

§ 3.4 协整理论与误差修正模型

当许多传统的计量经济学模型在 20 世纪 70 年代的经济动荡面前预测失灵时，误差修正模型却显示了它的稳定性和可靠性。对其原因进行深入分析之后发现，误差修正模型的非稳定的单整变量之间存在一种长期稳定关系。C.J.Granger 把这种长期稳定关系称为“协整关系”，于是，一种新的理论——协整理论诞生了。虽然协整理论诞生于误差修正模型之后，但在本节中，为了便于理解，我们首先介绍协整理论，然后引出误差修正模型。

将协整理论误差修正模型作为单方程计量经济学模型的扩展的理由在于，传统的计量经济学模型是以某种经济理论或对经济行为的认识来确立模型的理论关系形式，而在这里，则是从经济变量的数据中所显示的关系出发，确定模型包含的变量和变量之间的理论关系。这是 20 世纪 80 年代以来计量经济学模型建模理论的一个重大发展。

一、单整 (Integration)

1. 稳定序列

如果一个时间序列 x_t 是稳定的，则

(1) 其均值 $E(x_t)$ 与时间 t 无关；

(2) 其方差 $\text{Var}(x_t)$ 是有限的，并不随着 t 的推移产生系统的变化。

于是，时间序列 x_t 将趋于返回它的均值，以一种相对不变的振幅围绕均值波动。

如果一个时间序列 x_t 是非稳定的，则其均值、方差将随 t 而改变。例如，随机游动序列

$$x_t = x_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0, \delta^2)$$

若 $x_0 = 0$

$$\text{则 } x_t = \sum_{i=1}^t \varepsilon_i \quad \text{Var}(x_t) = t\delta^2$$

当 $t \rightarrow \infty$ 时， $\text{Var}(x_t) \rightarrow \infty$ ，均值就无意义了，实际上序列 x_t 返回曾经达到过的某一点的期望时间是无穷大。

一个稳定序列一般以用一个自回归移动平均表达式 $\text{ARMA}(p, q)$ 表示：

$$x_t = \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \xi_{t-1} + \dots + \xi_{t-q}$$

2. 单整

如果一个序列在成为稳定序列之前必须经过 d 次差分，则该序列被称为 d 单整。记为 $I(d)$ 。换句话说，如果序列 x_t 是非稳定序列， $\Delta^d x_t$ 是稳定序列，则序列 x_t 是 $I(d)$ 。其中

$$\Delta x_t = x_t - x_{t-1}, \Delta^2 x_t = \Delta(\Delta x_t), \dots, \Delta^d x_t = \Delta(\Delta^{d-1} x_t)$$

如果有两个序列分别为 d 阶单整和 e 阶单整，即

$$x_t \sim I(d), y_t \sim I(e), e > d$$

则二序列的线性组合是 e 阶单整序列，即

$$z_t = \alpha x_t + \beta y_t \sim I(\max(d, e))$$

3. 单整的检验

对于时间序列 x_t ，建立下列方程：

$$x_t = \rho x_{t-1} + \varepsilon_t \text{ 或者 } \Delta x_t = (\rho - 1)x_{t-1} + \varepsilon_t$$

如果 ρ 不显著为 0，则序列 x_t 至少为 1 阶单整 $I(1)$ 。问题在于如何判断 ρ 是否是显著为 0。

构造 t 统计量，但这时 t 统计量服从由 Dickey 和 Fuller 于 1979 年提出的 Dickey-Fuller 分布，即 DF 分布。像第二章中介绍的变量显著性检验 t 统计量的计算一样，计算得到 t 统计量的值；从 DF 分布表中查出给定显著性水平下的临界值；如果 t 统计量的绝对值大于临界值的绝对值，则拒绝 $\rho = 0$ 假设，序列 x_t 至少为 1 阶单整 $I(1)$ 。这就是 Dickey-Fuller 检验，也称单位根检验。

