基于混合模型和度量误差模型的林分生长收获模型研究

李永慈　编著
唐守正　主审

东北林业大学出版社

基于混合模型和度量误差模型的林分生长收获模型研究

李永慈 编著
雷相东 主审

东北林业大学出版社

图书在版编目（CIP）数据

基于混合模型和度量误差模型的林分生长收获模型研究/李永慈编著. —哈尔滨：东北林业大学出版社，2005.6

ISBN 7-81076-731-3

Ⅰ. 基… Ⅱ. 李… Ⅲ. 林分生长—模型—参数分析 Ⅳ. S758.5

中国版本图书馆 CIP 数据核字（2005）第 054473 号

责任编辑：张红梅

封面设计：彭 宇

NEFUP

基于混合模型和度量误差模型的林分生长收获模型研究

Jiyu Hunhe Moxing He Duliang Wucha Moxing De Linfen Shengzhang Shouhuo Moxing Yanjiu

李永慈 编著

唐守正 主审

东北林业大学出版社出版发行

（哈尔滨市和兴路 26 号）

哈尔滨市工大节能印刷厂印装

开本 850×1168 1/32 印张 5.125 字数 127 千字

2005 年 6 月 第 1 版 2005 年 6 月 第 1 次 印 刷

印数 1—1 000 册

ISBN 7-81076-731-3

S·420 定价：16.80 元

内容简介

林分生长与收获模型是描述林分生长规律的一组方程式，是开展各种森林经营活动的理论基础，模型精度直接影响经营决策，如何提高模型的估计精度一直受到人们的关注。在建立生长收获模型时，各林分因子的观测值中不可避免地带有度量误差，并且往往具有复杂的误差结构，这些都会影响模型的估计精度。混合模型方法和度量误差模型方法是近代统计分析方法。混合模型通过引入随机效应，可以解决因变量具有复杂误差结构的模型参数估计问题。度量误差模型可以解决自变量和因变量都存在度量误差时的参数估计问题。本书第二章以大岗山杉木优势高为对象，用混合模型方法研究了具有复杂误差结构的生长模型。第三章以 $v=ad^b$ 和 $y=a(1-e^{-bx})^c$ 为例，研究了度量误差对模型参数估计的影响；以模型 $v=ad^b$ 为例，对有度量误差的模型的参数估计方法进行了研究。全林整体模型是一个模型系统，模型之间相互关联。第四章首先研究了度量误差对全林整体模型的影响，然后对带有度量误差的全林整体模型的参数估计方法进行了研究。

目　录

第1章 绪 论

国际林联从20世纪70年代开始召开独立的有关森林生长和收获模型的国际会议，标志着“森林生长和收获模型”已经被国际承认为一个独立的研究方向。1987年在世界林分生长模型和模拟会议上提出林分生长模型和模拟的定义(Bruce D, et al, 1987)：林分生长模型是指一个或一组数学函数，它描述林木生长与林分状态和立地条件的关系；模拟是使用生长模型去估计林分在各种特定条件下的发展。

1.1 生长收获模型的研究进展

林分生长收获模型分类方法很多，按照唐守正(1993)的分类方法，将生长收获模型分为三类：第一类模型是全林生长模型；第二类模型是径级模型；第三类模型是单木模型。本书研究的是应用最广泛的全林生长模型，特点是模型方程中自变量是林分的平均因子或总计因子。全林模型又分为两类，一类是与密度无关的模型，例如早期欧美的收获表(胡希，1977)及林分生长过程表，都是与密度无关的正常收获表；另一类模型是与密度有关的模型。从20世纪30年代末开始把密度引入收获方程(Mackinney A L, et al, 1937; Schummack F X, 1939)。这类模型现在应用较广，但形式各不相同，采用的密度指标也各不相同，模型也由单个的模型发展到模型系。

