Semiparametric Estimation of Consumer Demand Systems with Micro Data

Abdoul Sam
Assistant Professor
Department of Agricultural, Environmental and Development Economics
The Ohio State University
Contact: sam.7@osu.edu

Yi Zheng
Ph.D. student
Department of Agricultural, Environmental and Development Economics
The Ohio State University
Contact: Zheng.485@osu.edu

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Portland, OR, July 29-August 1, 2007

Abstract

This article proposes a semiparametric two-step procedure for estimating a censored consumer demand system with micro data. The semiparametric estimator considered in the first step is suggested by Klein and Spady (1993). This estimator, used as a counterpart of the probit estimator in a conventional two-step model, does not make any distributional assumptions about the disturbances and so is exempt from model misspecification and plausible heteroscedasticity. In the second step, we motivate the choice of the Almost Ideal Demand System (AIDS) as an economic representation of consumers' demand behavior. Implementing our proposed semiparametric two-step procedure as well as Shonkwiler and Yen (1999)'s two-step model to a household meat consumption dataset from China generates the price and expenditure elasticities of demand. We also conducted the Horrowitz and Härdle (1994)'s specification test to our data and reject the null.

Key words: censoring, semiparametric estimator, consumer demand system, food expenditures.

1 INTRODUCTION

The increased reliance on cross-sectional household-level micro data to estimate consumer demand equations has spawned a growing literature on the econometric treatment of the censoring of dependent variables, which occurs when one or more commodities have a significant proportion of zero expenditures. Theorists have proposed full information maximum likelihood (FIML) models which account for the left censoring of the dependent variables in a system of equations (Wales and Woodland (1983), Lee and Pitt (1986, 1987), Amemiya (1974), Chiang and Lee (1992)). However, the practical potential of the FIML approach for the estimation of demand systems is limited by its computational cost when censoring occurs for several commodities, as it requires the evaluation of multidimensional integrals.

Less efficient methods in the same realm as Heckman's two-step sample selection approach (1979) have been proposed as computationally expeditious alternatives to FIML estimators (Heien and Wessells (1990), Shonkwiler and Yen (1999), Yen (2005), Yen and Lin (2006)). In these methods, probit regressions which determine the probabilities that households will make a purchase are obtained from a binary censoring rule. These probit regressions are used to compute the inverse mill ratio for each household, which are then inserted in the second step as instrumental variables. These methods are straightforward to implement and thus have gained significant attention in applied work. However, these Heckman-type approaches rely on a critical assumption that the error processes follow a joint normal distribution to recover consistent estimates of the demand system and therefore are prone to distributional misspecification. Specifically, when the underlying distribution between the error processes is normal then these methods yield estimates that are $\sqrt{n}$-consistent. On the other hand, if the wrong joint distribution is assumed then the parameter estimates are O(1). Furthermore, these Hackman-type models assume homoscedasticity in the disturbances, which is not always true especially in cross-sectional data. If heteroscedasticity emerges in the error terms, not surprisingly these approaches may yield erroneous elasticity estimates with potentially significant economic implications.

Drawing from recent advances in the nonparametric econometrics literature, this article proposes a semiparametric approach for the estimation of censored demand systems that is similar spirit to Heckman-type estimators but is exempt from distributional misspecification and accounts for potential heteroscedasticity in the disturbances. The suggested semiparametric approach consists of two steps of estimation. In the first step, a semiparametric estimator proposed by Klein & Spady (1993) is adopted as a counterpart of the probit estimator used in conventional Hackman-type procedures. The Klein and Spady (1993)'s estimator is both consistent and achieves the semiparametric efficiency bound, thus it has been applied in several empirical studies (Newey, Powell, and Walker (1990), Martins (2001)). Similar to Shonkwiler and Yen (1999)'s two-step method, in the second stage, the semiparametrically estimated link function as well as the index computed from the latent parameter estimates are incorporated in the demand equations which are then estimated by seemingly uncorrelated regression (SUR).

This paper is organized as follows. Our proposed semiparametric estimation model is constructed and explained in Section 2. Section 3 presents an empirical analysis of a consumer demand system with censored data. Specifically, the proposed semiparametric two-stage procedure as well as Shonkwiler and Yen (1999)'s parametric procedure are implemented using a cross-sectional dataset of 1,237 households from the Hainan province in China. Elasticity estimates are computed with respect to two procedures and then are comparatively discussed in aspects of their economic implications. Concluding comments are presented in Section 4. Introduction and explanation about Horrowitz and Hardle (1994)'s test are in Appendix B.

