Article

Bayesian Approach for Estimating the Probability of Cartel Penalization under the Leniency Program

Jihyun Park [1] [ID]**, Juhyun Lee [1] and Suneung Ahn [2,***

[1] Department of Industrial and Management Engineering, Hanyang University, Seoul 04763, Korea;
    pjh3226@hanyang.ac.kr (J.P.); ljh812@hanyang.ac.kr (J.L.)
[2] Department of Industrial and Management Engineering, Hanyang University ERICA, Ansan 15588, Korea
* Correspondence: sunahn@hanyang.ac.kr; Tel.: +82-031-400-5267

check for updates

Abstract: Cartels cause tremendous damage to the market economy and disadvantage consumers by creating higher prices and lower-quality goods; moreover, they are difficult to detect. We need to prevent them through scientific analysis, which includes the determination of an indicator to explain antitrust enforcement. In particular, the probability of cartel penalization is a useful indicator for evaluating competition enforcement. This study estimates the probability of cartel penalization using a Bayesian approach. In the empirical study, the probability of cartel penalization is estimated by a Bayesian approach from the cartel data of the Department of Justice in the United States between 1970 and 2009. The probability of cartel penalization is seen as sensitive to changes in competition law, and the results have implications for market efficiency and the antitrust authority's efforts against cartel formation and demise. The result of policy simulation shows the effectiveness of the leniency program. Antitrust enforcement is evaluated from the estimation results, and can therefore be improved.

1. Introduction

Cartels cause tremendous damage to perfect competition markets and consumers by effectually applying upward pressure on prices and downward pressure on quality; moreover, cartels are difficult to detect because of their tacit nature. In this way, cartels mitigate against perfect competition under which consumers are offered the best goods and services at the lowest possible prices. Antitrust authorities have continuously sought to maintain a free-market system against cartels, but with only partial and limited success.

In previous research, the probability of cartel detection was a key indicator for measuring the effectiveness of antitrust policies. Detection is the state in which unobserved cartels are caught by the antitrust authority. After introducing a leniency program as a new antitrust policy, both the number of cartel investigations and the probability of cartel detection increase. The higher the probability of cartel detection, the greater the expected penalties, and therefore, the likelihood of cartel formation will decrease. On this principle, it is possible to measure the deterrence effect according to the change in antitrust policy. This study uses the probability of cartel penalization as a key indicator.

The Markov transition process and the birth and death process model were widely used. Bryant and Eckard [1] constructed the birth and death process model to empirically analyze cartel data provided by the United States (US) Department of Justice, and estimated the probability of cartel detection in the US in 1961–1988 as between 13–17%. Using the same method, Combe et al. [2] estimated European Commission (EC) cartel detection probabilities of 12.9–13.2% for 1969–2007. When the birth and death model has two states of competition and collusion, the lifetimes and inter-arrival

times between the births of cartels were independent and had exponential distributions with means of λ^{-1} and θ^{-1}. The number of cartels at a particular time *t* follows a Poisson distribution with a mean of $\theta_T = (\theta/\lambda)\{1 - e^{-\lambda T}\}^2$. Both Bryant and Eckard [1] and Combe et al. [2] assumed that every cartel would eventually be caught and prosecuted. However, this assumption is not realistic, because some cases are not penalized, despite having been detected.

Further, Bryant and Eckard [1] and Combe et al. [2] do not take account of the unobservable cartel population. J. E. Harrington and Chang [3] sought to estimate the unobservable population by developing the birth and death model from that noted above. They concluded that cartel duration could be a good indicator of whether new competition law had a significant cartel-dissolution effect. Using Harrington and Chang [3]'s model, Zhou [4] analyzed the EC cartel data for 1985–2012, and concluded that the EU's new leniency program in 2002 had the effect of deterring cartels.

In the research of Bryant and Eckard [1], Combe et al. [2], Harrington and Chang [3], and Zhou [4], the probability of cartel detection—as derived from cartel duration—entailed the determination of the time-average probability from continuous variables. On the other hand, there is research indicating that the probability of cartel detection represents the ensemble-average probability obtained from discrete variables such as caseloads. The time-average probability is the average of a stochastic process that is obtained by selecting a sample path randomly, and taking the average of a period in a particular state on that sample path over the observation period. The ensemble-average probability is that mean of a quantity at time *t* that is estimated by the average of the ensemble of possible states of total sample paths in stochastic process theory [5,6].

Miller [7] formulated a cartel behavior model using the Markov process, and used the number of cartel cases as discrete variables. The model assumed that the cartel transition process is in a non-absorbing and first-order Markov chain in contrast with previous Markov models, and showed the change of the number of cartel detections before and after a leniency program. He concluded that the introduction of this leniency program in 1993 increased the detection and deterrence capabilities of competition enforcement. The previous research above [1–4,7] used Markov process models; this research had two notable points.

First, the duration of cartels and inter-arrival times between cartels follow exponential distributions. Verifying this assumption requires a hypothesis testing of the null hypothesis that "the distribution is exponential". The cumulative distribution function $\hat{F}(x)$ of durations and inter-arrival times is given by:

$$\hat{F}(x) = \frac{number\ of\ observations \leq x}{total\ number\ of\ observations}$$

Under the exponential distribution, $log(1 - \hat{F}(x))$ should be approximately linear in *x*. The result of these previous works indicates that the cartels' duration and inter-arrival times between cartels follow the exponential distribution; therefore, models can be applied to the Markov process [7].

Second, this research assumed that the cartel process was stationary for adopting the Markov process, and that the values could be analyzed when the cartel process attained a steady state; this is also unrealistic. In the research of Bryant and Eckard [1] and Combe et al. [2], the probability is the resultant value when it reaches a steady state. This kind of probability is called a time-independent probability. Otherwise, the form of estimators needs to be a time-dependent rather than time-independent, because the purpose of estimating the probability of cartel detection is evaluating the effects of various competition policies [8]. Thus, Hinloopen [8]'s research was an theoretical literature review for analyzing a subgame of collusion.

A new mathematical methodology has emerged recently in the form of a non-Markov process. Ormosi [9] estimated the annual probability of cartel detection by employing capture–recapture methods based on EC information in the period between 1981–2001. The methods of Ormosi [9], which are frequently used in ecology, reflect that transition parameters are not steady state, and that detection and survival rates are time-independent. However, there are two unreasonable assumptions. First, capture–recapture methods assume that temporary migrations between the two

states (compete–collude) do not exist; thus, they are regarded as robust design methods. The antitrust policy tends to vary broadly according to governmental power or social issues. Second, Ormosi [9] deduced a result from moving average methods, specifically in the moving average of three or five years. If the probability is used on the basis of a single year, the accuracy of the probability may decrease due to data insufficiency. The industry reacts immediately to changes in competition law; therefore, the probability needs to be estimated for the smallest unit of time.

