



上海對外經貿大學

SHANGHAI UNIVERSITY OF INTERNATIONAL BUSINESS AND ECONOMICS

题目：加密货币时间序列数据分析

院 系： 统计与信息学院

课 程 名 称： 应用时间序列

任 课 教 师： 李睿

学 生 姓 名： 陶盛皿

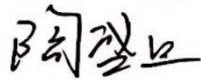
学 号： 19024075

专 业： 大数据 1901

考试诚信承诺书

本人郑重承诺：在 2021-2022 学年第二学期课程期末考试中，严格遵守学校《学生考试规定》，独立完成考试（论文、报告、作业等），不违纪，不作弊，如有违反，按学校规定接受处理。

学生签名：



日期：2022 年 6 月 1 日

摘要

“数字经济”是我国“十四五”规划和 2035 远景目标纲要的重要概念，随着“元宇宙”的概念兴起，投资者将更加关注加密货币（本文以比特币和以太坊为例），因此研究两币交易量、收盘价等变量的时间序列数据具有重要意义。本文通过季节自回归移动平均(SARIMA)模型、多元时序协整分析与建立误差修正模型(ECM-VAR)、深度学习方法建立 LSTM 模型进行时序分析与预测，得出：两币时序非平稳且具有季节性，两币交易量具有协整关系，即需求量存在长期均衡，可将 LSTM 和 SARIMA 模型结合分别预测收盘价和交易量以获得更高的准确度。

关键词：加密货币，SARIMA，ECM-VAR，LSTM

加密货币时间序列数据分析

引言

自 2020 年起新冠疫情突然爆发极大冲击了全球经济，形成了全球经济严峻态势。此外，随着云计算、移动互联网、大数据、人工智能等数字技术的快速创新与应用，数字经济正在成为全球经济社会发展的重要引擎。在全球逆势经济背景下，全球数字经济在逆势中加速腾飞。我国“数字经济”的概念和加快经济数字化发展在“十四五”规划和 2035 远景目标纲要中被提出。数字货币是数字经济中关键要素，比特币和以太坊属于数字货币的范畴，自 2018 年以来，投资者对其的兴趣呈爆炸式增长，目前其市值分别位列第一和第二¹。

数字经济是计算机科学与技术、互联网、传统经济融合，随着“元宇宙”的概念被提出，其延伸与发展了数字经济至虚拟世界²，其中加密货币由于需要可追溯、透明、去中心化的加密支付方式是实现虚拟现实（即“元宇宙”）的必要部分³。“元宇宙”的概念进一步提升了投资者对加密货币（本文以比特币和以太坊为例）的兴趣和关注程度，因此研究两币交易量、收盘价等变量的时间序列数据具有重要意义。

本文通过三种方法分析两币时序数据：首先，通过时间序列分析发现时序具有季节性，建立季节自回归移动平均(SARIMA)模型进行预测；通过多元时序分析研究两币序列的协整关系，建立 VAR 模型和误差修正模型(ECM-VAR)；最后，通过深度学习方法，建立 LSTM 模型进行时序预测。

一、数据集介绍

本数据来源于雅虎量化金融的 API 接口。通过接口获取了比特币和以太坊近三年（自 2019 年 6 月 2 日至 2022 年 6 月 2 日）的交易历史数据，本文针对这两种加密货币进行时间序列分析，建模目的是：预测这两种货币的未来走势。

（一）数据集介绍

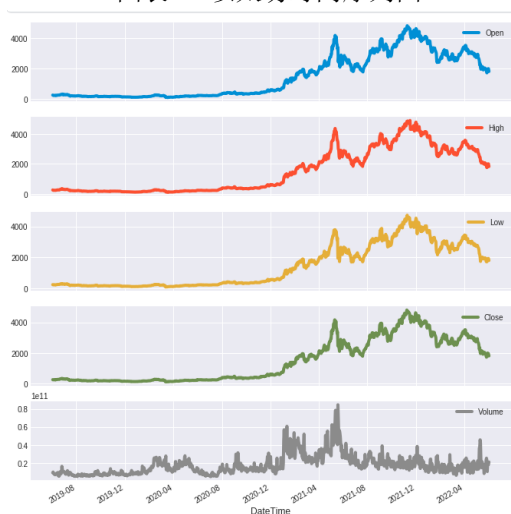
数据集包含比特币和以太坊近三年的历史交易，包含：交易时间、开盘价、最高价、最低价、收盘价、交易量，总计六个变量。

数据总量：1096。

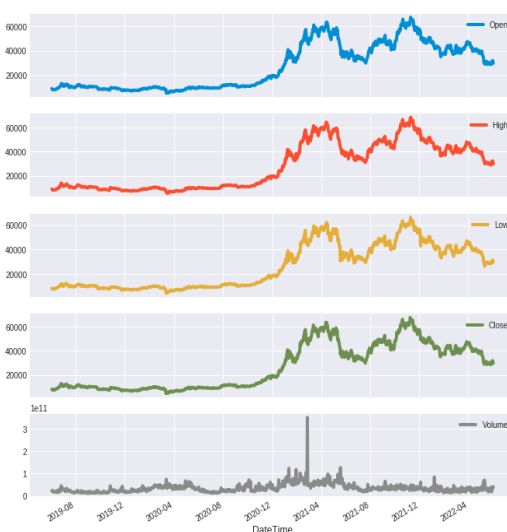
（二）时序图可视化

从比特币和以太坊的时序图可以看出近三年其市值总体上从呈现增长的趋势，自 2020 年迎来快速增长，但于 2021 年年底两种加密货币同步出现回落。近三年，其市值和交易量都呈现出剧烈波动、不稳定的状态。因而，需要使用时间序列分析方法，分析统计特征和数据之间的相依关系，拟合能反映其动态变化规律的模型进行科学预测。

