HITOTSUBASHI UNIVERSITY

Center for Financial Engineering Education (CFEE)

CFEE Discussion Paper Series No. 2009-2

On a Statistical Analysis of Implied Data

Hajime Takahashi

Center for Financial Engineering Education (CFEE)
Department of Economics, Hitotsubashi University
2-1 Naka, Kunitachi Tokyo, 186-8601 Japan
http://www.econ.hit-u.ac.jp/~finmodel

On a Statistical Analysis of Implied Data

by

Hajime Takahashi
Hitotsubashi University

November 20, 2009
Revised
February 5, 2010

Abstract

We propose a method of evaluating the accuracy of the implied default probabilities. We modify the model proposed by Duffie and Singleton [1999] to allow the parametric statistical analysis. The pseudo maximum likelihood estimator is defined and to justify our method we shall prove the consistency and the asymptotic normality of the estimator. The key step is to define a pseudo score vector and apply the method of Wald [1949] and a delta method. We also introduce the bootstrap for estimating the accuracies, which is similar to that for regression models. To implement our method to the real data, we shall recommend the bootstrap rather than asymptotic normality.

Key words and Phases. implied default probability, statistical model, parametric model, pseudo maximum likelihood estimator, consistency, asymptotic normality, bootstrap, delta method.

Section 1 Introduction.

We are interested in estimating default probabilities of individual private industry and the average/representative default probability of rating classes such as AA, BB, etc.. The problem has been an important issues in credit risk problem in the last few decades. And a naive method is to use the historical data, such as the binary data telling you if a certain company is dead or not at a given day, some finance document, account data of each company, etc., on the one hand. One of the problems of in a historical method is, of course, we don't have enough data. It may be extremely difficult to have a default data for the AAA companies (we should have said "might have been" rather than "may be" in the recent economic situation though !). On the other hand, we all know that the corporate bond issued by stable companies are traded at higher prices. This comes from the fact that traders in the market certainly take their "subjective" default probabilities of the companies into account when they price the corporate bonds. Thus, by considering backward, we may use the prices of the corporate bonds as "data" to estimate the default probabilities, see for example, Duffie and Singleton [1999], Hull [2006] Ch. 29. Indeed, as we shall discuss briefly below, the spreads of the interest rates may be determined largely by the credit rating as well as the default probability of the companies being considered. Although what we shall estimate is the probabilities under the equivalent Martingale Measure (MG measure for short), they are utilized in many places, such as pricing derivatives written on the bonds (cf. Hull [2006]).

Now, in the begging, most of the rating agencies may utilize the historical data to estimate the default probabilities of companies indirectly and these probabilities are the key elements for making up the rating table in the very first time it is determined. Once the rating table is set and companies are categorizes to one of the rating classes, and after a while, we may be interested in updating a representative default probabilities of theses classes. Now, when we reestimate the default probabilities of these classes, we may use the historical data again, where we will be faced with many difficult statistical problems, such as censoring (Cox [1972], Takahashi [2009]). We then transform the estimated probabilities into the ones under the equivalent MG measure, if our purpose is to update the price of derivatives.

On the other hand, the use of implied probabilities is free from these data collecting problems and transformations. With this in mind, we shall take the latter approach in this paper.

There are several literatures discussing ways to derive the formula obtaining the default probabilities from the spread, they do not discuss the accuracy of the estimated probabilities. We shall propose a way to estimate the accuracy of the estimated default probability from the interest spread.

The key idea is to consider the implied default probability as a pseudo data and we allow the ambiguity when the corporate bond price of a defaultable company is determined in the market. Hence we suppose that the observed spreads contain error terms. We shall briefly review the results of Duffie and Singleton [1999] in Section 2, the statistical model and method is presented in Sections 3 and 4. We shall present and prove the consistency and the asymptotic normality of our estimator to justify our method in Section 5. Finally in Section 5 we introduce a bootstrap method and some empirical results are given in Section 6.

Our method makes it possible to perform the statistical inference to the individual default probability and the "representative" probability of the rating classes.

Section 2. Preliminaries

Let (Ω, F, Pr) be the probability space, on which Standard Brownian motion $\{W(t), t \geq 0\}$ is defined. We also let $\{F_t = \sigma(W(s), s \leq t)\}$, $(F_t \subset F \quad \text{for all } t \geq 0)$ be a Brownian filtration. We further suppose that $\{W(t), t \geq 0\}$ is the only source of uncertainty in our model, and there are no arbitrage opportunities in the market, and the market is assumed to be complete. These are the standing assumptions throughout the rest of the paper and won't be stated again. With this in mind, we shall review some results on the relation between the default probability and the spread of the interest rates.

Let $P(t,T) = P_k(t,T)$ $(t \leq T \leq \mathsf{T})$ be the time t price of the defaultable zero coupon bond maturing at T, where the suffix k indicates the rating class to which the company of the interest is classified. The suffix k may be omitted if there is no confusions. We also let $P^*(t,T)$ be the price of the corresponding default free bond. It follows that the yield to maturity $R(t,T) = R_k(t,T)$ and $R^*(t,T)$ are given by

$$R(t,T) = \frac{-1}{T-t} \log P(t,T)$$
$$R^*(t,T) = \frac{-1}{T-t} \log P^*(t,T)$$

respectively under the no-arbitrage conditions with the continuos compounding. The spread between the risk free and risky interest rate is thus defined by

$$\begin{aligned} \Upsilon(t) = \Upsilon(t : P^*, P) &= R(t,T) - R^*(t,T) \\ &= \frac{-1}{T-t} \log \frac{P(t,T)}{P^*(t,T)} \end{aligned} \quad (1)$$

Now to determine $P^*(t,T)$ and $P(t,T)$, we let Q be the unique equivalent Martingale measure, then the standard theory of mathematical finance asserts that

$$\frac{P^*(t,T)}{B(t)} = E^Q\{\frac{1}{B(T)}|F_t\},$$

where $B(t) = \exp\{-\int_0^t r(u)du\}$ and $r(u)$ is an instantaneous risk free spot rate and E^Q denotes the expectation under the measure Q (Harrison and Pliska [1982]).

On the other hand, the derivation of $P(t,T)$ is troublesome. Several authors have discussed the problem, and we will review here the result of Duffie and Singleton [1999] among others. Now, to determine $P(t,T)$, the default probability, as well as the proportion and the time at which the recovery of the debt is made are playing important roles.

To fix the idea, let us write $\tau = \tau_k$ be the time of the default whose survival function is given by

$$G(t) = G_k(t) = Q\{\tau > t\}$$
$$= \exp\{-\int_0^t \lambda(u)du\}, \qquad (2)$$

where $\lambda(t)$ is the hazard function (rate). We note that some authors consider the hazard rate and thus the survival function is a random process, we do assume the hazard rate (function) is a usual non random function.

By assuming that $100 \times \delta\%$ of the debt is recovered at the time of default, we have

$$P(t,T) = E^Q\{\exp\{-\int_t^T R(u)du\}|F_t\}$$

where,

$$R(u) = r(u) + \lambda(u)(1-\delta)$$

(cf. Duffie and Singleton [1999]). If δ is also assumed to be non-random, it follows that

$$\begin{aligned} P(t,T) &= E^Q\left\{\exp\{-\int_t^T R(u)du\}|F_t\right\} \\ &= E^Q\left\{\exp\{-\int_t^T r(u) + \lambda(u)(1-\delta)du\}|F_t\right\} \\ &= E^Q\left\{\exp\{-\int_t^T r(u)du\} \cdot \exp\{-(1-\delta)\int_0^t \lambda(u)du\}|F_t\right\} \\ &= P^*(t,T)G(T|t)^{(1-\delta)}, \end{aligned}$$

where $G(T|t)$ is a version of the conditional survival function $Q\{\tau > T|F_t\}$.

Now to avoid tedious conditional arguments, we shall fix the present time t, and we may suppose without losing too much generality that $T = t$ and $t = 0$. This will help us to simplify the notations and argument significantly and, hence, the survival function may be written in terms of the spread,

$$G(t) = \exp\{-t\Upsilon(t : P^*, P)\}^{\frac{1}{1-\delta}} \qquad (3)$$

[***Note* 1**] If the recovery is made at the maturity day, we have a slightly different formula:

$$G^{\#}(t) = \exp\{-t\Upsilon(t : P^*, P) - \delta\}/(1 - \delta)$$
$$= \left[\frac{P(0,t)}{P^*(0,t)}\right]^{\frac{1}{1-\delta}} \quad (4)$$

(cf. Kusuoka et. al. [2001]).//

If either relation (3) and/or (**??**) hold exactly and δ is known as well as we can observe $G(t)$ or $G^{\#}(t)$ for every t, we are all set. We have an exact estimate of the "true" survival function for each t and the distribution function of τ under the equivalent martingale measure Q.

Section 3. Statistical Models.

Even if we suppose the theory is complete, there may be many sources of errors when $P(0,t)$ is determined in the market, such as a lack of our information and a noise in the market, the assumption of being able to determine the price $P(0,t)$ without error is unrealistic. Therefore, the implied survival function is calculated with some errors and it may be reasonable to consider a statistical model for the realized (or determined price of the defaultable bond) $P(0,t)$. On the other hand, we shall suppose that $P^*(0,t)$ is observed without any error, as we shall set it the base line of the analysis. Although we start from the continuos time model, we suppose that both $P(0,t)$ and $P^*(0,t)$ are observed in the discrete fashion and that we confine ourselves to the time points $\{t = t_i, i = 1, ..., n\} \subset (0, \mathsf{T}]$. Then our observed defaultable bond prices (the prices of risky bond) determined in the market, are assumed to be

$$P(0,t_i) + \sigma h_{t_i} \varepsilon_{t_i}, \qquad i = 1, 2, ..., n \tag{5}$$

where we have set $P(0,t_i)$ are the "*theoretically true*" prices of the defaultable bonds maturing at t_i, $\varepsilon_{t_i}'s$ *are* *i.i.d.* random variables *with mean* 0 and variance 1, and σ is an unknown constant. The term h_{t_i} has an important role in our model. Since, $P(0,0) = P^*(0,0) = 1$, we need to assume $h_0 = 0$, and it may be reasonable to suppose that h_t is a increasing in t. On the other hand, it will make the whole analysis unnecessary complicated if we assume h_t is an unknown function of t. We, thus, suppose that h_t is a known function of t, or more concretely, a function of the form $h_t = h(t, P^*(0,t))$ and as examples, we sometimes consider

$$h_t = [1 - P^*(0,t)]^\alpha \tag{6}$$

or

$$h_t = [1 - P^*(0,t)]^\alpha \, P^*(0,t), \tag{7}$$

for some $\alpha > 0$. It follows that we will observe random variable $S_t = S(t)$ in stead of $G_t = G(t)$, where we have set,

$$S_t = S(t) = \{ \exp \{ \log[\frac{P(0,t)}{P^*(0,t)} + \sigma \rho_t \varepsilon_t]\}\}^{\frac{1}{1-\delta}} \tag{8}$$

and

$$\rho_t = \frac{h_t}{P^*(0,t)}$$

If we specify the function h_t as above, we have

$$\rho_t = \frac{[1 - P^*(0,t)]^\alpha}{P^*(0,t)} \quad or \quad [1 - P^*(0,t)]^\alpha \tag{9}$$

It is convenient to rewrite the equation (8) as,

$$S_t^{1-\delta} = S(t)^{1-\delta} = G_t^{1-\delta} + \sigma \rho_{t_i} \varepsilon_t \tag{10}$$

If we denote the observed S_{t_i} by $s_{t_i} = s(t_i)$ $i = 1, ...n$, a non-parametric estimator of the survival function may be obtained by smoothing the sequence $\{s(t_i)$ $i = 1, ...n\}$ and we may consider some statistical properties of the estimated curves. In some situation, however, we have extra information about the shape of the survival function, and in this case we may assume a parametric model. Formally, we shall set

