# 小样本协整系统的检验[①]

杨宝臣, 王 艳, 张世英
(天津大学管理学院, 天津 300072)

摘要: 采用Monte Carlo 模拟方法, 给出了小样本容量下协整检验临界值的响应面(response surface)方程. 结果表明小样本协整检验的临界值不仅与样本容量有关, 还依赖于协整检验中的滞后阶数. 最后对中国城镇居民的食品消费进行了小样本协整分析.

关键词: 协整检验; 响应面方法; 小样本
中图分类号: O213 文献标识码: A 文章编号: 1007-9807(2002)01-0070-06

## 0 引 言

协整理论的研究大多数都是建立在大样本基础上的, 模型的建立普遍基于月份或季度数据. 文[1-5]等给出的协整关系的检验都是基于大样本的情形. 小样本下经济变量间协整关系的检验, 尤其是涉及到数据量较少的年度数据时, 就很薄弱了. 其主要原因是缺少用于小样本协整检验统计量的临界值, 而且相应的协整检验的势是未知的.

文[6]研究了小样本协整问题, 通过Monte Carlo 实验得出结论, 不同的数据生成过程(data generating process, DGP)会得到检验统计量的不同概率分布. 由不同模型生成的数据之间及由它们得出的小样本下统计量概率分布之间的差异是显著的. 文[7]以实际DGP 为基础, 应用靴攀(bootstrap)方法, 得到了单位根检验的临界值及检验的势分布, 其方法只是针对不同的实际模型, 分别进行相应的单位根及协整检验, 没有建立普遍适用的协整检验的临界值表.

文[8]应用响应面方法[9](RSM)得到了EG两步法在有限样本下的临界值, 对于任意样本容量下的临界值都可以通过该响应面方程计算出来. 文[10]也引入RSM, 研究了文[11-13]的迹统计量和最大特征根统计量在有限样本下的临界值, 得出的结论是, 在有限样本条件下, 协整检验的临界值不但与样本容量有关, 而且与模型的滞后阶数, 变量的个数有关.

本文研究小样本下的ADF 协整检验问题. 采用响应面方法, 研究了样本容量 $T$ 和滞后阶数 $k$ 对小样本下协整检验临界值的共同影响. 在研究了两变量协整系统基础上, 对多变量协整系统在小样本下的ADF 检验也进行了模拟研究, 给出了小样本条件下协整检验的响应面方程. 最后对中国城镇居民的食品消费进行了协整的实证分析.

## 1 小样本协整检验

### 1.1 两变量协整检验

先考察小样本下两变量协整系统的ADF 检验. 设DGP 为

$$x_t = x_{t-1} + \epsilon_{1t}, \quad \epsilon_{1t} \overset{iid}{\sim} N(0, 1)$$
$$y_t = y_{t-1} + \epsilon_{2t}, \quad \epsilon_{1t} \overset{iid}{\sim} N(0, 1) \tag{1}$$

在模拟分析中, 样本容量 $T$ 取以下值 $T$ = {15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100}, 滞后阶数 $k$ = {1, 2, 3, 4, 5}, 实验共 18 × 5 = 90 组, 对于每一组 $T$ 与 $k$ 实验分别重复 30 000 次.

令 $x_0 = 0, y_0 = 0$, 为了减少初值条件对数据

① 收稿日期: 2000-05-31; 修订日期: 2001-10-29.
基金项目: 国家自然科学基金资助项目(79400014).
作者简介: 杨宝臣(1966-), 男, 河北人, 博士, 副教授.

生成过程的影响, 采用重复删除法进行了系统模拟, 即每次采样去掉前 50 个样本数据, 而以第 51 个数据作为初始值, 以消除初始状态对数据生成过程的影响. 用 GAUSS 中的 RNDN 产生均值为 0, 标准差为 1 的伪正态随机变量 $\epsilon_{1t}$, $\epsilon_{2t}$ 各 $T + 50$ 个, 经过一阶求和后生成关于 $x_t$ 和 $y_t$ 的时间序列, 其分量分别为 $T + 50$ 个, 除去前 50 个观测值, 结果就得到所需要的时间序列 $\{x_t\}$ 和 $\{y_t\}$.

