

SHANGHAI UNIVERSITY

课程设计（论文）

**UNDERGRADUATE COURSE (THESIS)**

**题目:国际天然气价格的分析和预测**

|  |  |
| --- | --- |
| **学院** | **计算机工程与科学学院** |
| **专业** | **计算机科学与技术** |
| **学号** |  |
| **学生姓名** |  |
| **指导教师** | **武星** |
| **起讫日期** | **2022.03.25 – 2022.06.05** |

目录

[摘要 II](#_Toc8411)

[ABSTRACT II](#_Toc8927)

[第1章 绪论 1](#_Toc29372)

[§1.1 国际天然气价格分析与预测的背景及意义 1](#_Toc14614)

[§1.2 本文研究内容及目标 1](#_Toc23846)

[§1.2.1 研究内容 1](#_Toc210)

[（1） 国际天然气的变化及趋势； 1](#_Toc18085)

[（2） 国际天然气价格变化带来的影响； 1](#_Toc342)

[（3） 造成国际天然气价格上涨的因素； 1](#_Toc4280)

[§1.2.2 研究目标 1](#_Toc14430)

[（1） 绘制国际天然气变化曲线； 1](#_Toc17277)

[（2） 分析造成国际天然气变化的影响因素； 1](#_Toc20890)

[（3） 分析天然气价格上涨会带来的影响； 1](#_Toc14274)

[（4） 对未来天然气价格进行预测； 2](#_Toc14078)

[§1.3 本文组织结构 2](#_Toc15010)

[第2章 国际天然气价格的数据集的采集和构建 3](#_Toc10335)

[§2.1 数据集的采集与构建 3](#_Toc16155)

[第3章 国际天然气价格数据集的预处理 4](#_Toc15678)

[§3.1 提取数据集标签及特征 4](#_Toc5438)

[§3.2 处理重复值 4](#_Toc2484)

[§3.3 缺失值处理 5](#_Toc7139)

[§3.4 异常值处理以及提取每日平均价格 5](#_Toc10192)

[第4章 数据分析及可视化 6](#_Toc18702)

[§4.1 数据的总体分析 6](#_Toc12529)

[§4.2 研究天然气每日涨/跌幅对天然气价格变化的关系 7](#_Toc1638)

[§4.3 天然气每日峰谷值变化幅度与平均价格的关系研究 8](#_Toc20189)

[§4.4 对两个异常值区间数据的分析 10](#_Toc2664)

[§4.4.1 对2005年8月29日至2005年12月30日异常值区间的分析 10](#_Toc2725)

[§4.4.2 对2008年4月28日-2008年7月16日异常值区间的分析 10](#_Toc23618)

[§4.4.3 两组异常区间的整体分析 11](#_Toc28569)

[§4.5 数据分析结论整合 11](#_Toc24444)

[第5章 对当下天然气价格的分析与预测 13](#_Toc2768)

[§5.1 对当下天然气价格的分析 13](#_Toc28109)

[§5.2 对未来天然气价格的预测 13](#_Toc21530)

[第6章 总结与展望 15](#_Toc30938)

国际天然气价格的分析和预测

# 摘要

自从俄乌战争以来，时常能够看到有关欧洲地区天然气价格飙升的新闻。而这也对国际天然气价格产生了深远的影响，国际天然气市场波动剧烈，对各个国家产生了影响。在中国，部分地区的天然气也迎来了涨价，粗略判断，在未来，天然气价格还将持续上涨。为了了解天然气价格的变化趋势以及影响因素，本篇文章将对2000年1月4日至2022年5月9日的国际天然气价格变化进行分析，试图从历史天然气价格变化中获取重要信息，分析造成国际天然气价格变化的影响因素，并预测今后的国际天然气价格变化趋势。

关键词：国际，天然气，价格，涨幅，单日变化，数据分析，预测

Analysis and forecast of international natural gas prices

# ABSTRACT

Since the russia-Ukraine war, there have been frequent headlines about soaring gas prices in Europe. And this also has a profound impact on the international gas price, the volatility of the international gas market, has an impact on each country. In China, natural gas prices have also risen in some parts of the country, and crude judgment indicates that natural gas prices will continue to rise in the future. In order to understand the changing trends of natural gas prices as well as the influence factors, this article will be on January 4, 2000 to May 9, 2022 international gas prices change were analyzed, and tried to obtain important information from the history of natural gas price changes, the analysis of the influence factors of the international natural gas price changes, and predict the international natural gas price change trend in the future.

