

统计学 教学案例

高等教育“十二五”规划教材·经、管、法教学案例

主 编 张焕明



南京大学出版社

统计学 教学案例

高等教育“十二五”规划教材·经、管、法教学案例

主 编 张焕明

副主编 王玉梅 程如轩

 南京大学出版社

图书在版编目(CIP)数据

统计学教学案例 / 张焕明主编. — 南京: 南京大学出版社, 2012. 8

高等教育“十二五”规划教材. 经、管、法教学案例

ISBN 978-7-305-10183-0

I. ①统… II. ①张… III. ①统计学—案例—高等学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2012)第 153100 号

出版者 南京大学出版社
社 址 南京市汉口路 22 号 邮编 210093
网 址 <http://www.NjupCo.com>
出版人 左 健

丛 书 名 高等教育“十二五”规划教材·经、管、法教学案例
书 名 统计学教学案例
主 编 张焕明
责任编辑 陈家霞 王抗战 编辑热线 025-83596027

照 排 江苏南大印刷厂
印 刷 常州市武进第三印刷有限公司
开 本 787×1092 1/16 印张 11.75 字数 278 千
版 次 2012 年 8 月第 1 版 2012 年 8 月第 1 次印刷
ISBN 978-7-305-10183-0
定 价 25.00 元

发行热线 025-83594756
电子邮箱 Press@NjupCo.com
Sales@NjupCo.com(市场部)

* 版权所有,侵权必究

* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换

《高等教育“十二五”规划教材·经管法教学案例丛书》

编 委 会

主 任：王建刚

副主任：陈阿兴

委 员：王晶晶 石旭斋 任志安

任森春 邢孝兵 张焕明

单学勇 夏万军 程昔武

前 言

案例教学是在教师指导下,根据教学目的,组织学生通过对案例的调查、阅读、思考、分析、讨论和交流等活动,教授他们分析问题的方法,进而提高解决问题的能力,加深对基本原理和概念理解的一种开放式教学方法。案例教学突破了传统教学中教师讲授、学生接受的单项流动教学方式,具有教学的多向互动性、明确目的性、客观真实性、较强综合性、深刻启发性、突出实践性、学生主体性、结果多元化等特点,因此案例教学的使用领域越来越广泛,在许多专业资源建设中发挥重要作用。

本系列教学案例遵循“精选案例、内容广泛、深入浅出、注重能力培养”的宗旨,包括经济学教学案例、金融保险教学案例、贸易与商务教学案例、财政与公共管理教学案例、工商管理教学案例、财会审计教学案例、法学教学案例、统计学教学案例等教材。在案例的选取和深度的把握上,体现以下几方面特色:

1. 本系列教材主要面向高校经、管、法专业学生,使学生通过教材的学习,掌握基本理论知识,并提高分析问题和解决实际问题能力。
2. 本系列教材由安徽财经大学一批优秀教师精心编写,编写思路清晰、深入浅出、内容广泛。
3. 本系列教材中的案例均是精选出的典型案例,具有代表性,并且紧跟时代发展,选取具有时代性的案例。
4. 本系列教材力图彰显实用原则,启发学生自主学习。学生通过对案例的钻研和探讨,发挥主观能动作用,增强消化和运用知识与经验的能力,增强解决实际问题的本领。

总之,本系列教材依据学生培养方案,把激发学生兴趣,培养学生主体性、主动性、自主性,提高学生分析问题和解决实际问题能力及提高教师的业务水平作为教材的编写初衷。它有利于贯彻分类指导、因材施教的原则,有利于实践新的教学模式和推进教学方法、手段的创新,推动学生个性化学习,满足不同专业的发展需要。

《高等教育“十二五”规划教材·经、管、法教学案例丛书》编委会

2012年6月

目 录

案例 1 安徽省蔬菜价格波动及影响因素分析	1
一、模型的相关理论知识	1
二、案例分析	6
案例 2 我国节能减排效率的区域差异及变化研究	22
一、模型的相关理论知识	22
二、案例分析	24
附录	42
案例 3 安徽省粮食安全问题研究	47
一、模型的相关理论知识	47
二、案例分析	50
案例 4 中国地区经济增长收敛性分位数回归分析	69
一、模型的相关理论知识	69
二、案例分析	70
附录	82
案例 5 自然灾害对经济增长影响的实证研究	87
一、模型的相关理论知识	87
二、案例分析	88
案例 6 基于核密度估计的上证 A 股收益率分析	97
一、模型的相关理论知识	97
二、案例分析	99
附录	104

