.067) \quad (1.145)$$
$$R^2 = 0.9737 \quad X^* = 5500$$

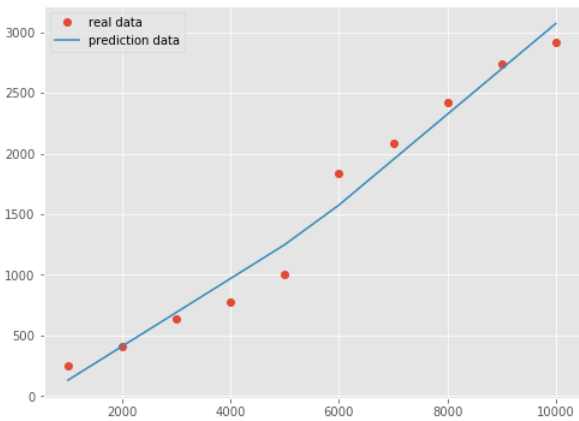


图 23 回归方程

5 多项式拟合

在本节，我们对高通股价收盘价进行多项式拟合。其数据集如下：

	Date	time	Close
0	1995-01-03	1	23.47
1	1995-01-09	2	20.54
2	1995-01-16	3	22.74
3	1995-01-23	4	27.88
4	1995-01-30	5	27.39

图 24 数据集

图像如下：

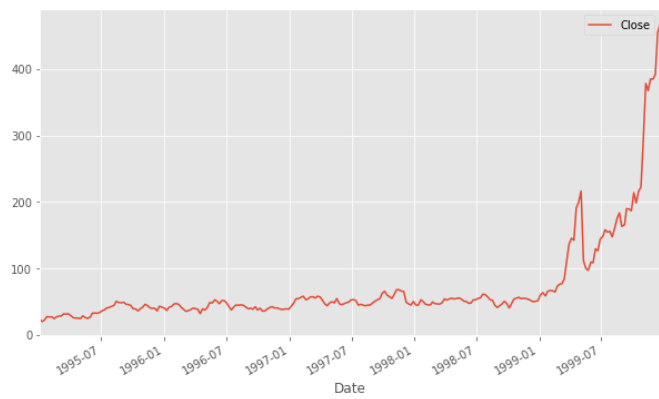


图 25 数据集图形

5.1 一阶多项式拟合

首先利用一阶多项式拟合得到：

```
=====
                        OLS Regression Results
=====
Dep. Variable:          Close      R-squared:                0.385
Model:                  OLS        Adj. R-squared:           0.382
Method:                 Least Squares   F-statistic:             161.3
Date:                   Thu, 11 Jan 2018   Prob (F-statistic):       4.94e-29
Time:                   10:30:01     Log-Likelihood:          -1411.3
No. Observations:       260          AIC:                     2827.
Df Residuals:           258          BIC:                     2834.
Df Model:                1
Covariance Type:        nonrobust
=====
               coef    std err          t      P>|t|    [0.025    0.975]
-----
Intercept    -4.6941     6.881     -0.682    0.496    -18.244     8.856
time          0.5805     0.046    12.701    0.000     0.491     0.671
=====
Omnibus:                 191.388   Durbin-Watson:           0.048
Prob(Omnibus):            0.000   Jarque-Bera (JB):        2042.237
Skew:                     2.939   Prob(JB):                 0.00
Kurtosis:                 15.408   Cond. No.                 302.
=====
```

图 26 回归分析

图像如下

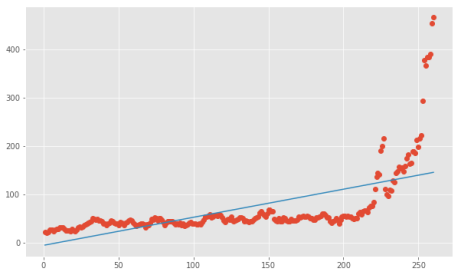


图 27 回归方程

5.2 二阶多项式拟合

接下来，用二阶多项式拟合得到：

OLS Regression Results						
=====						
Dep. Variable:	Close	R-squared:	0.622			
Model:	OLS	Adj. R-squared:	0.619			
Method:	Least Squares	F-statistic:	211.3			
Date:	Thu, 11 Jan 2018	Prob (F-statistic):	5.45e-55			
Time:	11:00:37	Log-Likelihood:	-1348.0			
No. Observations:	260	AIC:	2702.			
Df Residuals:	257	BIC:	2713.			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	72.6825	8.147	8.921	0.000	56.639	88.726
time	-1.1915	0.144	-8.266	0.000	-1.475	-0.908
np.power(time, 2)	0.0068	0.001	12.694	0.000	0.006	0.008
=====						
Omnibus:	148.250	Durbin-Watson:	0.077			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1176.728			
Skew:	2.177	Prob(JB):	3.00e-256			
Kurtosis:	12.470	Cond. No.	9.18e-04			
=====						