2 METHODOLOGY

We consider the standard empirical framework for a censored demand system, i.e.

$$\begin{aligned} Y_{ij} &= d_{ij}\left(g(X_{ij},\beta_j)+\epsilon_{ij}\right) \\ d_{ij} &= I(W'_{ij}\gamma_j+v_{ij}>0), \quad \text{for } i=1,2,...,n;\ j=1,2,...,J \end{aligned}$$

where $I(\omega)$ denotes an indicator function of the event ω, X_{ij} and W_{ij} are vectors of design variables for the j^{th} equation, Y_{ij} and d_{ij} are the response variables, β_j and γ_j are the model parameters, and ϵ_{ij} and v_{ij} are zero-mean and finite variance error processes. The unconditional mean of Y_{ij} is

$$\begin{aligned} E(Y_{ij}|X_{ij},W_{ij}) &= E(Y_{ij}|X_{ij},W_{ij};d_{ij}=1)Prob(d_{ij}=1) \\ &= \left(g(X_{ij},\beta_j)+E(\epsilon_{it}|v_{ij}>-W'_{ij}\gamma_j)\right)F_j({W_{ij}}'\gamma_j) \\ &= \left(g(X_{ij},\beta_j)+\lambda({W_{ij}}'\gamma_j)\right)F_j({W_{ij}}'\gamma_j) \qquad (1) \end{aligned}$$

where $F_j(W_{ij}'\gamma_j)$ is the unknown cumulative distribution function (link function) of the error term v_{ij}. It follows from (1) that

$$Y_{ij} = \left(g(X_{ij}, \beta_j) + \lambda(W_{ij}'\gamma_j)\right) F_j(W_{ij}'\gamma_i) + \eta_{ij}, \quad i = 1, 2, ...n; j = 1, 2, ...J \tag{2}$$

where $\eta_{ij} = Y_{ij} - E(Y_{ij}|X_{ij}, W_{ij})$. Let $\Phi(\cdot)$ and $\phi(\cdot)$ denote respectively the standard normal cumulative distribution and probability density functions. If the errors, ϵ_{ij}, v_{ij} are assumed to follow a bivariate normal distribution with covariance θ_j, the system of equations (2) becomes

$$Y_{ij} = \left(g(X_{ij}, \beta_j) + \theta_j \frac{\phi(W_{ij}'\gamma_j)}{\Phi(W_{ij}'\gamma_j)}\right) \Phi(W_{ij}'\gamma_i) + \eta_{ij}, \quad i = 1, 2, ...n; j = 1, 2, ...J \tag{3}$$

which corresponds to the system of demand equations derived by Shonkwiler and Yen. They propose that the β_j's in (3) be estimated in two steps. First, estimate γ_j by Probit to obtain $\widetilde{\gamma}_j$; then estimate the J equations (3) jointly as a system of seemingly unrelated regressions (SUR) after substituting $\widetilde{\gamma}_j$ for γ_j. As mentioned above, Shonkwiler and Yen's approach produces inconsistent estimates when the true unknown joint distribution departs from the normal.

Instead of assuming joint normality of the disturbances, our proposed approach utilizes Klein and Spady (1993)'s semiparametric method in the first step to estimate *both* the link function $F_j(.)$ and the parameter vector γ_j for each censored equation. The Klein and Spady estimator is semiparametric in that it does not make any assumption about the distribution of the error term in the binary selection equation, instead it estimates the distribution function nonparametrically using the Kernel method. However, it assumes a linear index function to circumvent the curse of dimensionality common to nonparametric approaches. Briefly, the Klein and Spady estimator of γ_j is obtained by maximizing the quasi-likelihood function

$$\ell(\gamma_j) = n^{-1} \sum_{i=1}^{n} \Big(d_{ij}\ log(\widehat{F_j}(W'_{ij}\ \gamma_j)) + (1 - d_{ij})\ log(1 - \widehat{F_j}(W'_{ij}\ \gamma_j) \Big)$$

where

$$\hat{F}_j(v_{sj}) = \frac{\sum_{l=1}^{n} d_{lj} K_h(v_{sj} - v_{lj})}{\sum_{l=1}^{n} K_h(v_{sj} - v_{lj})}, \qquad v_{sj} = W'_{sj}\ \gamma_j, \qquad K_h(u) = 1/h\ K(u/h)$$

and h is a non-stochastic smoothing parameter (see Klein and Spady for technical details). Klein and Spady show that the resulting estimator, $\widehat{\gamma}_j$, is both consistent and efficient. What's more, the KS estimator accommodates a certain form of heteroscedasticity by which the probit model is inconsistent.