**Keywords:** International, natural gas, price, rise, one-day change, data analysis, forecast

# 绪论

本章主要描述了进行国际天然气价格预测项目的背景及意义，进而提出了本文所要研究的内容及目标。

## 国际天然气价格分析与预测的背景及意义

天然气是一种清洁、低碳、高效的新能源，日益在人们生活中扮演着重要的角色，它具有传统化石能源不具有的众多优点。近年来，随着我国天然气需求的持续增加，天然气储量的不断增加，家家户户都用上了天然气，也成为了我们居民日常生活中不可缺少的一部分。

因此，天然气价格的变动不仅与我们的日常生活息息相关。能够影响到我们的生活质量，更是国家能源安全的重要一环。

同时，由于俄乌战争的爆发以及新冠疫情的影响，全球能源价格持续高涨，天然气更是不断冲击历史高位，一度突破十四年来的最高水平。在欧洲，由于俄罗斯的制裁，天然气价格更是一度突破1000%大关。可能国内目前短期来看，感觉并价格波动不明显，但在今后，随着各种调整，天然气价格仍有可能发生变动。 于是，为了更好地了解国际天然气价格的变化，更好地应对未来可能发生的天然气价格增长，对国际天然气价格进行分析与预测是一件很有必要的事情。

## 本文研究内容及目标

本文通过对国际天然气历史价格进行分析，试图找出造成国际天然气价格波动的影响因素，并对天然气价格造成的影响进行分析，最后对未来国际天然气价格的变化情况进行预测。

### 研究内容

本文具体研究内容有以下几个方面。

1. 国际天然气的变化及趋势；
2. 国际天然气价格变化带来的影响；
3. 造成国际天然气价格上涨的因素；

### 研究目标

针对本文的研究内容，制定了以下几项指标：

1. 绘制国际天然气变化曲线；
2. 分析造成国际天然气变化的影响因素；
3. 分析天然气价格上涨会带来的影响；
4. 对未来天然气价格进行预测；

## 本文组织结构

整篇论文分为五章。

第一章介绍了研究背景、研究意义，并提出了本文的研究内容以及研究目标。

第二章主要介绍了国际天然气价格的数据集的采集和构建

第三章对数据的预处理进行了详细的说明

第四章通过对数据进行分析以及可视化，获得各种结论。

第五章未来的国际天然气价格的变化进行了预测。

# 国际天然气价格的数据集的采集和构建

## 数据集的采集与构建

本文所获取的数据为结构化数据，数据来源于（https://www.kaggle.com）网站，该数据集记载了2000年1月4日至2022年5月9日的每日天然气历史数据，具有较高的准确性与可信度。

本次进行大数据分析的工具是Jupyter，采用python语言进行分析，运用了pandas、numpy、matplotlib、ARIMA等库。

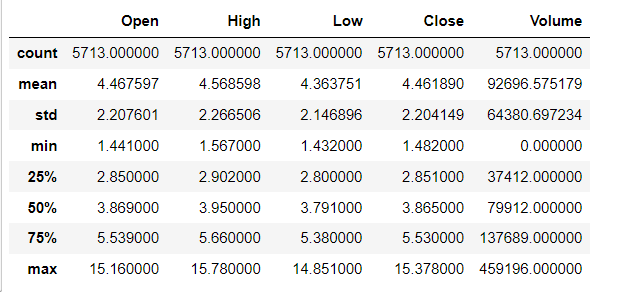
通过导入csv类型的数据集并查看数据大致信息，我们能够发现，该数据集共有5713行，7列的数据（详见图2-1），数据量较大，具有较高的可信度。

图2-1 数据集信息

其中，Open代表每日天然气开盘价格，High代表当日最高价格，Low代表当日最低价格，Close代表收盘价格，Volume代表成交量，还有一列表示日期（格式：2020-05-09）为图中并未显示。

# 国际天然气价格数据集的预处理

本章全面阐述了处理本数据集的方法。

## 提取数据集标签及特征

由于该数据集中包含了7类内容，因此，我们需要提取这7种类别的信息，将其转化为列表的形式，便于后续进行数据分析（图3-1）。

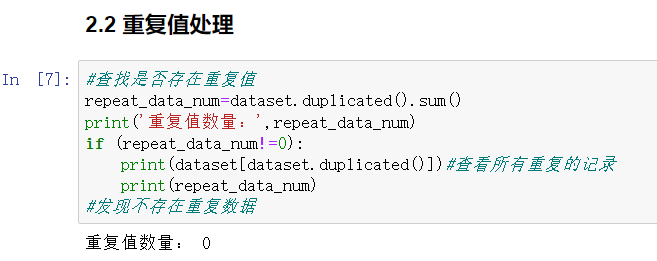


图3-1 提取标签与特征操作

## 处理重复值

由于在数据集中可能存在部分重复的数据，会对后续的分析以及预测产生影响，需要在预处理时将重复元素删去。

该数据集在处理重复元素时，发现重复元素数量为0（图3-2），具有较好的数字特征，因此不用执行任何操作。

图3-2 重复值处理结果

## 缺失值处理

在数据集中，可能存在统一行中，部分信息不存在（为NULL），而后续进行模型训练以及可视化分析时，要求数据不能存在未知元素，因此需要删去缺失值或是通过某些方式（用前后的平均值进行替代）将缺失值补全。

