

上海对外经贸大学

课程论文

论文名称： 新冠疫情下原油价格影响因素回归分析

课程名称： Stata 与金融实证分析

专 业： 金融学 1903

学 号： 19067080

学生姓名： 昌晓蕾

课程序号： 713.120.201

2022 年 05 月

新冠疫情下原油价格影响因素回归分析

1 引言

20 世纪 70 年代初发生的石油危机，给世界原油市场带来巨大冲击，石油价格剧烈波动，直接导致了石油期货的产生。石油期货诞生以后，其交易量一直呈现快速增长之势，已经超过金属期货，是国际期货市场的重要组成部分。石油在一定程度上代替黄金，成为美元的信用背书，有助于继续确保美元在国际货币体系中的核心地位。在过去的 40 多年里，美国用以信用担保的纸币换取世界各国的商品和资源，同时又以美债等美元计价资产吸收各国创造的财富和储蓄，美国以其特殊的经济金融地位，维持着石油美元环流，使美国长期呈现消费膨胀、外贸逆差和大量吸收外资并存的局面，美国经济亦得以在这种特殊的格局中增长。

而 2020 年新冠疫情在全球大范围传播后，全球原油期货交易急剧下跌，流入资金减少，说明新冠肺炎疫情对世界石油行业影响的不仅仅是油价，还冲击了原油期货交易本身。新冠肺炎疫情严重冲击了国际石油市场，急剧萎缩的需求带来了全球性的石油供应过剩，世界石油行业陷入困境。从今年 3 月中旬开始以来，国际石油价格持续走低并长时间维持在低位，特别是 2020 年 4 月 20 日出现的世界石油工业 160 年来首次负油价，是新冠肺炎疫情对世界石油行业严重冲击的最直接表现。与油价持续走低相一致的是，从 4 月份以来，原油期货交易活动也陷入清淡，流入原油期货交易活动的资金在减少。

因此本篇文章在此背景下通过 Stata 数据分析研究新冠疫情下原油价格的影响因素。

2 检测方法

2.1 OLS 普通最小二乘法

OLS 线性回归，即通过对一系列预测变量建立以 β 为参数的线性多元回归模型来预测响应变量的方法，其中，参数 β 只以一次方的形式出现。其基本原则为：最优拟

合曲线应该使各点到直线的距离的平方和（即残差平方和，简称 RSS）最小。OLS 线性回归模型的基本形式如下：

OLS 线性回归，即通过对一系列预测变量建立以 β 为参数的线性多元回归模型来预测响应变量的方法，其中，参数 β 只以一次方的形式出现。其基本原则为：最优拟合曲线应该使各点到直线的距离的平方和（即残差平方和，简称 RSS）最小。OLS 线性回归模型的基本形式如下：

$$Y_t = \alpha + \beta X_t + \mu_t \quad (n = 1, 2, \dots, n) \quad (\text{公式 1})$$

其中， t 表示观测数； Y_t 为预测变量（因变量）； X_t 为响应变量（自变量）； α 、 β 为需要最小二乘法去确定的参数（回归系数）； μ_t 为随机误差项

为了能够恰当地解释 OLS 模型的系数，数据必须满足以下统计假设：正态性，即对于固定的自变量值，因变量值成正态分布；独立性，即个体之间相互独立；线性相关性，即因变量和自变量之间为线性相关；同方差性，即因变量的方差不随自变量的水平不同而变化，即因变量的方差是不变的。

2.2 WLS 加权最小二乘法

WLS 加权最小二乘法是对原模型进行加权，使之成为一个新的不存在异方差性的模型，然后采用普通最小二乘法估计其参数的一种数学优化技术，适用于消除模型的异方差性。

加权最小二乘法采用指数权数 W^{n-i} ， $0 < W < 1$ ，加权以后求得的参数估计值满足：WLS 加权最小二乘法是对原模型进行加权，使之成为一个新的不存在异方差性的模型，然后采用普通最小二乘法估计其参数的一种数学优化技术，适用于消除模型的异方差性。

加权最小二乘法采用指数权数 W^{n-i} ， $0 < W < 1$ ，加权以后求得的参数估计值满足：

$$S = \sum_{i=1}^n W^{n-i} (y_i - \hat{y}_i)^2 = \min, (i = 1, 2, \dots, n) \quad (\text{公式 2})$$

以直线模型为例，其加权的剩余平方和为：

$$S = \sum_{i=1}^n W^{n-i} (y_i - a - bt)^2 \quad (\text{公式 3})$$

对上式分别求 a 和 b 的偏导数，得到标准方程组：

$$\sum W^{n-1} y_i = a \sum W^{n-i} + b \sum W^{n-1} t \quad (\text{公式 4})$$

$$\sum W_{n-1}ty = a\sum W^{n-i}t + b\sum W^{n-i}t^2 \quad (\text{公式 5})$$

对上述方程解出 a 和 b ，就得到加权最小二乘法直线模型。应用加权最小二乘法， W 的取值不同，解出的 a ， b 也不同，因此 W 值取多少，需要经分析后确定。

