**中国粮食安全与风险归因分析与破解路径**

# 摘要

近年来，外部的政治环境、经济环境以及自然环境都在急速变化，粮食安全影响日益加剧。本文通过自然环境因素与人文社会因素对粮食产量的作用关系，分析“非粮化”的动因与机理，构建种业危机预警指标体系，指导生产。

**针对问题一：**为定量刻画气候生态环境与粮食生成之间的联系，借助OLS模型通过自然环境因素与人文社会因素两个维度，度量温度、降雨量、农作物受灾面积对农作物产量的影响。使用VIF方法剔除数据中共线性数据，避免共线性的潜在影响。为验证结论的可靠性，本文借助VAR模型验证结论，对数据进行差分处理，经过平滑性检验后，建立粮食产量自回归模型。模型结果相互验证，故本文得出结论如下：**在当前气候条件下，适当的升温与降低降雨量有利于提高粮食的总产量，极端天气不利于粮食生产**。

**针对问题二：**基于时序和省域数据，经理论分析，利用偏最小二乘回归模型得到土地流转面积等**5个因子**是驱动耕地非粮化的主要因素；利用时间序列分析和空间自相关模型，可得在时序水平上，耕地“非粮化”水平总体呈现出波动变化，**未来3年会进一步趋于平稳**；空间格局分异特征显著，且驱动因子在不同区域存在空间异质性，**未来5年，空间格局分异继续深化**，但空间分布和驱动因素呈现更为复杂和多样化的趋势。

**针对问题三：**在相关研究的基础上，构建种业危机预警指标体系，由**数量安全、质量安全、市场安全、行业安全四个维度**，包含**10个方面**监测和评估中国种业的发展状况与潜在风险，使用基于熵权值的TOPSIS综合评价法，测度指标的权重值，**四个维度权重分为0.6317、0.3155、0.0388、0.0139**，结合历年数据，计算出中国13年种业危机评价指数，发现中国种业稳步发展。

**针对问题四：**根据对气候与粮食生产之间的因果联系、中国耕地“非粮化”的动因与机理和粮食种业“卡脖子”风险的主要影响因素的分析，同时构建种业危机预警统计指标体系，为指导农业生产，本研究提出**“三位三源四体”**。

**关键字：粮食安全、气候变化、耕地非粮化、种业测度**

# 问题提出

## 问题背景

粮食安全事关千家万户的民生福祉。气候变暖、地域冲突、逆全球化、疾病蔓延等多重挑战所导致的复合型风险时代已经到来，外部的政治环境、经济环境以及自然环境都在急速变化，外部环境的不确定性使得保障粮食安全的举措施行变得具有相当的难度。

在当前气象与生态环境运行机理日趋复杂化的背景下,气候条件的变异性已经直接威胁到农业生产的稳定性和安全性。粮食生产与生态环境保护存在密切的因果关联。随着全球变暖趋势持续,极端气象运动的不确定性表现明显,地表生态形态的测不准与不适宜日益复杂化,对粮食安全构成严重挑战。在未来20至50年内,气候变化的不确定性与一定程度上的不可预测性使得粮食生产进一步复杂化，并且会对全球地域性的第一产业生产、流通的全过程安全性带来严重的冲击。

提高粮食安全水平,核心生产要素要害是关键，要时刻树立种子安全意识和耕地红线意识。一方面，种子的健康和优良品质对于维护国家的粮食安全至关重要，但我国种业存在诸多“卡脖子”风险，必须未雨绸缪加强防范。另一方面，全球地域性的耕地危机进一步加剧，中国也不例外。城市化、工业用地扩张和环境污染导致耕地面积减少、土地质量下降，威胁粮食安全。

## 问题重述：

上述背景，提供了粮食安全研究的科学问题。收集相关数据，并通过数学建模与算法设计，完成以下相关研究：

* 定量刻画气候生态与粮食生产之间的因果联系，预测未来气候变化对农业生产的冲击效应。
* 分析中国耕地“非粮化”的动因与机理，定量描述中国种粮耕地变化的时空特征与演变趋势。
* 识别粮食种业“卡脖子”风险的主要影响因素，构建种业危机预警统计指标体系与破解路径。
* 根据研究结论，提供500字左右的“献策湖南”金点子1份。

# 问题分析

## 问题一分析

为定量刻画气候生态环境与粮食生成之间的联系，借助OLS模型通过自然环境因素与人文社会因素两个维度，度量温度、降雨量、农作物受灾面积对农作物产量的影响。利用农作物受灾面积、平均温度、农业总产值、平均降雨量、第一产业就业人员、农业机械总动力、农药使用量等数据构建农产品产量的预测方程。为了避免共线性的潜在影响，利用VIF(Variance Inflation Factor，方差膨胀因子)检测多元线性回归模型中自变量之间是否存在共线性(collinearity)的指标。

为了检验结论可靠性，本文通过建立VAR(Vector Autoregression，向量自回归模型)模型同时考虑多个时间序列变量之间的相互作用和影响，为了全面评估气候变化的不确定性与一定程度上的不可预测性对农业生产的冲击效应,本研究采用了ADF(Augmented Dickey-Fuller)单位根检验法,旨在检测时间序列数据是否存在单位根现象,即判断其是否为非平稳序列。借由双模型结论相互验证,我们可以综合分析气候变化的不确定性与一定程度上的不可预测性使得粮食生产的复杂影响。

## 问题二分析

在阅读大量文献的基础上，首先通过理论分析得到涵盖经济水平、社会发展、资源配置和政策环境的9个预选驱动因子，在经过数据预处理后，利用偏最小二乘回归模型得到推动非粮化时空演化的内在机制；再利用时间序列分析和空间自相关模型，并结合MATLAB和ArcGIS求解，我们就可以得到我们得到中国种粮耕地变化的时空特征与演变趋势。

## 问题三分析

深入剖析现有学术文献与行业报告，围绕粮食安全、种子质量、市场稳定性及行业健康发展四大核心领域，搭建中国种业危机指标体系，其中一级指标有4个，表征指标有10个。借助熵值理论,科学地确定了指标权重，为避免量纲的影响，采取极差标准化方法处理，使得正、逆向指标归一化映射在0和1之间。根据熵权法将指标数据归一化后的全部指标产出型权重系数，结合指标数据计算出2010-2023年中国种业危机评价指数。

## 问题四分析

通过前文对气候与粮食生产之间的因果联系、中国耕地“非粮化”的动因与机理和粮食种业“卡脖子”风险的主要影响因素的分析，着手建立了一套完备的种业危机预警统计指标体系。该指标框架旨在全面监测种业运行的各个环节。本文将从政府、企业、专业技术人员和农民四大主体、机制、产业、教育和政策三个方位和环境资源、经济资源和技术资源三种资源的三个角度对在气候冲击、耕地红线与种业危机环境下粮食生产提出建议。



图 2‑1研究框架图

# 模型假设

为使问题易于理解，本研究作出以下合理假设：

假设一：假设仅考虑作物的全生育期；

假设二：假设种植决策中仅考虑耕地总面积限制；

假设三：假设可以通过适当的指标和方法对粮食安全与风险进行测度；

假设四：假设粮食需求受到人口增长、经济发展等因素的影响；

假设五：假设粮食风险的发生只与特定因素相关，如干旱导致粮食减产、政策调整导致价格波动等。

# 符号说明

|  |  |  |
| --- | --- | --- |
| 名称 | 含义 | 单位 |
| AT | 平均气温 | ℃ |
| AR | 平均降雨量 | mm |
| CY | 作物产量 | Kg/亩 |
|  | 残差 | - |
|  | 残差平方和 | - |
|  | 方差膨胀因子 | - |
|  | 非粮化率 | % |
|  | 复种指数 | % |
|  | 非粮化面积 | 万公顷 |
|  | 用于种植粮食的土地面积 | 万公顷 |
|  | 用于耕种的土地面积 | 万公顷 |
|  | 农作物总播种面积 | 万公顷 |
|  | 耕地面积 | 万公顷 |
|  | 莫兰指数 | % |

# 问题一模型的建立与求解

## 数据来源与预处理

通过查阅相关文献、阅读行业报告，本文选取中国2010年-2023年粮食产量、农作物受灾面积、平均温度、农业总产值、平均降雨量、第一产业就业人员、农业机械总动力、农药使用量作为探究气候环境变化对粮食生产的影响，自然与人文数据说明及描述性统计见表 5‑1，数据分为两个维度刻画粮食生产，分为自然环境因素维度与人文社会因素维度。

表 5‑1自然与人文数据说明及描述性统计

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **数据变量** | **符号** | **含义** | **单位** | **mean** | **std** | **min** | **max** |
| **粮食产量** | Yield | 一年内的粮食  总产量 | 万吨 | 0.88 | 0.00 | 0.87 | 0.88 |
| **农作物受灾面积** | Disaster | 一年内受到旱涝灾害影响面积 | 千公顷 | 2.30 | 0.04 | 2.23 | 2.35 |
| **平均温度** | Temperature | 全国或时期内的平均气温 | 摄氏度 | 2.06 | 0.04 | 1.96 | 2.11 |
| **农业总产值** | Output | 农业产出的总价值 | 亿元 | 10.97 | 0.27 | 10.49 | 11.34 |
| **平均降雨量** | Rainfall | 全国或时期内的平均降雨量 | mm | 22.94 | 0.84 | -24.17 | -21.33 |
| **第一产业就业人员** | primary | 从事第一产业的就业人口数量 | 万人 | 10.38 | 0.11 | 10.18 | 10.49 |
| **农业机械总动力** | Machinery | 农业生产中使用机械设备总功率 | 万千瓦 | 11.55 | 0.06 | 11.44 | 11.62 |
| **农药使用量** | Pesticide | 一年内农业生产中使用农药总量 | 万吨 | 5.05 | 0.16 | 4.82 | 5.20 |

## OLS模型原理与建立

多元线性回归模型备受推崇,被广泛应用于各类研究。其基本原理是利用多个自变量的线性组合来预测因变量的值，同时分析这些自变量对结果的影响因素。多元线性回归的核心在于通过最小化残差平方和的思想来建立模型，以最佳拟合样本数据并估计模型参数。其公式原理如下：