但该数据集进行缺失值处理时，所有信息皆为可知，不存在未知元素（图3-3），数据较好，因此也不用进行任何处理。

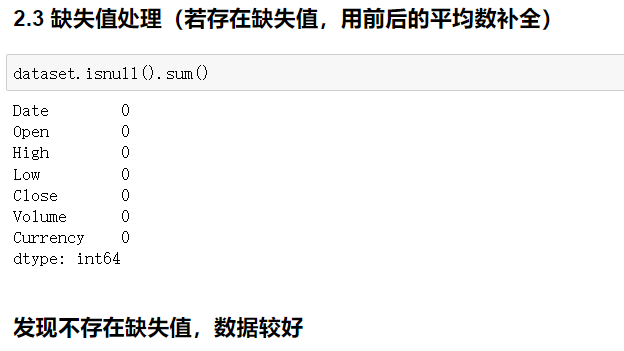


图3-3 缺失值处理

## 异常值处理以及提取每日平均价格

由于后续对价格进行可视化后发现，当日的平均价格、最高价格和最低价格整体来看并没有太大的区别，趋势近乎一致，因此，能够采用当日的平均价格来代替当日天然气的价格。异常值的判断也是基于平均价格。

本次研究利用了3α原理来进行异常值的判断。总计获取到133个异常值，平均价格为12.5，最大值为15.3，最小值为11.13，方差约为1，呈现左右对称的样子。而由于异常值中包含了许多重要的信息，因此，在实际操作过程中并没有将其直接删去。我发现，所有的异常值有着明显的分布特点，集中于两大区间内（2005年8月29日-2005年12月30日以及2008年4月28日-2008年7月16日）。后续将针对这两个数据域进行重点分析。

# 数据分析及可视化

本章是全文的重点，主要对数据进行了可视化与分析。

## 数据的总体分析

首先，我们需要对数据集的数据进行总体的分析，通过分析可以看到，自2000年以来，国际天然气的平均价格为4.67，标准差为2.2，历史最高价格为15.316，位于2005年12月13日，最低价格为1.5，位于2020年6月26日。

通过可视化绘制直方图，我们可以看到，大部分的天然气平均价格基本都位于2-6区间内。

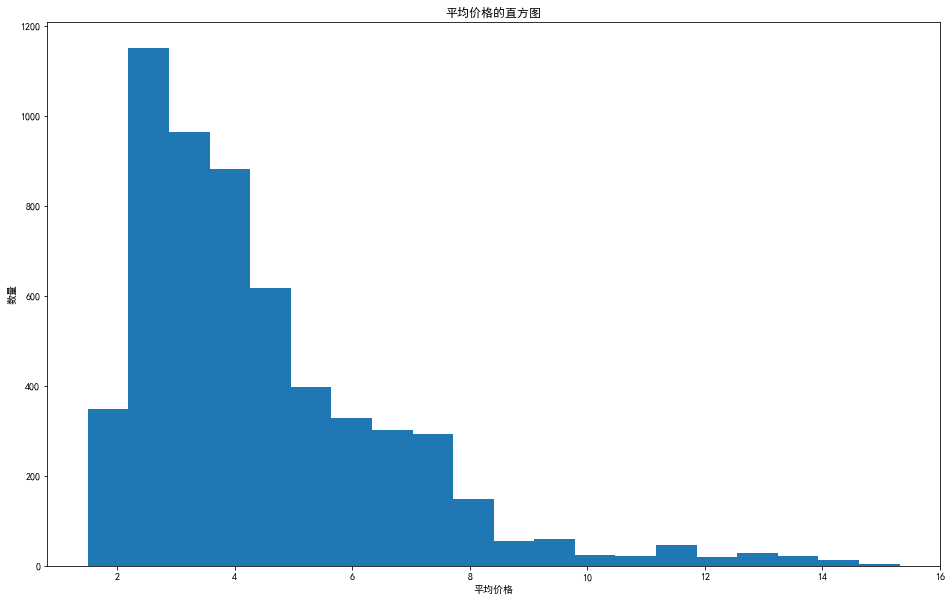


图4-1 平均价格直方图

通过绘制每日最高、最低、平均价格的折线图（图4-2）我们可以看到，每日最高、最低、平均天然气价格的变化趋势基本一致，因此，在接下来的分析中，我将用每日天然气的平均价格代替天然气的价格进行后续的分析。

