5 Empirical Results

Before estimating multinominal logistic model and AFT model in equation (1) and (3), preliminary data exploration will be presented in advance to provide a sketch of this survey data. Table 3 demonstrates the distribution of the preference group. Unsurprisingly, the number of respondents in group 2 is the minimum among the three and that in group 3 is the maximum. It is definitely a good thing to have a reasonably large number of observations in group 3, but we should also be cautious about the unbalanced data structure.

Table 3: Cross-Table for Question 218 and Question 219

		Purchase *Non-GM*		
		Possible	Impossible	Don't Know
Purchase *GM*	Possible	268($G = 3$)	8($G = 2$)	3
	Impossible	58 ($G = 1$)	18[a]	1
	Don't Know	3	1	9

[a] These 18 observations and Don't know responses will be deleted from further analysis due to their uninformative.

Conditional on $G = 3$, these 268 observations are then used to estimate AFT model in equation (3). Table 3 demonstrates how these respondents decide which salmon to purchase when the prices for GM-fed salmon and Non-GM-fed salmon are equal.

Table 4: Initial CV Response at Equal Price

	Non-GM-fed Salmon	GM-fed Salmin	Indifference	Don't Know
Frequency	123	92	50	3
Percentage	45.8%	34.3%	18.6%	1.1%

It is evident that most of them choose the Non-GM-fed option, followed by GM-fed. Unlike Don't Know response, we think Indifference is a choice that also conveys useful information. Two possible reasons may be used to explain such phenomenon. First, any one who chooses indifference might feel that these two items are essentially the same. Second, in an economic view point, either consuming GM-fed or Non-GM-fed salmon yields the same amount of utility. In that case, genetic modification is irrelevant to these respondent.

Don't Know responses are generally treated as uninformative response and investigators often omit those resposes. However, we suppose these Don't Know responses in Table 4 are made temporarily, they could be unfamiliar with such questionnaire or don't know how to make a decision if the prices are the same. We assume that their behaviors can be elicited if more options or more information are provided. Therefore, follow up questions for these Indifference and Don't Know responses could be randomly selected from one of the four options, including lower(raise) the price for GM-fed salmon and lower(raise) the price for Non-GM-fed salmon.

After the initial CV question are two follow up questions to which a respondent is asked to answer. Combining all the information obtained from these three questions, the observed WTPP interval will be determined accordingly. In chapter 3 we have illustrated how this interval is determined by three decision trees. Table 5 summarizes all possible outcomes in these decision trees .

Table 5: Observed Censoring Data with Frequencies

Censoring Interval	Count	Censoring Interval	Count
(-C,-113)	12	(0,38)	18
(-C,-67)	15	(0,22)	16
(-C,-22)	26	(0,8)	8
(-113,-75)	6	(37,75)	9
(-67,-45)	5	(22,45)	6
(-22,-15)	7	(8,15)	2
(-75,-37)	9	(75,113)	6
(-45,-22)	4	(45,67)	10
(-15,-8)	3	(15,22)	7
(-38,0)	12	113+	14
(-22,0)	10	67+	26
(-8,0)	10	22+	27

In order to have an idea about how WTPP is distributed, we conduct a non-parametric Turnbull (1976) estimation to estimate the survival function of WTPP.

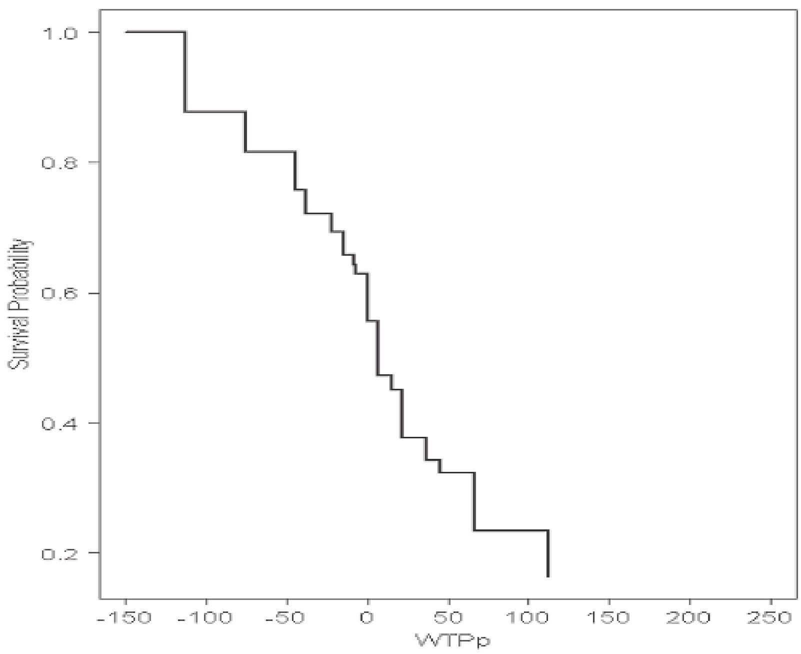

Figure 4: Turnbull Estimation of WTP for Premium

Figure 4 illustrates Turnbull estimates of willingness to pay for premium without considering the effects introduced by covariates. Based on this graph, it can be inferred that the median WTPP is about 10 but we can't infer the mean WTPP by this graph as WTPP is not non-negatively defined. Additional survival curves are stratified by covariates. Through these graphs, we would like to profile the respondents whose WTPP is relatively higher or lower. Moreover, it provides a guideline to detect variables significantly affecting WTPP.