案例 7 社区信息化服务满意度模型研究	106
一、模型的相关理论知识	106
二、案例分析	109
附录	121
案例 8 基于稳健主成分回归的统计数据可靠性评估方法	
——以 2008 年我国地区经济增长横截面数据为例	124
一、模型的相关理论知识	124
二、案例分析	127
案例 9 农民工家庭贫困水平:模糊收入线测度及代际传递性原因	135
一、模型的相关理论知识	135
二、案例分析	138
案例 10 中国新兴生物企业的生产效率及其不确定性	
——基于 DEA 和神经网络模拟的面板数据分析	150
一、模型的相关理论知识	150
二、案例分析	153
案例 11 高技术产业研发创新效率与全要素生产率增长	160
一、模型的相关理论知识	160
二、案例分析	161
案例 12 新兴产业与中国产业结构优化升级有序度研究	171
一、模型的相关理论知识	171
二、案例分析	173
后 记	179



安徽省蔬菜价格波动及影响因素分析

一、模型的相关理论知识

(一) 提出问题

蔬菜是生活中不可或缺的必需品,其价格波动水平自然时刻受到居民及政府主管部门的高度关注。因此,正确预测蔬菜价格的未来趋势成为诸多学者的研究兴趣。其基本的研究思路是,先将价格趋势分解为长期、周期、循环和不规则等成分;其次分析影响各成分的主要因素,基于正确预测各成分的变动从而正确预测价格变动趋势,呈现出总一分一总的特点。具体用到的统计方法包括时间序列的周期分解及动态时间序列的建模等。

(二) 平稳性检验

1. ADF 检验

若一个时间序列的均值或自协方差函数随时间而改变,则该序列为非平稳序列(即序列含有某种变动趋势)。对于随机过程 $\{y_t\}$,若 $y_t = \rho y_{t-1} + \epsilon_t$,其中 $\rho = 1$; $\{\epsilon_t\}$ 为一稳定过程,且 $E(\epsilon_t) = 0$, $\text{cov}(\epsilon_t, \epsilon_{t-1}) = u_t < \infty$,这里 $t = 0, 1, 2, \dots$,则称该过程为单位根过程。

ADF 检验(augmented dickey-fuller test)方法通过在回归方程右边加入因变量 y_t 的滞后差分项来控制高阶序列相关。

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \mu_t \quad t = 1, 2, \dots, T \quad (1-1)$$

$$\Delta y_t = \gamma y_{t-1} + a + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \mu_t \quad t = 1, 2, \dots, T \quad (1-2)$$

$$\Delta y_t = \gamma y_{t-1} + a + \delta t + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \mu_t \quad t = 1, 2, \dots, T \quad (1-3)$$

扩展定义将检验。

$$\begin{cases} H_0: \gamma = 0 \\ H_1: \gamma < 0 \end{cases} \quad (1-4)$$

也就是说,原假设为序列存在一个单位根;备择假设为不存在单位根序列 y_t 可能还包

含常数项和时间趋势项。

2. KPSS 检验

上面介绍的检验方法,零假设都是非平稳的。根据检验的一般原理,零假设是受到保护的一个假设,很难推翻。所以实证分析将许多 $I(0)$ 时间序列当成是 $I(1)$ 的。KPSS 首次提出以平稳为零假设。

分情况估计模型:

$$\text{情况 1:} \quad Y_t = \alpha + u_t$$

$$\text{情况 2:} \quad Y_t = \alpha + \delta t + u_t$$

计算残差用 e_t 表示。定义残差的部分和过程为:

$$S_t = \sum_{i=1}^t e_i \quad t = 1, 2, \dots, T \quad (1-5)$$

LM 统计检验量是:

$$LM = \frac{\sum_{t=1}^T S_t}{\hat{\sigma}_u^2} \quad (1-6)$$

其中

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^T e_i^2}{T}$$

查临界值表并判断。

(三) Granger 因果检验

在经济分析中经常要分辨谁是因,谁是果。要想验证因果关系,可以利用的只能是过去和现在的数据,一个广为使用的方法是 Granger 意义上的因果检验。

对变量进行 Granger 因果检验前一般先检验两者的平稳性,只有两个平稳的变量或协整的变量才可以进行 Granger 因果检验。对于不平稳的变量要对其进行适当变换,如进行差分变换等,使其满足平稳性条件再进行 Granger 因果检验。