通过了 1 阶单整检验后，再建立如下方程：

$$\Delta^2 x_t = (\rho - 1)\Delta x_{t-1} + \varepsilon_t$$

进行同样过程的检验，如果通过检验，则序列 x_t 至少为 2 阶单整 $I(2)$ ……直到不能通过检验为止。通过该检验，同时也就确定了序列 x_t 的单整的阶数。

一般讲，在经济数据中，表示流量的序列，例如以不变价格表示的消费额、收入等经常表现为 1 阶单整；表示存量的序列，例如以下不变价格表示的资产总值、储蓄余额等经常表现为 2 阶单整；由于价格指数的作用，也经常表现为 2 阶单整；而像利率等序列，经常表现为 0 阶单整。了解这些，对于选择什么变量进入模型是十分重要的。

在 DF 检验中，由于不能保证方程中的 ε_t 是白噪声，所以得到的 $(\rho - 1)$ 的估计值不是无偏的。于是 Dickey 和 Fuller 于 1979、1980 年对 DF 检验进行了扩充，形成了 ADF (Augment Dickey-Fuller) 检验。这是目前普遍应用的单整检验方法。由于其过程较为复杂，这里不作介绍。

二、协整 (Cointegration)

1. 定义及意义

如果序列 $X_{1t}, X_{2t}, \dots, X_{kt}$ 都是 d 阶单整，存在一个向量 $a = (a_1, a_2, \dots, a_k)$ ，使得 $Z_t = aX_t' \sim I(d - b)$ ，其中， $b > 0, X_t' = (X_{1t}, X_{2t}, \dots, X_{kt})'$ ，则认为序列 $X_{1t}, X_{2t}, \dots, X_{kt}$ 是 $(d - b)$ 阶协整，记为 $X_t \sim CI(d, b)$ ， a 为协整向量。

例如，居民收入时间序列 Y_t 为 1 阶单整序列，居民消费时间序列 C_t 也为 1 阶单整序列，如果二者的线性组合 $a_1 Y_t + a_2 C_t$ 构成的新序列为 0 阶单整序列，于是认为序列 Y_t 与 C_t 是 $(1,$

1) 阶协整。

由此可见，如果两个变量都是单整变量，只有当它们的单整阶相同时，才可能协整，例如上面的居民收入 Y_t 和居民消费 C_t ；如果它们的单整阶不相同，就不可能协整，例如居民消费 C_t 和居民储蓄余额 S_t （一般讲作为存量的居民储蓄余额 S_t 为 2 阶单整）。

三个以上的变量，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。例如，如果存在

$$W_t \sim I(1), V_t \sim I(2), U_t \sim I(2)$$

并且

$$P_t = aV_t + bU_t \sim I(1)$$

$$Q_t = cW_t + eP_t \sim I(0)$$

那么认为

$$V_t, U_t \sim CI(2,1)$$

$$W_t, P_t \sim CI(1,1)$$

从协整的定义可以看出协整的经济意义在于：两个变量，虽然它们具有各自的长波动规律，但是如果它们是协整的，则它们之间存在着一个长期稳定的比例关系。例如居民收入 Y_t 和居民消费 C_t ，如果它们各自都是 1 阶单整，并且它们是 (1, 1) 阶协整，则说明它们之间存在着一个长期稳定的比例关系，而这个比例关系就是消费倾向，也就是说消费倾向是不变的。从计量经济学模型的意义讲，建立如下消费函数模型：

$$C_t = a_0 + a_1 Y_t + \mu_t$$

变量选择是合理的，随机误差项一定是“白噪声”（即均值为 0、方差不变的稳定随机序列），模型参数有合理的经济解释。

反过来，如果两个变量，具有各自的长期波动规律，但是它们不是协整的，则它们之间就不存在着一个长期稳定的比例关系。例如居民消费 C_t 和居民储蓄余额 S_t ，由于它们单整阶数不同，所以它们不是协整的，则说明它们之间不存在着一个长期稳定的比例关系。从计量经济学模型的意义讲，建立如下消费函数模型：

