

# 基于主成分分析法与 VAR 模型的猪肉价格波动分析

PB19010450 和泳毅

## 摘要

从 2002 年至今，我国经历了 5 轮猪周期，每轮猪周期持续 3-4 年。每轮猪周期中猪肉市场波动幅度较大，这种不规则的波动会对猪肉产业的健康发展带来不利影响。本文建模分析解释了这种猪肉价格的周期波动现象，从供给端、需求端和外部因素分析了造成猪肉价格波动的原因，并根据分析结果提出了若干建议。其中主要有两个问题：一是应该选取什么因素来进行分析，二是如何形象地描述各个因素对猪肉价格的影响。

针对问题一，本文在猪肉供给端选取了仔猪的价格数据和养殖饲料玉米、大豆的价格数据。在猪肉需求端选取了猪肉替代品牛肉、羊肉和鸡肉的价格数据。在外部因素方面选取了消费者满意指数、中国经济政策不确定性指数和国内生产总值的数据。并通过相关性分析和主成分分析的方法分析了各个因素的贡献度，选取了其中的六个因素作为主因子进行分析。

针对问题二，本文采用 VAR (3) 模型进行建模，为形象地描述各个因素对猪肉价格的影响，使用脉冲响应分析来解释各个因子对猪肉价格的影响，并结合实际客观规律进行具体描述。

最后，本文根据模型分析，得出了如下结论：牛肉、羊肉的价格等猪肉替代品从需求端影响猪肉价格，大豆、玉米等猪饲料的价格从供给端影响猪肉价格，中国经济政策不确定性指数和国内生产总值的变化也反映了外部因素对猪肉价格的影响。

**关键字：**猪肉； 价格波动； 主成分分析； VAR 模型； 脉冲响应分析；

# 目录

<b>一、问题重述</b>	<b>3</b>
1.1 问题背景	3
1.2 要解决的具体问题	3
<b>二、问题分析</b>	<b>4</b>
2.1 供给端因素对猪肉价格波动的影响	4
2.2 需求端因素对猪肉价格波动的影响	5
2.3 外部因素对猪肉价格波动的影响	6
2.4 数据说明	7
2.5 主要模型与方法	7
<b>三、名词解释</b>	<b>8</b>
<b>四、模型建立与求解</b>	<b>8</b>
4.1 主成分分析	8
4.2 VAR 模型分析	10
4.3 脉冲响应分析	13
<b>五、误差分析</b>	<b>15</b>
<b>六、模型评价</b>	<b>15</b>
6.1 模型的优点	15
6.2 模型的缺点	15
<b>七、结论</b>	<b>16</b>
<b>参考文献</b>	<b>17</b>
<b>附录 A 编程语言与软件</b>	<b>18</b>
<b>附录 B 核心代码</b>	<b>18</b>

# 一、问题重述

## 1.1 问题背景

我国是世界上最大的猪肉生产国和消费国，在居民的肉类消费结构中，猪肉消费总量约占三分之二，是居民肉类摄入的主要来源。据国家统计局统计数据显示，近十年来我国猪肉产量、人均猪肉消费量总体呈增长趋势，猪肉市场波动幅度较大。作为居民生活必需品之一，猪肉价格波动与居民生活密切相关。2020 年中央一号文件指出：加强市场监测和调控，做好猪肉保供稳价工作。生猪的养殖和消费在农业生产中一直占有重要位置。所以解释猪肉价格的周期波动现象，分析政府和资本在其中的作用是有必要的。

## 1.2 要解决的具体问题

猪周期，本质是利润驱动的生猪供给周期。从 2002 年至今，我国经历了 5 轮猪周期，每轮猪周期持续 3-4 年。猪肉市场中价格的波动有包括正常的价格波动和非正常的价格波动。正常的价格波动是由消费者偏好差异和生产者生产决策引发的市场价格波动，即由供需规律引起的价格波动，有利于增强市场活力。非正常的价格波动是由于猪的疾病疫情、自然灾害、经济危机等产生的波动。这种不规则的波动会对猪肉产业的健康发展带来不利影响。

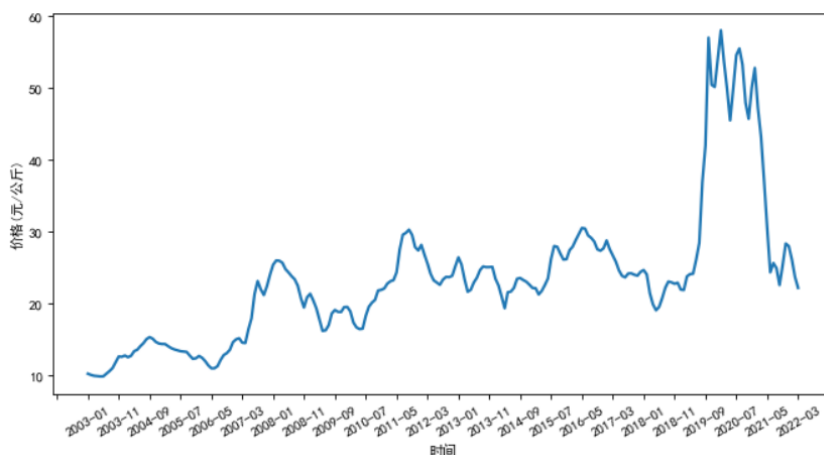


图 1 2003 年 1 月至 2022 年 3 月猪肉价格变化

从图 1 可以看出，2003 年 1 月至 2022 年 3 月我国猪肉市场价格月度走势呈现出剧烈波动的态势，尤其是 2007 年后波动较大。受甲型 H1N1 流感与猪蓝耳病等因素影响，2007 年 1 月至 2010 年 5 月，猪肉价格经历了一轮大幅的上涨，2008 年 2 月份达到最高 26.1 元/公斤的价格，变化最大幅度高达 41%。接下来受到政府调控影响，2010 年 6 月至 2015 年 3 月出现了第二轮“猪周期”，从最低 16 元/公斤的价格上涨到 30.3 元/公斤的最高价格，涨幅高达 89%。2015 年 4 月至 2018 年 5 月出现了第三轮“猪周期”，从最

高 31.3 元/公斤的价格跌为 19.5 元/公斤，下降幅度为 37%。第四轮“猪周期”从 2018 年 6 月开始一直持续到 2020 年，我国猪肉价格出现前所未有的上涨。受非洲猪瘟影响，2018 年 6 月至 2019 年 11 月，猪肉价格持续上涨，从 19.8 元/公斤上涨至 54.91 元/公斤，涨幅约为 177.32%；2019 年 12 月略微下跌至 51.09 元/公斤。2020 年开始，受新冠肺炎疫情影响，我国猪肉价格出现小幅度剧烈波动，后出现剧烈下降走势，直到 2022 年 3 月，猪肉价格下降至 27.85 元/公斤。

这种波动给生猪生产者和猪肉消费者带来较大的影响，也使得本来处于劣势的养殖者承担了更多的经营风险。因此，根据猪肉价格波动的内在机制，掌握猪肉价格的波动规律，可以一定程度上使得养殖者和消费者的利益得到保护。