[$\boldsymbol{Definition\ 1.\ Parametric\ Model}$] Let Θ be a subset of R^d, where d denotes the dimension of the parameter space. We also let

$$G(t:\ \theta) = Q\{\ \tau > t : \boldsymbol{\theta}\} \tag{11}$$

is a survival function parametrized by $d-demensional$ (unknown) vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)^{'} \in \Theta$, where $'$ denotes the transpose.//

In the standard statistical analysis, we relay on the samples generated by the assumed model (11) for performing statistical inferences. Although there are no such data here, a sequence of implied estimates of the survival functions are available and we shall use these estimates as "DATA" of the statistical inferences. We now define

[$\boldsymbol{Definition\, 2.\ Pseudo\, Data\ (RandomSample)}$] Let $\{S_{t_i} = S(t_i : \boldsymbol{\theta}), i = 1, ..., n\}$ be a sequence of implied survival functions defined in (10). And we denote the observed sequence by $\{s_{t_i}$ $i = 1, ...n\}$. We call $\{S_{t_i}$, $i = 1, ..., n\}$ a $\boldsymbol{Pseudo\ Random\ Sample}$, and $\{s_{t_i}$ $i = 1, ...n\}$ *a* $\boldsymbol{pseudo\ sample\ (data)}$.//

To estimate the unknown parameter vectors, we apply the least square criterion and we define *the pseudo maximum likelihood estimate* $\hat{\boldsymbol{\vartheta}}$ of $\boldsymbol{\theta}$ as follows;

[$\boldsymbol{Definition\ 3.\ Pseudo\ M.L.E.}$] In the parametric model (11), after having observed $\{s_{t_i},\ \ i = 1, ...n\}$, the $\boldsymbol{pseudo\ maximam\ likelihood\ estimate}$ $\hat{\boldsymbol{\vartheta}}$ of $\boldsymbol{\theta}$ is defined by

$$\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}(\boldsymbol{n}) = \arg_{\theta\ \in\ \Theta}[\min \sum_{i=1}^{n} \{\frac{G(t_i,\ \boldsymbol{\theta})^{1-\delta} - s_{t_i}^{1-\delta}}{\rho_{t_i}}\}^2] \tag{12}$$

We often omit n for simplifying the notations, unless it is necessary. The estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}\,(\boldsymbol{n})$ of $\boldsymbol{\theta}$ is also defined by (12), with $s_{t_i}^{1-\delta}$ replaced by

$S_t^{1-\delta}$. The drawback of this formulation is that t_i should be strictly bigger than 0. We shall come back to this point below.//

We shall justify our estimator (12) by showing that they are consistent and asymptotically normal in Sections 5. Now, to make the analysis easier, we shall transform the functions and variables so that the model has constant variances in time. Since, $\rho_0 = 0$, the careful readers have noticed that we have to impose some restriction when we define the homoscedastic models (12). Therefore, for some positive constant c, we let,

$$\Xi\ (t,\, \boldsymbol{\theta}\,) = \frac{G(t,\, \boldsymbol{\theta}\,)^{1-\delta}}{\rho_t} \quad for\ t > c \tag{13}$$

and

$$X_t = X(t) = \frac{S_t^{1-\delta}}{\rho_t} \quad and\ \ x_t = x(t) = \frac{s_t^{1-\delta}}{\rho_t} \quad for\ t > c$$

In practice, we may choose c

$$0 < c < \min\{t_i, i \geq 1\}.$$

With this new notations, (10) may be rewritten as

$$X(t) = \Xi\ (t,\, \boldsymbol{\theta}\,) + \sigma\ \varepsilon\ _t \tag{14}$$

and (12) becomes,

$$\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}\boldsymbol{(n)} = \arg_{\theta\ \in\ \Theta}[\min \sum_{i=1}^{n} \{\ \Xi\ (t_i,\, \boldsymbol{\theta}\,) - x_{t_i}\}^2] \tag{15}$$

[***Note* 2**] The choice of α is another interesting problem. At the maturity, both $P(0,0)$ and $P^*(0,0)$ are set equal to 1, it follows that there is no error determining $P(0,0)$ and $\rho_0 = 0$ follows immediately (cf.(9)). It suggests to us that $\alpha > 0$. If we take α large, then the weight of these data with small t become larger. We have made small data analysis on some AA company and it is found that the fitting of the Webull model is good when $\alpha = 0$. But, the choice of $\alpha = 1$ gives us a rather biased result. Based on the small data analysis we made, the mean squared error is decreasing as $\alpha \to 0$. We have observed that the MSE is stabilized in the range $\alpha \leq 0.3$. (cf. Picture 1)

[***Note* 3**] As to the value for δ, we shall simply follow the "tradition" that $\delta = 0.3$. The value also may be calculated implicitly from the "data", we do not take the position in this paper.

Section 4. Pseudo Score Vectors. Throughout the end of this paper, we shall suppose $t > c$, we do not repeat thee assumption unless it is necessary. The statistical model defined above contains several parameters, θ, α and δ. The role of α and δ are somehow different from θ, and they should be determined before the statistical analysis of the parameter θ is performed. Hence, we shall define a $d \times 1$ vector valued ***pseudo score function (vector)*** $\psi(\boldsymbol{\theta}, t, x)$ for the parameter vector $\boldsymbol{\theta}$, where we have assumed that both α and δ are given constants. Let,

$$\begin{aligned}\psi(\boldsymbol{\theta}, t, x_t) &= \frac{1}{2}[\frac{\partial}{\partial \boldsymbol{\theta}}\{\ \Xi\ (t, \boldsymbol{\theta}) - x_t\}^2] \\ &= [\frac{\partial}{\partial \boldsymbol{\theta}}\ \Xi\ (t, \boldsymbol{\theta})]\{\ \Xi\ (t, \boldsymbol{\theta}) - x_t\} \qquad (16) \\ &= \left(\left[\frac{\partial}{\partial \boldsymbol{\theta}_1}\ \Xi\ (t, \boldsymbol{\theta})\right], ..., \left[\frac{\partial}{\partial \boldsymbol{\theta}_d}\ \Xi\ (t, \boldsymbol{\theta})\right]\right)'\{\ \Xi\ (t, \boldsymbol{\theta}) - x_t\}\end{aligned}$$

If there is no confusions, we use the gradient and you may rewrite the above equation as,

$$\psi(\boldsymbol{\theta}, t, x_t) = [\nabla\ \Xi\ (t, \boldsymbol{\theta})]\{\ \Xi\ (t, \boldsymbol{\theta}) - x_t\}$$

where the gradient is taken with respect to the vector $\boldsymbol{\theta}$. And, the similar notation:

$$\psi(\boldsymbol{\theta}, t, x_t) = [\psi^{(1)}(\boldsymbol{\theta}, t, x_t), ..., \psi^{(d)}(\boldsymbol{\theta}, t, x_t)]'$$

will be used whenever it is necessary, where

$$\psi^{(l)}(\boldsymbol{\theta}, t, x_t) = \left[\frac{\partial}{\partial \boldsymbol{\theta}_l}\ \Xi\ (t, \boldsymbol{\theta})\right]\{\ \Xi\ (t, \boldsymbol{\theta}) - x_t\} \qquad (17)$$

We note that under the regularity conditions to be specified below, a ***pseuedo m.l.e.*** $\hat{\boldsymbol{\theta}}$ is obtained by the differentiation in (15), and it is given as a solution of

$$\Psi\ (\hat{\boldsymbol{\theta}}) = \Psi\ (\hat{\boldsymbol{\theta}}, \boldsymbol{n}) = \sum_{i=1}^{n} \psi(\hat{\boldsymbol{\theta}}, t, X_{t_i}) = \begin{pmatrix} \sum_{i=1}^{n} \psi^{(1)}(\hat{\boldsymbol{\theta}}, t_i, X_{t_i}) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} \psi^{(d)}(\hat{\boldsymbol{\theta}}, t_i, X_{t_i}) \end{pmatrix} = \mathbf{0}. \qquad (18)$$

We are now in the position to present the regularity conditions $[\boldsymbol{C1}]$ under which we shall obtain the basic properties of $\psi(\boldsymbol{\theta}, t, X_t) =$

$[\psi^{(1)}(\boldsymbol{\theta}, t, X_t), ..., \psi^{(d)}(\boldsymbol{\theta}, t, X_t)]'$ which are similar to that of the "**usual score**" vectors. The important relation in the usual score business, such as "the expectation of the negation of the second order derivatives and the square of the first order derivatives" do not hold here though.

[$\boldsymbol{Note\,4.\,Intuitive\,Justification\,\,of\,\,the\,Psudo\,score\,function}$]
The use of the pseudo score function may be "justified" by the following special case, where we will assume that the distribution of the error term ε_t is $N(0,1)$. Since, $X(t) = \Xi(t, \boldsymbol{\theta}) + \sigma\,\varepsilon_t$, it follows that the density function of $X(t)$ becomes

$$f(x:t,\,\theta) = \frac{1}{\sqrt{2\,\pi\,\sigma}}\exp\{\frac{-1}{2\,\sigma^2}(x - \Xi(t:\,\theta))^2\},$$

where σ is assumed to be known. And the score function is given by,

$$\frac{-1}{\sigma^2}[\frac{\partial}{\partial\,\theta}\,\Xi(t,\,\theta)]\{\,\Xi(t,\,\theta) - x\}$$

This is the same as our pseudo score function up to the constant multiplication. (cf. (16))//

[***Regularity Condition C1***]

1. The minimum in (12) is attained inside the parameter space Θ.
2. For each $t > 0$, $\Xi(\boldsymbol{\theta}, t)$ is three times continuously differentiable with respect to $\boldsymbol{\theta}$.//

We shall present some properties of the pseudo score vector function below;

[***Lemma* 1**]. Under the regularity condition [***C*1**],
(i). the expectation of the pseudo score vector $\psi(\boldsymbol{\theta}, t, S_t)$ is

$$E_\theta\{\psi(\boldsymbol{\theta}, t, X_t)\} = (0, ..., 0)' = \mathbf{0}, \text{ for all } \theta \text{ and } t > c.$$

(ii). The $(l, m)^{th}$ element of the variance covariance matrix, $Cov_\theta\{\psi(\boldsymbol{\theta}, s, X_s), \psi(\boldsymbol{\theta}, t, X_t)\}$ is given by

$$[Cov_\theta\{\psi(\boldsymbol{\theta}, s, X_s), \psi(\boldsymbol{\theta}, t, X_t)\}]^{(l,m)}$$
$$= \begin{cases} \sigma^2\left[\frac{\partial}{\partial\,\theta_l}\,\Xi(t,\,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\,\theta_m}\,\Xi(t,\,\boldsymbol{\theta})\right] \\ 0 \end{cases} \quad if \quad \begin{cases} s = t \\ else \end{cases} \tag{19}$$

In view of (19), we shall write:

$$Var_\theta\{\psi(\boldsymbol{\theta}, t, X_t)\} = \sigma^2[\nabla\,\Xi(t,\,\boldsymbol{\theta})][\nabla\,\Xi(t,\,\boldsymbol{\theta})]' \tag{20}$$

Proof. (i).From the equation (14), the l^{th} element of $E_\theta\{\boldsymbol{\psi}(\boldsymbol{\theta}, t, X_t)\}$ is

$$E_\theta\{\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}, t, X_t)\} = E_\theta\left\{\left[\frac{\partial}{\partial \theta_l}\Xi\ (t, \boldsymbol{\theta})\right]\{\ \Xi\ (t, \boldsymbol{\theta}) - X_t\}\right\}$$
$$= \left[\frac{\partial}{\partial \theta_l}\Xi\ (t, \boldsymbol{\theta})\right] E_\theta\{\sigma\epsilon_t\} = 0$$