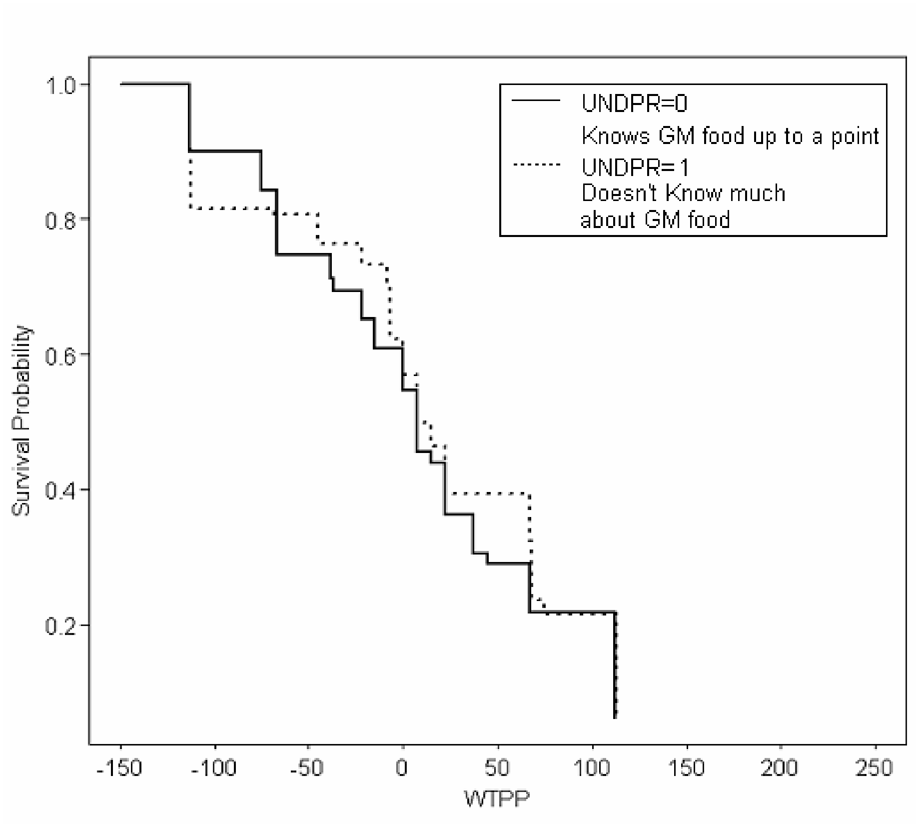

Figure 5: Turnbull Estimation of WTP for Premium by UNDPR

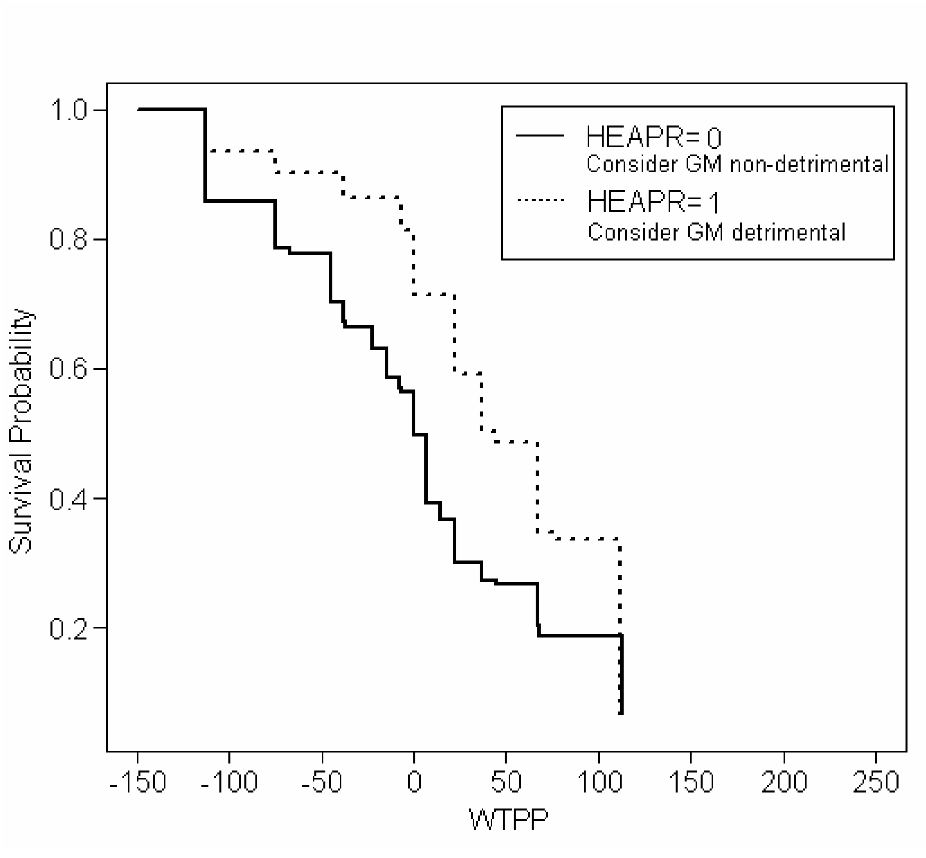

Figure 6: Turnbull Estimation of WTP for Premium by HEAPR

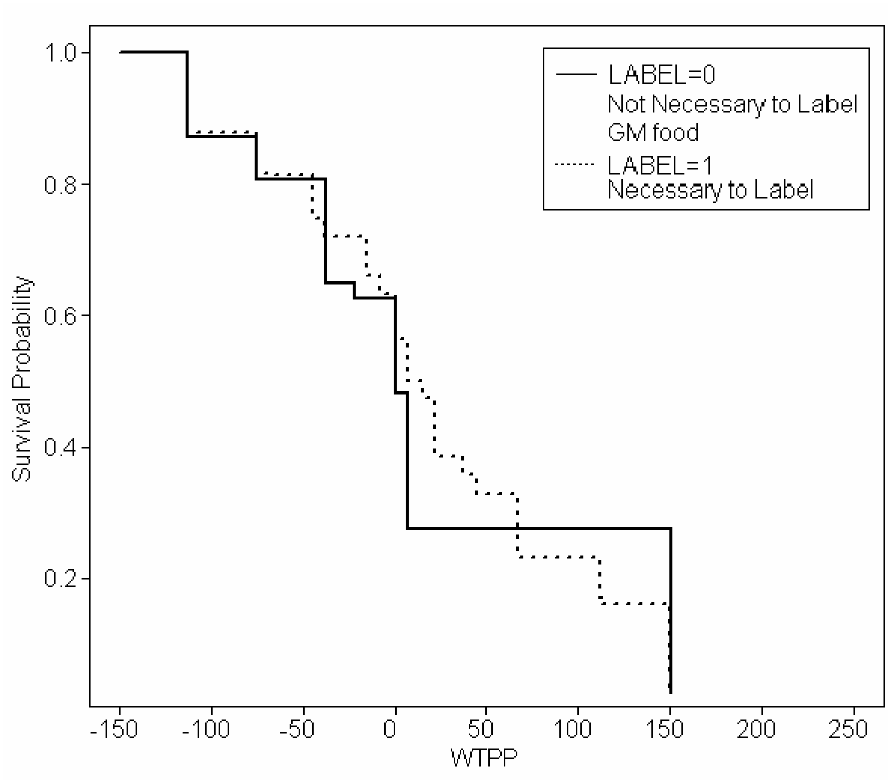

Figure 7: Turnbull Estimation of WTP for Premium by LABEL

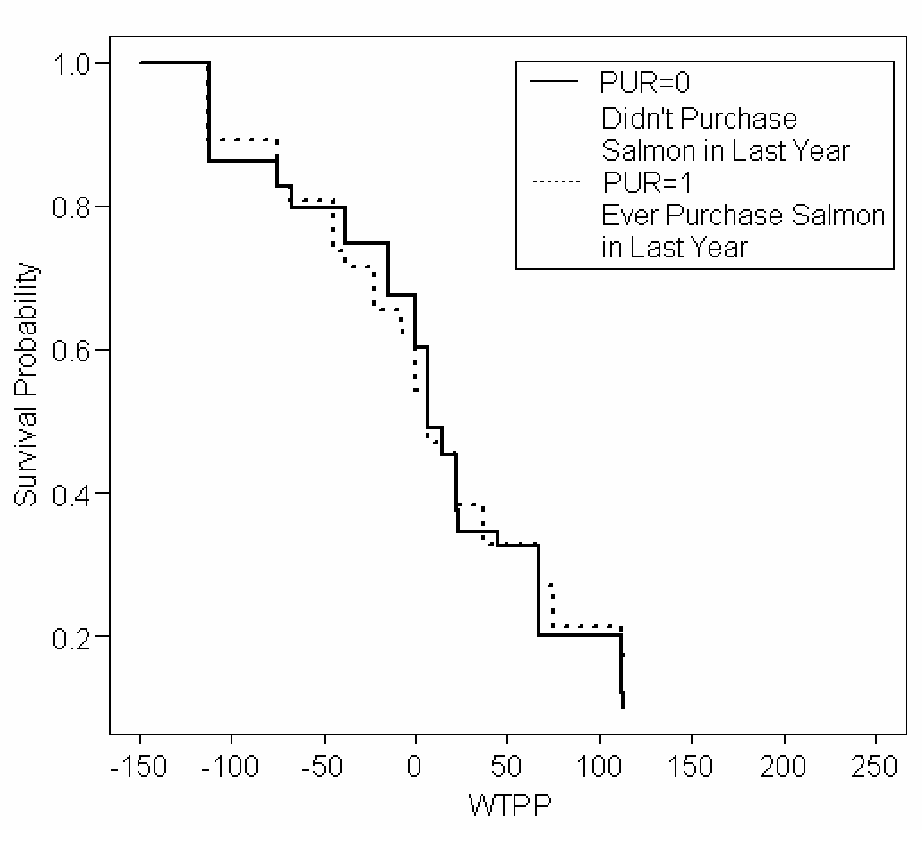

Figure 8: Turnbull Estimation of WTP for Premium by PUR

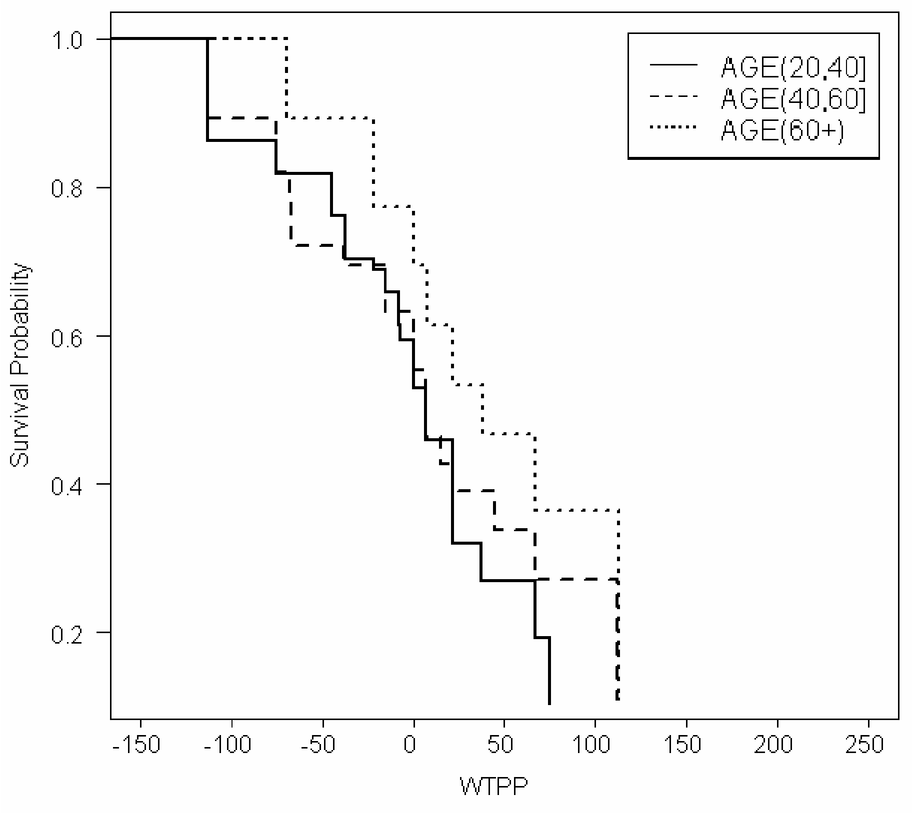

Figure 9: Turnbull Estimation of WTP for Premium by AGE

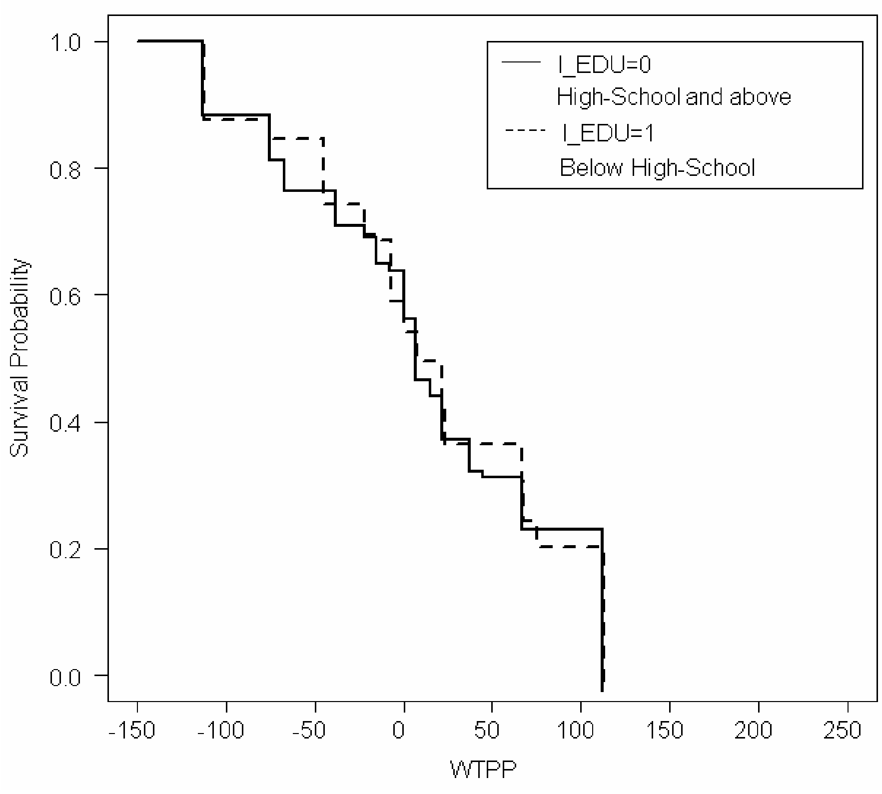

Figure 10: Turnbull Estimation of WTP for Premium by I-EDU

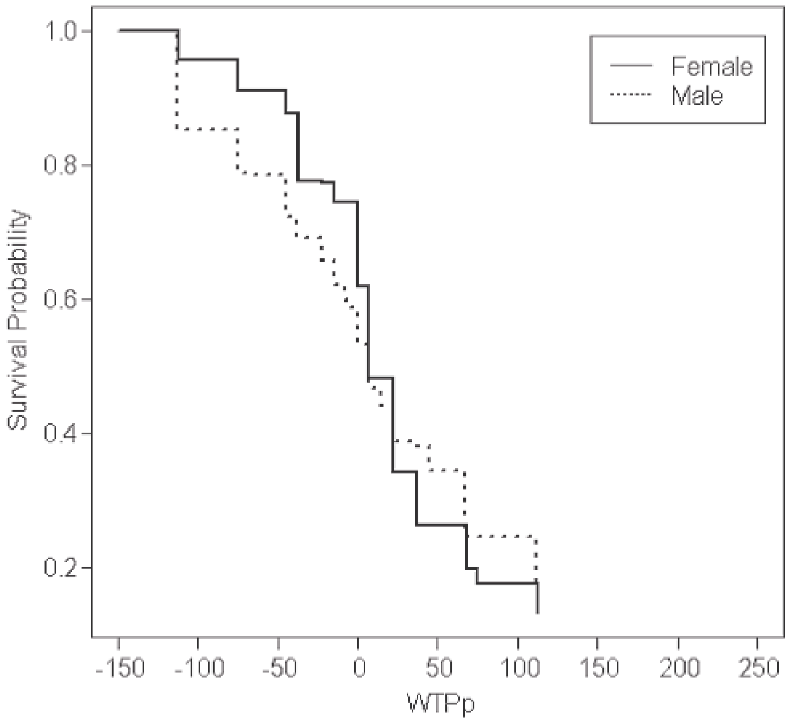

Figure 11: Turnbull Estimation of WTP for Premium by SEX