本文主要从供给、需求及外部环境 3 个方面对中国猪肉市场价格波动的主要影响因素进行定性分析，通过主成分分析法确定主要指标。利用 2003 年 1 月~2022 年 3 月中国猪肉价格月度数据，构建 VAR 模型实证分析主要指标对中国猪肉价格波动的影响程度。最后根据研究结论，提出了相应对策建议。

## 二、问题分析

猪肉产品产业链涉及方方面面，从产业链上游的育种、养殖再到中游的屠宰、加工及下游的流通、零售等，每个环节都会对最终的猪肉的价格产生影响。从猪肉价格供给端影响因素看，主要包括养殖规模、养猪成本、成长周期等；从猪肉价格需求端影响因素看，主要包括猪肉替代品价格和进口猪肉数量等；从猪肉价格外部影响因素看，主要包括政府政策、期货市场、能源价格、市场等。

### 2.1 供给端因素对猪肉价格波动的影响

在养殖成本方面，猪肉价格主要与养殖饲料价格和仔猪价格密切相关，而中国生猪养殖的主要饲料是玉米和大豆。

仔猪价格变化如图 2 所示，其波动幅度和猪肉价格波动大致相同，仔猪价格越低，猪肉价格也越低，具有明显的趋同性。

玉米和大豆两种主要饲料的价格变化如图 3 所示，其变动方向基本一致。并且当玉米和大豆价格上涨时，猪肉价格下跌。当玉米和大豆价格下跌时，猪肉价格上涨，但猪肉价格无论上涨还是下跌均具有滞后性。

在养殖规模方面，猪肉价格取决于养殖户的养殖规模，养殖规模包括生猪存栏量、出栏量和能繁母猪存栏量。近些年来，尽管中国一直在提倡生猪行业养殖规模化，但仍然有近一半的生猪养殖户还是散户。据国家统计局数据显示，2020 年一季度，国内猪肉价格一直维持在 50 元/公斤以上，其中，2020 年 2 月猪肉价格涨至 57.97 元/公斤。第一季度，全国生猪存栏量为 32120 头，出栏量仅 13129 头，市场供给严重不足，致使猪肉

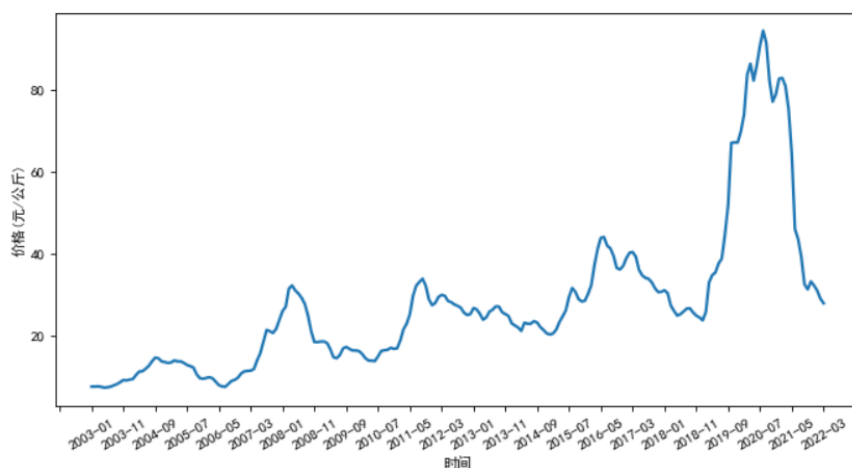


图2 2003年1月至2022年3月仔猪价格变化

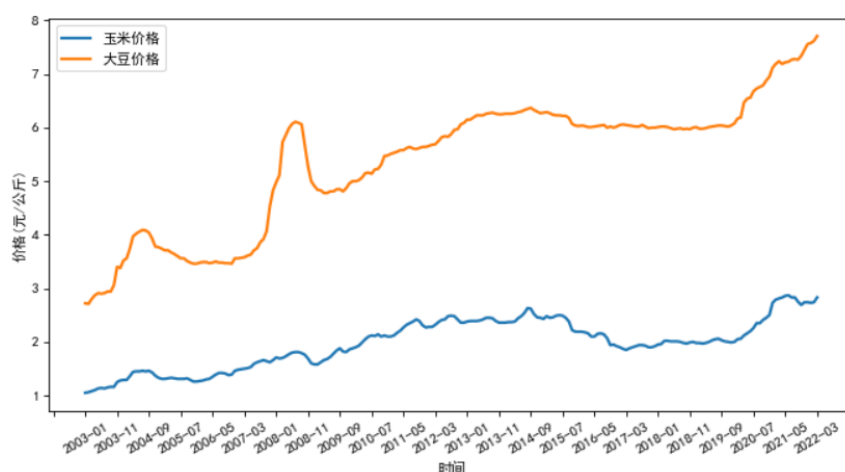


图3 2003年1月至2022年3月玉米、大豆价格变化

价格居高不下。

在生产周期方面，一般来说，生猪出栏至少需要6个月。育肥之后或出栏成为商品猪或成为后备母猪，而后备母猪继续育肥就可以成为能繁母猪，整个生产周期在1年左右。因此，能繁母猪数量及养殖户及时补栏对市场生猪供给有重要影响。近两年，国家大力推进生猪规模化养殖，在加速散户退出的同时，生猪屠宰量大幅增加，猪肉价格随之下降。

## 2.2 需求端因素对猪肉价格波动的影响

当猪肉价格过高并传导至需求端时，必然会引起需求端变化。需求端猪肉价格波动影响因素包括猪肉替代品价格等，其中猪肉的主要替代品为牛肉、羊肉和鸡肉。如图4所示，牛肉和羊肉价格变动幅度、变动方向基本一致，且两种替代品价格与猪肉价格波动具有相符性。说明当猪肉价格过高时，替代品价格变动可以在一定程度上影响猪肉价格变动。因此，要想促使猪肉市场平稳运行，必须让整个肉类市场处于平稳运行状态。

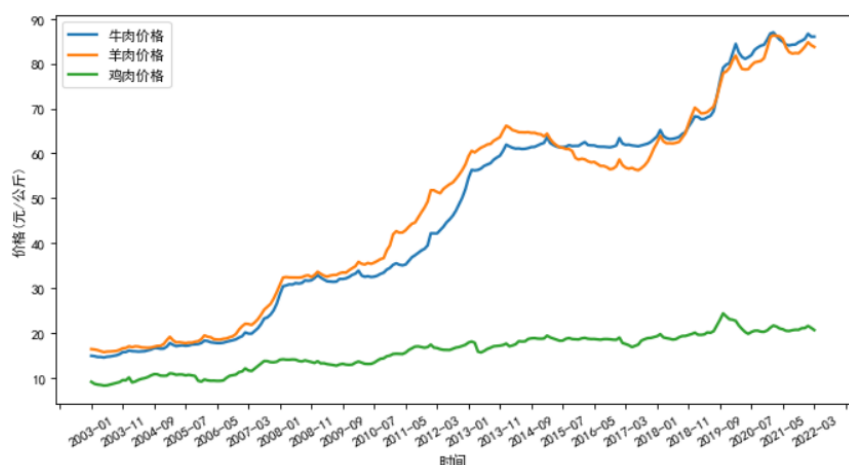


图 4 2003 年 1 月至 2022 年 3 月牛、羊、鸡肉价格变化

## 2.3 外部因素对猪肉价格波动的影响

猪肉价格除了受供需端因素影响外，还会受到许多外部因素的间接影响。外部因素对猪肉价格的影响本质上还是通过供需端影响因素产生的，如 2009 年-2014 年猪周期的政府政策和 2014 年-2018 年猪周期的环保政策等。图 5 描述了经济政策不确定性指数的变化走势，可以明显的看到走势变动幅度和变动方向和猪肉价格有着高度的一致性。