2.3 FGLS 可行广义最小二乘法

FGLS 可行广义最小二乘法是在广义最小二乘法中，先用模型残差的方差 - 协方差矩阵的估计值替代真实值，然后采取广义最小二乘法估计模型参数的一种估计方法。当用 Stata 操作后的模型存在自相关问题时，则可以使用 FGLS 估计这个特殊的双向固定效应模型。由于公式较为复杂，本文此处不再赘述。

3 实证分析

3.1 数据收集与模型建立

考虑到数据的可获得性、完整性与时效性，并结合新冠疫情爆发的时间点，我们通过查找 Federal Reserve Economic Data(联邦储备经济数据库)、World Health Organization（世界卫生组织官网）等渠道，选取 2020 年 1 月 3 日-2022 年 4 月 8 日美国西得克萨斯的轻质原油（即 WTI 原油）的价格作为因变量进行研究。该原油期货合约具有良好的流动性及很高的价格透明度，是世界原油市场上的三大基准价格之一，所有在美国生产或销往美国的原油在计价时都以轻质低硫的 WTI 作为基准，因此该原油的价格具有较好的代表性。

根据相关资料，我们用股票市场波动性传染病追踪器 *INFECTDISEMVTRACKD* (x_1) 衡量美国经济的不确定性；用美国新冠疫情每日新增病例 *New_cases* (x_2) 作为衡量美国疫情发展与控制情况的指标；用天然气现货价格 *GasSpotPrice* (x_3) 衡量能源市场其他产品的价格变动情况；用纳斯达克指数 *NasdaqCIndex* (x_4) 作为衡量美国资本市场整体走势的指标，运用这些数据进行回归分析。

采用的模型如下：

$$\ln wti = \beta_1 + \beta_2 * \ln x_1 + \beta_3 * \ln x_2 + \beta_4 * \ln x_3 + \beta_5 * \ln x_4 \quad (\text{公式 6})$$

其中， $\ln wti$ 代表 wti 原油价格增长率， $\ln x_1$ 代表股票市场波动性传染病追踪器增长率， $\ln x_2$ 代表美国新冠疫情每日新增病例增长率， $\ln x_3$ 代表天然气现货价格增长率， $\ln x_4$ 代表纳斯达克指数增长率。我们通过对该模型的回归分析，得出各个变量增长率与 wti 原油价格增长率的变动关系。

3.2 模型估计与检验

3.2.1 初始模型估计

我们用 OLS 最小二乘法对模型进行估计，得到回归结果如下：

. reg lnwti ln x1 ln x2 ln x3 ln x4						
Source	SS	df	MS	Number of obs	=	503
Model	75.8254607	4	18.9563652	F(4, 498)	=	627.72
Residual	15.0389194	498	.030198633	Prob > F	=	0.0000
				R-squared	=	0.8345
				Adj R-squared	=	0.8332
Total	90.8643801	502	.181004741	Root MSE	=	.17378

lnwti	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnx1	.0296454	.0147839	2.01	0.045	.0005989	.0586918
lnx2	-.0616942	.0048312	-12.77	0.000	-.0711863	-.0522022
lnx3	.3422412	.0365471	9.36	0.000	.2704357	.4140467
lnx4	1.632577	.0861307	18.95	0.000	1.463353	1.801801
_cons	-11.18297	.7779635	-14.37	0.000	-12.71146	-9.654473

图 1 OLS 最小二乘法模型估计

初步回归得到的模型为：

$$\ln wt_i = -11.18297 + 0.0296454 \ln x_1 - 0.0616942 \ln x_2 + 0.3422412 \ln x_3 + 1.632577 \ln x_4 \quad (\text{公式 7})$$

其中， $R^2=0.8345$ ，修正后的 $R^2=0.8332$ ， $F=627.72$

3.2.2 显著性和拟合优度检验

关于显著性检验，针对原假设： $\beta_2=\beta_3=\beta_4=0$ ，给定显著性水平 $\alpha=0.05$ ，自由度为 $k-1=3$ 和 $n-k=527-4=523$ ，由于 n 较大，我们选取一个间接临界值进行比较。由于自由度越低， F 越大，故我们在 F 分布表中查出临界值 $F_\alpha(3,200)=3.18$ ， $F=627.72 >> F(3,200)=3.18 > F(3,523)$ ，应拒绝原假设，表明从整体上看，模型的因变量与各自变量间的线形关系较为显著。

关于拟合优度检验，由表可知： $R^2=0.8345$ ，修正后的 $R^2=0.8332$ 。由于我们建立的回归模型属于解释型回归，更多关注模型整体显著性以及自变量的统计显著性和经济意义显著性。因此，即使 R 方接近于1的程度不明显，但我们仍可以认为该模型拟合情况良好。另外，各变量的系数正负也符合基本常识，即原油价格增长率（ $\ln wt_i$ ）

与美国经济不确定性增长率 ($\ln x_1$)、天然气现货价格增长率 ($\ln x_3$)、纳斯达克指数增长率 ($\ln x_4$) 成正比, 与美国新冠疫情每日新增病例增长率 ($\ln x_2$) 成反比。