Granger 因果关系检验的思路是,如果两个经济变量 x_t 与 y_t 在同时包含过去 x_t 与 y_t 信息的条件下,对 y_t 的预测效果比只单独由 y_t 的过去信息对 y_t 的预测效果更好,即变量 x_t 有助于变量 y_t 预测精度的改善,则认为 x_t 对 y_t 存在 Granger 因果关系。

检验 x_t 是否为引起 y_t 变化的原因的过程如下:

首先,检验“ x_t 不是引起 y_t 变化的原因”的原假设,对下面两个模型进行估计:

$$\text{无约束模型:} \quad y_t = \sum_{i=1}^s \alpha_i y_{t-i} + \epsilon_{1t} \quad (1-7)$$

$$\text{有约束模型: } y_t = \sum_{i=1}^s \alpha_i y_{t-i} + \sum_{i=1}^k \beta_i x_{t-i} + \varepsilon_{2t} \quad (1-8)$$

然后,检验“ y_t 不是引起 x_t 变化的原因”的原假设,交换 x_t 与 y_t ,做同样的回归估计,检验 y_t 的滞后项是不是显著的不为 0。要得到 y_t 是引起 x_t 变化的结论,我们必须拒绝“ y_t 不是引起 x_t 变化的原因”的原假设,同时接受“ y_t 是引起 x_t 变化的原因”的备择假设。

Granger 因果关系检验实质上是检验一个变量的滞后变量是否可以引入到其他变量方程中。一个变量如果受到其他变量的影响,则称它们具有 Granger 因果关系。

(四) B-N 分解模型

B-N 分解的目的是将序列分解为确定性趋势 DT_t 、随机性趋势 ST_t 与周期成分 C_t 之和,即:

$$y_t = DT_t + ST_t + C_t \quad (1-9)$$

其核心理论认为趋势成分保留于水平条件期望之中,通过累计使无限远期的水平条件期望形成趋势成分,而周期成分仅具有短期效应不包括于水平条件期望中,于是可以基于无限远期的水平条件期望来分解周期和随机趋势。

B-N 分解步骤如下: 第一步,确定序列是否趋势平稳,如果序列不是趋势平稳过程,而具有一阶单整的特征,则先从序列中分离出确定性趋势。由 y_t 为 $I(1)$ 可得 Δy_t 为 $I(0)$,根据 Wold 表述定理, Δy_t 可以表述为 ARMA 过程,而 ARMA 过程可以用 $MA(\infty)$ 表示,于是:

$$\Delta y_t = \mu + \phi(L)\varepsilon_t = \mu + \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \cdots \quad (1-10)$$

其中, μ 为 Δy_t 的长期均值, $\varepsilon_t \sim \text{iid}(0, \sigma^2)$, 按 B-N 分解理论, y_t 中的确定性趋势为:

$$DT_t = y_0 + \mu_t \quad (1-11)$$

其中, y_0 为初始时间值。

第二步,确定周期成分。首先,定义长期的水平条件期望以表示总趋势,即确定性趋势与随机趋势之和。记 $\hat{y}_t(j) = E(y_{t+k} | \cdots, y_{t-1}, y_t)$ 为基于当期(t 期)的所有信息为条件的未来 k 期(y_{t+k})的条件期望值,用 Δy_t 表示 y_t ,可以得到:

$$\hat{y}_t(k) = E(y_{t+k} | \cdots, y_{t-1}, y_t) = y_t + \Delta \hat{y}_t(1) + \cdots + \Delta \hat{y}_t(k) \quad (1-12)$$

将式(1-2)代入式(1-4)可得:

$$\hat{y}_t(k) = k\mu + y_t + \left(\sum_1^k \lambda_i\right)\varepsilon_t + \left(\sum_2^{k+1} \lambda_i\right)\varepsilon_{t-1} + \cdots \quad (1-13)$$

令 k 趋于无穷大,由此得到基于 t 期信息的长期的条件期望 $\lim_{k \rightarrow \infty} \hat{y}_t(k)$, 而式(1-13)右边的 $k\mu$ 为第 t 期至 $k+t(k \rightarrow \infty)$ 期的确定性趋势的增量,所以将其扣除后的剩余部分即为当期的总趋势成分,即:

$$\lim_{k \rightarrow \infty} (\hat{y}_t(k) - k\mu) = y_t + \left[\left(\sum_1^{\infty} \lambda_i \right) \epsilon_t + \left(\sum_2^{\infty} \lambda_i \right) \epsilon_{t-1} + \cdots \right] \quad (1-14)$$

由于周期定义为现期值与当期趋势值之差,故周期定义为:

$$C_t = - \left[\left(\sum_1^{\infty} \lambda_i \right) \epsilon_t + \left(\sum_2^{\infty} \lambda_i \right) \epsilon_{t-1} + \cdots \right] \quad (1-15)$$

