Application of Bayesian Methods to Structural Models and Stochastic Frontier Production Models

by

Kwang-shin Choi

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved April 2014 by the
Graduate Supervisory Committee:

Dr. Seung Ahn, Chair
Dr. Rajnish Mehra
Dr. Sungho Park

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

This dissertation applies the Bayesian approach as a method to improve the estimation efficiency of existing econometric tools. The first chapter suggests the Continuous Choice Bayesian (CCB) estimator which combines the Bayesian approach with the Continuous Choice (CC) estimator suggested by Imai and Keane (2004). Using simulation study, I provide two important findings. First, the CC estimator clearly has better finite sample properties compared to a frequently used Discrete Choice (DC) estimator. Second, the CCB estimator has better estimation efficiency when data size is relatively small and it still retains the advantage of the CC estimator over the DC estimator. The second chapter estimates baseball's managerial efficiency using a stochastic frontier function with the Bayesian approach. When I apply a stochastic frontier model to baseball panel data, the difficult part is that dataset often has a small number of periods, which result in large estimation variance. To overcome this problem, I apply the Bayesian approach to a stochastic frontier analysis. I compare the confidence interval of efficiencies from the Bayesian estimator with the classical frequentist confidence interval. Simulation results show that when I use the Bayesian approach, I achieve smaller estimation variance while I do not lose any reliability in a point estimation. Then, I apply the Bayesian stochastic frontier analysis to answer some interesting questions in baseball.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

FINITE SAMPLE PROPERTIES OF STRUCTURAL ESTIMATORS

1.1 Introduction

Structural estimation methods have been popularly used for discrete choice models especially for the practical reason. The maximum likelihood (ML) estimator for a discrete choice model is easy to compute when structural errors are type-I-extremely distributed because the likelihood functions has a closed form ML methods can be also used to estimate continuous choice models by discretizing choice variables. We refer to the ML estimator from discretized data as "discretized choice (DC) estimator. Recently, Imai and Keane (2004) proposed an alternative estimator that does not require discretization. We refer to this alternative estimator as continuous choice (CC) estimator. The main motivation of this chapter is to compare the finite-sample performances of the DC and CC estimators.

The closed-form likelihood function for the DC estimator is obtained under the assumption that structural errors are additive to utility and drawn from a type-I extreme distribution. The likelihood function is misspecified if structural errors violate distributional assumption or if observed choice or state variables contain severe measurement errors. One important advantage of using the CC estimator instead of the DC estimator is that it does not assign the strong distributional assumption on structural errors. This is so because the CC estimator is computed with approximated value functions. In addition, the CC estimator can control for possible measurement errors in observed variables. The likelihood function for the CC estimator is determined by the distribution of measurement errors. As long as correct distribution function is used for measurement errors and value functions are properly approximated, the CC estimator is expected to have desirable asymptotic properties.

Despite its flexibility, the CC estimator has not been popularly used in literature. One possible reason is that it requires heavier use of computation and complex programming than the DC estimator. For practitioners, it should be an important question whether the gains by using the CC estimator sufficiently outweigh the computational costs. To answer this question, we examine and compare the finite-sample performances of the DC and CC estimators by Monte Carlo simulations. To our knowledge, no study has examined the finite-sample performances of the DC and CC estimators. This chapter is an attempt to fill this gap in the literatures.

Our Monte Carlo simulation exercises are designed to address three particular issues. The first issue is how the DC and CC estimators perform for the cases ideal for the DC estimator. To address this issue, we consider the case in which structural errors are drawn from a type-I extreme distribution and data are not contaminated by measurement errors. For this case, the CC estimator is computed with the data contaminated with artificially generated measurement errors. The second issue is how sensitive the performances of the DC and CC estimators are to distributional assumptions. For this issue, we consider the case in which structural errors drawn from a log normal distribution with data that do not contain measurement errors. The third and final issue we address is how the two estimators are sensitive to measurement errors. The likelihood function for the DC estimator is obtained from the conditional probability of the discretized choice variables. Thus, measurement errors in choice and state variables may influence the distribution of the DC estimator differently.

In addition to the CC estimator, we also consider the Bayesian estimation of structural parameters combining flat priors and the likelihood function used for the CC estimator. The Bayesian estimator we consider is the posterior mean of a structural parameter. We refer to this estimator as continuous choice Bayesian (CCB) estimator. Asymptotically, the CC and CCB estimators are expected to have the same distributions. However, they may perform differently in finite samples. In addition, for small samples, the confidence interval obtained from the Bayesian posterior distribution could be more accurate than the one approximated by the asymptotic distribution of the CC estimator. Thus, we compare the finite-sample performances of the two estimators in the end.

Our main results are the following. First, for the cases ideal for the DC estimator (type-I-extremely distributed structural errors which are independent and additive to utility and no measurement errors), none of the DC and CC estimators show serious biases. Second, the DC estimator has a bias when structural errors are not additive to utility or drawn from a log normal distribution. The true parameter value is located outside of the 95% confidence interval constructed from the distribution of the DC estimates from simulated data. Third, the CC estimator provides more reliable inferences than the DC estimator when data contain measurement errors. In particular, the DC estimator is sensitive to the measurement errors in choice variables and less sensitive to those in state variables. Fifth and finally, the distribution of the CCB estimator has a narrower bell shape than that of the CC estimator when sample size is small. However, as sample size increases, the two estimators have similar distributions.

This chapter is organized as follows. Section 1.2 provides a brief literature review. Section 1.3 introduces the model used for our simulations and explains how data are generated. Section 1.4 describes how the DC, CC and CCB estimators are computed. Section 1.5 reports simulation results. Concluding remarks follow in Section 1.6.

1.2 Literature Review

Structural estimation became a main subject of econometrics through a series of papers such as Wolpin (1984) on fertility and child mortality, Miller (1984) on job matching and occupational choice, Pake (1986) on patent renewal and Rust (1987) on engine replacement. Those early papers shared one common setting that agents in models have only discrete choices. The reason behind that assumption is that it requires less computational burden compared to continuous choice models. Because computing power was not enough to handle the discrete choice models in 1980s, the application of structural estimation approaches on the continuous choice models, which require even stronger computing power for estimations, should have been left for future research.

The early expansion of structural estimation was made toward the improvement of methods intended to deal with more complex state variables. State variables are information used by agents to find the optimal choice that maximizes the utility of agents. The state variables can be discrete or continuous even if the choice variables are discrete. When the model of interest had a continuous or large discrete state space, this model could not be handled with structural estimators proposed in the 1980s. That is because to estimate this kind of model, we need to solve for optimal choices at every possible combination of state space values. However, it is simply impossible to solve for optimal choices at every state variable if the state variable is in continuous space. Even if state variables have discrete values, it is practically impossible to find a maximized value at every state variable, if the number of possible state variable value is too large. Keane and Wolpin (1994) and Rust (1997) are notable papers which provide practical solutions to this difficulty.

Another direction for the expansion of structural estimation was to provide estimators applicable to the model with continuous choice variables. Compared with a discretized choice model which uses discrete value transformed from continuous value, using continuous choice variable is attractive in that it maximizes the volume of information from a data set because discretization discards a certain amount of information. Another motivation for using continuous choice is that discretization procedures require subjective judgement. The use of continuous choice variable reduces the concern that the subjective discretization procedure can affect estimation results. Imai and Keane (2004) provides an estimation method which uses continuous choice variables. It is the first paper using continuous choice variable without discretization in structural estimation and also adds one valuable answer to an important question in labor economics. Technically, the paper uses Maximum Simulated Likelihood (MSL) to provide the most likelihood value used for Maximum Likelihood Estimation (MLE). The most noticeable feature of the MSL approach used in Imai and Keane (2004) is that it heavily depends on the distribution of measurement error assumed to be included in data set. This approach could solve difficult question of obtaining the likelihood value without discretizing continuous choice variables in structural estimation. However, rigorous theoretical proof of reliability, consistency or unbiasedness of this approach is not included in Imai and Keane (2004).

Bayesian Markov Chain Monte Carlo approach has been one of the more popular estimation methods for discrete choice models. A Bayesian approach is appealing in that it provides a posterior distribution as a result of estimation, which is especially attractive for the following two reasons. First, it increases the reliability of estimation. Structural estimation often has complicated estimation procedure and the distribution of the parameter is not specified in most cases. If we use posterior distribution for point estimations, this kind of concern is reduced because we can approximate the shape of parameter distribution. Second, the posterior distribution is a useful ingredient for other applications. One example of application is using posterior distributions to choose the best model among competing ones using loss functions, as explained in Gelman et al (2003). It is more attractive for structural estimation because the selection criteria for choosing the best model is often ambiguous. Schorfheide (2000) provides an example which applies the Bayesian loss function approach to choose the best model for structural estimation. However, the model used for Schorfheide (2000) has only one agent and the application of methods on more general forms of model and data have not been tried due to computational difficulties. The Bayesian approach has never been popular as a choice for structural estimation despite meaningful advantages due to a computational burden. Rare examples of papers applying Bayesian approach include Lancaster (1997), which estimated very simple model. Ching et al (2009) and Ching et al (2012) introduced a new method called the IJC algorithm which reduces the burden of Bayesian computations for discrete choice models. By reducing time spent on value function iteration, this new method can extend the range of models which can be estimated by the Bayesian method.

1.3 Model and Data Generation

1.3.1 Model

This chapter uses three models to compare properties of three structural estimators, the DC, CC and CCB estimators. The first model is deterministic model. We start with this model as a benchmark to compare the CC and DC estimators. When we use the DC estimator, we assume that structural errors in the economic model follows independent and additive type-I extreme distribution. Emax and likelihood value computation of the DC estimator depends on this assumption. The model is deterministic because, the errors satisfy this assumption, they do not affect marginal utility of consumption. We expect the DC estimator provides reliable property when error follows this assumption. We also check the properties of the CC estimator which does not have specific limitation on the type of error distribution. In the second model, we check the properties of the estimators when the additive assumption is violated. Because errors still follow independent type-I extreme distribution, we can check how the violation of 'additive' assumption affects on estimation result. In the third model, we examine the property of estimators, when errors do not follow extreme distribution. Errors are not additive and follow log normal distribution. We provide the estimation results to check flexibility of each estimator. For convenience, we call three models as model 1, model 2 and model 3.

Model 1 has following utility function.

$$u(c_{it}) = \frac{c_{it}^{1+\alpha}}{1+\alpha} \tag{1.1}$$

where c_{it} is the consumption level of agent i at time t, α is the degree of risk aversion. The α is the only parameter to be estimated in this model. Each agent i lives 10 years and solves the following problem to maximize utility.

$$\max_{\{c_{it}\}_{t=1}^{t=10}} E\left[\sum_{t=1}^{t=10} \delta^t u(c_{it}) + \varepsilon_{it}\right] \tag{1.2}$$

$$s.t. \quad A_{it+1} = R(A_{it} + g_{it} - c_{it}) \tag{1.3}$$

$$A_{it} \geq 0 \qquad \forall i, \quad t \tag{1.4}$$

Here, A_{it} is the asset value at time t. We use 4% for the interest rate and 95% for discount rate. Thus, $R = 1.04$ and $\delta = 0.95$. The transfer income from government g_t is common to all individual agents and fixed at 500.

Define:

$$V_t(A_{it}) = \max_{\{c_{it}\}_{t=1}^{t=10}} E\left[\sum_{t=1}^{t=10} \delta^t u(c_{it})|A_{it}\right] \tag{1.5}$$

subject to the restrictions (1.3) and (1.4). In Bellman form, the value function $V_t(A_{it})$ is equivalent to

$$V_t(A_{it}) = \max_{c_{it}} [u(c_{it}) + \delta Emax(A_{i,t+1})] \tag{1.6}$$

where $A_{i,t+1} = R(A_{it} + g_t - c_{it})$ and $Emax(A_{it}) = EV_t(A_{it})$. The optimal consumption and asset paths $\{c_{it}\}_{t=1}^{t=9}$ and $\{A_{it}\}_{t=2}^{t=10}$ of agent i are determined by solving the problem (1.6) from $t = 1$ with given initial asset value A_{i1}. The consumption level at $t = 10$ simply equals $c_{i10} = A_{i10} + g_{10}$

Model 2 has following utility function and error assumption.

$$u(c_{it}) = \frac{c_{it}^{1+\alpha}}{1+\alpha} \tag{1.7}$$

$$s.t. \quad A_{it+1} = R(A_{it} + g_t(1 + \varepsilon_{it}) - c_{it})$$

Different from model 1, this model has structural error ε_{it} which is not additive to utility function and violate condition for using the DC estimator. We assume that probability density function (pdf) of random shock ε_{it} is $exp(-\frac{\varepsilon_{it}}{\beta})exp(-\frac{\varepsilon_{it}}{\beta})$ with β equal to 15. Error value affects marginal utility of consumption and future Emax value.

Model 3 shares the same utility function with Model 2 as follows:

$$u(c_{it}) = \frac{c_{it}^{1+\alpha}}{1+\alpha} \tag{1.8}$$

$$s.t. \quad A_{it+1} = R(A_{it} + g_t(1+\varepsilon_{it}) - c_{it})$$

However, model 3 is different from model 2 in that ε_{it} has following distribution.

$$\varepsilon_{it} \sim LN(-0.0017, 0.12^2) \tag{1.9}$$

We use model three to check the effect of the type of error distribution on properties of estimators.

1.3.2 Data Generation

While model 1 has deterministic data generation procedure where uncertainty of the model only comes from initial asset generation, data generation procedure for model 2 and model 3 is stochastic because their consumption decision depends on given random shock ε_{it}. In this section, we first explain how to generate data from model 1 and will specify the difference in model 2 and model 3 data generation steps.

Model 1

With $\alpha = -0.7$, we generate consumption and asset value paths for 100 agents who live for 10 time periods for one data set. 1,000 different data sets are generated for

Monte Carlo study. To begin with, let us assume that the function $Emax(A_{it})$ is known. Then, we can generate asset values and consumption levels for each agent i as follows. First, we draw initial asset values A_{i1} from $N(2000, 500^2)$. Second, we find the optimal consumption level at time $t = 1$, c_{i1}, by solving (1.6) with $Emax(A_{i2})$ replaced by $Emax(R(A_{i1} + g_2 - c_{i1}))$. The asset value at the second period is determined by $A_{i2} = R(A_{i1} + g_2 - c_{i1})$. Third, we repeat the second step for each $t \in \{2, 9\}$ to find optimal c_{it} and $A_{i,t+1}$.

To use the above procedure, one needs to know the $Emax(A_{it})$ functions for all t and all possible values of A_{it}. For deterministic model 1 we consider here, it may not be too difficult to find the $Emax$ functions. However, for model 2 and 3 we consider later, it is difficult to find the $Emax$ functions. For this reason, we generate data using the $Emax$ functions approximated by the interpolation method of Keane and Wolpin (1994). The approximation procedure is discussed in section 1.4.

Model 2 and Model 3

Similar to data generation for model 1, we start from generating initial asset value A_{i1} from following distribution.

$$A_{i1} \sim N(2000, 500^2)$$

Then, each agent i is given random transfer $g_t(1 + \varepsilon_{it})$ at the beginning of each period. ε_{it} follows type-I extreme distribution in model 2 and log normal distribution in model 3. Next step is to approximate $Emax$ values using the method which will be illustrated later in section 1.4.2.1. Then, we solve for following equation for $t = 1, ...9$.

$$V_t(A_{it}) = \max_{c_{it}} [u(c_{it}) + \delta Emax(A_{i,t+1})] \tag{1.10}$$

where $A_{i,t+1} = R(A_{it} + g_t(1 + \varepsilon_{it}) - c_{it})$. The optimal consumption and asset paths $\{c_{it}\}_{t=1}^{t=9}$ and $\{A_{it}\}_{t=2}^{t=10}$ of agent i are determined by solving equation (1.10). The last period consumption, c_{i10}, is simply $c_{i10} = A_{i10} + g_{10}(1 + \varepsilon_{i10})$.

We generate data for $i = 1, ..., 100$ by repeating above sequence $1 - 3$. To illustrate generated data values, table 1.1 provides generated data for agent 1 from three different model settings. Mean values of generated data from each model is also provided in table 1.2.

Here are several things to be mentioned about this dataset illustrated in table 1.1. First, asset has continuous space. It means that we need an approximation procedure. Second, consumption, the only choice variable in the model, has also continuous space. To deal with continuous choice, we need to choose between discretizing those values and applying a recent approach such as the method from Imai and Keane (2004). We apply both approaches in this chapter to check the validity of each method and to compare their reliability and efficiency. Third, the fourth column of table 1 is asset contaminated by measurement errors. We need asset values contaminated by measurement errors for two reasons. First, we use them to check the property of the CC estimator by Imai and Keane (2004) which requires existence of measurement error. Second, we investigate how they affect estimation results from the DC estimator, which uses same estimation steps regardless of existence of measurement errors. For such cases, the measurement errors ξ_{it} are drawn from $N(0, 400^2)$.

	Model 1			Model 2			Model 3		
Period	A_{1t}	c_{1t}	$A_{1t}+\xi_{1t}$	A_{1t}	c_{1t}	$A_{1t}+\xi_{1t}$	A_{1t}	c_{1t}	$A_{1t}+\xi_{1t}$
1	2153	805	1983	2697	913	2445	1632	704	1539
2	1922	790	1840	2342	871	2188	1520	678	1763
...	...	...	...	...	...	...	...	...	...
10	125	625	9	245	832	45	112	548	25

Table 1.1: Generated Data of Agent 1

Model 1			Model 2			Model 3		
A_{i1}	A_{i10}	c_{it}	A_{i1}	A_{i10}	c_{it}	A_{i1}	A_{i10}	c_{it}
2016.1	184.3	805.4	1989.5	150.1	755	2018.5	172.2	736.5

Table 1.2: Mean Values of Generated Data Set

1.4 Estimation Methods

1.4.1 Model 1

Continuous Choice Estimator

The continuous choice (CC) estimator is proposed by Imai and Keane (2004). To replace BHHH used for Imai and Keane (2004), we applied the Constrained Optimization By Linear Approximation (COBYLA) algorithm introduced by Powell (1994) as an numerical optimizer for the CC estimator. This algorithm does not require finding derivatives of log likelihood functions and is particularly useful for the estimation of structural models with complicated likelihood functions. McKinnon (1998) proved that COBYLA provides more robust optimization results compared to the widely used the Nelder-Mead method which is also a non-differential search method.

Step 1. Approximating $Emax$ values for a given $\hat{\alpha}$

To approximate values, we start with the last time period ($t = 10$) when the function has the following simple form:

$$Emax(A_{i10}) = \max_{c_{i10}} \left(\frac{c_{i10}^{1+\widehat{\alpha}}}{1+\widehat{\alpha}} \right) \tag{1.11}$$

$$s.t. \qquad c_{i10} \leq A_{i10} + g_{i10}$$

where $\hat{\alpha}$ is a candidate parameter value, and A_{i10} is the asset value at the beginning of the last period and the only state variable for agent i. g_{10} value is fixed at 500. We search for $Emax(A_{i10})$ values for $0 \leq A_{i10} \leq 4000$ at the last period. A_{i10}'s range is limited to the interval [0, 40000], because asset values in generated data do not exceed the range. We approximate the $Emax(A_{i10})$ function using the interpolation method of Keane and Wolpin (1994).