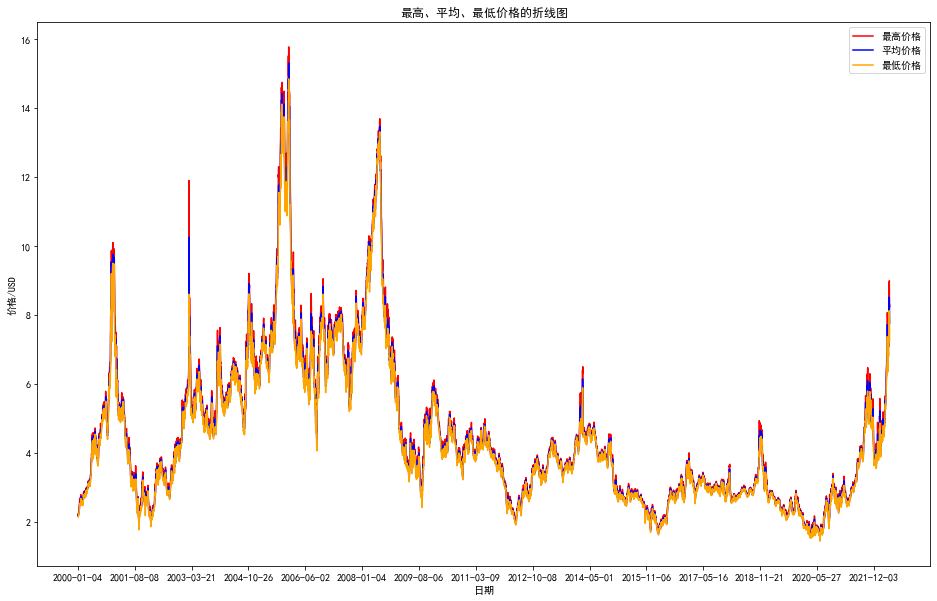


图4-2 天然气每日最高、平均、最低价格的折线图

因此，2000年1月4日-2020年5月9日的天然气平均价格的柱状图如下（图4-3）：

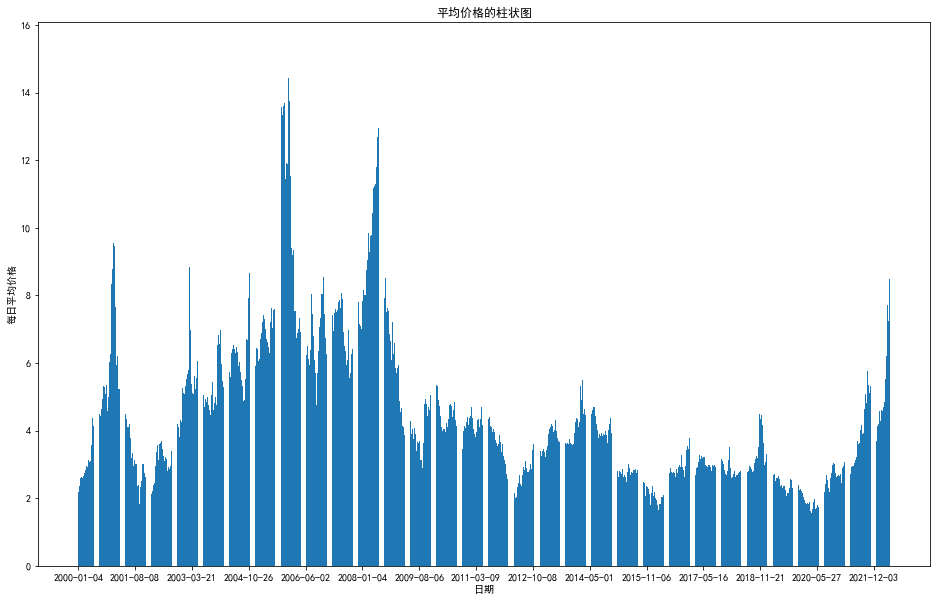


图4-3 天然气平均价格的柱状图

## 研究天然气每日涨/跌幅对天然气价格变化的关系

涨幅=(今日收盘价-昨日收盘价)/昨日收盘价\*100%,如果计算结果为负数,则为跌幅，这是天然气价格变化的重要指标之一。通过分析发现，天然气的最大涨幅为38.31%，位于2003年2月24日，而最大跌幅为19.06%，位于2004年3月18日。但通过与平均价格的柱状图进行比较后，并没有发展有相关联的地方，因此需要对其进行进一步的分析。

通过对涨/跌幅的可视化（图4-4）可以发现，天然气价格的单日涨跌幅与价格之间并不存在明显的关系，而是始终维持在0附近，整体较为平稳，有波动，但基本符合市场的变化规律。