$$C_t = a_0 + a_1 S_t + \mu_t$$

或者

$$C_t = a_0 + a_1 Y_t + a_2 S_t + \mu_t$$

变量选择是不合理的，随机误差项一定不是“白噪声”，模型参数没有合理的经济解释。

从这里，我们已经初步认识，检验变量之间的协整关系，在建立计量经济学模型中的重要性。而且，从变量之间是否具有协整关系出发选择模型的变量，其数据基是牢固的，其统计性质是优良的。从协整理论出发，在建立消费函数模型时，就不会选择居民储蓄余额作为居民消费的解释变量；但是，按照传统的计量经济学建模理论，从已经认识的经济理论出发

选择模型的变量，那么选择居民储蓄余额和居民收入共同作为居民消费的解释变量，不仅不感到奇怪，而且被认为是完全合理的，在本书第五章中将会看到，按照“生命周期消费理论”建立的消费函数模型正是这样的。

2. 协整的检验

(1) 两变量的 Engle-Granger 检验。为了检验两变量 Y_t, X_t 是否为协整, Engle 和 Granger 于 1987 年提出两步检验法。

第一步, 用 OLS 方法估计下列方程:

$$Y_t = aX_t + \varepsilon_t$$

得到

$$\begin{aligned}\hat{Y}_t &= \hat{a}X_t \\ \hat{e}_t &= Y_t - \hat{Y}_t\end{aligned}$$

称为协整回归。

第二步, 检验 \hat{e}_t 的单整性。如果 \hat{e}_t 为稳定序列, 则认为变量 Y_t, X_t 为 (1, 1) 阶协整;

如果 \hat{e}_t 为 1 阶单整, 则认为变量 Y_t, X_t 为 (2, 1) 阶协整……检验 \hat{e}_t 的单整性的方法即是上述的 DF 检验。

(2) 多变量协整关系的检验。上述 Engle-Granger 检验通常用于检验两变量之间的协整关系, 对于多变量之间的协整关系, Johansen 于 1988 年, 以及与 Juselius 于 1990 年提出了一种用极大或然法进行检验的方法, 通常称为 Jo-hansen 检验。

三、误差修正模型 (ECM)

误差修正模型 (Error Correction Model) 是一种具有特定形式的计量经济学模型, 它的主要形式是由 Davidson、Hendry、Srba 和 Yeo 于 1978 年提出的, 称为 DHSY 模型。为了便于理解, 我们通过一个具体的模型来介绍它的结构。

1. 误差修正模型

对于 (1, 1) 阶自回归分布滞后模型

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_t + \beta_3 z_{t-1} + \varepsilon_t \quad (3.4.1)$$

移项后得到

$$\begin{aligned}\Delta y_t &= \beta_0 + \beta_1 \Delta z_t + (\beta_2 - 1)y_{t-1} + \beta_3 z_{t-1} + \beta_1 z_{t-1} + \varepsilon_t \\ &= \beta_0 + \beta_1 \Delta z_t + (\beta_2 - 1) \left(y - \frac{\beta_1 + \beta_3}{1 - \beta_2} z \right)_{t-1} + \varepsilon_t\end{aligned} \quad (3.4.2)$$

方程 (3.4.2) 即为误差修正模型。其中 $y - \frac{\beta_1 + \beta_3}{1 - \beta_2} z$ 为误差修正项。

显然, (3.4.2) 实际上是一个短期模型, 反映了 y_t 的短期波动 Δy_t 是如何被决定的。如果变量 y 和 z 之间存在长期均衡关系, 即存在 $y = az$, 例如在 (3.4.1) 中, 若 $z = \bar{z}$, 那么 y 的均衡值与 \bar{z} 有下列均衡关系

$$\bar{y} = \frac{\beta_1 + \beta_3}{1 - \beta_2} \bar{z}$$

(3.4.2) 中的误差修正项正是与它相一致的。所以它反映长期均衡对短期波动的影响；(3.4.2) 中的差分项反映变量短期波动的影响。于是，被解释变量的波动被分成两部分：一部分为短期波动，一部分为长期均衡。