The approximated $Emax(A_{i10})$ function is obtained by the following procedure. First, the approximated functional form we use is given

$$\pi_{0,10} + \pi_{1,10} p\left(A_{i10}\right) + \pi_{2,10} (p\left(A_{i10}\right))^2 + \pi_{3,10} (p\left(A_{i10}\right))^3 \tag{1.12}$$

We estimate $\pi_{0,10}$, $\pi_{1,10}$, $\pi_{2,10}$ and $\pi_{3,10}$ by the least squares using a set of the $Emax(A_{i10})$ values on selected A_{i10} values. Here $p(A_{i10})$ is A_{i10}'s value after the Chebyshev polynomial transformation. We use the Chebyshev polynomial transformation, because if we use transformed values of A_{i10} selected on Chebyshev nodes, we can minimize the maximum value of errors occurred in approximation (Stewart, 1996). Table 1.3 provides selected data points used for the least squares estimation of equation (1.12). Using the estimated values, $\hat{\pi}_{0,10}$, $\hat{\pi}_{1,10}$, $\hat{\pi}_{2,10}$ and $\hat{\pi}_{3,10}$, we can

A_{10}	3999.8	3997.8	$\cdots$	2031.4	$\cdots$	0.247
$p(A_{10})$	0.999	0.998	$\cdots$	0.016	$\cdots$	-0.999
$Emax(A_{10})$	42.73	42.72	$\cdots$	36.14	$\cdots$	0.042

Table 1.3: Asset and Computed $Emax$ Values at Chebyshev Nodes at $t = 10$

approximate the $Emax(A_{i10})$ value for any state variable value A_{i10} by applying the following equation.

$$\widehat{Emax}(A_{i10}) = \hat{\pi}_{0,10} + \hat{\pi}_{1,10}p(A_{i10}) + \hat{\pi}_{2,10}(p(A_{i10}))^2 + \hat{\pi}_{3,10}(p(A_{i10}))^3 \tag{1.13}$$

At period $t = 9$, the procedure to approximate $Emax$ values is similar to the procedure for the last period except we need $Emax(A_{i10})$ to solve for the optimal value of c_{i9}. To approach this problem, we use equation (1.13) to approximate $Emax(A_{i10})$. With the approximated $Emax(A_{i10})$, $Emax(A_{i9})$ is given

$$\begin{aligned} Emax(A_{i9}) &= \max_{c_{i9}} \left[\frac{{c_{i9}}^{1+\alpha}}{1+\alpha} + V_{10}(A_{i10}) \right] \\ &= \max_{c_{i9}} \left[\frac{c_{i9}^{1+\alpha}}{1+\alpha} + Emax(A_{i10}) \right] \\ &\approx \max_{c_{i9}} \left[\frac{{c_{i9}}^{1+\alpha}}{1+\alpha} + \widehat{Emax}(R(A_{i9} + g_t - c_{i9})) \right] \end{aligned} \tag{1.14}$$

Solving the first order condition for problem (1.14), we can find the optimal value c_{i9} for a given state variable A_{i9}. Now, we can approximate $Emax(A_{i9})$ for selected values of A_{i9} by repeating the procedure used for $Emax(A_{i9})$. The function that we use is

$$\pi_{0,9} + \pi_{1,9}p\,(A_{i9}) + \pi_{2,9}p(A_{i9})^2 + \pi_{3,9}p(A_{i9})^3 \tag{1.15}$$

We again estimate the parameters $\pi_{0,9}$, $\pi_{1,9}$, $\pi_{2,9}$ and $\pi_{3,9}$ by the least squares approach using a set of the $Emax(A_{i9})$ values on selected values of A_{i9}. Then, the approximated $Emax(A_{i9})$ is given

$$\widehat{E\max}(A_{i9}) = \hat{\pi}_{0,9} + \hat{\pi}_{1,9}p(A_9) + \hat{\pi}_{2,9}(p(A_{i9}))^2 + \hat{\pi}_{3,9}(p(A_{i9}))^3 \tag{1.16}$$

t	$\hat{\pi}_0$	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$
2	187.58	18.796	-1.229	0.092
3	172.21	18.473	-1.265	0.098
4	155.97	18.069	-1.311	0.108
5	138.80	17.555	-1.369	0.119
6	120.61	16.880	-1.444	0.137
7	101.27	15.962	-1.541	0.166
8	80.60	14.642	-1.660	0.221
9	58.26	12.681	-1.709	0.371
10	34.55	9.489	-1.652	0.467

Table 1.4: Estimated π Values when $\hat{\alpha} = -0.7$

At A_{i8}, we will use equation (1.12) to find approximated $Emax(A_{i8})$ value and the repeat the similar approach to approximate the functions for other time periods. As an example, Table 1.4 provides the estimated π values for the candidate parameter $\hat{\alpha} = -0.7$. To check interpolation method's validity, we used R^2 as a standard. At each period, R^2 value is at least larger than 0.999 in this approximation setting.

Step 2. Finding likelihood

To find the log-likelihood for a given candidate parameter, we use the maximum simulated likelihood (MSL) procedure suggested by Imai and Keane (2004). A distinctive feature of the MSL procedure is that measurement errors in data take an important role. This feature is not often found in other likelihood methods. The estimation procedure begins with the assumption that data contain measurement errors. In fact, measurement errors are necessary for the MSL estimation because the likelihood function for the MSL estimator (CC estimator) is determined by the distribution of measurement errors. This feature leads us to an ironical situation where if data do not contain any measurement errors, we need to contaminate the data with artificially generated errors. We also check the effect from generated artificial error in result section.

We consider two cases, say, Cases A and B. Case A is the case in which data are contaminated by measurement errors, and Case B is the case in which data do not contain measurement errors. For both cases, we draw random numbers ξ_{it} from $N(0, \sigma_\xi^2)$ for each combination of i and t. Then, we add the measurement errors to true asset values A_{it} to obtain contaminated asset values A_{it}^D:

$$A_{it}^D = A_{it} + \xi_{it} \tag{1.17}$$

The estimation procedures are different for Cases A and Case B. For Case A, the ξ_{it} represent the measurement errors whose distribution is unknown, and thus σ_ξ^2, as well as α should be estimated. In contrast, for Case B, the errors ξ_{it} represent artificial measurement errors drawn from a known distribution ($N(0, \sigma_\xi^2)$ in our case) to construct the likelihood function. The variance of ξ_{it}, σ_ξ^2, is known and it needs not be estimated. Thus, for Case B, the only parameter to be estimated is α.

· **Steps for Case A**

1. Choose candidate parameter values of α and σ_ξ, say $\hat{\alpha}$ and $\hat{\sigma}_\xi$.

2. Fetch initial asset values (with measurement errors), A_t^D. Then, generate errors ξ_{it}^m from $N(0, \hat{\sigma}_\xi^2)$. Set $A_{i1}^m = A_{i1}^D - \xi_{i1}^m$ and consider it to be true initial asset value.

3. Starting with A_{i1}^m, find the optimal consumption level c_{i1}^m by solving

$$c_{i1}{}^m = \arg\max_{c_{i1}^m} \left[\frac{(c_{i1}{}^m)^{1+\alpha}}{1+\alpha} + \widehat{Emax}(A_{i2}{}^m) \right] \tag{1.18}$$

where $A_{i2}^m = R(A_{i1}^m + g_t - c_{i1}{}^m)$.

4. Apply $1-3$ for $t = 2, ..., 9$. Then, we can find the optimal consumption and asset value paths $\{c_{i2}{}^m, ..., c_{i9}{}^m\}$ and $\{A_{i3}{}^m, ..., A_{i10}{}^m\}$. The last period consumption is simply settled as $c_{i10}{}^m = A_{i10}{}^m + g_t$.

5. Repeat steps $1-4$ to generate simulated data for $m = 1, ..., 150$.

6. Estimate measurement errors by $\xi_{it}^m = A_{it}^D - A_{it}^m$.

7. Define

$$L_i^m(\hat{\alpha}, \hat{\sigma}_\xi) = \prod_{t=1}^{10} \left[\frac{1}{\hat{\sigma}_\xi \sqrt{2\pi}} \exp\left(-\frac{(\xi_{it}{}^m)^2}{2\hat{\sigma}_\xi^2} \right) \right] \tag{1.19}$$

8. Likelihood function value at the candidate parameter values $\hat{\alpha}$ and $\hat{\sigma}_\xi$ is given by

$$l(\hat{\alpha}, \hat{\sigma}_\xi) = \sum_{i=1}^{100} \left[ln \frac{1}{150} \sum_{m=1}^{150} L_i^m(\hat{\alpha}, \hat{\sigma}_\xi) \right] \tag{1.20}$$

· **Steps for Case B**

The steps needed for Case B are the same as those for Case A, except now that initial period measurement error ξ_{i1}^m is drawn from $N(0, \sigma_\xi^2)$. Because σ_ξ is known, it needs not be estimated.

Step 3. Computing the CC estimator. We use optimal candidate parameter which provides maximum likelihood by applying COBYLA.

Orig. value	d	Disc. value c_{it}^d
0∼650	1	600
650∼750	2	700
750∼850	3	800
850∼950	4	900
Over 950	5	1000

Table 1.5: Discretization Rule of Continuous Consumption Value

Step 4. Repeating steps $1-3$, 1,000 times. Find average and standard error of estimated values.

Discretized Choice Estimator

Rust (1987) proposed a maximum likelihood estimator for discrete choice models. We refer to the maximum likelihood estimator applied to discretized choice variables as the discretized choice (DC) estimator. For our experimental model (1), the DC estimator computation follows the steps below.

Step 1. Discretization

Discretize consumption levels. This procedure follows the rule given in table 1.5. As shown in table 1.2, the average of consumption levels is about 800 and most of consumption values are in the interval (550, 1050). The discretization rule we use is similar to those which have been used for the studies of labor hours such as Houser (2003), Rust and Phelan (2007) and Van der Klaauw and Wolpin (2008).

Step 2. Approximating Emax values.

We start from $t = 10$. For $t = 10$, the procedure to find the $Emax(A_{i10})$ values is same with the CC estimator. After finding the $Emax(A_{it})$ values on selected asset value A_{i10}, we apply the interpolation method used for the CC estimator to obtain $\widehat{Emax}(A_{i10})$

For the time periods $t = 9, 8, ..., 2$, we compute the $Emax(A_{it})$ values following the method of Rust (1987). For time t, $Emax(A_{it})$ for a given state variable valueA_{it} is

$$Emax(A_{it}) = \ln\left[\sum_{d=1}^{5} \exp(\widehat{V}_t(A_{it}, c_{it}^d))\right] \tag{1.21}$$

where

$$\begin{aligned} c_{it}^d &= 600 + 100(d-1) \\ \widehat{V}_t(A_{it}, c_{it}^d) &= \frac{(c_{it}^d)^{1+\alpha}}{1+\alpha} + \widehat{E\max}(A_{it+1}) \\ A_{it+1} &= R(A_{it} + g_t - c_{it}) \end{aligned}$$

For $t = 9$, we use $\widehat{E\max}(A_{i10})$ to find $\widehat{V}(A_{it}, c_{it}^d)$ at selected A_{it}. We can easily find $Emax(A_{it})$ when we have $\widehat{V}(A_{it}, c_{it}^d)$ values using equation (1.21) which comes from type-I extreme structural error distribution. Now we use the interpolation method to find $\widehat{Emax}(A_{i9})$ in the same way that we used for the CC estimator. To find $\widehat{Emax}(A_{it})$ for $t = 2, ...8$, we repeat the approach we used for $t = 9$ to find $\widehat{Emax}(A_{i9})$

Step 3. Computing log-likelihood.

To obtain the log-likelihood for a given parameter, we start from finding $Prob(c_{it}^D|A_{it})$, conditional probability of choosing consumption, c_{it}^D, given continuous asset value A_{it}. We use equation (1.22) to find the probability and this equation comes from multiple logit conditional probability which depends on structural error assumption of type-I extreme distribution (Train (2003)).

$$Prob(c_{it}^D|A_{it}) = \frac{exp(V_t(A_{it}, c_{it}^D))}{\sum_{d=1}^{5} exp(V_t(A_{it}, c_{it}^d))} \tag{1.22}$$

where D is 1, 2, 3, 4 or 5.

Combine ${c_{it}}^D$, discretized consumption value in dataset, and A_{it}, asset value from dataset, we obtain equation (1.23) that we use for computing the likelihood value for candidate $\hat{\alpha}$

$$L(\widehat{\alpha}) = \prod_{i=1}^{100}\prod_{t=1}^{10} Prob({c_{it}}^D|A_{it}) \tag{1.23}$$

Then, log-likelihood value is as follows:

$$l(\widehat{\alpha}) = \sum_{i=1}^{100}\sum_{t=1}^{10} \log Prob({c_{it}}^D|A_{it}) \tag{1.24}$$

Step 4. Find the candidate parameter which provides maximum likelihood value using numerical optimization methods. Like the CC estimator, the DC estimator uses COBYLA for numerical search.

Step 5. Repeating steps $1 - 4$, 1,000 times. Find average and standard error of estimated values.

Continuous Choice Bayesian Estimator

The Continuous Choice Bayesian (CCB) estimator uses the same likelihood finding step that we used for the CC estimator but finds optimal candidate value by applying the Metropolis algorithm which is the one of Bayesian MCMC convergence method. As an output of Metropolis algorithm, it provides the posterior distribution. This procedure can be summarized into following two steps.

Step 1. Finding likelihood value for a given candidate $\hat{\alpha}$.

Step 1 is same with likelihood finding steps used for the CC estimator. That is, this step is the combination of step 1 and step 2 of the CC estimator.

Step 2. Updating candidate parameter value and making posterior distribution

We use Metropolis algorithm to update candidate parameter. Steps for Metropolis algorithm has different detail according to the existence of the measurement error. Therefore we will separately describe update procedure for case A and case B. The setting of case A and case B is same with the CC estimator case.

· **Steps for Case A**

(1) Set initial candidate parameter α_0 and $\sigma_{\xi,0}$. We use $\alpha_0 = -0.9$ and $\sigma_{\xi,0} = 150$ for the initial values. We set these values far enough apart to show that Metropolis algorithm is reliable even if we do not have the proper knowledge on initial values. Similar approach is found in Ching et al (2009) where a initial value is set apart from the true parameter value. Using step 1, find $l(\alpha_0, \sigma_{\xi,0})$, log-likelihood of initial parameters α_0 and $\sigma_{\xi,0}$.

(2) Generate new parameter candidate $\hat{\alpha}_1$ from following distributional assumption.

$$\hat{\alpha}_1 \sim N(\alpha_0, {\sigma_\alpha}^2) \tag{1.25}$$

where σ_α is standard deviation for generation. We also define jumping distribution, $J_t(\hat{\alpha}_{t+1}|\alpha_t)$ as the probability density function value of $\hat{\alpha}_{t+1}$ when $\hat{\alpha}_{t+1}$ has the following distribution.

$$\hat{\alpha}_{t+1} \sim N(\alpha_t, {\sigma_\alpha}^2) \tag{1.26}$$

(3) Find $l(\hat{\alpha}_1, \sigma_{\xi,0})$, log-likelihood for $\hat{\alpha}_1$ and $\sigma_{\xi,0}$. Solve for update probability from α_0 to candidate $\hat{\alpha}_1$ as follows:

$$Prob(\alpha_0, \hat{\alpha}_1|\sigma_{\xi,0}) = min\left[\exp(l(\hat{\alpha}_1, \sigma_{\xi,0}) - l(\alpha_0, \sigma_{\xi,0})), 1\right] \tag{1.27}$$

Equation (1.27) comes from following update probability definition in Gelman et al (2003).

$$Prob(\alpha_0, \hat{\alpha}_1|\sigma_{\xi,0}) = min\left[r, 1\right] \tag{1.28}$$

In equation (1.28), r is jumping rule which is defined as follows:

$$r = \frac{p(\hat{\alpha}_1|\sigma_{\xi,0}, \mathbf{A}, \mathbf{c})}{p(\alpha_0|\sigma_{\xi,0}, \mathbf{A}, \mathbf{c})} \tag{1.29}$$

where $\mathbf{A}$ and $\mathbf{c}$ are the vectors of asset and consumption data generated from the true parameter. Ratio of posterior distribution, $\frac{p(\hat{\alpha}_1|\sigma_{\xi,0}, \mathbf{A}, \mathbf{c})}{p(\alpha_0|\sigma_{\xi,0}, \mathbf{A}, \mathbf{c})}$ is same with $\frac{L(\hat{\alpha}_1, \sigma_{\xi,0})}{L(\alpha_0, \sigma_{\xi,0})}$, the ratio of likelihood ($\hat{\alpha}_1$,$\sigma_{\xi,0}$) over (α_0,$\sigma_{\xi,0}$), when we use flat prior for α. Therefore, we have following equation

$$r = \frac{L(\hat{\alpha}_1, \sigma_{\xi,0})}{L(\alpha_0, \sigma_{\xi,0})} = \exp\left(l(\hat{\alpha}_1, \sigma_{\xi,0}) - l(\alpha_0, \sigma_{\xi,0})\right) \tag{1.30}$$

where $L(a, b)$ is likelihood of candidate parameter a and b.

Then, we generate a number λ_α from continuous uniform distribution between 0 and 1 and compare λ_α with $Prob(\alpha_0, \hat{\alpha}_1)$. Using λ_α, we update α_1, accepted parameter value at 1^{st} iteration for α, using following rule.

$$\begin{cases} Prob(\alpha_0, \hat{\alpha}_1|\sigma_{\xi,0}) & \geq & \lambda_\alpha & \rightarrow & Update & : & \alpha_1 = \hat{\alpha}_1 \\ Prob(\alpha_0, \hat{\alpha}_1|\sigma_{\xi,0}) & < & \lambda_\alpha & \rightarrow & Stay & : & \alpha_1 = \alpha_0 \end{cases}$$

(4) Update parameter $\sigma_{\xi,0}$ in similar manner used for (2) and (3). Using appropriate σ_{σ_ξ}, generate candidate parameter value $\hat{\sigma}_{\xi,1}$ from

$$\hat{\sigma}_{\xi,1} \sim N(\sigma_{\xi,0}, {\sigma_{\sigma_\xi}}^2)$$

(5) Find log-likelihood for $\hat{\sigma}_{\xi,1}$ and $\sigma_{\xi,0}$ with given α using Step 1. For α value, we use updated candidate α_1 from (3). Then we solve for update probability to set $\sigma_{\xi,1} = \hat{\sigma}_{\xi,1}$ given α_1 using equation (1.31) which comes from the assumption that we use flat prior for σ_ξ and jumping distribution of σ_ξ is symmetric. To obtain equation (1.31), we apply the same logic that we used for finding equation (1.27).

$$Prob(\sigma_{\xi,0}, \hat{\sigma_{\xi,1}}|\alpha_1) = [\exp(l(\hat{\sigma_{\xi,1}}, \alpha_1) - l(\sigma_{\xi,0}, \alpha_1)), 1] \tag{1.31}$$

Then we generate a number λ_{σ_ξ} from continuous uniform distribution between 0 and 1 to compare λ_{σ_ξ} with $Prob(\alpha_0, \hat{\alpha}_1)$. Using λ_{σ_ξ}, we update $\sigma_{\xi,1}$. parameter value at 1^{st} iteration for σ_ξ, as follows.

$$\begin{cases} Prob(\sigma_{\xi,0}, \hat{\sigma}_{\xi,1}|\alpha_1) & \geq & \lambda_{\sigma_\xi} & \rightarrow & Update & : & \sigma_{\xi,1} = \hat{\sigma}_{\xi,1} \\ Prob(\sigma_{\xi,0}, \hat{\sigma}_{\xi,1}|\alpha_1) & < & \lambda_{\sigma_\xi} & \rightarrow & Update & : & \sigma_{\xi,1} = \sigma_{\xi,0} \end{cases}$$

(6) Update accepted parameter set $\{\alpha_r, \sigma_{\xi,r}\}$ $(r \geq 1)$ in similar manner that we did in (2)$\sim$(5). The number of iterations is set to be sufficient for convergence of parameter.

1. Generate $\hat{\alpha}_{r+1}$ from α_r using $\hat{\alpha}_{r+1} \sim N(\alpha_r, {\sigma_\alpha}^2)$

2. Find log-likelihood for $\{\hat{\alpha}_{r+1}, \sigma_{\xi,r}\}$ and probability $Prob(\alpha_r, \hat{\alpha}_{r+1}|\sigma_{\xi,r}) = \exp(l(\hat{\alpha}_{r+1}, \sigma_{\xi,r}) - l(\alpha_r, \sigma_{\xi,r}))$

3. Generate a number λ_α from continuous uniform distribution between 0 and 1 then decide the value α_{r+1} using following rule.

$$\begin{cases} Prob(\alpha_r, \hat{\alpha}_{r+1}|\sigma_{\xi,r}) & \geq & \lambda_\alpha & \rightarrow & Update & : & \alpha_{r+1} = \hat{\alpha}_{r+1} \\ Prob(\alpha_r, \hat{\alpha}_{r+1}|\sigma_{\xi,r}) & < & \lambda_\alpha & \rightarrow & Stay & : & \alpha_{r+1} = \alpha_r \end{cases}$$

4. Generate $\hat{\sigma}_{\xi,r+1}$ from $\sigma_{\xi,r}$ using $\sigma_{\xi,r+1} \sim N(\sigma_{\xi,r}, {\sigma_\xi}^2)$

5. Compute log-likelihood for $\{\alpha_{r+1}, \hat{\sigma}_{\xi,r+1}\}$ and $\{\alpha_{r+1}, \sigma_{\xi,r}\}$.

   Then, find $Prob(\sigma_{\xi,r}, \hat{\sigma}_{\xi,r+1}|\alpha_{r+1}) = \exp(l(\hat{\sigma}_{\xi,r+1}, \alpha_{r+1}) - l(\sigma_{\xi,r}, \alpha_{r+1}))$

6. Generate a number λ_{σ_ξ} from continuous uniform distribution between 0 and 1 then decide the value $\sigma_{\xi,r+1}$ using following rule.

$$\begin{cases} Prob(\sigma_{\xi,r}, \hat{\sigma}_{\xi,r+1}|\alpha_{r+1}) \geq \lambda_{\sigma_\xi} \rightarrow Update : \sigma_{\xi,r+1} = \hat{\sigma}_{\xi,r+1} \\ Prob(\sigma_{\xi,r}, \hat{\sigma}_{\xi,r+1}|\alpha_{r+1}) < \lambda_{\sigma_\xi} \rightarrow Update : \sigma_{\xi,r+1} = \sigma_{\xi,r} \end{cases}$$

(7) Find mean and standard deviation from posterior distribution which is set of values $\{\alpha_r, \sigma_{\xi,r}\}_{r=b+1}^{B}$ where b is the number of initial burning and B is the number of iterations.