由式(1) 给出的 DGP 是形式最简单的数据生成过程, 并没有考虑到 $\epsilon_{1t}$ 含多余参数时的情况. 文[8] 考察了当 DGP 中的残差为 AR 或 MA 过程时的小样本 ADF 检验临界值变化情况. 实验结果表明, 只要 $k$ 足够大以消除残差中的自相关, 不同结构的残差得到的结果的差异是很小的.

由式(1) 生成的 $x_t$ 和 $y_t$ 都为一阶单整序列 $I(1)$, 如果 $x_t$ 和 $y_t$ 是协整的, 则存在一个常数 $\alpha$, 使 $z_t = y_t - \alpha x_t$ 为 $I(0)$; 如果它们之间不存在协整关系, 则 $x_t$ 和 $y_t$ 的任意线性组合将仍为一个一阶单整序列 $I(1)$.

协整检验的步骤:

**第 1 步**　估计协整回归方程

$$y_t = \alpha_1 + \alpha_2 x_t + z_t \tag{2}$$

得到估计系数 $\hat{\alpha}_1$ 和 $\hat{\alpha}_2$ 后, 计算 $\hat{z}_t = y_t - \hat{\alpha}_1 - \hat{\alpha}_2 x_t$.

本文采用更一般的 ADF 检验形式

$$\Delta \hat{z}_t = (\rho - 1)\hat{z}_{t-1} + \sum_{i=1}^{k-1} \sigma_i \Delta \hat{z}_{t-i} + u_t \tag{3}$$

$$\Delta \hat{z}_t = \mu + (\rho - 1)\hat{z}_{t-1} + \sum_{i=1}^{k-1} \sigma_i \Delta \hat{z}_{t-i} + u_t \tag{4}$$

$$\Delta \hat{z}_t = \mu + \gamma t + (\rho - 1)\hat{z}_{t-1} + \sum_{i=1}^{k-1} \sigma_i \Delta \hat{z}_{t-i} + u_t \tag{5}$$

其中 $u_t$ 为白噪声. 零假设 $H_0$: $\rho = 1$, 即 $z_t$ 有单位根; 备择假设 $H_1$: $|\rho| < 1$, 即序列为 $I(0)$ 序列或带有确定性趋势的 $I(0)$ 序列. 式(3)、(4)、(5) 三种情况分别为"无常数项和趋势项", "带常数项"及"带常数项和趋势项".

**第 2 步**　用得到的 $z_t$ 序列作为样本按模型(3)、(4) 和(5) 分别去估计值 $\rho$, 并计算出相应的 ADF 检验统计量在 1%, 5% 和 10% 显著性水平上的临界值.

在分析中, 响应变量为小样本下的 ADF 检验的临界值, 控制变量为样本容量 $T$ 和滞后阶数 $k$. 选择一个合适的函数形式作为响应面方程对得到正确的临界值估计是十分重要的, 而且需要满足一些约束条件, 即滞后阶数 $k$ 对临界值的影响必须随样本容量 $T$ 的增大而减小, 至 $T$ 趋于无穷时, 这种影响减至零. 经过反复的实验与筛选, 得到一个二阶多项式方程

$$CV = \alpha_0 + \alpha_1 (1/T') + \beta_1 (k/T') + \beta_2 (k/T'^2) + \epsilon \tag{6}$$

式中 $CV$ 表示对应于样本容量 $T$ 和滞后阶数 $k$ 的临界值, $T' = T - k$ 为有效的观测值个数, $\epsilon$ 为残差项.

注意到如果控制变量采用样本容量个数 $T$ 替代有效观测值个数 $T'$, 拟合优度会降低, 所以采用有效观测值个数作为控制变量拟合响应面方程是合理的.

当变量个数 $N = 2$ 时, 三种不同显著性水平下的 ADF 检验的响应面方程结果见表 1, 同时各种拟合结果的参数也列于表中, 包括相关系数 $R^2$, 残差平方和(SSE). 在所有情况下, $1/T'$ 与 $k/T'$, $k/T'^2$ 项的系数在统计上都是显著的.