设一个线性模型：

其中，为模型因变量，为模型自变量，为待估计的回归模型参数，是模型中的误差项。

OLS的目标是找到一组参数使得观测值与模型预测值之间的残差平方和最小。残差定义为观测值与模型预测值之间的差异，即：

残差平方和为：

要最小化残差平方和，可以对关于参数求偏导数，并令其等于零。

令残差平方和对参数的偏导数为零：

对于：

对于(其中)：

将这些方程整理为矩阵形式：，其中

* 是一个的矩阵，每行代表一个样本的自变量(包括截距项)，每列代表一个自变量；
* 是一个的向量，包含个样本的因变量数据；
* 是一个的向量，包含了模型的参数。

解这个方程，可以得到参数的估计值：

(5.6)

VIF(Variance Inflation Factor，方差膨胀因子)是用来检测多元线性回归模型中自变量之间是否存在共线性(collinearity)的指标。共线性意为一个解释变量可以被其他一个或多个解释变量的线性组合所近似表示，这将导致模型估计的不稳定性和解释力下降。

自变量的VIF计算度量公式如下：

(5.7)

其中，是将作为因变量，其他自变量作为自变量进行线性回归后得到的决定系数(R平方)越接近1，表示能够被其他自变量很好地解释，VIF值就越大，说明共线性越强。

## VAR模型原理与建立

在计量经济学领域,向量自回归模型(VAR)被广泛用于刻画多个时间序列变量之间的内在联系及其动态演化过程。其原理是认为当期变量都是往期所有变量的线性组合，通过将当前变量对其他变量的过去值进行回归分析，探究变量之间的联动关系。VAR模型最关键的两个参数N变量个数与K最带滞后阶数。

假设两个变量存在联系，通过内联建立两个VR模型，就可探究变量之前的关系。

(5.8)

其中,写出矩阵形式为：

(5.9)

设，经过精确计算，得到了变量滞后K阶的向量自回归模型，具体表达如下：

(5.10)

其中：

(5.11)

为了确保模型结果的准确性，需保证所得的的所有特质值都放在单位原中，其中一种验证这一条件的方法是采用增强型Dickey-Fuller(ADF)单位根检验。其原理如下：

考虑一个一阶差分自回归过程(AR(1))：

(5.12)

其中，是时间序列变量，在时间的值，是自回归系数，是白噪声误差。

ADF检验的目标是检验 𝜌=1 的原假设(即存在单位根，序列是非平稳的)，与备择假设𝜌<1*ρ*<1或𝜌>1*ρ*>1相对立(即序列是平稳的)。

ADF检验通常使用以下的回归模型来进行检验：

(5.13)

其中，是时间序列的一阶差分，是时间趋势项，是滞后一期的时间序列，是的一阶差分的滞后项，是截距，是时间趋势的系数，是的系数，是 的系数，是误差项。

## 结果分析

### 多元线性回归结果分析

对数据进行共线性检验结果如表 5‑2：

表 5‑2自然与人文数据VIF检验

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | 农作物受  灾面积 | 平均温度 | 农业  总产值 | 平均  降雨量 | 农业机械  总动力 | 农药  使用量 | 第一产业  就业人员 |
| **VIF** | 14.06106 | 2.071176 | 95.26832 | 2.141174 | 3.554746 | 13.50648 | 24.17378 |

分析结果可知但存在农作物受灾面积、农业总产值、农药使用量、第一产业就业人员四个变量数值大于10，说明数据存在严重的共线性，提出VIF值最大的变量农业总产值，重新进行检验。结果如表 5‑3：

表 5‑3调整后的自然与人文数据VIF检验

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | 农作物受灾面积 | 平均温度 | 平均  降雨量 | 第一产业就业人员 | 农业机械总动力 | 农药  使用量 |
| **VIF** | 7.736098 | 1.456972 | 1.225765 | 7.32979 | 2.188289 | 6.1965452 |

变量的VIF均在10之内，因此可以认为剔除农业总产值后数据不存在共线性。带入数据构建多元线性回归模型得到表 5‑4：

表 5‑4 OLS Regression Results

==============================================================================

Dep. Variable: 粮食产量 R-squared: 0.999

Model: OLS Adj. R-squared: 0.997

Method: Least Squares F-statistic: 783.7

Date: Sun, 12 May 2024 Prob (F-statistic): 1.56e-09

Time: 03:27:49 Log-Likelihood: 65.265

No. Observations: 14 AIC: -116.5

Df Residuals: 7 BIC: -112.1

Df Model: 6

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const 15.8528 0.132 119.885 0.000 15.540 16.165

农作物受灾面积 -0.0423 0.006 -6.557 0.000 -0.058 -0.027

平均温度 0.0135 0.002 7.242 0.000 0.009 0.018

平均降雨量 -0.0819 0.010 -7.929 0.000 -0.106 -0.057

农业机械总动力 9.395e-07 2.3e-07 4.085 0.005 3.96e-07 1.48e-06

农药使用量 0.0017 9.23e-05 18.945 0.000 0.002 0.002

第一产业就业人员-0.4799 0.015 -32.344 0.000 -0.515 -0.445

==============================================================================

Omnibus: 1.438 Durbin-Watson: 2.314

Prob(Omnibus): 0.487 Jarque-Bera (JB): 0.388

Skew: -0.397 Prob(JB): 0.824

Kurtosis: 3.182 Cond. No. 1.59e+07

==============================================================================

由于R-squared为0.999以及Prob<0.05模型有很强的准确性与可靠性，由数据回归表格可知粮食产量与其他变量可以建立模型如下：

(5.14)

回归分析中的结果表提供了宝贵的信息，其中系数部分尤为重要，它量化了自变量对因变量的影响强度。例如，农作物受灾面积的系数为-0.0423。这意味着，当农作物受灾面积变化一个单位时，粮食产量会减少0.0423个单位。换句话说，每增加一个单位的农作物受灾面积，粮食产量平均会减少0.0423个单位。

结合表达式分析，可知当前的环境气温条件下，适宜的温度上升对于我国粮食产量的增长具有积极作用，目前降雨对于大多地区已经发挥阻碍作用，除了气候变化的长期趋势外,偶发的极端天气现象同样对粮食生产构成了巨大威胁。

### VAR模型结果分析

为了保证数据的平滑性需要对为平滑数据进行差分处理，经过处理后得到的粮食产量等8个变量的ADF值和p值结果如表 5‑5：

表 5‑5 ADF检验结果

|  |  |  |
| --- | --- | --- |
| **变量** | **ADF Statistic** | **p-value** |
| **粮食产量** | -8.91211 | 1.1E-14 |
| **农作物受灾面积** | -6.84422 | 1.76E-09 |
| **平均温度** | -5.84562 | 3.69E-07 |
| **农业总产值** | -2.90547 | 0.044715 |
| **平均降雨量** | -7.98961 | 2.49E-12 |
| **农业机械总动力** | -3.22225 | 0.018732 |
| **农药使用量** | -4.33522 | 0.000386 |
| **第一产业就业人员** | -3.6961 | 0.004168 |

由表格分析可知，数据基本上在1%的可信度下(除了农药使用量是5%)，拒绝原假设，因此差分后的数据是平滑的。VAR结果如表 5‑6：

==================================

Model: VAR

Method: OLS

Date: Sat, 11, May, 2024

Time: 23:15:18

--------------------------------------------------------------------

No. of Equations: 7.00000 BIC: 58.0988

Nobs: 11.0000 HQIC: 57.6860

Log likelihood: -420.409 FPE: 1.32910e-2

AIC: 57.8456 Det(Omega\_mle): 7.22834e+24

--------------------------------------------------------------------

Results for equation 粮食产量

========================================================================

coefficient std. error t-stat prob

------------------------------------------------------------------------

const 98.630909 519.050961 0.019 0.084

========================================================================

表 5‑6 Summary of Regression Results

由结果Prob<0.1与FPE<0.05分别在10%与5%的置信水平下拒绝原假设，即本研究结果可信，检验结果见表 5‑7。

表 5‑7 VAR模型检验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Coefficient** | **Std. Error** | **t-stat** | **Prob** |
| **粮食产量** | 98.630909 | 519.050961 | 0.190 | 0.0849 |
| **农作物受灾面积** | -1311.090909 | 1450.663970 | -0.904 | 0.0366 |
| **平均温度** | 0.066818 | 0.529742 | 0.126 | 0.0900 |
| **平均降雨量** | -0.008958 | 0.050256 | 0.178 | 0.0859 |
| **农业机械总动力** | 730.748182 | 1560.258860 | 0.468 | 0.0640 |
| **农药使用量** | 0.088182 | 1.468275 | 0.060 | 0.0952 |
| **第一产业就业人员** | 75.363636 | 169.931017 | 0.443 | 0.0657 |

由表可知，平均温度和平均降水量的Prob值大于0.05，分析结果可知，极端天气与降雨增加不利于粮食的生产。

# 问题二模型的建立与求解

## 数据预处理

首先需要说明的是，本文在处理在利用所收集的数据进行时域分析时，着眼于全国范围内的耕地“非粮化”动因分析，因此采用连续有记录的六年的数据，这一方面考虑到在这段时间我国耕地“非粮化”率有明显的波动，由此可以进一步的分析出耕地“非粮化”现象的时域尺度上的驱动因素，并对其成因机理做出较为精确的分析；另外一方面也是考虑到所收集数据的不连续性。本文在进行空间尺度上耕地“非粮化”的分析时，对于所缺失的数据采用样条插值的办法补全数据后在进行分析。

## 耕地“非粮化”测度的建立

不同文献对耕地“非粮化”的内涵在理解上具有一些偏差，所以对于如何界定“非粮化”文献中各有论述，依据相关政策文件和已有的研究结果，本文认为，对于“非粮化”的定义首先需要利用权威且精确的统计数据，同时需要严格界定不同农作物的粮食意义，避免产生如种植经济作物用地的交叉来源，并且，为了进一步将所定义的“非粮化”指数与实际情况尽可能地贴合，就不可避免地考虑到实际农作物播种面积的重复性，基于上述理由，我们将非粮化指标定义如下：