图表 1 以太坊时间序列图

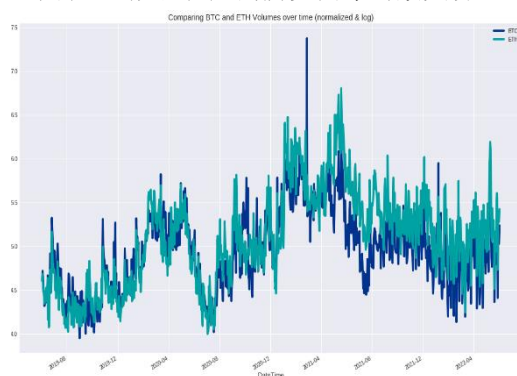


图表 2 比特币时间序列图

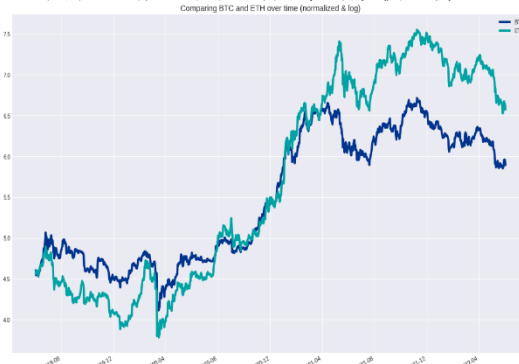


我们将数据规范化比较这 2 个货币的最高价和交易量进行对比，可以看到比特币和比特币呈现出剧烈波动性，较难发现趋势，将在下一节中进行季节差分。

图表 3 规范化后的历史两币对数交易量



图表 4 规范化后的两币对数收盘价

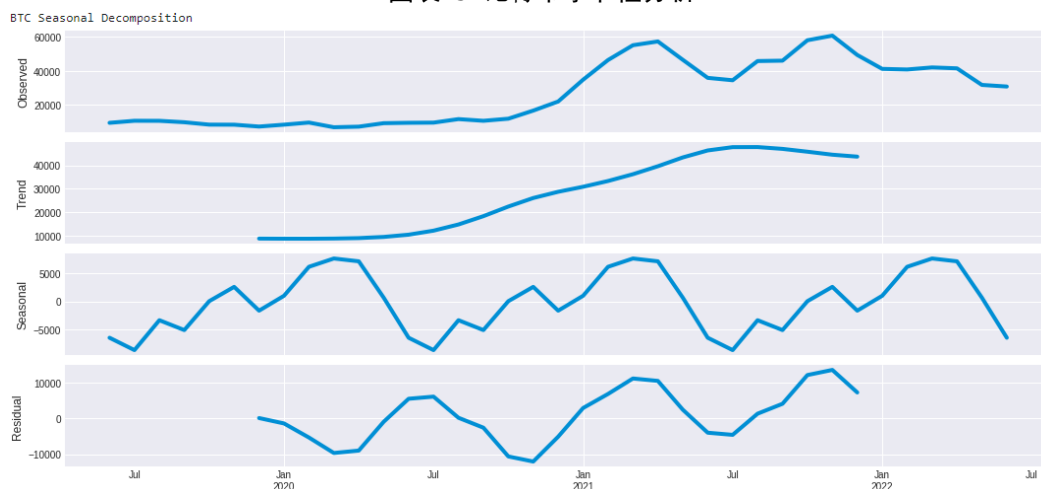


二、非平稳性序列

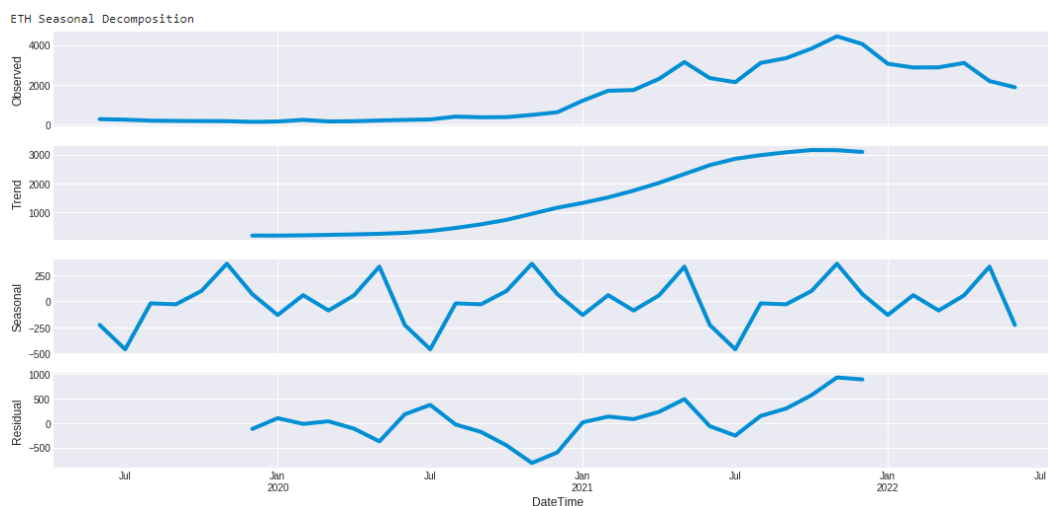
(一) 季节性差分

本数据集由于是每日数据，对规范化后的对数收盘价进行差分，即使 200 阶差分依旧难以平稳，因此对原始数据进行月度平均值的重采样后，使用 `python statsmodels.tsa.seasonal.seasonal_decompose` 函数进行季节性分解，通过可视化可以明显看到比特币和以太坊的收盘价具有显著的季节性。该函数使用了移动平均进行季节分解，第一行 **observed** 图是经过月度平均值重采样之后观测的实际值，可以看到和原始序列相比具有平稳性。第二行 **trend** 图是重采样后的整体变化趋势，比特币和以太坊月度均呈现出上升趋势。第三行 **seasonal** 图可以明显看出比特币和以太坊序列具有季节性。第四行则是残差图。接下去，采用单位根进行平稳性判断，采用 ACF 和 PACF 进行模型定阶。