Our two-step approach to estimate the demand system (2) proceeds as follows. First, obtain the estimates of γ_j and the link function $F_j(.)$ using Klein & Spady's (1993) method for each censored equation. Second, $\widehat{\gamma}_j$ and $\widehat{F_j}(\widehat{v_{ij}})$ are substituted for γ and $F_j(v_{ij})$ respectively in (2), and following Newey (1991) and Fraga and Martins (2001) $\lambda(\widehat{v_{ij}})$ is approximated with a series based on orthogonal polynomials of the first-stage index, i.e. $\lambda(\widehat{v_{ij}}) \simeq \sum_{k=1}^{K} \alpha_k \widehat{v_{ij}}^{k-1}$ where $\{\alpha_k\}$ are unknown coefficients. Hence, the second step consists of estimating the following system of nonlinear equations

$$Y_{ij} = \left(g(X_{ij}, \beta_j) + \sum_{k=1}^{K} \alpha_k \widehat{v_{ij}}^{k-1} \right) \widehat{F_j}(\widehat{v_{ij}}) + \xi_{ij}, \quad i = 1, 2, ...n; j = 1, 2, ...J \qquad (4)$$

3 EMPIRICAL ANALYSIS

3.1 *The Data Set*

In this section we apply our proposed econometric model developed in section 2 using a survey of household meat consumption carried out by the National Statistical Bureau in China (2003) over a one-year period. The survey, which was undertaken in Hainan Province of China, contains information on the purchases of various types of meat by each household, together with information on the characteristics of the household members. As in Yen & Lin (2006), we limit our empirical analysis to four popular meat products: beef, pork, fish, and poultry. The resulting sample data set contains 1,237 urban households. Pork and poultry are consumed by nearly all (over 99%) households in the sample, while about 93.5% of sample consume fish and 50.8% of the sample consume beef during the year. From the reported expenditure and quantity of each meat product consumed, price was derived as the unit value. Missing prices for nonconsuming households were replaced by regional averages as in Yen and Lin (2006). In addition to expenditure and prices, we also have three demographic variables which are respectively the number of wage earners in a household, educational level of household head, and household size. Definitions of variables and sample descriptive statistics are presented in table 1. It appears that pork is the most consumed food while beef is the most expensive one on average.

3.2 *Empirical Results for the Selection Equations*

We estimate two selection equations, one for beef, and the other for fish by Probit and KS estimators respectively. The dependent variables in the selection equations are dichotomous variables that take the value 1 if the household makes a purchase and zero otherwise. Beside of the expenditure and price explanatory variables, we also include the three demographic variables. Consequently, the specification of the first step selection equations are given by

$$w_j = \gamma_0 \log x + \gamma_1 \log p_b + \gamma_2 \log p_p + \gamma_3 \log p_f + \gamma_4 \log p_t + \gamma_5 NOWE + \gamma_6 HSIZE + \gamma_7 EDUC \tag{5}$$

where j represents beef and fish for which censoring occurs substantially (50.8% and 93.5% respectively). The estimation results are presented in table 2.

The probit estimates of the three demographic (non-price) variables are not significant at any conventional level except that the estimate of the education of household head is significant at the 5% level in the selection equation for fish. It is plausible that educated people consider the dietary benefits of eating fish. On the other hand, the Klein and Spady results indicate that the number of wage earners and household size are significant at the 1% level for purchase of beef. Besides, the KS estimate of the number of wage earners is also significant at the 5% level for purchase of fish. Both probit and KS estimates suggest that the household expenditure on meat and the price of beef are significant at the 1% level in the selection equation for beef and also for fish. Additionally, it appears that whether or not to consume fish does not depend on the price of fish instead it depends on the price of beef and the household total expenditure, however whether to consume beef depends on almost everything.

Compared to the probit estimation results, the KS estimates have considerably small variances which reveals higher efficiency of the KS estimator, but both two types of estimates suggest that the total expenditure has a positive below unity coefficient which indicates that the increasing total expenditure raises the probability of consumer purchasing beef and fish but more effectively with fish rather than beef. The significant (suggested by KS estimates) demographic variable, the number of wage earners in households brings down the probability of consumers purchasing beef but promotes the probability of purchasing fish. Another significant (also suggested by KS estimates) demographic variable, the size of household has a negative below unity coefficient to the selection of beef, which demonstrates that larger-size households have less probability to purchase beef.