· **Steps for Case B**

(1) Set initial candidate parameter α_0. Initial value is set as -0.9. Using step 1, approximate Emax values and find log-likelihood value for α_0.

(2) Generate new candidate parameter $\hat{\alpha}_1$ from following distributional assumption.

$$\hat{\alpha}_1 \sim N(\alpha_0, {\sigma_\alpha}^2)$$

(3) Find log-likelihood for $\hat{\alpha}_1$ using Step 1 and Step 2. Then we solve for probability to update α_0 to candidate $\hat{\alpha}_1$ as follows

$$Prob(\alpha_0, \hat{\alpha}_1) = [\exp(l(\widehat{\alpha}_1) - l(\alpha_0)), 1] \tag{1.32}$$

Equation (1.32) has simple form which has only likelihood values because we assume the flat prior and the symmetric jumping distribution as we did for Case A. Then, we generate a number λ from continuous uniform distribution between 0 and 1 to compare λ with $Prob(\alpha_0, \hat{\alpha}_1)$. Using λ, we update parameter value at 1^{st} iteration, α_1, using following rule.

$$\begin{cases} Prob(\alpha_0, \hat{\alpha}_1) & \geq & \lambda & \rightarrow & Update & : & \alpha_1 = \hat{\alpha}_1 \\ Prob(\alpha_0, \hat{\alpha}_1) & < & \lambda & \rightarrow & Stay & : & \alpha_1 = \alpha_0 \end{cases}$$

(4) Update parameter α_r $(r \geq 1)$ in similar manner used for (2) and (3). The number of iterations is set to be sufficient for convergence of parameter.

1. Generate $\hat{\alpha}_{r+1}$ from α_r using $\hat{\alpha}_{r+1} \sim N(\alpha_r, {\sigma_\alpha}^2)$

2. Find log-likelihood for $\hat{\alpha}_{r+1}$

   and probability $Prob(\alpha_r, \hat{\alpha}_{r+1}) = [\exp(l(\hat{\alpha}_{r+1}) - l(\alpha_r)), 1]$

3. Generate a number λ from continuous uniform distribution between 0 and 1 then decide the value α_{r+1} using following rule.

$$\begin{cases} Prob(\alpha_r, \hat{\alpha}_{r+1}) & \geq & \lambda & \rightarrow & Update & : & \alpha_{r+1} = \hat{\alpha}_{r+1} \\ Prob(\alpha_r, \hat{\alpha}_{r+1}) & < & \lambda & \rightarrow & Stay & : & \alpha_{r+1} = \alpha_r \end{cases}$$

(5) Find the mean and standard deviation of posterior distribution which is set of values $\{\alpha_r\}_{r=b+1}^{B}$ where b is the number of initial burning and B is the number of iterations.

1.4.2 Model 2 and Model 3

Continuous Choice Estimator

The CC estimator consists of following steps. The order of steps is same with the CC estimator for the estimation of model 1 but the detail of execution is different. The estimation of model 2 and model 3 applies the same steps except that the model 2 procedure includes an Emax approximation in step 1 as follows.

Step 1. Approximating $Emax$ for a given parameter

· **Model 2**

At last period $t = 10$, we solve for approximated $Emax$ value in following orders.

1. Generate $\varepsilon_{10,j}$ which follows type-I distribution with pdf, $exp(-\frac{\varepsilon_{it}}{\beta})exp(-\frac{\varepsilon_{it}}{\beta})$ with β equal to 15. We generate 30 errors for result in this chapter.

2. Find value of $V_{10}(A_{i10}, \varepsilon_{i10,j})$ for a given error value $\varepsilon_{10,j}$ as follows

$$V_{10}(A_{i10}, \varepsilon_{i10,j}) = \frac{[A_{i10} + g_{10}(1 + \varepsilon_{i10,j})]^{1+\hat{\alpha}}}{1 + \hat{\alpha}} \tag{1.33}$$

where $\hat{\alpha}$ is given parameter value. We use transformed Chebyshev nodes that we used for the model 1 case. Using $V_{10}(A_{10}, \varepsilon_{10,j})$ from equation (1.33), we find $Emax$ values as follows:

$$E\max(A_{i10}) = \sum_{j=1}^{30} \frac{V_{10}(A_{i10}, \varepsilon_{i10,j})}{30} \tag{1.34}$$

3. Set functional form to approximate $Emax(A_{i10})$ as follows.

$$\pi_{0,10} + \pi_{1,10}p\left(A_{i10}\right) + \pi_{2,10}(p\left(A_{i10}\right))^2 + \pi_{3,10}(p\left(A_{i10}\right))^3 \tag{1.35}$$

then, use the least squares method that we used for extreme error case to find the following equation.

$$\widehat{Emax}(A_{i10}) = \hat{\pi}_{0,10} + \hat{\pi}_{1,10}p(A_{i10}) + \hat{\pi}_{2,10}(p(A_{i10}))^2 + \hat{\pi}_{3,10}(p(A_{i10}))^3 \tag{1.36}$$

$p(A_{i10})$ is value after Chebyshev polynomial transformation. Now, we have $\widehat{Emax}(A_{i10})$ for all the asset values, A_{i10}.

To find $\widehat{Emax}(A_{it})$ for period $t = 9, 8, ..., 2$, we use following sequences.

(a) Prepare a set of measurement errors, $\{\varepsilon_{it,j}\}$ for $j = 1, ..., 30$ from type-I extreme distribution.

(b) Find value of $V_t(A_{it}, \varepsilon_{it,j})$ for a given error value $\varepsilon_{it,j}$ as follows.

$$V_t(A_{it}, \varepsilon_{it,j}) = \max_{c_{it}} \left[u(c_{it}) + \widehat{Emax}(A_{it+1}) \right]$$

$$= \max_{c_{it}} \left[u(c_{it}) + \widehat{Emax}(R(A_{it} + g_t(1 + \varepsilon_{it,j}) - c_{it})) \right] \quad (1.37)$$

Then, find Emax value at A_{it} as follows:

$$E\max(A_{it}) = \sum_{j=1}^{30} \frac{V_t(A_{it}, \varepsilon_{it,j})}{30} \quad (1.38)$$

(c) To find $\widehat{Emax}(A_{it})$, set functional form for $Emax(A_{it})$ as follows:

$$\pi_{0,t} + \pi_{1,t} p\left(A_{it}\right) + \pi_{2,t} (p\left(A_{it}\right))^2 + \pi_{3,t} (p\left(A_{it}\right))^3 \quad (1.39)$$

Then, estimate $\hat{\pi}_{0,t}$, $\hat{\pi}_{1,t}$, $\hat{\pi}_{2,t}$ and $\hat{\pi}_{3,t}$ for following equation.

$$\widehat{Emax}(A_{it}) = \hat{\pi}_{0,t} + \hat{\pi}_{1,t} p(A_{it}) + \hat{\pi}_{2,t} (p(A_{it}))^2 + \hat{\pi}_{3,t} (p(A_{it}))^3 \quad (1.40)$$

(d) Repeat above sequence in $(a) - (c)$ for $t = 8, ..., 2$.

· **Model 3**

Same with previous models, we start from approximating $Emax$ value in terminal period.

1. Find error values $\varepsilon_{10,l}$ which comes from Gauss-Hermite quadrature points $\{x_l\}_{l=1}^{20}$.

$$\varepsilon_{10,l} = \exp(\sqrt{2} x_l \sigma + \mu) \quad l = 1, ..., 20 \quad (1.41)$$

where $\mu = -0.0017$ and $\sigma = 0.12$ which are log normal distribution parameters used for structural error generation.

2. Find value of $V_{10}(A_{i10}, \varepsilon_{i10,l})$ for a given error value $\varepsilon_{10,l}$ as follows

$$V_{10}(A_{i10}, \varepsilon_{i10,l}) = \frac{[A_{i10} + g_{10}(1 + \varepsilon_{i10,l})]^{1+\hat{\alpha}}}{1 + \hat{\alpha}} \tag{1.42}$$

where $\hat{\alpha}$ is given parameter value. We also use Chebyshev nodes here. We use $V_{10}(A_{10}, \varepsilon_{10,l})$ from equation (1.42) to find $Emax$ values as follows:

$$E\max(A_{i10}) = \sum_{l=1}^{20} \frac{V_{10}(A_{i10}, \varepsilon_{i10,l})W(l)}{\sqrt{\pi}} \tag{1.43}$$

where W(l) is weight value for Gauss-Hermite quadrature points.

3. Set functional form to approximate $Emax(A_{i10})$ as follows.

$$\pi_{0,10} + \pi_{1,10}p\left(A_{i10}\right) + \pi_{2,10}(p\left(A_{i10}\right))^2 + \pi_{3,10}(p\left(A_{i10}\right))^3 \tag{1.44}$$

Estimate π's in equation (1.44) to find following equation.

$$\widehat{Emax}(A_{i10}) = \hat{\pi}_{0,10} + \hat{\pi}_{1,10}p(A_{i10}) + \hat{\pi}_{2,10}(p(A_{i10}))^2 + \hat{\pi}_{3,10}(p(A_{i10}))^3 \tag{1.45}$$

$p(A_{i10})$ is value after Chebyshev polynomial transformation. Now, we have $\widehat{Emax}(A_{i10})$ for all the asset values, A_{i10}.

To find $\widehat{Emax}(A_{it})$ for period $t = 9, 8, ..., 2$, we use following sequences.

(a) Prepare set of measurement errors, $\{\varepsilon_{it}\}$, on Gauss-Hermite quadrature points as we did for the last period. $\{\varepsilon_{it}\}$ is same for all i and t.

(b) Find value of $V_t(A_{it}, \varepsilon_{it,l})$ for a given error value ε_{it} as follows.

$$V_t(A_{it}, \varepsilon_{it,l}) = \max_{c_{it}} \left[u(c_{it}) + \widehat{Emax}(A_{it+1})\right]$$

$$= \max_{c_{it}} \left[u(c_{it}) + \widehat{Emax}(R(A_{it} + g_t(1 + \varepsilon_{it,l}) - c_{it}))\right] \tag{1.46}$$

Then, find Emax value at A_{it} as follows:

$$E\max(A_{it}) = \sum_{l=1}^{20} \frac{V_t(A_{it}, \varepsilon_{it,l})W(l)}{\sqrt{\pi}} \tag{1.47}$$

(c) To find $\widehat{Emax}(A_{it})$, set functional form for $Emax(A_{it})$ as follows:

$$\pi_{0,t} + \pi_{1,t} p\left(A_{it}\right) + \pi_{2,t} (p\left(A_{it}\right))^2 + \pi_{3,t} (p\left(A_{it}\right))^3 \tag{1.48}$$

Then, estimate $\hat{\pi}_{0,t}$, $\hat{\pi}_{1,t}$, $\hat{\pi}_{2,t}$ and $\hat{\pi}_{3,t}$ for following equation.

$$\widehat{Emax}(A_{it}) = \hat{\pi}_{0,t} + \hat{\pi}_{1,t} p(A_{it}) + \hat{\pi}_{2,t} (p(A_{it}))^2 + \hat{\pi}_{3,t} (p(A_{it}))^3 \tag{1.49}$$

(d) Repeat above sequence in $(a) - (c)$ for $t = 8, ..., 2$.

Step 2. Finding log-likelihood

We use similar MSL approach that we applied for the estimation of model 1 case. However, this approach is slightly different in that we need ε_{it}^m to solve for equation (1.50).

$$c_{it}^m = \arg\max_{c_{it}{}^m} \left[\frac{(c_{it}^m)^{1+\alpha}}{1+\alpha} + \widehat{Emax}(A_{it+1}^m) \right] \tag{1.50}$$

where $A_{it+1}^m = R(A_t^m + g_t(1 + \varepsilon_{it}^m) - c_{it}^m)$. Therefore, we need to generate ε_{it}^m at the beginning of each period. Aside from this difference, we use the same simulation procedure as for the extreme error case.

Step 3. Finding optimal candidate value.

We use numerical search method for this step. For contaminated data, we use two way optimization to maximize $l(\hat{\alpha}, \hat{\sigma}_\xi)$ while we use one way optimization for pure data case to maximize $l(\hat{\alpha}|\sigma_\xi = 200)$. Like extreme structural error case, we use COBYLA as numerical optimizer.

Discrete Choice Estimator

The DC estimator always assumes added type-I extreme distribution. Therefore, we do not give any change to the estimator according to the structural error distribution. For the estimation of model 2 and model 3, we follow the same steps in the DC estimator that we use for model 1.

Continuous Choice Bayesian Estimator

When we apply the CC or CCB estimator for a dataset with different structural error distribution, difference only exists in the likelihood finding step. For the estimation of data from model 2 and the model 3, the CCB estimator follow the same likelihood finding step of the CC estimator. Then the CCB estimator applies the Metropolis algorithm to find estimated value and posterior distribution.

1.5 Simulation Results

In this section, we provide results from the CC estimator and the DC estimator under various error conditions. We start this section by providing estimation results on data generated from model 1. Theoretically, this assumption supports reliable estimation result when we use the DC estimator. We use the result from this error setting as a benchmark for other error settings. We will proceed to the estimation of model 2. As we specified in model section, structural error is not additive to utility function. Then we check the simulation results of the estimators on data generated from model 3 where structural errors do not follow type-I extreme distribution. Next topic is about the change of estimation quality that comes from the existence of measurement errors. First, we add measurement errors to asset values in the dataset that we used for benchmark result. Second, we add measurement errors to consumption values and check the change in the estimation quality. For this setting, there is no

	DC		CC	
Parameter	*Est.V.*	*S.E.*	*Est.V.*	*S.E.*
Value	-0.708	0.045	-0.694	0.025

Table 1.6: Model 1, No Measurement Errors

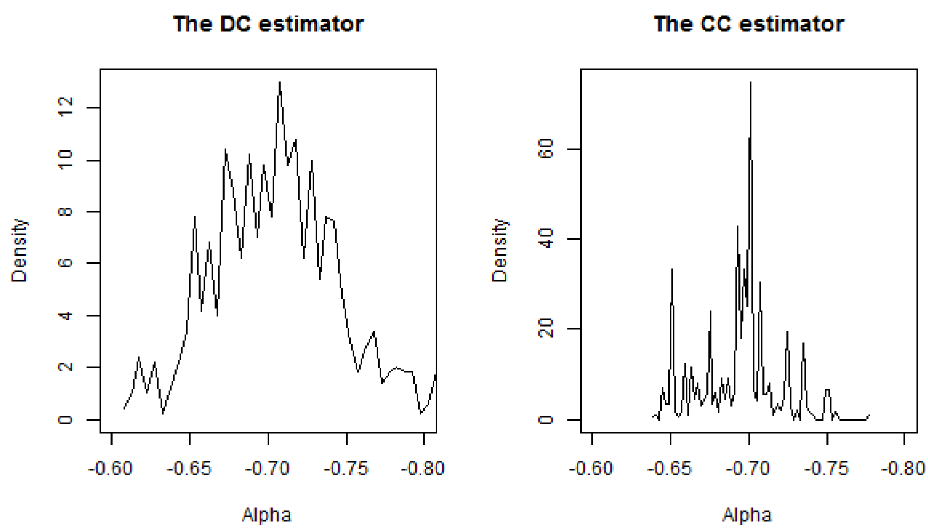

Figure 1.1: Model 1, No Measurement Errors

measurement errors on asset values. We execute this test to check whether the type of contaminated variable, state or choice, affects the estimation reliability. Finishing this section, we provide estimation results from the CCB estimator as a comparison.

1.5.1 Model 1 without Measurement Error

In table 1.6, we obtain the expected results from the benchmark test. Structural error assumption follows conditions assumed for the use of the DC estimator, so it is reasonable that the DC estimator provides reliable point estimation. The CC estimator also brings unbiased estimation result. There are two notable features of the benchmark estimation result, the first is that the CC estimator provides smaller

	DC		CC	
Parameter	*Est.V.*	*S.E.*	*Est.V.*	*S.E.*
Value	-0.470	0.014	-0.695	0.065

Table 1.7: Model 2, No Measurement Errors

variances even for the estimation of the models whose structural errors are set to be ideal for the use of the DC estimator and the second is that the CC estimator works well when we use the CC estimator on data set without measurement error. Because there is no measurement error in data set, we generated artificial measurement errors and added them to asset values. Estimation result shows that the CC estimator still provide reliable point estimation.

Figure 1.1 provides histograms of 1,000 estimated values. In the histogram from the CC estimator, estimated values are condensed around the true parameter value with uni-modal shape while distribution from the use of the DC estimator is more dispersed, which means that we can reduce the risk of obtaining biased estimated value when we use the CC estimator.

1.5.2 Model 2 without Measurement Error

When we estimate data generated from model 2, our focus is on checking the properties of estimators when the additive assumption of errors for the DC estimator is violated. Estimation results in table 1.7 shows that the use of the DC estimator on dataset which does not have additive structural error may cause serious bias problem. True parameter value is clearly rejected at 5% significance level. As shown in figure 1.2, the histogram of estimated values from the DC estimator fails to obtain estimated

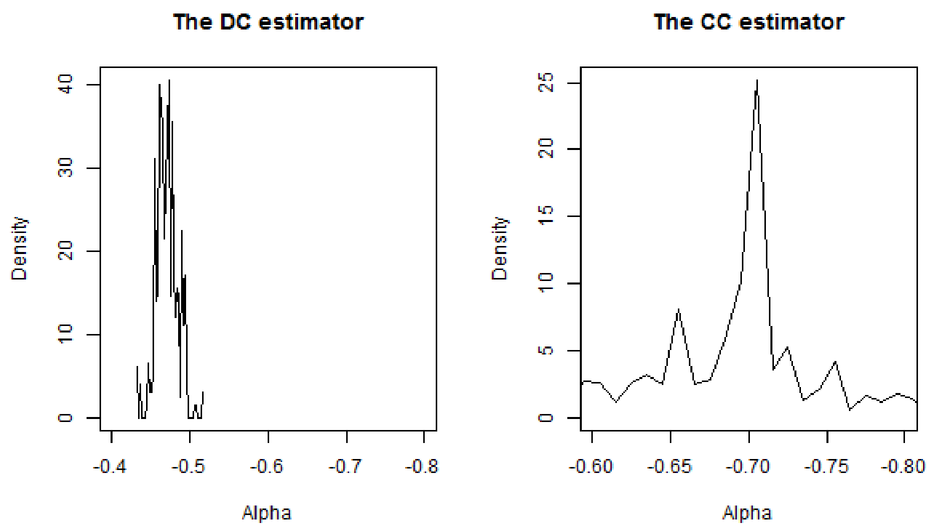

Figure 1.2: Model 2, No Measurement Errors

	DC		CC	
Parameter	*Est.V.*	*S.E.*	*Est.V.*	*S.E.*
Value	-0.580	0.037	-0.715	0.050

Table 1.8: Model 3, No Measurement Errors

value close to true parameter value. This result shows that it is very concerning to use the DC estimator if we do not have clear evidence that structural errors contained in data follow additive to utility. On the other hand, the CC estimator shows more reliable estimation results. In figure 1.2, estimated values are located around true parameter value and, in particular, the mode is close to -0.7. By the way, estimation results shows that standard error of estimation is much increased compared with model 1 case. This results seem reasonable when we consider that model 1 is deterministic and random shock from structural errors added uncertainty to model 2.

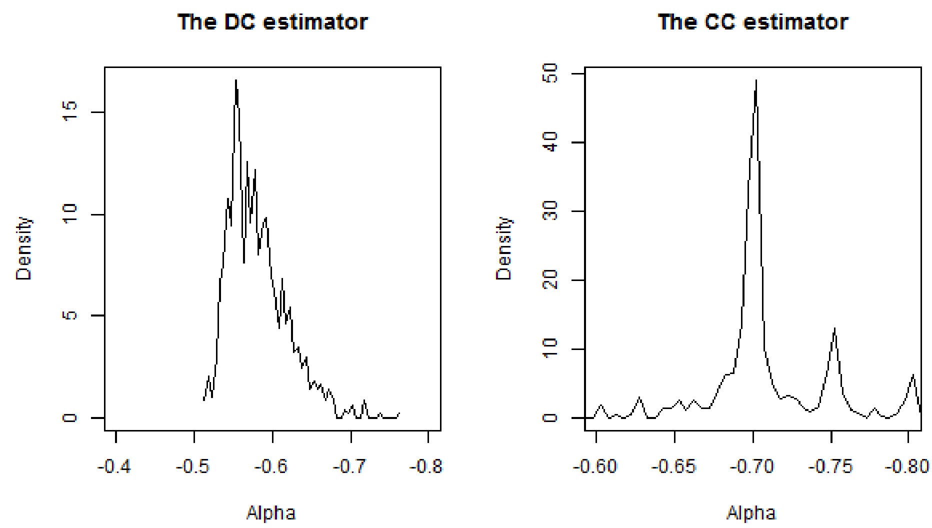

Figure 1.3: Model 3, No Measurement Errors

1.5.3 Model 3 without Measurement Error

We test estimators with this model setting for two reasons. The first goal is to check the reliability and efficiency of the DC estimator when error assumptions for the DC estimators are seriously biased. The second goal is to check the flexibility of the CC estimator which can adjust estimation approach according to the type and structure of errors. In model 3, error setting violates two important assumptions for the DC estimator. The structural error is not additive to utility function and does not follow type-I extreme distribution. The result can be meaningful in two ways. If we obtain good estimation result, it refutes criticism about the DC estimator under error conditions other than additive type-I error distribution. In opposite case, results will motivate us to find an alternative estimator for estimation of datasets which may not have additive type-I extreme error distribution. Estimation results provide the following findings.