用方程(6) 估计得到的临界值与文[2, 8] 进行比较. 例如当 $k = 5$, $T = 50$ 时, 估计得出的临界值为 $-3.652(1\%)$, $-2.862(5\%)$ 和 $-2.593(10\%)$, 而文[2] 为 $-2.62(1\%)$, $-2.25(5\%)$ 和 $-1.95(10\%)$, 文[8] 为 $-4.231(1\%)$, $-3.496(5\%)$, $-3.144(10\%)$, 本文的估计值显然与文[2, 8] 未考虑滞后阶数得到的临界值不同. 但随样本容量增大这些差别将逐渐缩小, 因为方程是 $1/T'$ 和 $k/T'^2$ 的函数.

**表 1　$N = 2$ 时临界值响应面的估计结果**

| 方程系数 | 无常数项和趋势项 | | | 带常数项 | | | 带常数项和趋势项 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $\alpha_0$ | -3.366 506<br>(0.012 200) | -2.784 632<br>(0.008 068) | -2.480 604<br>(0.006 09) | -3.557 953<br>(0.014 134) | -2.969 647<br>(0.008 264) | -2.672 650<br>(0.006 479) | -4.093 594<br>(0.019 446) | -3.502 505<br>(0.012 121) | -3.201 512<br>(0.009 686) |
| $\alpha_1$ | -10.738 151<br>(0.567 699) | -6.684 779<br>(2.375 419) | -5.342 555<br>(0.283 39) | -6.114 677<br>(0.657 662) | -3.067 052<br>(0.383 694) | -1.803 428<br>(0.301 466) | -9.288 471<br>(0.904 850) | -4.604 490<br>(0.564 015) | -3.382 569<br>(0.450 686) |

续表 1

| 方程系数 | 无常数项和趋势项 | | | 带常数项 | | | 带常数项和趋势项 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $\beta_1$ | 2.680 396<br>(0.171 946) | 2.158 801<br>(0.113 708) | 2.909 287<br>(0.085 83) | 2.103 123<br>(0.199 194) | 2.527 371<br>(0.116 214) | 2.354 348<br>(0.091 309) | 3.648 660<br>(0.274 063) | 2.191 316<br>(0.170 830) | 2.790 942<br>(0.136 505) |
| $\beta_2$ | - 12.487 412<br>(2.505 169) | - 4.808 964<br>(1.656 668) | - 2.059 357<br>(1.250 56) | - 35.944 307<br>(2.902 163) | - 16.707 123<br>(1.693 184) | - 11.368 299<br>(1.330 323) | - 67.055 275<br>(3.992 964) | - 27.790 451<br>(2.488 910) | - 17.675 995<br>(1.988 807) |
| $R^2$ | 0.926 | 0.898 | 0.930 | 0.955 | 0.876 | 0.871 | 0.971 | 0.924 | 0.836 |
| SSE | 0.035 | 0.023 | 0.017 | 0.040 | 0.023 | 0.018 | 0.055 | 0.034 | 0.027 |

注: 对于每一组 $T$ 与 $k$, 实验分别重复 30 000 次

## 1.2　多元协整检验

文[14] 提出了协整检验统计量临界值的大小不仅依赖于样本容量的个数, 还与协整回归方程中变量的个数有关. 因此, 考虑协整系统中存在多个变量的情形. 考虑以下DGP:

$$Y_t = Y_{t-1} + \epsilon_t$$

$$Y_t = (y_{t1}, y_{t2}, \cdots, y_{tN}), N > 2 \tag{7}$$

$\epsilon_t$ 为相互独立的标准正态分布的随机变量, 即 $\epsilon_{1t} \overset{iid}{\sim} N(0, 1)$. 向量 $Y_t$ 表示 $N$ 个时间序列的第 $t$ 期观测值, 每一个分量都是一个一阶单整 $I(1)$ 序列. 如果 $Y$ 的各分量之间存在协整关系, 则存在向量 $\alpha$, 使 $z_t = \alpha' Y_t$ 为 $I(0)$ 序列.

同两变量协整情况相同, 首先估计协整回归方程

$$y_{1t} = \alpha_1 + \sum_{j=2}^{N} \alpha_j y_{jt} + z_t, \ N > 2 \tag{8}$$

因此得到的系数的估计值 $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_N)$, 计算

$$\hat{z}_t = y_{1t} - \hat{\alpha}_1 - \hat{\alpha}_2 y_{2t} - \cdots - \hat{\alpha}_N y_{Nt} \tag{9}$$

利用 $\hat{z}_t$ 按式(3)～(5) 中三种不同的形式分别计算相应ADF 统计量在 1%, 5% 和 10% 显著性水平下的临界值.