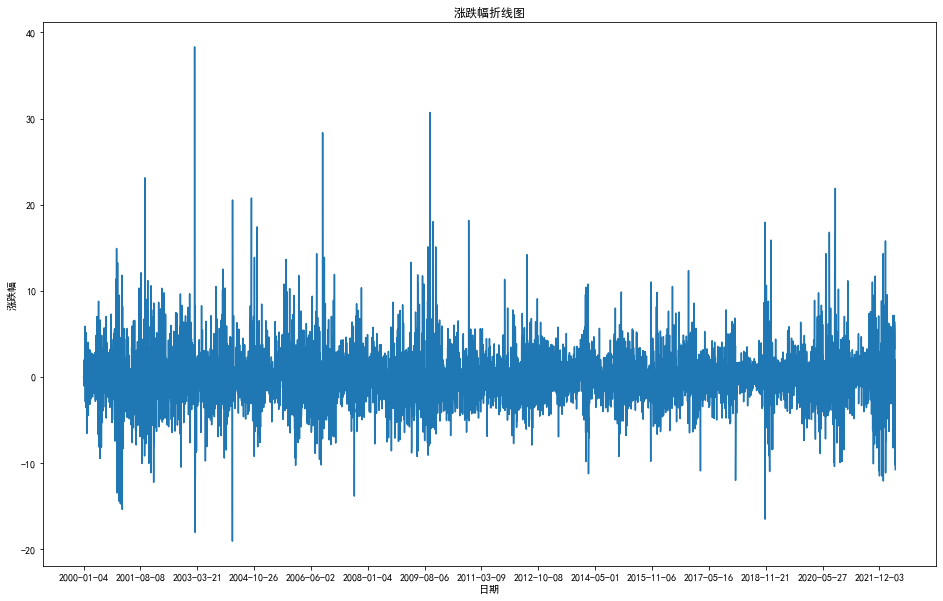


图4-4 天然气每日涨幅的折线图

## 天然气每日峰谷值变化幅度与平均价格的关系研究

除了天然气每日的涨/跌幅，天然气每日峰谷值之间的变化幅度也具有非常重要的意义，它代表了天然气价格的每日波动程度，一定程度上代表了天然气市场的不稳定程度。

通过研究发现，在该数据集中，单日变化程度最大为3.29，位于2003年2月25日（该日期非常重要，后续统一分析），平均单日变化幅度为0.204。经过可视化（图4-5）可以发现，单日天然气价格变化程度与天然气平均价格之间具有明显的相关性，曲线基本一致，需要我们进行更深一步的分析。

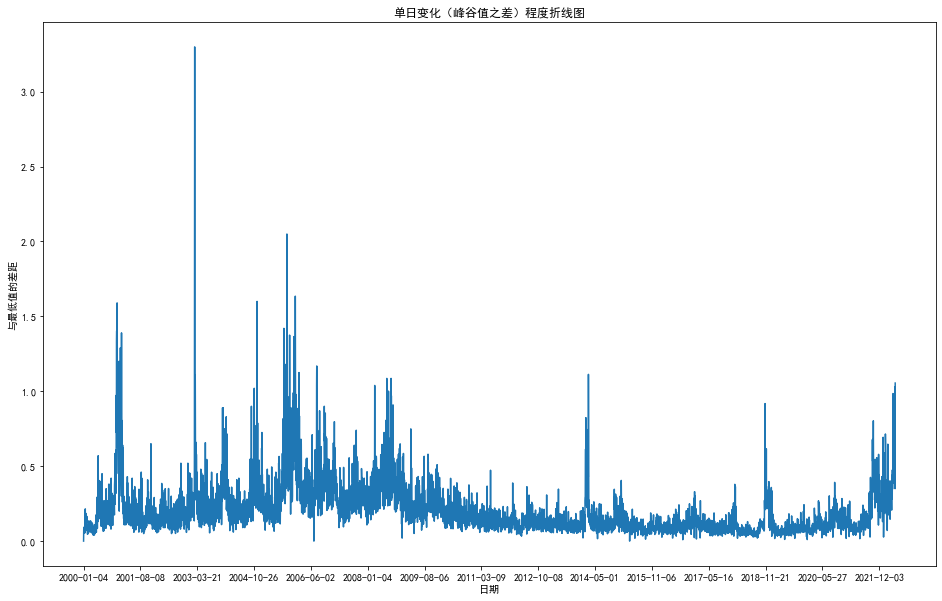


图4-5 单日变化（峰谷值之差）程度折线图

因此，我绘制了价格与单日变化程度的散点图，并对其进行拟合（图4-6）。可以发现，单日变化程度与天然气平均价格之间的确存在着线性关系，拟合的直线的方程为y=0.0543x-0.0375。

由此可以得出一个结论：国际天然气的平均价格越高，天然气价格的单日变化程度越大，天然气市场波动越大，购买的风险也越大。每日价格变化程度的趋势，在一定程度上预示着天然气价格的变化。