模型 (3.4.2) 可以写成

$$\Delta y_t = \beta_0 + \beta_1 \Delta z_t + \gamma ecm + \varepsilon_t \quad (3.4.3)$$

其中 ecm 表示误差修正项。由 (3.4.1) 可知，一般情况下 $|\beta_2| < 1$ ，所以有 $\gamma = \beta_2 - 1 < 0$ 。

我们可以据此分析 ecm 的修正作用：若 $(t-1)$ 时刻 y 大于其长期均衡解 $\frac{\beta_1 + \beta_3}{1 - \beta_2} z$, ecm 为

正， $\gamma \times ecm$ 为负，使得 Δy_t 减少；若 $(t-1)$ 时刻 y 小于其长期均衡解 $\frac{\beta_1 + \beta_3}{1 - \beta_2} z$, ecm 为负，

$\gamma \times ecm$ 为正，使得 Δy_t 增大。体现了长期均衡误差对 y_t 的控制。

2. ecm 与协整的关系

对于上述 (1, 1) 阶自回归分布滞后模型，如果

$$y_t \sim I(1), z_t \sim I(1)$$

那么，(3.4.2) 式左边

$$\Delta y_t \sim I(0)$$

右边的 $\Delta z_t \sim I(0)$ ，只有 y 与 z 协整，才能保证右边也是 $I(0)$ 。此时， $\frac{\beta_1 + \beta_3}{1 - \beta_2}$ 为协整

系数， $y_t - \frac{\beta_1 + \beta_3}{1 - \beta_2} z_t$ 即为均衡误差。

3. 从协整理论到误差修正模型

前面提到，实际上是先有误差修正模型，然后用协整理论去解释误差修正模型。那么在今天，我们就可以首先对变量进行协整分析，以发现变量之间的协整关系，即长期均衡关系，求出协整系数，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，加同其他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。

§ 3.5 单方程计量经济学模型的贝叶斯估计

贝叶斯 (Bayes) 统计是由 T.R.Bayes 于 19 世纪创立的数理统计的一个重要分支，20 世纪 50 年代，以 H.Robbins 为代表提出了在计量经济学模型估计中将经验贝叶斯方法与经典方法相结合，引起了广泛的重视，得到了广泛的应用。贝叶斯方法的数学描述较为复杂，内容也很广泛，在本节中主要介绍其基本概念，并尽可能用简单的语言介绍其用于计量经济学模型估计的过程，只涉及一些简单的内容，使读者对这种方法有一个概念性了解，为深入学

习与应用建立一个基础。

一、概念

1. 贝叶斯方法的基本思路

贝叶斯方法是与传统(也称经典的)计量经济学模型的估计方法相对的一种统计学方法。它的基本思路是：认为要估计的模型参数是服从一定分布的随机变量，根据经验给出待估参数的先验分布（也称为主观分布），关于这些先验分布的信息被称为先验信息；然后根据这些先验信息，并与样本信息相结合，应用贝叶斯定理，求出待估参数的后验分布；再应用损失函数，得出后验分布的一些特征值，并把它们作为待估参数的估计量。

贝叶斯方法与经典估计方法的主要不同之处是：

(1) 关于参数的解释不同。经典估计方法认为待估参数具有确定值，它的估计量才是随机的，如果估计量是无偏的，该估计量的期望等于那个确定的参数；而贝叶斯方法认为待估参数是一个服从某种分布的随机变量。