According to these graphical displays, one may conclude temporarily that HEAPR, SEX, UNDPR, and AGE are potential covariates while LABEL, PUR and I-EDU, are possibly not. Furthermore we can also infer that aged respondents who don't know much about GM foods, consider GM foods risky, and don't have high school degree are willing to pay more premium to purchase Non-GM instead of GM counterpart. These findings are helpful in the subsequent discussion on parametric models.

Parameter Estimates

In this sub-section, the estimated parametric models, including multinominal logistic model and AFT model, will be presented. To begin with, parameter estimates for the model without covariates are given in Table 6.

Table 6: Estimated Model without Covariates

Covariates	Point Estimate	Standard Deviation	Asy. $\chi^2(1)$	P-Value
		Multinominal Logistic Model		
$\hat{\alpha}_0$(Intercept)	-1.5305**	0.1448	111.719	≈ 0
$\hat{\beta}_0$(Intercept)	-3.5114**	0.3587	95.829	≈ 0
		AFT Model		
$\hat{\theta}_0$(Intercept)	5.3204**	0.0565	8852.7	≈ 0
$\hat{\sigma}$	0.4213**	0.0520	65.497	≈ 0
$\hat{q}$	1.5476**	0.3025	26.170	≈ 0
-2Log-Likelihood	1457.773(380.7907+1076.982)			
AIC	1467.773			

[a] *,**: Significant at $\alpha = 0.05$ and 0.01, respectively

[b] Sample size are 334 and 268 for respective models.

However, the parameter estimates associated with multinominal logistic model, $\hat{\alpha}$ and $\hat{\beta}$, rarely provide useful information since they tell us nothing but sample proportions of $G = 1$, $G = 2$, and $G = 3$. On the other hand, parameter estimates associated with AFT model help not only to estimate the overall mean of WTPP but