

协整检验及误差修正模型

分享者：邵登科

日期：2019年03月13日

地点：南京宜开数据分析技术有限公司

1

为什么进行协整检验

2

相关的概念

3

协整的概念

4

误差修正模型介绍

5

协整检验的方法

6

案例分析

1、为什么进行协整检验

● 为什么进行协整检验

1、经典计量经济模型常用到的数据有：

- 时间序列数据 (time-series data)
- 截面数据(cross-sectional data)
- 平行/面板数据 (panel data/time-series cross-section data)

2、时间序列数据是最常见，也是最常用到的数据。

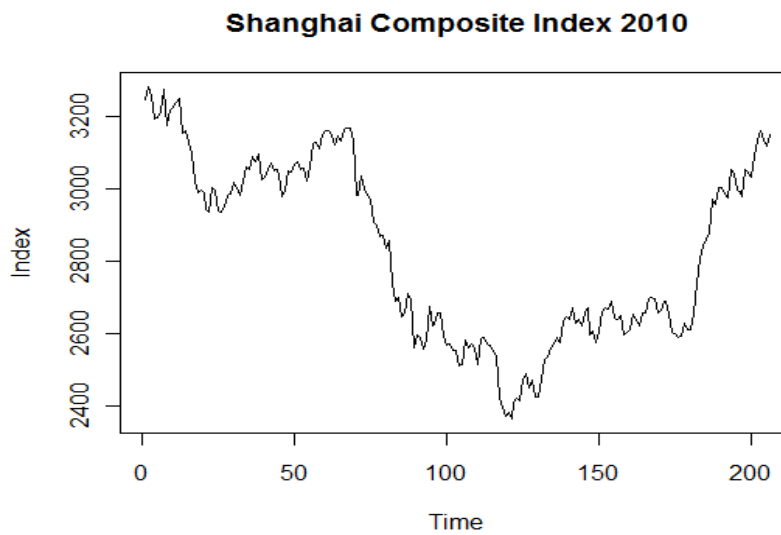
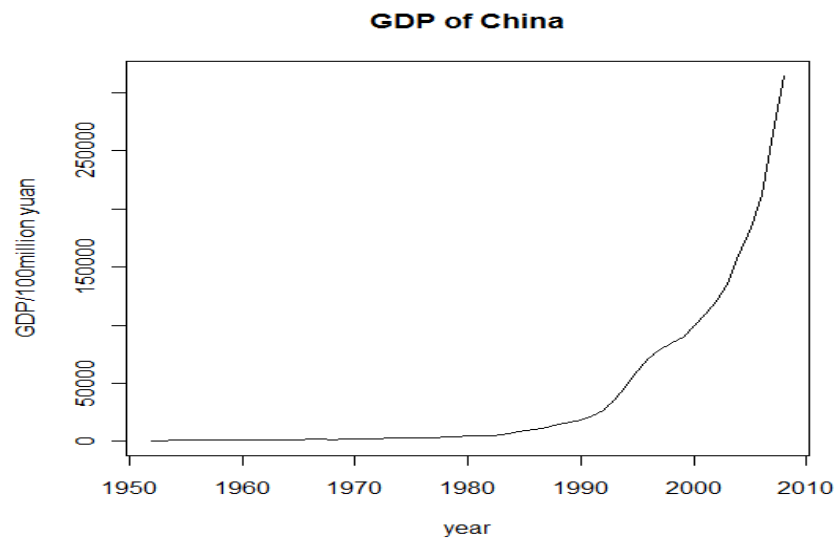
3、经典回归分析暗含着一个重要假设：数据是平稳的。



为什么进行协整检验

回归分析一个重要的前提假设：平稳性

但是，实际上大部分的宏观经济时间序列和金融时间序列都是非平稳的。



● 为什么进行协整检验

- 1、对于非平稳变量，大样本下的统计推断基础-“一致性”要求被破坏。不能使用经典回归模型，否则会出现‘伪回归’等诸多问题。**
 - 表现为两个本来没有任何因果关系的变量，却有很高的相关性。
 - 例如：如果有两列时间序列数据表现出一致的变化趋势（非平稳的），即使它们没有任何有意义的关系，但进行回归也可表现出较高的可决系数。
 - 金融、经济时间序列常常是非平稳的，对他们的建模必须考虑模型的设定问题，避免‘伪回归’问题。
- 2、协整理论的提出为非平稳的多变量时间序列分析提供了理论和方法，它把时间序列方法中对模型短期动态设定和数量经济学中长期均衡关系的确定特点融合。**
- 3、检验变量之间的协整关系，从变量之间是否具有协整关系出发选择模型的变量，其数据基础是牢固的，其统计性质是优良的。**