弗吉尼亚火炬松天然林可变密度模型(Sullivan A D, et al, 1972)：

$$\ln V = \beta_0 + \beta_1 L + \beta_2 T^{-1} + \beta_3 \ln G$$

其中：$\beta_0 \sim \beta_3$ 为参数；T 是林龄；L 为立地指数；G 是断面积；V 为林分蓄积。

佐治亚州湿地松人工林的可变模型：

$$\ln V = \beta_0 + \beta_1 L + \beta_2 / L + \beta_3 T^{-1} + \beta_4 \ln N$$

其中：N 为株数。

这是早期可变密度模型的例子。可以估计不同密度林分的现实材积，但不能预估今后该林分材积，也不能估计林分其他测树因子。为了估计其他测树因子，有必要引入其他关系。

印度黄檀可变密度表(Sharma R P.1973)：

$$\ln G = b_0 + b_1 \ln T \cdot \ln L + b_2 T + b_3 L + b_4 \ln N$$

上式加上 $V = f(T, L, G)$ 可估计各种不同密度林分的主要测树因子，出现了多因子同时预测问题。这些例子没有描述密度随林龄的变化，仍然是静态模型。Clutter(1963)给出了一个同时估计生长量和蓄积量的动态模型系：

$$\begin{cases} \ln V_1 = b_0 + b_1 L + b_2 (T_1^{-1}) + b_3 \ln G_1 & (1) \\ \ln G_2 = T_1 / T_2 \ln G_1 + a_0 (1 - T_1 / T_2) + a_1 L (1 - T_1 / T_2) & (2) \\ \ln V_2 = \ln V_1 + b_2 (T_2^{-1} - T_1^{-1}) + b_3 (\ln G_2 - \ln G_1) & (3) \end{cases}$$

其中：下标 1、2 表示不同时间点。

此方程系在下述两种意义下相容：固定 L，由 T_1、G_1、T_2 推出 V_2 与(1)式中直接将下标 1 换成 2 得到的 V_2 相同。由 T_1、G_1、T_2 推出 G_2，再由 T_2、G_2、T_3 推出 V_3，与 T_1、G_1、T_3 推出 V_3 相同。Clutter 首先提出了生长和收获模型的相容性问题。事实上生长收获模型是一个大的模型体系，这个系统中的各个模型从不同的侧面对生长过程进行了刻画，这些模型之间是相互关联的，然而通常都是用最小二乘法对这些模型分别进行估计，没有把它们作为一个系统来统一考虑进行参数估计。联立方程组是计量经济学方法，Phoebus J. Dhrymes(1994)的“Topics in Advanced Econometrics: Linear and Nonlinear Simultaneous

Equations"介绍了线性和非线性联立方程组在经济领域的最新研究成果,Furnival 和 Wilson(1971)建议用计量经济学中联立方程组的估计理论和方法研究生长收获模型系的参数估计问题。随着联立方程组参数估计软件的不断完善,不断有有关这方面的研究报道,例如:Murphy 和 Sternitzke(1979)、Murphy 和 Beltz(1981)对火炬松的生长收获模型系统进行了研究;Amateis 等(1984)比较了三个生长模型系统的参数估计方法;B. E. Borders 和 R. L. Bsiley(1986)用带限制的三阶段最小二乘法为沼泽松建立了相容的生长收获模型系统;Bruce E. Borders(1989)也对生长收获模型系统的参数估计方法进行了研究;Yuancai L.(2001)研究了林分材积、优势高和胸径生长模型的联合估计问题;Hubert Hasenauer 等(1998)用这种方法对单木生长模型系统的参数估计问题进行了研究。

唐守正(1991)提出的全林整体模型,是我们国内将生长收获模型作为系统来进行研究的一项重要研究成果。全林整体模型包括基本模型、基本函数式和派生模型。

(1)基本模型包括:

①断面积 G 和地位指数 L、密度指数 S、年龄 T 的关系:

$$G = b_1 \times L^{b_2} \times \{1 - \exp[-b_4 \times (S/1000)^{b_5} \times (T - t_0)]\}^{b_3}$$

其中:t_0 为平均树高达到胸高时的年龄;b_1, b_2, b_3, b_4, b_5 皆为参数。

②地位指数曲线即优势高和年龄的关系:

$$H_d = L\exp(-B/T + B/t_1)$$

其中:t_1 为基准年龄;L 和 B 为参数。

③优势高 H_d 和平均高 H_a 的关系:

$$H_a = (H_d - a_a)/b_a$$

其中:a_a 和 b_a 为参数。

④形高 H_f 和平均高 H_a 的关系:

$$H_f = [a_f + b_f/(H_a + 2)]H_a$$

其中：a_f 和 b_f 为参数。

⑤自稀疏模型：

$$\ln N = \ln S_f - \frac{1}{\gamma}\ln\left[\left(\frac{D}{D_0}\right)^{\beta\gamma} + \left(\frac{S_f}{N_1}\right)^{\gamma} - \left(\frac{D_1}{D_0}\right)^{\beta\gamma}\right]$$