(ii). For the covariance matrix, it is also straight forward that from the independence of X_s and X_t, for $t \neq s$, the $(l, m)^{th}$ element of the covariance matrix is given by

$$[Cov_\theta\{\boldsymbol{\psi}(\boldsymbol{\theta}, s, X_s), \boldsymbol{\psi}(\boldsymbol{\theta}, t, X_t)\}]^{(l,m)} = E_\theta\{\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}, s, X_s)\boldsymbol{\psi}^{(m)}(\boldsymbol{\theta}, t, X_t)\}$$
$$= \left[\frac{\partial}{\partial\theta_l}\Xi(s, \boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_m}\Xi(t, \boldsymbol{\theta})\right]$$
$$\times E_\theta\left(\{\Xi(s, \boldsymbol{\theta}) - X_s\}\{\Xi(t, \boldsymbol{\theta}) - X_t\}\right)$$

$$= \begin{cases} \sigma^2\left[\left(\frac{\partial}{\partial\theta_l}\Xi(t, \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_m}\Xi(t, \boldsymbol{\theta})\right)\right] & if \quad s = t \\ 0 & else \end{cases} \qquad //$$

The additivity of the variance matrix is also easily seen. We simply record the result in

[*Lemma* 2] For any $s \neq t$, we have

$$Var_\theta\{\boldsymbol{\psi}(\boldsymbol{\theta}, s, X_s) + \boldsymbol{\psi}(\boldsymbol{\theta}, t, X_t)\} = Var_\theta\{\boldsymbol{\psi}(\boldsymbol{\theta}, s, X_s)\} + Var_\theta\{\boldsymbol{\psi}(\boldsymbol{\theta}, t, X_t)\} \qquad //$$

Now, to discuss the asymptotic normality of $\hat{\boldsymbol{\theta}}(\boldsymbol{n})$, we shall expand $\Psi\ (\hat{\boldsymbol{\theta}}, \boldsymbol{n})$ into Taylor series in $\boldsymbol{\theta}$ about *the true parameter* ***vector*** $\boldsymbol{\theta}^0$, and for this purpose we need the gradient vector and the Hessian matrix of $\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}, t, X_t)$. The following lemmas are also simple, but we present them for the recording purpose as they are notationally very confusing and tedious; we should go slowly and carefully.

[*Lemma* 3] The m^{th} element of the gradient vector $\nabla\psi^{(l)}(\boldsymbol{\theta}, t, X_t)$, which is denoted by $[\nabla\psi^{(l)}(\boldsymbol{\theta}, t, X_t)]^{(m)}$, and the $(k, m)^{th}$ element $[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\psi^{(l)}(\boldsymbol{\theta}, t, X_t)]^{(k,m)}$ of the Hessian matrix $\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}, t, X_t)$ of $\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}, t, X_t)$ are given respectively by

$$[\nabla\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta},t,X_t)]^{(m)}$$
$$= \left[\frac{\partial}{\partial\theta_m}\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right]\{\Xi(t,\boldsymbol{\theta})-X_t\} + \left[\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_m}\Xi(t,\boldsymbol{\theta})\right]$$

$$\begin{aligned}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta},t,X_t)\right]^{(k,m)} &= \left[\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\theta_m}\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right]\{\Xi(t,\boldsymbol{\theta})-X_t\} \\ &+ \left[\frac{\partial}{\partial\theta_m}\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_k}\Xi(t,\boldsymbol{\theta})\right] \\ &+ \left[\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_m}\Xi(t,\boldsymbol{\theta})\right] \\ &+ \left[\frac{\partial}{\partial\theta_m}\frac{\partial}{\partial\theta_k}\Xi(t,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right] \qquad //\end{aligned} \tag{21}$$

It follows that the expectation under θ of the gradient vector and the Hessian matrix are given by,

$$E_\theta\left\{\left[\nabla\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta},t,S_t)\right]^{(m)}\right\} = \left[\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})^{1-\delta}\right]\left[\frac{\partial}{\partial\theta_m}\Xi(t,\boldsymbol{\theta})^{1-\delta}\right] \tag{22}$$

and

$$\begin{aligned}E_\theta\left\{\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta},t,X_t)\right]^{(k,m)}\right\} &= \left[\frac{\partial}{\partial\theta_m}\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_k}\Xi(t,\boldsymbol{\theta})\right] \\ &+ \left[\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_m}\Xi(t,\boldsymbol{\theta})\right] \\ &+ \left[\frac{\partial}{\partial\theta_m}\frac{\partial}{\partial\theta_k}\Xi(t,\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\right] \quad //\end{aligned} \tag{23}$$

Combining with Lemma 1, we have,

[*Lemma* 4] *For* $l, m = 1, ..., d$

$$E_\theta\left\{\left[\nabla\psi^{(l)}(\boldsymbol{\theta},t,X_t)\right]^{(m)}\right\} = \frac{1}{\sigma^2}\left[Var_\theta\{\psi(\boldsymbol{\theta},t,X_t)\}\right]^{(l,m)} // \tag{24}$$

Section 5. Asymptotic Properties.

In order to justify our estimating procedure (15) or (12), we shall prove the consistency and the asymptotic normality of our estimator $\hat{\boldsymbol{\theta}}$ under the additional regularity conditions [C2] and [C3] specified below. We shall first give a sketchy proof of the consistency by the method of Wald [1949]. For this purpose, we shall suppose that the parameter space Θ is compact subset of R^d. We also introduce the new notations. For any $\rho > 0$, let,

$$Q(t, X_t, \theta) = \{ \ \Xi \ (t, \ \boldsymbol{\theta} \ ) - X_t\}^2$$

and,

$$Q(t, X_t, \theta, \rho) = \inf_{|\theta' - \theta| < \rho} Q(t, X_t, \theta').$$

We shall now impose some regularity conditions for the consistency:

[Regularity Condition *C2*] (Conditions for the consistency)

1. For any pairs $\theta \neq \theta'$,

$$Q(t, X_t, \theta) \neq Q(t, X_t, \theta'), at\ leat\ one\ pair\ (t, X_t)$$

2. If $\lim_{i \to \infty} \theta_i = \theta$

$$\lim_{i \to \infty} Q(t, X_t, \theta_i) = Q(t, X_t, \theta)$$

for all (t, X_t) except on a set of measure zero under the true parameter vector θ^0. The null set may depend on θ, but not on the sequence $\{\theta_i, i = 1, 2, ...\}$.

3. For any pairs $\theta \neq \theta'$,

$$\Xi(t, \theta) \neq \Xi(t, \theta'), \ at\ leat\ one\ t \qquad (25)$$

//

We are now ready to present and give a heuristic proof of the consistency.

[Theorem 1] In addition to the conditions [C1] and [C2], we suppose that the parameter space is compact, then the pseudo m.l.e. $\hat{\boldsymbol{\theta}}$ is consistent estimator of $\boldsymbol{\theta}$: under the probability measure $\Pr_{\theta^0}$, we have

$$\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}^0 \quad as\ n \to \infty \qquad (26)$$

Proof (Sketch). Let, $\boldsymbol{\theta}^{0}$ be the true parameter vector, then it follows that,

$$\begin{aligned}\sigma^2 &= E_{\theta^0}\{Q(t, X_t, \theta^0)\} \\ &\leq E_{\theta^0}\{Q(t, X_t, \theta)\} \\ &= \sigma^2 + \left[\Xi\ (t,\, \boldsymbol{\theta}^{0}) - \Xi\ (t,\, \boldsymbol{\theta}\,)\}\right]^2\end{aligned} \tag{27}$$

First, we suppose the parameter space is finite,

$$\Theta = \{\theta^0, \theta^1, ..., \theta^K\}$$

Then, from (27), under the probability measure $\Pr_{\theta^0}$, the strong law of large numbers tells us that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Q(t_i, X_{t_i}, \theta^0) = \sigma^2$$

and for all $k = 1, 2, ..., K$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Q(t_i, X_{t_i}, \theta^k) = \sigma^2 + \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[\Xi\ (t_i,\, \boldsymbol{\theta}^{0}) - \Xi\ (t_i,\, \boldsymbol{\theta}^{k})\}\right]^2 \tag{28}$$

The left hand side of () is strictly larger than σ^2 by [C2]-3, it follows that

$$P_{\theta^0}\{\lim_{n\to\infty} \frac{\sum_{i=1}^{n} Q(t_i, X_{t_i}, \theta^k)}{\sum_{i=1}^{n} Q(t_i, X_{t_i}, \theta^0)} = \infty\} = 1 \text{ for all } \ k = 1, 2, ..., K \tag{29}$$

Hence, the consistency follows easily from (29).

We next consider the case where Θ is compact. Let, $\tilde{\Theta}$ be a compact subset of Θ, which does not contain the true parameter vector θ^0. Then for any $\theta \in \tilde{\Theta}$, there is a positive constant ρ_θ for which,

$$E_{\theta^0}\{Q(t, X_t, \theta, \rho_\theta)\} > E_{\theta^0}\{Q(t, X_t, \theta^0)\}. \tag{30}$$

The existence of such ρ_θ follows from [*C2*]-2, [*C2*]-3, and (25). Since, $\tilde{\Theta}$ is compact, there is a finite open covering of $\tilde{\Theta}$

$$\{O(\theta^{(l)}, \rho_{\theta^{(l)}}), l = 1, 2, ...L\}$$

where, $O(\theta, \rho)$ is a sphere with center θ,and radius ρ. Since,

$$\min_{\theta\in\tilde{\Theta}} \sum_{i=1}^{n} Q(t_i, X_{t_i}, \theta) \geq \min_{l=1,...,L} \sum_{i=1}^{n} Q(t, X_t, \theta^{(l)}, \rho_{\theta^{(l)}})$$

it suffices to prove

$$P_{\theta^0}\{\lim_{n\to\infty} \frac{\sum_{i=1}^{n} Q(t, X_t, \theta^{(l)}, \rho_{\theta^{(l)}})}{\sum_{i=1}^{n} Q(t_i, X_{t_i}, \theta^0)} = \infty\} = 1 \text{ for all } \ l = 1, 2, ..., L \tag{31}$$

By (30) and again by the strong law of large numbers, (31) follows readily and the theorem is proved. //

We shall next consider the asymptotic normality of the pseudo m.l.e. Although the assumption of compactness may be removed with little effort and the complete proof will be published elsewhere, we only have proved the consistency of our estimator under rather stringent conditions. The consistency under the more general conditions of the estimator is necessary for the asymptotic normality. Therefore, in the rest of this section, we shall suppose without proof the consistency under more general conditions. In fact, Regularity Condition [C3]-5 below includes the weak consistency of the estimator.