2、相关的概念

● 相关的概念

1、平稳性的概念

- 假定某个时间序列是由某一随机过程生成的，即假定时间序列 $\{X_t\}$ ($t=1, 2, \dots$) 的每一个数值都是从一个概率分布中随机得到，如果满足下列条件：
 - **均值 $E(X_t)=\mu$ 是与时间 t 无关的常数；**
 - **方差 $Var(X_t)=\sigma^2$ 是与时间 t 无关的常数；**
 - **协方差 $Cov(X_t, X_{t+k})=\gamma_k$ 是只与时期间隔 k 有关，与时间 t 无关的常数；**
- 则称该随机时间序列是平稳的 (stationary)，而该随机过程是一个平稳随机过程 (stationary stochastic process) 。

● 相关的概念

2、平稳性的检验方法

- **DF检验** (Dicky和Fuller于1976年提出)
- **ADF检验** (解决了时间序列高阶的自回归过程生成, 或者随机误差项并非是白噪声, 用OLS法进行估计均会表现出随机误差项出现自相关, 导致DF检验无效)

$$\Delta X_t = \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$$

模型1

$$\Delta X_t = \alpha + \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$$

模型2

$$\Delta X_t = \alpha + \beta t + \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$$

模型3

零假设 $H_0: \delta=0$ (X_t 为随机游走序列)

备择假设 $H_1: \delta<0$ (X_t 为平稳序列)

检验过程:

- 同时估计出上述三个模型的适当形式, 然后通过ADF临界值表检验零假设 $H_0: \delta=0$ 。
- 只要其中有一个模型的检验结果拒绝了零假设, 就可以认为时间序列是平稳的;
- 当三个模型的检验结果都不能拒绝零假设时, 则认为时间序列是非平稳的。

3、单整的概念

- 如果一个时间序列经过一次差分变成平稳的，就称原序列是一阶单整序列，记为 $I(1)$ 。
- 一般地，如果一个时间序列经过 d 次差分后变成平稳序列，则称原序列是 d 阶单整序列，记为 $I(d)$ 。
例如带截距项的随机游走序列，即为 $I(1)$ 序列。
- $I(0)$ 代表一平稳时间序列。

4、长期均衡

- 经济理论指出，某些经济变量间确实存在着长期均衡关系，这种均衡关系意味着经济系统不存在破坏均衡的内在机制，如果变量在某时期受到干扰后偏离其长期均衡点，则均衡机制将会在下一期进行调整以使其重新回到均衡状态。

- 假设X与Y间的长期“均衡关系”由式描述：

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$

- **该均衡关系意味着：**给定X的一个值，Y相应的均衡值也随之确定为 $\alpha_0 + \alpha_1 X$ 。

5、长期均衡的解释

- 在t-1期末，存在下述三种情形之一：
 - Y等于它的均衡值： $Y_{t-1} = \alpha_0 + \alpha_1 X_t$ ；
 - Y小于它的均衡值： $Y_{t-1} < \alpha_0 + \alpha_1 X_t$ ；
 - Y大于它的均衡值： $Y_{t-1} > \alpha_0 + \alpha_1 X_t$ ；
- 在时期t，假设X有一个变化量 ΔX_t ，如果变量X与Y在时期t与t-1末期仍满足它们间的长期均衡关系，即上述第一种情况，则Y的相应变化量为：

$$\Delta Y_t = \alpha_1 \Delta X_t + v_t \quad v_t = \mu_t - \mu_{t-1}$$

- 如果t-1期末，发生了上述第二种情况，即Y的值小于其均衡值，则t期末Y的变化往往会比第一种情形下Y的变化大一些；
- 反之，如果t-1期末Y的值大于其均衡值，则t期末Y的变化往往会小于第一种情形下的 ΔY_t 。
- 可见，如果 $Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$ 正确地提示了X与Y间的长期稳定的“均衡关系”，则意味着Y对其均衡点的偏离从本质上说是“临时性”的。
- 一个重要的假设就是：随机扰动项 μ_t 必须是平稳序列。如果 μ_t 有随机性趋势（上升或下降），则会导致Y对其均衡点的任何偏离都会被长期累积下来而不能被消除。