This paper seeks to estimate the probability of cartel penalization using a Bayesian approach and evaluate the impact of the leniency program as an antitrust policy. This study uses the conjugate family of the beta-binomial in that the cartel occurs in binomial events. The posterior mean of the beta distribution is the probability of cartel penalization in a year. This shows the trend of the probability of cartel penalization, and can then improve the antitrust policy using the measured impact of the leniency program. In this light, the present research makes three contributions.

First, this paper estimates the probability of cartel penalization for analyzing cartels in contrast to the probability of cartel detection as treated in previous research. The probability of cartel detection means the probability that unobserved cartels will be investigated, prosecuted, and penalized. However, the probability of cartel penalization means the penalized likelihood of investigated cartels through sufficient investigation. This is used as an indicator with which to evaluate the impact of the leniency program and the capability of antitrust authorities.

Second, the methodology of this paper makes up for the weak points of previous probability estimation methods. Previous methods have many unrealistic assumptions such as the analyzed cases being eventually caught/detected cases, the time-average probability, etc. We can improve on these assumptions by estimating the time-dependent ensemble-average probability based on the discrete data of caseloads, which is more practical than the time-average probability for the sensitive estimation of probability.

Third, this study shows that the Bayesian approach could play a practical role in modeling and analyzing the cartel situation. Although the Markov process model, which was commonly used in previous research, is an essential consideration "in steady-state probability", it is difficult to assume "in steady-state probability", because cartel cases continuously vary over time. The probability of cartel penalization estimated using the Bayesian approach does not need to consider "steady-state probability". The Bayesian approach for estimating probability can contain significant uncertainty, but has good predictive performance in itself [10]. The bias between the estimation probability and the actual value could be solved from the update procedure of the Bayesian approach. Therefore, we present reliable results using the non-informative prior and conjugate prior distribution when prior information is insufficient.

The paper is organized as follows: Section 2 defines the penalization probability and Bayesian probabilistic model; Section 3 presents an empirical study based on US cartel data; and Section 4 draws conclusions.

2. Bayesian Probabilistic Model

When faced with suspected cartel cases, a competition authority carries out an initial investigation to determine whether there are sufficient grounds to prosecute. Prosecuted cartels are penalized in the form of fines through a trial. Eventually, the three states of cartel cases are investigation, prosecution, and penalization [11]. The estimated probability of this study is based on investigation and penalization states. The probability of cartel penalization (ρ_t) is described as the proportion of the numbers of penalized cases to investigated cases for year t ($t = 1, 2, \cdots$).

The estimation of the penalization probability using the Bayesian approach involves two assumptions. First, the unit of case is an industry. Accordingly, the research of Bryant and Eckard [1] and Miller [7] is based on the analysis unit of the industry. Bos and Harrington [12] argued that firm-based analysis is more realistic; nonetheless, this study was analyzed based on the analysis unit of industry for easy analysis. In practice, cartels can participate in all firms of an industry. Second,

a cartel only arises as one event during a year. Every cartel is transferred to the competition as a result of punishment by the authorities. This is called the "Grim trigger strategy" [13,14]. Thereafter, if some player deviates from the cartel, the game cannot be colluded indefinitely.

This study constructed a Bayesian probabilistic model to estimate the probability of cartel penalization. The probability of cartel penalization is the posterior mean calculated from the posterior distribution. Inferring a posterior distribution requires determining the proper prior distribution. A Bayesian probabilistic model is comprised of a prior distribution to induce a posterior distribution, hyperparameters, and a likelihood function. A Bayesian sequential analysis of the dynamic Bayesian model can be used to reflect the latest trends of time-series data [15,16].

Two things should be considered to induce a posterior distribution from a prior distribution: the likelihood function and the parameters in the prior distribution, which are known as hyperparameters [17]. The natural conjugate priors are generally recommended in the Bayesian approach, because its functional form is similar to the likelihood distribution [18,19]. Therefore, we have to obtain the appropriate likelihood function to adopt the notion of natural conjugacy. Consider the following notations for the Bayesian probabilistic model.

ρ_t: The probability of cartel penalization cases in year t

n_t: The number of cartel investigation cases by the competition authority in year t

k_t: The number of cartel penalization cases by the competition authority in year t

When the investigated industry participating in a cartel is n, Figure 1 shows a binomial tree to demonstrate the process of cartel formation and demise in year t.
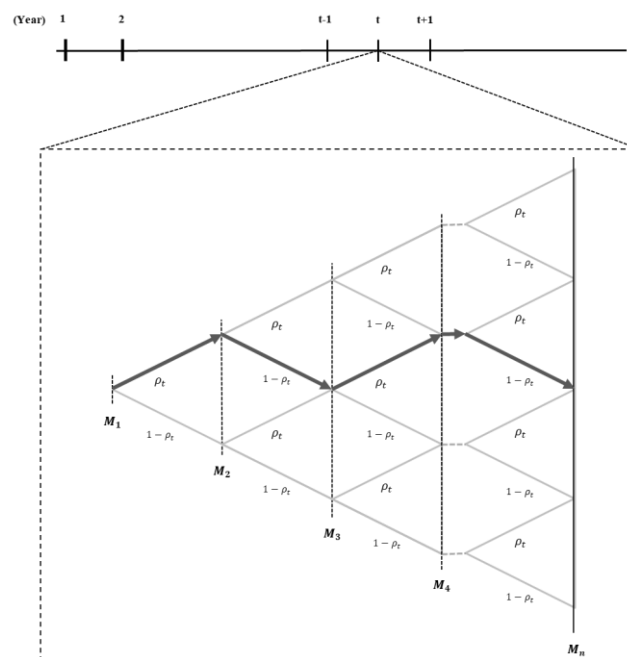

Figure 1. Estimating the probability of cartel penalization through a binomial tree.

In Figure 1, M_1, M_2, $\cdots$, M_n is the industry of investigated cartels in year t. Arrows in the path show whether the investigated cartels were finally penalized. When a route contains an arrow pointing to the right, this cartel will be finally penalized; otherwise, it is not penalized. For example, the industry M_2 is in the left direction; this means that industry M_2 will be not finally penalized as the probability ρ_t. This study wants to infer the probability of industry $n + 1$ penalization in path G; this probability is estimating the likelihood function based on the data from industry 1 to n, and the prior distribution

while inferring a posterior distribution from the Bayesian approach [13]. The expectation of a posterior distribution indicates the probability of cartel penalization.