(2)基本函数式：

①密度指数 S 和株数 N、平均值径 D 的关系：

$$S = N(D/D_0)^{\beta}$$

其中：D_0 为基准直径；在一定地区的某一树种 β 是常数。

②公顷断面积 G、公顷株数 N、平均直径 D 的关系：

$$G = \pi ND^2/40\,000$$

③形高蓄积公式：

$$M = G \times H_f$$

(3)派生模型：

①$D = c_1 \times L^{c_2} \times \{1 - \exp[-b_4 \times (S/1\,000)^{b_5} \times (T - t_0)]\}^{c_3} S^{c_4}$

②$N = a_1 \times L^{a_2} \times \{1 - \exp[-b_4 \times (S/1\,000)^{b_5} \times (T - t_0)]\}^{a_3} S^{a_4}$

全林整体模型能估计不同年龄、密度、立地林分的未来蓄积量和生长量，很好地解决了生长和收获的相容性问题。在此基础上展开了一系列的应用研究。李希菲等(1991)年建立了大青山主要树种的全林整体模型并进行了精度验证，洪玲霞(1993)给出了由全林整体生长模型推导林分密度控制图的方法，唐守正等(1995)对用全林整体模型计算林分纯生长量的方法及精度分析进行了研究。

1.2 生长收获模型参数估计中存在的两个问题

1.2.1 生长收获模型参数估计中存在的问题之一

为了建立生长收获模型，通常设置固定样地，对固定样地中

的观测对象进行多次观测。利用这些观测数据建立生长收获模型时，通常假定误差是服从独立同分布的。然而不同固定样地、不同时段上的观测数据很难满足独立同分布的条件。早在 1962 年 Buckman 首先注意到数据的序列相关性会影响模型的估计精度。Later、Curtis(1967)提出这些数据是多个相关序列而并非单个相关序列。Sullivan 和 Clutter(1972)也讨论了相关序列对估计精度的影响，并为模型指定了误差结构。Reynolds(1976)、Seegrist 和 Arner(1978)、Seegrist(1979,1980)都进行了相关的研究。Gregoire(1987)分析了多个样地在多个时间点上观测数据模型可能的误差结构，模型 $\boldsymbol{Y}=x\beta+e$ 为：

$$\begin{cases} \boldsymbol{Y}_{it} = \sum_{j=1}^{p} x_{itj}\beta_j + e_{it} \\ e_{it} = \alpha_i + \delta_t + v_{it}, 1 \leqslant i \leqslant n, 1 \leqslant t \leqslant T_i, 1 \leqslant j \leqslant p \end{cases}$$

其中：x_{itj} 是第 i 块样地第 j 个自变量在时刻 t 的值。$\boldsymbol{Y}_{it}$ 是因变量在第 i 块样地时刻 t 的观测值，误差 e_{it} 有三个来源组成：α_i 是样地之间的随机误差，包括样地的灌溉条件、土壤条件以及立木的生态和基因特性等因素所造成的影响，这些影响对不同样地是不同的，但在各个观测时刻是同分布的；δ_t 是时间效应误差，一般反映误差在时间上的传播，比如气候的影响，这些影响在不同的观测时刻是不同的，但是对于不同的样地是相同的；v_{it} 是样地和时间交互效应误差。误差的三个组成部分 α_i、δ_t 和 v_{it} 相互独立。由每一个误差组成特定的方差结构，可以组合成多种总方差结构。在下面考虑的四种情况中，记 $\boldsymbol{V}=cov(e)$，为了方便叙述，假定 $n=3, T_1=2, T_2=3, T_3=4$。

情况 1：样地误差独立且齐次，时间效应误差独立齐次，v_{it} 是纯随机误差。

$$E(\alpha_i\alpha_{i'})=\begin{cases}\sigma_a^2, & \text{如果 } i=i' \\ 0, & \text{如果 } i\neq i'\end{cases}$$

$$E(\delta_t\delta_{t'})=\begin{cases}\sigma_\delta^2, & \text{如果 } t=t'\\ 0, & \text{如果 } t\neq t'\end{cases}$$

$$E(v_{it}v_{it'})=\begin{cases}\sigma_v^2, & \text{如果 } i=i'\text{且 } t=t'\\ 0, & \text{如果 } i\neq i'\text{或 } t\neq t'\end{cases}$$

记：

$$e=(e_{11}\quad e_{12}\quad e_{21}\quad e_{22}\quad e_{23}\quad e_{31}\quad e_{32}\quad e_{33}\quad e_{34})'$$