To determine whether the normal distribution assumption made by the probit model is consistent with our data, we utilized the specification test proposed by Horrowitz and Härdle (1994). The test is based on the distance between the KS estimator and its probit counterpart, specifically the difference between the probit link, Φ and the nonparametric regression curve, F. Under the null hypothesis that the link function is specified correctly as a standard normal cumulative distribution function, the test statistics has the form

$$T_n = \sqrt{h} \sum_{i=1}^{n} u(x_i' \widehat{\beta}) \{ Z_i - \Phi(x_i' \widehat{\beta}) \} \{ \widehat{F}_i(x_i' \widehat{\beta}) - \Phi(x_i' \widehat{\beta}) \}$$

and asymptotically follows a normal distribution with zero mean and variance σ_T^2 (see Appendix for details). As mentioned above, $\widehat{F}_i(x_i'\widehat{\beta})$ is the CDF estimated by the KS method, $\Phi(x_i'\widehat{\beta})$ is the CDF estimated by the probit model, h is the bandwidth used in the semiparametric regression and chosen by cross validation, $u(x_i'\widehat{\beta})$ is a weighting function that downweights extreme index values, and Z_i is the binary dependent variable. The asymptotic variance σ_T^2 is replaced by a consistent estimate (see Appendix for its estimator).

The test results are presented in table 3. As expected, the probit link is clearly rejected at the 1% confidence level for both selection equations of beef and fish. Additionally, visual implication of the CDF and PDF plots (figure 1 and 2) show differences between the probit and KS estimates. Noticeably the plot for beef is bimodal based on KS estimates, a feature that cannot be captured by probit estimates. We conclude that the normality assumption of the probit model is not consistent with our data. In this way, using more sophisticated (semiparametric) approaches is necessary and can be more informative and reliable than a standard parametric approach.

3.3 *Estimated Demand Elasticities*

In the second stage, we estimate equation (2) using the AIDS functional form of the demand system $g(X_{ij}, \beta_j)$, specified as follows.

$$w_{ij} = a_j + b_j(\log \frac{x_i}{P}) + \sum_{k=1}^{J} \gamma_{jk} \log p_{ik}, \text{ for } i = 1, 2, ...n;\ j = 1, 2, ...J; k = 1, 2, ...J \quad (6)$$

$$\log P = a_0 + \sum_{j=1}^{J} a_j \log p_j + .5 \sum_{j=1}^{J} \sum_{k=1}^{J} \gamma_{jk} \log p_j \log p_k \quad (7)$$

where w_{ij} is the ith household's expenditure share of commodity j, x_i is the ith household's total expenditure, p_j stands for the price of jth commodity, and P is a price index specified as in (7). Incorporating equation (6) into (3) and (4) respectively gives the two estimating systems as

$$w_{ij}^p = \left(a_j + b_j(\log\frac{x_i}{P}) + \sum_{k=1}^{J} \gamma_{jk} \log p_{ik} + \theta_j \frac{\phi(W_{ij}'\widetilde{\gamma}_j)}{\Phi(W_{ij}'\widetilde{\gamma}_j)} \right) \Phi(W_{ij}'\widetilde{\gamma}_j) + \xi_{ij}^p \tag{8}$$

$$w_{ij}^s = \left(a_j + b_j(\log\frac{x_i}{P}) + \sum_{k=1}^{J} \gamma_{jk} \log p_{ik} + \sum_{k=1}^{K} \alpha_k (W_{ij}'\widehat{\gamma}_j)^{k-1} \right) \widehat{F_j}(W_{ij}'\widehat{\gamma}_i) + \xi_{ij}^s \tag{9}$$

where $\xi_{ij}^p = w_{ij}^p - E(w_{ij}^p)$ and $\xi_{ij}^s = w_{ij}^s - E(w_{ij}^s)$.We choose the number of polynomials contained in $\lambda(\widehat{v_{ij}})$ to be 2 by cross-validation method. Estimating (8) and (9) by Iterated Seemingly Unrelated Regression (ITSUR) yields parameter estimates which can be used to derive the demand elasticities.[1]

Because of the two-step estimation procedure, it is well known that the standard errors need to be adjusted to account for the added randomness due to the first step estimation. We circumvent this issue by bootstrapping our sample. Specifically, we obtained 100 bootstrap samples from our data; performed our multi-step estimation for each sample; and constructed standard error estimates for our parameters from the resulting distribution of bootstrapping parameter estimates. Table 4 presents the parametrically estimated elasticities and their standard errors, calculated by the *bootstrap method*. The semiparametrically estimated elasticities and their bootstrapping standard errors are in table 5.