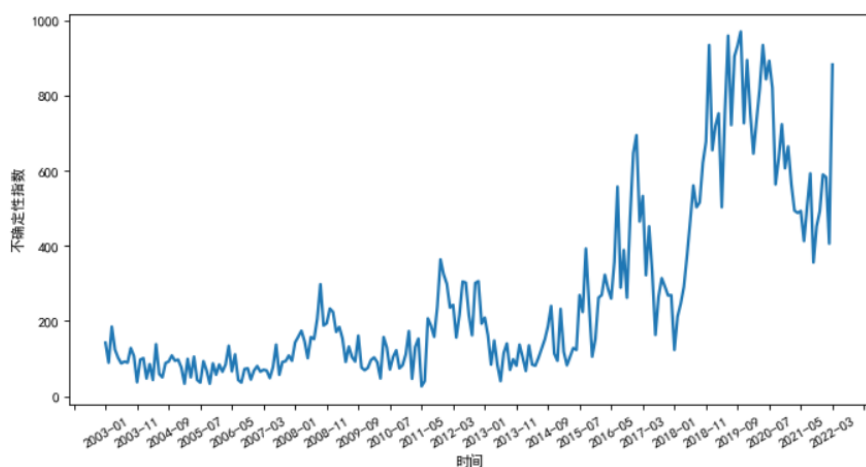


图 5 2003 年 1 月至 2022 年 3 月经济政策不确定性指数变化

随着国内外市场联系愈加密切，国际猪肉价格对国内猪肉价格的影响越来越大，且基于猪肉进口体量较大，国际市场猪肉价格变动必然会影响到进口猪肉成本，进而引发国内猪肉价格波动，所以外部的自然灾害、疫情都会对猪肉价格造成影响。此外，货币供应量、通货膨胀率、居民消费水平和居民收入等因素都会对猪肉市场需求产生影响。如图 6 描述了消费者景气满意指数的变化走势，可以看到走势变动幅度和变动方向和猪肉价格有着一定的相似性。

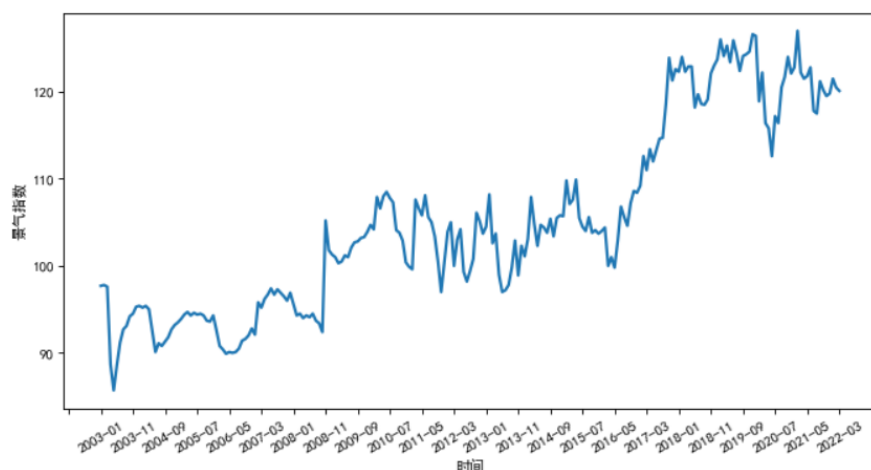


图 6 2003 年 1 月至 2022 年 3 月消费者满意指数变化

## 2.4 数据说明

根据以上分析，本文一共选取了 9 个解释变量：

- 需求因素方面，猪肉的三种替代品选取了牛肉、羊肉和鸡肉的价格数据，分别用  $X_1$ 、 $X_2$  和  $X_3$  表示；
- 供给因素方面，选取了仔猪价格数据用  $X_4$  表示，养殖饲料选取了玉米、大豆的价格数据分别用  $X_5$  和  $X_6$  表示；
- 外部因素方面，选取了消费者满意指数 CSI、中国经济政策不确定性指数 EPU 和国内生产总值 GDP 三个指标，分别用  $X_7$ 、 $X_8$  和  $X_9$  表示。

被解释变量为猪肉价格，用  $Y$  表示。

数据类型均为月度数据。其中，国内生产总值 GDP 为季度数据，对该季度数据进行算术平均处理转化为月度数据。时间区间为 2003 年 1 月-2022 年 3 月，猪肉、仔猪、玉米、大豆、牛肉、羊肉、鸡肉价格以及国内生产总值数据均来源于国家统计局，消费者满意指数 CSI 数据来源于 CSMAR 数据库，中国经济政策不确定性指数 EPU 来源于 Economic Policy Uncertainty 数据库。

## 2.5 主要模型与方法

首先采用主成分分析法将多个猪肉价格波动影响因素转化为主要影响因子，再选取关键性指标构建 VAR 模型。引入内生变量滞后值，并结合脉冲函数，分析随机扰动项对变量系统的动态冲击。

### 三、名词解释

1. 中国的经济政策不确定性指数：为了衡量中国的经济政策不确定性，Baker 等人利用通过统计新闻上与政策相关的经济不确定性的文章的频次计数，构建出月度中国经济政策不确定性指数 EPU；
2. 消费者满意指数：是描述消费者满意度指标，反映了消费者对当前经济生活的评价。由一些二级指标构成：对收入、生活质量、宏观经济、消费支出、就业状况、购买耐用消费品和储蓄的满意程度。该指数通过对城市消费者进行问卷调查，由国家统计局中国经济景气监测中心负责调查统计，以 1997 年底为基期，动态反映消费者满意指数的变化。指数数值越高，表明消费者满意度越高。

### 四、模型建立与求解

首先查看猪肉价格  $Y$  与各变量  $X_i$  的均值：

	$Y$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
均值	23.45	46.65	47.77	15.78	27.68	1.97	5.38	105.15	262.06	47270.74
方差	9.89	23.04	22.04	4.01	18.63	0.46	1.22	10.78	240.26	26507.44