To prove the asymptotic normality, we shall use a delta method which is similar to that of applied to the regression problems. And for this purpose, we shall first expand $\Psi\,(\hat{\boldsymbol{\theta}})$ about the true parameter vector $\boldsymbol{\theta}^{0}$. It follows that

$$0 = \Psi\,(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \boldsymbol{\psi}(\hat{\boldsymbol{\theta}}, t_i, X_{t_i}) \tag{32}$$
$$= \sum_{i=1}^{n} \psi\,(\boldsymbol{\theta}^{0}, t_i, X_{t_i}) + \frac{\partial}{\partial \boldsymbol{\theta}} \Psi\,(\boldsymbol{\theta}^{0})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}) + \boldsymbol{Rem},$$

where we have set

$$\frac{\partial}{\partial \boldsymbol{\theta}} \Psi\,(\boldsymbol{\theta}^{0}) = \begin{pmatrix} \sum_{i=1}^{n} \left[\nabla \psi^{(1)}(\boldsymbol{\theta}^{0}, t_i, X_{t_i})\right]' \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} \left[\nabla \boldsymbol{\psi}^{(d)}(\boldsymbol{\theta}^{0}, t_i, X_{t_i})\right]' \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_1} \boldsymbol{\psi}^{(1)}(\boldsymbol{\theta}^{0}, t_i, X_{t_i}), ..., \sum_{i=1}^{n} \frac{\partial}{\partial \theta_d} \psi^{(1)}(\boldsymbol{\theta}^{0}, t_i, X_{t_i}) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} \frac{\partial}{\partial \theta_1} \psi^{(d)}(\boldsymbol{\theta}^{0}, t_i, X_{t_i}), ..., \sum_{i=1}^{n} \frac{\partial}{\partial \theta_d} \psi^{(d)}(\boldsymbol{\theta}^{0}, t_i, X_{t_i}) \end{pmatrix}$$

and

$$\boldsymbol{Rem} = \frac{1}{2} \begin{pmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{0})' \sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{\psi}^{(1)}(\boldsymbol{\theta}^{(1)*}, t_i, X_{t_i})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{0}) \\ \cdot \\ \cdot \\ \cdot \\ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{0})' \sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{\psi}^{(d)}(\boldsymbol{\theta}^{(d)*}, t_i, X_{t_i})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{0}) \end{pmatrix}$$

,

$$= \frac{1}{2}\begin{pmatrix} (\hat{\theta} - \theta^0)' & 0 & \cdot & 0 & 0 \\ 0 & (\hat{\theta} - \theta^0)' & & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & (\hat{\theta} - \theta^0)' \end{pmatrix}\begin{pmatrix} \sum_{i=1}^n (\frac{\partial^2}{\partial \theta^2}\boldsymbol{\psi}^{(1)}(\theta^{(1)*}, t_i, X_{t_i}) \\ \sum_{i=1}^n (\frac{\partial^2}{\partial \theta^2}\boldsymbol{\psi}^{(2)}(\theta^{(2)*}, t_i, X_{t_i}) \\ \cdot \\ \cdot \\ \sum_{i=1}^n (\frac{\partial^2}{\partial \theta^2}\boldsymbol{\psi}^{(d)}(\theta^{(d)*}, t_i, X_{t_i}) \end{pmatrix}(\hat{\theta} - \theta^0)$$

$$= \frac{1}{2} I \otimes (\hat{\theta} - \theta^0)' \begin{pmatrix} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2}\boldsymbol{\psi}^{(1)}(\theta^{(1)*}, t_i, X_{t_i}) \\ \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2}\boldsymbol{\psi}^{(2)}(\theta^{(2)*}, t_i, X_{t_i}) \\ \cdot \\ \cdot \\ \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2}\boldsymbol{\psi}^{(d)}(\theta^{(d)*}, t_i, X_{t_i}) \end{pmatrix}(\hat{\theta} - \theta^0)$$

where $\theta^{(l)*}$ are intermediate vectors with $\| \theta^{(l)*} - \theta^0 \| \leq \| \hat{\theta} - \theta^0 \|$ for $l = 1, ..., d$ and $\| \cdot \|$ is the usual Euclidean distance, and $\otimes$ is the Kronecker's product. We shall now present the regularity conditions for the asymptotic normality:

[Regularity Condition *C3*]

1. (Conditions for the convergence of covariance matrices): The pseudo data will be sampled in a way that the time sequence $\{t_i, i = 1, 2, ...\}$ are selected so that there is a positive definite matrix $\Sigma(\theta^0)$ and $B(\theta^0)$ for which

(i)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Var\{\boldsymbol{\psi}(\theta^0, t_i, X_{t_i})\} = \lim_{n\to\infty} \sigma^2\Sigma(\theta^0, n) = \sigma^2\Sigma(\theta^0) \qquad (Say) \tag{33}$$

where $(l, m)^{th}$ element of $\Sigma(\theta^0, n) = \frac{1}{n\sigma^2}\sum_{i=1}^{n} Var\{\psi(\theta^0, t_i, S_{t_i})\}$ is defined by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \theta_l}\Xi(t_i, \boldsymbol{\theta}^0)\frac{\partial}{\partial \theta_m}\Xi(t_i, \boldsymbol{\theta}^0)$$

and,

(ii)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}^0, t_i, X_{t_i}))\right] = \lim_{n\to\infty} B(\theta^0, n) = B(\theta^0) \quad (Say) \tag{34}$$

where $(l,m)^{th}$ element of $\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial \mu^2}\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}^0, t_i, X_{t_i})\right]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^3}{\partial\theta_k\partial\theta_m\partial\theta_l}\Xi(t_i,\boldsymbol{\theta}^0)\right]\left\{\Xi(t_i,\boldsymbol{\theta}^0)-X_t\right\}+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_m\partial\theta_l}\Xi(t_i,\boldsymbol{\theta}^0)\right]\left[\frac{\partial}{\partial\theta_k}\Xi(t_i,\boldsymbol{\theta}^0)\right] \tag{35}$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_k\partial\theta_l}\Xi(t_i,\boldsymbol{\theta}^0)\right]\left[\frac{\partial}{\partial\theta_m}\Xi(t_i,\boldsymbol{\theta}^0)\right]+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_m\partial\theta_k}\Xi(t_i,\boldsymbol{\theta}^0)\right]\left[\frac{\partial}{\partial\theta_l}\Xi(t_i\boldsymbol{\theta}^0)\right]$$

and $(l,m)^{th}$ element of $B(\boldsymbol{\theta}^0, n)$ is defined by

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_m\partial\theta_l}\Xi(t_i,\boldsymbol{\theta}^0)\right]\left[\frac{\partial}{\partial\theta_k}\Xi(t_i,\boldsymbol{\theta}^0)\right]+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_k\partial\theta_l}\Xi(t_i,\boldsymbol{\theta}^0)\right]\left[\frac{\partial}{\partial\theta_m}\Xi(t_i,\boldsymbol{\theta}^0)\right] \tag{36}$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_m\partial\theta_k}\Xi(t_i,\boldsymbol{\theta}^0)\right]\left[\frac{\partial}{\partial\theta_l}\Xi(t_i,\boldsymbol{\theta}^0)\right]$$

Note that the difference between (35) and (36) comes from the weak law of large numbers and (38) below. The matrices $\Sigma(\boldsymbol{\theta}^0)$ and $B(\boldsymbol{\theta}^0)$ may depend on the sequence $\{t_i, i = 1, 2, ...\}$.

2. For every $\boldsymbol{\theta} \in \Theta$, the time sequence $\{t_i, i = 1, 2, ...\}$ are selected so that

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta_l\partial\theta_m}\Xi(t_i,\boldsymbol{\theta})\right]^2 \quad \textit{converges as } n\to\infty, \quad \textit{for all } l \textit{ and } m, \tag{37}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^3}{\partial\theta_k\partial\theta_m\partial\theta_l}\Xi(t_i,\boldsymbol{\theta}^0)\right]^2 \quad \textit{converges as } n\to\infty, \quad \textit{for all } k,\ l \textit{ and } m, \tag{38}$$

$$\tag{39}$$

3. Under the condition (33) the sequence

$$\left\{\boldsymbol{\psi}(\boldsymbol{\theta}^0, t_i, X_{t_i}) = [\boldsymbol{\psi}^{(1)}(\boldsymbol{\theta}^0, t_i, X_{t_i}), ..., \boldsymbol{\psi}^{(d)}(\boldsymbol{\theta}^0, t_i, X_{t_i})]', i = 1, 2, ...\right\}$$

obeys the multivariate central limit theorem;

$$\left[\sum_{i=1}^{n} Var\{\boldsymbol{\psi}(\boldsymbol{\theta}^0, t_i, X_{t_i})\}\right]^{-\frac{1}{2}} \sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{\theta}^0, t_i, X_{t_i}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}_d) \quad \textit{as } n\to\infty$$

where $\mathbf{0} = (\mathbf{0}, ..., \mathbf{0})'$ is the d dimensional *zero* vector and $\mathbf{I}_d$ is the $d\times d$ identity matrix.

4. For any bounded symmetric matrix $D^{(l)}, l = 1, ..., d$, the rate of convergence of the pseudo M.L.E. to the true parameter vector $\boldsymbol{\theta}^0$ is fast enough so that we have, $for\ each\ l = 1, ..., d$

$$\sqrt{n}\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)' \left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\boldsymbol{\psi}^{(l)}(\boldsymbol{\theta}^{(0)}, t_i, X_{t_i})\right]\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\right\} \xrightarrow{p} 0 \quad (40)$$

as $n \to \infty$. Here bounded matrix means that every element of the matrix is bounded. //

[Theorem 2] Under the conditions [C1],and [C3],

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}^{\mathbf{0}})^{-1}) \quad as\ n \to \infty$$

Proof. We first rearrange the terms (32) and multiply both side by $\sqrt{n}$, we have

$$\begin{aligned} &\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\ &= -\left[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\theta}}\Psi(\boldsymbol{\theta}^0, n)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{\theta}^0, t_i, X_{t_i}) - \left[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\theta}}\Psi(\boldsymbol{\theta}^0)\right]^{-1}\frac{1}{\sqrt{n}}\boldsymbol{Rem} \\ &= \mathrm{I} + \mathrm{II} \qquad (say). \end{aligned} \quad (41)$$

To analyses the denominator of I , note that the $(l, m)^{th}$ element of $\frac{1}{n}\left[\frac{\partial}{\partial \mu}\Psi(\boldsymbol{\theta}^0, n)\right]$ is (by Lemma 3)

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial \theta_l}\Xi(t_i, \boldsymbol{\theta})\right]\left[\frac{\partial}{\partial \theta_m}\Xi(t_i, \boldsymbol{\theta})\right] + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial \theta_l \partial \theta_m}\Xi(t_i, \boldsymbol{\theta})\right]\{\Xi(t_i, \boldsymbol{\theta}) - X_{t_i}\}.$$

The convergence of $\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial \theta_l \partial \theta_m}\Xi(t_i, \boldsymbol{\theta})\right]^2$ (cf. [C3]-2) guarantees that the second term in the above equation converges to zero in probability. Hence, as $n \to \infty$, we have

$$\left[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\theta}}\Psi(\boldsymbol{\theta}^0, n)\right] \backsim \left[\frac{1}{n\sigma^2}\sum_{i=1}^{n} Var\{\boldsymbol{\psi}(\boldsymbol{\theta}^0, t_i, X_{t_i})\}\right] \quad (42)$$

And eventually $\left[\frac{1}{n}\frac{\partial}{\partial \mu}\Psi(\boldsymbol{\theta}^0, n)\right]$ converges to $\left[\sigma^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}^{\mathbf{0}})\right]$ in probability as n gets larger. To analyses the numerator of I , in view of (42), we write I as,

$$-\left[\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\theta}}\Psi(\boldsymbol{\theta}^0, n)\right]^{-\frac{1}{2}}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n} Var\{\boldsymbol{\psi}(\boldsymbol{\theta}^0, t_i, X_{t_i})\}\right]^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{\theta}^0, t_i, X_{t_i})$$

Hence, combined with [C3]-3, we have

$$\mathrm{I} \xrightarrow{d} -\left[\boldsymbol{\Sigma}(\boldsymbol{\theta}^{\mathbf{0}})\right]^{-\frac{1}{2}} N(\mathbf{0}, \sigma^2 \boldsymbol{I}_d) \text{ as } n \to \infty$$

We shall next consider II. As above, the denominator converges to $\boldsymbol{\Sigma}(\boldsymbol{\theta}^{\mathbf{0}})$. Since $\theta^{(l)*}$ converges to θ^0 in probability fort each $l = 1, 2, ..., d$, the numerator is proved to be equivalent to

$$\sqrt{n}\left\{(\hat{\theta} - \theta^0)'\left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial \theta^2}\boldsymbol{\psi}^{(l)}(\theta^{(0)}, t_i, X_{t_i})\right]\right)(\hat{\theta} - \theta^0)\right\}$$

and thus by the condition [C3]-4, II converges to zero in probability as $n \to \infty$. We have proved the theorem. //

[***Notes on the conditions*** [***C1***] ***and*** [***C3***]] Condition [***C1***] is a standard one and it looks OK. But, as some readers may have already noticed that some of the conditions in [***C3***] are strong and look strange. If the time sequence $\{t_i,\ i = 1, 2, ...\}$ is chosen equally spaced, the convergence of $\frac{1}{n}\sum_{i=1}^{n} Var\{\boldsymbol{\psi}(\theta^0, t_i, X_{t_i})\}$ may be guaranteed by the following integral comparison test:

$$\frac{1}{n}\sum_{i=1}^{n} Var\{\boldsymbol{\psi}(\theta^0, t_i, X_{t_i})\} \to \int_c^{\mathsf{T}} |\frac{\partial}{\partial \theta_l}\Xi(t, \boldsymbol{\theta}^0)\frac{\partial}{\partial \theta_m}\Xi(t, \boldsymbol{\theta}^0)|dt < \infty,\ as\ n \to \infty$$

Similar integral representation will be possible for the other terms related to $\frac{\partial^2}{\partial \theta_m \partial \theta_l}\Xi(t, \boldsymbol{\theta})$, $\frac{\partial^3}{\partial \theta_k \partial \theta_m \partial \theta_l}\Xi(t, \boldsymbol{\theta})$, and $\left[\frac{\partial^2}{\partial \theta_l \partial \theta_m}\Xi(t, \boldsymbol{\theta})\right]^2$. //

Now, the main mission of Theorems 1 and 2 are to justify our estimation method and we do not intend to use the theorems to construct a confidence regions and/or testing in the application. For the practical purposes, we shall invoke bootstrap to evaluate the accuracy of the estimate. We shall discuss this in the next section more extensively.