因该式包含无穷项,在实际计算上存在困难,Newbold(1990)改进了上述分解方法,将式(1-10)带入式(1-15),将周期成分等价:

$$C_t = - \lim_{k \rightarrow \infty} [(\Delta \hat{y}_t(1) - \mu) + (\Delta \hat{y}_t(2) - \mu) + \cdots + (\Delta \hat{y}_t(k) - \mu)] \quad (1-16)$$

令 $w_t = \Delta y_t - \mu$, 有 $w_t: I(0)$, 由 Wold 表述定理, w_t 可以表述为零均值的 ARMA(p, q) 过程。估计出 $\hat{w}_t(j)$, 将其代入式(1-16), 得到周期表达式为:

$$C_t = - \lim_{k \rightarrow \infty} \left[\sum_{j=1}^k \hat{w}_t(j) \right] = - \left[\sum_{j=1}^q \hat{w}_t(j) + \sum_{j=q+1}^{\infty} \hat{w}_t(j) \right] \quad (1-17)$$

Newbold 巧妙地解决了式(1-17)中的“无穷和”, 得到最终周期计算公式为:

$$C_t = - \left[\sum_{j=1}^4 \hat{w}_t(j) + \varphi \hat{w}_t(4) / (1 - \varphi) \right] \quad (1-18)$$

其中 φ 为 w_t 的 AR 项系数。最后由式(1-9)得到随机趋势为:

$$ST_t = y_t - DT_t - C_t$$

(五) VAR 模型

自向量回归(VAR)是基于数据的统计性质建立模型, VAR 模型将系统中每一个内生变量作为系统中所有内生变量滞后值的函数来构造模型, 从而将单变量自回归模型推广到由多元时间序列变量组成的向量自回归模型。

VAR(p)模型的数学表达式是:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + Bx_t + \epsilon_t \quad t = 1, 2, \cdots, T \quad (1-19)$$

其中, y_t 是 k 维内生变量向量; x_t 是 d 维外生变量向量; p 是滞后阶数; T 是样本个数; $k \times k$ 维矩阵 $A_1, \cdots, A_p$ 和 $k \times d$ 维矩阵 B 是要被估计的系数矩阵; ϵ_t 是 k 维扰动向量, 它们相互之间可以周期相关, 但不与自己的滞后值相关及不与等式右边的变量相关。假设 Σ 是 ϵ_t 的协方差阵, 是一个 $k \times k$ 的正定矩阵, 式(1-19)可以展开表示为:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{bmatrix} = A_1 \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ \vdots \\ y_{kt-1} \end{bmatrix} + A_p \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \\ \vdots \\ y_{kt-p} \end{bmatrix} + \cdots + B \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{dt} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{kt} \end{bmatrix} \quad t = 1, 2, \cdots, T$$

即含有 k 个时间序列变量的 VAR(p)模型由 k 个方程组成。

(六) 脉冲响应函数

在实际应用中,由于 VAR 模型是一种非理论性的模型,它无须对变量作任何先验性的约束。因此在分析 VAR 模型时,往往不分析一个变量的变化对另一个变量的影响如何,而是分析当一个误差项发生变化,或者说模型受到某种冲击时对系统的动态影响,这种分析方法称为脉冲响应函数分析方法。

用时间序列模型来分析影响关系的一种思路,是考虑扰动项的影响是如何传播到各变量的。下面先根据量变量的 VAR(2) 模型来说明脉冲响应函数的基本思想。

$$\begin{cases} x_t = a_1 x_{t-1} + a_2 x_{t-2} + b_1 z_{t-1} + b_2 z_{t-2} + \epsilon_{1t} \\ z_t = c_1 x_{t-1} + c_2 x_{t-2} + d_1 z_{t-1} + d_2 z_{t-2} + \epsilon_{2t} \end{cases} \quad t = 1, 2, \dots, T \quad (1-20)$$

其中, a_i, b_i, c_i, d_i 是参数, 扰动项 $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})'$ 假设是具有下面这样性质的白噪声向量, 则有:

$$E(\epsilon_t) = 0 \quad \text{对于 } \forall t \quad i = 1, 2$$

$$Var(\epsilon_t) = E(\epsilon_t, \epsilon_t') = \sum = \{\sigma_{ij}\} \quad \text{对于 } \forall t$$

$$E(\epsilon_s, \epsilon_t) = 0 \quad \text{对于 } \forall t \neq s \quad i = 1, 2$$