	DC		CC	
Parameter	*Est.V.*	*S.E.*	*Est.V.*	*S.E.*
Model 1	-0.781	0.038	-0.680	0.085
Model 2	-0.471	0.014	-0.717	0.133

Table 1.9: Model 1 and Model 2, with Measurement Errors

The DC estimator shows significant decline in the estimation quality compared with benchmark model 1 setting. In table 1.8, the estimation result from the DC estimator provides large bias in point estimation with 95% confidence interval of $(-0.652, -0.510)$ which does not contain the true parameter value.

On the other hand, we do not find that a change of the error assumption has a significant effect on estimation performance of the CC estimator. The CC estimator shows reliable estimation result in unbiasedness standards because the CC estimator uses a structure that simply uses distribution of measurement error without further assumption. In a similar vein, the CC estimator may have highly reliable estimation result when it is used for different types of structural error distribution. Along with robustness shown with tests on data containing measurement errors, this result shows that the use of the CC estimator should be seriously considered when we estimate data with unknown error properties. However, further investigation is seriously needed with regards to the properties of the CC estimator with other structural error distributions to argue that the CC estimator proves its reliability; because we only tested two different type of structural error distributions in this chapter.

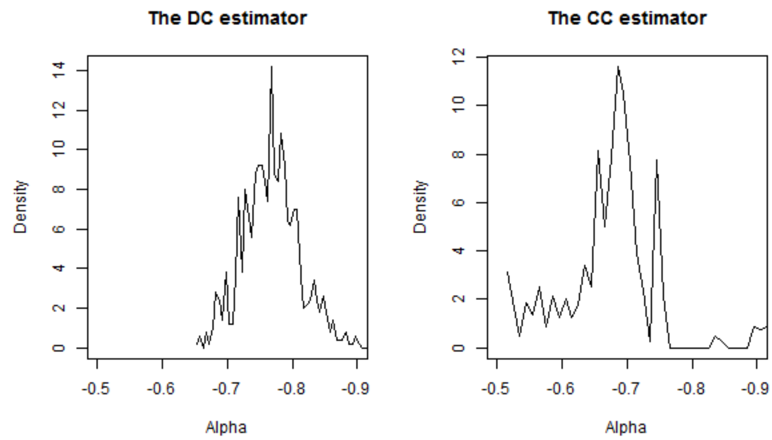

Figure 1.4: Model 1, with Measurement Errors

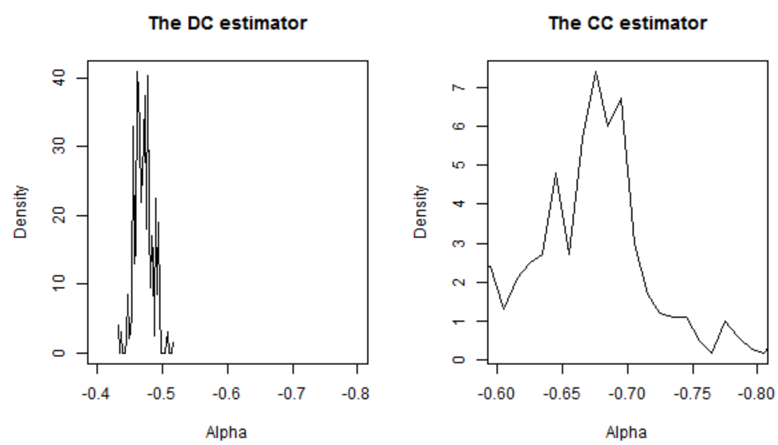

Figure 1.5: Model 2, with Measurement Errors

1.5.4 Model 1 and Model 2 with Measurement Errors

Estimation results in table 1.9 show that existence of measurement error significantly decreases estimation quality of the DC estimator. In the estimation of model 1, estimated value of α from the use of the DC estimator has 95% confidence interval $(-0.852, -0.704)$ and does not contain a true parameter value. Bias is a more serious problem when we use the DC estimator for model 2 data which contain measurement error. On the other hand, the CC estimator provides reliable result. One

# of Agent	Model 1			Model 3		
	100	200	500	100	200	500
Est. Value	-0.684	-0.688	-0.688	-0.687	-0.687	-0.688
S.E	0.026	0.022	0.019	0.027	0.023	0.019

Table 1.10: Estimation Result from Different Sample Size, the CC Estimator

possible explanation for this difference is that the DC estimator does not consider the existence of measurement error and has no way to measure that kind of error and adjust estimation as the CC estimator does. The third histogram in figure 1.4 is a histogram illustrating estimated value of measurement error. Estimated values are mostly located around the true parameter, demonstrating that the CC estimator provides reliable measurement error estimation.

Table 1.10 provides estimation results when we have different number of agents in dataset. Results show that efficiency of estimation improves as the number of sample sizes grows. Specifically, point estimation results show less bias as the number of agents goes up. We can not argue that the CC estimator is consistent with this limited finite sample test. However, this result indicates that the CC estimator has consistency as an estimator. Figure 1.6 provides histogram of estimated α values from data set with different number of agents. It shows that estimated values are more condensed around true parameter value as the number of agents increases for both error distribution cases. It also shows the distributional shape becomes more uni-modal. This means that the risk of obtaining seriously biased point estimation result from single application gets lower as sample size increases.

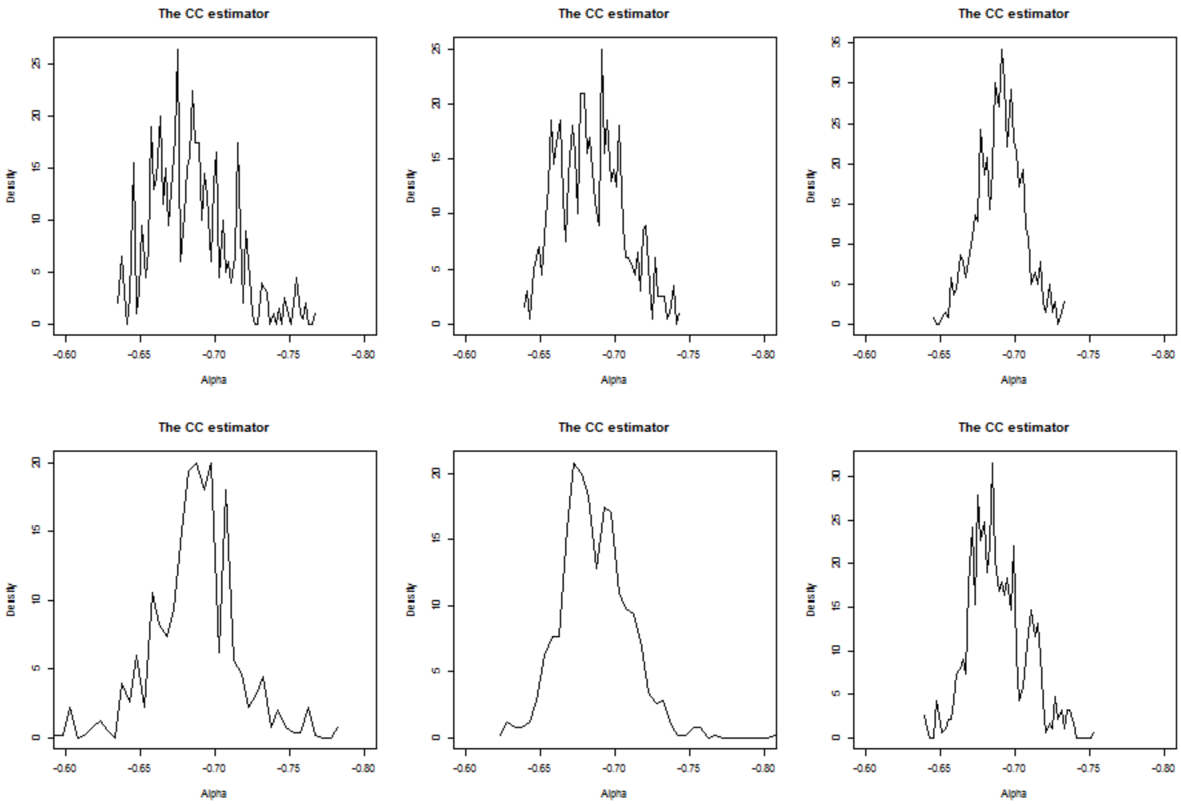

Figure 1.6: Estimation Result From Different Sample Size, the CC Estimator

	DC		CC	
Parameter	*Est.V.*	*S.E.*	*Est.V.*	*S.E.*
Value	-0.824	0.065	-0.695	0.021

Table 1.11: Model 1, with Measurement Errors on Consumption

1.5.5 Model 1 with Measurement Errors on Choice Variable

Previously in this section, we added the measurement error to asset which is the state variable. We will now check the properties of estimators when the choice variable contains measurement error. To contaminate consumption with measurement error, we generate a random shock of η_t which follows $\eta_t \sim N(0, 50^2)$ and add each

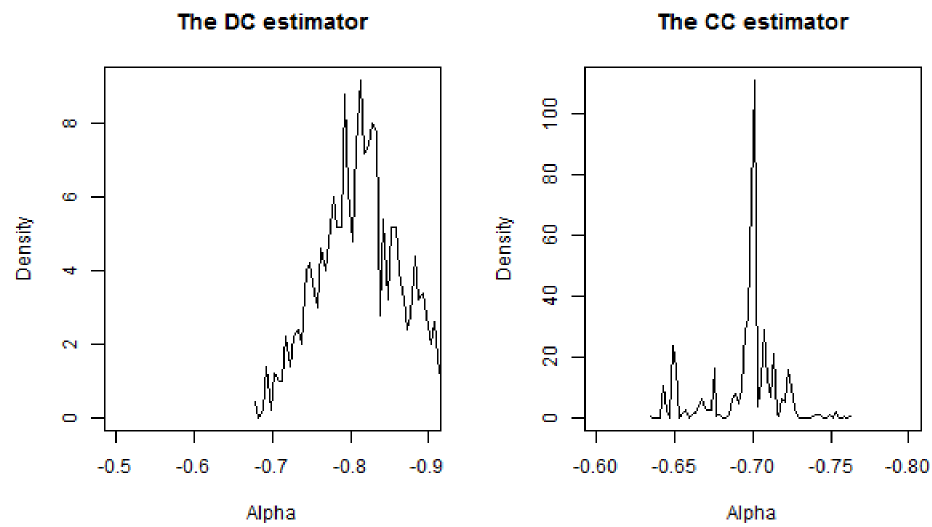

Figure 1.7: Model 1, with Measurement Errors on Consumption

measurement error to true consumption value c_t and data set has contaminated value of c_t^D as follows.

$$c_t^D = c_t + \eta_t$$

To find log likelihood for the CC estimator, we use almost identical steps that we used to obtain equation (1.19).

Estimation results in table 1.11 show that measurement error on choice variable leads to an even worse bias problem and may have larger decline of the DC estimator's estimation quality. This result also clearly demonstrates that the CC estimator is robust against the measurement errors contained in the choice variable. Actually, the histogram shape that we obtain when we apply the CC estimator for a dataset with contaminated choice variable looks better than histogram from the CC estimator application for a dataset that contains measurement errors in asset values. One possible explanation for the improvement of estimation results is that the initial value of the asset has no measurement error in this simulation case. Because the initial value of

	Model 1		Model 3	
Parameter	*alpha*	*sigma*	*alpha*	*sigma*
Est. Value	-0.689	98.11	-0.696	99.70
S.E	0.020	2.34	0.031	2.61

Table 1.12: The CCB Estimator on Model 1 and Model 3

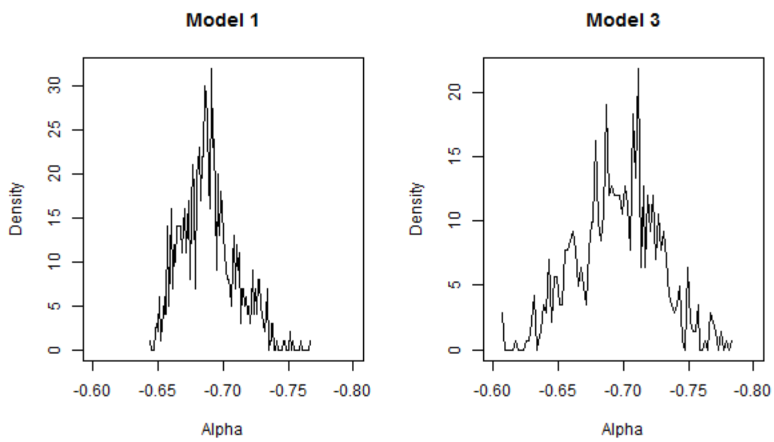

Figure 1.8: The CCB Estimator on Two Type of Structural Errors

the asset affects the whole likelihood finding procedure, the use of a true initial asset value may improve the CC estimator's reliability.

1.5.6 Continuous Choice Bayesian Estimator

This section provides results from the Continuous Choice Bayesian (CCB) estimator which is structural estimation method first applied in this paper. As a comparison to the CC estimator, we applies the CCB estimator on datasets with both types of structural error. Datasets also contains measurement errors. In table 1.11, the CCB estimator shows better estimation results in both unbiasedness and efficiency standards compared with the DC estimator. Any Notable features of the CCB estimator from this study are summarized as two following findings.

First, the CCB estimator provides reliable estimation results on data with both model 1 and model 2. Compared with the CC estimator, it has less bias problem for data from both models as shown in table 1.12. The shape of histogram is clearly uni-modal in the estimation of model 1. The histogram from model 2 estimation is close to uni-modal even though it is not as clear as model 1 estimation result. These results mean that we can expect the CCB estimator deliver reliable estimation result in single application.

Second, the CCB estimator requires heavy computational burden. To provide information about the cost of computation, Table 1.13 has measured time of estimation from approaches covered in this paper. Compared to the CC estimator, the CCB estimator consumes 35 times more time. To mitigate increased computational burden, we applied parallel programming approach [1] on the CC and CCB estimator. In table 1.12, '# Core' represents for the number of CPU cores used for estimation at the same time. When we used 6 cores, we could save about 65% of computing time for each estimators compared to typical single core processing. This result shows that the benefit of using parallel computing will be significant as the number of parameters increases for complex model. If we compare the histograms in figure 1.5 and figure 1.6, distribution shape from the CC estimator gets as good as or better than that from the CCB estimator as sample size increases. This result show that use of the CCB estimator is encouraged when we have relatively small sample size or we need posterior distribution as a result of estimation if we consider huge computational burden.

[1] We used OpenMP with gcc to speed up loop calculation

Estimator	DC	CC		CCB	
# Core	1 core	1 core	6 core	1 core	6 core
Model 1	.5s	52s	19s	34 m 1s	12 m 29s
Model 3	.9s	1m 02s	22s	35 m 55s	13 m 16s

Table 1.13: Measured Time of Estimation

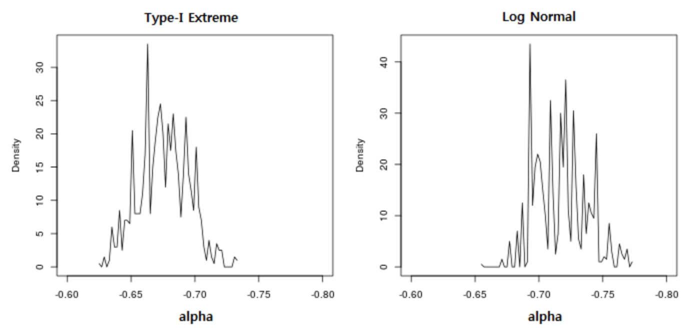

Figure 1.9: Posterior Distribution, Continuous Choice Bayesian Estimator

Figure 1.8 provides the example of the posterior distribution which is useful output from the CCB estimator. It comes from collecting accepted candidate $\hat{\alpha}$ and $\hat{\sigma}_\xi$ in each use of the CCB estimator after burning initial 100 values that were accepted before convergence.

1.6 Concluding Remarks

This study was conducted to check the finite sample properties of the structural estimators. Especially, the CC estimator has intuitively appealing factors as a structural estimator but this estimator has not been frequent choice for structural estimation researches. It is partly due to practical difficulty which comes from the demanding computational burden accompanied by the CC estimator application. However, more substantial obstacle blocking frequent use of this seemingly attractive estimator is that its property as an reliable estimator has not been thoroughly investigated.

Using Monte Carlo simulation study, we provide the following findings about finite sample properties of estimators.

First, the CC estimator provides reliable estimation results. This estimator does not show bias problem in estimation results on a dataset from different kind of error settings. As expected from its flexible setting which can be adjusted according to the expected model, its estimation quality is not significantly affected by the types of structural error. This estimator also shows reliable estimation result when substantial amount of measurement errors exists. On the other hand, these errors caused significant decline of estimation quality for the DC estimator that estimated parameter value from the DC estimator is rejected at 95% significance level. The reliability of the CC estimator is not affected by the types of variable (choice or state) which contains measurement errors, either. Another notable feature of the CC estimator is that the histogram of estimated values becomes more uni-modal and condensed around the true parameter value as the sample size increases. It means that we can expect improved estimation quality from the CC estimator when we have larger sample size.

Second, the DC estimator results raise concern for its reliability especially when this estimator is used under structural error conditions other than additive type-I extreme distribution. The DC estimator results clearly show that estimation quality declines seriously, when the DC estimator is used outside additive type-I extreme error distribution. Another concern for the use of the DC estimator is that this estimator can be affected by the existence of measurement errors, which is common in practice. The estimation results from the DC estimator show bias problem as the size of measurement errors contained in data set gets large. The decline of error quality become more significant when measurement errors contaminate choice variable.

Third, the CCB estimator provides mixed result. This estimator shows no bias problem under two different error settings tested and provides histogram of estimated values, which is close to uni-modal shape. It means that we can use the CCB estimator and obtain useful posterior distribution without concern for reliability. However, the CC estimator can provide estimation result as good as the CCB estimator as sample size grows. This means that use of the CCB estimator is only encouraged when we have small sample size or we need posterior distribution as an outcome of estimation because the CCB estimator requires heavy burden of computation compared to the CC or DC estimator. We could reduce estimation time to one third the original time when we applied parallel programming and this result shows that use of parallel programming can partly mitigate this increased computational burden.

Lastly, it is important to take into account the cost of using the continuous approaches, CC and CCB. As shown in measured time of estimation, when we apply these estimators, we clearly have increased computational burden. So far, advantages of simplicity provided by the DC estimator have been widely accepted by previous researches. However, advanced operation speed by recent processors and easier access to servers ideal for parallel programming remove significant portion of cost which has prevented the use of those methods, and this trend will grow, making such computing even cheaper. This study proves that the use of the continuous approaches are worth conductions.

Chapter 2

ESTIMATION OF MANAGERIAL EFFICIENCY IN BASEBALL: A BAYESIAN APPROACH

2.1 Introduction

After each baseball season, the Boston Globe ranks the baseball managers of 30 Major League Baseball (MLB) teams. In 2013 grading, John Farrell of the Boston Red Sox, which is the home team of the Boston area, was ranked as a top manager among all MLB teams. The Boston Globe explains that they used the opinions from a number of baseball people, including managers, coaches, scouts, players, and front office executives to formulate this ranking. However, not every baseball fan including me will agree on the ranking from the Boston Globe based only on the Boston Redsox's winning the 2013 world Series championship. I have two arguments against this ranking. One is that several baseball teams were seriously plagued by a disastrous series of injuries to their key players. Team performance of those teams can be worse than the Boston Redsox regardless of Managerial efficiencies. Another argument is that the Boston Redsox boasts of the fourth biggest payroll in MLB that it consists of better players than small payroll teams like the Tampa Bay Rays who spent the mere amount of $57,030,272 which is only one third of Boston's payroll but still could advance to the Divisional Series.