(2) 所利用的信息不同。经典方法只利用样本信息；贝叶斯方法要求事先提供一个参数的先验分布，即人们有关参数的主观认识，被称为先验信息，是非样本信息，在参数估计过程中，这些非样本信息与样本信息一起被利用。

(3) 对随机误差项的要求不同。经典方法，除了最大或然法，在参数估计过程中并不要求知道随机误差项的具体分布形式，但是在假设检验与艾是估计时是需要的；贝叶斯方法需要知道随相误差的具体分布形式。

(4) 选择参数估计量的准则不同。经典方法或者以最小二乘，或者以最大或然为准则，求解参数估计量；贝叶斯方法则需要构造一个损失函数，并以损失函数最小化为准则求得参数估计量。

2. 贝叶斯定理

贝叶斯定理是贝叶斯估计方法的理论基础。贝叶斯定理表达如下：

$$g(\theta/Y) = \frac{f(Y/\theta)g(\theta)}{f(Y)} \quad (3.5.1)$$

其中 θ 为待估参数； Y 为样本观测值信息，即样本信息； $g(\theta)$ 是待估参数 θ 的先验分布密度函数； $g(\theta|Y)$ 为 θ 的后验分布密度函数； $f(Y)$ 和 $f(Y|\theta)$ 是 Y 的密度函数。因为对 θ 而言， $f(Y)$ 可以认为是常数（样本观测值独立于待估参数）， $f(Y|\theta)$ 在形式上又同的或然函数 $L(\theta|Y)$ 一致（在最大或然法中已经了解），于是（3.5.1）可以改写为

$$g(\theta|Y) \propto L(\theta|Y) \cdot g(\theta)$$

即后验信息正比于样本信息与先验信息的乘积。（3.5.2）表明，可以通过样本信息对先验信息的修正来得到更准确的后验信息。得到后验分布的密度函数后，.就可以此为基础进行参数的点估计、区间估计与假设检验。

3. 损失函数

常用的损失函数有线性函数和二次函数。不同的损失函数，得到的参数估计值是不同的。在下面的估计过程中再作进一步的说明。

二、单方程计量经济学模型贝叶斯估计的过程

采用贝叶斯方法估计单方程计量经济学模型的过程可概括如下：

1. 确定模型的形式，指出待估参数 B

2. 给出待估参数 B 的先验分布

通常待估参数 B 是一个多维向量，即需要给出多参数的联合信息先验。如果对参数一无所知，可以认为 B 是均匀分布。实际上常用的也是均匀分布和共轭分布（尤以自然共轭分布为多）。

3. 利用样本信息，修正先验分布

利用贝叶斯定理，导出 B 的后验密度函数（密度函数简写为 p.d.f.）。到此为止，实际上已经得到 B 的所有信息。

4. 利用 B 的后验密度函数，进一步推断出 B 的点估计值，或进行区间估计与假设检验。

一般利用损失函数推断出 B 的点估计值，利用最高后验密度区间进行 B 的区间估计，利用最高后验密度区间或者利用后验优抛比进和假设检验。

5. 预测

根据求得的参数估计值进行预测。这并不是贝叶斯估计过程所必须，但对于检验估计结果是重要的。

三、正态线性单方程计量经济学模型的贝叶斯估计

下面以正态线性单方程计量经济学模型为例分析贝叶斯估计方法。选择正态线性单方程计量经济学模型，主要因为：（1）在第一章已经介绍，多元线性单方程计量经济学模型具有普遍性意义；（2）随计误差项是大量随机扰动之总和，根据中心极限定理，可以认为它是渐近正态分布；（3）计算简单，使用方便，并能完整地体现贝叶斯估计方法的主要内容。