应用式(6), 分别估计 $N = 3, 4, 5$ 下的ADF 临界值的响应面方程, 结果见表 2, 其中 S.E 分别表示每个参数的标准差.

表 2　$N = 3, 4, 5, 6$ 时临界值响应面的估计结果

| $N$ | 响应面方程 | $\alpha$ (%) | $\alpha_0$ | 标准差 S.E | $\alpha_1$ | 标准差 S.E | $\beta_1$ | 标准差 S.E | $\beta_2$ | 标准差 S.E |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 无常数项和趋势项 | 1 | - 3.901 599 | 0.015 635 | - 14.971 972 | 0.727 519 | 4.810 935 | 0.220 353 | - 19.419 715 | 3.210 431 |
| | | 5 | - 3.325 914 | 0.010 772 | - 9.382 795 | 0.501 220 | 3.870 052 | 0.151 810 | - 10.220 42 | 2.211 805 |
| | | 10 | - 3.029 270 | 0.008 587 | - 7.401 87 | 0.399 566 | 3.522 845 | 0.121 021 | - 7.529 447 | 1.763 224 |
| 3 | 带常数项 | 1 | - 3.967 799 | 0.017 275 | - 10.579 268 | 0.803 826 | 4.263 872 | 0.243 464 | - 41.267 087 | 3.547 160 |
| | | 5 | - 3.385 296 | 0.011 116 | - 6.015 794 | 0.517 245 | 3.319 929 | 0.156 664 | - 20.946 716 | 2.282 522 |
| | | 10 | - 3.092 103 | 0.008 497 | - 4.190 991 | 0.395 371 | 3.053 045 | 0.119 751 | - 16.513 091 | 1.744 712 |
| 3 | 带常数项和趋势项 | 1 | - 4.275 280 | 0.024 424 | - 9.292 095 | 1.136 500 | 4.736 282 | 0.344 225 | - 75.679 100 | 5.015 200 |
| | | 5 | - 3.664 531 | 0.013 108 | - 5.814 335 | 0.609 915 | 3.160 492 | 0.184 732 | - 32.663 704 | 2.691 459 |
| | | 10 | - 3.365 174 | 0.010 519 | - 3.961 433 | 0.489 445 | 2.631 782 | 0.148 244 | - 21.148 896 | 2.159 846 |
| 4 | 无常数项和趋势项 | 1 | - 4.281 729 | 0.018 142 | - 20.902 325 | 0.844 17 | 6.604 161 | 0.255 687 | - 18.644 406 | 3.725 230 |
| | | 5 | - 3.732 137 | 0.019 410 | - 13.273 459 | 0.903 183 | 5.650 421 | 0.273 558 | - 14.130 556 | 3.985 605 |
| | | 10 | - 3.439 591 | 0.021 342 | - 10.405 695 | 0.550 929 | 5.003 619 | 0.166 867 | - 9.766 540 | 2.431 166 |
| 4 | 带常数项 | 1 | - 3.760 024 | 0.021 342 | - 9.360 548 | 0.993 062 | 5.063 535 | 0.300 781 | - 23.900 824 | 4.382 231 |
| | | 5 | - 3.760 024 | 0.019 736 | - 12.387 44 | 0.738 221 | 7.429 824 | 0.253 878 | - 25.986 236 | 4.304 33 |
| | | 10 | - 3.622 029 | 0.013 422 | - 5.788 496 | 0.624 564 | 3.968 644 | 0.189 169 | - 23.994 685 | 2.756 106 |
| 4 | 带常数项和趋势项 | 1 | - 4.529 792 | 0.028 067 | - 12.187 445 | 1.305 986 | 6.256 605 | 0.395 560 | - 78.414 492 | 5.763 113 |
| | | 5 | - 3.911 970 | 0.016 056 | - 8.358 554 | 0.747 118 | 4.509 975 | 0.226 289 | - 33.845 875 | 3.296 917 |
| | | 10 | - 3.622 029 | 0.013 422 | - 5.788 496 | 0.624 564 | 3.968 644 | 0.189 169 | - 23.994 685 | 2.756 106 |