The goal of this paper is to provide a quantitative solution to the above question: "Which team has shown better managerial efficiency? To answer that question, we use a stochastic frontier function approach and estimate the managerial efficiency in baseball. As a result, we will first provide the value of estimated managerial efficiency of each team. Using this value, we can compare the managerial performance of each MLB team. Another focus of this paper is about the reliability of this comparison procedure. As Bera and Sharma (1999) pointed out, it is unfortunate that we have not given much attention to the reliability of comparison procedure. To approach this topic, we will estimate uncertainty in stochastic frontier production function model with two different methods. The first one is the Multiple Comparison with the Best (MCB) suggested by Horace and Schmidt (1996) and the other one is the Bayesian estimation approach suggested by Koop, Osiewalski, and Steel (1997) . Here is the reason why we focus on above two approaches. Bera and Sharma (1999) and Green (2007) suggest MCB as the most frequently used method to build confidence interval for the stochastic frontier function. Holloway (2005) argues that the Bayesian approach has an advantage as an alternative to the frequentist approach. The reason behind his argument is that the choice of distributional assumption for inefficiencies has little effect on estimation result as is mentioned in Koop, Osiewalski and Steel (1997). To check the validity of this argument, we check the estimated values and the variance under different distributional assumptions in this paper. Here is another reason why we focus on the application of Bayesian approach on baseball data. Sickles and Schmidt (1984) shows that when we apply stochastic frontier analysis on panel data, if we use panel data with fairly large number of period, estimated value of inefficiency will be precise. However in baseball, the typical contract length for the manager is three years and this contract is not easily extended. When we have panel dataset with the only small number of periods, Horace and Schmidt (2000) argue that the estimated variance of individual inefficiency is fairly large and confidence interval of efficiency should be large as a result. In this case, it is not easy to compare the managerial efficiency of two teams. In this paper, we apply the Bayesian approach and test whether it provides smaller variance value for inefficiency estimation. To validate the Bayesian approach, we will first show that the Bayesian approach provides reliable point estimation with a smaller variance using a simulation study. In simulation study, Bayesian approach does not necessarily obtain narrower confidence intervals. An informative prior is needed to provide narrower confidence interval. We will use data from the 2011 to 2013 MLB seasons with generated inefficiency values. Then, we will provide efficiency estimation results to answer interesting questions related to managerial efficiency in MLB. Here is additional reason why we choose baseball as a topic. Porter and Scully (1982) argues that baseball data is especially appealing target for frontier analysis because outputs and inputs are unambiguously measured, and production function is simply specified. They showed that managerial skill in baseball contributes very substantially to the production process.

This paper is organized as follows. Section 2.2 provides a brief description of related literatures. Section 2.3 describes the model used for estimation. Section 2.4 explains the data we use for this paper. Section 2.5 will compare the Bayesian approach and the frequentist approach mainly using simulation study. Section 2.6 answers various questions about baseball. Some concluding remarks follow in Section 2.7.

2.2 Literature Review

The contemporary model for stochastic frontier analysis which focuses on estimating inefficiencies in production was first suggested by Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977). This chapter is also based on basic frame works by Aigner, Lovell and Schmidt (1977). However, there had been research papers focusing on production inefficiencies before Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977) combined stochastic error concept to frontier function. Deterministic analysis of production inefficiency was first started by

Debreu (1951) and Farrell (1957).

Schmidt and Sickles (1984) provide the way to solve some difficulties originally presented in Aigner, Lovell, and Schmidt (1977) by using panel data. In their paper, they showed that if panel data is used and the number of periods is large enough, these three difficulties can be avoided:

1. A question of consistency in the estimation of technical inefficiency
2. Choice of distributional assumption for the distribution of technical inefficiency
3. Probable correlation between regressors and inefficiency.

Battese and Coelli (1992) and Battese and Coelli (1995) provide the method which is useful for a practical approach for the estimation of panel data. They also provides computational packages which can be conveniently used for estimation procedure. This paper also uses their R package 'Frontier' to provide estimation result and confidence interval provided by frequentist approach.

A meaningful suggestion for building the confidence interval was first made by Horace and Schmidt (1996). Horace and Schmidt (1996) introduces the MCB concept of Hochberg and Tamhane (1987) to find confidence interval of production inefficiencies and provide an application example. Hsu (1996) is more detailed reference for the MCB concept. Bera and Sharma (1999) argues that one of the main goals when production inefficiency is estimated is to compare inefficiency level among cross-sectional entities and confidence interval is useful for estimating reliability of the comparison. Horace and Schmidt (2000) provides a more user-friendly recipe to build confidence interval with MCB which was introduced in Horace and Schmidt (1996) with more various kind of application examples. In this chapter, we use method presented in Horace and Schmidt (2000) to build the frequentist confidence interval.

A Bayesian approach to stochastic frontier analysis was introduced by Broeck, Koop, Osiewalski and Steel (1994). This paper shows that when the number of periods in panel data is not large enough, the difficulty which comes from the choice of distributional assumption of technical efficiency can be reduced by the Bayesian approach. In this chapter, we additionally show that the estimation output from a Bayesian approach shows better efficiency compared to other approaches. As a illustration, we compare the interval of estimated technical efficiency estimation from a Bayesian approach with the frequentistss interval.

An empirical application of stochastic frontier approach has covered a wide range of topics in economics and other study areas. It is well summarized in Fried, Lovell and Schmidt (2007). Especially in sports economics, there have been works on estimating the technical efficiency of the coach or organization in team sports using a stochastic frontier function. Dawson, Dobson and Gerrard (2000) estimates the efficiency of coach in English Premier League from 1992 to 1998 using panel data stochastic frontier model. Rimler, Song and Yi (2010) estimates technical efficiency in Atlantic 10 conference in NCAA Basketball. This paper argues that the managerial efficiency difference is trivial and focus more on the contribution of player statistic on winning percentage.

In baseball, Porter and Scully (1982) uses a frontier model to estimate managerial efficiency and Ruggiero, Hadley and Gustafson (1996) evaluates managerial efficiencies using Data Envelopment Analysis method. Both papers are using the non-stochastic model so they inevitably have the drawbacks all the deterministic frontier models share.

2.3 Model

Our study estimates the managerial efficiency in baseball games using the stochastic frontier function with a Bayesian approach. To explain the goal and the organization of our study, we start by briefly introducing the concept of the stochastic frontier function and the reason behind using a Bayesian approach.

Stochastic frontier analysis was first introduced by Aigner, Lovell, and Schmidt (1977). In this paper, they suggest an approach to the estimation of frontier production functions. They define stochastic frontier function for firm i at period t, and it is given by:

$$y_{it} = f(x_{it}; \beta)e^{u_{it}} \tag{2.1}$$

where y_{it} is the maximum level of output, x_{it} is a vector of input, β is an unknown parameter vector and $e^{u_{it}}$ is error term. Then, they assume there exists a technical inefficiency as deviation of actual production from maximum level of output. With existence of technical inefficiency, stochastic frontier production function is given by:

$$y_{it} = f(x_{it}; \beta)\tau_i e^{u_{it}} \tag{2.2}$$

where $0 \leq \tau_i \leq 1$ is a measure of firm specific inefficiency. By using error term, we can fix the critical problem shared by all the deterministic frontier estimation models. That problem is that any deviation of an observation from the frontier must be attributed to an inefficiency because a deterministic model does not assume the existence of statistical noises or measurement errors.

A Model for estimating managerial efficiency comes from equation (2.2). Let's assume that production function $f(\cdot;\cdot)$ is Cobb-Douglas production function. Using log linear transformation, equation (2.2) becomes:

$$\ln y_{it} = \ln x_{it}'\beta + u_{it} - z_i, \qquad i = 1, \cdots, N, t = 1, \cdots, T \tag{2.3}$$

where $z_i = -\log \tau_i$.

Here i indexes teams and t indexes time periods while 1 period is 1 season in baseball. We make equation (2.4) by splitting x_{it} in equation (2.3) into $x_{o,it}$ and $x_{d,it}$.

We use (2.4) to estimate managerial efficiency and β's:

$$\ln y_{it} = \ln x'_{o,it}\beta_o + \ln x'_{d,it}\beta_d + u_{it} - z_i, \qquad i = 1, \cdots, N, t = 1, \cdots, T \tag{2.4}$$

where y_{it} is the ratio of run made by team i over run allowed by team i in season t $\left(\frac{Run_{made,it}}{Run_{allowed,it}}\right)$. $x_{o,it}$ is the vector of offensive numbers for team i in season t, including single, double, triple, homerun, steal, walk, strikeout. $x_{d,it}$ is vector of defensive numbers including single allowed , homerun allowed, walk allowed, strikeout made, error commited (by team i). Data values are yearly summation of each team's whole regular season games. β_o and β_d are coefficients for offensive and defensive variables.

Most of the previous papers estimating technical efficiency in team sports use winning percentage as the value of y_{it}. But we use run ratio as the value of y_{it} following Bill James (1981). In his book which started current sabermetrics, Bill James argues that difference between a run production and a run allowance is the best tool to describe the team's quality and each baseball team try to maximize the difference between a run production and a run allowance. Following this view, there have been various researches such as Pythagorean approach to find formula to link team's run difference and winning percentage. In other sports, this view is recently gaining more attention. For example, in basketball, point differential per 100 possessions (NetRtg) is regarded as the best source to judge the real quality of a team that National Basketball Association (NBA) official power ranking provides NetRtg along with peer review of basketball experts to rank the NBA teams' real quality.

For stochastic frontier analysis with a Bayesian approach, we use assumptions used by Koop, Osiewalski, and Steel (2007). They can be summarized as follows:

For $i = 1, , N$

1. $u_{it} \sim N(0, \eta^{-1})$ and the u_{it}s are independent to each other

2. Prior distribution for η: $\eta \sim \gamma(\underline{s}^{-2}, \underline{\nu})$

3. u_{it} and z_i are independent to one another

4. z_i and z_j are independent to each other when $i \neq j$.

5. Prior distribution for z_i: $z_i \sim Exp(\mu_z)$

6. Prior distribution for μ_z^{-1}: $\mu_z^{-1} \sim \gamma(\underline{\mu}_z^{-1}, \underline{\nu}_z)$

7. Prior distribution for β: $\beta \sim N(\underline{\beta}, \underline{V})$

where μ_z, $\underline{s}^{-2}$, $\underline{\nu}$, $\underline{\mu}_z^{-1}$, $\underline{\nu}_z$, $\underline{\beta}$ and $\underline{V}$ are hyperparameters.

Among all the assumptions, the most critical one is the assumption of $z_i \sim Exp(\mu_z)$ because choice for the distribution of technical efficiency always has been the most difficult part in the application of stochastic frontier analysis. Van den Broeck et al. (1994) argues the exponential is the least sensitive prior distribution in a study of the most commonly used models and we follow their argument for our study.

From equation (2.4) and additional assumptions, likelihood function is given as:

$$p(y|\beta, \eta, z) = \prod_{i=1}^{N} \frac{eta^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp\left[-\frac{\eta}{2}(y_i - X_i\beta + z_i\iota_T)'(y_i - X_i\beta + z_i\iota_T) \right] \right\} \tag{2.5}$$

where $X_i = [X_{i1} \ldots X_{iT}]'$, $X_{it} = [x'_{o,it}, \ldots, x'_{d,it}]$ and $\beta = [\beta_o \quad \beta_d]$

Using likelihood function (2.5) and distributional assumptions, we are ready to make posterior distribution of parameters. Starting from β, we multiply likelihood function with assumption $\beta \sim N(\underline{\beta}, \underline{V})$ then, posterior distribution of β is given as:

$$\beta|y, \eta, \mu_z \sim N(\overline{\beta}, \overline{V}) \tag{2.6}$$

where $\overline{V} = \left(\underline{V}^{-1} + \eta \sum_{i=1}^{N} X'_i X_i \right)^{-1}$ and $\overline{\beta} = \overline{V} \left(\underline{V}^{-1}\underline{\beta} + \eta \sum_{i=1}^{N} X_i' \left[y_i + z_i\iota_T \right] \right)$

To find posterior distribution of η, we use likelihood function (2.5) and assumption 2 where prior distribution of η is $\eta \sim \gamma(\underline{s}^{-2}, \underline{\nu})$. Then, posterior distribution of η is given as:

$$\eta|y,\beta,z,\mu_z \sim Gamma(\overline{s}^{-1},\overline{\nu}) \tag{2.7}$$

where $\overline{\nu} = TN + \underline{\nu}$ and $\overline{s}^2 = \frac{\sum_{i-1}^{N}(y_i+z_i\iota_T-X_i\beta)'(y_i+z_i\iota_T-X_i\beta)+\underline{\nu}\underline{s}^2}{\overline{\nu}}$

The posterior distribution of μ_z^{-1} can be found in the similar way when we use assumption 6 and the likelihood function (2.5). The posterior distribution of μ_z^{-1} is:

$$\mu_z{}^{-1}|y,\beta,\eta,z \sim Gamma(\overline{\mu}_z{}^{-1},\overline{\nu}) \tag{2.8}$$

where $\overline{\nu}_z = 2N + \underline{\nu}_z$ and $\overline{\mu}_z{}^{-1} = \frac{2N+\underline{\nu}_z}{2\sum_{i=1}^{N} z_i+\underline{\nu}_z\underline{\mu}_z}$.

Now, we find the posterior distribution to generate technical inefficiency, z_i. Using Bayes' theorem, we know

$$p(z|y,\beta,\eta,\mu_z) \propto p(y|z,\beta,\eta,\mu_z)p(z|\beta,\eta,\mu_z)$$

We already have likelihood function (2.5) and the assumption for the prior distribution of z_i as $z_i \sim Exp(\mu_z)$. Then, posterior distribution of z_i comes from multiplication of likelihood function (2.5) and the prior distribution of z_i as follows:

$$p(z_i|y_i,X_i,\beta,\eta,\mu_z) \propto \phi\left(z_i|\overline{X_i\beta} - \overline{y}_i - (T\eta\mu_z)^{-1},(T\eta)^{-1}\right) I\left(z_i \geq 0\right) \tag{2.9}$$

where $\overline{y}_i = \frac{\sum_{t=1}^{T} y_{it}}{T}$ and $\overline{X}_i$ is a row vector containing the average value of each explanatory variable. $I(z_i \geq 0)$ is the indicator function. We should note that (2.9) uses assumption 4 which makes this formula simpler.

Using equations for posterior distributions from (2.6) to (2.9), we can find posterior distributions for coefficient β's, parameter values η, μ_z and technical efficiencies z_i's. We use Gibbs sampling to generate posterior distributions.

2.4 Data

Our study estimates data from Major League Baseball from 1969 to 2013 season. We set 1969 as the starting year for the data because MLB added four teams (Kansas

City Royals, Milwaukee Brewers, San Diego Padres and Montreal Expos) to the organization and made the fundamental changes in organization. Data comes from a baseball reference site: www.baseball-reference.com. During the period, the 1972, 1981, 1994 and 1995 seasons are omitted from the data because the number of games was reduced severely due to the labor disputes between players and MLB organization on those seasons. The number of teams has changed over the period because new teams have joined. Here is the summary of the number of teams:

Year	Number of seasons	Number of teams	Team added
1998-2013	16	30	AZD, TBR
1993-1997	3 (except for 94,95)	28	FLA, COL
1977-1992	15 (except for 81)	26	SEA, TOR
1969-1976	7 (except for 72)	24	

Table 2.1: The Number of Teams over 1969-2013 Seasons

Because there are the different number of teams, the estimation procedure for seasons from 1969 to 2013 should use an approach for the unbalanced case.

In the data set, there are three kind of values used for the analysis. We take log on both input and output values.

1. Output values - Run ratio. As described in section 2.3, a team's run ratio is the fraction of the summation of run produced by a team over the summation of run allowed by a team in a season.

2. Input values

2. a. Offensive input for run production - Single, Double & Triple, HR, Steal, Walk, and K are used as the values. We use Double & Triple together as an input value rather than each of Double and Triple because the number of triple is too small to use it as an additional explanatory variable. We decide to use Double & Triple to

increase the efficiency of estimation. Batting average, on base percentage (OBP), slugging percentage (SLG), and OPS (OBP + SLG) are the values decided by single, double, triple and homerun so we will not use them as explanatory variables to prevent multicollinearity problem.

2. b. Defensive input for run allowance - Single allowed, HR allowed, Walk allowed, K made, Error, Double play. Other important defensive variables such as ERA and WHIP are omitted from inputs because use of them can cause multicollinearity problem because ERA and WHIP are strongly related with other explanatory variables. Descriptive statistics of input data is provided baseball seasons (from 1969 to 1976) to modern baseball seasons (from 2007 to 2013) in table 2.2.

Input	1969 - 2013		1969 - 1976		2007-2013	
	mean	sd	mean	sd	mean	sd
Single	996	66	1023	66	958	70
2B & 3B	297	39	251	30	317	30
Homerun	148	39	120	31	160	32
Stolen Base	107	41	93	43	99	30
Walk	535	69	545	70	520	64
Strikeout	973	157	855	105	1146	125
Error	121	24	143	21	98	16

Table 2.2: Descriptive Statistics of Input Data (Yearly Value)

We can find several characteristic changes from the comparison of values from the old baseball and the modern baseball. At first, the number of extra base hits increased by large margin. There is huge increase in the average number of homeruns (+33.3%) and Double & Triples (+26.3%). On the other hand, the number of single hits decreased (-6.4%). Second, the number of strikeouts increased significantly ($+34.0\%$). When we combine the first and second change in descriptive numbers, it is highly possible that offensive players in the modern baseball are focusing more on producing extra base hits compared to old players. Because they are trying to load more power on each of their swings, they can make increased number of extra hits at the cost of more strikeouts. Third, this table shows that the number of errors decreased by large (-31.5%). This feature is actually related to the first and the second characteristics mentioned above. The decreased number of errors means that it is more difficult to make hits due to the improved fielding ability of defensive players. Increased difficulty of making a hit is also verified in the decreased number of single hits. Therefore, it is more difficult to produce runs from the series of hits in the modern baseball. This is the reason why the modern baseball players focus more on power hitting to produce extra base hits which can make runs without making a series of hits. Another possibility is that the power hitting tendency is the reason for the decreased number of errors. Longballs from power hitting tend to be more related to outfielders and errors happen more frequently in infield play than outfield play. This characteristic change of baseball also affects managerial efficiency and the change will be explained in detail in section 2.6.

We perform a graphical check on data using Quantile Quantile plot (QQ plot). In figure 2.1, the dependent variable, run ratio, shows normality and other explanatory variables do not show extreme skewness, either. Therefore, this dataset do not violate the necessary assumptions required for stochastic frontier approach, which is related to ordinary least squares requirements.

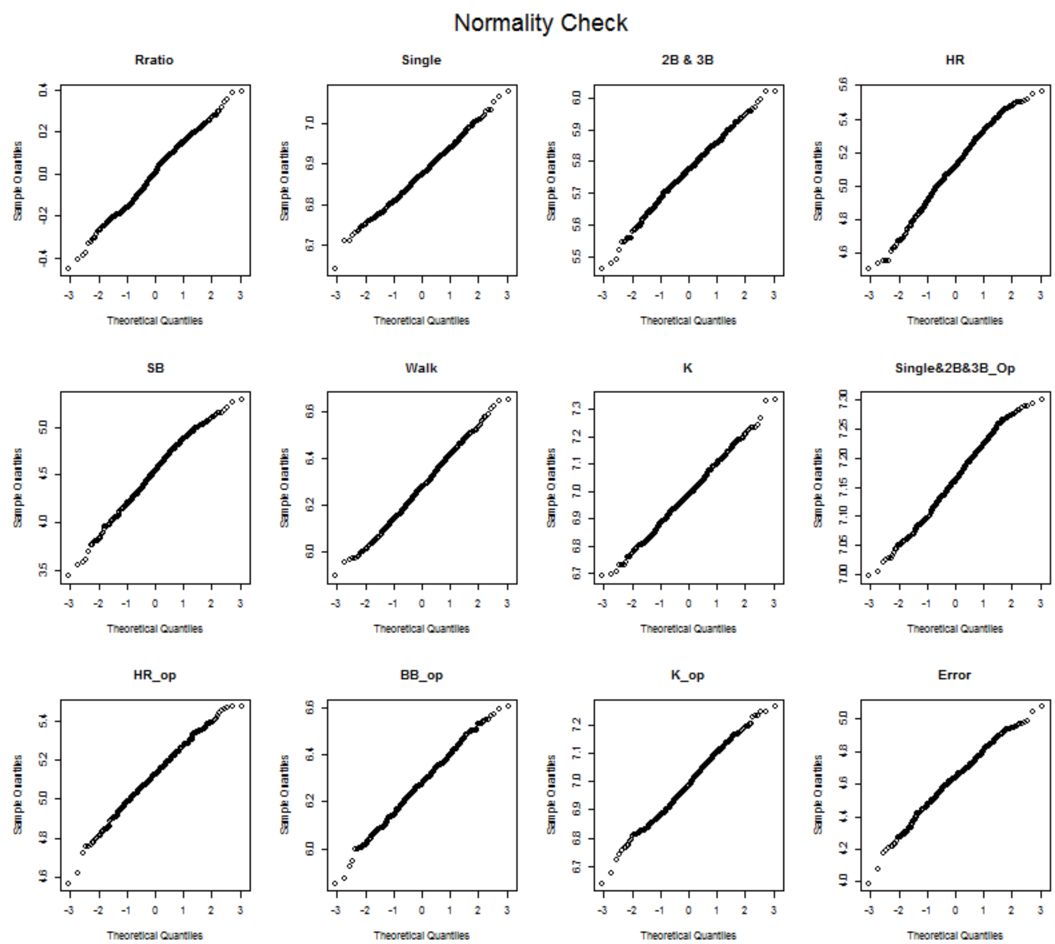

Figure 2.1: QQ Plot of Input and Output Data (1969 2013)

2.5 Comparison of Estimation Approach

2.5.1 Simulation Study

The goal of this section is to provide empirical evidence that the Bayesian approach can provide narrower confidence intervals for managerial inefficiencies compared to MCB approach when we analyze panel data with relatively small number of periods. We also try to find the reason why Bayesian approach provides narrower confidence intervals. We have an interest in this result because the typical contract term for a baseball manager is merely 3 years. We started this paper with the question about the managerial efficiency of John Farrell and the length of his first contract term as a manger was a mere 2 years with Toronto Blue Jays. Therefore, finding estimator which provides a relatively smaller confidence interval when we have small T value is important for the study of managerial efficiency in baseball. Simulation study consists of following steps.