3.2.3 多重共线性检验

对于多重共线性, 我们采取的判断指标为方差膨胀因子。方差膨胀因子的计算公式如下:

$$\text{Var}(\widehat{\beta}_k|X) = \frac{\sigma^2}{(1-R_k^2)S_k} \quad (\text{公式 8})$$

$$\text{VIF}_k = \frac{1}{1-R_k^2} \quad (\text{公式 9})$$

由上述公式可知, 方差膨胀因子构成了变量估计值方差的一部分, 如果方差膨胀因子过大, 将会导致解释变量的系数方差增大, 将会导致检验难以通过。

因此, 我们用命令"*estat vif*"在*stata*中检验该模型的方差膨胀因子。结果显示, $\ln x_1, \ln x_2, \ln x_3, \ln x_4$ 的方差膨胀因子均较小 (<10), 说明该模型不存在多重共线性。

. estat vif		
Variable	VIF	1/VIF
lnx4	4.89	0.204438
lnx3	3.39	0.295069
lnx1	1.64	0.608404
lnx2	1.49	0.670894
Mean VIF	2.85	

图 2 方差膨胀因子检验

3.2.4 异方差性检验

对于异方差性, 我们首先在*stata*中用命令 "*estat imtest, white*" 对该模型进行怀特检验, 检验结果如下图:

```
. estat imtest,white
```

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14) = 169.62
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	169.62	14	0.0000
Skewness	9.77	4	0.0445
Kurtosis	3.28	1	0.0701
Total	182.67	19	0.0000

图3 怀特检验

在给定 $\alpha = 0.05$ 的状况下,怀特检验的卡方统计量 $\chi^2(14) = 169.62$, 对应 p 值为 0, 没有通过“方程不存在异方差”的原假设, 说明方程存在异方差性, 仍需进一步对模型进行修正。

其次, 对于异方差性, 我们还可以通过作残差图进行判断, 基本原理为: 通过观察残差值是否在一个数值附近波动, 如图像中所有散点沿 x 轴的变动幅度不大, 则说明不存在异方差性; 如图像中所有散点沿 x 轴的变动幅度较大, 则说明存在异方差性。

如下图可知, 股票市场波动性传染病追踪器增长率 ($\ln x_1$)、美国新冠疫情每日新增病例增长率 ($\ln x_2$) 沿 x 轴的方差并没有明显的变化, 表明不存在异方差性; 而天然气现货价格增长率 ($\ln x_3$)、纳斯达克指数增长率 ($\ln x_4$) 沿 x 轴的数值波动则较大, 表明模型整体仍存在异方差性。