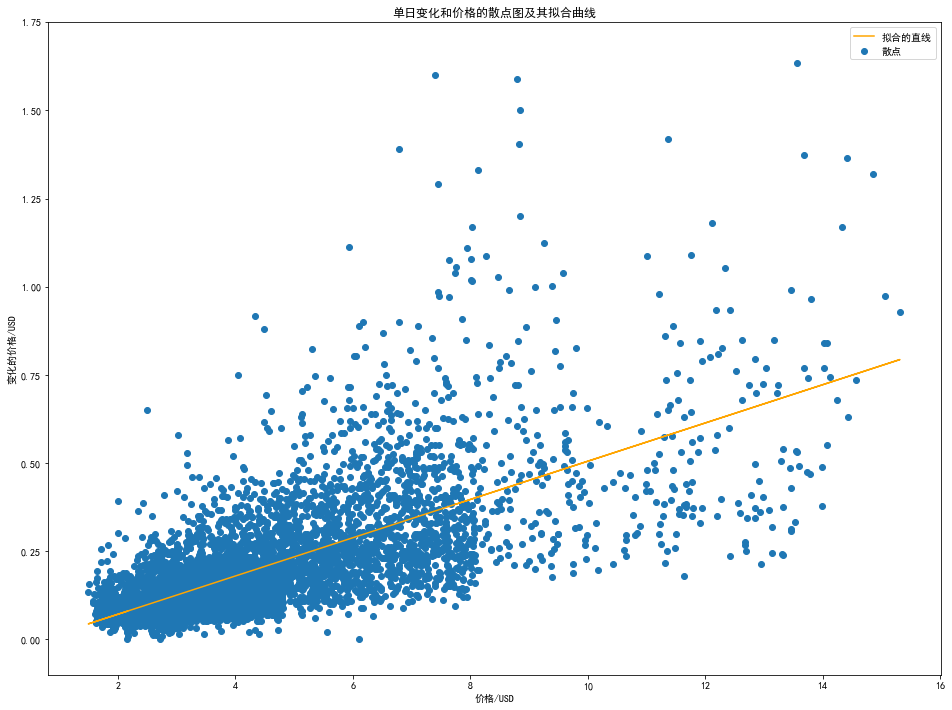


图4-6 天然气价格单日变化散点图及拟合曲线

## 对两个异常值区间数据的分析

### 对2005年8月29日至2005年12月30日异常值区间的分析

2005年8月29日-2005年12月30日中存在85个异常值，由于上文证明了单日变化程度与价格之间具有较为紧密的关系，因此通过分析可知：单日变化的平均值为0.701，最小值为0.179，最大值为2.04（出现于2005年9月28日）。

通过可视化可以看出（图4-7），这些异常值的的平均价格普遍在10以上，每日的变化程度（将数值乘以10进行处理，便于区分）也非常大，而涨跌幅则一如既往地稳定，再次证明了涨跌幅与价格的关系并不大。

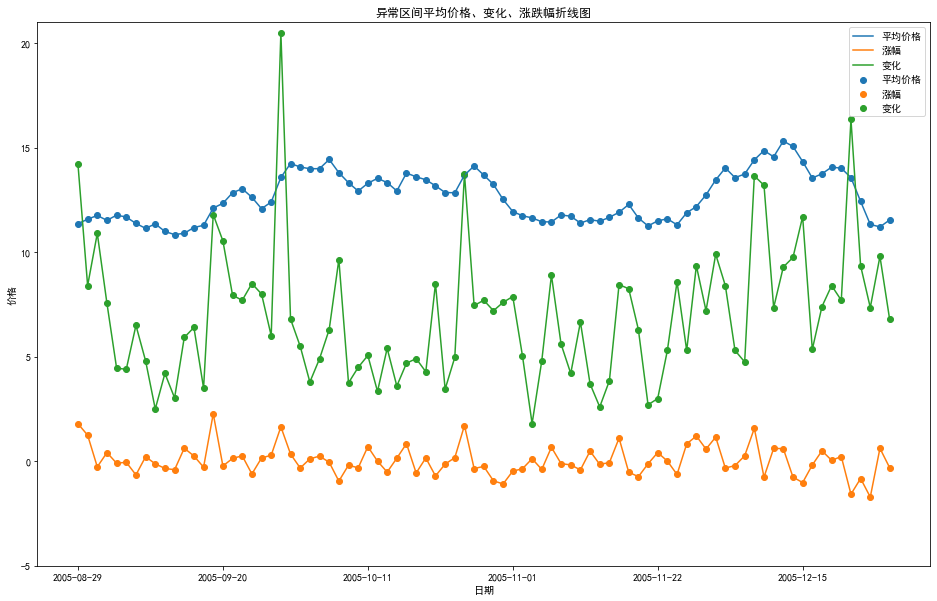


图4-7 2005年8月29日至2005年12月30日区间内的平均价格、价格变化、涨跌幅的折线图

### 对2008年4月28日-2008年7月16日异常值区间的分析

同理，需要对2008年4月28日-2008年7月16日区间内的数据进行分析，分析可得：天然气平均价格为12.1，变化的平均值为0.41，变化的最大值为0.808，最小值为0.213。