also provide the information about the distribution of WTPP. The estimated mean is about 10.92 and the density function and survival function are shown in Figure 12.

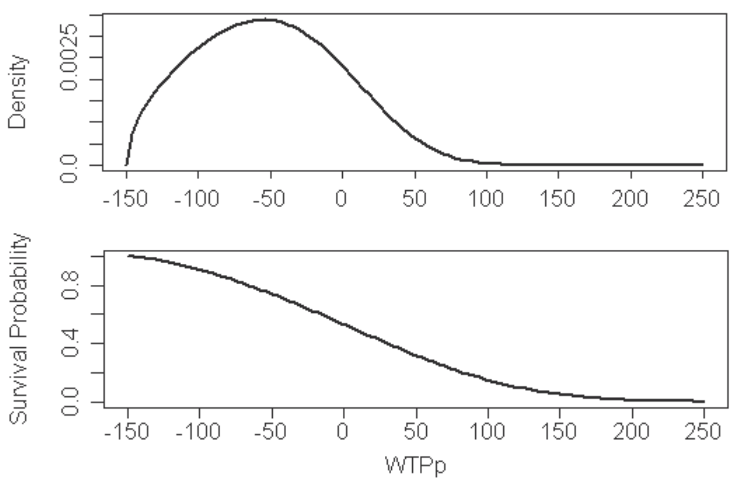

Figure 12: Estimated Density Function and Survival Function

Before introducing covariates into multinominal logistic model, it should be noticed that there are only 8 observations belonging to the second group. Owing to

data sparsity, it is not suggested to introduce too many covariates into the logarithm of the ratio $P(G = 2|\boldsymbol{s})$ to $P(G = 3|\boldsymbol{s})$. Hence, the maximum number of covariates is temporarily set to be two in this case.
A forward search is performed and model selection technique is then used to identify the best-fitted model. model.