(6.1)

其中，为非粮化面积，是可用于耕种的土地总面积，是用于种植粮食的土地面积，则表示土地的复种率。需要注意的是，本文语境中所提到的“粮食”一词均表示“谷类作物”。为了统一度量各个地区、各个时间段“非粮化”程度的差异性，我们定义非粮化率如下：

(6.2)

在上述公式中，我们定义为非粮化率，通过查阅文献可知复种指数定义如下：

(6.3)

在定义中，：农作物总播种面积，：耕地面积。

## 耕地“非粮化”的驱动因素和成因机制分析

### 预选驱动因子的选取

中国耕地的非粮化是一个与经济、社会、资源和政策密切相关的现象，而这些影响因素相互作用共同对于不同地区的非粮化水平造成了不同的影响。经济增长一方面使得社会得以进步，但是在另外一方面确实也增加了资源配置的压力。政策的宏观调控对于引导农村经济有着一定的影响，不仅仅是这样，政策宏观调控在对于第一产业全要素的构成优化中起到了重要作用，加速了构成要素之间的流通。不言而喻的是，GDP增长进一步增加了城乡居民收入的不平衡，由此进一步降低了种粮在市场环境下的盈利性，从而影响了农民种植热情，最终加剧了中国耕地非粮化的风险。从城市化加速的方面来看，一方面，城乡收入分配鸿沟进一步明显化，导致农业生产者的薪资水平较为低下，而种粮在市场环境下的盈利性改变了农产品需求，影响了种植结构。社会的进步导致资本市场中流通的劳动力商品减少。同时，大机器化生产,部分满足了非粮作物种植对人力资源的需求替代。然而,特殊的自然环境条件和有限的耕地资源,进一步限制了农业生产者对于第一产业的生产能力。外部经济和社会发展吸引效应导致农村地区衰退，促进农业生产者所构建的集体经济产业得到进一步发展，加速土地流转。在多方利益的共同驱动下,农业生产者对于第一产业进行内部的生产结果结构优化，进一步导致耕地要素的不变资本进一步催生出非粮价值量。各利益相关方的诉求与期望,推动了这一结构性变革的过程。

值得深入探讨的是,非粮食作物种植的扩张现象与经济、社会、资源条件以及政策导向等多重因素密切相关。这些方面的元素相互影响、相互关联,共同推动了不同地区非粮化种植格局的形成与发展。

**宏观层面：**

* 在市场中，经济和产业结构的非周期性增长导致资源的流通结构优化,同时也对现有资源分配格局产生了新的冲击。
* 政策调控则可以对乡村地区的经济发展轨迹施加引导作用,由此改善资源在城乡之间的适当分布。

**微观层面：**

* 国内生产总值的增加扩大了城乡居民的收入鸿沟,进一步导致农业生产者对第一产业的结构性经济效益产生一定程度上的怀疑，进一步放大农业生产者的心理消极，加剧“非粮化”。
* 经济进步助推城市化，吸引乡村劳动力投身非农行业,加大收入差距,改变农产品需求,影响种植结构。
* 特殊的自然环境条件和有限的耕地资源,与上述因素共同推动了非粮食作物种植的扩张。
* 外部经济社会发展对乡村地区产生了一定程度的虹吸效应,导致乡村人口外流、资源外流。为振兴乡村,相关政策吸引资本下乡,加速了土地流转。在各方利益的驱使下,非粮化种植结构转变逐步加速。

根据上述分析，参考现有研究成果，结合全国时序、地域在不同尺度上数据的差异获得，我们可以确定确定非粮化的初选驱动因子如下表所示，上述诸多因素,包括经济增长、社会变迁、资源禀赋以及政策环境等多个层面,能够较为全面地解释非粮食作物种植扩张的深层动因。这些驱动力进行复杂耦合运动,共同推动了不同区域非粮化生产格局的形成与发展。其中： 是种植效益因子，代表农业生产者在第一产业中的利益可获得性水平； 是耕地面积， 是粮食作物播种面积，同样代表资源禀赋对非粮化的影响，是农业机械总动力，代表资本要素中的技术要素对第一产业结构变化的影响， 是年末总人口数量，是城镇人口数量，是乡村人口数量，是乡村从业人员数量，这些变量用于表征人力资本对非粮化的影响，是全国土地流转面积，代表土地要素在市场下的流通对中国耕地非粮化的冲击。

### 非粮化种植驱动力的偏最小二乘回归分析

运用偏最小二乘回归模型,本节对影响耕地利用向非粮食作物种植转变的关键驱动因素及其作用机理进行了探讨分析，剖析了各驱动要素对非粮化现象的作用路径及影响程度。

鉴于预设的自变量驱动因子较多,且根据前期理论分析,这些因子之间存在一定的多重共线性关系。另一方面,可获取的观测数据样本量又相对有限。所以我们综合考虑使用偏最小二乘方法来建立中国耕地非粮化的内在动因模型，从而揭示各驱动要素对被解释变量的作用路径及影响程度。

我们接下来考虑建立非粮化率与预选驱动因子的偏最小二乘回归模型。 设预选驱动因子的次标准化观测数据矩阵为：

(6.4)

非粮化率的次标准化观测数据矩阵为：

(6.5)

**Step1：提取预选驱动因子和非粮化率的第一对成分**

首先需要从预选驱动因子变量组和非粮化率变量组中分别提取出第一对成分。设选驱动因子变量组的线性组合记为,被解释变量非粮化率的第一成分记为，则有：

(6.6)

是非粮化率的线性组合：

(6.7)

**为了分析的需要，我们要求以下两点：**

* 和各自最大化提取预选驱动因子和非粮化率的变异特征。
* 和的相关性达到最大。

为了对上述两点做出精确的度量，我们从预选驱动因子和非粮化率的次标准化观测数据矩阵出发，可以得到第一对成分的得分向量，不妨将其记为：，则我们有：

(6.8)

以及：

(6.9)

值得注意的是,和之间的协方差可以通过它们的标准化向量和的内积来计算,从而，我们将上述两点要求转化成了带约束的优化问题，以此获得更加准确可靠的回归结果：

(6.10)

通过使用罚函数法，我们可以将关于驱动因子和非粮化率的有约束极值问题转化为关于罚函数的无约束极值问题。因此，去找向量和，使得最大化。在使用罚函数的有关技术和进行数学推到之后，我们可以得到的计算公式：

(6.11)

**Step2：建立解释因子对被解释因子的回归方程**

设回归模型为：

(6.12)

其中，，分别代表模型中的未知参数，而和则表示模型中的残差阵。根据最小二乘法，我们可以得到效应负荷量的表达式：

(6.13)

**Step3：利用残差阵重复上述步骤**

设,那么残差阵就可以表示为, 。注意到如果残差阵中元素的绝对值近似为0, 那么我们自然地认为用第一个成分建立的回归模型的精度已经满足偏最小二乘回归分析的需要了，那么我们就可以停止抽取成分，算法到此为止。若不然，我们考虑使用残差阵和代替 和重复以上步骤，就可以得到：

(6.14)

注意到为预选驱动因子和非粮化率的第二对成分的得分向量，而

(6.15)

分别表示预选驱动因子和非粮化率的第二对成分的负荷量。则我们有：

(6.16)

**Step4：得到偏最小二乘的回归方程式**

设阶的数据阵的秩为，其中，那么一定存在个成分，使得下式成立：

(6.17)

再将带入到中去，我们就可以得到个关于非粮化率变量的回归方程为：

(6.18)

**Step5:交叉有效性检验**

与主成分分析类似，利用偏最小二乘法并只需要选取前个成分就可以得到较好的回归模型。对于我们可以利用交叉有效性检验做出先验的估计。而在本文中由于对影响非粮化率的预选驱动因子进行了理论上的分类，在这里我们不妨取.

由此，利用所建立的非粮化率与预选驱动因子的偏最小二乘回归模型，就可以对中国耕地“非粮化”的动因与机理展开定量分析，并为定量描述中国种粮耕地变化演变趋势做准备。

接下来，我们来求解上述非粮化率与预选驱动因子的偏最小二乘回归模型。

**Step1：将数据进行标准化**

令表示各个指标值的标准化数据，那么我们有：

(6.19)

其中 对于我们有标准化指标：

(6.20)

其中

**Step2：求出预选驱动因子和非粮化率的相关系数矩阵**

下面大型矩阵给出了这十个变量的简单相关系数矩阵：

从相关系数矩阵中可以看出：非粮化率与城乡居民可支配收入比、粮食作物播种面积、农业机械总动力、年末总人口、城镇人口、乡村从业人员成负相关关系，而非粮化率与耕地面积、乡村人口、全国土地流转面积成正相关关系。

**Step3：提取驱动因子与非粮化率的成分**

利用MATLAB软件，求得的各对成分分别为：

(6.21)

注意到前三个成分的自变量解释率为，所以我们只需要提取出三个成分即可。

**Step4：求三个成分对预选驱动因子与耕地“非粮化”之间的回归方程**

求得预选驱动因子和非粮化率与、、之间的回归方程分别为：

(6.22)

以及：

(6.23)

**Step5：求预选驱动因子与耕地“非粮化”之间的回归方程**

将Step3中成分带入Step4中的回归方程，得到标准化之后的预选驱动因子与耕地“非粮化”率之间的回归方程为：

(6.24)

我们注意到如果显示出非粮化率与驱动因子之间的回归方程，那么驱动因子前面的系数将会变得极小，这将不利于我们的讨论，不失一般性，类似上面的讨论，我们得到预选驱动因子与耕地“非粮化”面积之间的回归方程为：

(6.25)

我们将利用图表来进一步阐述耕地非粮化的动因与演变机理。我们首先给出标准化了的非粮化率与驱动因子之间回归系数直方图，如图 6‑1：



图 6‑1非粮化率与驱动因子之间回归系数直方图

从图中我们非常直观地观察到城乡居民可支配收入比、耕地面积、粮食作物播种面积、农业机械总动力、以及土地流转面积在时域尺度上是驱动耕地非粮化地主要因素，而其它变量与之相比则不显著。