图 28 回归分析

图像如下

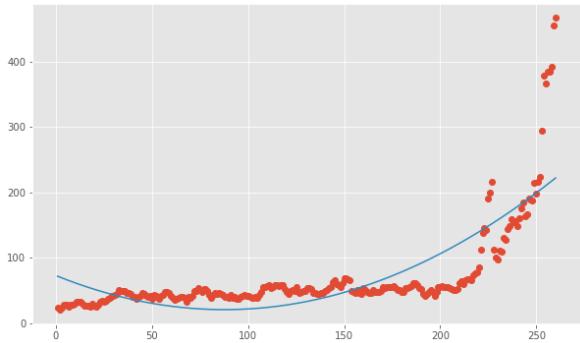


图 29 回归方程

5.3 三阶多项式拟合

之后，用三阶多项式拟合得到：

OLS Regression Results						
=====						
Dep. Variable:	Close	R-squared:	0.815			
Model:	OLS	Adj. R-squared:	0.813			
Method:	Least Squares	F-statistic:	375.2			
Date:	Thu, 11 Jan 2018	Prob (F-statistic):	2.22e-93			
Time:	11:03:13	Log-Likelihood:	-1255.3			
No. Observations:	260	AIC:	2519.			
Df Residuals:	256	BIC:	2533.			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	-10.8543	7.670	-1.415	0.158	-25.959	4.250
time	2.6128	0.254	10.286	0.000	2.113	3.113
np.power(time, 2)	-0.0296	0.002	-13.094	0.000	-0.034	-0.025
np.power(time, 3)	9.29e-05	5.69e-06	16.326	0.000	8.17e-05	0.000
=====						
Omnibus:	117.694	Durbin-Watson:	0.150			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	652.996			
Skew:	1.754	Prob(JB):	1.60e-142			
Kurtosis:	9.926	Cond. No.	2.72e-07			
=====						

图 30 回归分析

图像如下

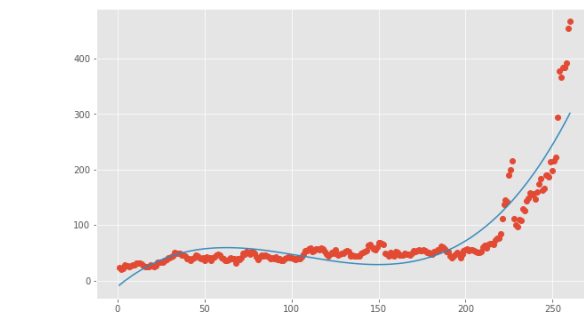


图 31 回归方程

5.4 四阶多项式拟合

最后，用四阶多项式拟合得到：

OLS Regression Results						
=====						
Dep. Variable:	Close	R-squared:	0.891			
Model:	OLS	Adj. R-squared:	0.890			
Method:	Least Squares	F-statistic:	523.1			
Date:	Thu, 11 Jan 2018	Prob (F-statistic):	1.39e-121			
Time:	11:04:20	Log-Likelihood:	-1185.9			
No. Observations:	260	AIC:	2382.			
Df Residuals:	255	BIC:	2400.			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	49.7829	7.420	6.709	0.000	35.171	64.395
time	-1.9548	0.392	-4.983	0.000	-2.727	-1.182
np.power(time, 2)	0.0488	0.006	8.009	0.000	0.037	0.061
np.power(time, 3)	-0.0004	3.51e-05	-10.661	0.000	-0.000	-0.000
np.power(time, 4)	8.94e-07	6.66e-08	13.415	0.000	7.63e-07	1.03e-06
=====						
Omnibus:	80.297	Durbin-Watson:	0.246			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	383.605			
Skew:	1.160	Prob(JB):	5.03e-84			
Kurtosis:	8.479	Cond. No.	7.87e-09			
=====						