As seen in table 4, Shonkwiler and Yen's parametric estimation results suggest that all uncompensated own-price elasticities are negative, below unity (except that pork has a subtly above unity own-price elasticity), and significant at the 1% level. All significant (at the

[1]Demand elasticities are calculated by differentiating the unconditional mean of expenditure shares.

10% level or lower) uncompensated cross-price elasticities are negative (except between beef and pork), suggesting gross complementarity among the meat products. The uncompensated cross-price elasticities between pork and fish, and between pork and poultry are not significant. Expenditure elasticities are below unity for beef, pork, and fish but above unity for poultry, which indicates that the first three meat products are normal goods but isn't poultry. Unlike uncompensated cross-price elasticities, the significant (at the 1% level) compensated elasticities indicate net substitution between beef and pork, between fish and pork and between poultry and pork, and net complementarity between beef and fish and between beef and poultry. All compensated own-price elasticities are negative and significant at the 1% level, and also smaller than their uncompensated counterparts due to the positive expenditure elasticities. The semiparametric estimation results shown in table 5 suggest very similar statements about both the uncompensated and compensated price elasticities but very different expenditure elasticities, i.e. the total expenditure elasticities are above unity for beef, fish, and poultry but only below unity for pork.

4 CONCLUDING REMARKS

The use of micro survey data has been popular in estimating consumer demand equations, thus interest in censored data has continued to grow. For the application such that zero observations occur in one equation, direct ML estimation of the Tobit model would be straightforward under the normality assumption. For a large system with many censored equations a two-step estimator though statistically inefficient, is a computationally expeditious alternative to the full information ML estimator as it avoids evaluating multidimensional integrals. However, the conventional two-step procedure generates inconsistent estimates if wrong joint distribution is assumed. This paper contributes to the censored demand system literature by incorporating the recently advanced semiparametric estimation methodology to the conventional two-step econometric framework. This semiparametric methodology appears particularly attractive in model specification regarding the underlying distribution generating the disturbances and in its ability to accommodate a certain form of heteroscedasticity which likely happens in cross-sectional data.

The proposed semiparametric two-step model is applied to an empirical analysis with a survey data set of meat product consumption in China (2003). For the demand system where only a subset of equations is censored (beef and fish), selectivity terms are included only for equations with zero observations. The AIDS functional form of the demand system was used to obtain elasticity estimates. Although the proposed semiparametric and Shonkwiler and Yen's procedure produce very similar price elasticities for the current application, the differences among these models are worthy of further investigation in other applications.

5 APPENDIX A: Tables and Figures

Table 1. Variable Definitions and Sample Statistics (Sample Size: 1,237)

Variable	Mean	Std Dev
Quantities (Kg. per person per annum)		
Beef (Consuming households: 50.8 % of sample)	2.64	5.60
Pork	42.90	27.32
Fish (Consuming households: 93.5%of sample)	11.76	13.33
Poultry	18.34	17.52
Expenditures (Yuan per person per annum)		
Beef (Consuming households: 50.8% of sample)	36.66	77.46
Pork	461.82	290.69
Fish (Consuming households: 93.5% of sample)	83.29	97.97
Poultry	208.67	187.75
Prices (Yuan/Kg.)		
Beef	14.38	2.404
Pork	10.83	1.12
Fish	7.32	1.96
Poultry	12.16	3.75
NOWE (number of wage earners)	1.48	0.89
HSIZE (size of household)	3.05	0.87
EDUC (educational level of household head)	5.34	1.63

Source: Urban Household Survey, China's National Statistical Bureau, 2003.

Table 2. Estimates for the Sample Selection Model

	Probit				Klein-Spady			
	Beef		Fish		Beef		Fish	
Variables	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Intercept	-3.740***	1.053	-0.562	1.739				
lx	0.600***	0.065	0.735***	0.096	0.163***	0.015	0.505***	0.056
lpb	-0.836***	0.200	-0.768**	0.404	-0.073***	0.009	-0.365***	0.112
lpp	0.623*	0.391	-0.287	0.598	0.037*	0.023	-0.092	0.109
lpf	0.675***	0.175	-0.379*	0.285	0.113***	0.011	-0.103	0.110
lpt	-0.261**	0.134	0.111	0.199	-0.019***	0.007	0.041	0.043
NOWE	-0.025	0.050	0.001	0.083	-0.020***	0.004	0.059**	0.028
HSIZE	0.000	0.049	0.090	0.086	-0.036***	0.004	0.003	0.021
EDUC	-0.011	0.024	0.073**	0.038				

Note: 1. Triple(***), double(**), and single(*) asterisks indicate significance at the 1%, 5% and 10% levels, respectively.