可以看到数据之间存在量级的差异，为了消除不同量级带来的影响，使用 z-score 标准化对原始数据做处理：

$$z = \frac{x - \mu}{\sigma},$$

其中， $\mu$  为变量均值， $\sigma$  为变量方差。

#### 4.1 主成分分析

首先，对指标变量进行相关性分析，以此判定 9 个指标变量间是否存在相关关系。使用 Python 对指标变量作相关系数矩阵并作图查看相关程度（见图 7）。

相关系数的绝对值介于 0-1 之间。通常来说，相关系数越接近 1，表示两个变量之间的相关程度就越强。可以看到大部分变量间的相关系数都在 0.7 及以上，具有高度的相关性。

其次，在进行主成分分析之前，需要通过 KMO 统计量和 Bartlett 球形检验对数据结构进行总体分析。其中，KMO 检验是从比较变量之间的简单相关系数和偏相关系数的相对大小出发来进行的检验。当所有变量之间的偏相关系数的平方和，远远小于所有变量之间的简单相关系数的平方和时，变量之间的偏相关系数很小，KMO 值接近 1，变量



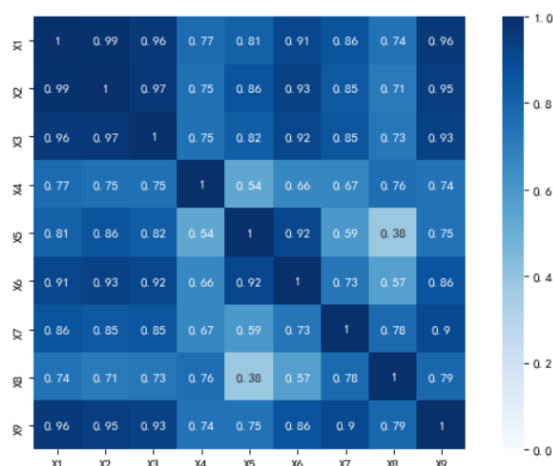


图 7 猪肉价格指标变量相关矩阵热力图

适合进行主成分分析。KMO 值的计算公式为:

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} \alpha_{ij.1,2,\dots,k}^2},$$

这里  $r_{ij}$  表示两变量间的相关系数,  $\alpha_{ij.1,2,\dots,k}^2$  表示偏相关系数。显然, 当  $\alpha_{ij.1,2,\dots,k}^2 \approx 0$  时,  $KMO \approx 1$ , 当  $\alpha_{ij.1,2,\dots,k}^2 \approx 1$  时,  $KMO \approx 0$ 。KMO 的取值介于 0-1 之间, KMO 越大越适合做主成分分析。实际分析中, KMO 统计量在 0.7 以上时效果比较好

其次, 如果变量间彼此独立, 则无法从中提取公因子, 也就无法应用因子分析法。Bartlett 球形检验主要检验数据的分布以及各变量间的独立性。该检验设立各变量间一定程度上独立的原假设, 若检验的 p-value 小于 0.05, 原假设被拒绝, 原始变量之间存在相关性, 适合做主成分分析。

使用 SPSS 对变量  $X_1 - X_9$  进行 KMO 统计量和 Bartlett 球形检验, 结果如下:

表 1 KMO 统计量与 Bartlett 球形检验结果

KMO 统计量		0.852
Bartlett 球形检验	近似卡方	4205.132
	自由度	36
	p-value	0.000

由表 1 可知, KMO 统计量大于 0.7, 适合进行主成分分析, Bartlett 球形检验结果拒绝原假设, 说明主成分分析模型整体构建状况良好, 可以据此展开下一步统计分析。

使用 Python 和 SPSS 计算相关系数矩阵的特征值和特征向量, 按照累计方差贡献率大于 85% 标准选取主因子。结果见表 2, 第一个变量特征值为 7.424, 第二个变量特征值为 0.843, 这两个特征值的累计方差贡献率达到 91.857%, 基本能够反映原始数据的

大部分信息，并且观察提取后的成分矩阵，可以看到这两个因子能表示大多数成分 85% 以上的信息，所以可以认定为主因子。

**表 2 累计方差分解与主成分提取分析结果**

成分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%	总计	方差百分比	累积%
1	7.424	82.485	82.485	7.424	82.485	82.485	4.322	48.027	48.027
2	0.843	9.372	91.857	0.843	9.372	91.857	3.945	43.829	91.857
3	0.355	3.947	95.804						
4	0.148	1.644	97.448						
5	0.086	0.956	98.404						
6	0.059	0.655	99.059						
7	0.051	0.569	99.627						
8	0.03	0.336	99.964						
9	0.003	0.036	100						

为分析各个因子对猪肉价格波动影响程度的大小，根据主成分提取表及成分矩阵，可得到各因子对猪肉价格的贡献度（见表 3）。基于上述分析及表 3 各因子的贡献度，可以发现玉米和大豆价格从供给方面对猪肉价格产生了较大影响，牛肉和羊肉价格则从需求方面对猪肉价格产生了较大影响，此外，中国经济政策不确定性指数 EPU 和国内生产总值 GDP 从外部因素对猪肉价格产生了一定程度影响。综上，下面将这些变量分别纳入 VAR 模型进行分析，即主要分析  $X_1, X_2, X_5, X_6, X_8, X_9$ 。

**表 3 各因子对猪肉价格的贡献度**

变量	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
贡献度	0.971	0.982	0.956	0.753	0.951	0.953	0.835	0.922	0.943

## 4.2 VAR 模型分析

首先为消除季节性因素影响，采用 Seasonal Decompose 平滑分离方法将时间序列分解为趋势项、季节项、残差项，剔除季节因素产生的影响，例如猪肉价格的时间序列分解如图 8：

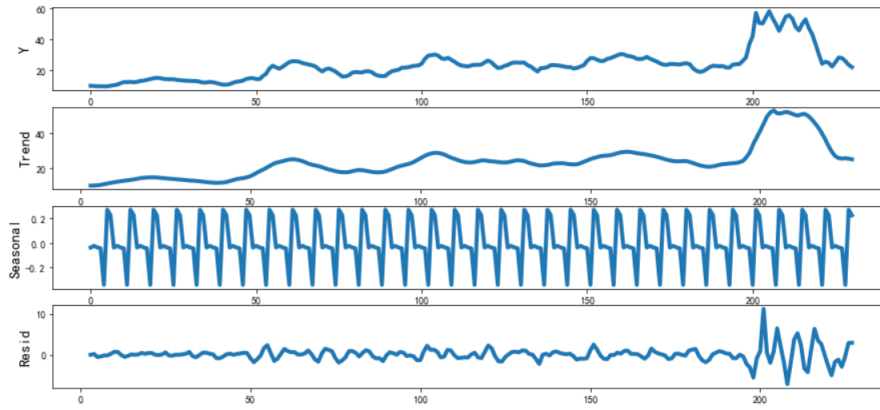


图8 猪肉价格时间序列分解图

为了防止异方差及多重共线性的情况造成影响，对各变量作对数处理。其次，在构建模型前要先对数据进行平稳性检验，采用 ADF 方法对  $\ln Y$ 、 $\ln X_1$ 、 $\ln X_2$ 、 $\ln X_5$ 、 $\ln X_6$ 、 $\ln X_8$  和  $\ln X_9$  进行单位根检验。ADF 检验就是判断序列是否存在单位根：如果序列平稳，就不存在单位根；否则，就会存在单位根。所以，ADF 检验的原假设就是存在单位根，如果得到的显著性检验统计量小于三个置信度（10%，5%，1%），则对应有（90%，95%，99%）的把握来拒绝原假设，即序列平稳。根据各变量的时序图确定检验方程具体形式，ADF 检验结果如下：