2.1. Likelihood Function and Prior Distribution

The variable n_t is the number of cartel cases investigated in year t, and each case follows the Bernoulli process with an independent and identical distribution. Therefore, the Bernoulli random variable X_i with one case shown is given by:

$$X_i = \begin{cases} 1 & \text{if penalizing with probability } \rho_t \\ 0 & \text{if non} - \text{penalizing with probability } 1 - \rho_t, \end{cases}$$

where i is the number of cartel firm ($i = 1, \cdots, n_t$) and $0 < \rho_t < 1$. The probability mass function of the random variable, which is known as the Bernoulli probability, is given by:

$$f(x_i|\rho_t) = \rho_t^{x_i}(1 - \rho_t)^{1-x_i}. \tag{1}$$

Once the number of cases n_t is investigated, and k_t is penalized in year t, the joint probability mass function of cartel cases is given by:

$$\begin{aligned} L(\rho_t|x_1, \cdots, x_{n_t}) = f(x_1, \cdots, x_{n_t}|\rho_t) & = \prod_{i=1}^{n_t} f(x_i|\rho_t) \\ & = \prod_{i=1}^{n_t} \rho_t^{x_i}(1 - \rho_t)^{1-x_i} \\ & = \rho_t^{\sum x_i}(1 - \rho_t)^{n_t - \sum x_i} \\ & = \rho_t^{k_t}(1 - \rho_t)^{n_t - k_t}. \end{aligned} \tag{2}$$

The probability of cartel penalization has a value between 0 and 1. In Equation (2), $f(\rho_t)$ is a binomial form as the prior distribution, because there are only two final states of a cartel: whether it has been penalized or not. Thus, we use the beta distribution as a prior distribution based on the natural conjugacy [17,20]. The prior distribution $f(\rho_t)$ is the beta distribution with hyperparameters α and β; thus, the probability density function is given by:

$$f(\rho_t) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \rho_t^{\alpha-1} \cdot (1 - \rho_t)^{\beta-1}, \tag{3}$$

where $\alpha > 0$ and $\beta > 0$ are the hyperparameters. The function $\Gamma(\cdot)$ is a gamma function, which is defined as:

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx. \tag{4}$$

Note that when α is a positive integer, $\Gamma(\alpha) = (\alpha - 1)!$.

2.2. Bayesian Estimation

In the Bayesian approach, the posterior distribution is given by:

$$f(\rho_t|x_1, \cdots, x_{n_t}) = \frac{f(x_1, \cdots, x_{n_t}, \rho_t)}{f(x_1, \cdots, x_{n_t})}. \tag{5}$$

The joint probability distribution $f(x_1, \cdots, x_{n_t}, \rho_t)$ in Equation (5), which reflects the multiplicative laws of probability in Equations (2) and (3), is:

$$\begin{aligned} f(x_1, \cdots, x_{n_t}, \rho_t) & = f(x_1, \cdots, x_{n_t}|\rho_t) \cdot f(\rho_t) \\ & = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \rho_t^{k_t+\alpha-1} \cdot (1 - \rho_t)^{n_t-k_t+\beta-1}. \end{aligned} \tag{6}$$

The marginal probability distribution $f(x_1, \cdots, x_{n_t})$, which is calculated by the law of total probability, is given by:

$$
\begin{aligned}
f(x_1, \cdots, x_{n_t}) &= \int_0^1 f(x_1, \cdots, x_{n_t}, \rho_t) d\rho_t \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \rho_t^{k_t+\alpha-1} \cdot (1-\rho_t)^{n_t-k_t+\beta-1} d\rho_t \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \int_0^1 \rho_t^{k_t+\alpha-1} \cdot (1-\rho_t)^{n_t-k_t+\beta-1} d\rho_t,
\end{aligned}
\tag{7}
$$

where $\int_0^1 \rho_t^{k_t+\alpha-1} \cdot (1-\rho_t)^{n_t-k_t+\beta-1} d\rho_t = \frac{\Gamma(\alpha+k_t) \cdot \Gamma(n_t-k_t+\beta)}{\Gamma(\alpha+\beta+n_t)}$.

Suppose that the initial probability (ρ_t) is 0.5 meaning whether the investigated or the non-investigated case for eliminating the dependence on the prior information. The hyperparameters α and β are 1 as a non-informative prior. Therefore, the posterior distribution is a beta distribution with the parameters $\alpha + k_t$ and $\beta + n_t - k_t$. The posterior distribution of Equation (5) is represented by:

$$
\begin{aligned}
f(\rho_t | x_1, \cdots, x_{n_t}) &= \frac{f\left(x_1, \cdots, x_{n_t}, \rho_t\right)}{f\left(x_1, \cdots, x_{n_t}\right)} \\
&= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \rho_t^{k_t+\alpha-1} \cdot (1-\rho_t)^{n_t-k_t+\beta-1}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+k_t) \cdot \Gamma(n_t-k_t+\beta)}{\Gamma(\alpha+\beta+n_t)}} \\
&= \frac{\Gamma(\alpha+\beta+n_t)}{\Gamma(\alpha+k_t)\Gamma(\beta+n_t-k_t)} \cdot \rho_t^{k_t+\alpha-1} \cdot (1-\rho_t)^{n_t-k_t+\beta-1}.
\end{aligned}
\tag{8}
$$

The posterior mean $E[\rho_t | x_1, \cdots, x_{n_t}]$ from Equation (8) is:

$$
\begin{aligned}
E[\rho_t | x_1, \cdots, x_{n_t}] &= \int_0^1 \rho_t \cdot f(\rho_t | x_1, \cdots, x_{n_t}) \, d\rho_t \\
&= \int_0^1 \rho_t \frac{\Gamma(\alpha+\beta+n_t)}{\Gamma(\alpha+k_t)\Gamma(\beta+n_t-k_t)} \cdot \rho_t^{k_t+\alpha-1} \cdot (1-\rho_t)^{n_t-k_t+\beta-1} d\rho_t \\
&= \frac{\Gamma(\alpha+\beta+n_t)}{\Gamma(\alpha+k_t)\Gamma(\beta+n_t-k_t)} \cdot \int_0^1 \rho_t^{(k_t+\alpha+1)-1} \cdot (1-\rho_t)^{(n_t-k_t+\beta)-1} d\rho_t \\
&= \frac{\Gamma(\alpha+\beta+n_t)}{\Gamma(\alpha+k_t)\Gamma(\beta+n_t-k_t)} \cdot \frac{\Gamma(\alpha+k_t+1) \cdot \Gamma(n_t-k_t+\beta)}{\Gamma(\alpha+\beta+n_t+1)} \\
&= \frac{\alpha+k_t}{\alpha+\beta+n_t}.
\end{aligned}
\tag{9}
$$

3. Empirical Study

3.1. Data

This study uses data from the Workload statistics published by the Antitrust Division of the Department of Justice (DOJ) for the period between 1970–2009 [21]. The information is shown in Table 1. It contains the annual statistics of penalized cases and investigated cases by the criminal enforcement and civil enforcement of district courts, with respect to the laws of Sherman §1-Restraint of Trade, Sherman §2-Monopoly, and Clayton §7-Mergers. The antitrust division prosecutes in the form of criminal enforcement cases if the cartels, which are known as "hardcore cartels," are determined by preliminary examination to have an especially injurious impact on the industry; otherwise, it prosecutes in the form of civil enforcement cases. This study does not consider the appellate cases and the cases of contemporary criminal–civil enforcement at the same time, due to a few of applicable cases.