正态线性单方程计量经济学模型又分为随计误差项方差已知和方差未知两种情况。作为方法的演示分析，我们只讨论方差已知的情况。

1. 有先验信息的后验分布

对于正态线性单方程计量经济学模型

$$Y = XB + \mu$$

其中 $\mu \sim N(0, \sigma^2 I)$ ，其他变量及参数的含义同前。

为方便起见，在下面的讨论中将只涉及正态分布的“核”，即其指数部分，而忽略其常数部分，不影响讨论结果。

选择 B 的先验分布为自然共轭分布，即 B 的密度函数和它的或然函数，以及二者结合后产生的函数服从同一分布。

B 的自然共轭先验密度函数为正态密度函数

$$g(B) \propto e^{-\frac{1}{2}(B-\bar{B})' \sum_b^{-1} (B-\bar{B})} \quad (3.5.3)$$

其中 \bar{B} 为待估参数先验均值， \sum_b 为待估参数先验协方差矩阵。

B 的或然函数 $L(B|Y)$ 等同于它的联合密度函数，即

$$L(B|Y) \propto e^{-\frac{1}{2\sigma^2}(Y-XB)'(Y-XB)} \quad (3.5.4)$$

利用贝叶斯定理，得到 B 的后验密度函数为

$$g(B|Y) \propto g(B) \cdot L(B|Y) \propto \exp \left\{ -\frac{1}{2\sigma^2} ((A^{\frac{1}{2}}B - A^{\frac{1}{2}}\bar{B})'(A^{\frac{1}{2}}B - A^{\frac{1}{2}}\bar{B}) + (Y - XB)'(Y - XB)) \right\} \quad (3.5.5)$$

其中 $A = \sigma^2 \bar{\Sigma}_B^{-1}$

令 $W = \begin{pmatrix} A^{\frac{1}{2}}\bar{B} \\ Y \end{pmatrix}_{(k+n) \times 1}$ $G = \begin{pmatrix} A^{\frac{1}{2}} \\ X \end{pmatrix}_{(k+n) \times n}$

于是有：

$$g(B|Y) \propto \exp \left\{ -\frac{1}{2\sigma^2} (W - GB)'(W - GB) \right\} \quad (3.5.6)$$

用 $\bar{\bar{B}}$ 表示待估参数后验均值， $\bar{\bar{\Sigma}}_B$ 表示待估参数后验协方差矩阵。并且应用下句结论：

$$(W - GB)'(W - GB) = (B - \bar{\bar{B}})'G'G(B - \bar{\bar{B}}) + (W - G\bar{\bar{B}})'(W - G\bar{\bar{B}}) \quad (3.5.7)$$

其中

$$\begin{aligned} \bar{\bar{B}} &= (G'G)^{-1}G'W = (A + X'X)^{-1}(A\bar{B} + X'X\beta) \\ \beta &= (X'X)^{-1}X'Y \end{aligned}$$

β 是 B 的 OLS 估计值。将 (3.5.7) 代入 (3.5.6)，代入时舍去 (3.5.7) 的右边第二项，因为其不含 B，可以作为常数项处理，只考虑核。得到

$$\begin{aligned} g(B|Y) &\propto \exp \left\{ * \frac{1}{2\sigma^2} (B - \bar{\bar{B}})'G'G(B - \bar{\bar{B}}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (B - \bar{\bar{B}})'(A + X'X)(B - \bar{\bar{B}}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (B - \bar{\bar{B}})'(\bar{\Sigma}_B^{-1} + X'X/\sigma^2)(B - \bar{\bar{B}}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (B - \bar{\bar{B}})'(\bar{\Sigma}_B^{-1})(B - \bar{\bar{B}}) \right\} \end{aligned} \quad (3.5.8)$$

式中

$$\bar{\bar{\Sigma}}_B^{-1} = \bar{\Sigma}_B^{-1} + X'X/\sigma^2 \quad (3.5.9)$$

$$\begin{aligned} \bar{\bar{B}} &= (A + X'X)^{-1}(A\bar{B} + X'X\beta) \\ &= (\bar{\Sigma}_B^{-1} + X'X/\sigma^2)^{-1}(\bar{\Sigma}_B^{-1}\bar{B} + (X'X/\sigma^2)\beta) \\ &= \bar{\bar{\Sigma}}_B^{-1}(\bar{\Sigma}_B^{-1}\bar{B} + (X'X/\sigma^2)\beta) \end{aligned} \quad (3.5.10)$$