表4 ADF 检验结果

变量	检验形式 (C,T,K)	ADF 值	5% 临界值	结论
$\ln Y$	(1,0,1)	-3.601	-2.874	平稳
$\ln X_1$	(1,0,1)	-5.056	-2.874	平稳
$\ln X_2$	(1,0,1)	-4.377	-2.874	平稳
$\ln X_5$	(1,0,1)	-3.505	-2.874	平稳
$\ln X_6$	(1,0,1)	-6.899	-2.874	平稳
$\ln X_8$	(1,1,1)	-5.831	-3.430	平稳
$\ln X_9$	(1,1,1)	-4.786	-3.430	平稳

其中 C 为常数项，T 表示时间趋势，K 表示滞后阶数。可见，变量均通过了平稳性检验，说明所选基础数据都是平稳的。

下面开始建立 VAR 模型，VAR 模型即向量自回归模型，由标准自回归模型演变而来，其假设变量之间存在一个线性的依赖关系，即输出变量如  $y_t$  与输入的历史变量如  $y_{t-1}, y_{t-2}, \dots$  存在一个线性表达式。在标准自回归模型中：给定单一时间序列  $\mathbf{y} \in \mathbb{R}^T$ ，

其时间间隔的数量为  $T$ ，则对于任意第  $t$  个时间间隔，存在如下的线性表达式：

$$y_t = \sum_{k=1}^d a_k y_{t-k} + \epsilon_t, t = d+1, \dots, T, \quad (1)$$

其中， $a_k, k = 1, 2, \dots, d$  表示回归系数，常数  $d$  表示自回归模型的滞后阶数，也可以将  $d$  简单地理解成当前时间点关联过去时间点的数量。

在自回归模型中，我们的目标是从观测数据中学习出参数  $a_k, k = 1, 2, \dots, d$ 。现假设观测数据为  $\mathbf{y} \in \mathbb{R}^T$ ，对式 (1) 进行改写，令  $\mathbf{v}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-d}) \in \mathbb{R}^d, \mathbf{z} = (y_{d+1}, y_{d+2}, \dots, y_T) \in \mathbb{R}^{T-d}$ ，有：

$$\mathbf{z} \approx Q\mathbf{a}, \quad Q = \begin{bmatrix} \mathbf{v}_{d+1}^\top \\ \vdots \\ \mathbf{v}_T^\top \end{bmatrix} \in \mathbb{R}^{(T-d) \times d}, \quad (2)$$

如果进一步将  $\epsilon_t$  作为高斯噪声，采用最小二乘法，则回归系数  $\mathbf{a}$  的最优解为

$$\begin{aligned} \mathbf{a} &= \arg \min_{\mathbf{x}} \frac{1}{2} \sum_{t=d+1}^T (y_t - \mathbf{x}^\top \mathbf{v}_t)^2 \\ &= \arg \min_{\mathbf{x}} \frac{1}{2} (\mathbf{z} - Q\mathbf{x})^\top (\mathbf{z} - Q\mathbf{x}) \\ &= \arg \min_{\mathbf{x}} \frac{1}{2} (\mathbf{x}^\top Q^\top Q\mathbf{x} - \mathbf{z}^\top Q\mathbf{x} - \mathbf{x}^\top Q^\top \mathbf{z}) \\ &= (Q^\top Q)^{-1} Q^\top \mathbf{z}. \end{aligned} \quad (3)$$

而针对多元时间序列数据，VAR 模型采用了一种更为灵活的时序建模策略：给定多元时间序列数据为  $\mathbf{y} \in \mathbb{R}^{N \times T}$ ，则对于任意第  $t$  个时间间隔，存在如下的线性表达式：

$$\mathbf{y}_t = \sum_{k=1}^d A_k \mathbf{y}_{t-k} + \boldsymbol{\epsilon}_t, t = d+1, \dots, T, \quad (4)$$

其中， $A_k \in \mathbb{R}^{N \times N}, k = 1, 2, \dots, d$  表示 VAR 模型的系数矩阵，常数  $d$  表示自回归模型的滞后阶数， $\boldsymbol{\epsilon}_t$  可视为高斯噪声。

同式 (2)，VAR 模型可以改写为：

$$\mathbf{Z} \approx Q\mathbf{A}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{y}_{d+1}^\top \\ \vdots \\ \mathbf{y}_T^\top \end{bmatrix} \in \mathbb{R}^{(T-d) \times N}, \quad Q = \begin{bmatrix} \mathbf{v}_{d+1}^\top \\ \vdots \\ \mathbf{v}_T^\top \end{bmatrix} \in \mathbb{R}^{(T-d) \times (Nd)}. \quad (5)$$

采用最小二乘法，系数矩阵  $\mathbf{A}$  的最优解为

$$\begin{aligned} \mathbf{A} &= \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Z} - Q\mathbf{X}\|_F^2 \\ &= \arg \min_{\mathbf{X}} \frac{1}{2} \text{tr}((\mathbf{Z} - Q\mathbf{X})^\top (\mathbf{Z} - Q\mathbf{X})) \\ &= \arg \min_{\mathbf{X}} \frac{1}{2} \text{tr}(\mathbf{X}^\top Q^\top Q\mathbf{X} - \mathbf{Z}^\top Q\mathbf{X} - \mathbf{X}^\top Q^\top \mathbf{Z}) \\ &= (Q^\top Q)^{-1} Q^\top \mathbf{Z}. \end{aligned} \quad (6)$$

构建 VAR 模型首先要确定滞后阶数，根据 AIC-SC 准则来选择滞后阶数，尝试不同滞后阶数使得 AIC-SC 值最小，最终选取的滞后阶数为 3，使用 EViews 构建 VAR(3) 模型。

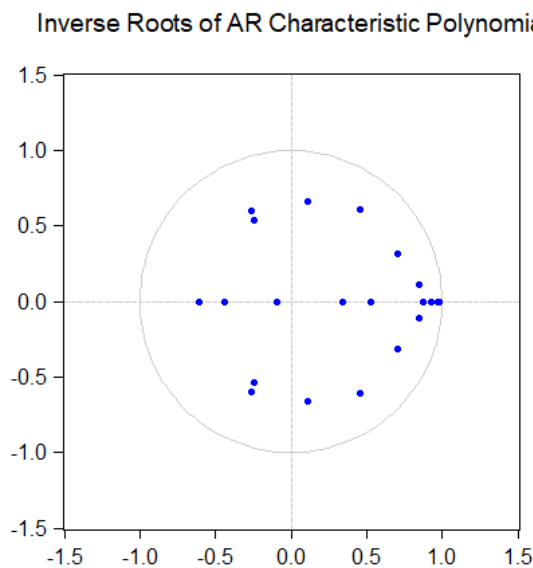


图 9 AR 根检验图

通过对各阶系数的联合显著性及残差检验，同时使用 EViews 绘制 AR 根图（见图 9），所有特征根都落在单位圆内，说明该模型不存在大于 1 的特征根，证明 VAR(3) 模型是稳定的。

