A Selectivity Model of Household Pork Consumption Behavior in Taipei, Taiwan

Jau-Rong Li

Presented at Western Agricultural Economics Association 1997 Annual Meeting
July 13-16, 1997
Reno/Sparks, Nevada

July 1997

A Selectivity Model of Household Pork Consumption Behavior in Taipei, Taiwan

by

Jau-Rong Li*

March 1996; revised February 1997

Abstract: This paper studies branded pork consumption behavior in Taipei, Taiwan. Since branded pork is only available in some markets, the consumer decision process involves two stages; the first concerns which market to purchase, the second is what product to purchase. A bivariate probit model with sample selection is used.

JEL codes: C35, D12, C31
Keywords: Sample Selection Model, Taiwan, Pork Consumption, Agricultural Product Advertisement, Chinese Agricultural Standards.

The author is a graduate student in the Department of Economics, Utah State University, Logan, UT 84322-3530. Phone:(801)755-9460. E-mail: slnzt@cc.usu.edu

Introduction

Agricultural products are commonly seen as indifferentiated and inelastic in both supply and demand. Because of these special characteristics, it is sometimes difficult to differentiate between products supplied by different producers. Producers, producer groups and government therefore often try to promote brands with a positive image among consumers. In order to strengthen the competitiveness of domestic agricultural products in a global market, the Taiwan government set up a high-quality standard, the CAS (Chinese Agricultural Standards) mark in 1988. By this standard, firms or producers meeting strict sanitary standards are authorized to put the mark on their products, thereby, differentiating their products from lower quality competitors. Promotional advertising is supported and funded by the government. Small producers usually cannot afford advertising expenses, therefore they have particular incentives to improve product quality and gain the CAS mark. In general, the CAS mark implies a product of higher quality, and elicits higher prices.

In this study, we focus on consumers' preferences and consumption behavior of pork, both fresh and frozen. Hog production is the most important industry in Taiwan's livestock sector. It has the highest value in the domestic food market[1] and provides animal protein for national nutrition goals[2]. The hog industry also earns a large amount of foreign exchange yearly. Hence, it is crucial for government and producers to have better insight into domestic consumption patterns.

---

1 For more detailed statistics about pork, please refer to the Taiwan Agricultural Year Book ,published by Department of Agriculture and Forestry, Taiwan Province, Republic of China.

2 More information can be found in the Taiwan Food Balance Sheet, published by the Council of Agriculture, the Executive Yuan, Republic of China.

There are two kinds of pork markets in Taiwan. One is traditional markets; the other is supermarkets. Since there is no refrigeration in traditional markets, and all CAS-mark pork is either frozen or chilled, there is therefore no branded pork available in the traditional markets. However, only a portion of the frozen or chilled pork sold in supermarkets is branded and has the CAS mark. Those who shop exclusively in traditional markets will never purchase CAS-mark pork. That is, only those who choose to go to supermarkets could be observed as CAS pork purchasers. Consequently, there exists a sample selectivity problem where one tries to describe consumer preferences using survey data, either of supermarket patrons or the full consumer population. The appropriate model in this situation is a bivariate probit regression with sample selection.

One purpose of this paper is to identify the highest potential consumers for advertised and branded pork in Taiwan. Another purpose is to see if advertising affects consumer demand for pork. A comparison between a univariate probit model and a model corrected for sample selection is also presented. The paper is organized as follows. Section II develops the theoretical model and the derivation of the estimators for the sample selection model. Section III shows the primary data descriptions and some statistics. Section IV contains the empirical results of both models mentioned above. The final section offers conclusions and policy implications associated with the empirical work.

Theoretical Framework

When estimating regression models, it is commonly assumed that the sample is randomly selected. The problems associated with nonrandom sampling are widely discussed in the literature (Olsen 1980; Berk 1982; Lee 1982) . Since observations are

selected based on researchers' interest, they are generally not independent of the outcome variables, the conventional regression methods yields sample selection bias. That is, elementary statistical approaches might lead to biased and inconsistent estimates of the parameters associated with the explanatory variables.