Step 1. Prepare data to be estimated. To generate artificial run ratio which will be used as the dependent variable, we need the following values:

1. Explanatory variables (Offensive and Defensive variables)

2. Two stochastic error terms for production error and inefficiencies

3. Coefficient for Explanatory variables

First, we use the explanatory variables from 2011 to 2013 MLB seasons. Same with data used for the estimation procedure, explanatory variables are the values of following variables: single hit, sum of double and triple, homerun, steal, walk, strikeout, error, single allowed, the sum of single, double and triple allowed, homerun, walk allowed, and strikeout made. Second, we generate stochastic error terms, production random shock from simple distribution as follows:

$$u_{it} \sim N(0, 0.05^2)$$

Third, we generate team-specific managerial inefficiencies from following distribution.

$$z_i \sim Exp(\lambda)$$

where rate parameter $\lambda = 0.05$ to make average efficiency value be set close to 0.95 which is close to the average estimated efficiency value from 1969 to 2013 season.

Fourth, we find the coefficients for explanatory variables from ordinary least square (OLS) estimation of true run ratio (dependent variable) and explanatory variables. Now, we can generate artificial run ratio value from explanatory variable, stochastic error terms and estimated coefficient from OLS. Lastly, We generate 1,000 different datasets with different sets of stochastic error term u_{it} by repeating procedures above.

Step 2. Estimate managerial inefficiency from generated datasets from step 1.

For Step 2, we use two different stochastic frontier estimation approaches. First, we will follow estimator suggested by Battese and Coelli (1992) to find frequentists inefficiency point estimation value. Then, we build the confidence interval of inefficiencies by applying MCB from Horace and Schmidt (2000). Second, we will estimate efficiency with Bayesian estimator following Koop, Osiewalski and Steel (1997). Third, we repeat estimation of efficiencies using both the frequentist and Bayesian approach on 1,000 different datasets from step 1 and report mean value of 95% confidence interval of efficiencies.

Figure 2.2 provides box plots which provide 95% confidence intervals of 30 MLB baseball team from the third procedure of step 2. In figure 2.2, we can find the following things. First, the Bayesian estimation clearly provides the narrower confidence interval of production inefficiency. Second, estimated inefficiencies from two different methods are close to each other. Especially, order of each estimated values are almost identical. We have additional chance to check the order of efficiencies from two different methods in section 2.5.2. Table 2.3 provides the table of generated inefficiency values and estimation result (95% confidence interval) from the frequentist approach and the Bayesian approach.

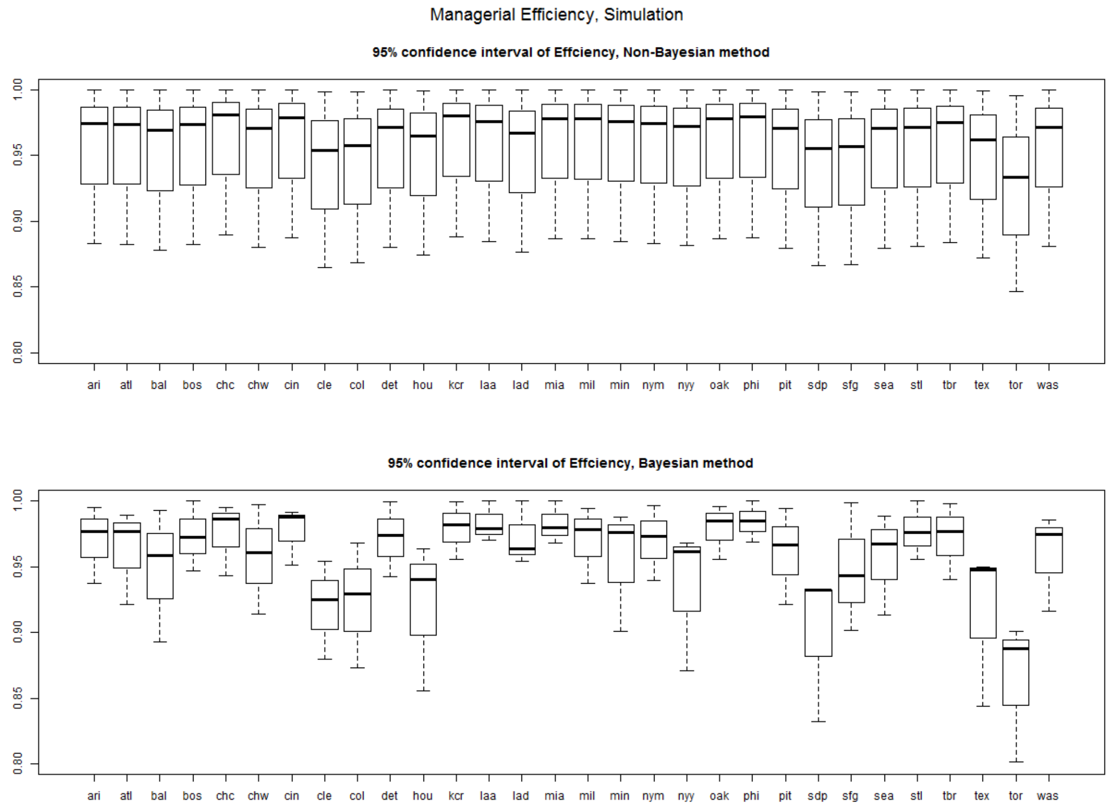

Figure 2.2: Box Plot of Estimated Efficiencies - Simulated Data (1,000 Times)

Result in table 2.3 shows that both approaches do not show the serious problem in estimation. None of estimated values are rejected at typical 5% significance level. However, confidence intervals from the Bayesian approach are clearly narrower than intervals from frequentist approach. At the same time, the sum of difference between true inefficiency and mean of estimated inefficiencies from the Bayesian approach is -0.435 which is smaller than -0.604 from the frequentist approach. This result shows that the Bayesian approach provides estimation performance at least as precise as frequentist approach for this dataset.

To find the reason behind the narrower confidence interval from the Bayesian approach, we checked the simulation result when we apply a Bayesian approach without

Team	Data	Frequentist			Inform. Bayesian		
	efficiency	lb	mean	ub	Lb	mean	ub
ARI	0.967	0.883	0.974	1.000	0.937	0.977	0.995
ATL	0.962	0.883	0.974	1.000	0.922	0.977	0.989
BAL	0.945	0.878	0.969	1.000	0.893	0.958	0.993
BOS	0.952	0.882	0.973	1.000	0.947	0.972	1.000
CHC	0.980	0.889	0.981	1.000	0.943	0.987	0.995
CHW	0.945	0.880	0.971	0.999	0.914	0.961	0.997
CIN	0.976	0.887	0.979	1.000	0.951	0.988	0.991
CLE	0.906	0.865	0.954	0.998	0.879	0.925	0.954
COL	0.910	0.868	0.958	0.999	0.873	0.929	0.968
DET	0.954	0.880	0.971	1.000	0.943	0.974	1.000
HOU	0.923	0.874	0.965	0.999	0.855	0.940	0.964
KCR	0.967	0.888	0.980	1.000	0.956	0.982	1.000
LAA	0.970	0.884	0.976	1.000	0.970	0.979	1.000
LAD	0.969	0.877	0.967	0.999	0.954	0.964	1.000
MIA	0.969	0.887	0.978	1.000	0.968	0.979	1.000
MIL	0.968	0.886	0.978	1.000	0.937	0.978	0.995
MIN	0.969	0.885	0.976	1.000	0.901	0.976	0.987
NY	0.959	0.883	0.974	1.000	0.939	0.973	0.996
NYY	0.950	0.881	0.972	1.000	0.871	0.961	0.968
OAK	0.975	0.887	0.978	1.000	0.955	0.985	0.996
PHI	0.979	0.887	0.979	1.000	0.969	0.985	1.000
PIT	0.949	0.879	0.970	1.000	0.921	0.967	0.994
SDP	0.907	0.866	0.955	0.998	0.832	0.932	0.933
SFG	0.918	0.867	0.957	0.999	0.902	0.943	0.999
SEA	0.954	0.880	0.970	1.000	0.913	0.967	0.989
STL	0.960	0.881	0.972	1.000	0.955	0.976	1.000
TBR	0.963	0.884	0.975	1.000	0.941	0.977	0.998
TEX	0.924	0.872	0.962	0.999	0.844	0.950	0.948
TOR	0.863	0.846	0.933	0.996	0.802	0.887	0.901
WAS	0.956	0.881	0.972	1.000	0.916	0.974	0.985

Table 2.3: Estimation Result: Simulation Study

uninformative prior for each team's efficiency value. We use uniform distribution as a prior distribution for this test instead of exponential distribution. In this case, we need to apply Metropolis-within-Gibbs approach as an updating procedure to generate posterior distribution of managerial inefficiencies because we cannot find closed form posterior generating function. Table 2.4 provides the simulation result. 95 % interval of estimated efficiency values are containing true efficiency values for all 30 teams. However, we can not conclude that interval lengths are narrower than those from frequentist approach. It shows that the use of exponential prior (informative

Team	Data	Uninform. Bayesian		
	efficiency	lb	mean	ub
ARI	0.967	0.861	0.947	0.998
ATL	0.962	0.865	0.950	0.999
BAL	0.945	0.840	0.936	0.998
BOS	0.952	0.861	0.949	0.999
CHC	0.980	0.882	0.961	1.000
CHW	0.945	0.840	0.936	0.998
CIN	0.976	0.884	0.960	1.000
CLE	0.906	0.800	0.903	0.989
COL	0.910	0.811	0.912	0.991
DET	0.954	0.858	0.945	0.998
HOU	0.923	0.818	0.923	0.997
KCR	0.967	0.870	0.953	0.999
LAA	0.970	0.872	0.953	0.999
LAD	0.969	0.848	0.940	0.998
MIA	0.969	0.864	0.950	0.999
MIL	0.968	0.866	0.951	0.999
MIN	0.969	0.863	0.951	0.999
NY	0.959	0.856	0.943	0.998
NYY	0.950	0.842	0.938	0.997
OAK	0.975	0.878	0.958	1.000
PHI	0.979	0.875	0.955	0.999
PIT	0.949	0.849	0.940	0.998
SDP	0.907	0.808	0.912	0.992
SFG	0.918	0.820	0.920	0.993
SEA	0.954	0.852	0.941	0.998
STL	0.960	0.859	0.946	0.999
TBR	0.963	0.867	0.952	0.999
TEX	0.924	0.829	0.926	0.996
TOR	0.863	0.761	0.875	0.978
WAS	0.956	0.861	0.946	0.998

Table 2.4: Simulation Result: Bayesian Approach with Uninformative Prior

prior) affected narrower confidence intervals of the Bayesian approach in table 2.3.

2.5.2 Frequentist and Bayesian Approach on Real Data

In this section, we apply a frequentist and Bayesian approach on the real data from MLB. We estimate panel data from 1998 to 2013 seasons. From 1998 season, MLB has the current system of 30 teams by adding the Arizona Diamondbacks and the Tampa Bay Devil Rays. The purpose of this estimation is to compare the result from two different approaches. Our interest will span estimated values of coefficients for offensive and defensive inputs and a efficiency estimation.

Coefficient estimation results in table 2.5 shows that estimated values are not strongly affected by the type of approach. However, the efficiency estimation result in table 2.6 contains more complicated results.

Coefficient	Frequentist	Bayesian
Single	0.684	0.684
2B & 3B	0.329	0.346
Homerun	0.322	0.323
Steal	0.034	0.033
Walk	0.256	0.259
Strikeout	-0.029	-0.022
S & 2B &3B allowed	-1.054	-1.039
Homerun allowed	-0.333	-0.327
Walk allowed	-0.292	-0.289
Strikeout made	-0.032	-0.014
Reached on error	-0.070	-0.071

Table 2.5: Coefficient Estimation Result

First, estimated values and the order of estimated efficiencies from the frequentist and Bayesian approach are similar to each other. However, variance of estimated efficiencies are fairly different and the Bayesian approach consistently provides smaller variance values. We already verified this characteristic from the simulation study in section 2.5.1. Figure 2.3 shows difference in variance more clearly.

Result in section 2.5 can be summarized as follows. First, the frequentist and the Bayesian approach do not show critical difference in point estimation. Second, the Bayesian approach provides smaller variance when we estimate managerial ineffi-

Team	Frequentist				Bayesian			
	lb	mean	ub	rank	lb	mean	ub	rank
ARI	0.942	0.964	0.986	28	0.947	0.962	0.978	29
ATL	0.968	0.990	1.000	1	0.985	1.000	1.000	1
BAL	0.944	0.965	0.987	26	0.952	0.967	0.983	25
BOS	0.930	0.952	0.973	30	0.924	0.939	0.955	30
CHC	0.944	0.966	0.988	25	0.951	0.966	0.981	27
CHW	0.952	0.974	0.997	15	0.969	0.985	1.000	14
CIN	0.957	0.979	1.000	8	0.975	0.991	1.000	8
CLE	0.950	0.972	0.994	17	0.962	0.977	0.993	18
COL	0.950	0.972	0.994	19	0.961	0.976	0.991	19
DET	0.945	0.966	0.989	23	0.954	0.970	0.985	23
HOU	0.961	0.983	1.000	6	0.981	0.996	1.000	6
KCR	0.966	0.988	1.000	3	0.985	1.000	1.000	3
LAA	0.962	0.984	1.000	5	0.984	0.999	1.000	5
LAD	0.951	0.973	0.995	16	0.966	0.982	0.997	16
MIA	0.944	0.966	0.988	24	0.952	0.968	0.983	24
MIL	0.953	0.975	0.997	14	0.967	0.982	0.998	15
MIN	0.956	0.978	1.000	10	0.975	0.991	1.000	9
NY	0.956	0.977	1.000	12	0.974	0.989	1.000	11
NYY	0.946	0.968	0.990	22	0.955	0.970	0.986	22
OAK	0.966	0.988	1.000	2	0.985	1.000	1.000	2
PHI	0.949	0.971	0.993	20	0.960	0.975	0.991	20
PIT	0.959	0.981	1.000	7	0.980	0.995	1.000	7
SDP	0.950	0.972	0.994	18	0.963	0.978	0.994	17
SFG	0.953	0.975	0.998	27	0.969	0.985	1.000	26
SEA	0.942	0.964	0.986	13	0.951	0.966	0.982	13
STL	0.963	0.985	1.000	4	0.984	1.000	1.000	4
TBR	0.942	0.964	0.986	29	0.949	0.964	0.980	28
TEX	0.956	0.977	1.000	11	0.972	0.988	1.000	12
TOR	0.957	0.979	1.000	9	0.974	0.990	1.000	10
WAS	0.947	0.969	0.991	21	0.958	0.974	0.989	21

Table 2.6: Efficiency Estimation Result From 1998 to 2013 Seasons

ciency. In section 2.6, we will try to answer questions related to managerial efficiencies in baseball using the Bayesian approach.

2.6 Estimation of Managerial Inefficiency for Baseball Questions

2.6.1 Evolution of Baseball over Time

In this section, we estimate the role of each offensive and defensive inputs in baseball to produce the run ratio of run production over run allowed. To find the contribution from each inputs, we first estimate coefficient values of each offensive

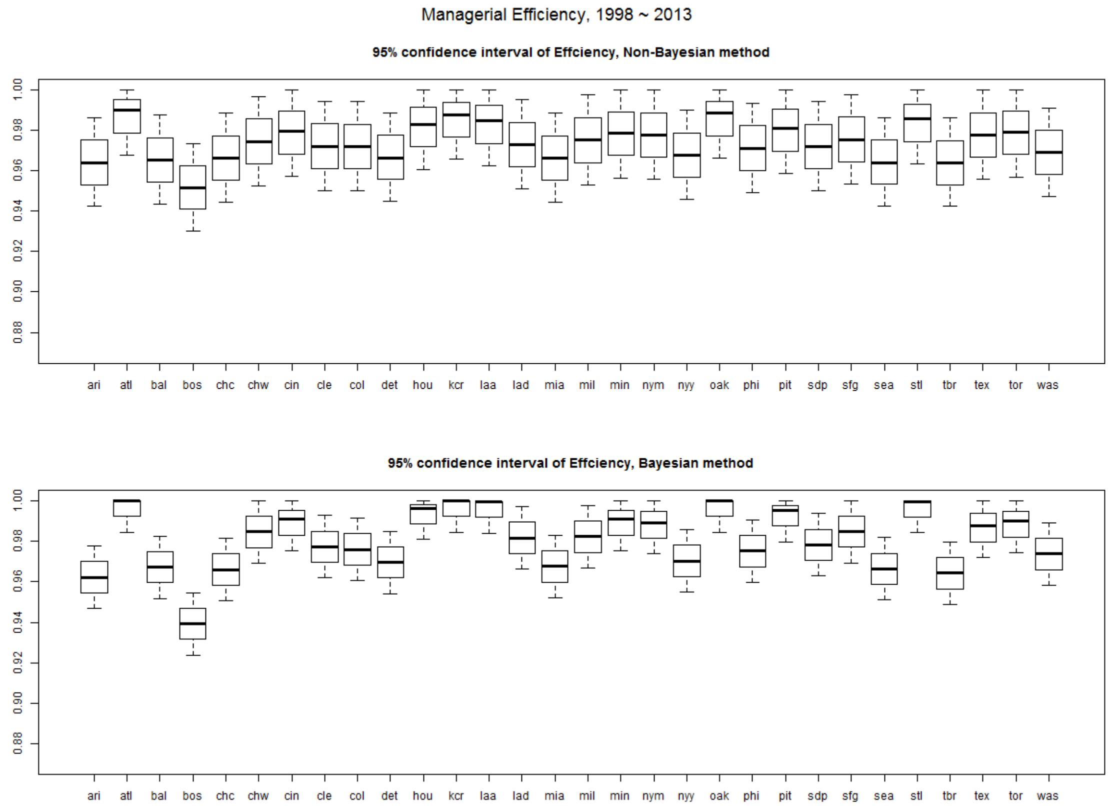

Figure 2.3: Box Plot of Estimated Efficiencies - Data from 1998 to 2013 Season

and defensive inputs. Then we find how managerial efficiencies have changed over the history of baseball due to the value change of each coefficients.

Posterior distributions of coefficient estimation from the old baseball (from 1969 to 1976) and the modern baseball (from 2007 to 2013) are provided in figure 2.4 and figure 2.5.