图表 5 比特币季节性分析



图表 6 以太坊季节性分析



(二) 单位根检验

1. 比特币

对比特币月度重采样后的收盘价使用 `scipy.stats.adfuller` 进行单位根检验，参数 `autolag` 设置为 `AIC`，即自动以最小化相应的信息标准确定滞后数。首先，参数 `regression` 设置为 `ct`，即具有常数项和时间趋势项的 ADF 检验。P 值为 0.306338，大于 0.05，需要对模型常数项进行进一步检验。其次参数 `regression` 设置为 `c`，即具有常数项和没有时间趋势项的 ADF 检验。P 值为 0.441503，大于

0.05，仍然需要对模型进行进一步检验。最后，参数 regression 设置为 nc，即没有常数项和没有时间趋势项的 ADF 检验。P 值为 0.426658，大于 0.05，因此不拒绝原假设。综上，认为原始序列非平稳具有单位根。

接着，对该序列进行一阶差分。首先，参数 regression 设置为 ct，即具有常数项和时间趋势项的 ADF 检验。P 值为 0.013859，小于 0.05，因此拒绝原假设，认为序列平稳。

图表 7 比特币一阶差分后单位根检验结果

```
result_format is
Test Statistic          -3.857666
p-value                 0.013859
Lags Used               1.000000
Number of Observations Used 34.000000
Critical Value (1%)     -4.253043
Critical Value (5%)     -3.548590
Critical Value (10%)    -3.207062
The maximized information criterion if autolag is not None. 512.329728
dtype: float64
```

2. 以太坊

对以太坊月度重采样后的收盘价使用相同方法和 AIC 信息选择标准进行单位根检验。首先，参数 regression 设置仍为 ct。P 值为 0.274388，大于 0.05，因此需要对模型常数项进行进一步检验。其次参数 regression 设置为 c，即具有常数项和没有时间趋势项的 ADF 检验。P 值为 0.018951，小于 0.05，因此拒绝原假设。综上，认为序列在具有常数项和没有时间趋势项下的单位根检验平稳。

图表 8 以太坊具有常数项没有时间趋势的单位根检验结果

```
result_format is
Test Statistic          -3.218268
p-value                 0.018951
Lags Used               8.000000
Number of Observations Used 28.000000
Critical Value (1%)     -3.688926
Critical Value (5%)     -2.971989
Critical Value (10%)    -2.625296
The maximized information criterion if autolag is not None. 395.547639
dtype: float64
```

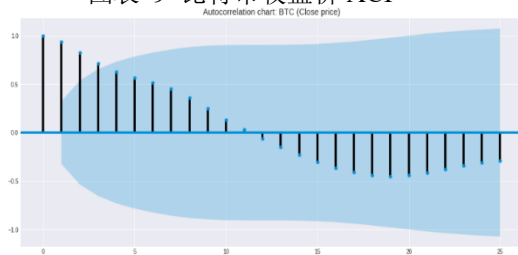

(三) 自相关函数和部分自相关函数

自相关函数 (ACF) 衡量一个序列在不同的滞后下与其自身相关的情况；部分自相关函数 (PACF) 可以解释序列对其过去滞后的回归。通过可视化可以快速判断截尾或者拖尾的阶数 (K 阶截尾是指：在大于某个常数 K 快速趋于零；拖尾是指：始终有非零取值，即不在 K 大于某个常数后恒等于零或在零附近随机波动)。这是快速基于序列平稳性前提的快速定阶法。

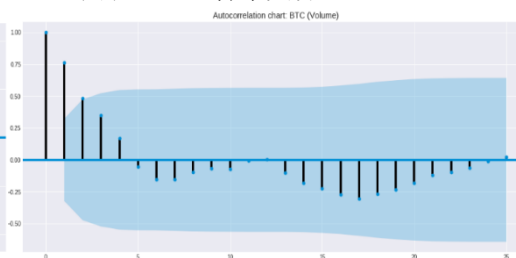
在上一节中比特币一阶差分后序列和以太坊序列均已通过单位根检验序列平稳，可以使用 ACF 和 PACF 进行定阶，即通过可视化后的值是否落入 2 倍标准差的置信区间来确定阶数。

1. ACF

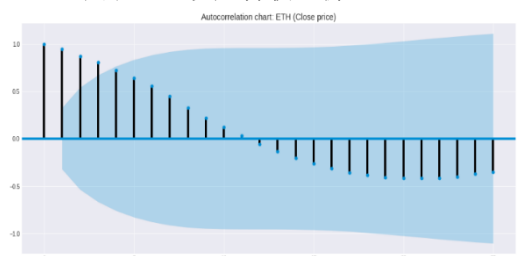
图表 9 比特币收盘价 ACF



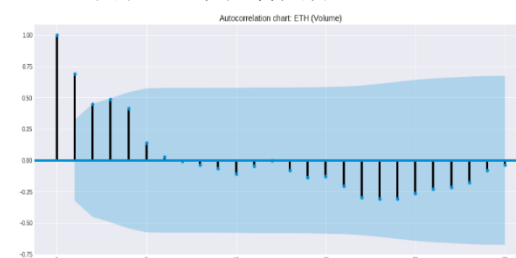
图表 10 比特币交易量 ACF



图表 11 以太坊收盘价 ACF



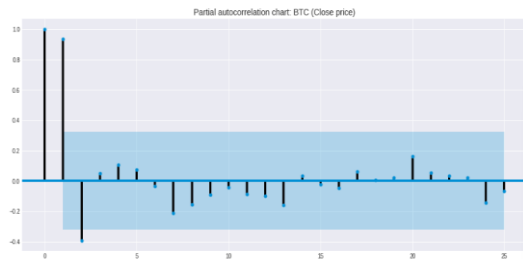
图表 12 以太坊交易量 ACF



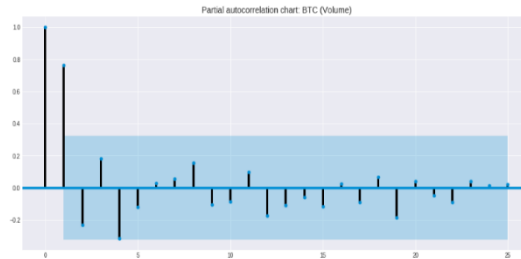
比特币收盘价 ACF 呈现 4 阶拖尾性，交易量 ACF 呈现 2 阶拖尾性；以太坊收盘价 ACF 呈现 5 阶拖尾性，交易量 ACF 呈现 3 阶拖尾性。