Table 1. Statistic of cartel data by the Department of Justice (DOJ) between 1970 and 2009.

Years	1970–1979	1980–1989	1990–1999	2000–2009
Number of investigation cases	4155	2505	3145	1881
Total penalization cases	544	859	780	535
(Civil cases)	351	210	161	129
(Criminal cases)	193	649	619	406
Total fines ($thousand)	47,712	187,548	1,612,993	4,222,407

3.2. Time-Series Analysis

Prior to the model application, a time-series analysis was implemented to eliminate spurious relations. This study, alternatively, employed the augmented Dickey–Fuller (ADF) unit root test to confirm the stability of the time-series data (details are provided in Appendix A).

If the result shows that the time-series data is unstable, the difference stationary process is needed. The representative method for stabilizing time-series data is order difference or log order difference. However, using order difference, it is possible that the meaning of original data will be lost, leading to different conclusions in the economy [22]. Economic variables such as price, currency, and stock index cannot be used to verify the stability of time-series data, because they are commonly non-stationary data [23].

3.3. Results

The empirical study, using the model defined in Section 2, drew an annual beta distribution for the probability of cartel penalization. The results are summarized in Table 2, and Figure 2 illustrates the distribution for every year.

Figure 2 shows that the probability distributions tend to increase over time. Beta distributions converge on a specific range with Bayesian updating [17]. Indeed, the result shows the convergence of the present distribution on the specific range at around 0.22. We were able to calculate the posterior mean by Equation (9). Figure 3, accordingly, illustrates the annual expected probability of cartel penalization.

Table 2. The probability of cartel penalization through Bayesian sequential analysis.

Years (t)	Investigation Cases (n_t)	Penalization Cases (k_t)	Prior α	Prior β	Posterior α	Posterior β	The Expected Probability of Cartel Penalization (æ_t)
1970 (1)	473	53	1	1	54	421	
1971 (2)	593	51	54	421	105	963	0.11368
1972 (3)	465	53	105	963	158	1375	0.09831
1973 (4)	538	61	158	1375	219	1852	0.10307
1974 (5)	338	57	219	1852	276	2133	0.10575
1975 (6)	381	29	276	2133	305	2485	0.11457
1976 (7)	374	64	305	2485	369	2795	0.10932
1977 (8)	484	46	369	2795	415	3233	0.11662
1978 (9)	290	68	415	3233	483	3455	0.11376
1979 (10)	407	62	483	3455	545	3800	0.12265
1980 (11)	377	89	545	3800	634	4088	0.12543
1981 (12)	255	93	634	4088	727	4250	0.13427
1982 (13)	262	109	727	4250	836	4403	0.14607
1983 (14)	245	99	836	4403	935	4549	0.15957
1984 (15)	257	80	935	4549	1015	4726	0.17050
1985 (16)	254	77	1015	4726	1092	4903	0.17680
1986 (17)	307	98	1092	4903	1190	5112	0.18215
1987 (18)	270	27	1190	5112	1217	5355	0.18883
1988 (19)	216	55	1217	5355	1272	5516	0.18518
1989 (20)	220	132	1272	5516	1404	5604	0.18739
1990 (21)	178	77	1404	5604	1481	5705	0.20034
1991 (22)	178	81	1481	5705	1562	5802	0.20610
1992 (23)	176	113	1562	5802	1675	5865	0.21211
1993 (24)	224	84	1675	5865	1759	6005	0.22215
1994 (25)	269	58	1759	6005	1817	6216	0.22656
1995 (26)	249	86	1817	6216	1903	6379	0.22619
1996 (27)	436	59	1903	6379	1962	6756	0.22978
1997 (28)	454	64	1962	6756	2026	7146	0.22505
1998 (29)	408	89	2026	7146	2115	7465	0.22089
1999 (30)	373	69	2115	7465	2184	7769	0.22077
2000 (31)	261	64	2184	7769	2248	7966	0.21943
2001 (32)	225	61	2248	7966	2309	8130	0.22009
2002 (33)	192	50	2309	8130	2359	8272	0.22119
2003 (34)	218	43	2359	8272	2402	8447	0.22190
2004 (35)	171	46	2402	8447	2448	8572	0.22140
2005 (36)	217	42	2448	8572	2490	8747	0.22214
2006 (37)	204	48	2490	8747	2538	8903	0.22159
2007 (38)	186	40	2538	8903	2578	9049	0.22183
2008 (39)	172	58	2578	9049	2636	9163	0.22173
2009 (40)	164	83	2636	9163	2719	9244	0.22341
2010 (41)							0.22728