为了考察我们得到的标准化了的非粮化率与驱动因子之间回归方程的精度，我们将所有的耕地非粮化率的观测值与预测值都在一张图中呈现出来，如图 6‑2，考察其是否都在对角线附近均匀分布。



图 6‑2耕地非粮化率的观测值

从上图中我们可以看到，耕地非粮化率的观测值与预测值都在对角线附近均匀分布，由此我们可以得出我们所建立的回归方程模型的拟合精度非常高。对于非粮化面积与驱动因子的回归模型，如图 6‑4、图 6‑5所示。



**图 6‑3非粮化面积回归系数直方图**

这进一步说明了城乡居民可支配收入比、耕地面积、粮食作物播种面积、农业机械总动力、以及土地流转面积在时域尺度上是驱动耕地非粮化地主要因素，而其它变量与之相比则不显著以及我们所建立的回归方程模型的拟合精度非常高的结论。



图 6‑4非粮化面积拟合效果图

我们将上述分析总结如下：

首先对农业生产者在第一产业中的生产结构影响因素进行分析，如图 6‑5，包括经济效益、资源配置、科技进步以及政策环境等方面，并选取相应的变量进行考察。



图 6‑5“非粮化”影响因素分析图

随着生产要素成本的上升,农作物生产面临着压力,其中劳动力价格上涨成为主要因素。理性的种植人员会根据种植粮食作物还是非粮食作物的收益差异调整种植方案。

城乡发展深刻影响了农村劳动力流向非农业领域。各种类型的农业生产者在种植目标和倾向上存在差异。资本市场中劳动力商品的减少导致从事第一产业生产活动的农业生产者调整生产要素的结构化配置，利用直接现实的大机械力来弥补第一产业人工劳动力的空缺,或由于资金约束放松，所以资本要素会将第一产业生产活动的生产结构进行逐利化,从而第一产业生产种植结构也会逐利化。

科技的进步促进了第一产业的内部结构优化。不同的农作物对于机械力的使用程度以及使用方式都不大相同,从而影响了耕地上粮食作物和非粮食作物的种植规模,进一步加深第一产业耕地要素上的粮食不变资本与非粮不变资本的内在结构性变化。

政策环境,有助于降低种粮主体的资金约束。此类支持政策能够提高生产者的种粮意愿,而对种植结构的不当引导可能导致"非粮化"趋势。

## 中国耕地“非粮化”的时序变化特征描写

根据浙大卡特企研中国涉农研究数据库(CCAD)、国家统计局、EPS数据平台所收集到的数据，利用所建立的“非粮化”统计测度，我们可以直观清晰地察觉到中国耕地“非粮化”的时序变化特征。考虑的离散数据的不连续性，从而导致数据的光滑性条件不强，因此，我们利用耕地非粮化率对于时间的一阶差分，即：耕地非粮变化率，来反映中国耕地非粮化的时序特征，见表 6‑1非粮作物种植比例的变化值。

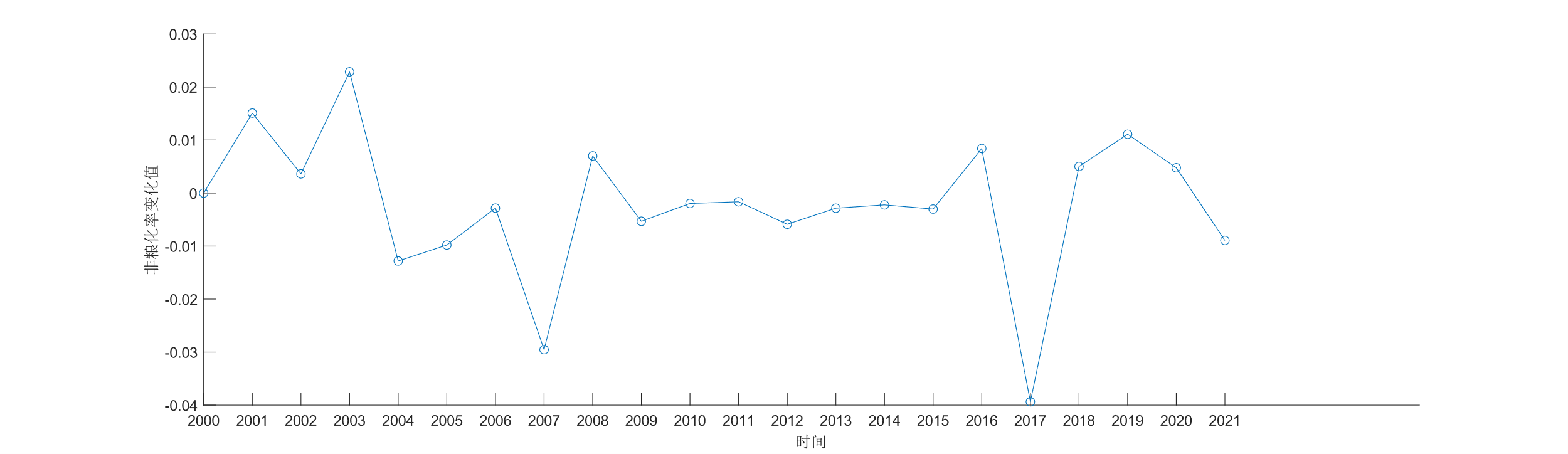


图 6‑6非粮化率变化值

自2000年至2021年，中国耕地上非粮作物的种植比率和面积发生了变化，具体数据可见于图 6‑7。尤其是自2016年以来，“非粮化”的趋势显著增加。在2003年之前，耕地的非粮化率先上升后下降再上升；而2003年至2004年期间，中国耕地非粮化率迎来了第一个大幅度的下降阶段。

表 6‑2 2000-2021年非粮化率变化和非粮面积变化统计表

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **年份** | 2000—2001 | 2001—2002 | 2002—2003 | 2003—2004 | 2004—2005 | 2005—2006 | 2006—2007 |
| **非粮率变化** | 0.0151 | 0.0036 | 0.0229 | -0.0128 | -0.0098 | -0.0028 | -0.0295 |
| **非粮面积变化** | 2081.1 | 57.3 | 2435.1 | -1402.4 | -588.6 | 280.7 | -6207.5 |
| **年份** | 2007—2008 | 2008—2009 | 2009—2010 | 2010—2011 | 2011—2012 | 2012—2013 | 2013—2014 |
| **非粮率变化** | 0.007 | -0.0053 | -0.002 | -0.0016 | -0.0059 | -0.0029 | -0.0022 |
| **非粮面积变化** | 2331.1 | 220.1 | 586.1 | 442.9 | -463.8 | 55 | -15.7 |
| **年份** | 2014—2015 | 2015—2016 | 2016—2017 | 2017—2018 | 2018—2019 | 2019—2020 | 2020—2021 |
| **非粮率变化** | -0.003 | 0.0084 | -0.0394 | 0.005 | 0.0111 | 0.0048 | -0.0089 |
| **非粮面积变化** | -105.1 | 1517.9 | -6688.3 | 663.9 | 1852.9 | 1439 | -1005 |

在2004年到2006年，耕地非粮化的开始缓慢回升，但是始终保持负增长，在这段时间内，第一产业内部结构中的“非粮化”现象并没有被统计数据明显地反映出来。然而，2007年前后出现了异常的增长，造成上述情况的潜在原因,很可能源于当年发生的极端气象灾害事件，例如2008年农作物受灾面积高达3999万hm²，绝收面积为403.2万hm²，将第一产业的相对脆弱性进一步明显化。

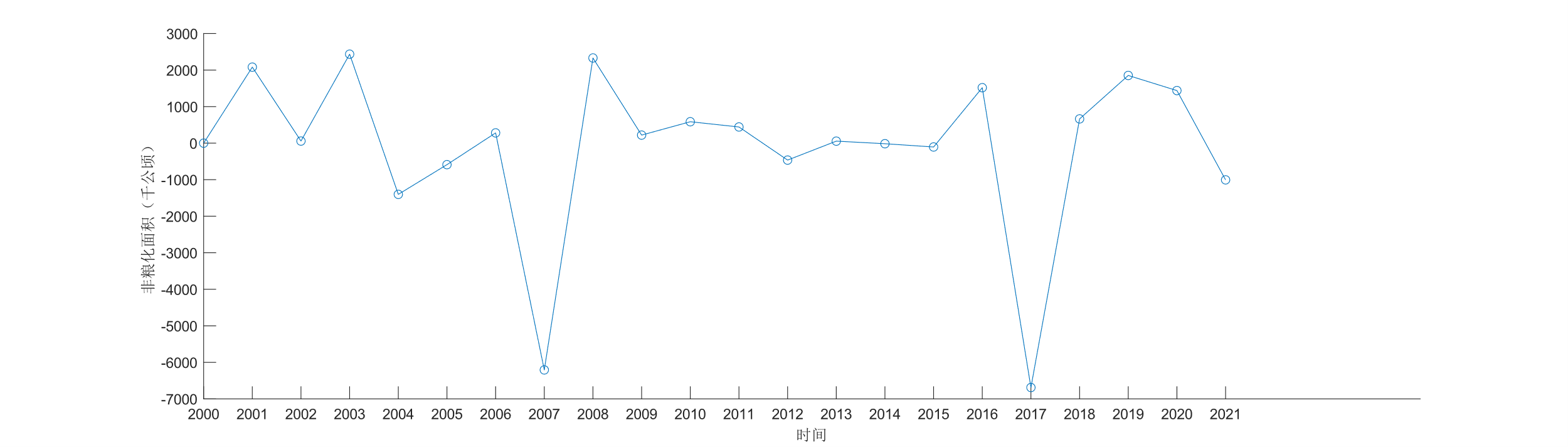


图 6‑7非粮种植面积变化

自2011年至2016年，非粮作物种植的变化呈现出趋于稳定的态势。尽管非粮种植面积在一定程度上有所增加，但总体上非粮面积比例呈现负增长趋势，这说明在这段时期中国的粮食生产相对稳定，第一产业内部结构中的“非粮化”现象并没有被统计数据明显地反映出来，详见图67。然而，自2016年开始，非粮作物种植迅速增加，非粮面积和比例均显著增长。2017年和2018年的非粮比例相较上一年分别增长了一定幅度，显示出中国耕地的“非粮化”水平在此期间明显提升。而自2018年至2021年，中国耕地的非粮作物种植比率和面积呈现出倒“U”形态，表明非粮化趋势已经出现阈值效应。