图 32 回归分析

图像如下

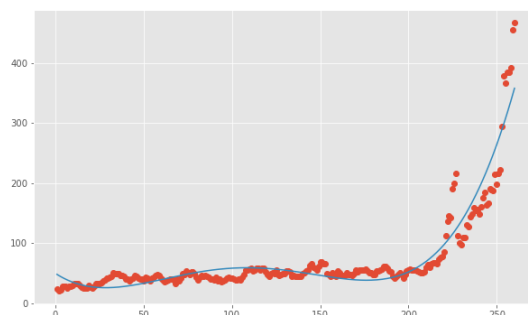


图 33 回归方程

6 时间序列

在本节，我们利用时间序列分析方法对美国 GDP 进行分析。
美国每季度的 GDP 数据如下：

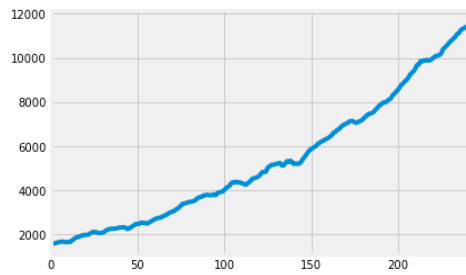


图 34 数据集图形

取对数之后：

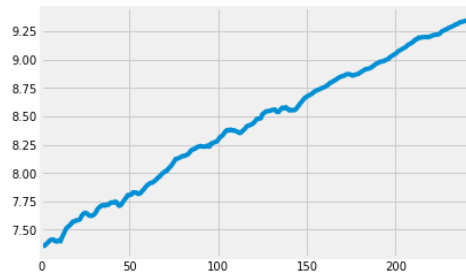


图 35 数据集图形

其自相关，偏自相关图如下：

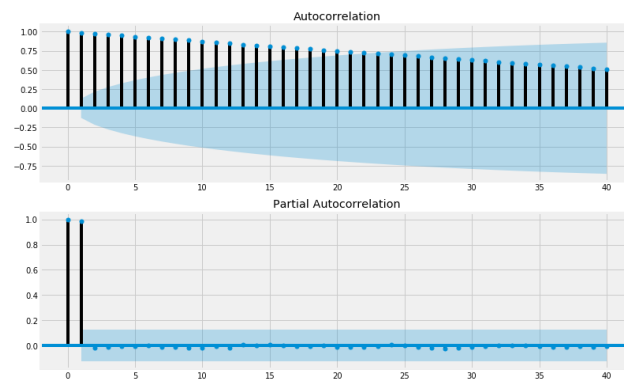


图 36 自相关检验

发现不符合要求，于是我们对源数据进行一阶差分后进行检验：

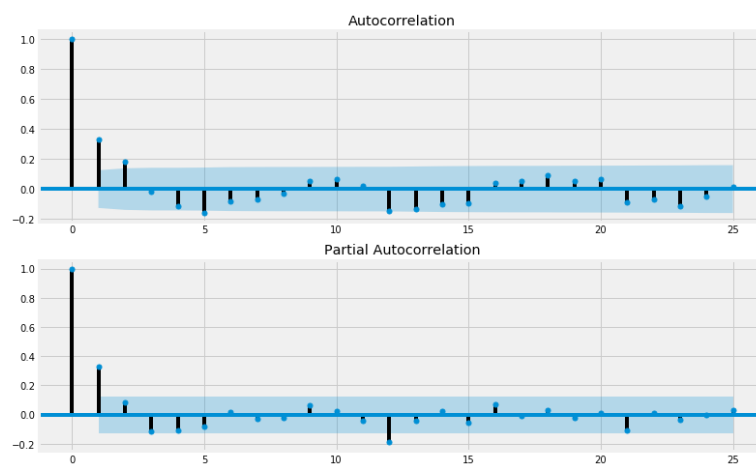


图 37 自相关检验

之后，我们即使用 SARIMA 模型进行拟合。其统计结果如下：

Statespace Model Results						
=====						
Dep. Variable:	y	No. Observations:	243			
Model:	SARIMAX(0, 1, 2)	Log Likelihood	785.034			
Date:	Thu, 11 Jan 2018	AIC	-1562.069			
Time:	13:35:27	BIC	-1548.097			
Sample:	0	HQIC	-1556.441			
	- 243					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

intercept	-1.475e-05	1.54e-05	-0.959	0.338	-4.49e-05	1.54e-05
ma.L1	-0.7094	0.470	-1.508	0.131	-1.631	0.212
ma.L2	-0.2917	0.133	-2.200	0.028	-0.552	-0.032
sigma2	8.823e-05	4.27e-05	2.068	0.039	4.61e-06	0.000
=====						
Ljung-Box (Q):	58.59	Jarque-Bera (JB):	36.74			
Prob(Q):	0.03	Prob(JB):	0.00			
Heteroskedasticity (H):	0.20	Skew:	0.04			
Prob(H) (two-sided):	0.00	Kurtosis:	4.91			
=====						

图 38 结果

检验其残差:

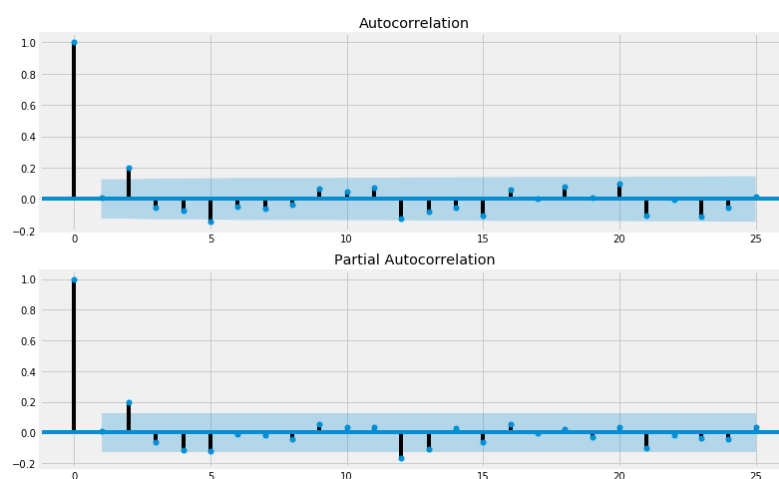


图 38 残差

源码

本文的代码可以在我的 Github 上获取:

<https://github.com/stxupengyu/Econometrics-Example>