假定上述系统从 0 期开始活动, 且设 $x_{-1} = x_{-2} = z_{-1} = z_{-2} = 0$, 又设于第 0 期给定了扰动项 $\epsilon_{10} = 1, \epsilon_{20} = 0$, 并且其后均值为 0, 即 $\epsilon_{1t} = \epsilon_{2t} = 0 (t = 1, 2, \dots)$, 称此为第 0 期给 x 以脉冲。下面讨论 x_t 与 z_t 的响应, $t=0$ 时:

$$x_0 = 1, z_0 = 0$$

将其结果代入式(1-20), $t=1$ 时:

$$x_1 = a_1, z_1 = c_1$$

将其结果代入式(1-20), $t=2$ 时:

$$x_2 = a_1^2 + a_2 + b_1 c_1, z_2 = c_1 a_1 + c_2 + d_1 c_1$$

继续这样计算下去, 设所求的结果为:

$$x_0, x_1, x_2, x_3, x_4, \dots$$

以上称为由 x 的脉冲引起的 x 的响应函数。

同样所求得的结果为:

$$z_0, z_1, z_2, z_3, z_4, \dots$$

以上称为由 x 的脉冲引起的 z 的响应函数。

二、案例分析

(一) 案例背景材料

蔬菜是人们日常饮食不可缺少的食品,其在营养均衡以及饮食健康中起着举足轻重的作用。蔬菜对人体的影响是多方面的,人类所需的六大营养素中的纤维素、矿物质和维生素主要来源于蔬菜。作为关系民生的主要消费品,蔬菜跟粮食一样关系着整个国家的生计,而蔬菜价格的高低也影响着每个人的生活水平。

蔬菜业是我国农业的重要组成部分,是关系“三农”问题的重大产业,是提高农民收入,发展农村经济,加快农业发展的基础。据联合国粮农组织(FAO)2007年统计,我国蔬菜播种面积和产量分别占世界的43%和49%,均居世界第一;2007年,我国蔬菜的播种面积达2.5亿亩,总产量5.65亿吨,人均产量420公斤。2009年产量达6.19亿吨。到2010年,我国蔬菜总产量达到7.37亿吨,出口844.6万吨,同比增长5.0%,出口额99.8亿美元,同比增长45.2%;进口15.0万吨,同比增长54.4%,进口额2.8亿美元,同比增长54.6%。贸易顺差97.0亿美元,同比扩大45.0%。我国已经成为名副其实的蔬菜产销大国,蔬菜价格直接影响了农民的收入、农村的稳定以及我国经济的发展。

然而,2007年以来,受外界环境因素的影响,蔬菜价格波动剧烈。2009年以来,CPI迅速上涨,并且一直处于高位难以下降,2011年国家宏观调控的首要目标为稳定物价。在物价飞涨的今天,蔬菜价格问题日益严峻。

要解决蔬菜问题,就必须弄清楚蔬菜价格上涨的本质,即到底是什么因素影响蔬菜价格过快上涨,以及这些因素在实质上对蔬菜价格产生多大的冲击。鉴于我国蔬菜价格问题研究的相对不足,以及不断变化的外部环境,本案例采用B-N分解法,从定量分析的角度出发,以安徽省蔬菜价格为代表,研究蔬菜价格的形成机制,将蔬菜价格分解成确定性趋势、随机性趋势以及周期性趋势。在此基础上,根据VAR模型,分析影响蔬菜价格的主要波动因素。

(二) 研究的现状及存在的问题

在国内,蔬菜价格以及农产品价格一直是备受关注的问题,学术界对其进行了深入研究。赵留彦(2007)经过实证分析认为,农产品价格的过快上涨从根本上是由国内经济过热,农资价格上涨造成的,不是农产品价格上涨导致通货膨胀,而是通货膨胀引起农产品价格的上扬。鲁沂(2007)分析认为,居民收入提高导致的需求拉动,国际农产品价格和国内生产资料价格上涨,以及全球粮食减产和油价上涨是农产品价格上涨的主要因素。薛淑珍、王保忠(2008)认为,经济发展周期变化、需求上升拉动价格上涨、货币供应问题、国内农业生产资料价格持续上涨、全球粮食减产和石油价格飙升这六个方面原因造成我国农产品价格上涨。马晓河(2008)认为,货币溢出效应、经济周期影响、成本推动、国际市场农产品价格传导是农产品价格上涨的主要原因。孔婷婷、许北鸥(2009)通过对农产品价格变动的因素分析认为,一方面,耕地面积总体减少,可耕地减速更快、农村青壮年劳动力转移,导致农业生产投入