则有：

$$V_1=\mathrm{cov}(e)=\Sigma_{a1}+\Sigma_{\delta 1}+\Sigma_{v1}$$

其中：

$$\Sigma_{a1}=\sigma_a^2\begin{pmatrix}\underset{2\times 2}{\mathbf{1}} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \underset{3\times 3}{\mathbf{1}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \underset{4\times 4}{\mathbf{1}}\end{pmatrix}=\sigma_a^2\begin{pmatrix}1&1&0&0&0&0&0&0&0\\1&1&0&0&0&0&0&0&0\\0&0&1&1&1&0&0&0&0\\0&0&1&1&1&0&0&0&0\\0&0&1&1&1&0&0&0&0\\0&0&0&0&0&1&1&1&1\\0&0&0&0&0&1&1&1&1\\0&0&0&0&0&1&1&1&1\\0&0&0&0&0&1&1&1&1\end{pmatrix},$$

$$\Sigma_{\delta 1}=\sigma_\delta^2\begin{pmatrix}I_{T_1\times T_1} & I_{T_1\times T_2} & I_{T_1\times T_3}\\ I_{T_2\times T_1} & I_{T_2\times T_2} & I_{T_2\times T_3}\\ I_{T_3\times T_1} & I_{T_3\times T_2} & I_{T_3\times T_2}\end{pmatrix}=\sigma_\delta^2\begin{pmatrix}1&0&1&0&0&1&0&0&0\\0&1&0&1&0&0&1&0&0\\1&0&1&0&0&1&0&0&0\\0&1&0&1&0&0&1&0&0\\0&0&0&0&1&0&0&1&0\\1&0&1&0&0&1&0&0&0\\0&1&0&1&0&0&1&0&0\\0&0&0&0&1&0&0&1&0\\0&0&0&0&0&0&0&0&1\end{pmatrix},$$

$$\Sigma_{v1}=\sigma_v^2 I_9$$

这里 $I_{s\times t}$ 表示主对角线元素为 1 其他位置元素为 0 的 $s\times t$ 阶矩

阵。这种误差结构 V_1 中有三个未知参数 σ_u^2、σ_δ^2 和 σ_v^2。

情况 2：样地误差独立且齐次，时间效应形成 AR(1)模型，v_{it} 是纯随机误差。

在情况 2 中有关样地误差和样地与时间的交互作用与情况 1 中的描述相同，只是对时间效应的描述由原来的独立同分布变为一阶自回归序列，则有：

$$V_2 = \mathrm{cov}(e) = \Sigma_{u1} + \Sigma_{\delta 2} + \Sigma_{v1}$$

其中：Σ_{u1} 和 Σ_{v1} 同情况 1。而：

$$\Sigma_{\delta 2} = \sigma_\delta^2 \begin{pmatrix} P_{T_1T_1} & P_{T_1T_2} & P_{T_1T_3} \\ P_{T_2T_1} & P_{T_2T_2} & P_{T_2T_3} \\ P_{T_3T_1} & P_{T_3T_2} & P_{T_3T_3} \end{pmatrix}$$

其中：$P_{T\tilde{T}}$ 是 $T \times \tilde{T}$ 维的一阶自回归误差结构矩阵。例如：

$$P_{34} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \end{pmatrix}$$

因此，这种误差结构 V_2 具有四个参数 σ_u^2、σ_δ^2、σ_v^2 和 ρ。

情况 3：第三种误差结构的定义直接针对误差项，而没有将误差项再分解。在第三种误差结构中：

$$\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + u_{it}$$

$$E[u_{it}] = 0$$

$$E[u_{it}^2] = \sigma_1^2$$

$$E[u_{it}u_{js}] = 0$$

$$E[\varepsilon_{i,t-1}u_{it}] = 0$$

$$E[\varepsilon_{it}\varepsilon_{jt}] = \begin{cases} \dfrac{\sigma_i^2}{1-\rho_i^2}, i = j \\ 0, i \neq j \end{cases}$$

记：

$$\varepsilon_i = \{\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT_i}\}'$$