The sample selection model was first used to analyze labor market by Heckman (1974). One example given by Heckman (1979) is the estimation of different subpopulations' earnings. One may be interested in estimating the earnings of migrants. The estimated parameters for such an equation would be biased because one can not observe the wages that nonmigrants would have earned had they decided to become migrants. In our pork consumption behavior model, the equation for deciding where to purchase pork and what kind of pork to consume both have dichotomous responses. The selection equation which predicts where consumers choose to procure pork is of significance because once the choice to go to traditional markets is made, one does not have the option to purchase CAS pork.

This paper considers the observed dichotomous responses to whether or not an individual chooses to purchase pork at a supermarket, and whether he/she chooses CAS frozen pork or not. Let Y_{1i}^{*} represents the propensity, for individual *i*, to go to the supermarket rather than the traditional market given that the respondent is responsible for purchasing food in the household. The relationship between the observed response Y_{1i} and response propensity can be represented as

$$Y_{1i} = \begin{cases} 0 & if \quad Y_{1i}^{*} \le 0 \\ 1 & if \quad Y_{1i}^{*} > 0 \end{cases}$$

Let Y_{2i} be the corresponding propensity of choosing the pork with the CAS mark. Consider a random sample of N observations. The decision-making equations for individual *i* are:

(1) $$Y_{1i}^{*} = X_{1i}\beta_1 + U_{1i}$$

(2) $$Y_{2i} = X_{2i}\beta_2 + U_{2i} \qquad (i=1....N)$$

Where X_{ji} is a $1\times K_j$ vector of exogenous regressors which influence the dependent variables and β_j is a $K_j \times 1$ vector of parameters. We make the following assumptions about the error terms : $E(U_{ji})=0$; $E(U_{ji}U_{j'i''})=\sigma_{jj'}$ when $i=i''$ and $E(U_{ji}U_{j'i''})=0$ when $i \neq i''$ and $j, j' = 1, 2$.

The assumptions on the error terms imply that an individual's decision is correlated with his previous decision, but is independent of others' decision. Assume the joint density of U_{1i} and U_{2i} is $h(U_{1i}, U_{2i})$. $Y_1 = 1$ if the respondent goes to supermarket, and $Y_1 = 0$ if not, and Y_2 is observed only when $Y_1=1$ occurs. The observations for the second equation are thus from a subsample of the population. The decision procedure can be expressed by equations (3) and (4).

(3) $$E(Y_{1i}|X_{1i}) = X_1\beta_1 \qquad i=1...N$$

(4) $$E\left(Y_{2i}|X_{2i}, \textit{sample selection rule}\right) = X_{2i}\beta_2 + E\left(U_{2i}\left|\begin{matrix}\textit{sample}\\ \textit{selection rule}\end{matrix}\right.\right)$$

If the conditional expectation of U_{2i} is zero, the regression function for the selected subsample is the same as the population regression function and sample selection bias doesn't exist and conventional estimation methods can be used on the selected sample, i.e., equation 2. The only cost of having an incomplete sample is a loss in efficiency (Heckman

1979). When the expectation of U_{2i} is not zero, however, equation (4) can be rewritten in more precise form as equations (5) and (6).

(5) $$E\left(Y_{2i}\middle|X_{2i},Y_{1i}^{*}\geq 0\right)=X_{2i}\beta_2+E\left(U_{2i}\middle|Y_{1i}^{*}\geq 0\right)$$

(6) $$E\left(Y_{2i}\middle|X_{2i},Y_{1i}^{*}\geq 0\right)=X_{2i}\beta_2+E(U_{2i}\middle|U_{1i}\geq -X_{1i}\beta_1)$$