## 中国耕地“非粮化”的时序演变趋势

长期来看，在时序演变尺度上，第一产业耕地要素流通转移“非粮化”有一定规律，但短期受政策环境调控、市场供需结构调整等因素影响，预测较困难。预测耕地“非粮化”有多种方法，包括生长曲线和指数平滑等经典方法，但对短期波动预测能力却是有限的。又考虑到AR自回归模型结合了耕地“非粮化”时间序列的依赖性和随机波动，对经济短期趋势预测准确度较高。基于上述原因，本文利用AR自回归模型来探讨中国耕地“非粮化”的时序演变趋势。

设原始的中国耕地“非粮化”率的时间序列为，我们首先对的平稳性做出检验。对于显著性水平0.05，由

(6.26)

算得，计算得统计量，可以得到上分位数的值为，所以，可以认为这是一个非平稳的时间序列，因为，时间序列有上升趋势。

为了构造平稳的时间序列，对于中国耕地“非粮化”率原始时间序列作一阶差分运算，进而得到时间序列模型为：

(6.27)

其中、是二阶递推方程中需要确定的系数，是与时间耦合的随机扰动项。

对上述时间序列模型的求解，我们可以得到中国耕地“非粮化”的时序演变趋势如图 6‑8、图 6‑9：

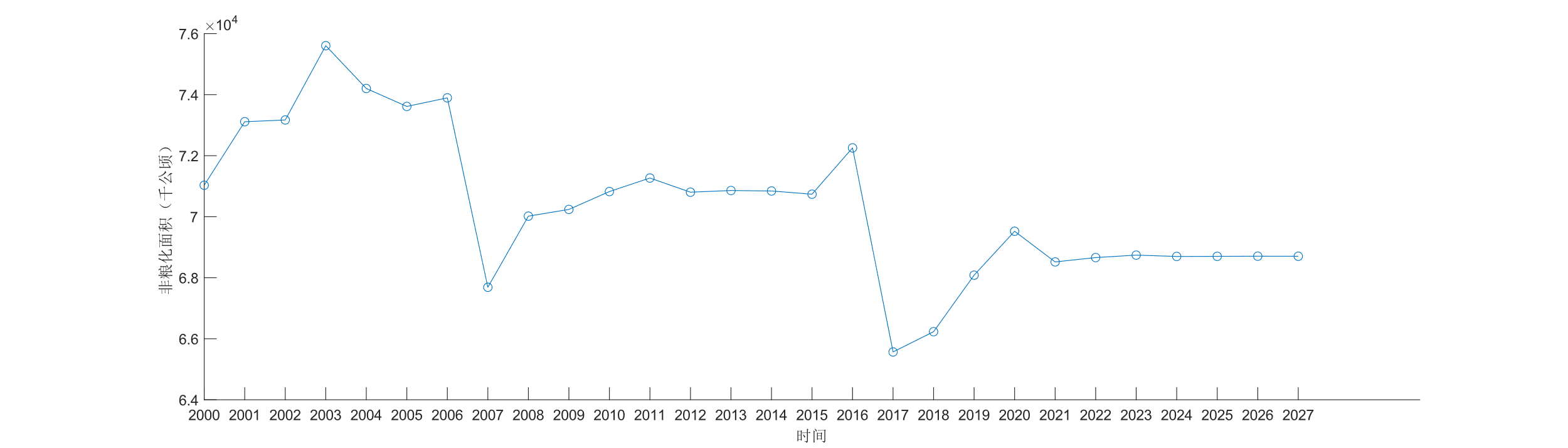


图 6‑8非粮化率演变趋势

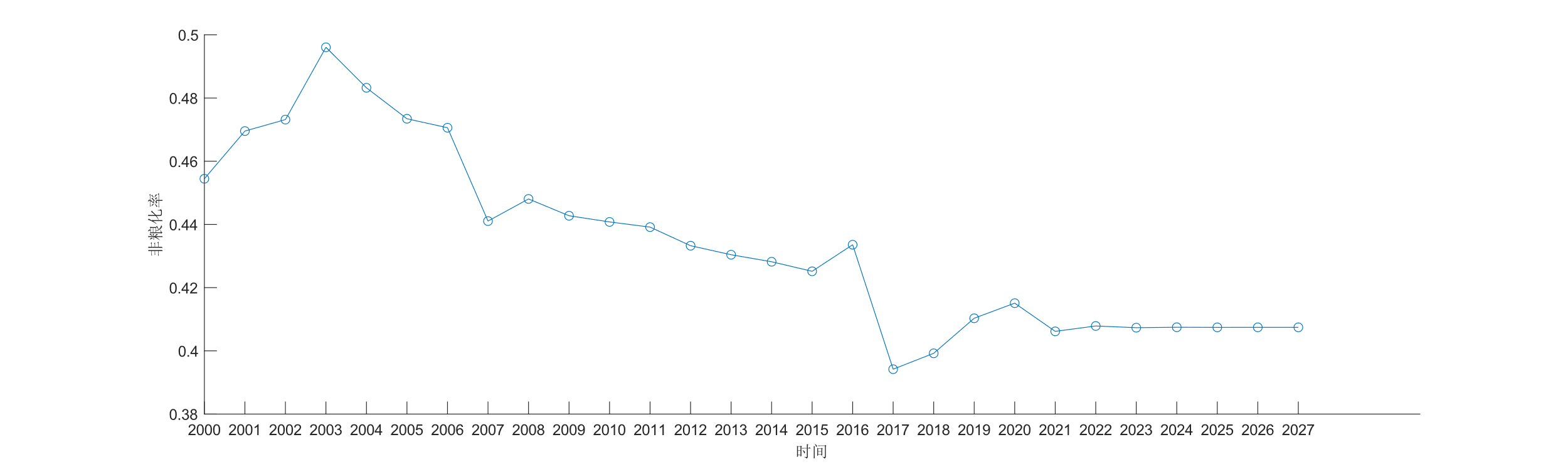


图 6‑9非粮化面积演变趋势

在未来五年，在影响中国耕地非粮化的四大驱动因素的共同作用下，中国耕地“非粮化”率会趋于平稳。结合前文所分析的中国耕地非粮化驱动因子和成因，我们将未来五年的中国耕地非粮化时序演变趋势总结如下：

根据上文内容，未来五年的中国耕地非粮化时序演变趋势可能受到多种因素的影响，包括但不限于人工成本、种粮比较收益、劳动力非农就业、机械动力水平以及政策环境等。首先，如果种粮成本持续上升而收益相对较低，不难发现，这件导致会有更多的农业生产者不愿意将耕地来种植粮食。其次，劳动力的非农就业也是影响非粮化趋势的一个重要因素，这可能导致耕地资源的闲置或转向非粮作物的种植。此外，机械动力水平的发展也可能对非粮化趋势产生影响，这有助于稳定粮食生产并降低非粮化的风险。然而，如果机械动力水平的提升不足以弥补劳动力减少带来的负面影响，非粮化趋势可能会继续加剧。最后，在探讨影响农业生产者参与要素流通的诸多因素中,政策环境无疑扮演着关键角色。政府通过制定相关政策和制度来调控耕地利用和粮食生产。这些政策措施的实施将直接影响农民的种植决策和耕地利用方式。综上所述，未来五年的中国耕地非粮化时序演变趋势将受到多种因素的共同影响。为了保障农业生产者对于第一产业经济效益地结构信任，政府需要继续加强政策调控和监管力度，提高农业生产者的第一产业结构化经济收入水平和心理预期收入。

## 中国耕地“非粮化”的空间尺度变化特征描写

为了进一步探究中国种粮耕地变化的时空演变特征，我们考虑建立空间自相关模型。

用 表示莫兰指数，其计算公式如下：

(6.28)

(6.29)

其中，为研究中国耕地的总划分位点数目；和分别为第与中国耕地的总划分位点的观测值；为全部中国耕地的总划分位点的均值；为每一个中国耕地的总划分位点的空间权重矩阵；若与为克罗内克下标符号。

将所收集到的省域和时序数据导入到ArcGIS 10.8中，得到如下所示中国耕地非粮化空间格局分异图：

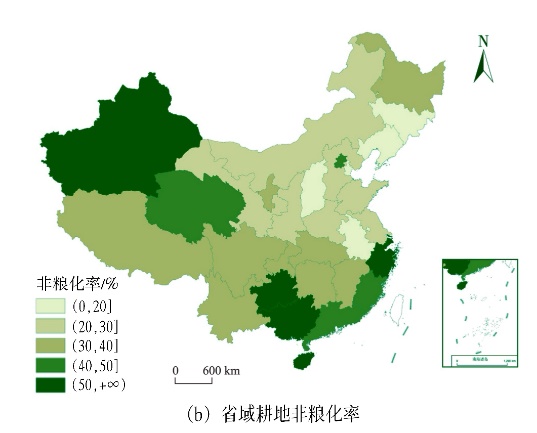
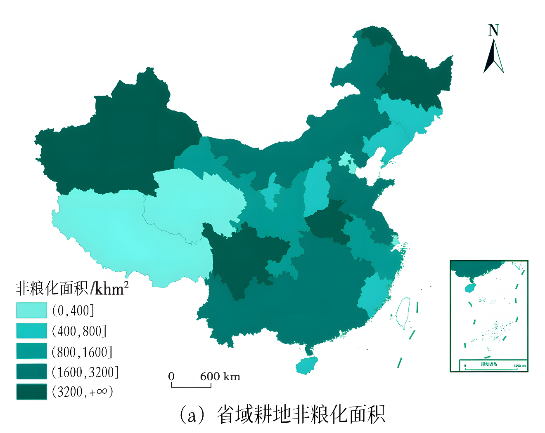