Source: Workload Statistics, Department of Justice in the United States (US).
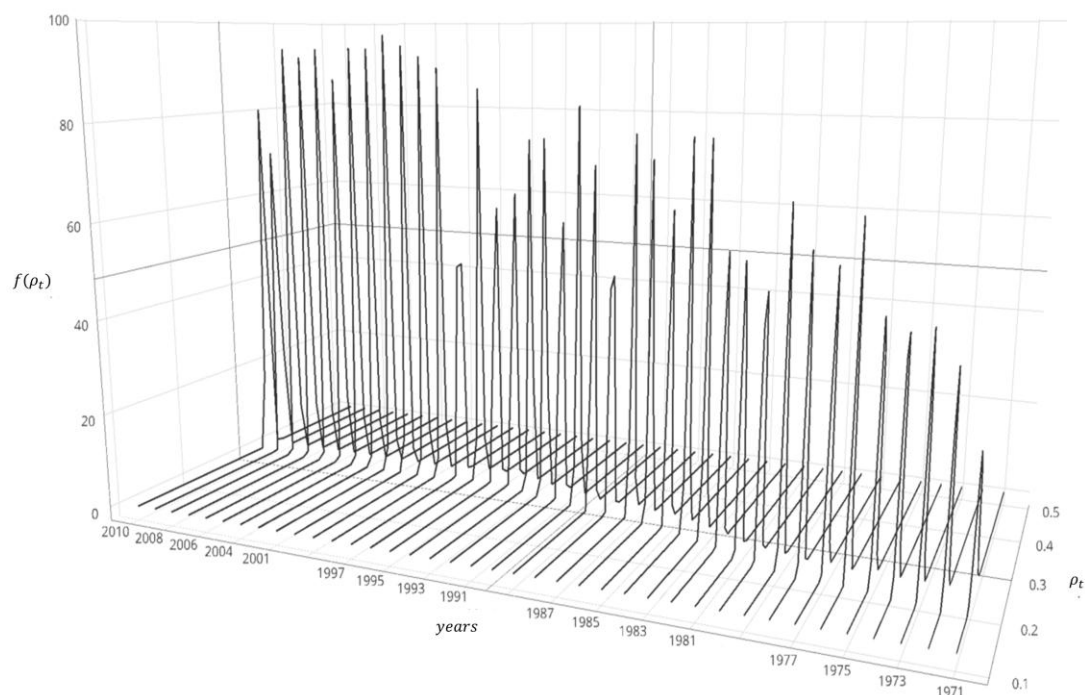
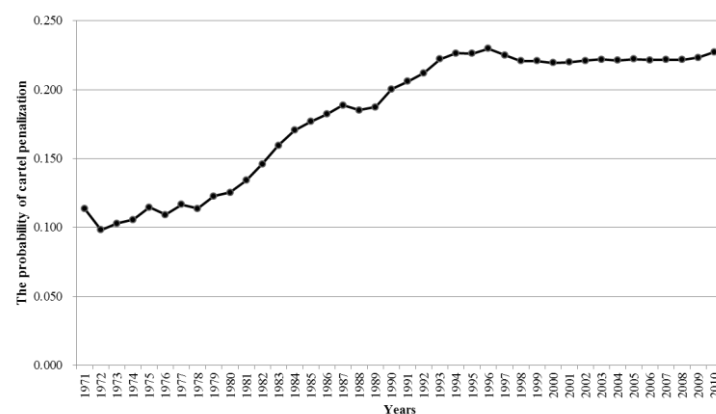**Figure 2.** Annual beta distributions of the probability of cartel penalization.

Figure 3. The annual probability of cartel penalization.

In the late 19th century, the United States was confronted with a very significant change: large-scale manufacturing interests emerged, in great numbers, and enjoyed excessive economic power. In response, the Interstate Commerce Act in 1887 began a shift towards federal rather than state regulation of big business. This was followed by the Sherman Antitrust Act in 1890, which is the basis of US competition laws. Later, the Clayton Antitrust Act in 1914 was enacted to prohibit price discrimination, corporate mergers, and interlocking directorates.

We can now show how the change of probability of cartel penalization impacted upon the antitrust laws in the analysis periods. The Antitrust Penalty and Procedure Act in 1974, which was known as the Tunney Act, required that prospective mergers and acquisitions obtain approval from the DOJ. In 1976, the Hart–Scott–Rodino Antitrust Improvements Act was passed, and in 1978, the leniency program was instituted. At this notable time, the probability of cartel penalization was increasing. At the peak of cartel penalization probability, in 1994, the DOJ reformed the leniency program. The reformed version of the program included an additional amnesty for those who cooperate with investigations.

Figure 3 indicates that the probability after 1994 has been steady and stable. The reform of competition laws clearly had an impact on the industry.

3.4. Model Comparison

Chang and Harrington [24] constructed a Markov process model to consider the stochastic formation and demise of cartels. By numerical analysis, they estimated the impact of the leniency program on the steady-state rate. Figure 4, in the form of the analysis results, plots the change in the rate of penalized cartels according to the proportion of prosecuted cases.
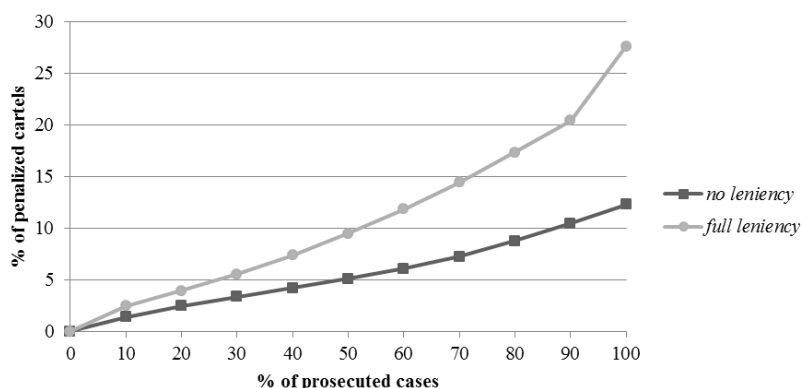

Figure 4. Effects of the proportions of penalized cartels according to the probability of prosecuted cases.

The proportion of probable prosecution cases, as reflects the 1970–2009 Workload statistics, was about 20~40%. In this value, the rate of penalized cartels is estimated about 5~10%.

The estimated probability of cartel penalization of this study and Bryant and Eckard [1]'s results are similar in their proportion of penalization to investigation. However, the present approach is the ensemble-average probability using discrete data, whereas that of Bryant and Eckard [1] is the time-average probability using continuous data. Cartel analysis is more commensurate with discrete data than with continuous data, because the form of Workload statistics data, as announced annually by the DOJ, is discrete. With our similar definition of probability, we could draw a box plot in the overlapped analysis period 1962–1988.

Figure 5 shows that the Bayesian probabilistic model estimates 0.114 for the top 25th percentile, and 0.1737 for the top 75th percentile, which are statistically significant. These are close to Bryant and Eckard [1]'s estimates, which fell between 0.128 and 0.174.
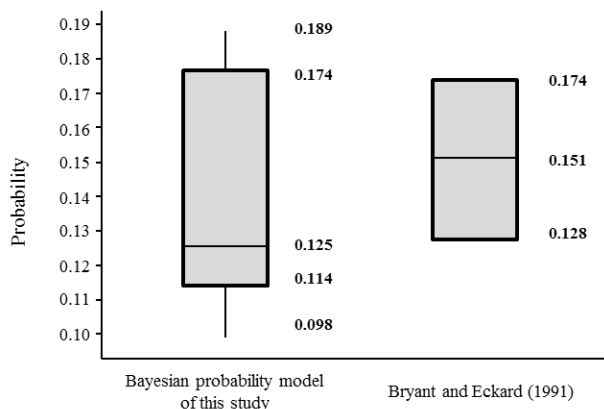

Figure 5. Box plots of the Bayesian probabilistic model and Bryant and Eckard's (1991) model.