2. PACF

图表 13 比特币收盘价 PACF



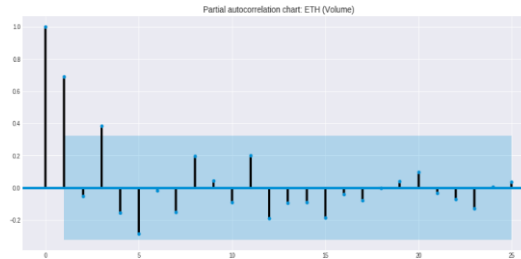
图表 14 比特币交易量 PACF



图表 15 以太坊收盘价 PACF



图表 16 以太坊交易量 PACF



比特币收盘价 PACF 呈现 2 阶截尾性，交易量 PACF 呈现 2 阶截尾性；以太坊收盘价 PACF 呈现 2 阶截尾性，交易量 PACF 呈现 2 阶拖尾性。

(四) SARIMA 模型建模与预测

SARIMA 是季节性差分自回归滑动平均模型。本文处理后的序列适用该模型进行建模⁴。为进一步确定阶数。通过使用迭代的方式判断是否是最佳 AIC 值进行定阶，具体定阶代码过程如下：

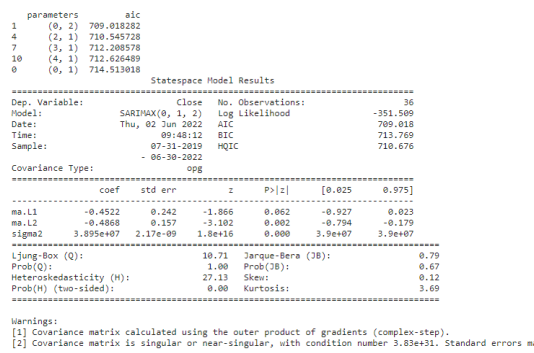
图表 17 定阶代码

```
best_aic = float("inf")
warnings.filterwarnings('ignore')
for param in parameters_list:
    try:
        model = SARIMAX(btc_month.Close, order=(param[0], d, param[1])).fit(dispatch=-1)
    except ValueError:
        print('bad parameter combination:', param)
        continue
    aic = model.aic
    if aic < best_aic:
        best_model = model
        best_aic = aic
        best_param = param
results.append([param, model.aic])
```

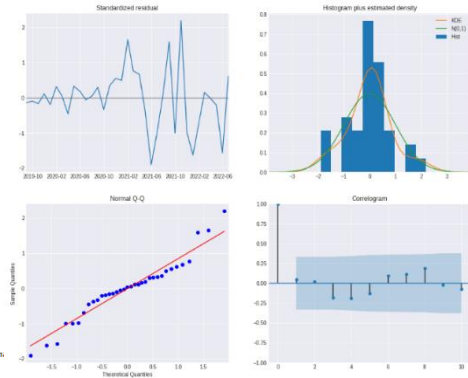
1. 比特币建模与预测

比特币收盘价模型为：SARIMA(0,1,2)。模型验证采用 plot_diagnostics 可视化模型诊断图的方式进行模型验证，以确保模型所做的假设合理。诊断图结果如下所示：

图表 18 比特币收盘价定阶结果



图表 19 比特币收盘价模型诊断图



右上角图 KDE 曲线可知，残差正态分布，通过左下角 QQ 图残差（蓝色点）的有序分布遵循标准正态分布的红色直线，同样表明了残差的正态性，右下角 ACF 图 1 阶截尾表明时序残差与自身得知后值具有较低相关性。综上，模型的残差不相关且具有正态性，故模型拟合有效，可以用于预测。预测结果如下：

图表 20 比特币月收盘价预测



2. 以太坊建模与预测

以太坊收盘价模型为：SARIMA(0,1,4)。模型验证同样的方法。右上角图 KDE 曲线可知，残差近似正态分布，通过左下角 QQ 图残差同样表明了残差的近似正态性，右下角 ACF 图 1 阶截尾表明时序残差与自身得知后值具有较低相关性。综上，模型的残差不相关且近似正态性，故模型可以用于预测。