于是，(3.5.8) 正好是均值为 $\bar{\bar{B}}$ 、方差为 $\bar{\bar{\Sigma}}_B$ 的多元正态分布的核，即 B 的后验密度

函数为

$$(B|Y) \sim N(\bar{\bar{B}}, \bar{\bar{\Sigma}}_B) \quad (3.5.11)$$

将协方差矩阵的逆 $\bar{\bar{\Sigma}}_B^{-1}$ 定义为精确度矩阵，那么可以看出：（1）后验精确度矩阵 $\bar{\bar{\Sigma}}_B^{-1}$ 是先验精确度矩阵 $\bar{\bar{\Sigma}}_B^{-1}$ 与样本信息精确度矩阵 $\mathbf{X}'\mathbf{X} / \sigma^2$ 之和，故后验精确度总是高于先验精确度总是高于先验精确度；（2）后验均值 $\bar{\bar{B}}$ 是先验均值 \bar{B} 与样本信息 OLS 估计值 β 的加权平均和，权数为高自的精确度。

2. 无先验信息的后验分布

在对待估参数一无所知的情况下，仍然可以用贝叶斯方法求得待估参数的后验分布。这时可以认为待估参数的所有元素服从 $(-\infty, +\infty)$ 上的均匀分布，且互不相关。于是， B 的先验密度函数为

$$g(B) = g(\beta_1) \cdot g(\beta_2) \cdots g(\beta_k) \propto c$$

B 的或然函数 $L(B|Y)$ 仍然与 (3.5.4) 相同，则后验密度函数与或然函数形式相同：

$$\begin{aligned} g(B|Y) &\propto g(B) \cdot L(B|Y) \propto L(B|Y) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (Y - \mathbf{X}B)'(Y - \mathbf{X}B) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} ((B - \beta)' \mathbf{X}'\mathbf{X} (B - \beta) + (Y - \mathbf{X}B)'(Y - \mathbf{X}B)) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (B - \beta)' \mathbf{X}'\mathbf{X} (B - \beta) \right\} \end{aligned} \quad (3.5.12)$$

得到的 B 的后验密度函数仍然是正态的，均值为 β ，协方差矩阵为 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ ，即

$$(B|Y) \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \quad (3.5.12)$$

从形式上看，无信息先验得到的后验分布均值与样本信息的 OLS 估计相同，但二者有不同的含义。贝叶斯结论 (3.5.13) 中的 B 是随机的，而均值 β 在样本确定后是固定的；样本信息的结论中， B 作为期望值，而 β 是随机变量，即有 $\beta \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ 。

事实上，也可以直接由 (3.5.8)、(3.5.9)、(3.5.10) 得到无信息先验下的后验密度，只要把无信息先验作为有信息先验的一种特殊情况，即无信息先验的精确度为 0。即

令 $\bar{\bar{\Sigma}}_B^{-1} = 0$ ，代入 (3.5.9)、(3.5.10) 可得到

$$\bar{\bar{\Sigma}}_B^{-1} = \mathbf{X}'\mathbf{X} / \sigma^2$$

从而 $\bar{\bar{\Sigma}}_B = \sigma^2 \mathbf{X}'\mathbf{X}$

$$\bar{\bar{B}} = (X'X / \sigma^2)^{-1} \cdot (X'X / \sigma^2) \cdot \beta = \beta$$