Section 6. Bootstrap.

When we implement our method to the real data, it is easy to obtain the pseudo m.l.e. $\hat{\boldsymbol{\vartheta}}$ of $\boldsymbol{\theta}$ by (18) numerically. And, theoretically, we can derive a confidence region of $\boldsymbol{\theta}$ from Theorem 1. The approximated $100(1-\alpha)\%$ confidence region (interval) is given by

$$(\boldsymbol{\theta} - \hat{\theta})'\boldsymbol{\Sigma}(\hat{\theta})(\boldsymbol{\theta} - \hat{\theta}) \leq \frac{\sigma^2}{n} c_{1-\alpha}, \tag{43}$$

where $c_{1-\alpha}$ is defined by $\Pr\{Z'Z \leq c_{1-\alpha}\} = 1-\alpha$, and Z is a d dimensional standard normal random vector. The difficulties arise when we calculate (approximate) $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The $(l,m)^{th}$ element of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ may be approximated by $\int_c^T \frac{\partial}{\partial\theta_l}\Xi(t,\boldsymbol{\theta})\frac{\partial}{\partial\theta_m}\Xi(t,\boldsymbol{\theta})dt$, and most of the cases these integrals are not given in the closed form and, therefore, the confidence region may be obtained only numerically. Likewise, we can obtain the confidence intervals of linear combinations of the elements of $\boldsymbol{\theta}$ only numerically in many of the applied aspects. We also note that the accuracy of the central limit theorem depends heavily on the sample size, and in our situation, we cannot expect a good normal approximation, as we do not have enough samples.

Hence, when we come up to analyses the real data, we recommend the reader to use bootstrap rather than asymptotic theory. Although, we have not fully justified the validity of the bootstrap, the results of previous section suggests that the bootstrap will works on this problem too. The method we apply here is similar to that of bootstrap for regression model. We simply describe the procedure as follows. We shall suppose the parametric model for continuity of the arguments, the assumption may be removed without too much effort. Needless to say, we suppose the model (14).

[*Bootstrap*]We let $\{x_{t_i} = x(t_i)\ i = 1, ...n\}$, be a sequence of the pseudo data, where $x_t = x(t) = \frac{s_t^{1-\delta}}{\rho_t}$. As in (15), we estimate the base line estimate $\hat{\boldsymbol{\vartheta}}$ by

$$\hat{\boldsymbol{\vartheta}} = \arg_{\theta \in \Theta}[\min \sum_{i=1}^{n}\{\ \Xi\ (t_i,\ \theta) - x_{t_i}\}^2]$$

where $\Xi\ (t,\ \boldsymbol{\theta}) = \frac{G(t,\theta)^{1-\delta}}{\rho_t}$. We estimate $\{\sigma\ \varepsilon\ _{t_i}\ i = 1, ..., n\}$ by

$$e_{t_i}^* = \Xi\ (t_i, \hat{\boldsymbol{\vartheta}}) - x_{t_i}, \qquad i = 1, ..., n \tag{44}$$

And we define the set E of "*standerdized*" residuals, from which the bootstrap sampling will be made (cf. Wu [1986]):

$$E = \{e_{t_1}, ..., e_{t_n}\}$$

where

$$e_{t_i} = \sqrt{\frac{n}{n-1}}\left(e_{t_i}^* - \bar{e}^*\right),\ and\ \bar{e}^* = \frac{1}{n}\sum_{i=1}^{n} e_{t_i}^*.$$

0^{th} *Step.* Set

$$m = 1$$

1^{st} *Step.* We choose n elements at random from E with replacement, denote it $E^{*(m)}$

$$E^{*(m)} = \{e_{t_1}^{*(m)}, ..., e_{t_n}^{*(m)}\}$$

2^{nd} *Step.* Construct a bootstrap pseudo sample by

$$x_{t_i}^{*(m)} = \Xi\ (t_i, \hat{\boldsymbol{\vartheta}}) + e_{t_i}^{*(m)} \quad i = 1, ..., n$$

3^{rd} *Step* Estimate $\boldsymbol{\theta}$ from the pseudo data $\{x_{t_i}^{*(m)}, i = 1, ...n\}$ and denote it as $\hat{\boldsymbol{\vartheta}}^{*(m)}$

$$\hat{\boldsymbol{\vartheta}}^{*(m)} = \arg_{\theta\ \in\ \Theta}[\min \sum_{i=1}^{n} \{\ \Xi\ (t_i,\ \boldsymbol{\theta}\ ) - x_{t_i}^{*(m)}\}^2]$$

4^{th} *Step.* with $m = m + 1$, Repeat the Step 1st to 3rd M times, and denote the estimate by

$$\hat{\boldsymbol{\vartheta}}^{*(1)}, ..., \hat{\boldsymbol{\vartheta}}^{*(M)}$$

and we define bootstrap estimate of *by*

$$\tilde{\boldsymbol{\vartheta}}^{(Boots)} = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\vartheta}}^{*(m)} \tag{45}$$

5^{th} *Step.* Construct the estimate of bias and variance of $\hat{\boldsymbol{\vartheta}}$ by

$$Bias^{(Boots)}(\hat{\boldsymbol{\vartheta}}) = \tilde{\boldsymbol{\vartheta}}^{(Boots)} - \hat{\boldsymbol{\vartheta}} \tag{46}$$

and

$$Var^{(Boots)}(\hat{\boldsymbol{\vartheta}}) = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\boldsymbol{\vartheta}}^{*(m)} - \tilde{\boldsymbol{\vartheta}}^{(Boots)})(\hat{\boldsymbol{\vartheta}}^{*(m)} - \tilde{\boldsymbol{\vartheta}}^{(Boots)})' \tag{47}$$

respectively. We also define the bootstrap estimate of the distribution function of $\hat{\boldsymbol{\vartheta}}$ by

$$1 - G^{(Boots)}(t_1, ..., t_d) = \frac{1}{M} \#\{m : \hat{\boldsymbol{\vartheta}}^{*(m)} \leq (t_1, ..., t_d)'\} \tag{48}$$

Of course, the marginal distribution may be obtained much easier. And, the naive confidence region for θ is given by the (**??**). In the next section, we shall apply our estimation procedure together with the bootstrap to the real data.

Section 7. Numerical Examples.

In this section, we shall apply our statistical model and the estimation procedure to the bond prices of ACOM, TOYOTA as examples for the individual companies. We also study the representative default probabilities of classes AAA and BBB. We shall estimate the default probabilities of these companies as well as the some rating classes as of April 15, 2003. Although, when we implement the method, regularity conditions C1, C2 and C 3 are not fully investigated, the results in this section suggest us the validity of our method numerically.

The parametric models we shall consider are (i) Exponential and (ii) Webull models. We denote the survival function of the exponential $G_{exp}(t,\theta)$ and of the Weibull by $G_{Wei}(t,\boldsymbol{\theta})$. Here, we have set

$$G_{exp}(t,\theta) = \exp\{-t\theta\}, \quad t \geq 0$$
$$G_{Wei}(t,\boldsymbol{\theta}) = \exp\{-(t\theta_1)^{\theta_2}\}, \quad t \geq 0, \text{ and } \boldsymbol{\theta = (\theta_1, \theta_2)}$$

respectively. The detail of the analysis are presented in the pictures and the tables, we shall simply summarize the rough observation. We have first apply both Exponential and Weibull distributions to the data from ACOM and TOYOTA with $\alpha = \frac{1}{3}$ and $\delta = 0.3$. The MSE from the Exponential distribution are larger than that of Weibull and this can be also seen visually in the Pictures 2 and 3 for ACOM, and Pictures 6 and 7 for TOYOTA. It is, of course, a natural consequence, as Weibull allows two parameters, whereas the exponential has only one. We, thus, apply Weibull for the rest of the analysis (AAA and BBB). The pictures of the estimated survival functions for AAA and BBB are presented in the Pictures 10 and 14 respectively. The bootstrap estimate of the histogram of each parameters for ACOM are given in the Pictures 4 and 5. The corresponding normal curve is superimposed in Picture 4 to see how the asymptotic normality of the estimator is attained by intuitive manner. The results from ACOM and BBB (Pictures 15 and 16) look good in the sense the histograms are well approximated by the normal (We do need to test them statistically though). The results from TOYOTA are disastrous (Pictures 8 and 9). Also, Pictures 12 and 13 show us that the distributions are rather biased to the left. These may be explained by the number of "data". We only have 3 for TOYOTA and 7 for AAA. This small analysis shows the statistical method we proposed seem to attain the satisfactory results.

Acknowledgement: The author thanks to Reiko Tobe for her superb help obtaining the all numerical results in Section 7. Thanks are also go to both Katsumi Shimotsu and Eiji Kurozumi for their comments on the consistency and bootstrap.

Ref erences

[1] Cox, D.R. [1972] "Regression Models and Life-Tables" *Journal of the Royal Statistical Society,* Ser. B 34, 187-202

[2] Duffie, D. and K. Singleton [1999] "Modeling Term Structures of Defaultable Bonds" *Revie of Financial Studies* 12, 678-720

[3] Efron, B. [1979] "Bootstrap Methods: Another Look at the Jackknife" *Annals of Statistics* 7, 1-26

[4] Harrison, M. and S. Pliska [1981] "Martingales and Stochastic Integrals in the Theory of Continuous Trading" *Stochastic Processes and Their Applications,* 11, 215-260.

[5] Hull, J. C. [2006] "*Options, Futures, and Other Derivatives*, 6th Ed. " Pearson Education, Inc., New Jersey

[6] Kusuoka, N., K. Aonuma, and H. Nakagawa [2001] "*Credit Risk Models*" (Japanese) Kinyu Zaisei Jijyou Kenkyu-kai, Tokyo

[7] Takahashi, H. [2009] "Statistical Analysis of Implied Data" (Japanese) CFEE Discussion Paper Series No. 2009-1 (Hitotsubashi University Kunitachi JApan)

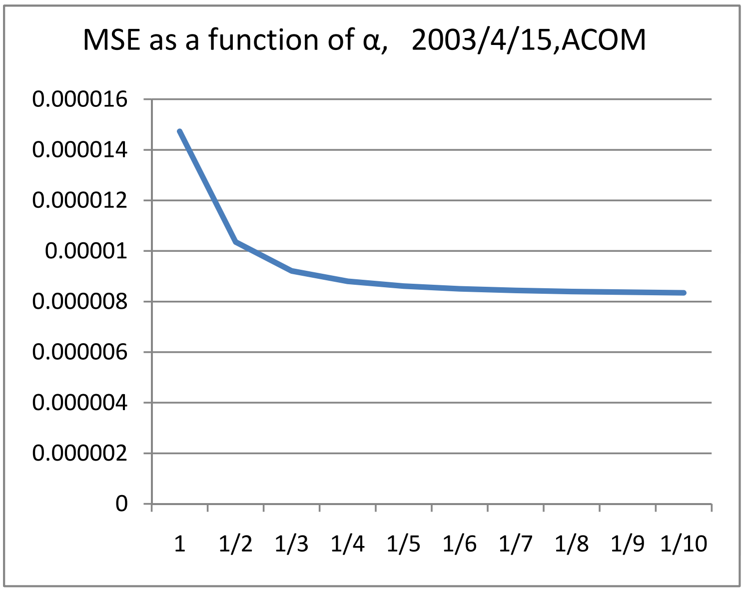
[8] Wald, A. [1949] "Note on the consistency of the maximum likelihood estimate" *Annals of Mathematical Statistics,* 20, 595-601.