图表 21 以太坊收盘价定阶结果

```
parameters      aic
2 (0, 4) 519.403082
1 (0, 2) 520.986483
0 (0, 1) 521.710282
4 (1, 4) 521.891008
9 (3, 0) 522.174780

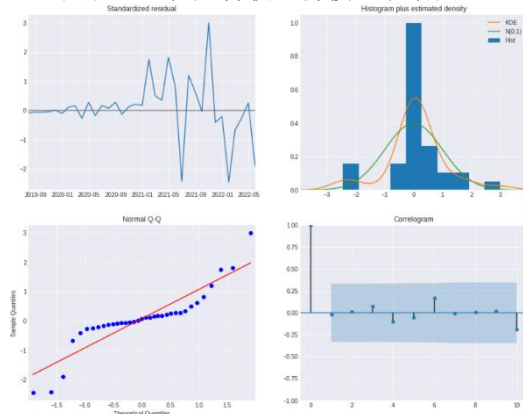
Statespace Model Results
=====
Dep. Variable:      Close  No. Observations:  36
Model:              SARIMAX(0, 1, 4)  Log Likelihood: -254.792
Date:               Thu, 02 Jun 2022  AIC: 519.403
Time:               12:47:22  BIC: 527.321
Sample:             06-30-2019  HQIC: 522.167
                   - 05-31-2022

Covariance Type:    opg
=====
              coef  std err      z      P>|z|  [0.025  0.975]
-----
ma.L1      0.6421    5.314    0.121    0.904   -9.774   11.058
ma.L2     -0.3358    0.544   -0.617    0.537   -1.402    0.731
ma.L3     -0.2084    3.944   -0.053    0.958   -7.938    7.521
ma.L4      0.5065    3.226    0.157    0.875   -5.817    6.830
sigma2     1.064e+05  6.74e+05  0.158    0.875  -1.21e+06  1.43e+06

Ljung-Box (Q):          9.58  Jarque-Bera (JB):          8.58
Prob(Q):                1.00  Prob(JB):              0.01
Heteroskedasticity (H): 97.22  Skew:              -0.09
Prob(H) (two-sided):    0.00  Kurtosis:          5.42

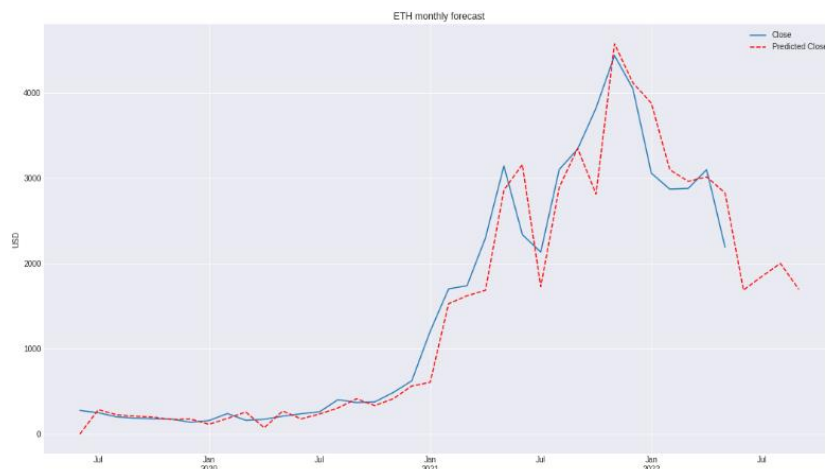
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

图表 22 以太坊收盘价模型诊断图



该模型预测结果如下图所示：

图表 23 以太坊收盘价预测结果



三、多元时间序列协整分析

在第一节第二部分的原始规范化后对数交易量和对视收盘价可视化时序中，可以猜测以太坊和比特币可能具有一定相关性，在这一节中，将检验这两个序列具有协整关系的原假设。

(一) 收盘价协整关系检验

使用 R 对于对数收盘价和一阶差分对数收盘价进行单位根检验 (在上述过程中已知有常数项，故参数 `type` 为 `c('c')`)，结果显示对数收盘价 P-值分别为：0.6253 和 0.7844 大于 0.05，不能拒绝原假设，认为序列具有单位根是非平稳序列，继而分别进行一阶差分对数收盘价两序列单位根检验中 P-值均为 0.01 小于 0.05，故认为收盘价的两个序列不存在单位根，是平稳序列，即两序列均满足一阶单整，因此可以进行协整关系检验。

图表 24 比特币对数收盘价单位根检验

```
> adfTest(btc.close,type = 'c')

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -1.1648
P VALUE:
0.6253

Description:
Fri Jun 03 09:45:23 2022 by user: Tracy Tao

> adfTest(diff(btc.close),type = 'c') # 比特币收盘价一阶单整

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -23.5752
P VALUE:
0.01

Description:
Fri Jun 03 09:45:34 2022 by user: Tracy Tao

warning message:
In adfTest(diff(btc.close), type = "c") :
  p-value smaller than printed p-value
```

图表 25 以太坊对数收盘价单位根检验

```
> adfTest(eth.close,type = 'c')

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -0.7392
P VALUE:
0.7844

Description:
Fri Jun 03 09:46:17 2022 by user: Tracy Tao

> adfTest(diff(eth.close),type = 'c') # 以太坊收盘价一阶单整

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -23.0105
P VALUE:
0.01

Description:
Fri Jun 03 09:46:20 2022 by user: Tracy Tao

warning message:
In adfTest(diff(eth.close), type = "c") :
  p-value smaller than printed p-value
```

使用利用最大特征值似然比检验方法进行协整检验，统计量为：

$L_{max}(m_0) = -(T - kp) \sum_{i=m_0+1}^k \ln(1 - \lambda_{m_0+1})$ 。使用 MTS.VARorder 进行 VAR 模型定阶，AIC 准则下阶数为 8，原假设 $H_0: m = 0$ ，备择假设 $H_a: m = 1$ 。

图表 26 对数收盘价最大特征值协整检验

```
> library(urca)
> m2=ca.jo(bnd,k=8, ecdet=c("const")) ##式 (5.80) #基于最大特征值的检验
> summary(m2)

#####
# Johansen-Procedure #
#####

Test type: maximal eigenvalue statistic (lambda max) , without linear trend
and constant in cointegration

Eigenvalues (lambda):
[1] 1.217131e-02 2.336185e-03 -4.297933e-19

Values of teststatistic and critical values of test:

      test 10pct  5pct  1pct
r <= 1 |   2.55   7.52   9.24 12.97
r = 0  |  13.34  13.75  15.67 20.20