则：

$$E[\varepsilon_i\varepsilon_i'] = \Omega_{3i} = \frac{\sigma_i^2}{(1-\rho_i^2)}\boldsymbol{P}_i$$

$$\boldsymbol{P}_i = \begin{pmatrix} 1 & \rho_i & \rho_i^2 & \cdots & \rho_i^{T_i-1} \\ \rho_i & 1 & \rho_i & \cdots & \rho_i^{T_i-2} \\ \rho_i^2 & \rho_i & 1 & \cdots & \rho_i^{T_i-3} \\ \vdots & & & \vdots & \vdots \\ \rho_i^{T_i-1} & \rho_i^{T_i-2} & \rho_i^{T_i-3} & \cdots & 1 \end{pmatrix}$$

其中：$\boldsymbol{P}(i)_{T_iT_i}$ 是以 ρ_i 为相关系数的 $T_i \times T_i$ 阶的一阶自回归误差结构矩阵。这种误差结构 V_3 中有 $2n$ 个参数、σ_i^2 和 ρ_i，($i=1, \cdots, n$)。

情况 4：四种误差结构与第三种误差结构的区别在于

$$E[u_{it}u_{js}] = \begin{cases} \sigma_{ij}, t = s \\ 0, t \neq s \end{cases}$$

因此：

$$E[\varepsilon_{it}\varepsilon_{jt}] = \begin{cases} \dfrac{\sigma_i^2}{(1-\rho_i^2)} = Q_{it}, i = j \\ \dfrac{\sigma_{ij}^2}{(1-\rho_i\rho_j)} = Q_{ij}, i \neq j \end{cases}$$

$$E[\varepsilon\varepsilon'] = \boldsymbol{\Omega}_4 = \begin{pmatrix} Q_{11}P_{11} & Q_{12}P_{12} & \cdots & Q_{1n}P_{1n} \\ Q_{21}P_{21} & Q_{22}P_{22} & \cdots & Q_{2n}P_{2n} \\ \vdots & \vdots & & \vdots \\ Q_{n1}P_{n1} & Q_{n2}P_{n2} & \cdots & Q_{nn}P_{nn} \end{pmatrix}$$

$$\boldsymbol{P}_{ij} = \begin{pmatrix} 1 & \rho_j & \rho_j^2 & \cdots & \rho_j^{T_j-1} \\ \rho_i & 1 & \rho_j & \cdots & \rho_j^{T_j-2} \\ \rho_i^2 & \rho_i & 1 & \cdots & \rho_j^{T_j-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_i^{T_i-1} & \rho_i^{T_i-2} & \rho_i^{T_i-3} & \cdots & 1 \end{pmatrix}$$

这种误差结构 V_4 中有 n^2+n 个参数、σ_{ij}^2 和 $\rho_i(i=1,\cdots,n)$。

Gregoire(1987)给出了这些误差结构中参数的估计方法，然后用广义最小二乘法估计模型的参数，并对不同误差结构下模型的参数估计精度进行了比较，结果表明当误差结构未知时，广义最小二乘估计结果反而不如一般的最小二乘估计。

唐守正(2002)用模拟实验的方法也对这个问题进行了研究，结果表明当我们选取的方差结构与真正的结构相同时，对于较大的样本，一般来说，广义最小二乘的结果优于通常最小二乘的结果。当我们选取的方差结构与真正的结构不同时(虽然相当接近)时，结果则不然，甚至样本越大广义最小二乘的结果越差，因此，谨慎使用广义最小二乘，当对方差结构有较大把握且样本数较多时使用广义最小二乘。

从理论分析上讲，在已知模型的误差结构时，使用广义线性模型可以得到模型系数的最优无偏估计值。但在实际问题中，由于下述三个原因，有时采用广义线性模型解法反而不如通常最小二乘解法。原因之一：当不知模型真正的方差参数时，用估计的方差参数代替真正的方差参数，而采用两步解法，因而增大了估计的误差。原因之二：由于模型对实际问题的近似程度不同，广义线性模型的解依赖于误差矩阵的结构，因此误差矩阵结构估计的错误可能造成广义线性模型解出现较大的偏差。原因之三：所谓的最优无偏估计，是统计意义下的结论。因此只有在大量重复或大样本时，其优良性才比较明显。

Gregoire(1987)提出的用多个样地在多个时间点上的观测数据建立生长收获模型时的四种误差结构，对提高生长收获模型的精度起了重要的作用，然而他给出的参数估计方法是不可行的，因为这些估计方法并没有使生长收获模型因为考虑了误差结构而提高精度。

混合模型是近代统计分析方法，它通过引入随机效应和误差效应，解决具有复杂误差结构的模型参数估计问题。从 20 世纪 80 年代起，国外就开始用混合模型方法研究生长收获模型。并且