### 4.3 脉冲响应分析

在进行脉冲响应分析之前，首先要对变量进行 Granger 因果关系检验。Granger 因果关系的内容是：若在包含了变量 X、Y 的过去信息的条件下，对变量 Y 的预测效果要优于只单独由 Y 的过去信息对 Y 进行的预测效果，则认为变量 X 是引致变量 Y 的 Granger 原因。即如果 X 的变化可以解释 Y 的变化，则 X 和 Y 之间存在 Granger 因果关系。Granger 检验通过受约束的  $F$  检验完成，结果如下：

表 5 Granger 因果关系检验结果

Y	X	F 值	滞后阶数	P 值	Y	X	F 值	滞后阶数	P 值
ln Y	ln X <sub>1</sub>	16.4999	3	0.0009	ln Y	ln X <sub>6</sub>	9.7114	3	0.0212
ln Y	ln X <sub>2</sub>	16.9941	3	0.0007	ln Y	ln X <sub>8</sub>	9.7627	3	0.0018
ln Y	ln X <sub>5</sub>	13.2949	3	0.0040	ln Y	ln X <sub>9</sub>	28.5432	3	0.0000

其中，牛肉和羊肉价格是猪肉价格波动的 Granger 原因，大豆和玉米价格是猪肉价格波动的 Granger 原因，中国经济政策不确定性指数和国内生产总值是猪肉价格波动的

Granger 原因。

为了更形象展现各变量间的动态关系，使用脉冲响应函数来分析模型解释变量冲击对被解释变量的影响。图 10-图 13 均为基于 VAR (3) 模型、采用正交化方法通过 EViews 模拟的脉冲响应函数图。其中，横轴表示新息冲击作用的滞后期数，纵轴表示响应变量对脉冲变量的响应程度。

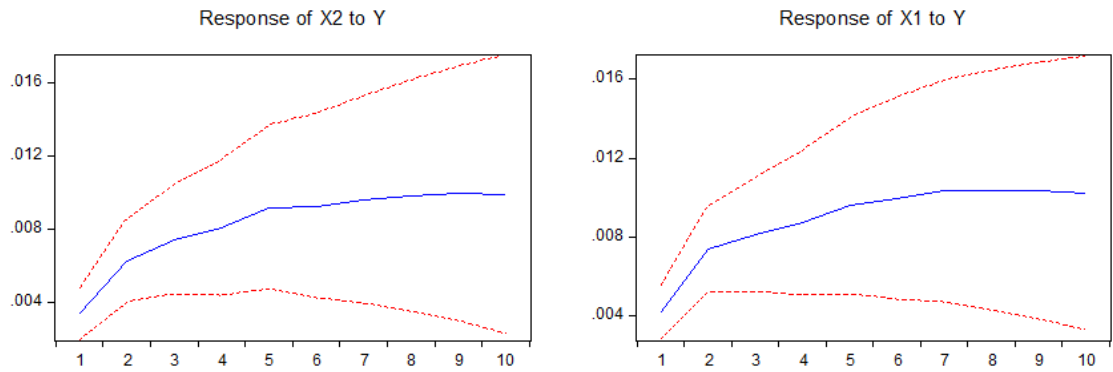


图 10  $\ln X1$ 、 $\ln X2$  对  $\ln Y$  的脉冲响应图

由图 10 可知，牛肉和羊肉的价格上涨会导致猪肉价格同幅度上涨，与格兰杰因果检验结果一致。猪肉替代品的价格升高也可以一定程度反映猪肉价格市场，反映整个肉类价格市场。并且到了第 5 期后，趋势有所平缓，牛羊肉的价格对猪肉价格的影响不再增加，这可能是由于外部因素的介入，如政府调控。

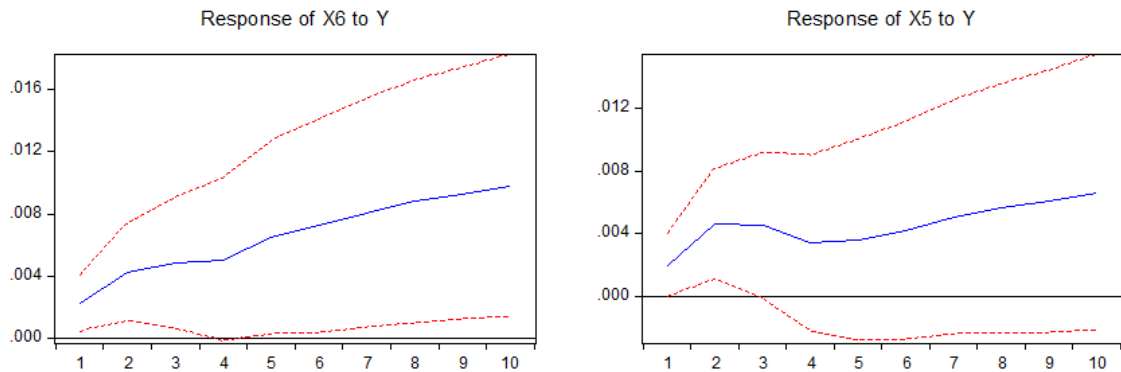


图 11  $\ln X5$ 、 $\ln X6$  对  $\ln Y$  的脉冲响应图

由图 11 可知，大豆和玉米的价格上涨会导致猪肉价格同幅度上涨，与格兰杰因果检验结果一致。大豆和玉米作为猪养殖的主要饲料，与前文分析一致，饲料是构成生猪养殖成本的主要因素之一，饲料价格越高，猪肉价格越高。并且大豆和玉米价格的持续上涨会长期影响到猪肉价格。

由图 12 可知，中国经济政策不确定性指数和国内生产总值的变化会导致猪肉价格波动剧烈。这是可以理解的，在国内生产总值发送波动时，会直接影响到国内经济市场，会多肉类市场造成多方面的复杂影响，使其价格发生剧烈波动。在面临非洲猪瘟疫情、