Eigenvectors, normalised to first column:
(These are the cointegration relations)

      btc.close.l8 eth.close.l8 constant
btc.close.l8   1.0000000   1.0000000  1.000000
eth.close.l8  -0.6043351   0.7726981 -1.043278
constant      -5.8741039 -16.3140807 -3.705471

weights w:
(This is the loading matrix)

      btc.close.l8 eth.close.l8 constant
btc.close.d  0.005219204 -0.0008558612 -6.365191e-17
eth.close.d  0.023404775 -0.0007711647 -3.052553e-16
```

在原假设成立的条件下计算出检验统计量的值为 13.34，均小于置信水平下的临界值不拒绝原假设，认为对数收盘价之间不存在协整关系。

(二) 交易量协整关系检验

通过两币对数收盘价和以太坊一阶差分对数收盘价进行单位根检验 (参数 type 未变动)，结果显示比特币对数收盘价 P-值均为：0.01 小于 0.05 拒绝原假设，故认为对数交易量的两个序列不存在单位根，是平稳序列，能够在一阶单整前提下进行协整关系检验。

图表 27 对数交易量单位根检验

```
> adfTest(btc.volume,type = 'c')

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -9.8059
P VALUE:
0.01

Description:
Fri Jun 03 10:09:48 2022 by user: Tracy Tao

Warning message:
In adfTest(btc.volume, type = "c") : p-value smaller than printed p-value
> adfTest(eth.volume,type = 'c')

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -5.9395
P VALUE:
0.01

Description:
Fri Jun 03 10:09:51 2022 by user: Tracy Tao

Warning message:
In adfTest(eth.volume, type = "c") : p-value smaller than printed p-value
```

使用同样的方法进行 VAR 模型定阶，AIC 准则下阶数为 13，统计量、原假设和备择假设不变。结果显示：

图表 28 对数交易量最大特征值协整检验

```
#####
# Johansen-Procedure #
#####

Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegration

Eigenvalues (lambda):
[1] 1.282564e-02 1.917769e-03 7.430270e-19

values of teststatistic and critical values of test:

      test 10pct  5pct  1pct
r <= 1 |   2.08  7.52  9.24 12.97
r = 0  |  13.99 13.75 15.67 20.20

Eigenvectors, normalised to first column:
(These are the cointegration relations)

      btc.close.l13  eth.close.l13  constant
btc.close.l13      1.0000000      1.0000000  1.0000000
eth.close.l13     -0.6081418      0.3916174 -1.135355
constant         -5.8501498     -13.5149387 -3.176629

weights W:
(This is the loading matrix)

      btc.close.l13  eth.close.l13  constant
btc.close.d      0.006309681 -0.0010271531 -1.767520e-16
eth.close.d      0.025384287 -0.0009010686 -7.248077e-16
```

在原假设成立的条件下计算出检验统计量的值为 13.99，大于传统 10%水平下的临界值 13.75 拒绝原假设。然后继续对备择假设进行检验，此时得到检验统计量的值为 2.08，均小于临界值，故接受备择假设，认为对数交易量之间存在一个协整关系。得到协整向量，构造协整序列：

$$w_t = Const + \beta_1 BX_t = Const + \beta_1 X_{t-1}$$

对其进行单位根检验 P 值 0.01 小与 0.05 说明协整序列平稳，对数交易量两序列具有长期均衡关系。

综上，对数收盘价的两序列不存在协整关系，对数交易量两序列具有长期稳定关系，因此两币需求量长期均衡，存在通过两币投资组合进行套利的可能性。

(三) 对数交易量误差修正模型估计

通过上述分析已知对数交易量序列存在协整关系，因此在此基础上可以创建误差修正项，将误差修正项看作一个解释变量，连同其它反映短期波动的解释变量一起建立短期模型⁵。

先对收盘价两序列使用 python statsmodel.formula.api.ols 最小二乘回归，即建立误差修正模型： $Volume_{BTC} = Volume_{ETH} * \beta_1 + \beta_0$ ，回归结果如下：校正决定系数 (Adj. R-squared) 为 0.591，说明模型中 $Volume_{ETH}$ 能解释 $Volume_{BTC}$ 59.10% 的变异程度，模型为 $Volume_{BTC} = Volume_{ETH} * 1.3889 + 1.01e10$ 。

进而通过拟合后模型计算残差，加入残差项进行最小二乘回归拟合 ECM 模型， $e_t = Volume_{BTC} - Volume_{ETH} * \beta_1 + \beta_0$ ，ECM 拟合结果如下：

最终模型形式为： $Volume_{BTC} = Volume_{ETH} * \beta_1 + \beta_0 + e_t$ 。

图表 29 误差修正模型拟合结果

```

=====
OLS Regression Results
=====
Dep. Variable:      A      R-squared:      0.591
Model:              OLS    Adj. R-squared:  0.591
Method:             Least Squares    F-statistic: 1582.
Date:               Thu, 02 Jun 2022  Prob (F-statistic): 1.20e-214
Time:               13:45:13    Log-Likelihood: -26993.
No. Observations:   1095    AIC: 5.399e+04
Df Residuals:       1093    BIC: 5.400e+04
Df Model:           1
Covariance Type:    nonrobust
=====
               coef      std err      t      P>|t|      [0.025      0.975]
-----
Intercept      1.01e+10      7.2e+08      14.018      0.000      8.68e+09      1.15e+10
B               1.3889      0.035      39.773      0.000      1.320      1.457
=====
Omnibus:                2056.641    Durbin-Watson:      1.220
Prob(Omnibus):           0.000    Jarque-Bera (JB):    4413509.878
Skew:                    13.126    Prob(JB):             0.00
Kurtosis:                312.912    Cond. No.             4.00e+10
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4e+10. This might indicate that there are
strong multicollinearity or other numerical problems.
Intercept      1.009776e+10
B               1.388910e+00
dtype: float64

```

图表 30 ECM 模型结果