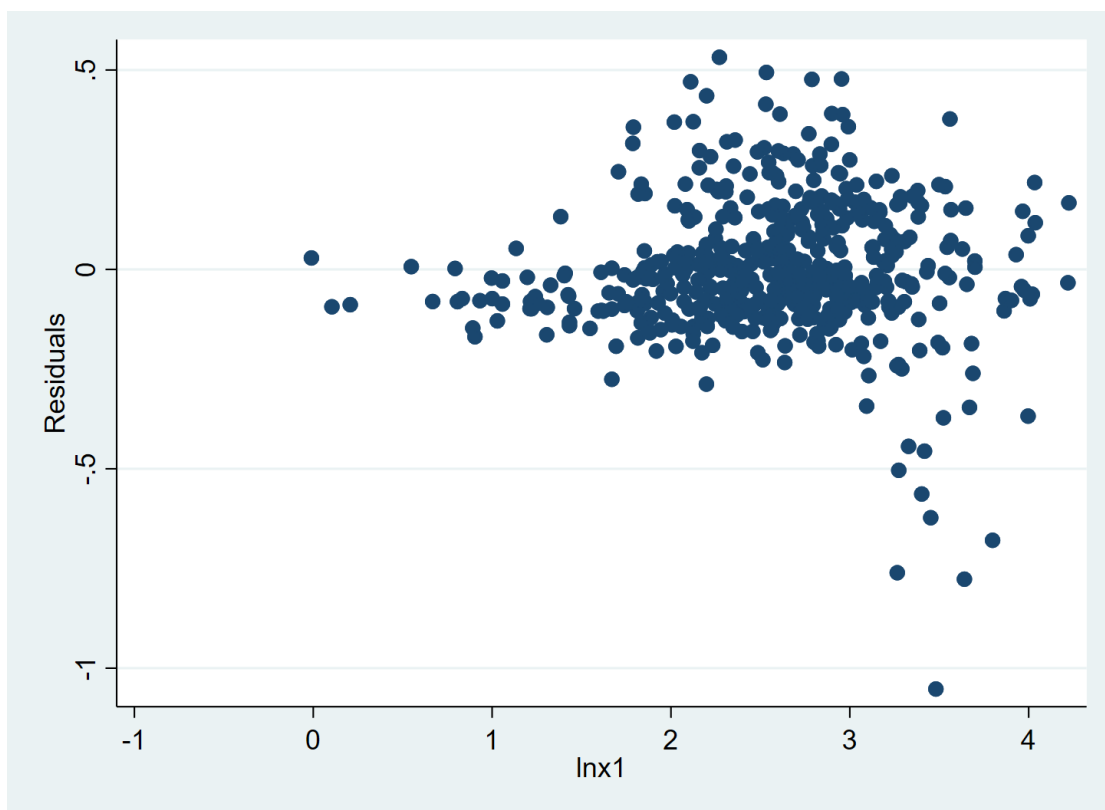


图 4 $\ln x_1$ 与残差的散点图

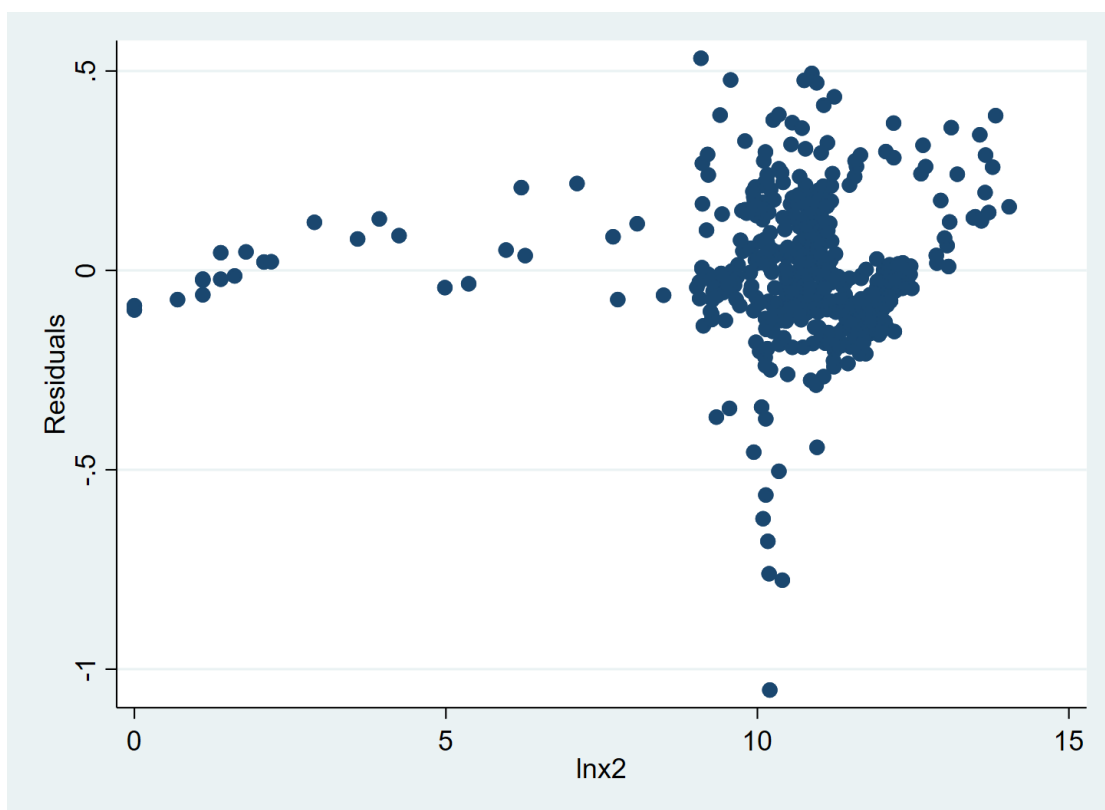


图 5 $\ln x_2$ 与残差的散点图

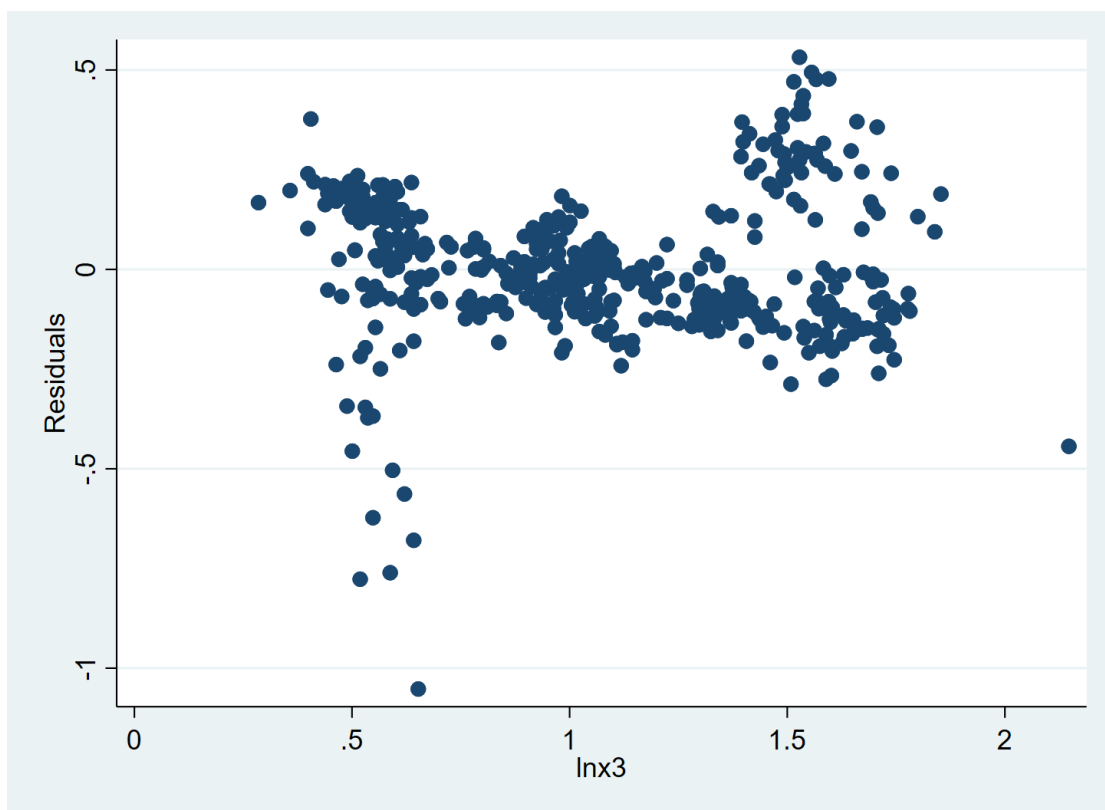


图 6 $\ln x_3$ 与残差的散点图

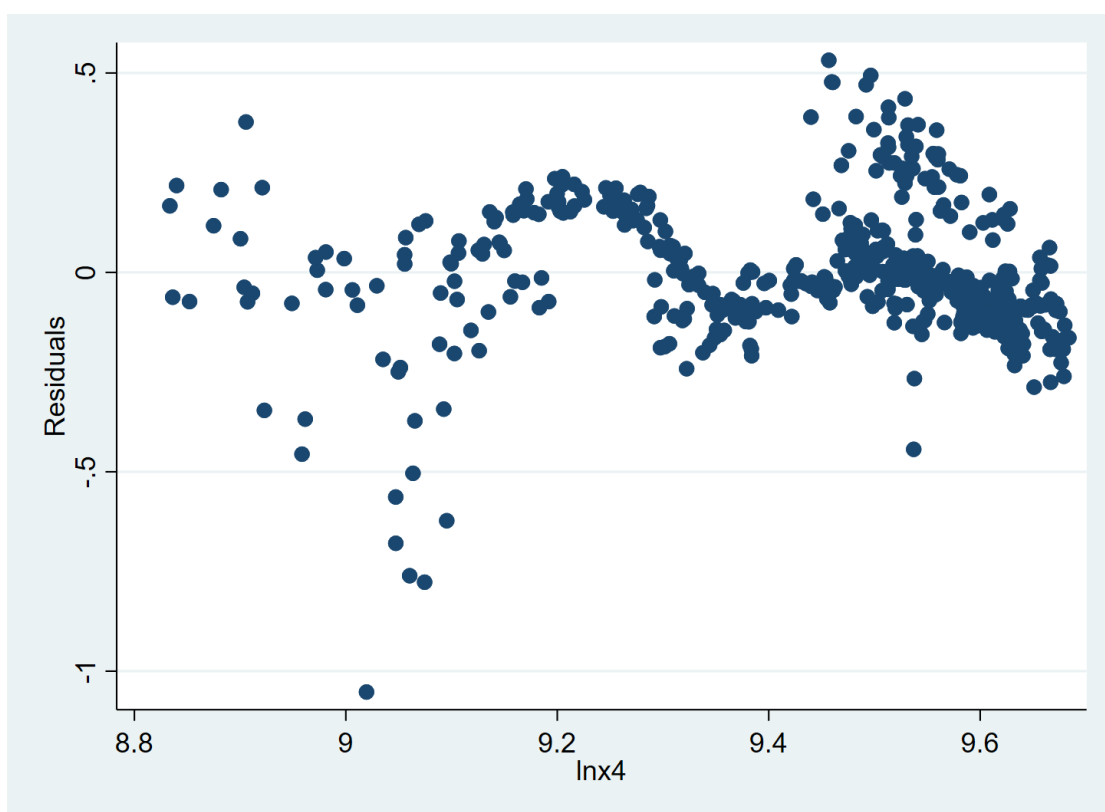


图 7 $\ln x_4$ 与残差的散点图

对于模型的异方差性修正，我们采取 WLS 加权最小二乘法对模型进行修正。修正

后的结果如下：

<pre>. regress lnwti ln x1 ln x2 ln x3 ln x4[aw=1/e3f],robust (sum of wgt is 90,060.2190925885)</pre>						
Linear regression			Number of obs = 503 F(4, 498) = 424.49 Prob > F = 0.0000 R-squared = 0.8272 Root MSE = .15972			
lnwti	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lnx1	.0405317	.0090549	4.48	0.000	.0227413	.0583222
lnx2	-.058019	.0032461	-17.87	0.000	-.0643966	-.0516413
lnx3	.3702502	.0446497	8.29	0.000	.2825251	.4579753
lnx4	1.512371	.1051151	14.39	0.000	1.305847	1.718895
_cons	-10.14437	.9404527	-10.79	0.000	-11.99211	-8.296627

图 8 WLS 加权最小二乘法模型估计

由图可知，修正后的模型为：

$$\ln wt_i = -10.14437 + 0.0405317 * \ln x_1 - 0.058019 * \ln x_2 + 0.3702502 * \ln x_3 + 1.512371 * \ln x_4 \quad (\text{公式 } 10)$$