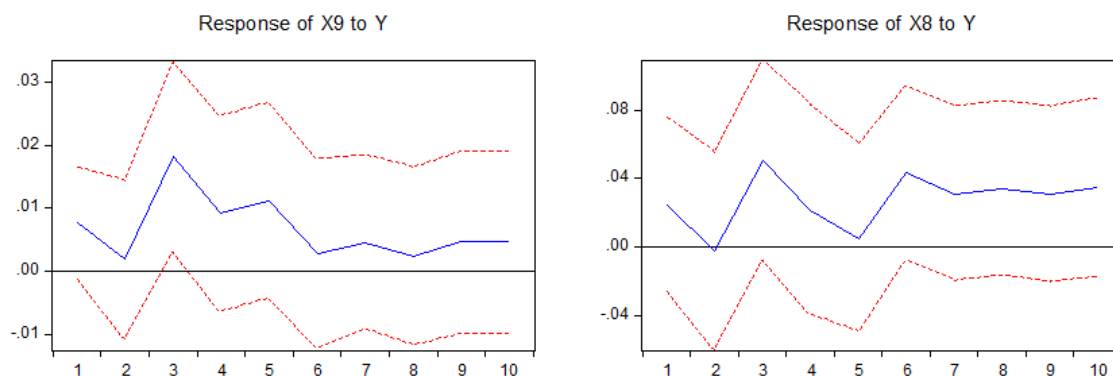


图 12  $\ln X_8$ 、 $\ln X_9$  对  $\ln Y$  的脉冲响应图

新冠肺炎疫情时，国家会加大对肉类食品的管控力度，实施严苛防疫标准，这会在一定程度上减少生猪的供应、降低人们对猪肉的需求。猪肉属于必需品，缺乏需求弹性，经济政策不确定性指数上升必然导致猪肉价格剧烈波动。

## 五、误差分析

由于国内生产总值数据为季度数据，本文采用算术平均法将其转化为月度数据，会导致一定的误差。并且处理季节性数据时采用了平滑分离的方法，该方法不能很好的剔除不规则变动要素，会对后续建立 VAR 模型并分析造成一定的影响。

## 六、模型评价

### 6.1 模型的优点

1. 本文采用的主成分分析方法可以一定程度上缩小影响因子维度，突出主因子对猪肉价格的影响，有利于后续 VAR 模型建模；
2. 本文采用的 VAR 模型在使用的数据上是平稳的，有利于后续进行脉冲响应分析；
3. 本文采用脉冲响应分析，提供了直观的可视化结果，简单形象地描述了影响因子对猪肉价格的影响，并且解释了 Granger 因果关系检验的结果；
4. 本文模型分别从供给端、需求端和外部因素对猪肉价格波动进行了因子分析，很好地解决了主要问题，并且根据分析结果可以提出建议与对策，使得养殖者和消费者的利益得到保护。

### 6.2 模型的缺点

1. 本文模型建立条件严格，不仅需要验证变量的相关性，还需要进行 ADF 检验、AR 根检验等平稳性检验，较为繁琐。且不同数据的平稳性不同，对不平稳序列还需要

- 做差分等处理，通用性不高；
2. 本文对数据的处理不够精细；
  3. 本文模型对滞后阶数的选取是在一定范围内基于准则的经验判断，可能不是最佳值。

## 七、结论

基于 2003 年 1 月-2022 年 3 月猪肉价格波动影响因素相关月度数据，利用主成分分析和 VAR 模型，研究了各变量对猪肉价格波动的影响，得到以下结论：

1. 猪肉价格波动在供给方面受到大豆、玉米等饲料价格的变化影响，需求方面受到猪肉替代品牛肉、羊肉价格的变化影响，外部因素受到中国经济政策变化以及中国经济水平变化的影响；
2. 大豆、玉米等饲料的价格主要通过供给关系来影响猪肉价格变化，从脉冲分析结果看，其对猪肉价格的波动效应在短期作用可能存在方向相反的情况，在长期作用方向一致，符合供求关系；
3. 牛肉、羊肉作为猪肉的替代品，其价格上涨会使得猪肉消费量增加，进而导致猪肉需求增加，价格随之上涨。但是其价格长期上涨后，对猪肉价格的影响会逐渐减弱，这符合政府宏观调控的结果；
4. 中国经济政策不确定性指数和国内生产总值反应了猪肉市场的外部环境，外部环境波动越剧烈，市场不确定性因素越高，猪肉价格波动也更加剧烈。

综上，本文提出如下政策建议：

1. 提高养殖户及相关行业从业人员对不确定性政策因素的重视程度，培养养殖户的自我调控能力，充分认识外部环境对猪肉消费市场的影响，从源头上减轻猪肉价格剧烈波动的可能性；
2. 构建现代生猪养殖体系，提高规模化养殖水平。政府推行全面调控措施的同时，需结合各自所在区域的价格波动特点，制定更加有效的调控措施。增强重点产区生猪疫情的预防与控制，建立科学有效的突发疫情预警防控体系和保障体系。同时也要加强应对生猪疫情能力，改善生猪养殖环境，最大程度减少非洲猪瘟、猪蓝耳病等疫情对生猪产业所造成的损失；
3. 建立价格检测预警机制，建立与生猪相关产品的价格数据库并加以模型进行预测，时刻更新数据，提高预测准确度。加强对猪肉市场的监督，防止出现不良竞争导致的价格波动。同时需要监控猪饲料市场价格变化，猪肉价格变化相较于猪饲料价格变化具有滞后性。
4. 完善我国猪肉储备体系，在我国生猪出栏量较大时，国家收储猪肉，稳定猪肉价格；当市场猪肉供给短缺时，猪肉价格上涨，国家投放猪肉，以缓解市场猪肉紧缺的局面。同时需要关注猪肉替代品价格，及时对肉类市场进行宏观调控；



5. 加大扶持资金的投入力度。增加养殖保险资金投入量，减弱突发因素对养殖户造成的影响，促进行业健康、持续、绿色、高质量发展。

## 参考文献

- [1] Baker S R ,Bloom N ,Davis S J . Measuring economic policy uncertainty[J]. LSE Research Online Documents on Economics, 2015.
- [2] 唐利群, 陈小珍. 疫情背景下中国猪肉价格波动的主要影响因素研究[J]. 价格月刊:1-11.
- [3] 李苏, 宝哲. 我国猪肉价格波动特征及预测研究[J]. 价格理论与实践,2020,(06):80-83+153.
- [4] 尚海洋, 寇莹. 猪肉价格周期波动的时空差异特征及影响因素研究[J]. 资源开发与市场,2021,37(04):445-451.
- [5] 熊靓. 中国猪肉价格变动因素与稳定对策研究[J]. 价格月刊,2021(03):36-41.
- [6] 孙滨, 谢冰清. 我国猪肉价格波动的实证分析[J]. 热带农业工程,2020,44(04):60-63.
- [7] 王长琴, 周家俊, 周德. 基于 VAR 模型的猪肉储备政策效果分析[J]. 江苏农业科学,2019,47(18):334-338.
- [8] 禹旭涛, 方永美. 基于 VAR 模型的猪肉价格波动分析[J]. 山西科技,2019,34(02):1-5.
- [9] 郑健壮, 武朝艳, 石爱林. 基于 VAR 模型的我国猪肉价格波动影响因素的实证研究[J]. 中国畜牧杂志,2022,58(02):259-265.
- [10] 刘艳芳. 国际食品价格对我国的传导效应分析——基于主成分分析和 VAR 模型的实证分析[J]. 中国猪业,2013,8(03):15-18.