图 6‑10中国耕地非粮化空间格局分异图

由上图我们可以清晰地看到，中国耕地非粮化的地域分布呈现出明显的特征，主要表现在地域集中性、空间分布不均、尺度差异和驱动因素多样性等方面。首先，地域集中性是非粮化现象的显著特征之一。特定地区如新疆、贵州和东南沿海等地的非粮化率较高，呈现出明显的地域特点，这些区域在非粮化现象上相对集中。其次，从全国范围来看，非粮化面积主要集中于中南部和北部地区，但高非粮化率的区域分布较为零散，没有形成统一的连片区域，呈现出空间分布不均的特征。第三，非粮化的空间格局在不同尺度下存在显著差异。省域和市域尺度的非粮化分布和驱动因素往往有所不同，显示出明显的尺度效应。最后，非粮化的地域分布形态受到多种因素的共同影响，包括经济增长、社会发展、政策调控以及资源配置等。这些因素在不同地区的作用程度和方式各不相同，这表示出了非粮化空间格局的复杂性。中国耕地非粮化的空间格局表现出地域集中性、空间分布不均、尺度差异和驱动因素多样性等特征。这些特征反映了中国耕地利用在不同地区、不同尺度下的复杂性和差异性。因此，为了有效管理和控制非粮化现象，需要因地制宜，具体问题产出不同的第一产业耕地要素的合理分配，制定差异化的政策和措施，将中国耕地非粮化的程度进行进一步的控制。

结合对于中国耕地非粮化率的成因分析，我们可以进一步对不同地区耕地非粮化空间格局做出如下论述：

**东部地区：**

东部地区以其经济发达、城镇化程度高、土地资源相对紧张而著称。在这一地区，非粮化的主要推动力是经济水平的提高，农业生产逐渐向非粮作物转变。此外，农业机械总动力的增加也在一定程度上推动了非粮化进程，但其贡献相对较低。随着东北向西南的地域延拓，东部地区的非粮化程度逐渐递增，这可能反映了东北地区人口相对稀少、经济水平较低以及城乡收入差距不大的情况，使得大规模机械化种植对经济总收入的贡献更为显著。

**中部地区：**

中部地区人多耕地少，农业劳动力有一定的剩余。这种地区的非粮化主要受到剩余农村劳动力的影响，加深了粮食种植与非粮食种植的结构性变化。虽然经济水平对非粮化仍然起着重要作用，但其影响不及剩余农村劳动力。随着由北到南的地域延拓，中部地区的非粮化程度逐渐增大，这可能反映了中部地区人口相对密集，由此导致农业生产者劳动力的结构影响显著。

**西部地区：**

西部地区地理环境复杂,人口分布较为集中,但整体意义上的经济发展呈现较低水平上的复杂性，城乡差距较大。在这一地区，非粮化的主要推动力是经济水平。与东部和中部地区不同的是，西部地区的非粮化受到农村剩余劳动力的影响较小。随着由东北向西北和东南进行地理延拓，西部地区的非粮化程度逐渐有所降低，这可能反映了该地区人口稀少、经济发展水平较低的情况，使得提高农业机械总动力更容易改变当地的农业种植结构。

**东北地区：**

东北地区人口相对稀少，但具有较好的农业生产历史条件和自然条件，经济水平相对较低，城乡收入差距相对较小。在这一地区，非粮化的主要推动力是经济水平和城乡收入差距。东北地区的非粮化受到城乡居民可支配收入比、农业机械总动力和乡村劳动力人数的影响。东北地区的非粮化趋势在空间上呈现出南北差异，这可能反映了该地区乡村劳动力对非粮化的影响作用更大，中部地区人口密集，农业劳动力对非粮化的影响更显著。

## 中国耕地“非粮化”的空间尺度演变趋势

结合中国耕地非粮化的驱动因子和成因机理分析，基于省域统计数据，利用ArcGIS 10.8做出如下中国耕地非粮化空间分析趋势演变示意图：

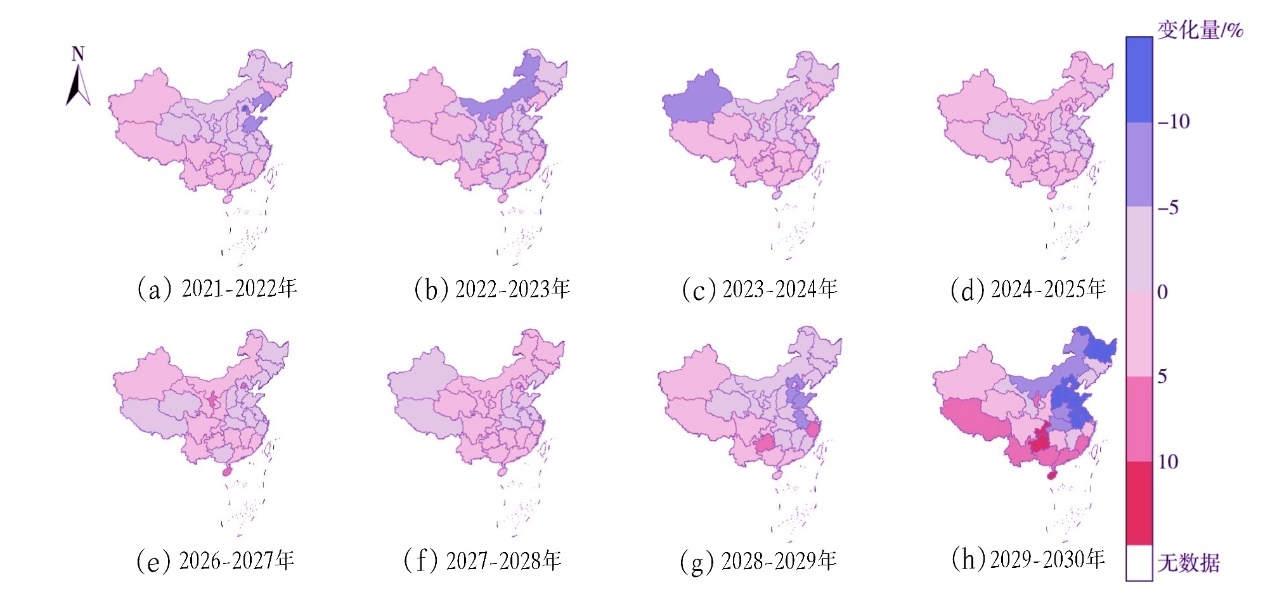


图 6‑11未来5年各省非粮化种植率变化空间分布

未来5年中国耕地非粮化空间格局分异特征可能呈现以下趋势：

**空间集聚性增强：**

非粮化现象可能会在某些特定区域进一步集中，将更加突出空间集聚特征。这些区域可能具备发展特色农业或瓜果蔬菜等产业的自然条件优势，从而吸引更多的非粮化种植。

**南北差异持续：**

预计非粮化面积在中南部和北部地区仍将保持相对集中，但高非粮化率区域可能会在新疆、贵州和东南沿海等地进一步凸显，这些地区的非粮化趋势可能更为显著。

**驱动因素复杂化：**

未来5年，影响非粮化空间格局的驱动因素可能更加复杂多样。经济增长、社会发展、政策调控以及资源配置等多重因素将继续相互作用，共同影响非粮化的空间分布和变化。

**尺度差异持续存在：**

省域尺度的非粮化分布和影响因素可能各有特点，需要针对不同尺度制定差异化的管理和调控策略。 需要注意的是，以上预测基于当前的趋势和影响因素分析，而在现实情况下，则需要考虑更多因素，如政策调整、自然灾害等。因此，为了更准确地把握未来非粮化空间格局分异特征，需要持续进行监测和研究，及时应对各种变化和挑战。

# 问题三模型的建立与求解

## 数据来源与预处理

为构建研究中国种业危机预警统计指标体系，本文数据参考了中国种字研究数据库[[1]](#footnote-1)、Resset行业数据库[[2]](#footnote-2)、国家数据[[3]](#footnote-3)等国内权威数据库，部分数据使用专利之星、知网当学术平台通过关键字检索得到。

## 中国种业危机预警统计指标构建

中国种业危机涵盖范围广阔，涉及领域丰富。在我国正处于科学技术创新的黄金时期的背景下，自主研发与创新能力的持续增强为种业带来了新的机遇与挑战。因此，对于种业危机的全面评价[2]，必须采取多层面、多维度的分析方法。

本研究在深入剖析现有学术文献与行业报告的基础上，围绕粮食安全、种子质量、市场稳定性及行业健康发展四大核心领域，为了及时发现并应对种业发展中可能出现的危机,本研究着手构建了一套完备的种业危机预警统计指标体系。该指标框架旨在全面监测种业运行的各个环节。其中，粮食安全维度着重考量种子的供应能力，以满足我国人民对粮食不断攀升的需求；种子质量维度则聚焦于育种技术的发展水平，评估种子在育种过程中的整体表现；市场稳定性维度强调种子品种与播种质量，要求种子具有高纯度、饱满完整、无病虫害且活力旺盛的特性；而行业健康发展维度则侧重于评估中国种业市场的整体健康状态。

基于上述四个关键维度，本研究设计了一个全面的预警指标体系，目的在于对中国种业的发展现状和潜在风险进行系统性监测与评估。如表7-1所示，这一指标体系为相关决策部门提供了坚实的数据支撑和决策参考

表 7‑1预警指标体系

|  |  |  |
| --- | --- | --- |
| **一级指标** | **表征指标** | **属性** |
| 数量安全A | A1:种业领域内不同种类植物种粮类别的数量 | + |
| 要求种子的供应满足我国人民日益增长的粮食需要 | A2:种业领域内粮食作物的产量 | + |
| A3:我国农作物种子进口情况 | - |
| 质量安全B | B1:种业专利发表量:IPC分类号为A01的专利年度发表量 | + |
| 中游的育种方面 | B2:种业领域内的学术研究论文发表量 | + |
| 市场安全C | C1:中国转基因种子种植面积及同比情况 | + |
| 要求种子纯度高、饱满完整、健全无病虫、活力强 | C2:中国转基因作物种子单产 | + |
| C3:第一产业从业者规模 | + |
|  | C4:农作物种植的土地面积 | + |
| 行业安全D | D1:种业市场的总体规模和容量 | + |
| 中国种业市场行业健康状态 |  | + |