```

=====
OLS Regression Results
=====
Dep. Variable:      A      R-squared:      1.000
Model:              OLS    Adj. R-squared:  1.000
Method:             Least Squares    F-statistic: 1.004e+33
Date:               Thu, 02 Jun 2022  Prob (F-statistic): 0.00
Time:               13:46:19    Log-Likelihood: 10670.
No. Observations:   1095    AIC: -2.133e+04
Df Residuals:       1092    BIC: -2.132e+04
Df Model:           2
Covariance Type:    nonrobust
=====
               coef      std err      t      P>|t|      [0.025      0.975]
-----
Intercept      1.01e+10      8.31e-07      1.21e+16      0.000      1.01e+10      1.01e+10
B               1.3889      4.03e-17      3.45e+16      0.000      1.389      1.389
residual       1.0000      3.49e-17      2.86e+16      0.000      1.000      1.000
=====
Omnibus:                1052.725    Durbin-Watson:      0.144
Prob(Omnibus):           0.000    Jarque-Bera (JB):    147953.007
Skew:                    3.982    Prob(JB):             0.00
Kurtosis:                59.386    Cond. No.             4.00e+10
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4e+10. This might indicate that there are
strong multicollinearity or other numerical problems.
Intercept      1.009776e+10
B               1.388910e+00
residual       1.000000e+00
dtype: float64

```

继而进一步使用 R 进行 ECM-VAR(13)模型估计。

图表 31 ECM-VAR(13)

```

> m6 <- ECMvar1(diff(bnd2),13,diff(wt))
alpha:
      btc.volume eth.volume
[1,]      5.82      4.64
standard error
      [,1] [,2]
[1,] 0.395 0.389
AR coefficient matrix
AR( 1 )-matrix
      btc.volume eth.volume
btc.volume      1.95      0.164
eth.volume      2.58     -1.094
standard error
      [,1] [,2]
[1,] 0.234 0.0472
[2,] 0.230 0.0465
AR( 2 )-matrix
      btc.volume eth.volume
btc.volume      1.54      0.155
eth.volume      2.31     -1.202
standard error
      [,1] [,2]
[1,] 0.227 0.0740
[2,] 0.224 0.0729
AR( 3 )-matrix
      btc.volume eth.volume
btc.volume      1.20      0.217
eth.volume      2.04     -1.189
standard error
      [,1] [,2]
[1,] 0.220 0.0956
[2,] 0.217 0.0942
AR( 4 )-matrix
      btc.volume eth.volume
btc.volume      0.916      0.327
eth.volume      1.747     -1.052
standard error
      [,1] [,2]
[1,] 0.212 0.113
[2,] 0.209 0.111
AR( 5 )-matrix
      btc.volume eth.volume
btc.volume      0.673      0.353
eth.volume      1.445     -0.894
standard error
      [,1] [,2]
[1,] 0.202 0.125
[2,] 0.199 0.123
AR( 6 )-matrix
      btc.volume eth.volume
btc.volume      0.605      0.299
eth.volume      1.298     -0.808
standard error
      [,1] [,2]
[1,] 0.190 0.131
[2,] 0.187 0.129

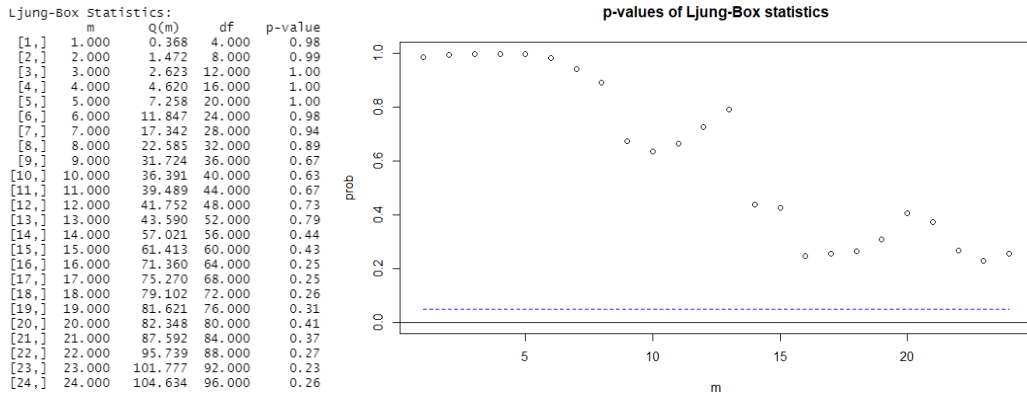
AR( 7 )-matrix
      btc.volume eth.volume
btc.volume      0.609      0.345
eth.volume      1.156     -0.570
standard error
      [,1] [,2]
[1,] 0.175 0.131
[2,] 0.172 0.129
AR( 8 )-matrix
      btc.volume eth.volume
btc.volume      0.528      0.371
eth.volume      0.963     -0.363
standard error
      [,1] [,2]
[1,] 0.156 0.125
[2,] 0.153 0.124
AR( 9 )-matrix
      btc.volume eth.volume
btc.volume      0.515      0.242
eth.volume      0.818     -0.303
standard error
      [,1] [,2]
[1,] 0.134 0.113
[2,] 0.132 0.112
AR( 10 )-matrix
      btc.volume eth.volume
btc.volume      0.413      0.213
eth.volume      0.607     -0.169
standard error
      [,1] [,2]
[1,] 0.109 0.0960
[2,] 0.107 0.0945
AR( 11 )-matrix
      btc.volume eth.volume
btc.volume      0.268      0.161
eth.volume      0.390     -0.067
standard error
      [,1] [,2]
[1,] 0.0793 0.0744
[2,] 0.0781 0.0732
AR( 12 )-matrix
      btc.volume eth.volume
btc.volume      0.134      0.0418
eth.volume      0.171     -0.0365
standard error
      [,1] [,2]
[1,] 0.0460 0.0475
[2,] 0.0454 0.0468
-----
Residuals cov-mtx:
      btc.volume eth.volume
btc.volume 0.04186017 0.03125644
eth.volume 0.03125644 0.04060158