3.5. Impact of Leniency Program

This study utilized a policy simulation to analyze the impact of competition policies [25,26]. In policy evaluation research, the impact of policy implementation is indicated as value-added. In other words, the impact is described as the difference of outcomes between implementing the policy and otherwise. The leniency program has been deemed an effective antitrust policy for detecting and deterring cartels in many countries. In general, the leniency program provides partial or total exemption for penalty to a cartel member who voluntarily reports information or agreements that prove helpful to the antitrust authorities. Under the leniency program, a firm or individual in a cartel is bound to first confess involvement for avoiding conviction or fines. The optimal policy is found by evaluating the impact of the leniency program. It is given by:

$$\frac{BX_{1992} - AX_{1992}}{AX_{1992}} \times 100 . \tag{10}$$

The impact of the leniency program (%) is the difference between the penalization probability under both it and non-leniency. The leniency program was originally launched in 1978 in the US, and was reformed in 1993. In Equation (10), BX_{1992} is the 1992 penalization probability estimated on the basis of the leniency program's implementation in 1978, and AX_{1992} is the penalization probability in 1992 estimated on the basis of the leniency program's non-implementation. The estimated probability BX_{1992} was calculated as 0.21211 by the Bayesian probabilistic model, and AX_{1992} was calculated as 0.1328 by the ordinary least squares estimation method of regression. The impact of the leniency program by the policy simulation, finally, is 65.39%. This can be seen in Figure 6.
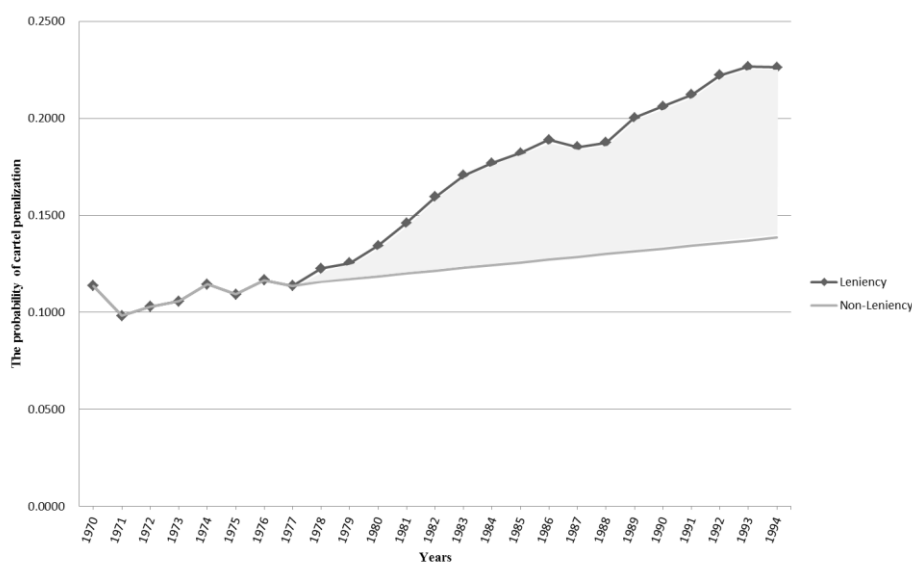

Figure 6. The increment of the probability of cartel penalization in US.

There has been much research that has analyzed the effectiveness and efficiency of the leniency program (i.e., Miller [7], Chang and Harrington [24], and Brenner [27]). The result of this study is similar to those of the research of Chang and Harrington [24] and Miller [7], which is based on US data; the implication was that the leniency program is a very effective policy. Chang and Harrington [24] argue that the occurrence of cartels decreased by about 70%, and the deterrence capability of the antitrust authority increased by about 60% after introducing the leniency program. Miller [7], through Poisson regression analysis, estimated the impact of the leniency program every half year using US data for the years 1985 to 2005. In the results, the detection capability increased by about 60%, and the deterrence capability improved by about 40%.

4. Conclusions

This study attempted to estimate the probability of cartel penalization using a Bayesian approach. Bryant and Eckard [1], Combe et al. [2], Harrington and Chang [3], and Zhou [4] estimated the probability of cartel detection in the form of the time-average probability from continuous data. However, the probability of cartel penalization of this study was estimated in the form of the ensemble-average probability from Workload statistics. Bryant and Eckard [1], Combe et al. [2], Harrington and Chang [3], Zhou [4], and Miller [7] all assumed that the duration of cartels and the inter-arrival times between cartels follow exponential distributions, and that the stochastic process for cartel cases is stationary. However, we built a Bayesian probabilistic model, as it did not need to consider a stationary process. This study made two assumptions: an industry-based analysis, and the grim trigger strategy. On the basis of the 1970–2009 Workload statistics from the US Department of Justice, the determined probability of cartel penalization reflected a sensitive response according to the change of antitrust policy. The result of the policy simulation of the impact of the leniency program was about 65%. The results are similar with the results of Chang and Harrington [24] and Miller [7], and similar to that of Bryant and Eckard [1]; indeed, the common finding among all of the studies, including the current study, was that the leniency program is a very effective policy.

This study evaluated the impact of antitrust policy and, therefrom estimated the probability of cartel penalization. From the antitrust authority standpoint, it provides an improved optimal policy, and from the corporate standpoint, it provides more effective decision-making. Certainly, the present paper has several limitations. First, further studies on realistic situations in specific countries and industries are needed. New antitrust policies recently have been introduced, such as for example, Amnesty Plus, punitive damage, class action, and consent order. These were also considered in further study.

Author Contributions: J.P., J.L. and S.A. conceived and designed the research; J.P. and J.L. conducted the research and drafted the manuscript. S.A. supervised the overall work. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

An ADF unit root test of maximum time lag 10 based on the Schwarz information criterion is performed using E-Views software. The regression of the time series for the test is

$$y_t = \delta y_{t-1} + u_t, \tag{A1}$$

where u_t is the white noise error term, following the normal distribution of mean 0 and variance σ^2.

The case of $\delta = 1$ in Equation (A1) indicates that the model has a unit root with a random walk. Time lags usually account for one-third of the total time series [22]. Accordingly, in the ADF unit root test, the time series is 30, and so the maximum time lag is 10. In any ADF unit root test, the procedure is important [28,29]. Such procedures are the model including the constant and time trend ($y_t = \beta_0 + \beta_1 t + \delta y_{t-1} + u_t$), the model including the constant ($y_t = \beta_1 t + \delta y_{t-1} + u_t$), and the model including nothing ($y_t = \delta y_{t-1} + u_t$).