2. The intercept cannot be identified by nonparametric estimators; the last predictor variable is fixed at its probit estimate.

Table 3. Results for Horrowitz and Härdle Test

	Statistic	p-value
Beef	-2.37	0.01
Fish	935	0.00

Table 4. Parametric Elasticity Estimates

	Price of				Total
Product	Beef	Pork	Fish	Poultry	Expenditure
Uncompensated elasticities					
Beef	-0.73***	0.46**	-0.06	-0.37***	0.68***
	(0.10)	(0.21)	(0.17)	(0.13)	(0.10)
Pork	0.04*	-1.02***	0.02	-0.01	0.97***
	(0.03)	(0.06)	(0.02)	(0.03)	(0.02)
Fish	-0.39***	-0.09	-0.62***	-0.33***	0.96***
	(0.13)	(0.13)	(0.10)	(0.07)	(0.05)
Poultry	-0.15***	-0.08	-0.11***	-0.71***	1.06***
	(0.05)	(0.08)	(0.03)	(0.07)	(0.05)
Compensated elasticities					
Beef	-0.70***	0.87***	0.01	-0.19*	
	(0.10)	(0.21)	(0.16)	(0.13)	
Pork	0.08***	-0.44***	0.12***	0.24***	
	(0.03)	(0.06)	(0.02)	(0.03)	
Fish	-0.35***	0.48***	-0.53***	-0.08	
	(0.13)	(0.13)	(0.10)	(0.07)	
Poultry	-0.11**	0.55***	0.00	-0.43***	
	(0.05)	(0.08)	(0.03)	(0.07)	

Note: Bootstrapping standard errors are in parentheses. Triple(***), double(**), and single(*) asterisks indicate significance at the 1%, 5% and 10% levels, respectively.

Table 5. Semiparametric Elasticity Estimates

	Price of				Total
Product	Beef	Pork	Fish	Poultry	Expenditure
Uncompensated elasticities					
Beef	-0.94***	0.51**	-0.89*	-0.42**	1.05***
	(0.17)	(0.27)	(0.58)	(0.19)	(0.10)
Pork	0.04*	-1.01***	0.01	-0.01	0.97***
	(0.03)	(0.06)	(0.02)	(0.03)	(0.02)
Fish	-0.06	-0.10	-0.69***	-0.36***	1.03***
	(0.10)	(0.13)	(0.10)	(0.07)	(0.06)
Poultry	-0.17***	-0.10	-0.14***	-0.69***	1.10***
	(0.06)	(0.08)	(0.03)	(0.07)	(0.06)
Compensated elasticities					
Beef	-0.89***	1.13***	-0.78*	-0.15	
	(0.17)	(0.26)	(0.59)	(0.19)	
Pork	0.09***	-0.44***	0.11***	0.24***	
	(0.03)	(0.06)	(0.02)	(0.03)	
Fish	-0.01	0.51***	-0.59***	-0.10*	
	(0.10)	(0.13)	(0.10)	(0.07)	
Poultry	-0.13***	0.55***	-0.03	-0.40***	
	(0.05)	(0.07)	(0.03)	(0.08)	

Note: Bootstrapping standard errors are in parentheses. Triple(***), double(**), and single(*) asterisks indicate significance at the 1%, 5% and 10% levels, respectively.