det(sse) = 0.0007226243
AIC = -7.14138
BIC = -6.913304

```

对该模型使用 MTSdiag 函数进行模型检验，Ljung-Box 统计量 P 值统计上均显著，因此模型拟合较好。并且所有参数在 5%水平下均统计上显著，不需要进行模型简化，因此 ECM-VAR(13)能反映对数交易量序列之间的短期波动关系。

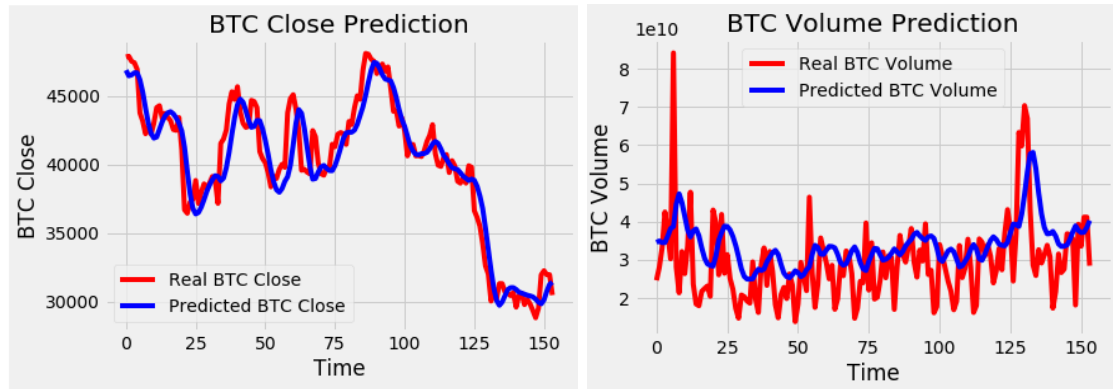
图表 32 ECM-VAR 模型的残差序列检验结果



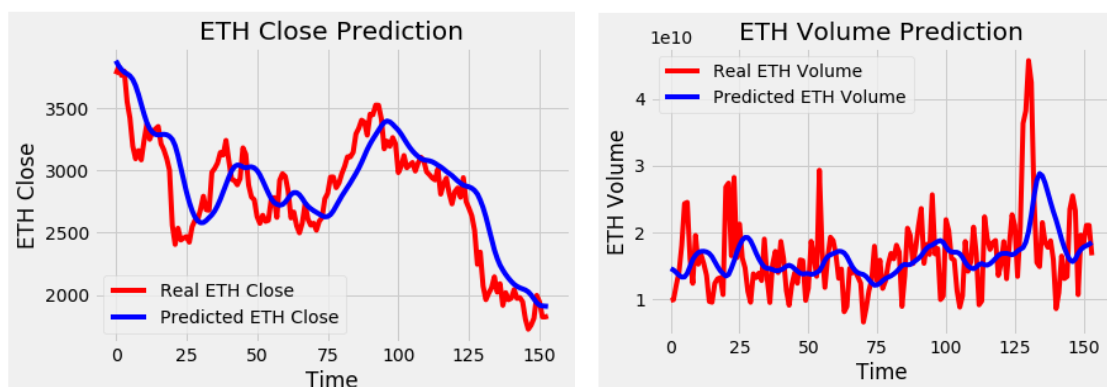
四、深度学习模型：LSTM

面对时间序列的预测问题，可以用本文使用的传统 SARIMA 模型，也可以深度学习模型预测，本文使用长短期记忆网络 (LSTM)。一般深度学习模型比传统时序模型准确度高⁶。LSTM 对比特币和以太坊收盘价、交易量预测结果如下：

图表 33 比特币 LSTM 预测结果



图表 34 以太坊 LSTM 预测



在对于收盘价预测时准确度较高，而对于交易量预测偏差较大。由于上文所述，通过协整分析认为两币交易量存在长期均衡关系，则可以将 LSTM 和 SARIMA 模型结合分别预测收盘价和交易量。

五、总结

本文首先对两币原始时序数据进行规范化取对数，对交易量和收盘价两变量可视化，从时序图直观可知时序剧烈波动认为其为非平稳序列。由于原始数据为每日数据，多次差分后依旧非平稳，难以发现时序趋势，故对原始数据进行月度平均重采样，发现时序具有季节性，建立 SARIMA 模型进行预测。从对数交易量和收盘价时序图猜测，时间可能具有相关性，继而进行多元时序分析，验证猜想。对收盘价和交易量进行协整检验，得出交易量存在协整关系，故两币需求量具有长期均衡，由此对交易量建立 ECM-VAR 模型。最后使用深度学习方法的 LSTM 模型分别对交易量收盘价进行预测，得出：LSTM 模型针对收盘价预测效果较好。综上，对于两币时序可将 LSTM 和 SARIMA 模型结合分别预测收盘价和交易量。

六、参考文献

-
- ¹ 知乎. 全球加密货币市值前十位介绍[DB/OL].(2021-02-17)[2022.06.03].
<https://zhuanlan.zhihu.com/p/351078487>.
- ² 腾讯新闻.从元宇宙看数字经济[EB/OL].(2021-10-01)[2022.06.02].
<https://new.qq.com/omn/20211001/20211001A0690O00.html>.
- ³ 小李财经.为什么说加密货币是实现元宇宙的关键? [DB/OL].(2021-11-12)[2022.06.02]. <https://zhuanlan.zhihu.com/p/432429667>.
- ⁴ 王莹,韩宝明,张琦,李得伟.基于 SARIMA 模型的北京地铁进站客流量预测[J].交通运输系统工程与信息,2015,15(06):205-211.DOI:10.16097/j.cnki.1009-6744.2015.06.031.
- ⁵ 陈守东,韩广哲,荆伟.主要股票市场指数与我国股票市场指数间的协整分析[J].数量经济技术经济研究,2003(05):124-129.
- ⁶ 王鑫,吴际,刘超,杨海燕,杜艳丽,牛文生.基于 LSTM 循环神经网络的故障时间序列预测[J].北京航空航天大学学报,2018,44(04):772-784.DOI:10.13700/j.bh.1001-5965.2017.0285.