## 基于熵权法的Topsis综合评价法

为了提高综合评价的客观性和科学性,本研究将熵权法与TOPSIS(技术于排序参考理想解的相似性)相结合,构建了一种新的评价模型。传统TOPSIS方法虽然在分析权重的复杂运动时具有较低水平意义下的便捷性,但在确定指标权重时存在一定的主观随意性。而熵权法则能够根据数据的波动情况,自动计算出各指标的权重值,避免了主观赋权的缺陷。因此,我们将熵权法与TOPSIS模型进行精确化联合分析,利用其科学合理的赋权结果,替代了原有的主观权重设置。熵权法的应用增强了指标权重的客观性和准确性。如图7-1所示，这一过程为中国种业危机预警统计指标体系的构建提供了清晰的框架和方法论。

由于数据的量纲不同，可能使得数据值比较大的数据，权重占比比较大，为了确保评价结果的客观性和可比性，本研究采用了极差标准化方法对各项指标进行了必要的数据预处理。该方法通过消除量纲差异和统一数值范围，使得不同指标能够在一个共同的基础上进行比较。通过极差标准化，能够将原始数据转换到[0, 1]的标准化区间内，原来的逆向指标将被化为正向指标，并对结果起到正向作用。



图 7‑1中国种业危机预警统计指标体系建立过程

令第i个数据对象的第j个指标为，经过归一化处理后的数值为，具体过程及公式如下。

1. **归一化过程**

正向指标：

(7.1)

逆向指标：

(7.2)

1. **熵值定义**

(7.3)

第j项指标熵值计算方式：

(7.4)

1. **权重系数**

第j项指标权重值：

(7.5)

(7.6)

## 指标权重的求解与计算

根据熵权法将指标数据归一化后的全部指标产出型权重系数，见表 7‑2。

表 7‑2种业危机预警指标权重系数表

|  |  |  |  |
| --- | --- | --- | --- |
| **指标** | **熵** | **差异性系数** | **权重** |
| **不同种类植物种粮类别的数量** | 0.9701 | 0.0299 | 0.1245 |
| **粮食作物的产量(万吨)** | 0.9991 | 0.0009 | 0.0039 |
| **农作物种子进口量(吨)** | 0.8792 | 0.1208 | 0.5033 |
| **种业行业专利发表数** | 0.9315 | 0.0685 | 0.2855 |
| **期刊论文发布量** | 0.9928 | 0.0072 | 0.0300 |
| **转基因种子种植面积及同比情况** | 0.9959 | 0.0041 | 0.0171 |
| **转基因作物种子单产** | 0.9997 | 0.0003 | 0.0011 |
| **第一产业从业者人数** | 0.9951 | 0.0049 | 0.0202 |
| **农作物种植的土地面积** | 0.9999 | 0.0001 | 0.0004 |
| **种业市场的总体规模和容量** | 0.9967 | 0.0033 | 0.0139 |

根据前文计算出的指标系数权重，结合指标数据计算出的中国种业危机评价指数得到的评价结果如表 7‑3：

表 7‑3中国种业危机评价指数

|  |  |
| --- | --- |
| **年份** | **中国种业危机评价指数** |
| **2010** | 7717.74 |
| **2011** | 8515.23 |
| **2012** | 9390.76 |
| **2013** | 11027.83 |
| **2014** | 12662.38 |
| **2015** | 16038.01 |
| **2016** | 17640.61 |
| **2017** | 19612.63 |
| **2018** | 52003.02 |
| **2019** | 47154.16 |
| **2020** | 58750.05 |
| **2021** | 47899.04 |
| **2022** | 55925.18 |
| **2023** | 72244.90 |

总体来说，我国种业安全呈现较好趋势，如图 7‑2所示。由数据可知，2010-2018年我国种业危机评价指数不断上升，意味着种业领域发展态势不断向好；但在2019年出现了显著下降，且2018-2021年出现波动现象，新冠疫情的爆发给我国农业和食品行业带来了极大波动，粮食的生产、消费、流通与储备均受到冲击。由此可见，我国粮食系统在遭遇外部冲击时的风险抵御能力，以及在遭受损害后的自我恢复能力有待提高，构建一个自主可控且能够有效防范、应对风险冲击的创新性农业系统具有重大意义。近年来，国家不断出台农业相关政策，2022年开始，我国种业邻域状况持续回升，2023年达到近10年来的顶峰。

图 7‑2中国种业危机评价指数折线图

# 问题四的研究内容

我们通过前文的模型建立与求解，首先定量刻画了气候生态与粮食生产之间的因果联系，除了评估当前气候复杂化运动对农业生产的内容和结构的影响外,本研究还着眼于未来气候变化给农业带来的潜在冲击；其次研究了中国耕地“非粮化”的动因与机理，定量描述了中国种粮耕地变化的时空特征与演变趋势；最后识别了粮食种业“卡脖子”风险的主要影响因素，构建种业危机预警统计指标体系与破解路径。通过上述研究提出以下“三位三源四体”建议。

## 建构政府、企业、专业人员和农民四大主体的协同作用机制

政府起主导作用，**加强公众气象服务**，帮助农民在气候变化冲击下降低损失；**规范土地流转制度，预防土地“非粮化”**；引导企业主动创新发展模式,推进粮食产业的转型升级。一方面,要鼓励企业加大粮食储备力度,完善储存设施,确保应急时的供给能力。另一方面,还需着力于优化粮食产业链的中后端环节,提高粮食深加工的附加值，满足多元化的市场需求，推动设备数字化转型；企业起筑底保障作用，积极构建调控保障机制，**提高对于市场异常波动下粮食安全的韧性指数**，为农民生活提供物质保障；专业技术人员起带动作用，积极响应国家政策，加大农村地区的技术推广力度，建立完善的农业咨询服务体系，解决农业生产技术难题，展示示范先进实用技术，**提高科学种粮水平**；农民起主力军作用，在政府的补贴下、企业的帮扶下、专业技术人员的指导下，**提高种植的积极性和主动性，多种粮，种好粮**。

## 创新产业、教育和政策三个方位的运营模式

第一方位需强化粮食产业链建设，优化供产销关系，第二方位需注重公众价值观的培养，加强技术教育，第三方位需完善优粮优价市场购销与应急保障机制，提升农民生产积极性。

## 健全环境资源、经济资源和技术资源三种资源的维护体系

环境资源包括耕地资源、水资源和气候资源等，合理分配耕地资源，解决农田和水污染问题，从根本上解决粮食数量和质量问题；经济资源包括政府补贴资源、企业支持资源以及贸易盈利资源，落实强农惠农富农政策，提升农民生产积极性；技术资源包括种植技术资源，数字化设备资源以及储备、运输和加工技术资源，在此基础上提高粮食产量与优化粮食产品质量。

# 模型的评价与推广

## 模型的优点

OLS模型：计算简单，能很好的处理线性关系

VAR模型：通俗易懂，能轻松的对风险进行评估

偏最小二乘回归模型：对样本的需求较少

基于熵权的 TOPSIS 综合评价法：能够考虑各属性之间的权重，是评价结果更有效可靠

## 模型的缺点

OLS模型：无法计算具有非线性关系的变量

VAR模型：低估了极端事件发生的概率

偏最小二乘回归模型：容易受到异常值的影响

基于熵权的 TOPSIS 综合评价法：需对指标进行清洗，数据处理较麻烦

## 模型的应用

熵权TOPSIS综合评价模型是一种创新的多属性决策分析方法,它巧妙地融合了熵权法和TOPSIS(技术于排序参考理想解的相似性)技术的优势。传统TOPSIS方法在确定属性权重时存在一定的主观随意性,而熵权法则能够基于数据的熵值信息,自动计算出各属性的客观权重。通过将熵权法与TOPSIS模型进行高效率的联合分析,不仅解决了权重赋值的主观性问题,更能有效考虑属性间的相关关系,使决策结果更加准确可靠。该模型的创新之处,正在于有机结合了两种优秀方法的长处,弥补了单一方法的不足,从而提高了多属性决策的科学性和合理性。在诸多实践应用中,熵权TOPSIS模型已展现出其卓越的决策支持能力。在智能制造技术创新能力评价、粮食平房仓架空层评价、农业绿色发展等多领域有着重大作用

# 参考文献

1. 陈浮,刘俊娜,常媛媛,等.中国耕地非粮化空间格局分异及驱动机制[J].中国土地科学,2021,35(09):33-43.
2. 韩洁,高道明,田志宏.中国农作物种子进出口贸易状况分析[J].世界农业,2015,(11):5-10.DOI:10.13856/j.cn11-1097/s.2015.11.002.
3. 孟菲,谭永忠,陈航,等.中国耕地“非粮化”的时空格局演变及其影响因素[J].中国土地科学,2022,36(01):97-106.