Comparison of two figures provides meaningful information about different characteristic of two baseball periods. The most important information is that making single hit was more important in old baseball. The coefficient of single is clearly larger than coefficient of single for the modern baseball. It is also verified in table 2.7 which provides coefficients estimation results. On the other hand, figures show that

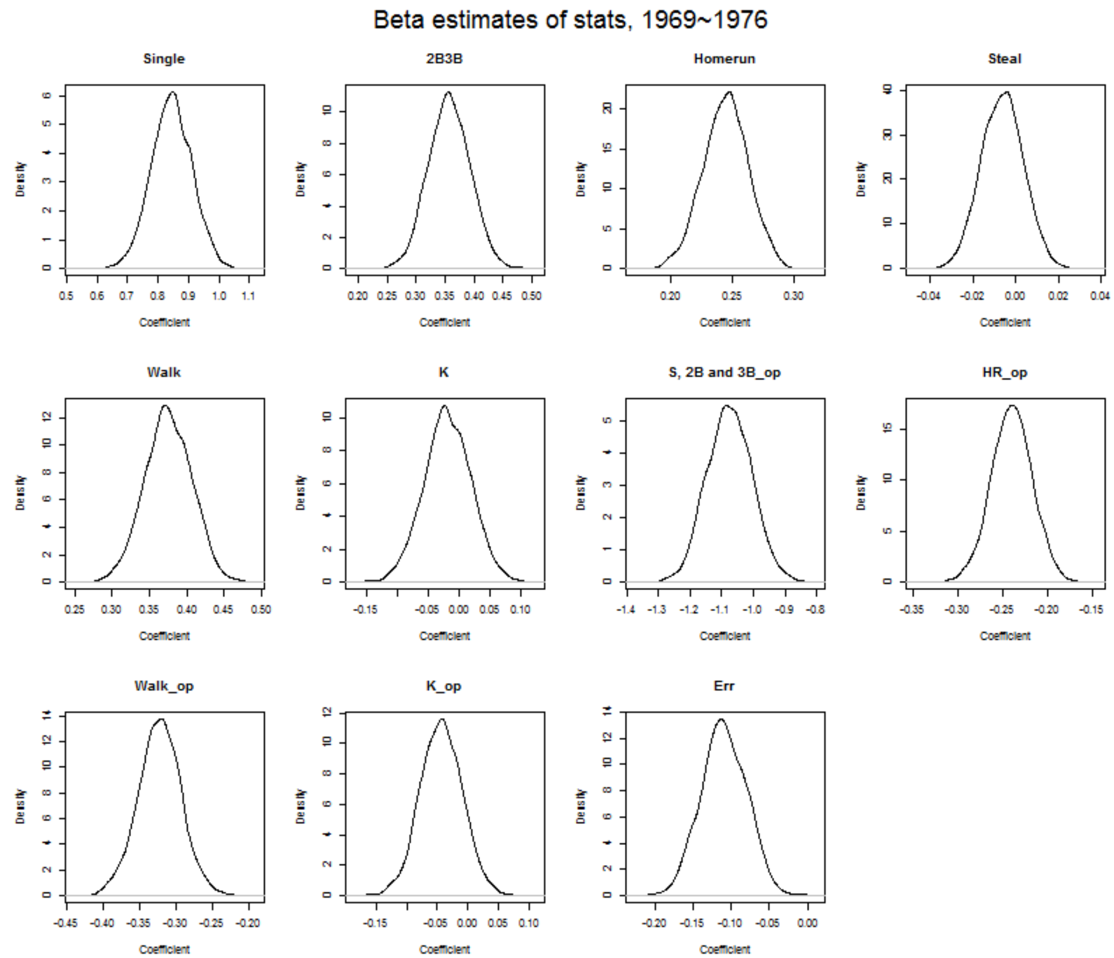

Figure 2.4: Posterior Distribution of Coefficients - Data from 1969 to 1976 Season

the production of extra base hit such as double, triple and homerun has become more important in the modern baseball. One of the noticeable tendency of MLB teams in the modern baseball is filling line up with hitters with more power even though they do not have a good contact skill. These results prove that this kind of strategy makes sense from sabermetric point of view. One of the possible reason behind this result is that it is more difficult to produce serial hits in the modern baseball. This characteristic change of game is also related to the improved fielding ability which was already mentioned in section 2.4 when we showed decreased number of errors in

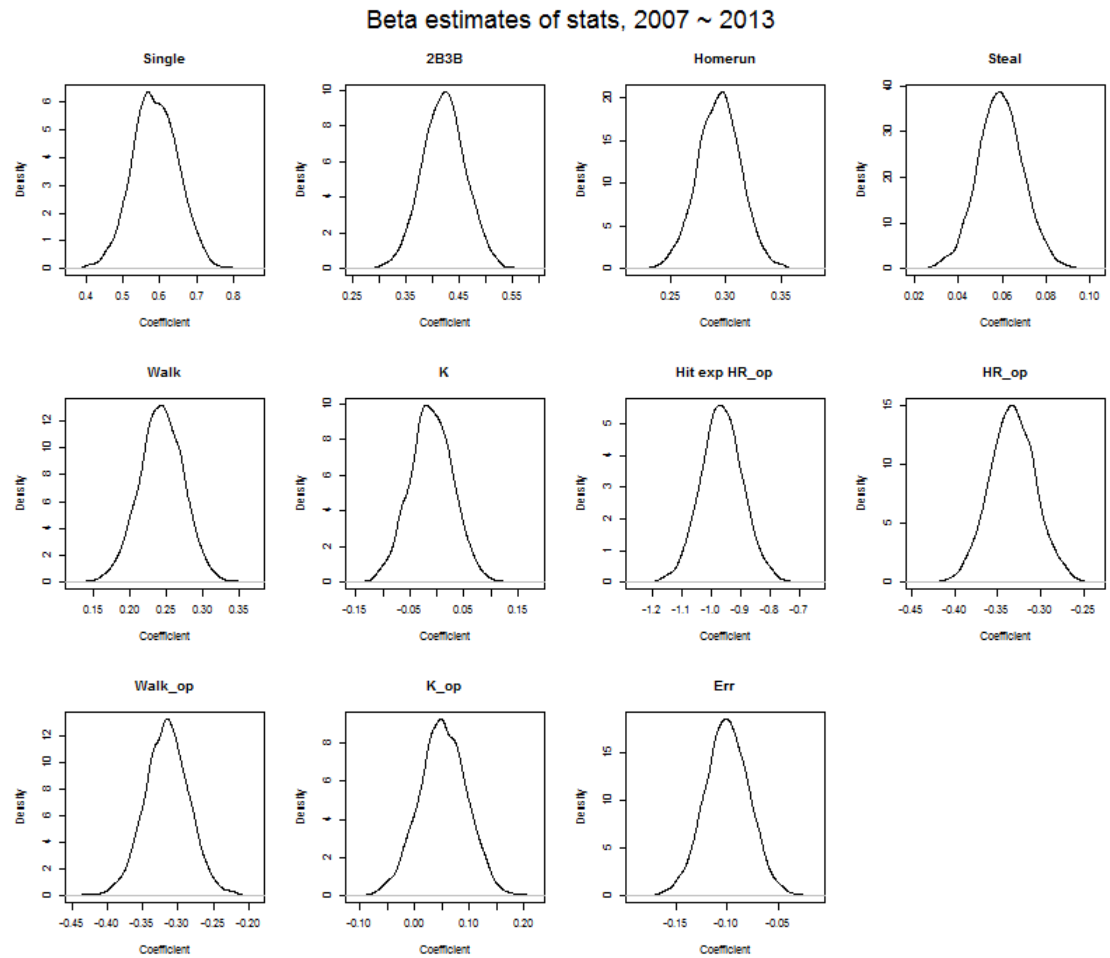

Figure 2.5: Posterior Distribution of Coefficients - Data from 2007 to 2013 Season

large margin.

From table 2.7, we can provide the amount of change in run ratio from each offensive and defensive values. For example, when single increased by 10%, run ratio will increase by 0.0844 ($= 0.1 * 0.844$). From table 2.2 which provides mean offensive variable numbers, 10% increase in modern baseball is about 96 singles. Therefore, we conclude that if a team produce 96 additional singles while other offensive and defensive variables remain at the same level, there will be 10% increase in run ratio.

How can we apply this sabermetric information on the management? We use

	1969-1976		2007-2013	
Coefficient	mean	sd	mean	sd
Single	0.844	0.068	0.588	0.061
2B & 3B	0.357	0.036	0.418	0.040
Homerun	0.245	0.019	0.293	0.020
Steal	-0.006	0.010	0.059	0.011
Walk	0.375	0.030	0.243	0.031
Strikeout	-0.019	0.038	-0.010	0.039
S & 2B &3B allowed	-1.075	0.072	-0.962	0.072
Homerun allowed	-0.240	0.023	-0.334	0.027
Walk allowed	-0.323	0.029	-0.316	0.032
Strikeout made	-0.043	0.034	0.052	0.045
Reached on error	-0.109	0.029	-0.101	0.021

Table 2.7: Coefficient Estimation Result

a free agent transaction for the application example. After the 2010 season, Boston Redsox acquired outfielder Carl Crawford with the annual average salary of 21 million and Washington Nationals made the contract with outfielder Jayson Worth with the annual average salary of 18 million. Both of them are considered to have a good defensive skill so we assume their defensive values are at the same level. According to the depth chart of each team, Carl Crawford will replace Darnell McDonald to play a left fielder and Jayson Worths back up will be Jerry Hairston Jr. in a right fielder position. The additional offensive production over replacement player from Crawford and Worth is in table 2.8. For the numbers 2010 season records are used. The difference in salary over replacement player is also provided

	Boston Redxos			Washington		
Input	Crawford	McDonald	Difference	Worth	Hairston	Difference
Single	184	86	98	164	105	59
2B&3B	43	18	25	48	15	33
HR	19	3	16	27	10	17
SB	47	9	38	13	9	4
BB	46	30	16	82	31	51
SO	104	85	19	147	54	93
Salary	21 mil	0.47 mil	20.53 mil	18 mil	2 mil	16 mil

Table 2.8: Additional Offensive Production of Crawford and Worth

In table 2.8 the additional offensive production provided by Crawford increases the run ratio by 15.24% when the coefficient estimates from 2007 to 2013 season in table 2.7 are used. When we use the same approach, Worth increases run ratio by 13.63%. The Boston Redsox spent 1.35 million to increase 1% higher run ratio, while the Washington Nationals invested 1.17 million for 1% higher run ratio. Therefore, we can conclude that the Nationals made more cost efficient investment. While Worth is still playing for the Nationals, the Boston Red Sox traded Crawford away to the Dodgers during 2013 season. The Red Sox even needed to throw in several young prospects in the deal to make the Dodgers to take the contract with Crawford.

The characteristic change of baseball game strongly affects managerial strategy. Extra hits has added value in modern baseball so players who can produce with extra power take more spots in line up. However, these kind of hitters have the typical shortcomings which comes from their swinging tendency. To give more power to their swinging, they tend to hit the ball to the side of the field from which he bats. It seriously limit their direction of hitting and defensive shift has become very important part of managing. The result is that this new tendency in baseball more frequent use of defensive shift gives more variance to managerial efficiency of modern baseball managers because baseball managers should make additional decision other than traditional ones. As you clearly see in the comparison of figure 2.6 and figure 2.7, managerial efficiency level shows more variance in the modern baseball. One possible reason behind the increased volatility is that teams are more aggressively searching for the manager who maximizes the run difference (run made - run allowed) with limited offensive and defensive inputs and it leads to the shorter contract terms for field managers.

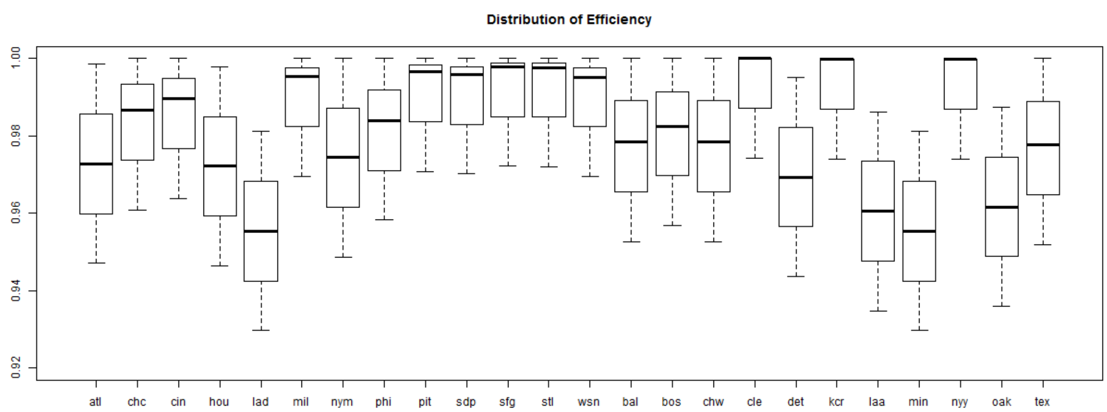

Figure 2.6: Managerial Efficiency - Data from 1969 to 1976 Season

2.6.2 Steroid Era and Moneyball

In the book "Moneyball (2003)" by Michael Lewis, Billy Beane, the general manager of Oakland Athletics, hires Art Howe as a manager who would understand that a field manager is not the boss to implement the ideas of front office with full control. The idea of Beane is anything that increases the offense's chance of making an

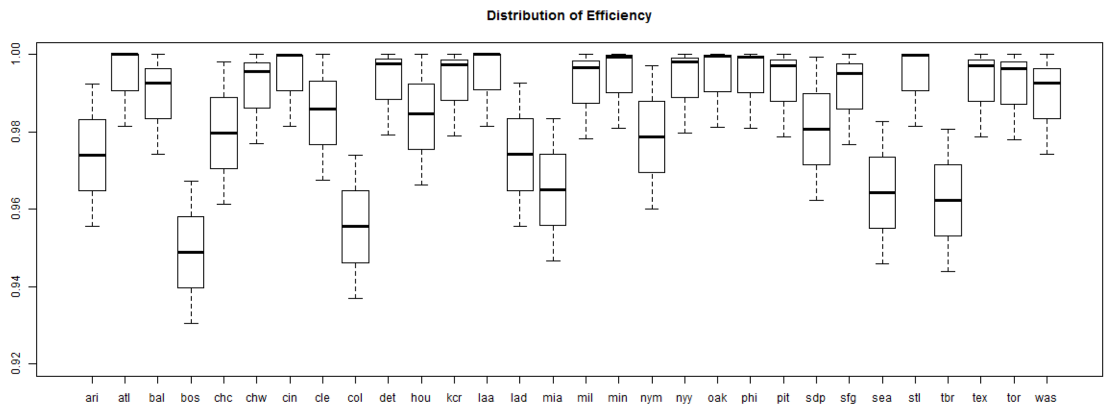

Figure 2.7: Managerial Efficiency - Data from 2007 to 2013 Season

out is bad. So, offensive strategies such as sacrifice bunt, hit and run, and steal are considered to be against efficiency and the manager should keep extremely passive stance to be more effective. Additionally, Beane has the model which argues that an extra point of on-base percentage is worth three times an extra point of slugging percentage. Based on this model, Athletics front office showed an obsession for a players ability to get on base. Art Howe manages Athletics from 1996 to 2002. Figure 2.8 has the distribution of managerial efficiencies over this period.

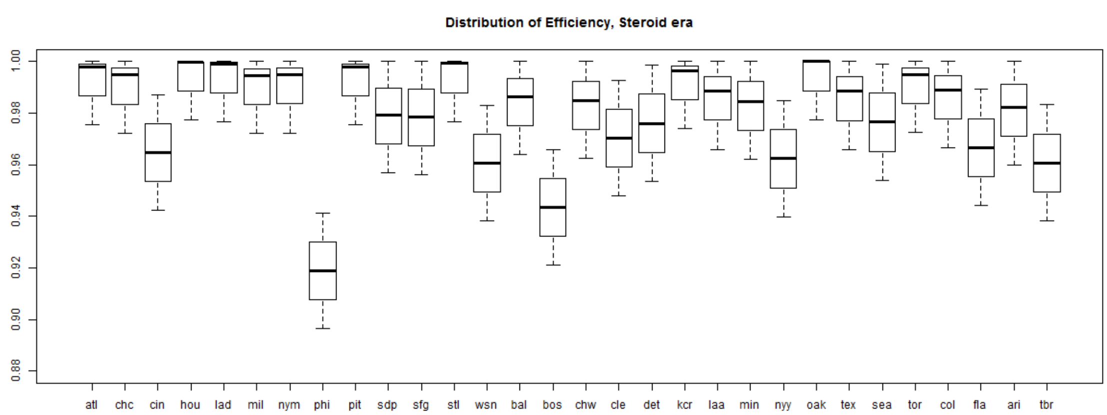

Figure 2.8: Managerial Efficiency - Data from the Steroid Era

Figures 2.8 shows that Athletics had the top notch managerial efficiency in this period. This period is called "The Steroid era" in baseball history. ESPN defines the steroid era as follows: "A period of time in Major League Baseball when a number of players were believed to have used performance-enhancing drugs, resulting in increased offensive output throughout the game". There is no clear guideline on the start or end time of steroid era. Our study set the 1996 season as the start time for the era because many players such as Mark McGwire began producing 50 plus homeruns from 1996 season, which was unprecedented in baseball. Especially, total team extra base hits numbers had a big increase from 1996 season. We also use 2002 season as the end point because the MLB office began test on Performance Enhancement Drug (PED) from the 2003 season. Even after PED test started, there happened intermittent PED scandals, team offensive values began regressing back to the level which was more common before the steroid era from 2003. Data from the steroid era clearly shows different numbers in several categories compared to other periods. Table 2.8 is provided to show the characteristic of the steroid era.

It is easily verified that the number of double and homerun are increased over the steroid era compared to other two periods. It is possible that increase of extra base hits come from the use of the steroid. These differences lead to critical change on the coefficients of offensive inputs provided in table 2.9.

Most characteristic feature of the steroid era is that the coefficient of walk is critically higher in the steroid era. As is mentioned in Moneyball (2003), Beane tried to draft the player with the exceptional skill on getting more number of walks. Typical example is Kevin Youkilis who was praised in Moneyball (2003) as an example of the ideal type of player who can draw more walks. Result in table 2.10 justifies that Beane's draft strategy was right one for the steroid era. The value of walk become lower after the steroid era. Another noticeable finding in table 2.10 is that the effect

Input	1986-1992		Steroid		2007-2013	
	mean	sd	mean	sd	mean	sd
Single	996	62	987	62	958	70
2B & 3B	284	28	322	28	317	30
Homerun	133	34	177	34	160	32
Stolen Base	126	34	107	34	99	30
Walk	531	75	565	75	520	64
Strikeout	926	92	1055	92	1146	125
Error	125	17	114	17	98	16

Table 2.9: Descriptive Statistics of Input and Output Data (Yearly Value)

from error is limited during the steroid era compared to other baseball seasons. Beane also recognized this characteristic of steroid era and used this feature to minimize the team payroll. He made contract with players who could not find their places due to the weak fielding ability even though they have good offensive skills. In Moneyball, Beane mentions Jeremy Giambi as the typical player in this category. Beane argues that he did not concern that Giambi is prone to a fielding error because a fielding error is not a important factor in baseball and Giambi is very good at drawing walks. Beane's strategy worked extremely well in the steroid era like tailor-made suit and table 2.10 partly explains why he was so successful during that era. However, Beane, who has worked as a general manager since 1996, and Oakland Athletics had not been stellar after the steroid era. Figure 2.9 shows that managerial efficiency of Athletics is not staying in the top level after the steroid era. One possible explanation for this change of performance is the decreased impact of walk as is shown in table 2.10.

	1986-1992		Steroid		2003-2013	
Coefficient	mean	sd	mean	sd	mean	sd
Single	0.769	0.075	0.768	0.058	0.588	0.061
2B & 3B	0.272	0.035	0.364	0.036	0.418	0.040
Homerun	0.261	0.020	0.285	0.018	0.293	0.020
Steal	0.047	0.013	0.027	0.010	0.059	0.011
Walk	0.285	0.031	0.314	0.026	0.243	0.031
Strikeout	0.009	0.042	-0.005	0.041	-0.010	0.039
S & 2B &3B allowed	-0.961	0.076	-1.078	0.065	-0.962	0.072
Homerun allowed	-0.249	0.024	-0.318	0.022	-0.334	0.027
Walk allowed	-0.304	0.034	-0.265	0.027	-0.316	0.032
Strikeout made	-0.035	0.040	0.003	0.037	0.052	0.045
Reached on error	-0.110	0.027	-0.035	0.021	-0.101	0.021

Table 2.10: Coefficient Estimation Result

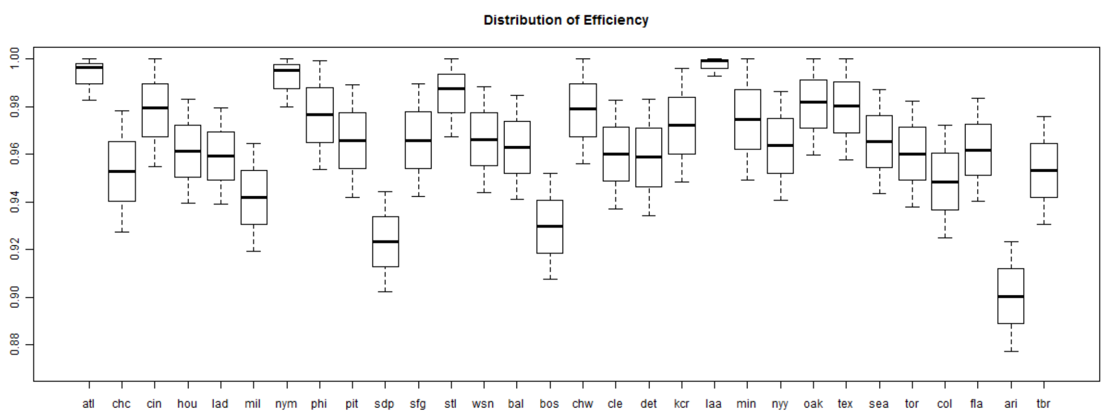

Figure 2.9: Managerial Efficiency - Data from 2003 to 2010 Season

2.6.3 Personal Evolution of Managerial Efficiency: Tony La Russa Case Study

This section compares the managerial efficiency of Tony La Russa, the former manger of St. Louis Cardinals. Tony La Russa started his job as a MLB manager in 1979 for the White Sox. In his career, he has managed the three MLB teams, White Sox, Athletics, and Cardinals and the length of his service is 35 years. By comparing the managerial efficiency in his career over different periods, we show that the efficiency, even from a same manger, can vary over time. Figure 2.10 provides the managerial efficiency over the period from 1979 to 1985 seasons when Tony La Russa managed the White Sox.

