设立双对数模型消除量纲的影响

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 06/24/20 Time: 19:55
Sample: 2001 2017
Included observations: 17

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-13.13508	3.463622	-3.792297	0.0026
LOG(X1)	2.778341	0.664735	4.179623	0.0013
LOG(X2)	-0.760149	0.286397	-2.654181	0.0210
LOG(X3)	-0.293647	0.745133	-0.394087	0.7004
LOG(X4)	0.197749	0.119079	1.660655	0.1227

R-squared	0.939835	Mean dependent var	7.851729
Adjusted R-squared	0.919780	S.D. dependent var	0.614905
S.E. of regression	0.174161	Akaike info criterion	-0.417750
Sum squared resid	0.363983	Schwarz criterion	-0.172687
Log likelihood	8.550872	Hannan-Quinn criter.	-0.393390
F-statistic	46.86275	Durbin-Watson stat	0.723474
Prob(F-statistic)	0.000000		

R^2=0.94，修正的 R^2=0.91，模型整体拟合效果较好，模型对网络消费的解释力度达到了 94%，F 统计量等于 46.86，在 5%的显著性水平下它是显著的，说明模型整体上是显著的。

LogX3 和 logX4 的 t 检验在统计意义上不显著，logx3 的符号和我们的预期是相反的，这显然与我们的实践经验是不相符合的，这是严重多重共线性导致的典型现象，我们用相关系数法和方差扩大因子法检验多重共线性。

相关系数法检验多重共线性

	LOG(X1)	LOG(X2)	LOG(X3)	LOG(X4)
LOG(X1)	1.000000	0.946221	0.930585	0.832379
LOG(X2)	0.946221	1.000000	0.947904	0.939013
LOG(X3)	0.930585	0.947904	1.000000	0.945609
LOG(X4)	0.832379	0.939013	0.945609	1.000000

相关系数矩阵每两个解释变量之间的相关系数都很大，均大于 0.8，解释变量之间可能存在严重的多重共线性。

方差扩大因子法

Variance Inflation Factors
Date: 06/24/20 Time: 20:04
Sample: 2001 2017
Included observations: 17

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	11.99668	6723.732	NA
LOG(X1)	0.441873	24782.33	**33.04084**
LOG(X2)	0.082023	4794.374	**44.14903**
LOG(X3)	0.555224	2607.482	**34.18106**
LOG(X4)	0.014180	681.9208	**35.15306**

每一个解释变量的方差扩大因子都远大于 10，解释变量之间存在比较严重的多重共线性。

处理多重共线性的问题——逐步回归

被解释变量对解释变量逐个做一元回归

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.627133	1.341714	-5.684618	0.0000
LOG(X1)	1.548392	0.134126	11.54433	0.0000

R-squared	0.898834	Mean dependent var	7.851729
Adjusted R-squared	0.892090	S.D. dependent var	0.614905
S.E. of regression	0.201994	Akaike info criterion	-0.251024
Sum squared resid	0.612025	Schwarz criterion	-0.152999
Log likelihood	4.133702	Hannan-Quinn criter.	-0.241280
F-statistic	133.2715	Durbin-Watson stat	0.498756
Prob(F-statistic)	0.000000		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.524339	0.816518	3.091589	0.0074
LOG(X2)	0.524083	0.079954	6.554772	0.0000

R-squared	0.741223	Mean dependent var	7.851729
Adjusted R-squared	0.723972	S.D. dependent var	0.614905
S.E. of regression	0.323061	Akaike info criterion	0.688181
Sum squared resid	1.565529	Schwarz criterion	0.786206
Log likelihood	-3.849542	Hannan-Quinn criter.	0.697925
F-statistic	42.96503	Durbin-Watson stat	0.373093
Prob(F-statistic)	0.000009		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.216611	0.598741	5.372295	0.0001
LOG(X3)	1.611849	0.206841	7.792691	0.0000

R-squared	0.801918	Mean dependent var	7.851729
Adjusted R-squared	0.788712	S.D. dependent var	0.614905
S.E. of regression	0.282648	Akaike info criterion	0.420899
Sum squared resid	1.198345	Schwarz criterion	0.518924
Log likelihood	-1.577644	Hannan-Quinn criter.	0.430643
F-statistic	60.72604	Durbin-Watson stat	0.687050
Prob(F-statistic)	0.000001		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.839669	0.419118	13.93325	0.0000
LOG(X4)	0.223036	0.045246	4.929444	0.0002