续表 2

| $N$ | 响应面方程 | α (%) | $\alpha_0$ | 标准差 S.E | $\alpha_1$ | 标准差 S.E | $\beta_1$ | 标准差 S.E | $\beta_2$ | 标准差 S.E |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 无常数项和趋势项 | 1 | -4.635 735 | 0.021 091 | -24.005 624 | 0.981 412 | 8.289 911 | 0.297 252 | -26.565 554 | 4.330 818 |
| | | 5 | -4.098 316 | 0.032 518 | -17.865 938 | 1.496 383 | 8.149 249 | 0.453 227 | -32.386 286 | 6.603 305 |
| | | 10 | -3.798 313 | 0.016 041 | -12.701 873 | 0.746 427 | 6.488 877 | 0.226 079 | -14.553 347 | 3.293 866 |
| 5 | 带常数项 | 1 | -4.672 689 | 0.022 261 | -19.134 605 | 1.035 857 | 8.097 197 | 0.313 742 | -48.295 880 | 4.571 077 |
| | | 5 | -4.108 383 | 0.018 327 | -11.827 841 | 0.852 774 | 6.578 644 | 0.258 290 | -27.077 642 | 3.763 158 |
| | | 10 | -3.811 818 | 0.016 306 | -9.796 227 | 0.758 743 | 6.126 018 | 0.229 810 | -20.067 649 | 3.348 215 |
| 5 | 带常数项和趋势项 | 1 | -4.786 863 | 0.025 209 | -16.193 421 | 1.173 032 | 7.803 923 | 0.355 290 | -75.510 248 | 5.176 409 |
| | | 5 | -4.207 208 | 0.017 942 | -9.680 842 | 0.834 881 | 6.055 418 | 0.252 871 | -40.213 767 | 3.684 203 |
| | | 10 | -3.912 037 | 0.015 388 | -7.495 009 | 0.716 004 | 5.543 628 | 0.216 877 | -29.395 472 | 3.159 792 |
| 6 | 无常数项和趋势项 | 1 | -4.640 878 | 0.021 130 | -23.769 008 | 0.983 214 | 8.280 772 | 0.297 798 | -27.032 709 | 4.338 770 |
| | | 5 | -4.093 828 | 0.017 674 | -15.656 407 | 0.822 412 | 7.100 859 | 0.249 094 | -18.243 338 | 3.629 178 |
| | | 10 | -3.912 868 | 0.015 600 | -7.629 997 | 0.725 904 | 5.609 507 | 0.219 863 | -29.946 564 | 3.203 304 |
| 6 | 带常数项 | 1 | -4.682 733 | 0.021 650 | -18.682 446 | 1.007 408 | 8.118 645 | 0.305 126 | -49.010 089 | 4.445 536 |
| | | 5 | -4.114 059 | 0.018 067 | -11.986 421 | 0.840 706 | 6.722 860 | 0.254 635 | -27.770 511 | 3.709 905 |
| | | 10 | -3.816 515 | 0.015 818 | -9.430 826 | 0.736 033 | 6.160 254 | 0.222 931 | -21.513 200 | 3.248 000 |
| 6 | 带常数项和趋势项 | 1 | -4.804 878 | 0.026 436 | -15.192 262 | 1.230 119 | 7.861 203 | 0.372 581 | -79.595 848 | 5.428 326 |
| | | 5 | -4.207 232 | 0.018 186 | -10.142 028 | 0.846 203 | 6.161 485 | 0.256 300 | -39.632 279 | 3.734 164 |
| | | 10 | -3.912 868 | 0.015 600 | -7.629 997 | 0.725 904 | 5.609 507 | 0.219 863 | -29.946 564 | 3.203 304 |

注: 表 2 中, $N$ 表示变量个数, α 表示检验的显著性水平.

## 2　中国城镇居民消费函数的协整分析

### 2.1　数据分析

中国实行改革开放政策以来, 中国城镇居民的消费结构发生了显著的变化. 食品消费支出在总的生活消费支出中的比重呈明显的下降趋势. 但是, 城镇居民的食品消费与其生活费的收入之间是否存在一种长期的均衡关系, 即二者之间是否存在协整关系? 由于统计数据数量的限制, 这里进行小样本条件下的协整分析.