3、协整的概念

● 协整的概念

协整关系描述了经济系统的长期均衡关系。具体来说，它描述了两个或多个非平稳时间序列的均衡关系，虽然每个时间序列的矩，如均值、方差或协方差等随时间变化，但这些序列的某种线性组合（均衡关系）的矩具有时不变的特征。

定义：对于m维向量时间序列 $\{X_t\}$ ，如果

- 1、 $\{X_t\}$ 的分量序列为I(d)序列；
- 2、存在一个向量 $\alpha \neq 0$ ，使得 $\alpha^T X_t \sim C(d-b), b > 0$

则称 $\{X_t\}$ 为分量序列在 (d,b) 阶协整关系，记为 $\{X_t\} \sim CI(d,b)$ ，而 α 成为协整向量。

该定义需要注意两点：

- 1、 $\{X_t\}$ 的分量序列是具有相同阶数的单整序列。
- 2、定义描述了 $\{X_t\}$ 中非平稳分量之间的长期线性均衡关系（ $\alpha^T X_t$ ）。

● 协整的概念

例如1：两个I(1)序列 $\{x_t\}$ 和 $\{y_t\}$,

$$\{x_t\} \sim I(1) \quad \{y_t\} \sim I(1)$$

他们之间的关系为：

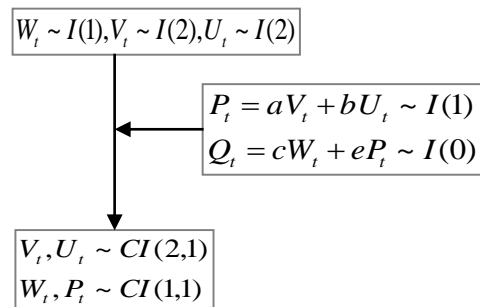
$$y_t = \alpha x_t + u_t \quad u_t \sim I(0)$$

这两个变量之间的长期均衡关系可以表示为：

$$y_t = \alpha x_t$$

由于 $u_t = \alpha^T X_t = (1 - \alpha) \begin{bmatrix} y_t \\ x_t \end{bmatrix} = y_t - \alpha x_t \sim I(1 - 1) \sim I(0)$

所以 $\{x_t\}$ 和 $\{y_t\}$ 具有协整关系 $CI(1,1)$ ，而协整向量为 $(1, -\alpha)$



如果两个变量都是单整变量，只有当它们的单整阶数相同时，才可能协整；如果它们的单整阶数不相同，就不可能协整。3个以上的变量，如果具有不同的单整阶数，有可能经过线性组合构成低阶单整变量。

4、误差修正模型的介绍

● 误差校正模型介绍

误差校正模型 (error-correction model, ECM) 是协整关系的一种重要表示形式, 它克服了伪回归问题, 有效的描述了经济变量序列之间的长期 (静态) 表现和短期 (动态) 表现。

定义: 称 m 维向量时间序列具有误差校正表现形式, 如果他可以表示为

$$A(L)(1 - L)X_t = -\gamma Z_{t-1} + \varepsilon_t$$

其中, ε_t 是 m 维平稳随机序列, $A(L)$ 是关于滞后算子 L 的矩阵多项式, 并且 $A(0) = I_m$, $A(1)$ 的元素都是有限的, $Z_t = \alpha^T X_t$ (α 是序列 X_t 的协整矩阵), 而 $Z_t = (z_{1t}, z_{2t}, \dots, z_{rt})^T$ 是模型的误差校正项, γ 是非0的 $m \times r$ 维参数矩阵。