进行可视化后（图4-8），我同样发现，在异常的区间内，价格的变化仍然保持在一个较高的水平，大约是平常的一倍，而涨跌幅则对此并无非常大的影响。

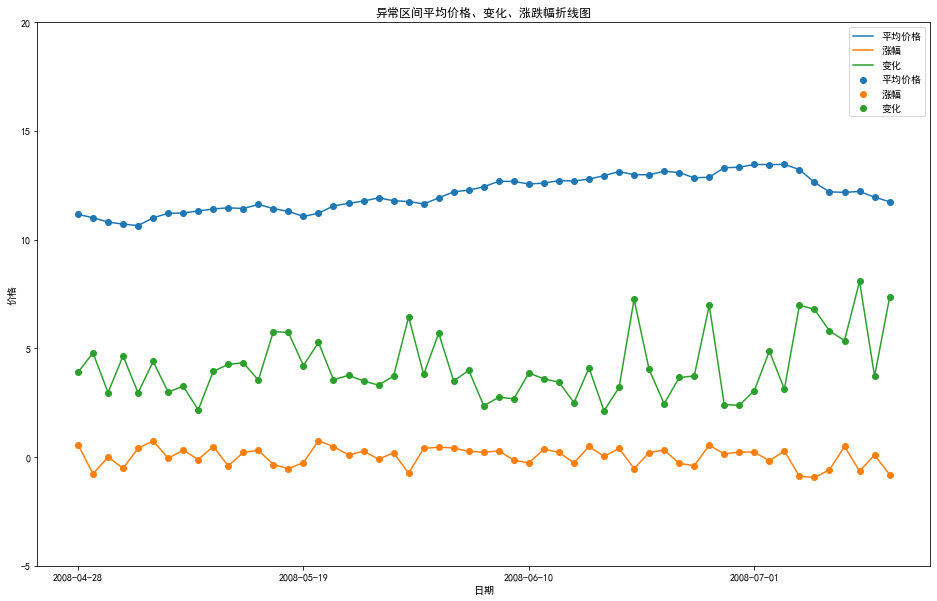


图4-8 2008年4月28日-2008年7月16日区间内的平均价格、价格变化、涨跌幅的折线图

### 两组异常区间的整体分析

通过对两组异常值集中的区间进行分析，我们发现，异常值的毎日変化的平均值为0.58，约为普通时段（0.2）的3倍之多，而平均价格12.5也为普通时期（4.46）的近乎3倍。这再次证明了价格与每日变化之间存在着较强的关系。

通过对异常值的分析以及对价格的分析，我们可以初步得出一个结论：毎日変化的值若是达到了0.58，意味着国际天然气的价格将在近期面临一个剧烈的变化，需要做好准备。

## 数据分析结论整合

通过对数据进行分析及可视化，我们能够发现许多重要的信息。

首先，上文研究中曾找到了出现价格（历史最高价格位于2005年12月13日，最低价格位于2020年6月26日）、涨幅（最大涨幅位于2003年2月24日，最大跌幅位于2004年3月18日）、单日变化（2005年9月28日）的最大值的日期。通过研究，我发现在历史上，在这几个日期附近，国际上均发生过非常重要的大事。

例如，2005年爆发了伦敦恐怖爆炸，同年8月25日，美国曾遭受了来自加勒比海的5级飓风“卡特里娜”，造成了巨大的破坏；2003年3月20日，美国掀起了伊拉克战争......而在异常值的区间内，也发生过许多国际大事，例如2006年的中东战争，2007年的次贷危机......而在这些日期中，战争、自然灾害与恐怖袭击出现的最为频繁。

因此，我们可以推断，战争和自然灾害是导致天然气价格上涨的重要因素之一！

同时，上文中也提到过：单日价格变化也是天然气价格增长带来的影响，在一定程度上，通过判断单日价格的变化，能够猜测数今后一段时间内天然气价格的变化趋势。而一旦单日价格变化幅度超过了0.58，我们便可以认定天然气价格在近期会出现剧烈的波动（增长或下降），国际上也出现了大事。

# 对当下天然气价格的分析与预测

## 对当下天然气价格的分析

通过对2000年1月1日-2022年5月9日的数据进行分析，我们可以发现：近年天然气的平均价格为3.57，保持在一个较为低的水平，但最大值在近期却达到了8.515，变化幅度的平均值为0.2，较为正常，但变化幅度的最大值却达到了1.02。这说明在未来的一段时间内，天然气将呈现增长的趋势。再加上最近俄乌战争的爆发，势必会造成天然气价格以及变化幅度的新一轮增长。