其中， $R^2=0.8272$ ， $F=424.49$ ，模型总体以及单个变量的 p 值均接近于 0，表明修正后的模型接受原假设，即不存在异方差性。

3.2.5 自相关性检验

对于自相关性，我们首先在 *stata* 中用命令 “*wntestq e1*” 对该模型进行 q 检验，检验结果如下图：

<pre>. wntestq e1 (note: time series has 4 gaps)</pre>	
Portmanteau test for white noise	
Portmanteau (Q) statistic =	3328.6448
Prob > chi2(40) =	0.0000

图 9 q 检验

在给定 $\alpha = 0.05$ 的状况下, Q 检验的统计量 $\chi^2(40) = 55.7585$, 对应 p 值为 $0.000 < 0.05$, 拒绝“方程扰动项不存在自相关”的原假设, 说明方程扰动项存在自相关。

另外, 我们还通过命令 “`predict e1, r`” “`twoway scatter e1 L.e1 || lfit e1 L.e1`” 绘制了残差与残差滞后的散点图, 来观察模型是否存在自相关性。

由下图可知, 时间趋势与残差散点图存在一定的正相关关系, 因此, 我们可以初步认为模型存在一阶正自相关。

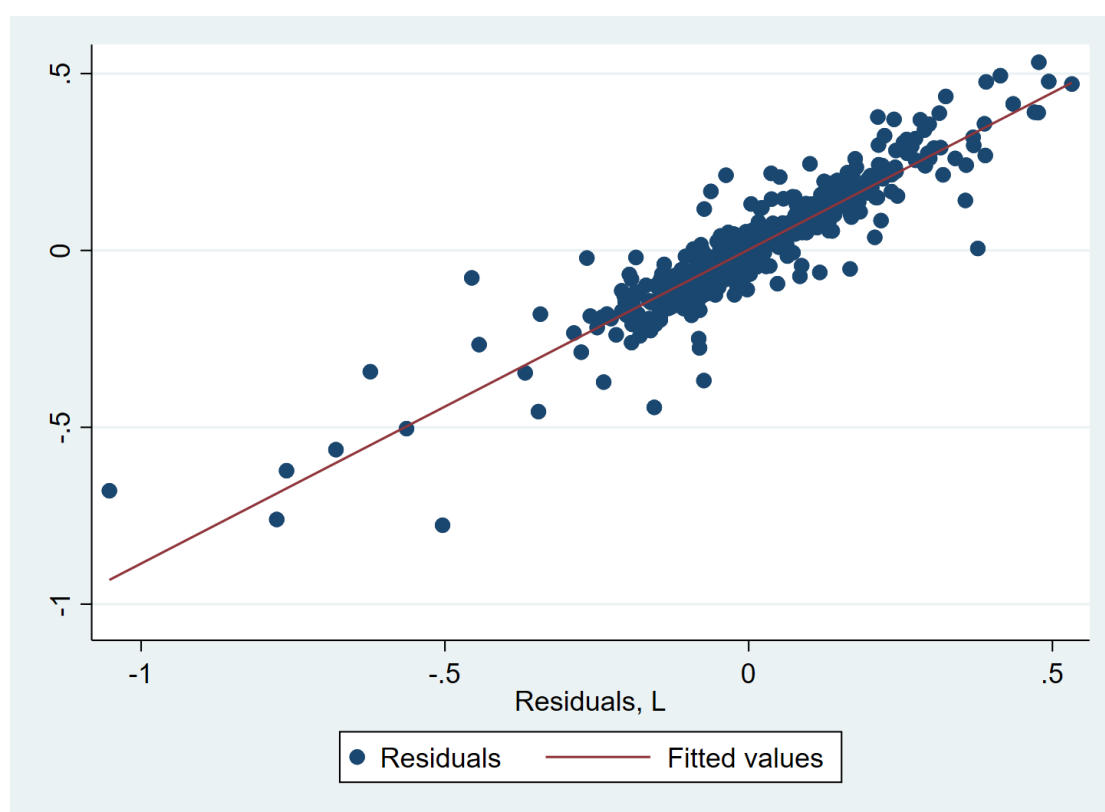


图 10 残差与残差滞后的散点图

除此之外, 我们还运用命令 “`ac e1`” “`pac e1`” 绘制了残差的自相关图及偏自相关图, 结果如下图所示。其中, 灰色的区域表示的是置信度水平 $\alpha = 0.05$ 下自相关的置信区间, 如果图像在灰色区域之内, 能够比较好地拒绝“模型存在自相关”这一原假设; 如果图像超过灰色区域, 则不能拒绝原假设, 即模型存在自相关。由图可知, 偏自相关和自相关的第一根线明显超出了灰色区域, 因此判定模型存在一阶自相关。