下降,劳动生产率增长缓慢;另一方面,国际市场石油和粮食价格的带动影响、生产资料价格上涨,以及天气灾害等导致粮食作物大量减产。邵威豹等(2011)应用脉冲响应函数和方差分解方法对影响我国农产品价格波动的因素进行了实证分析得出,农产品本身价格和通货膨胀对农产品价格存在波动冲击,并且有持久效应。孙倩、穆月英(2011)认为,季节因素、蔬菜供给量和上市成本是影响蔬菜价格的主要因素。

对于影响蔬菜和农产品价格的因素,可谓众说纷纭。从总体上看,影响蔬菜产量的因素主要有天气、种植面积、国家政策、生产技术、市场价格等,而影响其价格波动的主要原因有农业生产成本、通货膨胀水平、货币供应量(市场流通性)、物流价格、极端天气、游资这几个方面。但是前人研究价格波动的文献大都采用定性分析,定量分析运用得较少,更没有系统分析各因素对蔬菜价格的影响程度,仅仅进行定性分析缺少说服力。为了弥补前人研究的不足,本案例首先借助 Beveridge 和 Nelson 分解将价格分解为确定性趋势、周期成分和随机性趋势。在随机性趋势基础上,用向量自回归模型的动态分析、脉冲响应函数和方差分解来研究各个因素对农产品价格的冲击和影响。

(三) 模型构建前的准备

1. 两个假设

本案例从研究的背景以及前辈们对蔬菜价格研究的不足出发,提出下列假设:

- (1) 选定基期后,通过同比增长率数据计算得出的蔬菜价格合理;
- (2) 在 B-N 分解中,确实存在确定性趋势和周期成分;

对这两点的解释是:

首先,由于统计部门并不公布具体的蔬菜价格,而只是发布同比增长率数据,因此我们假定某年为基期(价格取 100,之后对数据进行对数处理),通过月度价格指数计算得出年度价格。如果价格指数是真实可靠的,那么通过这种方法得到的价格合理。在前人的研究中,也采用过这种技术处理方法。

其次,对于某个时间序列变量是否含有确定性趋势,并不能仅仅根据数据的短期特征来判断。尽管对于宏观变量往往在经济快速发展过程中都有一个向上增长的趋势,但在不同阶段往往会有较快增长,而在另一个阶段,增长比较缓慢。因此,我们的假设是蔬菜价格序列属于平稳增长序列,带有确定性趋势。此外,对于周期成分,我们采用“峰峰法”得到周期,假设该方法得到的周期合理。

2. 变量的选择

由于近期蔬菜价格备受关注,因此本案例选择蔬菜价格作为自变量。此外,考虑到安徽是农业大省,在蔬菜种植种类上品种丰富,其产品在当地自销后还往周边地区输出。总体而言,选取安徽省蔬菜价格作为研究对象是合理的。对于影响蔬菜价格因素的考虑,根据大量查阅文献,在总结前人经验的基础上,选取了三个变量,分别为货运价格、热钱和天气。

第一,货运价格考察了流通环节的成本。2009 年以来,我国物流运价快速上涨,货运作作为目前我国市场物品运输的主体,其价格引起了很多商品价格的浮动。由于运输成本最终转嫁给消费者,因此蔬菜购买价格中必然包含了货运成分,也就是货运价格高低会引起蔬菜价格的变动。货运价格在蔬菜价格变动中到底起了多大作用,还需具体分析。

第二,热钱也可能成为蔬菜价格上涨的主要因素。近几年,伴随农产品的上涨,空闲资金炒作的因素不可忽视。从2009年12月份大蒜开始,到2010年的玉米、绿豆、苹果、糖、生姜等,这些农产品价格飞涨的背后,游资炒作是幕后推手,直接带动了整个农产品价格的上涨。国家发改委明确表示,要坚决打击炒作农产品等行为。然而游资的特点是快进快出,获利性强,这两年的蔬菜价格上涨会不会是游资带动的?考虑到我国游资规模难以获取,因此选取了国际热钱规模来代替。

第三,我国最近极端气候接连发生,如2008年初的冰冻,2009年夏季的洪涝,以及2010年上半年西南大旱和2010年底到2011年上半年江淮河流域的大旱等,这些极端天气,往往导致蔬菜产量急剧下降,伴随产量下降的是价格的上升。因此,选取天气作为因变量进行分析。