Assume that $h(U_{1i}, U_{2i})$ is a bivariate normal density. The second term on the right hand side of equation (6) can be expressed as $E(U_{2i}\middle|U_{1i}\geq -X_{1i}\beta_1)=\beta_\lambda\lambda_i$ where $\beta_\lambda=\frac{\sigma_{12}}{(\sigma_{11})^{\frac{1}{2}}}$; $\lambda_i=\frac{\phi(Z_i)}{1-\Phi(Z_i)}=\frac{\phi(Z_i)}{\Phi(-Z_i)}$ where $Z_i=-\frac{X_{1i}\beta_1}{(\sigma_{11})^{\frac{1}{2}}}$. σ_{12} is covariance of U_{1i} and U_{2i} and σ_{11} is the variance of U_{1i}. β_λ is a scalar value population regression coefficient and reflects the degree to which selection is a problem. The higher β_λ is, the more serious the selectivity problem. ϕ and Φ are, respectively, the density and cumulative distribution functions for a standard normal variable.

The conditional regression functions for selected samples are thus:

(7) $$Y_{2i}=E\left(Y_{2i}\middle|X_{2i},Y_{1i}\geq 0\right)+V_{2i}=X_{2i}\beta+\frac{\sigma_{12}}{(\sigma_{11})^{\frac{1}{2}}}\cdot\lambda_i+V_{2i}$$

where $$E\left(V_{2i}\middle|X_{2i},\lambda_i,U_{1i}\geq -X_{1i}\beta\right)=0,$$

Heckman's estimation is the most popular selectivity bias correction method, but not the only one. There is literature proposing different approaches to cure sample selectivity problems. The Heckman estimator, like all parametric methods, imposes restrictive assumptions on the data generating process. Thus, the robustness of the estimators relies on the accuracy of the assumptions. Various studies of selectivity bias,

including semi- and nonparametric approaches, have been developed (Olsen 1980; Lee 1982; Duan and Li 1987).

Data Descriptions and Empirical Results

The data used in this study came from a survey of 547 families in Taipei calculated in 1993. Respondents were the family member with primary responsibility for purchasing food for each household. Excluding those observations with obvious recording errors or missing answers, 484 observations are included in our empirical analysis. Of these 484, 89.7% are female and 66.3% (30.0%) usually go to traditional markets (supermarkets) to purchase pork. 66.3% of those who usually go to supermarkets choose CAS frozen pork. 66.9% of the total respondents usually purchase pork in the morning. Since the traditional markets are only open in the morning, those who are used to purchasing pork in the morning could conceivably participate in both markets. By contrast, consumers holding daytime jobs have no choice but go to the supermarkets. The average weekly expenditures were 975 New Taiwan (NT) dollars for fresh pork, 350 NT dollars for frozen pork, of which 143 NT dollars was spent on frozen pork with the CAS mark. It is clear that people in Taipei are accustomed to purchasing fresh pork and spend more money on it, in aggregate. Further details about consumers' preferences for pork can be found in Li (1994).

Table 1. Monthly Household Expenditures on Pork

Expenditure (NT.$)	Frequency (%)		
	Fresh Pork	Frozen pork	CAS Frozen Pork
Never purchase	28 (7.4%)	82 (23.8%)	66 (29.1%)
Less than 200	50 (13.2%)	88 (25.5%)	66 (29.1%)
200-400	60 (15.8%)	62 (18%)	44 (19.4%)
400-600	57 (15%)	39 (11.3%)	23 (10.1%)
600-800	40 (10.6%)	14 (4.1%)	3 (1.3%)
800-1000	42 (11.1%)	12 (3.5%)	2 (0.9%)
1000-1200	23 (6.1%)	13 (3.8%)	8 (3.5%)
More than 1200	79 (20.8%)	35 (10.1%)	15 (6.6%)

Table 1 displays the frequency distributions of monthly expenditures on pork by household. More than three quarters of respondents spent less than 400 NT dollars on CAS pork per month, and more than 60% of them spend more than 400 NT dollars on fresh pork. 88.4% of respondents have heard about the CAS mark, while 20% don't clearly understand its unique attributes.