协整关系序列误差校正模型的一般形式为:

$$\varphi(L)\Delta X_t = -\Pi X_{t-k} + \varepsilon_t$$

$$\varphi(L)\Delta X_t = -r\alpha^T X_{t-k} + \varepsilon_t$$

其中, r 为 $m \times n$ 维的系数矩阵, α 为 $m \times r$ 维的 X_t 的协整矩阵。

● 误差校正模型介绍

例如1：两个I(1)序列 $\{x_t\}$ 和 $\{y_t\}$ ，他们之间的关系为：

$$y_t = \alpha_0 + \alpha_1 x_t + u_t \quad u_t \sim I(0)$$

由于 $u_t \sim I(0)$ ，所以 $\{x_t\}$ 和 $\{y_t\}$ 具有协整关系 $CI(1,1)$ ，而协整向量为 $(1, -\alpha_1)$

一阶误差修正模型(first-order error correction model)的形式：

$$\Delta x_t = \beta_{10} + \beta_{11} \Delta x_{t-1} + \beta_{12} \Delta y_{t-1} + \gamma_1 ECM_{t-1} + \varepsilon_{1t}$$

$$\Delta y_t = \beta_{20} + \beta_{21} \Delta x_{t-1} + \beta_{22} \Delta y_{t-1} + \gamma_2 ECM_{t-1} + \varepsilon_{2t}$$

其中： $ECM_t = y_t - \hat{\alpha}_1 x_t - \hat{\alpha}_0$

● 误差修正模型介绍

误差修正模型的建立的条件：

首先，对变量进行协整检验，以发现变量之间的协整关系，即长期均衡关系，并以这种关系构成误差修正项。

然后，建立短期模型，将误差修正项看作一个解释变量，连同其它反映短期波动的解释变量一起，建立短期模型，即误差修正模型。

5、协整检验的方法



协整检验的方法

协整检验的方法：

1、EG两步检验法

Engle和Granger于1987年提出两步检验法，为了检验变量之间是否为协整。

2、JJ检验法

Johansen于1988年，以及与Juselius一起于1990年提出了一种用向量自回归模型进行检验的方法，通常称为Johansen检验，或JJ检验，是一种进行多重 $I(1)$ 序列协整检验的较好方法。

● 协整检验的方法-EG检验

EG两步法（二维向量时间序列）：

第一步，对两个I(1)变量 X_t 和 Y_t 进行静态回归：

$$Y_t = \alpha_0 + \alpha_1 \times X_t + \mu_t$$

利用观测数据，通过普通最小二乘进行拟合，得到最小二乘估计量和残差序列：

$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 \times X_t \quad \hat{u}_t = Y_t - \hat{Y}_t$$

若 $\hat{u}_t \sim I(0)$ ，则和是CI(1,1) 协整，协整向量为 $\hat{\alpha} = (1, \hat{\alpha}_1)^T$ ，以残差序列 \hat{u}_t 作为 Z_t 误差校正项的估计。即 $\hat{Z}_t = \hat{u}_t$ 。

第二步，将估计的协整向量 $\hat{\alpha}$ 带入时间序列 $\{X_t, Y_t\}$ 的误差校正模型中，利用最小二乘法估计模型的动态均衡关系。

$$\Delta x_t = \beta_{10} + \beta_{11}\Delta x_{t-1} + \beta_{12}\Delta y_{t-1} + \gamma_1 ECM_{t-1} + \varepsilon_{1t}$$

$$\Delta y_t = \beta_{20} + \beta_{21}\Delta x_{t-1} + \beta_{22}\Delta y_{t-1} + \gamma_2 ECM_{t-1} + \varepsilon_{2t}$$

● 协整检验的方法-EG检验

协整关系的检验：

最基本的检验方法是对静态回归模型得到的残差进行单位根检验，即检验残差序列 \hat{u}_t （ $\hat{u}_t = \hat{Y}_t - \hat{\alpha}_0 - \hat{\alpha}_1 \times X_t$ ）的单整阶数。检验的零假设和备择假设分别为：

$$H_0: \hat{u}_t \sim I(1), \quad H_1: \hat{u}_t \sim I(0)$$

在零假设下，残差序列中存在一个单位根，如果零假设成立，说明的 $\{X_t, Y_t\}$ 分量序列之间不存在协整关系。这样，对静态回归算产生的残差序列 \hat{u}_t 进行单位根检验，就可以查验出序列中 $\{X_t, Y_t\}$ 分量序列之间协整关系的存在性。