(四) 数据的来源与量化处理

1. 数据的来源

(1) 安徽省蔬菜价格数据。本案例选取了2006年6月份到2011年3月份安徽省蔬菜价格指数(2005年=100),进行累计价格计算,获得月度价格,取自然对数后,进行季节调整,得到序列 y_t ,见图1-1。数据来源于安徽省统计局。

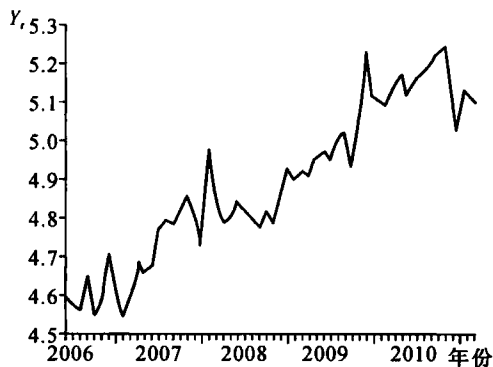


图1-1 安徽省2006年6月份以来蔬菜价格(季节调整后)

(2) 其他数据来源。首先,本案例选取2006年6月开始编制的中国物流与采购行业网站上的货运市场每千米每吨价格数据(见图1-2),在分析时进行标准化处理。

其次,由于游资规模数据难以统计,选取国际上普遍使用的国际热钱规模估计公式(刘洪洋,2008),即热钱=外汇储备增量-贸易顺差-外商直接投资(FDI)。根据外汇储备增量、贸易顺差、外商直接投资(FDI)的月度数据(来源:商务部网站、海关总署),计算出热钱数据,以此代替游资数据,见图1-3。

最后,对于天气因素,一般使用虚拟变量,即在极端天气月份取1,其他月份取0。本案例中对2008年的1、2月份,2009年的6、7月份,2010年10、11月份取1,其他月份取0。

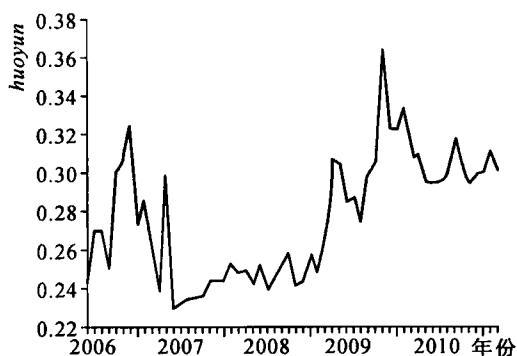


图 1-2 未经标准化前货运市场价格走势图

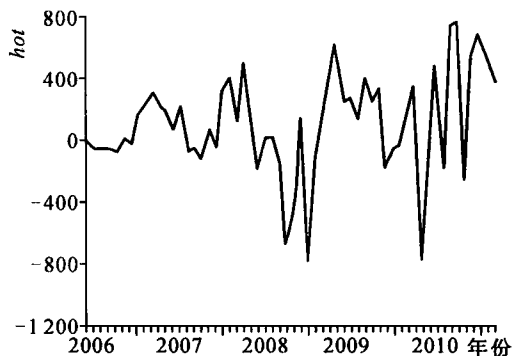


图 1-3 未经标准化前的热钱走势图

2. 选取的观察变量与因素变量之间的关系

从图 1-1 可以看出，观察变量即蔬菜价格在经历了 2006 年较平稳的走势后，2007 年开始快速上涨，并在 2008 年初南方冰冻自然灾害影响下创出了历史最高水平。这说明了天气因素很有可能是导致蔬菜价格突然上涨的一个关键因素。2008 年，经济危机影响导致 CPI 和蔬菜价格都有所回落，但是 2009 年我国 4 万亿的投资拉动和物流产业振兴规划，使物流价格快速上涨（见图 1-2），引起 2009 年末蔬菜价格的暴涨（见图 1-1）。2010 年后，蔬菜价格开始震荡走势，货运价格也有所下降，因此猜测两者之间有因果关系存在。

由图 1-3 可以看出，流入我国的热钱规模波动比较剧烈，2008 年前有较小规模的热钱流入，截至 2007 年底，2003 年到 2007 年热钱规模净流入 3 121 亿美元。到 2008 年，由于金融危机的影响，热钱开始大规模流出，2009 年后又存在回流现象。由于流入我国的热钱流向尚不明确，如 2003—2004 年，中国房地产业正处于膨胀期，房地产业年收益率高达 50%，境外热钱大量涌入房地产市场。2005—2006 年，完成了股改的中国股市开始了一波声势浩大的牛市，海外热钱开始转战 A 股市场。2008 年下半年，美国出现次贷危机，流动性不足，热钱开始逐步撤离中国市场。2009 年底，人民币激流勇进式的快速上涨又将部分热钱从危机影响稍缓的美国吸引回了中国，并且一直延续至今。因此相比较于货运价格和天气因素，国际热钱规模对我国蔬菜价格的影响不是很确切，但是并不能否定游资炒作导致蔬菜价格上涨的猜测。