# 代码附件

**第一题**

**数据预处理代码**

1. #异常识别
2. **import** pandas as pd
4. # 1. 检测缺失值
5. # 加载数据
6. data = pd.read\_excel("数据.xlsx")
8. # 检测缺失值
9. missing\_values = data.isnull().sum()  # 统计每列的缺失值数量
10. **print**("缺失值统计：\n", missing\_values)
12. # 2. 检测异常值
13. # 以数值列为例
14. numerical\_columns = data.select\_dtypes(include=['int', 'float']).columns
16. # 使用箱线图检测异常值
17. **def** detect\_outliers(data):
18. outliers = pd.DataFrame(columns=['feature', 'number\_of\_outliers'])
19. **for** column **in** numerical\_columns:
20. Q1 = data[column].quantile(0.25)
21. Q3 = data[column].quantile(0.75)
22. IQR = Q3 - Q1
23. lower\_bound = Q1 - 1.5 \* IQR
24. upper\_bound = Q3 + 1.5 \* IQR
25. outliers = outliers.append({'feature': column, 'number\_of\_outliers': data[(data[column] < lower\_bound) | (data[column] > upper\_bound)].shape[0]}, ignore\_index=True)
26. **return** outliers
28. outliers = detect\_outliers(data)
29. **print**("异常值统计：\n", outliers)
30. #异常处理
31. import pandas as pd
32. from sklearn.impute import SimpleImputer
33. from sklearn.preprocessing import RobustScaler
35. # 1. 加载数据
36. data = pd.read\_excel("数据.xlsx")
38. # 2. 检测并替换异常值
39. def replace\_outliers(data):
40. numerical\_columns = data.select\_dtypes(include=['int', 'float']).columns
41. for column in numerical\_columns:
42. Q1 = data[column].quantile(0.25)
43. Q3 = data[column].quantile(0.75)
44. IQR = Q3 - Q1
45. lower\_bound = Q1 - 1.5 \* IQR
46. upper\_bound = Q3 + 1.5 \* IQR
47. data[column] = data[column].mask((data[column] < lower\_bound) | (data[column] > upper\_bound), data[column].median())
48. return data
50. data = replace\_outliers(data)
52. # 3. 使用插值法处理缺失值
53. imputer = SimpleImputer(strategy="linear")  # 这里使用线性插值，也可以选择其他插值方法
54. data\_filled = pd.DataFrame(imputer.fit\_transform(data), columns=data.columns)
56. # 4. 特征缩放
57. scaler = RobustScaler()  # 使用RobustScaler来缩放数据，对异常值更加鲁棒
58. data\_scaled = scaler.fit\_transform(data\_filled)
60. # 将缩放后的数据放回DataFrame
61. data\_filled = pd.DataFrame(data\_scaled, columns=data\_filled.columns)
63. # 输出处理后的数据
64. print(data\_filled.head())

**多元线性回归代码**

1. **import** pandas as pd
2. **import** numpy as np
3. **import** matplotlib.pyplot as plt
4. plt.rcParams['font.sans-serif']=['SimHei'] #用来正常显示中文标签
5. plt.rcParams['axes.unicode\_minus']=False #用来正常显示负号
6. **from** statsmodels.stats.outliers\_influence **import** variance\_inflation\_factor
7. **import** statsmodels.api as sm
9. data = pd.read\_excel("数据.xlsx")
11. # # 加载数据
12. # import random
14. # 提取自变量
15. X = data[['农作物受灾面积', '平均温度', '平均降雨量', '农业机械总动力', '农药使用量', '第一产业就业人员']]
16. # data["平均温度"] = temp
17. # 添加截距项
18. X = sm.add\_constant(X)
20. # 计算VIF
21. vif\_data = pd.DataFrame()
22. vif\_data["Variable"] = X.columns
23. vif\_data["VIF"] = [variance\_inflation\_factor(X.values, i) **for** i **in** range(X.shape[1])]
25. **print**(vif\_data)
27. # 提取自变量和因变量
29. # X = data[['农作物受灾面积', '平均温度', '平均降雨量','农业机械总动力', '农药使用量', '第一产业就业人员']]
30. y = data['粮食产量']
32. # 添加截距项
33. X = sm.add\_constant(X)
35. # 创建线性回归模型
36. model = sm.OLS(y, X)
38. # 拟合模型
39. results = model.fit()
41. # 打印回归结果摘要
42. **print**(results.summary())
43. data
44. data.describe().to\_excel("describe.xlsx")

**数据可视化**

1. **import** pandas as pd
2. **import** matplotlib.pyplot as plt
3. **import** seaborn as sns
5. # 查看数据的前几行
6. **print**(data.head())
8. # 描述性统计
9. **print**(data.describe())
11. # 相关系数矩阵
12. correlation\_matrix = data.corr()
13. **print**(correlation\_matrix)
15. # 可视化数据分布
16. sns.pairplot(data)
17. plt.show()
19. # 可视化单个变量的分布
20. sns.histplot(data['粮食产量'], kde=True)
21. plt.title('Distribution of 粮食产量')
22. plt.show()
24. # 可视化变量之间的关系
25. sns.scatterplot(x='粮食产量', y='农作物受灾面积', data=data)
26. plt.title('Relationship between 粮食产量 and 农作物受灾面积')
27. plt.show()

**VAR模型**

1. **import** pandas as pd
2. **import** numpy as np
3. **from** statsmodels.tsa.api **import** VAR
4. **from** statsmodels.tools.eval\_measures **import** rmse
5. **from** statsmodels.tsa.stattools **import** adfuller
7. # 1. 数据收集和准备
9. # 读取数据
10. data = pd.read\_excel('数据.xlsx')
11. # data = np.log(data)
12. # 对于不平稳的变量进行一阶差分
13. data['农作物受灾面积'] = data['农作物受灾面积'].diff()
14. data['粮食产量'] = data['粮食产量'].diff()
15. data['平均温度'] = data['平均温度'].diff()
16. data['农业总产值'] = data['农业总产值'].diff()
17. data['平均降雨量'] = data['平均降雨量'].diff()
18. data['农业机械总动力'] = data['农业机械总动力'].diff()
19. data['农药使用量'] = data['农药使用量'].diff()
20. data['第一产业就业人员'] = data['第一产业就业人员'].diff()
22. data['粮食产量'] = data['粮食产量'].diff()
23. data['平均温度'] = data['平均温度'].diff()
24. data['农药使用量'] = data['农药使用量'].diff()
25. data['第一产业就业人员'] = data['第一产业就业人员'].diff()
27. data['粮食产量'] = data['粮食产量'].diff()
28. data['农药使用量'] = data['农药使用量'].diff()
30. # # 删除差分后产生的NaN值
31. data.dropna(inplace=True)
33. # 检查时间序列的平稳性
34. **for** column **in** data.columns:
35. result = adfuller(data[column])
36. **print**(f'ADF Statistic for {column}: {result[0]}, p-value: {result[1]}')
38. # 继续对其他变量进行差分操作
40. # 筛选出平稳的变量
41. stable\_variables = ['粮食产量', '农作物受灾面积', '平均温度', '平均降雨量', '农业机械总动力', '农药使用量', '第一产业就业人员']
42. data\_stable = data[stable\_variables]
44. # 3. 模型拟合
45. # 拟合VAR模型
46. model = VAR(data\_stable)
48. # 选择合适的滞后阶数
49. lag\_order = model.select\_order()
51. # 获取最佳的滞后阶数
52. best\_lag = lag\_order.selected\_orders['aic']
54. # 拟合模型
55. results = model.fit(best\_lag)
57. # 4. 模型诊断
58. # 检查残差是否满足白噪声
59. **print**("Ljung-Box test for white noise in residuals:\n", results.test\_whiteness())
61. # 5. 结果解释
62. # 打印模型系数
63. **print**("Model coefficients:\n", results.summary())

66. # 检查模型拟合是否成功
67. **if** results **is** None:
68. **print**("Error: Model fitting failed!")
69. **else**:
70. # 模型拟合成功，打印模型系数
71. **print**("Model coefficients:\n", results.params)
73. # 获取模型的预测值
74. predicted\_values = results.fittedvalues
76. # 获取模型的观测值
77. observed\_values = data\_stable[best\_lag:]
79. # 计算残差的均方根误差
80. **print**("RMSE of residuals:", rmse(observed\_values, predicted\_values))

**第二题**

1. **import** pandas as pd
2. **import** numpy as np
3. **import** math
5. **def** nml(series):  # 正向指标归一化 减最小值的min-max方法
6. normalized\_values = (series - series.min()) / (series.max() - series.min())
7. **return** pd.Series(normalized\_values, name=series.name)
9. **def** nml\_max(series):  # 负向指标归一化
10. normalized\_values = (series.max() - series) / (series.max() - series.min())
11. **return** pd.Series(normalized\_values, name=series.name)
13. **def** nmlzt(df, nml\_nmlmax):  # 归一化函数，对正负向指标分别调用nml()和nml\_max()
14. normalized\_df = pd.DataFrame()
15. **for** column **in** df.columns:
16. **if** nml\_nmlmax[column] > 0:
17. normalized\_df[column] = nml(df[column])
18. **else**:
19. normalized\_df[column] = nml\_max(df[column])
20. **return** normalized\_df
22. **def** pij(df):  # 求信息熵公式中的p，这里直接用取值除以取值总和，而不是数量的比例
23. pij\_values = df.copy()
24. **for** column **in** df.columns:
25. total\_sum = pij\_values[column].sum()
26. pij\_values[column] = pij\_values[column] / total\_sum
27. **return** pij\_values
29. **def** entropy(series):  # 计算信息熵
30. length = len(series)
31. **def** ln(x):
32. **if** x > 0:
33. **return** math.log(x)
34. **else**:
35. **return** 0
36. entropy\_sum = sum([i \* ln(i) **for** i **in** series])
37. **return** -(1 / ln(length)) \* entropy\_sum
39. **def** calculate\_weight(dfij):  # 求e、d、w并返回
40. entropy\_df = dfij.copy()
41. weight\_df = pd.DataFrame(index=entropy\_df.columns, dtype='float64')
42. entropy\_list = [entropy(entropy\_df[column]) **for** column **in** entropy\_df.columns]
43. weight\_df['熵'] = entropy\_list
44. weight\_df['差异性系数'] = 1 - np.array(entropy\_list)
45. total\_sum = weight\_df['差异性系数'].sum()
46. weight\_list = [i / total\_sum **for** i **in** weight\_df['差异性系数']]
47. weight\_df['权重'] = weight\_list
48. **return** weight\_df
50. **def** out(normalized\_df, weight\_df):  # 计算最终结果
51. weights = weight\_df['权重']
52. result = np.matmul(normalized\_df, weights)
53. **return** result

1. <https://r.qiyandata.com/data_center?database=CCAD&sonKey=CCAD_NZDB&part=CCAD_NZDB_CSK&code=CCAD_TSTJ_NUM_NZ2023> [↑](#footnote-ref-1)
2. https://inddb.resset.com/login/login.action?loginName=resset001&loginPwd=resset001 [↑](#footnote-ref-2)
3. https://data.stats.gov.cn/easyquery.htm?cn=C01 [↑](#footnote-ref-3)