## 对未来天然气价格的预测

由于要判断数值随时间的变化趋势，本次预测采用了ARIMA模型，将前5200项数据作为训练集，进行训练（训练结果如图5-1），将最后512项数据作为预测对象进行预测。同时，对最后512项数据进行拟合。

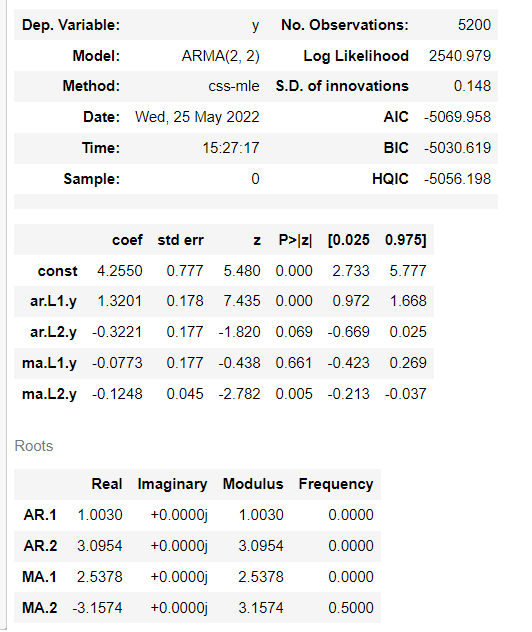


图5-1 ARIMA模型训练结果

经过可视化后（图5-2），可以看到，近期天然气价格以及每日变化幅度正在剧烈攀升，拟合后的曲线函数为：y=3.24x^3-1.54x^2+8.37x+1.69。而通过模型训练的预测曲线则代表了通常情况下的价格增长状况，但如今增长程度却远远高于该曲线，也可以说明当今天然气价格正面临新一轮的增长。

同时，通过拟合的曲线的函数，若保持相同的速率增长，预计在7月份左右，天然气的价格将会突破10大关。

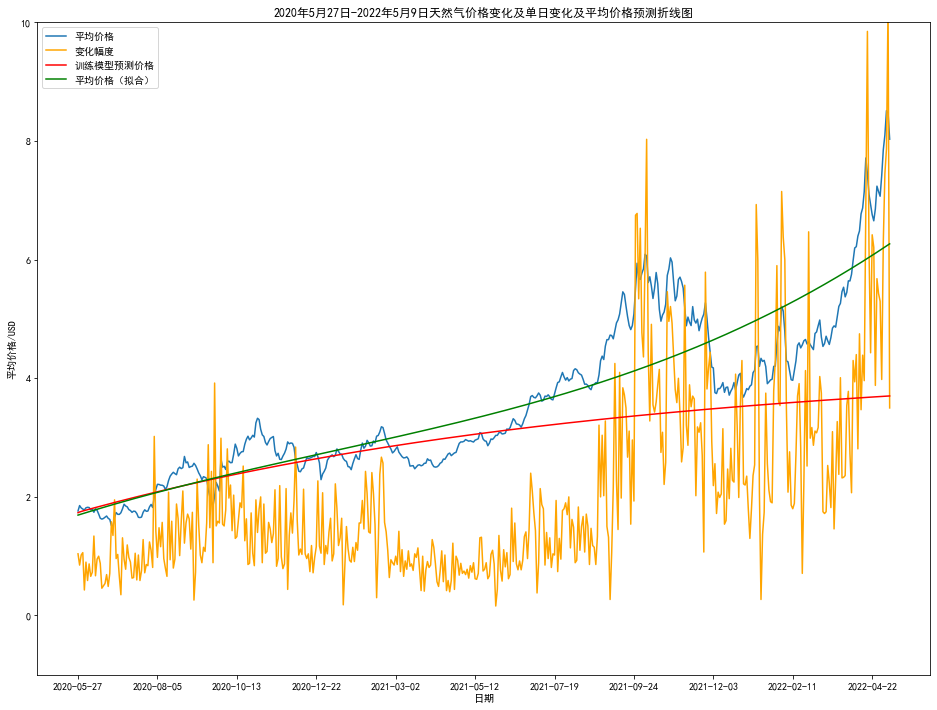


图5-2 2020年5月27日-2022年5月9日天然气价格变化及单日变化及平均价格预测折线图

# 总结与展望

通过数据分析，我们成功实现了对国际天然气价格的大数据分析，知晓了影响天然气价格变化的历史原因，对天然气价格增长所造成的影响作了分析，通过对异常区间的分析，给出了一个毎日変化程度的指标。同时，通过对近期天然气价格进行分析，预测了未来一段时间天然气的价格变化。

但在模型构建上仍存在众多不足，数学的知识仍不够充分。希望能够在未来充分学习相关的数学知识，完善模型，使其能够真正地吻合实际的变化。