按照不同的收入假说, 可以建立不同形式的消费函数模型. 把消费函数的一般形式表示为

$$EXP_t = f(INC_t, INC_{t-1}, \cdots; EXP_{t-1}, EXP_{t-2}, \cdots; \cdots)$$

采用中国城镇居民年人均食品消费 ($EXP_t$) 与人均生活费收入 ($INC_t$) 进行小样本协整分析. 样本区间为 1978～1996 年. 为了消除通货膨胀率的影响, 本文将数据 $EXP_t$ 和 $INC_t$ 分别除以全国消费品零售物价总指数 $PI_t$ (1978 年为 100), 得到实际的年人均食品消费及生活费收入数据, 再对它们进行自然对数变换, 得到分析中需要的序列. 数据取对数后并不改变原来的协整关系. 这里对数据取对数后再进行差分, 实际上等于增长速度的对数值, 而增长速度一般是 $I(0)$ 序列;

首先对数据进行单整检验, 采用 ADF 检验对 $\ln EXP_t$ 和 $\ln INC_t$ 进行单位根检验, 检验结果见表 3.

表 3　单位根检验

| 变　量 | ADF 检验 | 检验类型 |
|---|---|---|
| $\ln EXP_t$ | 0.438 94 | 带常数项和趋势项 |
| $\ln INC_t$ | 0.915 43 | |
| $\Delta\ln EXP_t$ | -3.265 4* | |
| $\Delta\ln INC_t$ | -3.795 4* | |

注: * 5% 显著性水平上拒绝单位根假设, 临界值 5% = -3.082

由表 3 可以看出, 序列 $\ln EXP_t$ 和 $\ln INC_t$ 均为一阶单整序列, 即经过差分以后序列为平稳的.

### 2.2　协整分析

由于只可能存在一个协整关系, 采用 EG 两步法来检验序列的协整关系.

1) 首先估计长期均衡关系, 对 (2.1) 应用 OLS 法, 得到

$$\ln EXP_t = 0.667\,3 + 0.768\,4 \ln INC_t \quad (10)$$

S. E. (0. 113 43) (0. 013 75)

2) 进行残差的平稳性检验, 设

$$z_t = \ln EXP_t - 0.6673 - 0.7684 \ln INC_t \quad (11)$$

对残差序列进行平稳性检验, 这里由于样本容量较小为 19, 无法从文[1] 或文[13] 等给出的临界值表获得相应的临界值. 而应用本文给出的小样本ADF 检验的响应面方程, 临界值就可以方便地计算出来. 因为本文的小样本检验方法涉及到滞后阶数, 先用 SC (Schwarz Criterion) 准则确定滞后阶数 $k$. SC 准则在 $k = 2$ 时取得最小值, 所以选 $k = 2$ 为滞后阶数. 当 $T = 19, k = 2, N = 2$ 时, 在模型(5) 的 5% 显著性水平下, 临界值的估计值为CV = - 3. 304, 对式(3) 的残差序列进行单位根检验得(表 4).

表 4 残差序列的单位根检验表

| 变量 | ADF 检验 | 5% 水平临界值 | 检验类型 |
|---|---|---|---|
| 残差 $z_t$ | - 3. 535 7* | 3. 304 | 带常数项和趋势项 |

注: * 5% 显著性水平上拒绝非平稳假设

由表 4 知, 食品消费支出 $\ln EXP_t$ 与生活费收入 $\ln INC_t$ 之间存在协整关系, 其协整向量为(1, - 0. 768 4).

3) 第 2 阶段就可以建立误差校正模型 ECM 了. 采用 Hendry 一般到特殊(general to simple) 的建模方法, 得估计的误差校正模型为

$$\Delta\ln EXP_t = 0.03442 + 0.54521\Delta\ln INC_t - 0.81321(\ln EXP - 0.7684 \ln INC)_{t-1} \quad (12)$$

S. E. (0. 008 604) (0. 089 522) (0. 165 43)

$R^2 = 0.8678, F(2, 14) = 28.351$, S. E. = 0. 017 63, $DW = 2.05$, RSS = 0. 004 83

($R^2$, $F$ 和DW 是常规统计量, S. E 为估计方程的标准误差, RSS 为残差平方和)

统计检验结果表明方程(12) 具有很好的拟合效果.