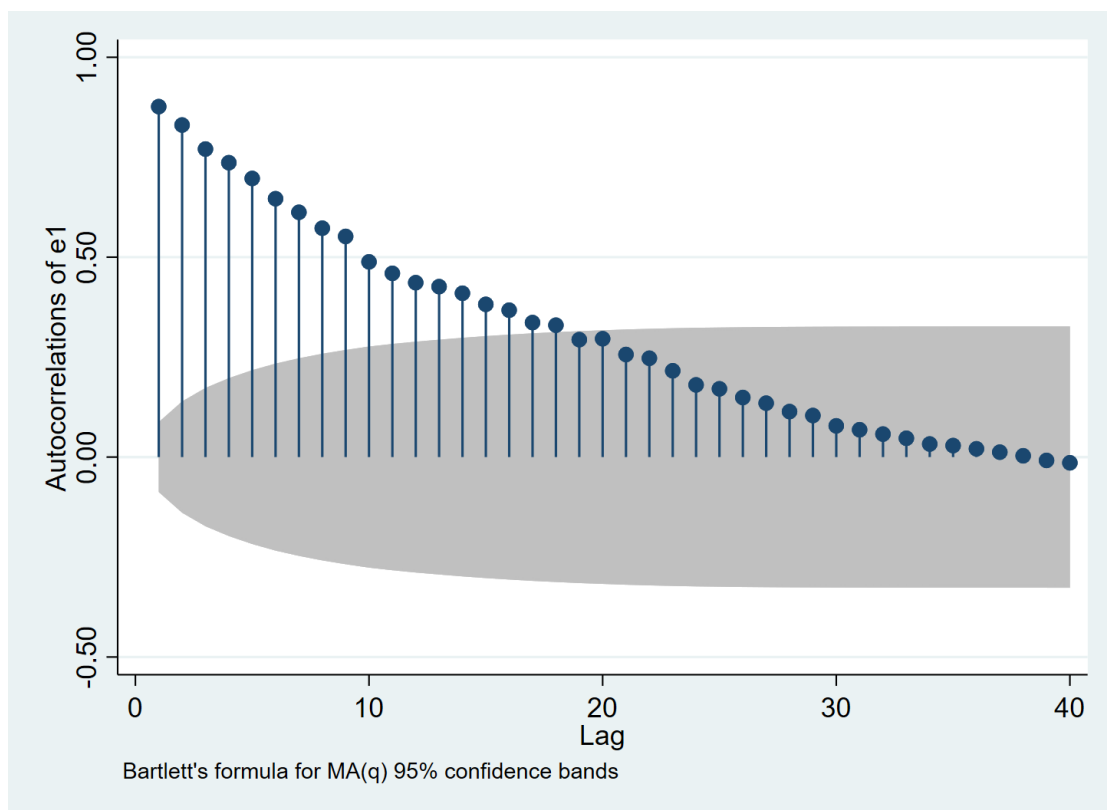


图 11 残差的自相关图

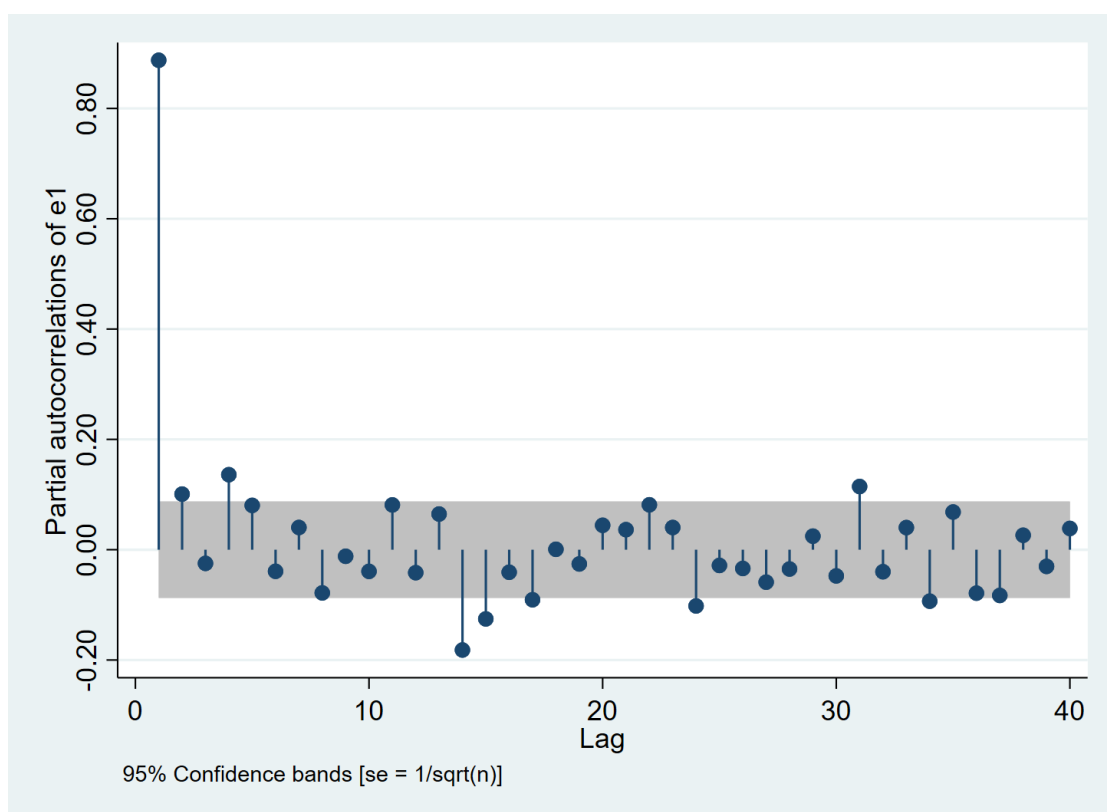


图 12 残差的偏自相关图

对于模型的自相关性，我们采取 FGLS 可行广义最小二乘法对模型进行修正。可

行广义最小二乘法的原理是不断试错迭代，最终得到一个合理的 ρ 值，即可消除模型自相关性。修正后的结果如下：

```
. prais lnwti ln x1 ln x2 ln x3 ln x4

Number of gaps in sample: 4
(note: computations for rho restarted at each gap)

Iteration 0: rho = 0.0000
Iteration 1: rho = 0.8874
Iteration 2: rho = 0.9376
Iteration 3: rho = 0.9585
Iteration 4: rho = 0.9680
Iteration 5: rho = 0.9720
Iteration 6: rho = 0.9736
Iteration 7: rho = 0.9743
Iteration 8: rho = 0.9746
Iteration 9: rho = 0.9747
Iteration 10: rho = 0.9747
Iteration 11: rho = 0.9747
Iteration 12: rho = 0.9747
Iteration 13: rho = 0.9747
Iteration 14: rho = 0.9747
Iteration 15: rho = 0.9747
```

图 13 ρ 值的迭代

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	503
Model	2.44752025	4	.611880063	F(4, 498)	=	208.41
Residual	1.46207355	498	.002935891	Prob > F	=	0.0000
				R-squared	=	0.6260
				Adj R-squared	=	0.6230
Total	3.90959381	502	.007788035	Root MSE	=	.05418

lnwti	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnx1	-.0017479	.004065	-0.43	0.667	-.0097345	.0062387
lnx2	-.0083937	.0055975	-1.50	0.134	-.0193914	.0026039
lnx3	.0273229	.0330329	0.83	0.409	-.0375781	.0922238
lnx4	.6338058	.1244131	5.09	0.000	.3893665	.8782451