当 $\hat{\alpha}_1$ 为已知情况下，可以之间应用DF检验的临界值对残差序列进行单位根检验。

当 $\hat{\alpha}_1$ 为未知情况下，必须首先从静态回归中估计出 $\hat{\alpha}_1$ ，从而得到残差序列的估计值，它为被检验的序列。

由于静态回归估计采用最小二乘法，使得模型的残差平方和最小，估计的残差序列比实际随机的残差序列要小，使得 $\hat{\alpha}_1$ 使得 $\{X_t, Y_t\}$ 最大可能一致运动，从而使得检验出的协整关系可能大于实际协整关系。

● 协整检验的方法-EG检验

多变量协整关系的检验—扩展的E-G检验：

多变量协整关系的检验要比双变量复杂一些，主要在于协整变量间可能存在多种稳定的线性组合。假设有4个一阶单整I(1)变量Z、X、Y、W，它们有如下的长期均衡关系：

$$Z_t = \alpha_0 + \alpha_1 W_t + \alpha_2 X_t + \alpha_3 Y_t + \mu_t$$

非均衡误差项 μ_t 应是I(0)序列：

$$\mu_t = Z_t - \alpha_0 - \alpha_1 W_t - \alpha_2 X_t - \alpha_3 Y_t$$

● 协整检验的方法-EG检验

多变量协整关系的检验—扩展的E-G检验：

然而，如果Z与W，X与Y间分别存在长期均衡关系：

$$Z_t = \beta_0 + \beta_1 W_t + v_{1t} \quad X_t = \gamma_0 + \gamma_1 Y_t + v_{2t}$$

则非均衡误差项 v_{1t} 、 v_{2t} 一定是稳定序列 $I(0)$ 。于是它们的任意线性组合也是稳定的。例如

$$v_t = v_{1t} + v_{2t} = Z_t - \beta_0 - \beta_1 W_t + X_t - \gamma_1 Y_t$$

一定是 $I(0)$ 序列。

由于 v_t 像 u_t 一样，也是Z、X、Y、W四个变量的线性组合，由此 v_t 式也成为该四变量的另一稳定线性组合。

$(1, -\alpha_0, -\alpha_1, -\alpha_2, -\alpha_3)$ 是对应于 μ_t 式的协整向量， $(1, -\beta_0 - \gamma_0, -\beta_1, 1, -\gamma_1)$ 是对应于 v_t 式的协整向量。

● 协整检验的方法-EG检验

多变量协整关系的检验—扩展的E-G检验

检验程序：

- 1、对于多变量的协整检验过程，基本与双变量情形相同，即需检验变量是否具有同阶单整性，以及是否存在稳定的线性组合。
- 2、在检验是否存在稳定的线性组合时，需通过设置一个变量为被解释变量，其他变量为解释变量，进行OLS估计并检验残差序列是否平稳。
- 3、如果不平稳，则需更换被解释变量，进行同样的OLS估计及相应的残差项检验。
- 4、当所有的变量都被作为被解释变量检验之后，仍不能得到平稳的残差项序列，则认为这些变量间不存在 (d,d) 阶协整。



误差修正模型介绍

采用Engle-Granger两步法建立误差修正模型

第一步，各变量进行单位根检验。

第二步，进行协整回归（OLS法），检验变量间的协整关系，估计协整向量（长期均衡关系参数）；

第二步，若协整性存在，则以第一步求到的残差作为非均衡误差项加入到误差修正模型中，并用OLS法估计相应参数。

需要注意的是：在进行变量间的协整检验时，如有必要可在协整回归式中加入趋势项，这时，对残差项的稳定性检验就无须再设趋势项。

另外，第二步中变量差分滞后项的多少，可以残差项序列是否存在自相关性来判断，如果存在自相关，则应加入变量差分的滞后项。

6、案例分析

● 案例分析

例 中国居民消费的误差修正模型

经济理论指出，居民消费支出是其实际收入的函数。

1、以中国国民核算中的居民消费支出经过居民消费价格指数缩减得到**中国居民实际消费支出**时间序列（C）

2、以支出法GDP对居民消费价格指数缩减近似地代表**国民收入**时间序列(GDP)