由以上分析可知, 自改革开放以来, 我国城镇居民的消费结构虽然发生了很大的变化, 食品支出在总的消费支出中的比重呈逐年下降的趋势, 但是由于饮食结构的不断调整, 在外就餐的机会增加等因素的影响, 食品支出仍然占有一定的比例. 它与城镇居民的人均生活收入之间仍然存在长期的均衡关系. 这种长期均衡的偏差有81. 32% 需要由下一年度来调整.

## 3 结束语

本文应用响应面方法对ADF 检验在小样本下的临界值进行了研究. 已有的一些研究忽略了ADF 检验中滞后阶数对小样本下临界值的影响, 而实验证明了在响应面方程中加入适当的滞后阶数的必要性. 本文亦给出了多变量协整系统下ADF 检验的响应面方程. 对于给定的样本容量和滞后阶数, 由响应面方程就可以方便地计算出准确的ADF 检验临界值了, 这对于实际应用提供了极大的方便. 此外, 如果响应面方程还是回归变量个数N 的函数, 将N 作为控制变量之一增加入响应面方程, 则响应面方程的个数会减少一半, 这就明显提高了响应面方程的普遍适用性, 可以采用神经网络方法建立响应面方程(另文报告).

对中国居民生活费收入与食品消费的协整分析, 表明两者之间的长期均衡关系的存在性.

## 参 考 文 献:


[1] Dickey D A, Fuller W A. Distribution of the estimation for autoregressive time series with a unit root[J]. Journal of the American Statistical Association, 1979, 74: 427-431

[2] Dickey D A, Fuller W A. Likelihood ratio statistics for autoregression time series with a unit root[J]. Econometrica, 1981, 49: 1057-1072

[3] Engle R F, Granger C W J. Co-integration and error correction: representation, estimation and testing[J]. Econometrica, 1987, 55: 251-276

[4] 孙青华, 张喜彬, 张世英. 非线性协整关系的存在性研究[J]. 管理科学学报, 2000, 3(3): 65-74

[5] 张喜彬, 孙青华, 张世英. 非线性协整关系及其检验方法[J]. 系统工程学报, 1999, 14(1): 57-68

[6] Blangiewicz M, Charemza W W. Cointegration in small samples: empirical percentiles, drifting moments and customized testing[J]. Bulletin, 1990, 52: 303-315
[7] Harris R ID. Small sample testing for unit roots[J]. Bulletin, 1992, 54: 615-625
[8] MacKinnon J G. Critical values for cointegration test in long-run economic relationships: Reading in Cointegration. Oxford: Oxford University Press, 1990. 266-276
[9] Mayers R H, Khuri A I, Carter W H. Response surface methodology: 1966-1988[J]. Technometrics, 1989, 31: 137-165
[10] Cheung Y W, Lai K S. Finite sample sizes of Johansen's likelihood ratio tests for cointegration[J]. Bulletin, 1993, 55: 313-328
[11] Johansen S, Juselius K. Maximum likelihood estimation and inference on cointegration-with application to the demand for money[J]. Bulletin, 1990, 52: 169-210
[12] Johansen S. Statistical analysis of cointegration vectors[J]. Journal of Economic Dynamics and Control, 1988, 12: 231-254
[13] Johansen S. Estimation and hypothesis testing of cointegration vectors in gaussian vector autogressive models [J]. Econometrica, 1991, 59: 1551-1580
[14] Engle R F, Yoo B S. Forecasting and testing in co-integration systems[J]. Journal of Econometrics, 1987, 35: 143-159

# Cointegration test in small sample

*YANG Bao-chen, WANG Yan, ZHANG Shi-ying*
School of Management, Tianjin University, Tianjin 300072, China

**Abstract**: In this paper, we develop the response surface equations for the critical value for cointegration test in small sample using Monte Carlo simulation. The simulation results indicate that the critical values are related with the number of the sample data as well as the lagged order for cointegration test. We also present an empirical analysis for the food consumption of the Chinese residents in cities and towns

**Key words**: cointegration test; response surface method; small sample