## 附录 A 编程语言与软件

编程语言为 Python 3.8，编程软件为 Pycharm。

统计分析软件为 SPSS 25。

时间序列处理软件为 EViews 10。

论文排版使用  $\text{\LaTeX}$ 。

## 附录 B 核心代码

```
import math
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import datetime
import seaborn as sns
from factor_analyzer import *
from sklearn import preprocessing
from scipy.stats import bartlett
import numpy.linalg as nlg
import matplotlib.ticker as ticker
import statsmodels.api as sm
from statsmodels.tsa.api import VAR
from statsmodels.tsa.stattools import grangercausalitytests

# 数据读取
df1 = pd.read_excel('猪牛羊鸡月度数据.xlsx', sheet_name='Sheet1')
df2 = pd.read_excel('猪牛羊鸡月度数据.xlsx', sheet_name='Sheet2')
df3 = pd.read_excel('猪牛羊鸡月度数据.xlsx', sheet_name='Sheet3')
df4 = pd.read_excel('猪牛羊鸡月度数据.xlsx', sheet_name='Sheet4')
df5 = pd.read_excel('猪牛羊鸡月度数据.xlsx', sheet_name='Sheet5')
df6 = pd.read_excel('饲料价格月度数据.xlsx', sheet_name='Sheet1')
df7 = pd.read_excel('饲料价格月度数据.xlsx', sheet_name='Sheet3')
df8 = pd.read_excel('消费者景气指数.xlsx')
df9 = pd.read_excel('经济政策不确定性指数.xlsx')
df10 = pd.read_excel('国内生产总值.xlsx')

# 数据处理
df1['时间'] = df1['时间'].apply(lambda x: datetime.datetime.strptime(x, "%Y-%m"))
df2['时间'] = df1['时间']
df3['时间'] = df1['时间']
df4['时间'] = df1['时间']
df5['时间'] = df1['时间']
df6['时间'] = df1['时间']
df7['时间'] = df1['时间']
```

```

df8['时间'] = df1['时间']
df9['时间'] = df1['时间']
df10['时间'] = df1['时间']

df11 = pd.DataFrame()
df11['时间'] = df1['时间']
df11['国内生产总值'] = 0
k = -1
for i in range(df12.shape[0]):
    k = k+1 if i%3 == 0 else k
    df11['国内生产总值'].iloc[i] = df10['国内生产总值当季值(亿元)'].iloc[k]/3

df = pd.DataFrame()
df['时间'] = df1['时间']
df['Y'] = df1.iloc[:,1]
df['X1'] = df2.iloc[:,1]
df['X2'] = df3.iloc[:,1]
df['X3'] = df4.iloc[:,1]
df['X4'] = df5.iloc[:,1]
df['X5'] = df6.iloc[:,1]
df['X6'] = df7.iloc[:,1]
df['X7'] = df8.iloc[:,1]
df['X8'] = df9.iloc[:,1]
df['X9'] = df11.iloc[:,1]

### 绘图部分省略 ###

# z-score标准化处理
std_scale = preprocessing.StandardScaler().fit(df.iloc[:,1:])
df_std = df.copy()
df_std.iloc[:,1:] = std_scale.transform(df.iloc[:,1:])

# 相关系数矩阵
df_corr = df_std.iloc[:,2:].corr()

# KMO测度
def kmo(dataset_corr):
    corr_inv = np.linalg.inv(dataset_corr)
    nrow_inv_corr, ncol_inv_corr = dataset_corr.shape
    A = np.ones((nrow_inv_corr, ncol_inv_corr))
    for i in range(0, nrow_inv_corr, 1):
        for j in range(i, ncol_inv_corr, 1):
            A[i, j] = -(corr_inv[i, j]) / (math.sqrt(corr_inv[i, i] * corr_inv[j, j]))
            A[j, i] = A[i, j]
    dataset_corr = np.asarray(dataset_corr)
    kmo_num = np.sum(np.square(dataset_corr)) - np.sum(np.square(np.diagonal(A)))
    kmo_denom = kmo_num + np.sum(np.square(A)) - np.sum(np.square(np.diagonal(A)))

```

```

    kmo_value = kmo_num / kmo_denom
    return kmo_value

print("\nKMO测度:", kmo(df_corr))

# 巴特利特球形检验
df_corr1 = df_corr.values
print("\n巴特利特球形检验:",
      bartlett(df['X1'],df['X2'],df['X3'],df['X4'],df['X5'],df['X6'],df['X7'],df['X8'],df['X9']))

# 求特征值和特征向量
eig_value, eigvector = nlgeig(df_corr) # 求矩阵R的全部特征值，构成向量
eig = pd.DataFrame()
eig['names'] = df_corr.columns
eig['eig_value'] = eig_value
eig.sort_values('eig_value', ascending=False, inplace=True)
print("\n特征值\n:", eig)
eig1=pd.DataFrame(eigvector)
eig1.columns = df_corr.columns
eig1.index = df_corr.columns
print("\n特征向量\n", eig1)

# 求公因子个数m,使用前m个特征值的比重大于85%的标准，选出了公共因子是两个
for m in range(1, 9):
    if eig['eig_value'][:m].sum() / eig['eig_value'].sum() >= 0.85:
        print("\n公因子个数:", m)
        break

# 成分矩阵
A = np.mat(np.zeros((9, 2)))
i = 0
j = 0
while i < 2:
    j = 0
    while j < 9:
        A[j:, i] = math.sqrt(eig_value[i]) * eigvector[j, i]
        j = j + 1
    i = i + 1
a = pd.DataFrame(A)
a.columns = ['factor1', 'factor2']
a.index = df_corr.columns
print("\n成分矩阵\n", a)
fa = FactorAnalyzer(n_factors=2)
fa.loadings_ = a
# print(fa.loadings_)
print("\n特殊因子方差:\n", fa.get_communalities())
var = fa.get_factor_variance()

```

```

print("\n解释的总方差（即贡献率）:\n", var)

# 选取主因子
df_v = df[['时间', 'Y', 'X1', 'X2', 'X5', 'X6', 'X8', 'X9']]

# 剔除季节性影响，对数处理
n = len(df_v.columns)
for i in range(1,n):
    a = df_v[[df_v.columns[0], df_v.columns[i]]]
    timestamp=pd.to_datetime(a.时间,format='%Y-%m')
    a.index=timestamp
    a.drop('时间',axis=1,inplace=True )

    first = sm.tsa.seasonal_decompose(a,model='additive', extrapolate_trend='freq')
    TEMP = first.resid.values + first.trend.values
    std_scale = preprocessing.MinMaxScaler().fit(TEMP.reshape(-1, 1))
    TEMP = std_scale.transform(TEMP.reshape(-1, 1)) + 0.01

    df_v[df_v.columns[i]] = np.log(TEMP)

# ADF检验
adfuller(df_v['Y'],maxlag=1, regression="CT")

# VAR模型定阶
model = VAR(df_v.iloc[:,1:])
for i in [1,2,3,4,5,6,7,8]:
    result = model.fit(i)
    print('Lag Order =', i)
    print('AIC : ', result.aic)
    print('BIC : ', result.bic)
    print('FPE : ', result.fpe)
    print('HQIC: ', result.hqic, '\n')

# Granger检验
print(grangercausalitytests(df_v[['Y', 'X1']], maxlag=4))

```