不断有有关这方面的研究报道，然而我们国内这方面的研究还很少。唐守正(2002)用混合模型表示了前面四种误差结构，并对Gregoire的第三种和第四种误差结构进行了改进。在用混合模型表示这四种结构时，对于第三种和第四种误差结构引入了样地随机效应。这项研究给出了四种误差结构的线性混合模型的表示形式，但是并没有对它们的参数估计精度进行研究。用混合模型方法对于多样地多时间段的数据建立生长模型时，究竟采用哪一种误差结构最有效？如何确定随机效应？应该选择什么形式的误差效应结构？本书的第二章对这些问题进行了研究。

第二章以大岗山实验局的五个不同密度下优势高实验数据为研究对象，针对Gregoire(1987)提出的用多个样地在多个时间点上的观测数据建立生长收获模型时的四种误差结构，采用唐守正给出的这四种误差结构的线性混合模型表示，建立了混合对数舒马克生长模型，并对各种误差结构下的模型估计精度进行比较，以确定最合适的误差结构形式。以建立混合对数舒马克生长模型的最优误差结构形式中的随机效应和误差效应为依据，通过建立Logistic混合生长方程，给出一条建立混合生长模型的技术路线。

1.2.2 生长收获模型参数估计中存在的问题之二

为了建立生长收获模型，经常需要测量林分因子，林分因子的测量不可避免地存在误差，Roiko－Jokela(1978)、Kujala(1979)、Daamen(1980)、Ihalainen(1978)、Paivine, et al(1992)对林分测量因子的度量误差进行了研究，研究结果表明：胸高断面积的标准误差为0.5～8.5 mm，树高的标准误差为0.5～8.5 dm。

用通常回归方法建立生长收获模型时，总是认为自变量的观测值不含有任何误差，而因变量的观测值含有误差。因变量的误差可能有各种来源，例如抽样误差、观测误差等。但是在实际问题中，往往自变量的观测值也可能含有各种不同的误差。比如在模型系中内生变量总是带有度量误差的(R J Carronll, et al,

1995)，再比如在全林整体模型中，优势高 H_d 和平均高 H_a 的关系是 $H_a=(H_d-a_a)/b_a$，用通常的最小二乘法建立这个模型时，认为因变量优势高是存在度量误差的，而将平均高作为没有误差的变量来对待，事实上平均高和优势高一样也存在度量误差。

对于优势高和平均高这样简单的一元线性关系 $y=\beta_0+\beta_1 x$（唐守正，2002)(这里参数 β_0 和 β_1 是未知参数)，如何用(y,x)的一组观测数据(X_i,Y_i)，$i=1,\cdots,n$，来估计未知参数 β_0 和 β_1。一般来说，对每个样本单元的观测值(X_i,Y_i)并不等于它的真值(x_i,y_i)，称观测值对真值的差为度量误差。因此，$X_i=x_i+u_i$，$Y_i=y_i+e_i$，这里(u_i,e_i)是度量误差。

当 x 的观测数据没有度量误差，即 $X_i=x_i$；而 y_i 的观测数据 Y_i 包含有度量误差，即 $Y_i=y_i+e_i$，则 $y=\beta_0+\beta_1 x$ 有如下的线性模型：

$$Y_i=\beta_0+x_i\beta_1+e_i \quad (1\leqslant i\leqslant n)$$

在 y_i 的观测误差 $e_i(1\leqslant i\leqslant n)$为独立同分布的随机变量的条件下，我们知道 β_0 和 β_1 的最小二乘估计量：

$$\begin{cases}\hat{\beta}_1=\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})\\ \hat{\beta}_0=\bar{Y}-\bar{x}\hat{\beta}_1\end{cases}$$

是参数 β_1 和 β_0 的最小方差线性无偏估计量。

反过来，如果 y 的观测数据没有度量误差，即 $Y_i=y_i$，而 x_i 的观测数据 X_i 包含有度量误差，即 $X_i=x_i+e_i$，则有如下的线性模型：

$$X_i=\alpha_0+\alpha_1 y_i+e_i \quad (1\leqslant i\leqslant n)$$

其中：

$$\alpha_0=\frac{\beta_0}{\beta_1},\alpha_1=\frac{1}{\beta_1}$$

在 x_i 的观测误差 $e_i(1\leqslant i\leqslant n)$为独立同分布的随机变量条件下，我们得到参数 α_0 和 α_1 的最小方差线性无偏估计量：