Table 7: Estimated Parametric Model I

Covariates	Point Estimate	Standard Deviation	Asy. $\chi^2(1)$	P-Value
	Multinominal Logistic Model			
$\hat{\alpha}_0$(Intercept)	-2.3310**	0.252	85.049	≈ 0
$\hat{\alpha}_1$(HEAPR)	1.3566**	0.303	19.968	< 0.01
$\hat{\alpha}_2$(I-EDU)	0.5548	0.306	3.284	0.069
$\hat{\beta}_0$(Intercept)	-2.9859**	0.359	59.428	≈ 0
$\hat{\beta}_1$(HEAPR)	-0.8534	1.081	0.622	0.430
$\hat{\beta}_2$(I-EDU)	-6.8856	16.914	0.165	0.683
	AFT Model			
$\hat{\theta}_0$(Intercept)	4.9246**	0.1364	1302.7	≈ 0
$\hat{\theta}_1$(HEAPR)	0.2677**	0.0833	10.306	0.001
$\hat{\theta}_2$(UNDPR)	0.0792	0.0684	1.341	0.246
$\hat{\theta}_3$(AGE)	0.0063*	0.0028	4.988	0.025
$\hat{\sigma}$	0.4309**	0.0472	83.379	≈ 0
$\hat{q}$	1.4203**	0.2646	28.798	$< 10^{-7}$
-2Log-Likelihood	1406.8218(347.5108+1059.311)			
AIC	1430.8218			

[a] *,**: Significant at $\alpha = 0.05$ and 0.01, respectively.

One should take heed of the unstable parameter estimates of β_1 and β_2, particularly β_2, as their variances are relatively larger than the others. Data sparsity should be account for such instability. Since fitting two covariates for the logarithm of the ratio is still too much, we consider a model without fitting any covariate for it. Moreover, interaction effect will be considered for the ratio $P(G = 1|\boldsymbol{s})$ to $P(G = 3|\boldsymbol{s})$ as well as AFT model .

Once again, we perform a forward selection to identify the best-fitted model based on minimum AIC criteria. A improved model is shown below.

Table 8: Estimated Parametric Model II

Covariates	Point Estimate	Standard Deviation	Asy. $\chi^2(1)$	P-Value
		Multinominal Logistic Model		
$\hat{\alpha}_0$(Intercept)	-3.5079**	0.654	28.763	$< 10^{-7}$
$\hat{\alpha}_1$(HEAPR)	1.4800**	0.313	22.361	$< 10^{-5}$
$\hat{\alpha}_2$(I-EDU)	0.6758*	0.311	4.711	0.029
$\hat{\alpha}_3$(AGE)	0.0246*	0.012	3.974	0.046
$\hat{\beta}_0$(Intercept)	-3.5115**	0.358	95.786	≈ 0
		AFT Model		
$\hat{\theta}_0$(Intercept)	4.9246**	0.1364	1302.7	≈ 0
$\hat{\theta}_1$(HEAPR)	0.2677**	0.0833	10.306	0.001
$\hat{\theta}_2$(UNDPR)	0.0792	0.0684	1.341	0.246
$\hat{\theta}_3$(AGE)	0.0063*	0.0028	4.988	0.025
$\hat{\sigma}$	0.4309**	0.0472	83.379	≈ 0
$\hat{q}$	1.4203**	0.2646	28.798	$< 10^{-7}$
-2Log-Likelihood	1409.378(350.067+1059.311)			
AIC	1431.378			

[a] *,**: Significant at $\alpha = 0.05$ and 0.01, respectively.

Albeit we have tried to introduce interaction effects into both multinominal logistic model and AFT model, parameter estimates manifest that none of them reduce -2log-likelihood significantly. Hence the model with interaction effects would not be preferred.

Another noticeable point is that the parameter estimate $\hat{\beta}_0$ rarely provides additional information about the probability of observing group 2. In the following, we propose another model in which group 2 is neglected. Subsequently, a logistic regression will be constructed to model the logarithm of the ratio $P(G = 1|\boldsymbol{s})$ to $P(G = 3|\boldsymbol{s})$. The final model is shown as follow.

Table 9: Estimated Parametric Model III

Covariates	Point Estimate	Standard Deviation	Asy. $\chi^2(1)$	P-Value
		Logistic Regression Model		
$\hat{\alpha}_0$(Intercept)	-3.6988**	0.6669	30.760	< 0.0001
$\hat{\alpha}_1$(UNDPR)	0.5033	0.3130	2.585	0.108
$\hat{\alpha}_1$(HEAPR)	1.4623**	0.3130	21.831	< 0.0001
$\hat{\alpha}_2$(AGE)	0.0249*	0.0125	3.937	0.047
$\hat{\alpha}_3$(I-EDU)	0.5336	0.3190	2.798	0.094
		AFT Model		
$\hat{\theta}_0$(Intercept)	4.9246**	0.1364	1302.7	≈ 0
$\hat{\theta}_1$(HEAPR)	0.2677**	0.0833	10.306	0.001
$\hat{\theta}_2$(UNDPR)	0.0792	0.0684	1.341	0.246
$\hat{\theta}_3$(AGE)	0.0063*	0.0028	4.988	0.025
$\hat{\sigma}$	0.4309**	0.0472	83.379	≈ 0
$\hat{q}$	1.4203**	0.2646	28.798	$< 10^{-7}$
-2Log-Likelihood	1332.395(273.084+1059.311)			
AIC	1354.395			

^{a} *,**: Significant at $\alpha = 0.05$ and 0.01, respectively.