（五）蔬菜价格的趋势周期分解

1. 数据特征检验

对数据进行 B-N 分解之前，必须确保数据序列为趋势平稳过程，这直接关系到周期分解的方法和结论。如果 y_t 序列为趋势平稳过程，早期的周期分解是直接从该序列中退化趋势，即将 y_t 对常数和趋势回归而形成线性趋势，而退除趋势后的剩余部分（即回归残差）就是周期。也就是说，当期的回归残差就是短期的周期成分。显然，这种分解周期和趋势的方法潜在地排除了 y_t 中的随机趋势。因此，对 y_t 进行平稳性检验，分别使用 KPSS 趋势平稳与 ADF 单位根检验，结果见表 1-1。

表 1-1 y_t 的趋势平稳检验和单位根检验

检验方法	检验方程形式	原假设	统计量	临界	结论
KPSS	$y_t = 4.5608 + 0.0115t + \varepsilon_t$	趋势平稳	0.1574	0.1460	否
ADF	$y_t = 0.2978 - 0.0572y_{t-1} - 0.3878\Delta y_{t-1} - 0.4701\Delta y_{t-2} + \varepsilon_t$	单位根	-1.2676	-2.8600	是

注：表中的临界值均为 5% 显著性水平下的临界值。

由表 1-1 可知， y_t 序列不是趋势平稳过程，是非平稳的 $I(1)$ 过程，这一结果表明 y_t 序列中包含随机趋势。KPSS 检验方程中的截距项显著不为 0，说明 y_t 为带截距项的 $I(1)$ 过程，表明序列中存在确定性趋势。ADF 检验方程中的差分滞后项代表一种短期的调整行为，暗含了 y_t 序列中还存在平稳的随机过程，即真实的周期成分。因此要计算真实周期，需要将随机趋势和确定性趋势从 y_t 序列中分解出来。

综上所述， y_t 序列不是趋势平稳过程，是带截距项的 $I(1)$ 过程，可以对其进行趋势周期分解。

2. 蔬菜价格的确定性趋势

根据 B-N 分解的思路框架，样本 y_t 序列的初始值 $y_0 = 4.5923$ ， $\mu = 0.0090$ ，根据式 (1-3) 计算确定性趋势 DT_t ，结果见图 1-4。为了便于分析，将计算的确定性趋势季节数据 DT_t 转换为年度数据 NDT_t ，计算其年度增长率。确定性趋势的年度值占年度实际价格序列 NYT_t 的百分比记为 S_t ，结果见表 1-2。

根据本案例的分解和计算结果可知，安徽省蔬菜价格确定性趋势的年均月增长速度达到 0.9%，度量波动的标准差为 0.0754。这一结果说明，近几年，安徽省蔬菜价格呈现平稳的递增型确定性趋势，表明了蔬菜在居民消费中占重要的地位。由于蔬菜价格还受各类随机冲击的影响，确定性趋势大于或者小于实际价格，表现为 S_t 围绕 100% 上下波动。当 $S_t > 100\%$ ，说明随机冲击对价格有抑制作用，而最近几年出现 $S_t < 100\%$ ，则说明外部冲击促进了价格的上涨。从表 1-2 可以看出，2006 年后，确定性趋势所占百分数都小于 1，说明外部的冲击促进了价格的上涨，这跟这几年的外部环境有极大关系，如 2008 年的金融危机和汶川地震、2009 年的 4 万亿投资以及 2010 年的极端天气等。

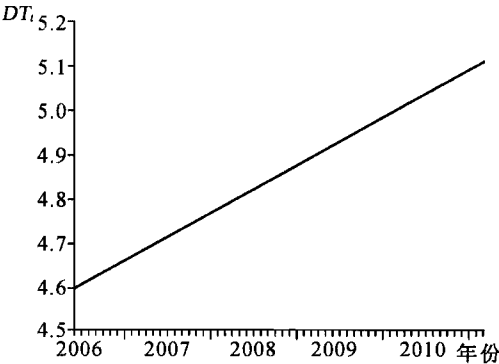


图 1-4 蔬菜价格的确定性趋势