Table 2. Variable Descriptions

Variable	Descriptions
Y_1	Y_1=1 if primarily shop in supermarkets. 0 otherwise.
Y_2	Y_2=1 if choose CAS pork, 0 otherwise.
TIME	*TIME*=1 if the respondent doesn't purchase food in the morning.
GEN	Gender of the food purchaser, *GEN*=1 if male, 0 otherwise.
AGE	Age of the respondent
FR	The proportion of household monthly food expenditure to income.
HOUSE	*HOUSE*=1 if the respondent is not in job market.
OFFICE	Office=1 if the respondent is a government official, teacher, or public official.
BUSI	Busi=1 if the respondent is a business man
EDU1	*EDU1*=1 if the respondent is graduated from junior high school or below, 0 otherwise.
EDU2	*EDU2*=1 if the respondent is graduated from senior high school, 0 otherwise.
EDU3	*EDU3*=1 if the respondent is graduated from college and above, 0 otherwise.
COOK	Number of meals consumed at home each week.
PEOP	The number of people who usually eat at home together.
KNOW	*KNOW*=1 if the respondent understands the CAS mark, 0 if not.
ADV	*ADV*=1 if knew of CAS through advertising, 0 if not.

Table 2 presents the descriptions of variables to be used in the following empirical models. Y_1 and Y_2 are the corresponding variables of the two decision stages. Y_1 = 1 if the

respondent usually shops in supermarkets for pork, 0 otherwise. Y_2 = 1 if the respondent chooses the CAS-mark pork whenever she/he shops in supermarkets, 0 otherwise. The variables *FR* is the ratio between monthly food expenditure and household income. This variable represents the proportion of income which is usually spent on food. *COOK* is the number of meal cooked at home per week and *PEOP* is the number of family members. Larger ratio between income and food, higher frequency of cooking at home and family size imply a particular life style of each household and thus may have some effect on consumption behavior. The *TIME* variable indicates when the respondent usually shops for food. This variable may be especially important in our model due to the limited operating time of traditional markets, which are open only in day time. Those who usually shop at night generally cannot go to traditional markets. *KNOW* and *ADV* are two independent variables associated with the knowledge of CAS mark and acceptance of advertising information. The estimates of these two variables enable us to draw inference about the effects of advertising on agricultural products. The *OFFICE*, *BUSI* and *HOUSE* variables represents the occupation of the respondent. We specify the four group, i.e., officials, businessmen, housewives and others, because the CAS is promoted and supported by government, thus those who work in the government institutes may have higher tendency to cooperate. Businessmen may be busier and pay less attention to advertising or other promotion activities. Other variables are basic demographic characteristics of the respondents, e.g., education levels and gender.

Table 3 shows the separate univariate probit regression results for the two decision stages. This analysis does not take into account the non-random sampling mechanism that generates an affirmative response to equation 2, nor the fact that the equations' residuals

are highly likely correlated. Nonetheless, the results in Table 3 are interesting comparative tools. Both models have a high share of correct predictions and the log-likelihood ratio tests show that the joint effects of the independent variables are significant for both equations.

Table 3. Separate Univariate Probit Regressions for Market and CAS Pork Selections