Model Validation

To examine the adequacy of model fitting, particularly AFT model, goodness-of-fit test or other model checking method should be performed. However, for such model specification, a well-proposed testing criteria is few and far between. Therefore, the model fitting will be assessed by two graphical display. The first graph is to compare the survival curve estimated by our parametric AFT model and the one by non-parametric Turnbull method. Our parametric model in Table 9 is said to be fitted adequately provided that these two curves don't differ severely. The second graphical display we are using to assess model adequacy is the added variable plot that has been well developed in normal linear models (Cook and Weisberg, 1982). Added variable plot is a graph that provides visual information on the numerical calculation of the coefficient of a candidate term and a visual assessment of the *net effect* of the candidate regressor, that is, the effect of a regressor in a subpopulation in which all other regressors are held constant. If the added variable plot exhibits a quadratic or periodic pattern, our linear form is then somewhat disputable.

A point needs to be addressed before drawing added variable plot. The added variable plot is originally designed for linear model where there is no censored data. For censored response variable, however, the residual is often calculated by subtracting the lower bound of the response variable by the fitted value provided the response variable is either right censored or interval censored. It follows immediately that the residual is also interval censored or right censored. To implement this model checking technique, we need to transform our censored data into complete data and a conditional expectation technique[7] will be employed to solve this issue.

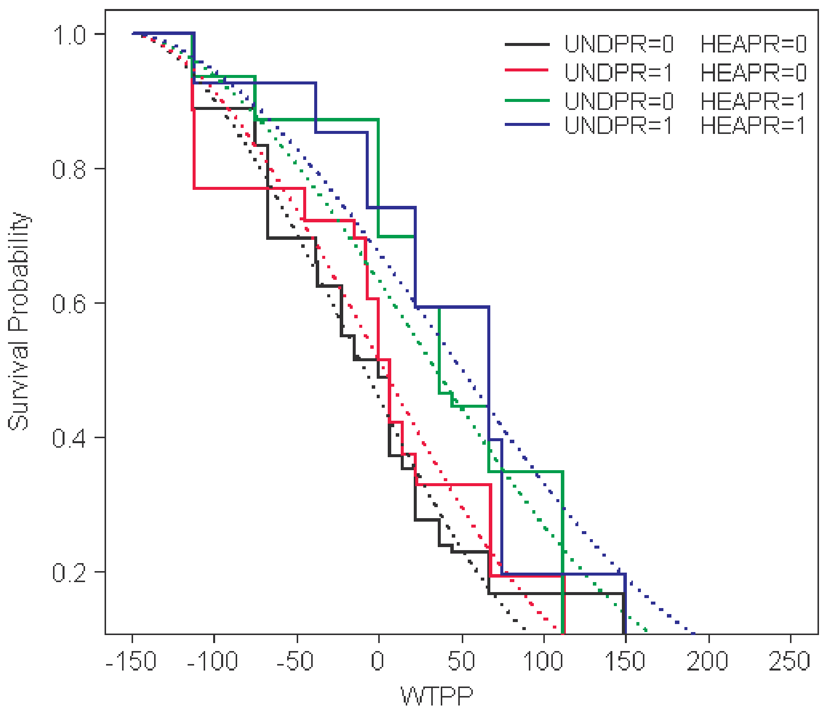

Figure 13: Comparison between Non-parametric Turnbull Estimation and Parametric Model

[7]Refer to Appendix E for more detail

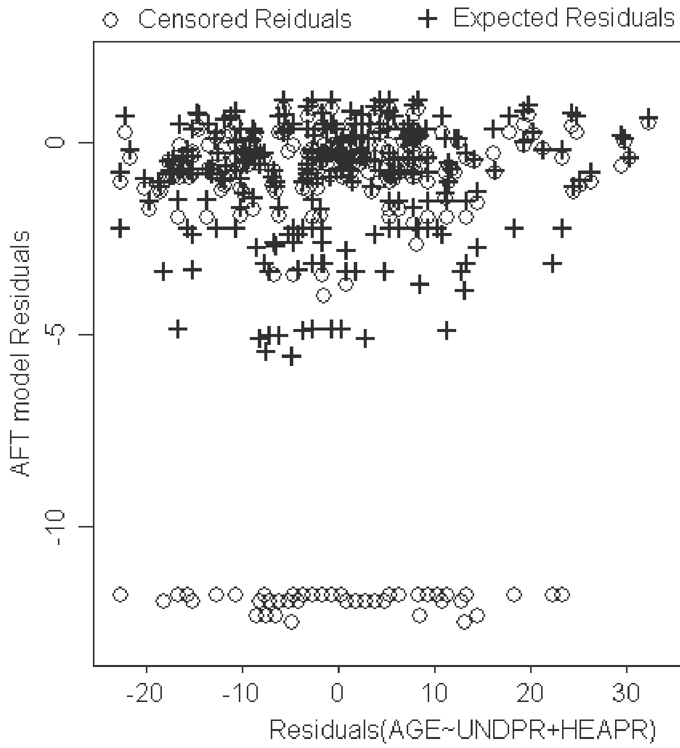

Figure 14: Added Variable Plot

Figure 13 divulges that the survival curves estimated by AFT model and Turnbull estimator are close enough whereas, the added variable plot in Figure 14 doesn't reveal a particular pattern which suggests the covariate AGE should be introduced into our model in a non-linear form. One may argue that the expected residuals are not stochastically scattered about zero, more points lie below zero. Under conventional linear model, we do expect residuals will lie around zero since the error term is assumed to be normally distributed with zero mean. However, it is not the case in our study. Under our model specification, the error term is assumed to be log-generalized gamma distribution. Since the estimated shape parameter $\hat{q}$ is about 1.42, it would be normal to expect more residuals falling below zero.