[9] Wu, C.F.J.[1986] "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis" *Annals of Statistics* 14, 1261-1295.

Pictures and Tables

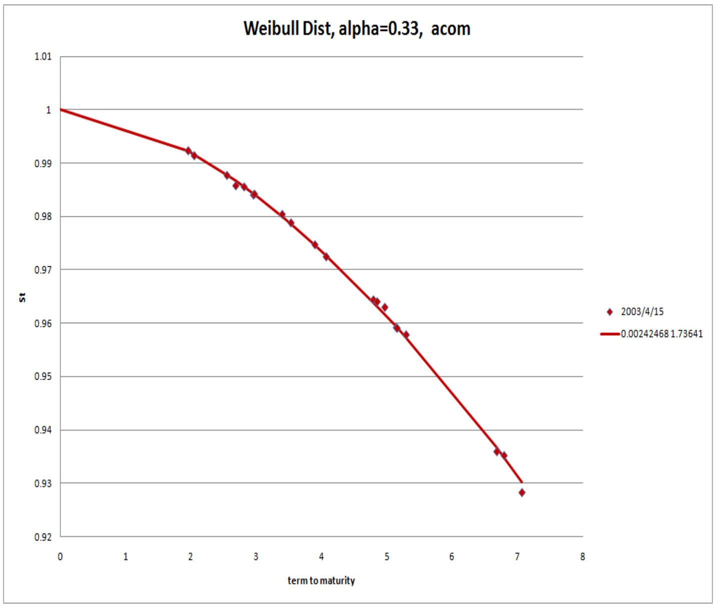
Ⅰ ACOM

Ⅰ－i Selection of α

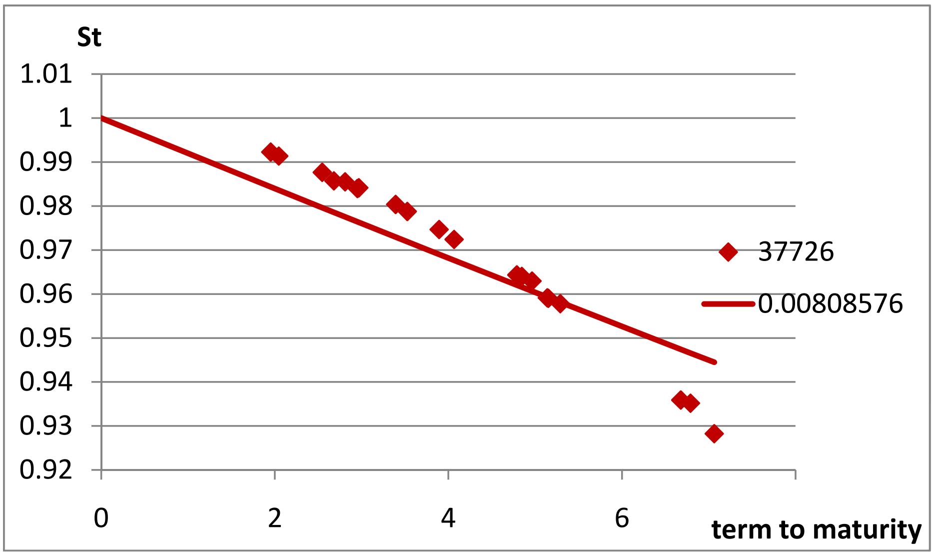


Picture 1 MSE as a function of α

Ⅰ－ii Survival Functions:

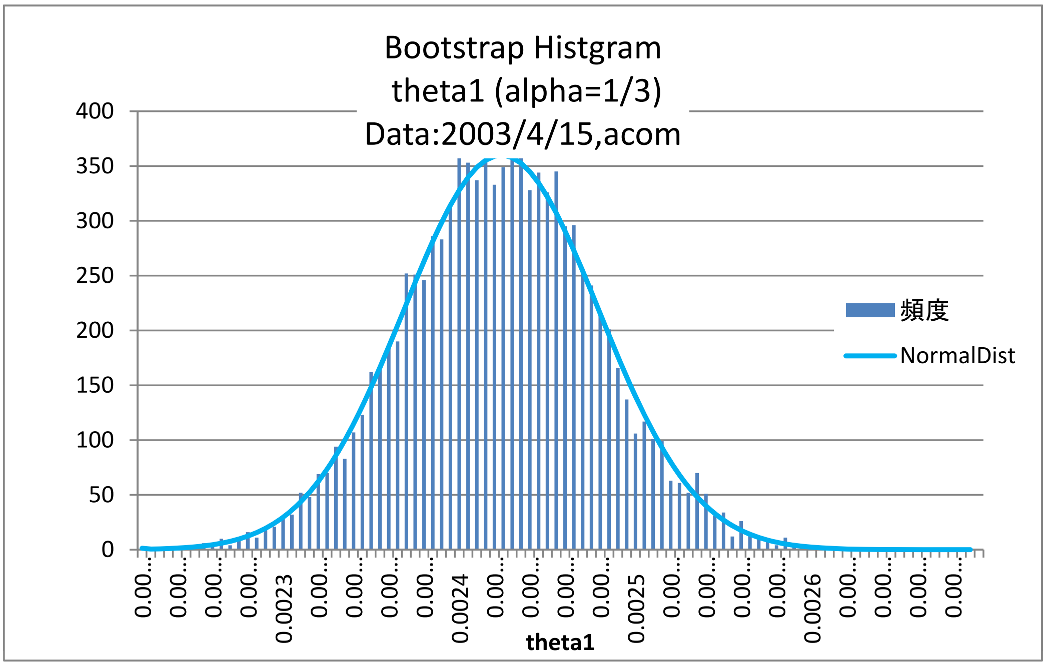


Picture 2 Weibull Distribution

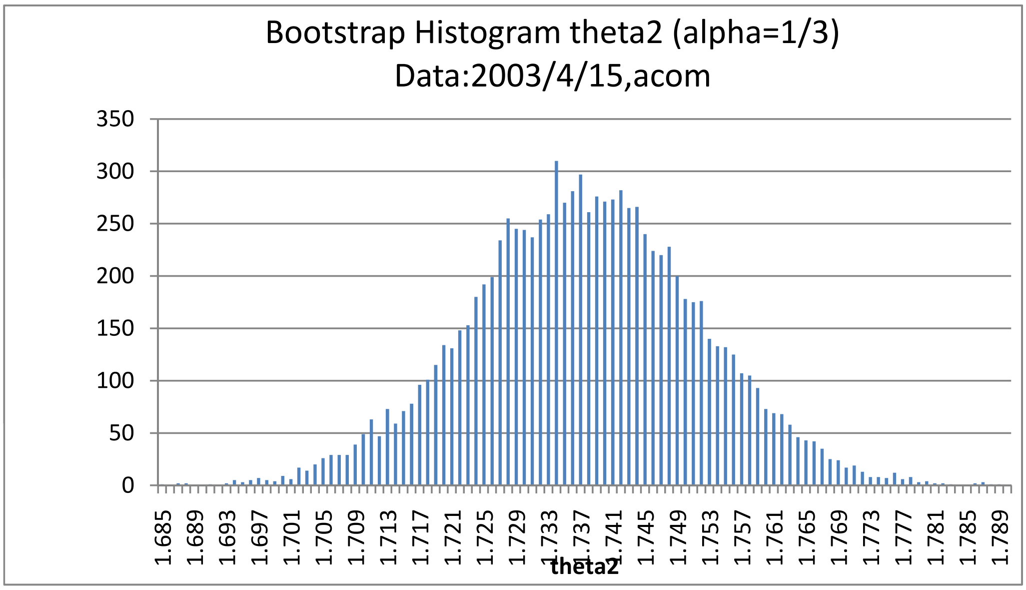


Picture 3 Exponeential Distribution

Ⅰ−iii Bootstrap Distribution of θ_1^ and θ_2^



Picture 4 Bootstrap Distribution of θ_1^



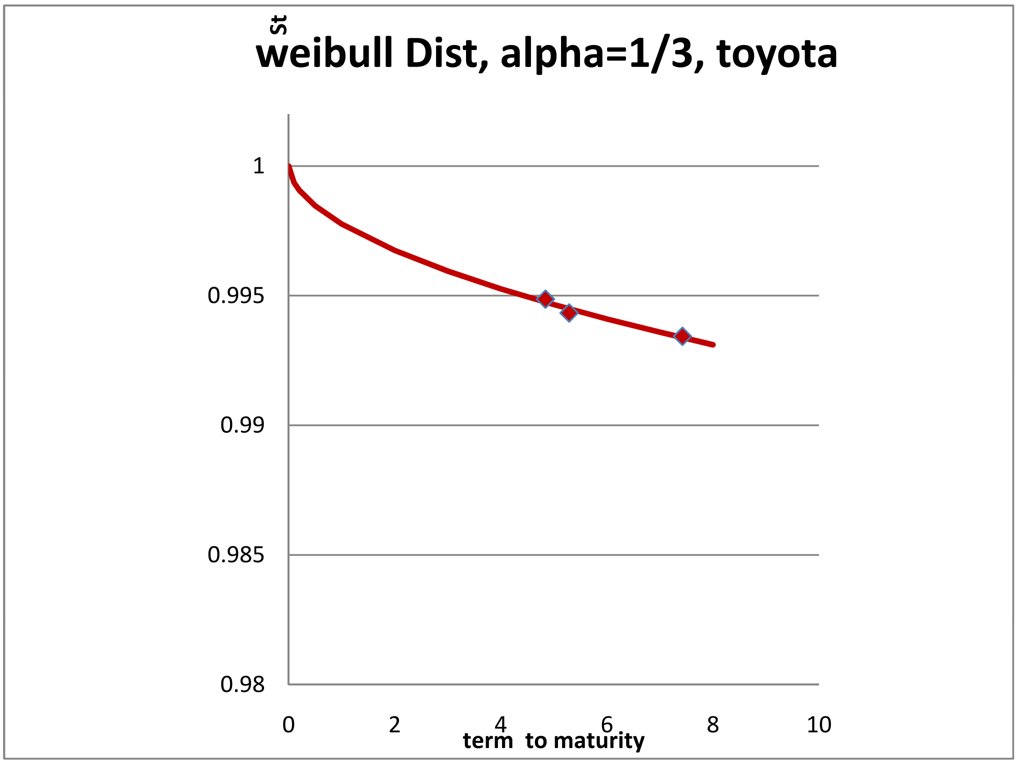
Picture 5 Bootstrap Distribution of θ_2^

	θ_1	θ_2
tilde(theta	0.002421	1.7369783
hat(theta)	0.002425	1.73641
Bias(hat(tl	-3.3E-06	0.0005683
Var(hat(th	3.07E-09	0.0002015
	5.54E-05	0.0141958