There are information criteria for ADF unit root tests: the AIC (Akaike information criterion), and the above-noted SIC (Schwarz information criterion). SIC, which supplements the AIC with the Bayesian view, is mainly used in empirical analysis, and is also known as the Bayesian information criterion [30].

$$AIC = e^{2k/n} \frac{RSS}{n}, SIC = n^{k/n} \frac{RSS}{n}, \tag{A2}$$

where k is the number of regressors, n is the number of observations and RSS (residual sum of squares) is the sum of square error between the data. The null hypothesis for the ADF unit root test is "including a unit root ($\delta = 1$)." Initially, the present study used the ADF unit root test with the model including the constant and time trend based on the detection cases data. The results are provided in Table A1.

Table A1. ADF unit root test with the model including constant and time trend based on the detection cases data.

			t-Statistic	Prob.
Augmented Dickey-Fuller test statistic			-2.981691	0.1501
Test critical values:	1% level		-4.211868	
	5% level		-3.529758	
	10% level		-3.196411	
Variable	**Coefficient**	**Std. Error**	**_t_-Statistic**	**Prob.**
Detection cases (-1)	-0.391532	0.131312	-2.981691	0.0051
Constant	156.6549	60.87809	2.573255	0.0143
@TREND (1970)	-2.307219	1.308371	-1.763428	0.0863

Table A1 shows that the p-value of the ADF test statistic, 0.1501, is greater than the significance level (0.05). This means that the null hypothesis cannot be rejected (the detection cases data has a unit root). Testing of the constant and time trend can show variable *Constant* and *@TREND* in the below of Table A2. The p-value of the constant is about 0.0143, smaller than the significance level (0.05). That is, the null hypothesis "no constant ($\beta_0 = 0$)" can be rejected. The p-value of the trend is 0.0863, again greater than the significance level (0.05). That is, the null hypothesis "no time trend ($\beta_1 = 0$)" also cannot be rejected. The time series data on the detection cases includes the unit root as well as the. Because of the lack of any time trend, we progress to the next step, which is the ADF unit root test with the model including only the constant. The results of this test are summarized in Table A2.

Table A2. ADF unit root test with the model including constant based on the detection cases data.

			t-Statistic	Prob.
Augmented Dickey-Fuller test statistic			-2.343469	0.1641
Test critical values:	1% level		-3.610453	
	5% level		-2.938987	
	10% level		-2.607932	
Variable	**Coefficient**	**Std. Error**	**_t_-Statistic**	**Prob.**
Detection cases (-1)	-0.245224	0.104641	-2.343469	0.0246
Constant	66.25393	33.75776	1.962628	0.0572

Table A2 shows that the p-value of the ADF test statistic is 0.1641, greater than the significance level (0.05). This result means that the data has a unit root. The p-value for constant is 0.0572, again greater than significance level (0.05). That is, the null hypothesis ($\beta_0 = 0$) cannot be rejected. The time series data on the detection cases includes the unit root. Because of no constant, we progress to the final step, which is the ADF unit root test with the model including nothing. The results of the ADF root test are summarized in Table A3.

Table A3. ADF unit root test with the model including nothing based on the detection cases data.

			t-Statistic	Prob.
Augmented Dickey-Fuller test statistic			−1.396253	0.1487
Test critical values:	1% level		−2.625606	
	5% level		−1.949609	
	10% level		−1.611593	
Variable	**Coefficient**	**Std. Error**	***t*-Statistic**	**Prob.**
Detection cases (−1)	−0.052658	0.037714	−1.396253	0.1707
R-squared	0.038594	Mean dependent var		−7.923077
Adjusted R-squared	0.038594	S.D. dependent var		77.49139
S.E. of regression	75.98132	Akaike info criterion		11.52416
Sum squared resid	219380.1	Schwarz criterion		11.56681
Log likelihood	−223.7211	Hannan-Quinn criter.		11.53946
Durbin-Watson stat	2.689882			

Table A3 shows that the Durbin-Watson statistic is 2.689882 where $k = 1$ and $n = 30$. The significance level (0.05) of these variables sets up as $d_L = 1.352$, $d_U = 1.489$. The null hypothesis "serially uncorrelated" can be rejected, because DW statistics (d) is included between $4 − d_L$ and 4. The data on detection cases presents an eventually negative correlation. *p*-value of the ADF test statistic is 0.1487, greater than the significance level (0.05). This result means that the data has a unit root. In conclusion, the time series data on the detection cases includes the unit root and does not include constant and time trend. In the sequence analysis, we also use an ADF unit root test with the model including the constant and time trend based on the penalization cases data. The results are summarized in Table A4.

Table A4. ADF unit root test with the model including constant and time trend based on the penalization cases data.

			t-Statistic	Prob.
Augmented Dickey-Fuller test statistic			−2.189536	0.4808
Test critical values:	1% level		−4.234972	
	5% level		−3.540328	
	10% level		−3.202445	
Variable	**Coefficient**	**Std. Error**	***t*-Statistic**	**Prob.**
Penalization cases (−1)	−0.472117	0.215624	−2.189536	0.0365
D (Penalization cases (−1))	−0.158801	0.238570	−0.665637	0.5107
D (Penalization cases (−2))	−0.242738	0.210821	−1.151395	0.2587
D (Penalization cases (−3))	0.243527	0.182994	1.330796	0.1933
Constant	38.11916	18.03466	2.113660	0.0430
@TREND (1970)	−0.233903	0.352704	−0.663171	0.5123

Table A4 shows that the *p*-value of the ADF test statistic, 0.4808, which is very much greater than the significance level (0.05). This means that the null hypothesis cannot be rejected (the penalization cases data has a unit root). The *p*-value of the constant is about 0.0043, smaller than the significance level (0.05). The *p*-value of the trend is 0.5123, greater than the significance level (0.05). The time series data on the penalization cases includes the unit root as well as the constant with the model including the constant and time trend. Because of the lack of any time trend, we progress to the next step, which is the ADF unit root test with the model including only the constant. The results of this test are summarized in Table A5.