	First Stage			Second Stage		
X_1	Coefficients	t-statistic	Marginal Effect[3]	Coefficients	t-statistic	Marginal Effect
Constant	1.335	2.088**		-3.182	-2.873**	
TIME	0.710	5.045**	0.272	0.058	0.222	0.022
GEN	0.378	1.730*	0.145	-0.196	-0.585	-0.074
AGE	-0.043	-3.484**	-0.016	-0.001	-0.032	0.000
EDU1	0.095	0.327	0.036	1.098	1.586	0.415
EDU2	0.145	0.841	0.056	0.544	1.669*	0.206
EDU3	0.387	2.041**	0.148	-0.128	-0.403	-0.048
FR	0.561	1.044	0.215	-0.782	-0.530	-0.296
HOUSE	0.750	-2.606**	0.287	1.122	2.213**	0.424
OFFICE	-0.094	3.063**	-0.036	0.446	1.150	0.169
BUSI	-0.053	-0.393	-0.020	0.345	0.838	0.131
COOK	-0.039	-1.850*	-0.015	0.071	1.466	0.027
PEOP	-0.138	-0.763	-0.053	0.062	0.578	0.024
ADV	0.131	-0.636	0.050	1.581	3.647**	0.598
KNOW	0.109	0.761	0.042	1.055	3.321**	0.399
LR-statistic		126**		56.32**		
Correct Predictions		74.57%		73.5%		

Note: * significant at α=10%. ** significant at α=5%.

The results of the bivariate probit regression with sample selection are shown in Table 4. In order to get efficient estimates, maximum likelihood is used to estimate the coefficients in the bivariate selectivity probit model. There are some points worth noticing in this table. First, the correlation between the disturbances of the two equations is -0.98, which suggests there does exist a serious sample selectivity problem.[4] Therefore,

[3] Marginal effects are estimated on variable means.

[4] Bek and Ray (1982) suggest that a correlation over 0.8 suggests serious sample selection problems.

inattention to sample selectivity bias may lead to incorrect conclusions. Second, comparing the second stage results in Table 4 with those in Table 3, the marginal effects of *ADV* and *KNOW* are lower if model is correctly specified. That is, when one mistakenly specifies consumer choice as an uncorrelated univariate probit model, one will overstate the effects of advertising and knowledge on consumer purchasing behavior with respect to branded pork.

Table 4. Bivariate Probit Regressions with Sample Selection Results

	First Stage			Second Stage		
	Coefficient	t-statistic	Marginal Effect	Coefficient	t-statistic	Marginal Effect
Constant	1.652	2.625**		-2.005	-1.909*	
TIME	0.646	4.483**	0.219	-0.278	-1.197	-0.065
GEN	0.497	2.132**	0.168	-0.318	-1.136	-0.074
AGE	-4.98E-02	-4.552**	-0.017	1.38E-02	0.894	0.003
EDU1	8.68E-02	0.259	0.029	0.676	1.131	0.158
EDU2	0.141	0.767	0.048	0.286	0.950	0.067
EDU3	0.371	1.870*	0.126	-0.259	-0.955	-0.061
FR	0.587	0.944	0.199	-0.932	-0.675	-0.218
HOUSE	-0.677	-2.637**	-0.230	1.238	2.812**	0.290
OFFICE	0.723	2.825**	0.245	-0.123	-0.385	-0.029
BUSI	-0.165	-0.643	-0.056	0.280	0.866	0.066
COOK	-5.01E-02	-1.647*	-0.017	8.36E-02	1.929*	0.020
PEOP	-4.03E-02	-0.684	-0.014	7.40E-02	0.724	0.017
ADV	-0.173	-0.747	-0.059	1.303	2.516**	0.305
KNOW	0.158	0.844	0.054	0.647	2.534**	0.151
ρ_{12}	-0.981	-4.301**				

Note: * significant at $\alpha = 10\%$. ** significant at $\alpha = 5\%$.

The corrected estimation results convey useful information. In the first stage, which estimates the probability of an individual purchasing pork in supermarkets, *TIME*, *GEN*, *EDU3*, and *OFFICE* have significantly positive effects. This implies that those who do not or cannot purchase food in the morning have a higher probability of choosing supermarkets as a place to purchase pork. This is likely due to the business hours of traditional markets, since traditional markets are open in the morning only. Men also have

higher probability than women to shop in supermarkets. This is probably due to the images of traditional markets: dirty, noisy and crowded. Moreover, better educated respondents and the households with higher proportion of food expenditure tend to shop more in supermarkets. Government officials, teacher or public officials also have higher propensities to purchasing pork in supermarkets. Therefore, the officials and teachers become highly potential supermarket customers. The results suggest that it may be more efficient for the government and pork processors to emphasize the promotion of the CAS pork on the latent supermarket consumers. On the other hand, *AGE*, *COOK* and *HOUSE* have negative effects on the probability. The older the purchaser, the less likely is he/she to choose supermarkets. Those families with more meals cooked at home also have lower probability to go to supermarkets. Those who are not employed in the job market have a higher tendency to go to traditional markets. Supermarket grocery shopping is a normal good heavily influenced by consumer demographics.