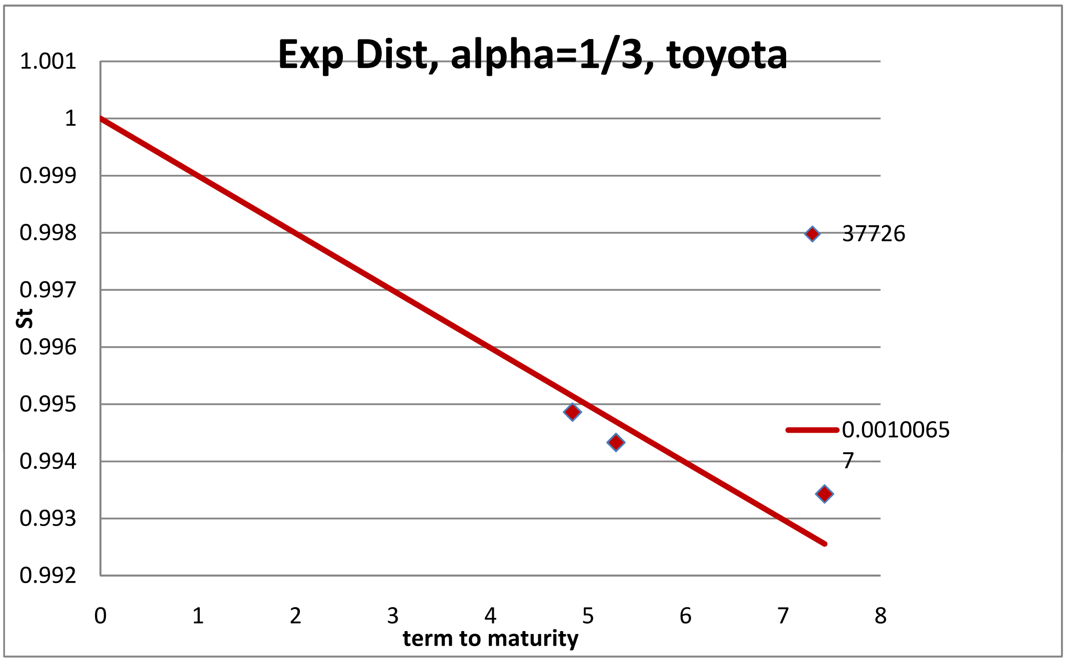
Table 1 Summery Statistics

Ⅱ　Toyota

Ⅱ－ⅰ　Survival Functions

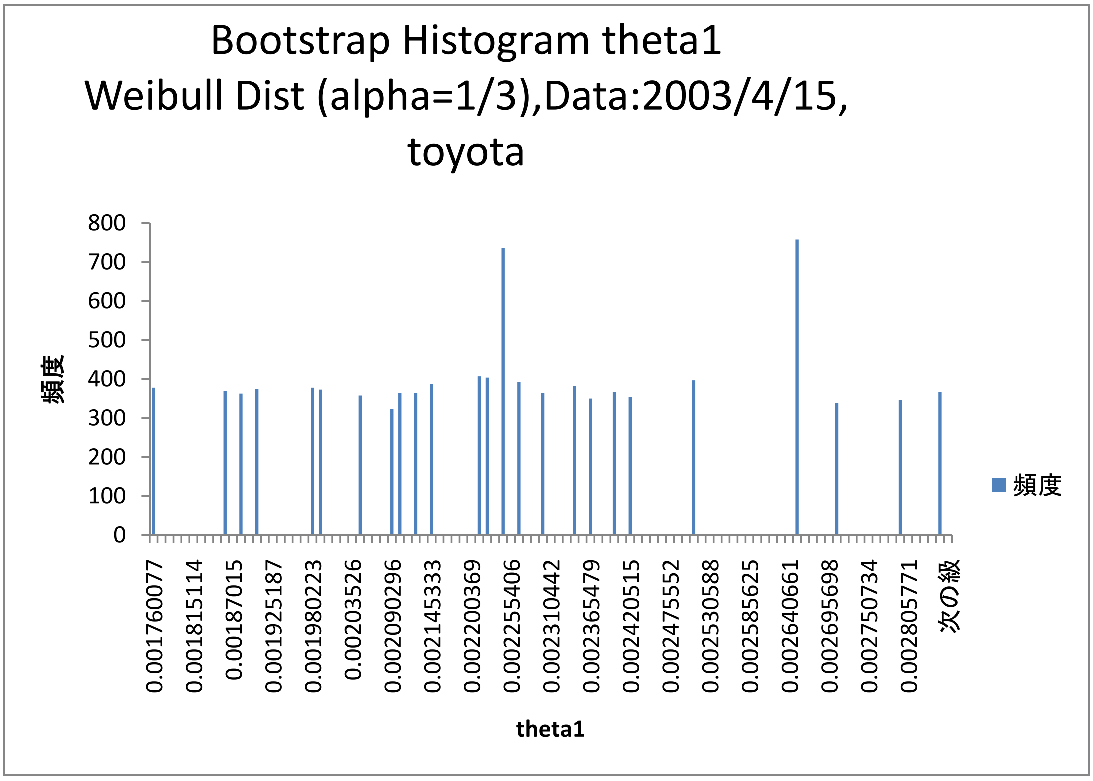


Picture 6 Weibull Distribution

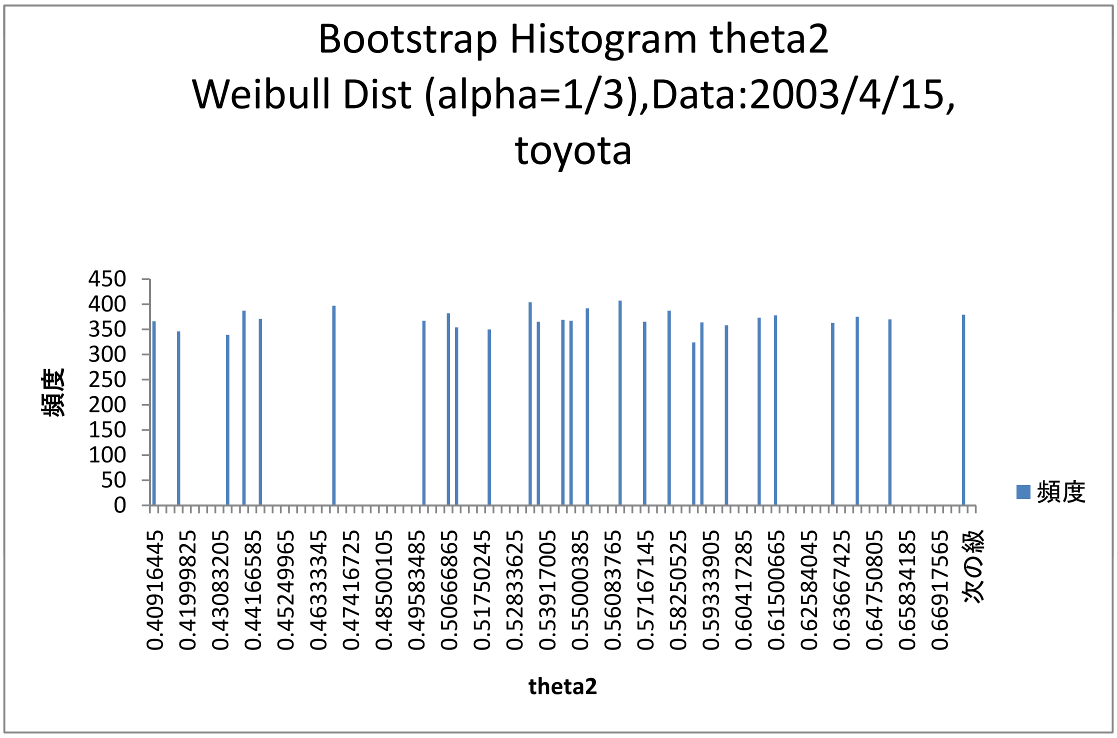


Picture 7 Exponential Distribution

Ⅱ－ⅱ　Bootstrap Distribution of $\theta\_1\hat{}$ and $\theta\_2\hat{}$



Picture 8 Bootstrap Distribution of　$\theta\_1\hat{}$



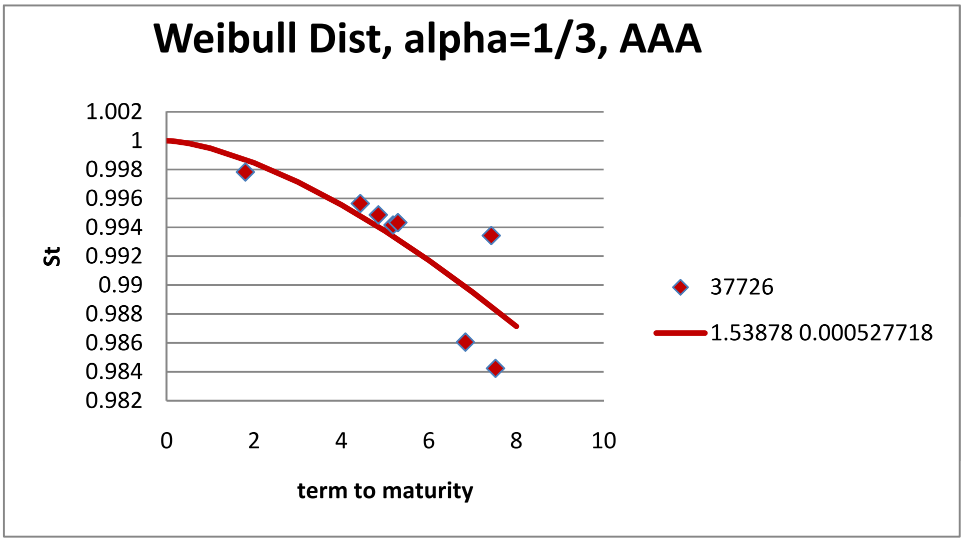
Picture9 Bootstrap Distribution of　$\theta\_2\hat{}$

	θ_1	θ_2
tilde(theta	0.002255	0.542792
hat(theta)	0.002237	0.542834
Bias(hat(th	1.75E-05	-4.20E-05
Var(hat(th	8.22E-08	0.005312
SD(hat(the	2.87E-04	7.29E-02