Table A5. ADF unit root test with the model including constant based on the penalization cases data.

			t-Statistic	Prob.
Augmented Dickey-Fuller test statistic			−2.131969	0.2339
Test critical values:	1% level		−3.626784	
	5% level		−2.945842	
	10% level		-2.611531	
Variable	**Coefficient**	**Std. Error**	***t*-Statistic**	**Prob.**
Penalization cases (−1)	−0.450100	0.211119	−2.131969	0.0410
D (Penalization cases (−1))	−0.154786	0.236329	−0.654959	0.5173
D (Penalization cases (−2))	−0.229258	0.207934	−1.102552	0.2787
D (Penalization cases (−3))	0.260693	0.179510	1.452247	0.1565
Constant	31.58062	14.96390	2.110454	0.0430
R-squared	0.474430	Mean dependent var		0.611111
Adjusted R-squared	0.406615	S.D. dependent var		27.22633
S.E. of regression	20.97285	Akaike info criterion		9.052581
Sum squared resid	13635.67	Schwarz criterion		9.272514
Log likelihood	−157.9465	Hannan-Quinn criter.		9.129343
F-statistic	6.995902	Durbin-Watson		2.098929
Prob (F-statistic)	0.000391			

Table A5 shows that the Durbin-Watson statistic is 2.098929 where $k = 1$ and $n = 30$. The significance level (0.05) of these variables sets up as $d_L = 1.352$, $d_U = 1.489$. The null hypothesis "serially uncorrelated" cannot be rejected, because DW statistics (d) is included between d_U and $4 - d_U$. The data on penalization cases eventually resulted in no correlation. It shows that the *p*-value of the ADF test statistic is 0.2339 greater than the significance level (0.05). This result means that the data has a unit root. The *p*-value for constant is 0.043, greater than the significance level (0.05). That is, null hypothesis ($\beta_0 = 0$) can be rejected. Therefore, we finish the steps. The time series data about penalization cases includes unit root and constant.

References

1. Bryant, P.G.; Eckard, E.W. Price Fixing—The Probability of Getting Caught. *Rev. Econ. Stat.* **1991**, *73*, 531–536. [CrossRef]
2. Combe, E.; Monnier, C.; Legal, R. Cartels: the Probability of Getting Caught in the European Union. Bruges European Economic Research (BEER) Papers 12/March 2008. Available online: http://aei.pitt.edu/58577/ (accessed on 8 June 2018).
3. Harrington, J.E.; Chang, M.H. Modeling the Birth and Death of Cartels with an Application to Evaluating Competition Policy. *J. Eur. Econ. Assoc.* **2009**, *7*, 1400–1435. [CrossRef]
4. Zhou, J. *Evaluating Leniency with Missing Information on Undetected Cartels: Exploring Time-Varying Policy Impacts on Cartel Duration*; TILEC Discussion Paper; Tilburg University—Tilburg Law and Economics Center (TILEC): Tilburg, The Netherlands, 2011.
5. Gallager, R.G. *Stochastic Processes: Theory for Applications*; Cambridge University Press: Cambridge, UK, 2013.
6. Li, X.R. *Probability, Random Signals, and Statistics*; CRC Press: Boca Raton, FL, USA, 1999.
7. Miller, N.H. Strategic Leniency and Cartel Enforcement. *Am. Econ. Rev.* **2009**, *99*, 750–768. [CrossRef]
8. Hinloopen, J. Internal cartel stability with time-dependent detection probabilities. *Int. J. Ind. Organ.* **2006**, *24*, 1213–1229. [CrossRef]
9. Ormosi, P.L. A Tip of the Iceberg? The Probability of Catching Cartels. *J. Appl. Econ.* **2014**, *29*, 549–566. [CrossRef]
10. Armstrong, J.S. Judgmental bootstrapping: Inferring experts' rules for forecasting. In *Principles of Forecasting*; Springer: Berlin, Germany, 2001; pp. 171–192.
11. Harrington, J.E. Behavioral screening and the detection of cartels. In *European Competition Law Annual, 2006*; Hart Publishing: Oxford, UK, 2006; pp. 51–69.

12. Bos, I.; Harrington, J.E. Endogenous cartel formation with heterogeneous firms. *Rand J. Econ.* **2010**, *41*, 92–117. [CrossRef]

13. Friedman, J.W. A non-cooperative equilibrium for supergames. *Rev. Econ. Stud.* **1971**, *38*, 1–12. [CrossRef]

14. Harrington, J.E. Optimal corporate leniency programs. *J. Ind. Econ.* **2008**, *56*, 215–246. [CrossRef]

15. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer: Berlin, Germany, 1985.

16. Da-Silva, C.Q.; Migon, H.S.; Correia, L.T. Dynamic Bayesian beta models. *Comput. Stat. Data Anal.* **2011**, *55*, 2074–2089. [CrossRef]

17. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; CRC Press: Boca Raton, FL, USA, 2013.

18. Ahn, S.E.; Park, C.S.; Kim, H.M. Hazard rate estimation of a mixture model with censored lifetimes. *Stoch. Environ. Res. Risk Assess.* **2006**, *21*, 711–716. [CrossRef]

19. Koop, G.; Poirier, D.J. Bayesian-Analysis of Logit-Models Using Natural Conjugate Priors. *J. Econ.* **1993**, *56*, 323–340. [CrossRef]

20. Guérin, F.; Dumon, B.; Usureau, E. Reliability estimation by Bayesian method: Definition of prior distribution using dependability study. *Reliab. Eng. Syst. Saf.* **2003**, *82*, 299–306. [CrossRef]

21. Antitrust Division. *10-Year Workload Statistics*; Department of Justice: Washington, DC, USA, 1970–2009.

22. Gujarati, D. *Basic Econometrics*; MeGraw-Hill: New York, NY, USA, 2008.

23. Nelson, C.R.; Plosser, C.R. Trends and random walks in macroeconmic time series: Some evidence and implications. *J. Monet. Econ.* **1982**, *10*, 139–162. [CrossRef]

24. Chang, M.-H.; Harrington, J.E. *The Impact of a Corporate Leniency Program on Antitrust Enforcement and Cartelization*; Department of Economics, The Johns Hopkins University: Baltimore, MD, USA, 2008.

25. Pindyck, R.S.; Rubinfeld, D.L. *Econometric Models and Economic Forecasts*; McGraw-Hill: New York, NY, USA, 1981.

26. Sims, C.A. Policy Analysis with Econometric-Models. *Brook. Pap. Econ. Act.* **1982**, *1982*, 107–164. [CrossRef]

27. Brenner, S. An empirical study of the European corporate leniency program. *Int. J. Ind. Organ.* **2009**, *27*, 639–645. [CrossRef]

28. Dickey, D.A.; Rossana, R.J. PRACTITIONERS' CORNER Cointegrated Time Series: A Guide to Estimation and Hypothesis Testing. *Oxf. Bull. Econ. Stat.* **2009**, *56*, 325–353. [CrossRef]

29. Dolado, J.J.; Jenkinson, T.; Sosvilla-Rivero, S. Cointegration and Unit Roots. *J. Econ. Surv.* **1990**, *4*, 249–273. [CrossRef]

30. Koehler, A.B.; Murphree, E.S. A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order. *Appl. Stat. J. R. Stat. Soc. Ser. C* **1988**, *37*, 187–195. [CrossRef]