所以

$$(B|Y) \sim N(\beta, \sigma^2 (X'X)^{-1})$$

与 (3.5.13) 相同。

3. 点估计

在得到贝叶斯估计的后验密度函数后，即可以此为出发点，进行点估计。其思路是利用损失函数并使平均损失最小。为此需要确定一个损失函数。

先讨论一般情况。假设 \hat{B} 为 B 的估计量，损失函数为 $Lo(B, \hat{B})$ ，表示参数为 B 时采用 \hat{B} 为估计量所造成的损失，故总存在 $Lo(B, \hat{B}) \geq 0$ 。于是满足 $\text{Min}(Lo(B, \hat{B}))$ 的 \hat{B} 即是需要的估计量。由于损失函数依赖样本，故取加权平均损失，权数为后验密度函数 $g(B|Y)$ ，表示为

$$E(Lo(B, \hat{B})|Y) = \int Lo(B, \hat{B}) \cdot g(B|Y) dB \quad (3.5.14)$$

能使 (3.5.14) 所表示的加权平均后验损失最小的 \hat{B} 的值，即为 B 的贝叶斯点估计值。

下面以二次损失函数为例说明点估计过程。

损失函数的二次型为

$$Lo = (B - \hat{B})' M (B - \hat{B}) \quad (3.5.15)$$

M 为一个正定矩阵， \hat{B} 是 B 的一个估计量。为使 $E(Lo)$ 最小，

$$\begin{aligned} E(Lo) &= E(B - \hat{B})' M (B - \hat{B}) \\ &= E\left\{ (B - E(B)) - (\hat{B} - E(B)) \right\}' M \left\{ (B - E(B)) - (\hat{B} - E(B)) \right\} \\ &= E\left\{ (B - E(B))' M (B - E(B)) + (\hat{B} - E(B))' M (\hat{B} - E(B)) \right\} \end{aligned} \quad (3.5.16)$$

上述推导过程中出现的交叉项为 0，即

$$\begin{aligned} E\left\{ (B - E(B))' M (\hat{B} - E(B)) \right\}' &= E\left\{ (B - E(B))' M (\hat{B} - E(B)) \right\} \\ &= 0 \cdot M \cdot (\hat{B} - E(B)) \\ &= 0 \end{aligned}$$

考察 (3.5.16) 式，第 1 项不含 \hat{B} ，又因为 M 为一个正定矩阵，故第 2 项

$$(\hat{B} - E(B))' M (\hat{B} - E(B)) \geq 0$$

所以使 (3.5.16) 式最小的 \hat{B} 为

$$\hat{B} = E(B)$$

即二次损失函数的点估计值为后验均值。

4. 区间估计

类似于点估计，可以根据 B 的后验密度函数进行区间估计。这里需要引入最高后验密度区间的概念：区间内每点的后验密度函数值大于区间外任何一点的后验密度函数值，这样的区间称为最高后验密度区间（HPD 区间）。

对于单参数 ($k=1$) 模型的区间估计，可参照下式：

$$\int_a^b g(\beta|Y) d\beta = 1 - \alpha \quad (3.5.17)$$

其中 $(1-\alpha)$ 为置信水平。这种区间估计与经典样本信息理论中的区间估计是一致的。特别在正态分布情况下，只要稍作变换，查正态分布表便可很容易地得到结论。

对于多参数 ($k>1$) 模型的区间估计，可根据 B 的后验密度函数，求出 B 的每一元素的边际后验密度函数，再按照单参数情况求出每一参数的最高后验密度区间。

需要说明的是，参数的最高后验密度区间在形式上与经典样本信息理论中的置信区间是一致的，但解释并不相同。在贝叶斯估计中参数的最高后验密度区间的含义是参数以 $(1-\alpha)$ 的概率位于该区间内；经典样本信息理论中的置信区间的含义是该区间以 $(1-\alpha)$ 的概率包含参数。含义上的区别源于是否把待估参数作为随机变量。

5. 假设检验

可以用上述的最高后验密度区间进行假设检验，但常用的方法是利用后验优势比检验。