3、时间段为1978~2000

年份	C	GDP	年份	C	GDP	年份	C	GDP
1978	3810	7809	1985	7579	14521	1992	11325	23509
1979	4262	8658	1986	8025	15714	1993	12428	27340
1980	4581	8998	1987	8616	17031	1994	13288	29815
1981	5023	9454	1988	9286	17889	1995	14693	31907
1982	5423	10380	1989	8788	16976	1996	16189	34406
1983	5900	11265	1990	9113	18320	1997	17072	36684
1984	6633	12933	1991	9977	20581	1998	18230	39008

(1) 对数据lnC与lnGDP进行单整检验

容易验证lnC与lnGDP是一阶单整的，它们适合的检验模型如下：

$$\Delta^2 \ln C_t = 0.056 - 0.744 \Delta \ln C_{t-1}$$

$$(2.76) \quad (-3.23)$$

$$\Delta^2 \ln GDP_t = 0.13 - 1.54 \Delta \ln GDP_{t-1} + 0.81 \Delta^2 \ln GDP_{t-1} + 0.59 \Delta^2 \ln GDP_{t-2} + 0.58 \Delta^2 \ln GDP_{t-3}$$

$$(3.81) \quad (-4.01)$$

$$(2.66)$$

$$(2.26)$$

$$(2.54)$$

● 案例分析

(2) 检验lnC与lnGDP的协整性，并建立长期均衡关系

首先，建立lnC与lnGDP的回归模型

$$\ln C_t = 0.047 + 0.923 \ln GDP_t$$

(0.30) (57.48)

$$R^2=0.994 \quad DW=0.744$$

发现有残差项有较强的一阶自相关性。考虑加入适当的滞后项，得lnC与lnGDP的分布滞后模型

$$\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1}$$

(1.63) (6.62) (4.92) (-2.17)

$$R^2=0.994 \quad DW=1.92$$

自相关性消除，因此可初步认为是lnC与lnGDP的长期稳定关系。

● 案例分析

其次，残差项的稳定性检验：

$$\Delta \hat{e}_t = -0.9975 \hat{e}_{t-1}$$

(-4.32)

$$R^2=0.994 \quad DW=2.01 \quad LM(1)=0.04 \quad LM(2)=1.34$$

$$t=-4.32 < -3.64 = ADF_{0.05}$$

说明lnC与lnGDP是 (1, 1) 阶协整的，下式即为它们长期稳定的均衡关系：

$$\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1}$$

(3) 建立误差修正模型

- 以稳定的时间序列 \hat{e}_t 做为误差修正项，可建立如下

误差修正模型：

$$\Delta \ln C_t = 0.686 \Delta \ln GDP_t + 0.784 \Delta \ln C_{t-1} - 0.484 \Delta \ln GDP_{t-1} - 1.163 \hat{e}_{t-1} \quad (**)$$

(6.96) (2.96) (-1.91) (-3.15)

$$R^2=0.994 \quad DW=2.06 \quad LM(1)=0.70 \quad LM(2)=2.04$$

由式 $\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1}$

可得lnC关于lnGDP的长期弹性：

$$(0.698-0.361)/(1-0.622)=0.892;$$

由 (**) 式可得lnC关于lnGDP的短期弹性： 0.686

(4) 预测1999年居民实际消费支出

由式 $\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1}$

给出1998年关于长期均衡点的偏差：

$$\begin{aligned}\hat{e}_{98} &= \ln(18230) - 0.152 - 0.698 \ln(39008) - 0.662 \ln(17072) \\ &\quad + 0.361 \ln(36684) = 0.0125\end{aligned}$$

由式 $\Delta \ln C_t = 0.686 \Delta \ln GDP_t + 0.784 \Delta \ln C_{t-1} - 0.484 \Delta \ln GDP_{t-1} - 1.163 \hat{e}_{t-1}$

预测1999年的短期波动：

$$\begin{aligned}\Delta \ln C_{99} &= 0.686(\ln(41400) - \ln(39008)) + 0.784(\ln(18230) - \ln(17072)) \\ &\quad - 0.484(\ln(39008) - \ln(36684)) - 1.163 \times 0.0125 = 0.048\end{aligned}$$

于是 $\ln C_{99} = 0.048 + \ln C_{98} = 0.048 + \ln(18230) = 9.859$

$$C_{99} = e^{9.859} = 19125$$