Table 3 Summery Statistics

Ⅲ AAA

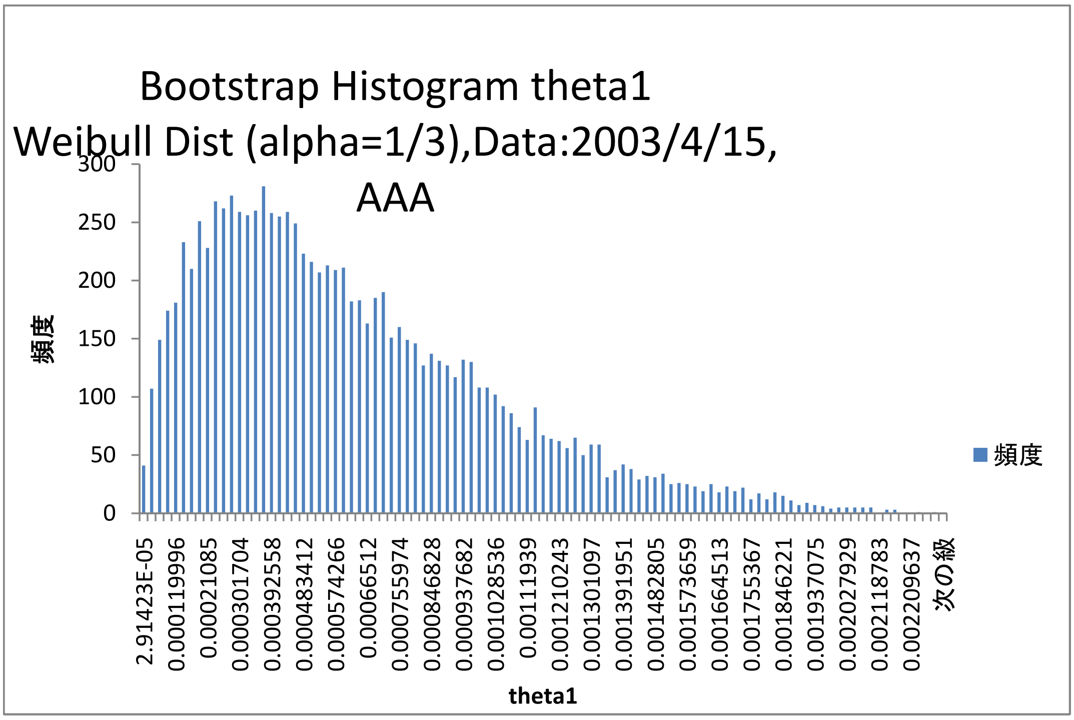
Ⅲ- ⅰ Survival Distribution



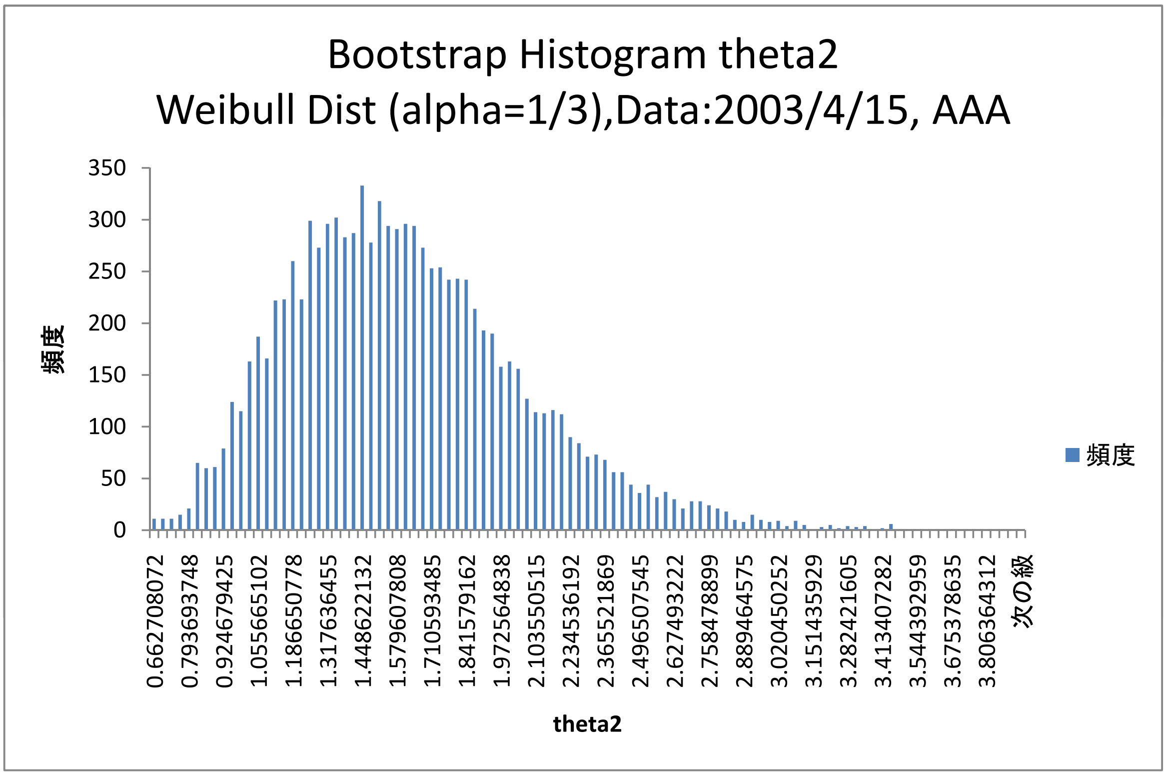
Picture 10 Weibull Distribution

	θ_1	θ_2
tilde(theta	0.000607	1.600027
hat(theta)	0.0005277	1.53878
Bias(hat(th	7.94E-05	0.061247
Var(hat(th	1.67E-07	0.199537
SD(hat(the	4.09E-04	4.47E-01

Table 4 Summery Statistics



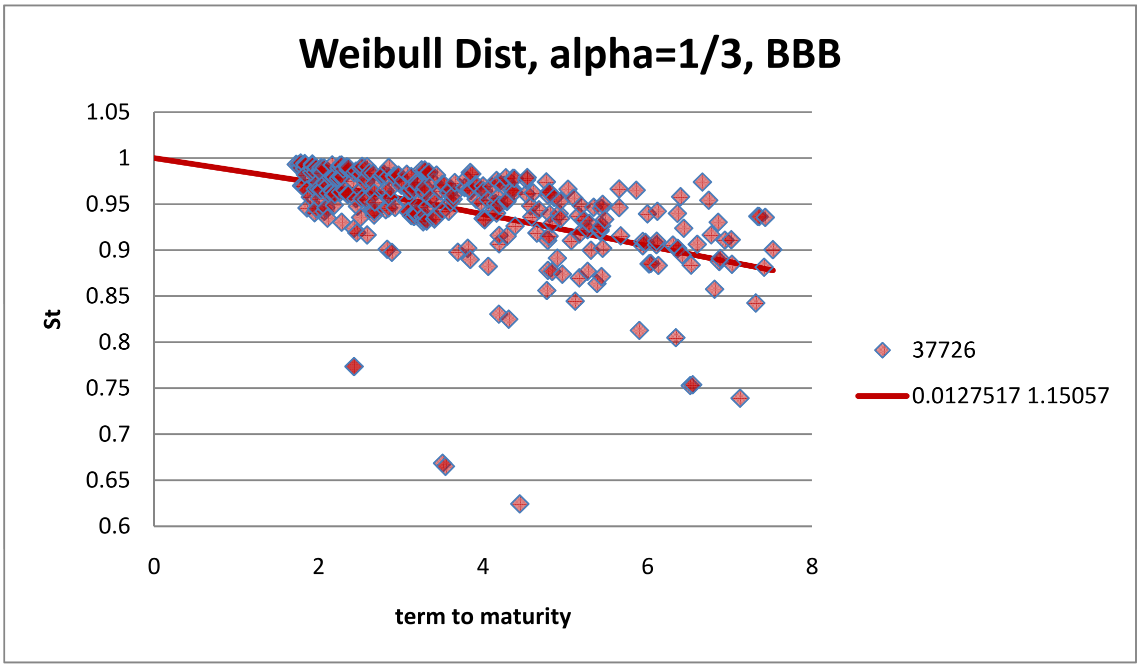
Picture 12 Bootstrap Distribution of θ _1^



Picture 13 Bootstrap Distribution of θ _2^

Ⅳ　BBB

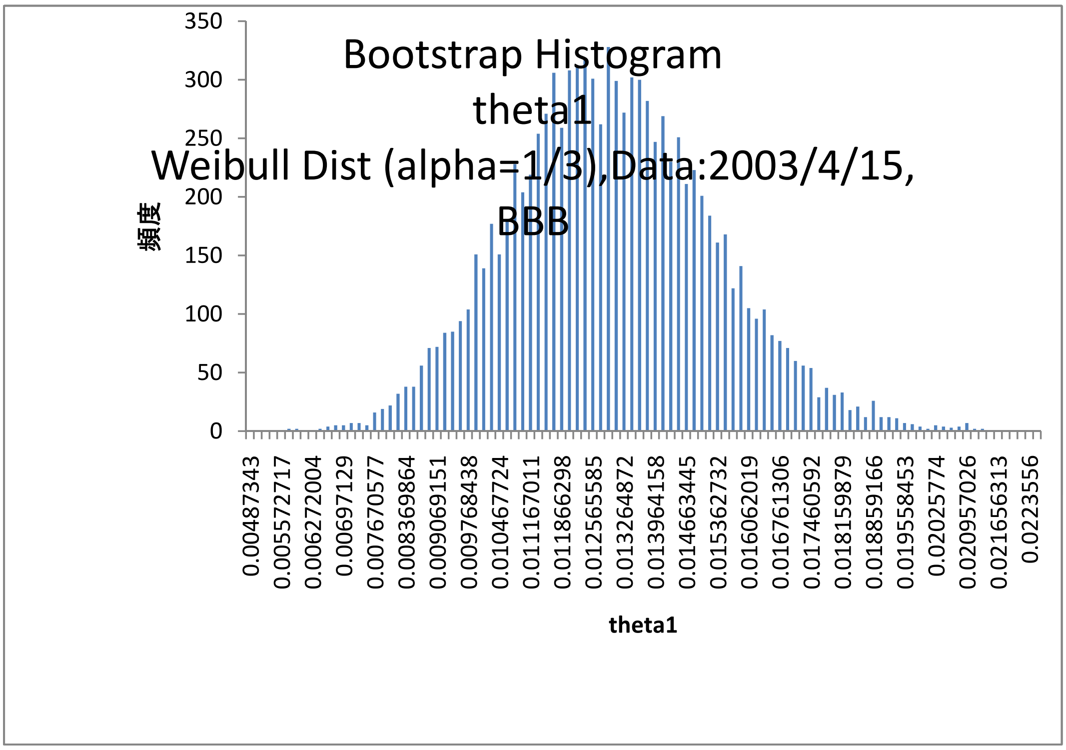
Ⅳ－ⅰ Survival Distribution



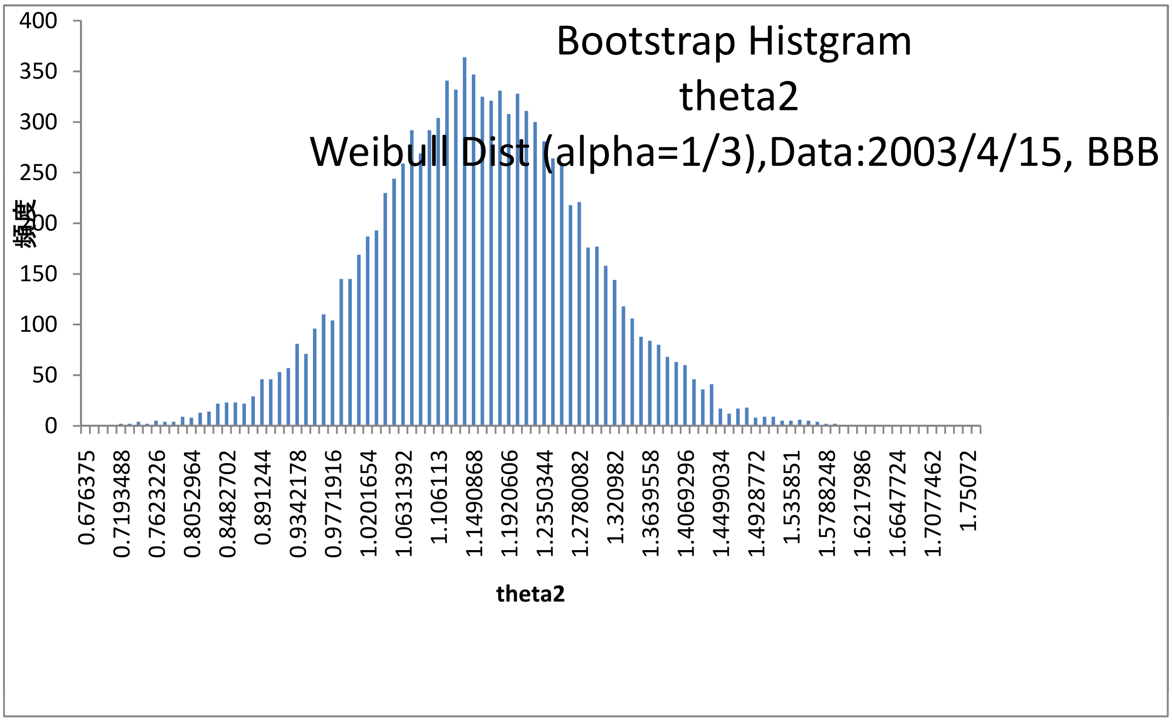
Ⅳ－ⅱ　Bootstrap Distribution

BBB	2003/4/15
Number of Repetion I	m=10000
	alpha=1/3
	weibull

		theta1	theta2
bootstrapEstimator	tilde(theta)'	0.012926	1.151379
MLE	hat(theta)	0.012752	1.15057
Bias	Bias(hat(th	0.000174	0.000809
Variance	Var(hat(the	5.27E-06	0.016635
StandardDiviation	SD(hat(the	0.002297	0.128982



Picture 14 Bootstrap Distribution of θ_1^



Picture 15 Bootstrap Distribution of θ_2^