Figure 1: Cumulative Distribution Function

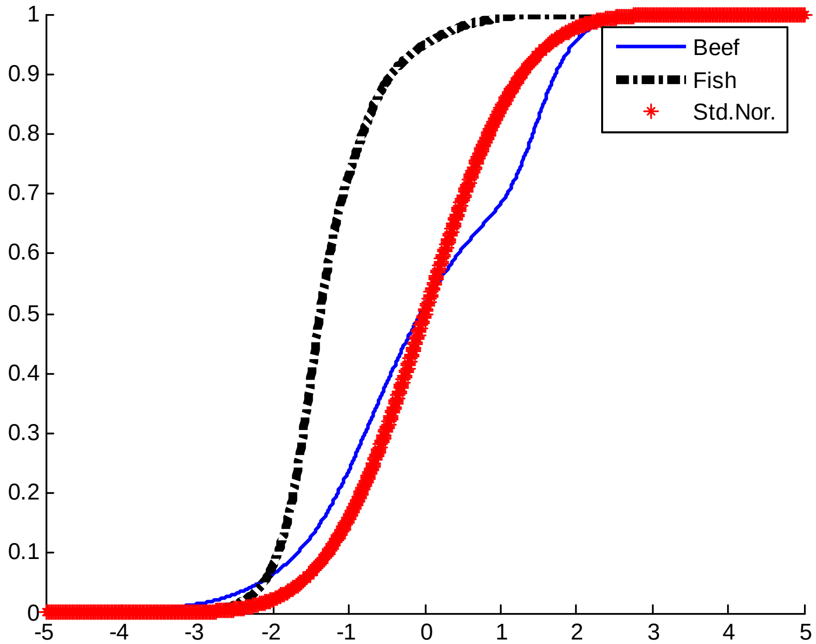
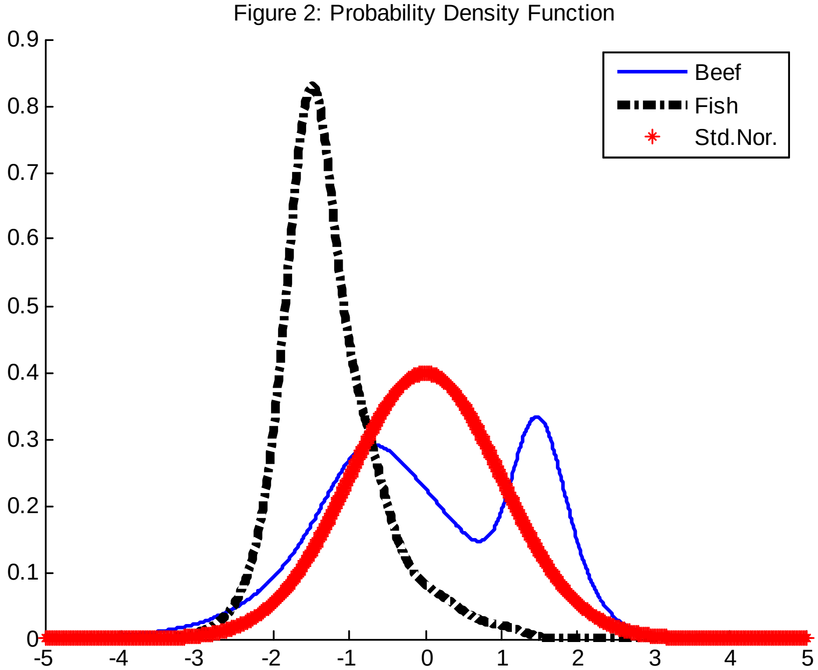


Figure 2: Probability Density Function



6 APPENDIX B: Horrowitz and Härdle Test (1994)

Horrowitz and Härdle (1994) proposed a procedure for testing the adequacy of a probit (parametric) model against a semiparametric alternative that can be used for binary response models. In this paper, the authors suggest testing the specification of a single-index model according to the hypothesis:

$$\begin{aligned} H_0 &: \quad E(Z|X'\beta = v) = F(v) \\ H_1 &: \quad E(Z|X'\beta = v) = H(v) \;\; where\ H(v)\ is\ an\ unknown\ function \end{aligned}$$

When the link is a probit one, under the null and some regularity conditions the test statistic has the following property

$$T = \frac{T_n}{\widehat{\sigma_T}} = \frac{\sqrt{h}\sum_{i=1}^{n} u(x_i'\widehat{\beta})\{Z_i - \Phi(x_i'\widehat{\beta})\}\{\widehat{F}_i(x_i'\widehat{\beta}) - \Phi(x_i'\widehat{\beta})\}}{\sqrt{\frac{2C_k}{n}\sum_{i=1}^{n}\frac{\{u(x_i'\widehat{\beta})\widehat{F}_h(x_i'\widehat{\beta})[1-\widehat{F}_h(x_i'\widehat{\beta})]\}^2}{\widehat{P_h}(x_i'\widehat{\beta})}}} \sim N(0,1)$$

where

$$C_k = \int_{-\infty}^{\infty} K(x)^2 dx = \int_{-\infty}^{\infty} \phi(x)^2 dx = \int_{-\infty}^{\infty} (\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2})^2 dx = \frac{1}{2\sqrt{\pi}}$$

$$u(x_i'\widehat{\beta}) = \{ {}^{1 \quad if \quad 98\%\ of\ x_i'\widehat{\beta}}_{0 \quad else}$$

Note: $\widehat{F}_h(x_i'\widehat{\beta})$ is the nonparametric CDF estimator; $\widehat{P_h}(x_i'\widehat{\beta})$ is the nonparametric estimator of the probability density function

References

[1] Amemiya, T. 1974. "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal." *Econometrica* 42: 999-1012.