R-squared	0.618315	Mean dependent var	7.851729
Adjusted R-squared	0.592869	S.D. dependent var	0.614905
S.E. of regression	0.392351	Akaike info criterion	1.076812
Sum squared resid	2.309091	Schwarz criterion	1.174837
Log likelihood	-7.152900	Hannan-Quinn criter.	1.086556
F-statistic	24.29941	Durbin-Watson stat	0.323789
Prob(F-statistic)	0.000182		

比较 4 个一元回归结果我们知道，log(y)对 log(x1)做回归的这个模型的 R^2最高，把这个模型作为基础回归模型，再引入第二个回归变量。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-10.82473	2.623287	-4.126401	0.0010
LOG(X1)	2.081981	0.401796	5.181685	0.0001
LOG(X2)	-0.210184	0.149758	-1.403488	0.1823

R-squared	0.911312	Mean dependent var	7.851729
Adjusted R-squared	0.898643	S.D. dependent var	0.614905
S.E. of regression	0.195765	Akaike info criterion	-0.265018
Sum squared resid	0.536536	Schwarz criterion	-0.117980
Log likelihood	5.252649	Hannan-Quinn criter.	-0.250402
F-statistic	71.92878	Durbin-Watson stat	0.599881
Prob(F-statistic)	0.000000		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.637461	2.691627	-2.465966	0.0272
LOG(X1)	1.398239	0.376789	3.710936	0.0023
LOG(X3)	0.177826	0.415256	0.428232	0.6750

R-squared	0.900142	Mean dependent var	7.851729
Adjusted R-squared	0.885877	S.D. dependent var	0.614905
S.E. of regression	0.207728	Akaike info criterion	-0.146390
Sum squared resid	0.604112	Schwarz criterion	0.000647
Log likelihood	4.244319	Hannan-Quinn criter.	-0.131775

F-statistic	63.09972	Durbin-Watson stat	0.516050
Prob(F-statistic)	0.000000		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.728517	2.188718	-3.531070	0.0033
LOG(X1)	1.560886	0.250476	6.231669	0.0000
LOG(X4)	-0.002607	0.043501	-0.059926	0.9531

R-squared	0.898860	Mean dependent var	7.851729
Adjusted R-squared	0.884412	S.D. dependent var	0.614905
S.E. of regression	0.209057	Akaike info criterion	-0.133633
Sum squared resid	0.611868	Schwarz criterion	0.013404
Log likelihood	4.135882	Hannan-Quinn criter.	-0.119017
F-statistic	62.21112	Durbin-Watson stat	0.502835
Prob(F-statistic)	0.000000		

引入 log(x2)之后，修正的可决系数为 0.899，比原先只有一个解释变量时的可决系数 0.892 更高，说明修正的可决系数得到了改进，并且引入 log(x2)之后并不影响 log(x1)的显著性，因此考虑保留变量 log(x2)，引入 log(x3)和 log(x4)后修正的可决系数并没有得到改进，所以这两个变量不予考虑。接下来考虑引入第三个变量。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.191300	2.686328	-3.421511	0.0046
LOG(X1)	1.870519	0.402950	4.642057	0.0005
LOG(X2)	-0.367928	0.172592	-2.131778	0.0527
LOG(X3)	0.724702	0.451011	1.606840	0.1321

R-squared	0.926008	Mean dependent var	7.851729
Adjusted R-squared	0.908933	S.D. dependent var	0.614905
S.E. of regression	0.185562	Akaike info criterion	-0.328533
Sum squared resid	0.447631	Schwarz criterion	-0.132483
Log likelihood	6.792533	Hannan-Quinn criter.	-0.309046
F-statistic	54.23157	Durbin-Watson stat	0.858600
Prob(F-statistic)	0.000000		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-12.15128	2.321635	-5.233930	0.0002
LOG(X1)	2.573394	0.400345	6.427937	0.0000
LOG(X2)	-0.704177	0.240483	-2.928183	0.0118
LOG(X4)	0.159129	0.065412	2.432704	0.0302

R-squared	0.939056	Mean dependent var	7.851729
Adjusted R-squared	0.924992	S.D. dependent var	0.614905
S.E. of regression	0.168407	Akaike info criterion	-0.522538
Sum squared resid	0.368694	Schwarz criterion	-0.326487
Log likelihood	8.441570	Hannan-Quinn criter.	-0.503050
F-statistic	66.77041	Durbin-Watson stat	0.732388
Prob(F-statistic)	0.000000		

引入 log(x4)后修正的可决系数得到了改进，log(x1),log(x2)仍然是显著的，引入

log(x3)后虽然修正的可决系数得到了改进，但是改进的力度没有引入 log(x4)的大，并且 log(x3)后，log(x2)的显著程度并没有很明显的改进，所以经过对比分析后，我们引入了 log(x4)这个变量