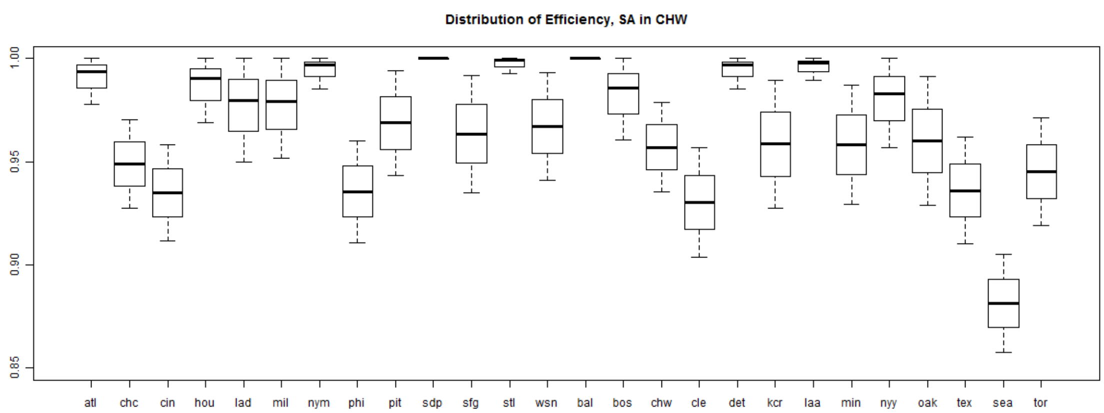

Figure 2.10: Managerial Efficiency - Data from 1979 to 1985 Season

In figure 2.10, the managerial efficiency of the White Sox was placed among the lower class and ranked at 21st place. During 1986 season, he was acquired by Athletics. Figure 2.11 shows the managerial efficiency of Tony La Russa with Athletics.

During this period, Tony La Russa and Athletics provided very good performance and Athletics was ranked at 5th place in managerial efficiency. Before Tony La Russa era, the managerial efficiency of Athletics was ranked at 19th among team. This shows that Tony La Russa played the important role in improving managerial

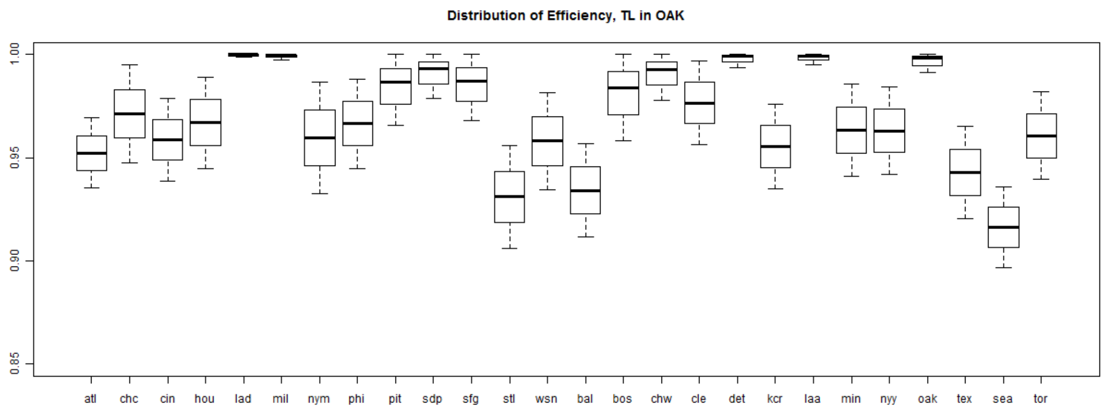

Figure 2.11: Managerial Efficiency - Data from 1987 to 1992 Season

efficiency of the Athletics. During this period, Tony La Russa and Athletics made three appearances on World Series out of 6 years. After 1992, team owner Walter Haas Jr. who paid even highest payroll in baseball went away and new owners of Athletics started to tighten team's payroll. Tony La Russa was then already one of the most acclaimed managers on the field and acquired by the Saint Louis Cardinals. He had been the manager of Cardinals since 1996 and figure 2.12 provides the managerial efficiency from 1996 to 2010 seasons.

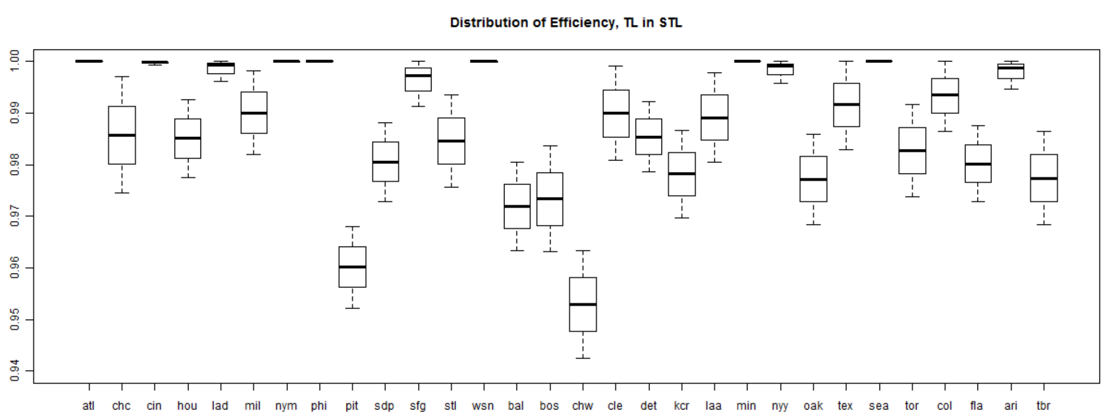

Figure 2.12: Managerial Efficiency - Data from 1996 to 2010 Season

In this period, the managerial efficiency of the Cardinals is ranked at 19th. Even though, the Cardinals has made two playoff berths and won the World Series in 2006 with Cardinals, but the managerial efficiency of the Cardinals was not staying at the top level among the teams. From the longitudinal analysis of managerial efficiencies over three teams, we find that the field manager's contribution to managerial efficiency can vary over time and it is highly affected by the characteristic of the team.

2.6.4 Is John Farrell the Most Efficient Manager in the 2013 Baseball Season?

In this section, we estimate managerial efficiency in 2013 MLB season. The data used for this section is daily game statistics of the 2013 season for 30 MLB teams. We applied the same Bayesian stochastic frontier approach for this analysis. However, we use daily data values for this study while we used yearly data values for the previous studies. We need to modify the dependent variable of the model because we cannot use a run ratio of the game finished like 2:0. Therefore, we use run difference as the dependent variable for this study. It is also impossible to use log value of output variables because some of them has zero values in daily data. Therefore we use input values without taking log on them. Figure 2.13 briefly shows that the Boston Redsox did not have the best managerial efficiency in the 2013 season. This result does not deny the peer review provided by the Boston Globe. However, result shows that New York Yankees's Joe Girardi provided better managing performance in spite of seriously damaged Yankees line up due to serial injuries to their key players such as Derek Jeter, Mark Teixeira, and Mariano Rivera.

Table 2.11 more clearly shows that the managerial efficiency of the Yankees who ranked as the second was managed more efficiently than the Boston Redsox. The interesting part of this result is that personal evolution of managerial efficiency is replayed in this estimation result. Managers who took two bottom spots in the rank

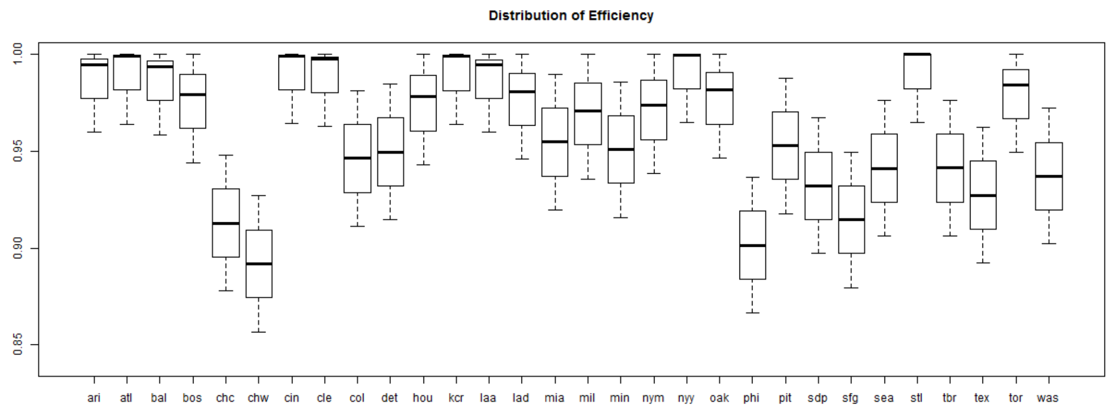

Figure 2.13: Managerial Efficiency - 2013 Season

Team	lb	mean	ub	Rank
ARI	0.960	0.995	1.000	7
ATL	0.964	0.999	1.000	4
BAL	0.959	0.994	1.000	9
BOS	0.944	0.979	1.000	13
CHC	0.878	0.913	0.948	28
CHW	0.857	0.892	0.927	30
CIN	0.964	0.999	1.000	3
CLE	0.963	0.998	1.000	6
COL	0.911	0.946	0.981	21
DET	0.915	0.950	0.985	20
HOU	0.943	0.978	1.000	14
KCR	0.964	0.999	1.000	5
LAA	0.960	0.995	1.000	8
LAD	0.946	0.981	1.000	12
MIA	0.920	0.955	0.990	17
MIL	0.936	0.971	1.000	16
MIN	0.916	0.951	0.986	19
NY	0.939	0.974	1.000	15
NYY	0.965	1.000	1.000	2
OAK	0.946	0.982	1.000	11
PHI	0.866	0.901	0.937	29
PIT	0.918	0.953	0.988	18
SDP	0.897	0.932	0.967	25
SFG	0.906	0.941	0.976	23
SEA	0.879	0.915	0.950	27
STL	0.965	1.000	1.000	1
TBR	0.906	0.941	0.976	22
TEX	0.892	0.927	0.962	26
TOR	0.949	0.984	1.000	10
WAS	0.902	0.937	0.972	24

Table 2.11: Estimated Managerial Efficiency in the 2013 Season

are Robin Ventura for the Chicago White Sox and Ryne Sandberg for the Philadelphia Phillies. Even though they are well-recognized as likely being good managers in the future, team managerial efficiencies with these first time managers stayed at a lower level.

2.7 Concluding Remarks

In this chapter, we have estimated the efficiency of baseball manager using a stochastic frontier model with the Bayesian approach. Data used for estimation is yearly data from 1969 to 2013 season and daily data in 2013 season.

Main finding of our study is that we could obtain the narrower confidence interval when we applied the Bayesian approach to stochastic frontier analysis compared to the interval found by the frequentist approach. This is the case when we use informative prior. However, Bayesian approach does not obtain narrower confidence intervals if uninformative prior is used. To illustrate the comparison, we build confidence interval of estimated efficiencies using both the Bayesian and the frequentist estimator on artificially generated data. Another finding is that the narrower confidence intervals are related to the use of informative prior distribution.

Our study also provides reasonable answers to the questions in baseball. It shows characteristic change of baseball over two different periods from the comparison of the classical baseball and the modern baseball. The specific features of baseball during the limited period which is characterized as the steroid era is analyzed in this paper. From the results, we could find the clue to interesting baseball question: "Why Billy Beane, the hero of sensational Moneyball, is not showing his old performance in steroid era?" The case study of the one of the baseball managing legend "Tony La Russa" shows that he was not born as legendary manager but evolved into the legend as he accumulated experience. The 2013 season estimation provides the evidence that this estimation of efficiency has strong relationship with the work of the front office in real baseball world from the following result. The managers who showed low efficiency were replaced with very high rate by the front office. 50% of managers who were ranked in the bottom 10 managerial efficiency were fired. It means that this estimation model has power to provide tools needed by the front office for decision making.

Here is the last question for this paper. The Boston Redsox was the only 13th in managerial efficiency during the 2013 MLB season. However, the Redsox still could win the championship. At the same time, the Boston Redsox did not have the highest payroll. Then, how the Boston Redsox could beat all the other teams. While studying on managerial efficiency topic, I came to find that there are two kind of production efficiencies in Baseball. The first efficiency is the one we covered in this chapter. This efficiency is mainly decided by the field strategies. This efficiency is affected by how to choose optimal field manager and organize most suitable players for team's strategies. The second and seemingly more important efficiency is to make more offensive and defensive production out of limited payroll cost. The key to answer the reason behind the Boston Red Sox 2013 championship seems to be finding a way to estimate the second efficiency.

REFERENCES

Aguirregabiria, V. and P. Mira, "Dynamic discrete choice structural models: A survey", Journal of Econometrics **156**, 1, 38–67 (2010).

Aigner, D., C. A. Lovell and P. Schmidt, "Formulation and estimation of stochastic frontier production function models", Journal of econometrics **6**, 1, 21–37 (1977).

Battese, G. E. and T. J. Coelli, *Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India* (Springer, 1992).

Battese, G. E. and T. J. Coelli, "A model for technical inefficiency effects in a stochastic frontier production function for panel data", Empirical economics **20**, 2, 325–332 (1995).

Bellman, R., B. Kashef and R. Vasudevan, "Dynamic programming and bicubic spline interpolation", Journal of Mathematical Analysis and Applications **44**, 1, 160–174 (1973).

Bellman, R. E., *Dynamic Programming* (Courier Dover Publications, 2003).

Bera, A. K. and S. C. Sharma, "Estimating production uncertainty in stochastic frontier production function models", Journal of Productivity Analysis **12**, 3, 187–210 (1999).

Ching, A. T., S. Imai, M. Ishihara and N. Jain, "A practitioner's guide to bayesian estimation of discrete choice dynamic programming models", Quantitative Marketing and Economics **10**, 2, 151–196 (2012).

Coelli, T. J., D. S. P. Rao, C. J. O'Donnell and G. E. Battese, *An introduction to efficiency and productivity analysis* (Springer, 2005).

Dawson, P., S. Dobson and B. Gerrard, "Estimating coaching efficiency in professional team sports: Evidence from english association football", Scottish Journal of Political Economy **47**, 4, 399–421 (2000).

Debreu, G., "The coefficient of resource utilization", Econometrica: Journal of the Econometric Society pp. 273–292 (1951).

Diermeier, D., M. Keane and A. Merlo, "A political economy model of congressional careers", The American Economic Review **95**, 1, 347–373 (2005).

Eckstein, Z. and K. I. Wolpin, "Why youths drop out of high school: The impact of preferences, opportunities, and abilities", Econometrica **67**, 6, 1295–1339 (1999).

Farrell, M. J., "The measurement of productive efficiency", Journal of the Royal Statistical Society. Series A (General) pp. 253–290 (1957).

Fizel, J., E. Gustafson and L. Hadley, *Baseball economics: Current research* (Greenwood Publishing Group, 1996).

Flury, T. and N. Shephard, "Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models", Econometric Theory **27**, Special Issue 05, 933–956 (2011).

Fried, H. O., C. K. Lovell and S. S. Schmidt, "Efficiency and productivity", The measurement of productive efficiency and productivity growth pp. 3–91 (2008a).

Fried, H. O., C. K. Lovell and S. S. Schmidt, *The measurement of productive efficiency and productivity growth* (Oxford University Press, 2008b).

Gotz, G. A., P. A. F. (U.S.), R. Corporation and U. States, *A dynamic retention model for Air Force officers theory and estimates* (RAND, Santa Monica, CA, 1984).

Greene, W. H., "The econometric approach to efficiency analysis", The measurement of productive efficiency and productivity growth pp. 92–250 (2008).

Hadley, L., M. Poitras, J. Ruggiero and S. Knowles, "Performance evaluation of national football league teams", Managerial and Decision Economics **21**, 2, 63–70 (2000).

Hochberg, Y. and A. C. Tamhane, *Multiple Comparison Procedures* (John Wiley & Sons, Inc., New York, NY, USA, 1987).

Holloway, G., D. Tomberlin and X. Irz, "Hierarchical analysis of production efficiency in a coastal trawl fishery", Applications of Simulation Methods in Environmental and Resource Economics pp. 159–185 (2005).

Horrace, W. C. and P. Schmidt, "Confidence statements for efficiency estimates from stochastic frontier models", Journal of Productivity Analysis **7**, 2-3, 257–282 (1996).

Horrace, W. C. and P. Schmidt, "Multiple comparisons with the best, with economic applications", Journal of Applied Econometrics **15**, 1, 1–26 (2000).

Houser, D., "Bayesian analysis of a dynamic stochastic model of labor supply and saving", Journal of Econometrics **113**, 2, 289–335 (2003).

Hsu, J., *Multiple comparisons: theory and methods* (CRC Press, 1996).

Imai, S., N. Jain and A. Ching, "Bayesian estimation of dynamic discrete choice models", Econometrica **77**, 1865–1899 (2009).

Ishihara, M., *Dynamic Demand for New and Used Durable Goods without Physical Depreciation*, Ph.D., University of Toronto (Canada), Canada (2011).

James, B., *1981 BASEBALL ABSTRACT The 5th Annual Edition* (Bill James, 1981).

Keane, M. P., P. E. Todd and K. I. Wolpin, "Chapter 4 - the structural estimation of behavioral models: Discrete choice dynamic programming methods and applications", in "Handbook of Labor Economics", edited by Orley Ashenfelter and David Card, vol. Volume 4, Part A, pp. 331–461 (Elsevier, 2011).

Keane, M. P. and K. I. Wolpin, "The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence", The Review of Economics and Statistics **76**, 648–672 (1994).

Keane, M. P. and K. I. Wolpin, "The career decisions of young men", The Journal of Political Economy **105**, 3, 473–522 (1997).

Koop, G., J. Osiewalski and M. F. Steel, "Bayesian efficiency analysis through individual effects: Hospital cost frontiers", Journal of Econometrics **76**, 1, 77–105 (1997).

Koop, G., D. J. Poirier and J. L. Tobias, *Bayesian econometric methods*, vol. 7 (Cambridge University Press, 2007).

Lancaster, T., "Exact structural inference in optimal job-search models", Journal of Business & Economic Statistics **15**, 2, 165–179 (1997).

Lee, Y. H., "Team sports efficiency estimation and stochastic frontier models", Handbook ofsports economic research, edited by Iohn Fizel. Armonk, NY: ME. Sharpe pp. 209–220 (2006).

Lewis, M., *Moneyball: The Art of Winning an Unfair Game* (W. W. Norton & Company, 2004).

McKinnon, K. I. M., "Convergence of the Nelder-Mead simplex method to a nonstationary point", SIAM Journal on Optimization **9**, 1, 11 (1998).

Meeusen, W. and J. Van den Broeck, "Efficiency estimation from cobb-douglas production functions with composed error", International economic review pp. 435–444 (1977a).

Meeusen, W. and J. Van den Broeck, "Technical efficiency and dimension of the firm: Some results on the use of frontier production functions", Empirical economics **2**, 2, 109–122 (1977b).

Miller, R. A., "Job matching and occupational choice", Journal of Political Economy **92**, 6, 1086–1120 (1984).

Nelder, J. A. and R. Mead, "A simplex method for function minimization", The Computer Journal **7**, 4, 308–313 (1965).

Norets, A., *Bayesian inference in dynamic discrete choice models*, Ph.D., The University of Iowa (2007).

Pakes, A., "Patents as options: Some estimates of the value of holding european patent stocks", Econometrica **54**, 4, 755–784 (1986).

Porter, P. K. and G. W. Scully, "Measuring managerial efficiency: the case of baseball", Southern Economic Journal pp. 642–650 (1982).

Powell, M. J., *A direct search optimization method that models the objective and constraint functions by linear interpolation* (Springer, 1994).

Rimler, M. S., S. Song and T. Y. David, “Estimating production efficiency in mens NCAA college basketball: A bayesian approach”, Journal of Sports Economics **11**, 3, 287–315 (2010).

Rust, J., “Optimal replacement of GMC bus engines: An empirical model of harold zurcher”, Econometrica **55**, 5, 999–1033 (1987).

Rust, J. and C. Phelan, “How social security and medicare affect retirement behavior in a world of incomplete markets”, Econometrica **65**, 4, 781–831 (1997).

Schmidt, P. and R. C. Sickles, “Production frontiers and panel data”, Journal of Business & Economic Statistics **2**, 4, 367–374 (1984).

Schorfheide, F., “Loss function-based evaluation of DSGE models”, Journal of Applied Econometrics **15**, 6, 645–670 (2000).

Stewart, G. W., *Afternotes on numerical analysis* (Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1996).

Su, C.-L. and K. L. Judd, “Constrained optimization approaches to estimation of structural models”, Econometrica **80**, 5, 2213–2230 (2012).

Train, K., *Discrete choice methods with simulation* (Cambridge university press, 2009).

van der Klaauw, W. and K. I. Wolpin, “Social security and the retirement and savings behavior of low-income households”, Journal of Econometrics **145**, 1-2, 21–42 (2008).

Wolpin, K. I., “An estimable dynamic stochastic model of fertility and child mortality”, Journal of Political Economy **92**, 5, 852–874 (1984).

Zhou, Y., “Failure to launch in two-sided markets: A study of the U.S. video game market”, SSRN Scholarly Paper ID 2163948, Social Science Research Network, Rochester, NY (2013).