[2] Andrews, Donald W.K., and Marcia M.A. Schafgans. 1998. "Semiparametric Estimation of the Intercept of a Sample Selection Model." *The Review of Economic Studies* 65(3): 497-517.

[3] Chiang, J., and L.F. Lee. 1992. "Discrete/Continuous Models of Consumer Demand with Binding Nonnegativity Constraints." *Journal of Econometrics* 54: 79-93.

[4] Deaton, A, and J. Muellbauer. 1980. "An Almost Ideal Demand System. " *American Economic Review* 70: 312-326

[5] Dong, Diansheng, Brian W. Gould, and Harry M. Kaiser. 2004. "Food Demand in Mexico: An Application of the Amemiya-Tobin Approach to the Estimation of A Censored Food System." *American Journal of Agricultural Economics* 86: 1094-1107.

[6] Green, Richard, William Hahn, and David Rocke. 1987. "Standard Errors for Elasticities: A Comparison of Bootstrap and Aysmptotic Standard Errors." *Journal of Business and Economic Statistics* 5(1): 145-149.

[7] Heien, Dale, and Cathy Roheim Wessells. 1990. "Demand Systems Estimation with Microdata: A Censored Regression Approach." *Journal of Business & Economics Statistics* 8(3): 365-371.

[8] Horowitz, Joel L., and Wolfgang Hardle. 1994. "Testing a Parametric Model against a Semiparametric Alternative." *Econometric Theory* 10(5): 821-848.

[9] Ker, Alan P., and A. Tolga Ergun. 2004. "On the Revelation of Private Information in the U.S. Crop Insurance Program." Research paper, University of Arizona.

[10] Klein R.L., and R.H. Spady. 1993. "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica* 61: 387-421.E

[11] Lazaridis, Panagiotis. 2004. "Demand Elasticities Derived from Consistent Estimation of Heckman-type Models." *Applied Economics Letters* 11: 523-527.

[12] Lee, Lung-Fei, and Mark M. Pitt. 1987. "Microeconometric Models of Rationing, Imperfect Markets, and Non-negativity Constraints." *Journal of Econometrics* 36: 89-110.

[13] Martins, M.F.O. 2001. "Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labor Force in Portugal." *Journal of Applied Econometrics* 16: 23-39.

[14] Meyerhoefer, Chad D., Christine K. Ranney, and David E. Sahn. 2005. "Consistent Estimation of Censored Demand Systems Using Panel Data." *American Journal of Agricultural Economics* 87: 660-672.

[15] Newey, Whitney K., James L. Powell, and James R. Walker. 1990. "Semiparametric Estimation of Selection Models: Some Empirical Results." *American Economic Association: Papers and Proceedings* 80(2): 324-328

[16] Perali, Federico, and Jean-Paul Chavas. 2000. "Estimation of Censored Demand Equations from Large Cross-Section Data." *American Journal of Agricultural Economics* 82: 1022-1037.

[17] Shonkwiler, J. Scott, and Steven T. Yen. 1999. "Two-Step Estimation of A Censored System of Equations." *American Journal of Agricultural Economics* 81: 972-982.

[18] Tauchmann, Harald. 2005. "Efficiency of Two-Step Estimators for Censored Systems of Equations: Shonkwiler and Yen Reconsidered." *Applied Economics* 37: 367-374.

[19] Tobin, J. 1958. "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26: 24-36

[20] Wales T.J., and A.D. Woodland. 1983. "Estimation of Consumer Demand Systems with Binding Non-negativity Constraints." *Journal of Econometrics* 21: 263-285.

[21] Yen, Steven T., and Biing-Hwan Lin. 2006. "A Sample Selection Approach to Censored Demand Systems." *American Journal of Agricultural Economics* 88: 742-749.

[22] Yen, Steven T.. 2005. "A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations." *American Journal of Agricultural Economics* 87: 453-466.

[23] Yen, Steven T., Kamhon Kan and Shew-Jiuan Su. 2002. "Household Demand for Fats and Oils: Two-step Estimation of a Censored Demand System." *Applied Economics* 14: 1799-1806.