_cons	-1.985812	1.153944	-1.72	0.086	-4.253012	.2813869
rho	.9747323					

Durbin-Watson statistic (original) 0.159440

Durbin-Watson statistic (transformed) 2.111041

图 14 FGLS 可行广义最小二乘法模型估计

由图可知，修改后的模型 p 值为 0，DW 值为 2.111041，即此时模型不存在自相关性。虽然 R^2 有所减小，但重要的是修改后的模型不存在自相关性，因此修正后的模型更具有说服力。

综上，最终得到的模型为：

$$\ln wti = -0.1985812 - 0.0017479 * \ln x_1 - 0.0083937 * \ln x_2 + 0.0273229 * \ln x_3 + 0.6338058 * \ln x_4 \quad (\text{公式 11})$$

其中， $R^2=0.6260$ ， $F=208.41$

模型的经济意义为：在其他变量保持不变的情况下，股票市场波动性传染病追踪器增长率（ $\ln x_1$ ）每增加个 1 单位，则 wti 原油价格增长率平均减少 0.0017479 个单位；美国新冠疫情每日新增病例增长率（ $\ln x_2$ ）每增加个 1 单位，则 wti 原油价格增长率平均减少 0.0083937 个单位；天然气现货价格增长率（ $\ln x_3$ ）每增加个 1 单位，则 wti 原油价格增长率平均增加 0.0273229 个单位；纳斯达克指数增长率（ $\ln x_4$ ）每增加个 1 单位，则 wti 原油价格增长率平均增加 0.6338058 个单位。

4 结论分析

根据拟合得出的模型来看，原油价格增长率与股票市场波动性传染病追踪器增长率呈负相关，即原油价格增长率与美国金融市场的不确定性增长率呈负相关，因此，当宏观经济不确定性增加时将导致原油价格下降；此外，原油价格的增长率也与美国新冠疫情每日新增病例增长率呈负相关，说明倘若新冠疫情感染人数得到控制将有利于原油市场价格的稳定，由此可见新冠疫情确实给宏观经济带来了不可忽略的影响；而原油价格的增长率与天然气价格的增长率正向相关，说明原油与天然气价格基本稳固同变化，体现出原油天然气市场的一体性。

最后，对原油价格增长率正向影响最大的是纳斯达克指数的增长率。因纳斯达克综合指数是所有在纳斯达克交易的股票的资产加权指数，当纳斯达克指数上升时，证明美国股票交易市场活跃，因此这种资本市场的繁荣会在很大程度上促进原油价格的上升，反之若资本市场发生动荡可能会导致原油价格骤跌。

长期以来，原油期货市场一直是全球重点金融市场之一，由于政治、经济和自然环境等方面的极端事件冲击，自其发展以来已经发生了多次原油价格危机。且若美国原油价格起伏较大时势必会对世界石油市场产生一定影响。因此，倘若要维持原油价

格的稳定，那么就要尽量减少宏观经济的不确定性，控制新冠疫情的传播范围和感染人数，稳定同市场中天然气的价格和资本市场、股票市场的繁荣。