Moreover, the points which are roughly scattered about a horizontal line indicate that the term AGE will have a coefficient close to zero. In sum, to some extent, there is no significant evidence against our final model. Substantiated by these two graphs, our model should be acceptable and therefore can be employed for further inference.

In evaluating WTP for premium, the overall impression is that aged respondents with greater perception and more knowledge toward GM food are willing to pay the highest premium than the others. Previous studies often suggest that gender difference and education level contribute great influence on willingness to pay. Nonetheless, neither of them are significant in this study. We are trying to figure out if there is any relationship between our significant variables and these two variables suggested by empirical studies.

The following table provides us with a general guideline to see if there exist any particular pattern between these variables.

Table 10: Cross Table for Cognitive Variable and Socioeconomic Variable

		HEAPR=0	HEAPR=1	UNDPR=0	UNDPR=1
SEX=0	I-EDU=0	42	11	32	21
	I-EDU=1	24	9	13	20
SEX=1	I-EDU=0	117	54	113	58
	I-EDU=1	42	35	33	44

By setting up a logistic model with HEAPR as response with SEX and I-EDU as regressors, we find that these two covariates are both significant at $\alpha = 5\%$. Male and high school graduates have higher inclination to consider GM food hazardous. Similarly, we also find that I-EDU is a significant factor in determining whether UNDPR is 0 or 1. It is probably these two risk perception variables that absorb the effects of SEX and I-EDU such that these two covariates which are often considered important are not significant in our study.

Estimate of Mean and Median of WTPP

According to parameter estimates in Table 8, we could calculate the mean and median WTPP for each category. Age, however, is treated as continuous variable in AFT model and therefore can't be enumerated exhaustively. Age 30 and 40 will be provided to exemplify the calculation of means and medians

Table 11: Estimated Mean and Median of WTPP

	AGE=30		AGE=40	
	HEAPR=0	HEAPR=1	HEAPR=0	HEAPR=1
	MEAN			
UNDPR=0	-15.762	25.354	-7.104	36.664
	(8.180)	(13.674)	(6.895)	(13.957)
UNDPR=1	-4.856	39.600	4.504	51.828
	(9.772)	(17.252)	(8.811)	(17.930)
	MEDIAN			
UNDPR=0	-19.219	20.839	-10.784	31.857
	(8.283)	(11.925)	(7.555)	(11.860)
UNDPR=1	-8.594	34.718	0.525	46.631
	(9.342)	(16.683)	(8.744)	(17.244)

,

[a] In the parentheses are standard error of estimated WTPP

Table 11 lists mean and median willingness to pay for non-GM premium instead of GM counter part in several conditions. The table also exhibits standard errors of the estimated mean and median premium computed by multivariate Delta method[8] and bootstrap[9]. According to this sample, an overall mean WTPP is about 17.83, which means that consumers in Taiwan are willing to pay a non-GM premium about 12% of the average market price.

Fu *et al.* (2004) and Kaneko and Chern (2005) obtained comparable premium values of 50% and 28.4%. It is surprising that our estimate are roughly 4 and 2.5 times smaller than those of Fu and Kaneko, respectively. The reason why our overall mean of WTPP appears to be far smaller than those found in Fu *et al.* (2004) and Kaneko and Chern (2005) could be contributed to following three factors.

One reason is that the definition of valid or informative respondents is not coherent. In particular, Fu used only those respondent choosing (1)Fed with Non-GM soybean derivatives in the in the initial CV question whereas Kaneko implemented all respondents. Our method, unlike the others, used those respondents whom are classified $G = 3$. Therefore, under these different definitions, it would not be too unexpected to yield such distant results.

Another reason is that we built our analysis on modeling WTPP function directly while those of Fu and Kaneko are utility oriented. Recall that these two philosophies bring about identical consequence provided that WTP and utility functions

[8]Refer to Appendix C for more detail.

[9]The lack of closed form expression of median has led us to estimate standard error of it non-parametrically.

are both linear (Bateman, 2002). We perform a simulation study to substantiates the assertion that the expected WTPP estimated by exponential WTPP function has a inclination to be less than that estimated by linear function[10]. As a result, our exponential WTPP function possibly has a lower estimate of WTPP.

The other reason is about the sensitivity in distributions, the assumption we imposed on stochastic component. If we fit the model under with normality assumption , the overall mean of WTPP dramatically jump to 44.20 which is 29.4% of the base price. However, we assume in our case a log-generalized gamma distribution which includes normal distribution as a special case. If a normal model could fit this data well, log-generalized gamma model should not fit the data poorly. Nonetheless, it is not the case the other way round. In this study, parameter estimate suggest that the stochastic component is not likely to be normally distributed. It could be misleading if normality assumption is adopted.

[10]Refer to Appendix D for more detail.