At the second stage, once the consumer has decided to go to supermarkets, his information about the CAS mark affects his decision as to what sort of pork to purchase. The *HOUSE* and *COOK* variables have significant positive effects, while these two have negative effects in the first stage. This indicates that once a housewife chooses to go to supermarkets, she is more likely to purchase the CAS-mark pork. Similarly, the households in which more meals are cooked at home have higher probabilities to choose CAS marked pork. That is, once those low potential supermarket shoppers decide to go to supermarkets, they appear to be high potential CAS pork purchasers. The results also indicate that both advertising and an understanding of the CAS mark have a positive

influence on consumer's probability of buying branded pork. In addition, officials are more likely to choose CAS pork.

Conclusions

A unique feature of this study is the application of an appropriate correction method for the nonrandom sample selection mechanism. We found the univariate probit regression applied to the purchase decision may overestimate the marginal effects of some explanatory variables on the selected equation, most notably the impact of advertising on the probability that a consumer purchases branded pork. Thus, we explicitly incorporate the fact that one cannot buy CAS pork outside supermarkets, using a bivariate model with correction for sample selection. Since the correlation between the disturbances of the first equation and the selection equation is extremely strong, there exist significant differences between the corrected model and the uncorrected one.

Some implications of this study are as follows. First, most people in Taipei still purchase pork in traditional markets. Thus the government may need to introduce CAS pork into traditional markets if further adoption of high quality pork is desired. Second, our research found that advertising does promote CAS pork consumption. Government or producer groups might, however, target promotion activities more precisely at high potential consumers, including males, younger and better educated food purchasers, the families in which the mother/wife works full time, and those who are unable to shop in the morning.

References

Berk, R.A. and S.C. Ray. 1982. "Selection Biases in Sociological Data" *Social Science Research*, 11:352-98.

Council of Agriculture, Executive Yuan, Republic of China. 1993. *Taiwan Food Balance Sheet*.

Davis, G.C. and M.K. Wohlgenant. 1993. "Demand Elasticities from A Discrete Choice Model: The Natural Christmas Tree Market". *American Journal of Agricultural Economics*. 75:730-38.

Department of Agriculture and Forestry, Taiwan province, Republic of China. 1993. *Agricultural Year Book.*

Duan, N., and K.C. Li.. 1987, "Distribution-Free and Link-Free Estimation for the Sample Selection Model." *Journal of Econometrics.* 35:25-35.

Gronau, R.. 1974. "Wage Comparisons - A Selectivity Bias" *Journal of Political Economy.* 82:1119-43.

Heckman, J.J.. 1974. "Shadow Price, Market Wages, and Labor Supply" *Econometrica* 42:679-94.

_____. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variable and a Simple Estimator for Such Models", *The Annals of Economic and Social Measurement*, 5:475-92.

_____. 1979. "Sample Selection Bias as a Specification Error" *Econometrica* 47(1): 153-60.

Lee, L.F.. 1982. "Some Approaches to the Correction of Selectivity Bias" *Review of Economic Studies* 49:355-72.

Li, J.R.. 1994 "An Analysis on Consumer Behaviors of Pork- A Case Study of Taipei, Taiwan" *Master Thesis*, National Taiwan University.

Olsen, R.J.. 1980. "A Least Squares Correction for Selectivity Bias". *Econometrica* 48(7):1815-20