

Interdependent Preferences for Seafood Consumption

Joaquín A. Millán,
Polytechnic University of Madrid, Spain

Abstract. This paper considers interdependence in seafood consumption exploiting data on consumer expenditures in Spain, using dominance models. Interdependence is modeled by making parameters in both the participation and the consumption equations dependent upon current consumption in the reference group of a particular consumer. The estimates of a parsimonious model using latent variables reflecting interdependencies seem better in parameters as important as expenditure elasticity or correlation between decision to purchase seafood and the level of seafood consumption.

Keywords: interdependent preferences, seafood consumption, demand

1. INTRODUCTION

Demographic and lifestyle changes are among the factors that have materially affected the structure of food distribution and the demand for food products. These demographic characteristics are studied using survey data because more variation is available than in time series data. The main problem to using survey data is the significant proportion of households that report zero consumption. Some studies focused in the probability of non zero occurrences, as in Nayga and Capps (1995) use discrete logit models to study the probability of consuming fish from home or away from home.

To address the issue of non zero quantitative responses, the tobit model has gained increasing popularity among demand analysts who use household or individual survey data to study consumption pattern and behavior. As an example, Lázaro et al. (2000) use tobit models for explaining the impact of male and female unemployment on the consumption patterns of Spanish households. However, the tobit model is very restrictive in the sense that the variables and parameters deciding the probability of consumption also determine the level of consumption.

Cragg (1971) proposed the double hurdle model, which allows separate stochastic processes for the participation and consumption decisions. Several models have been developed and applied in the same line by, among others, Blundell and Meghir (1987), Haines et al. (1988), Gould (1992), Blaylock and Blisard (1992), Yen (1993), Jensen and Yen (1996), Yen and Huang (1996), or Bertail et al. (1999). There is a previous literature on seafood demand studies, including Lin and Milon (1993) and Yen and Huang (1996), that have focused their analyses on sociodemographic and economic factors affecting seafood consumption, using double hurdle models. Manrique and Jensen (1998) use double hurdle models for explaining Spanish household demand for seafood products.

Some studies are interested in 'lifestyles', approaching this issue from different perspectives and disciplines. Øystein et al. (1998) revise several of these approaches and study determinants of seafood consumption in Norway using ordered probit models. Other approaches have tackled the problem of characterizing consumption patterns endogenizing preferences through habit formation and preference interdependence. Kapteyn et al. (1997) model interdependence in consumer behavior, exploiting aggregate cross-section data without zero consumption, by making parameters in the demand system dependent upon current quantities in the reference group of a household, addressing the identification issues stemming from the 'reflection problem' formulated by Manski (1993,1995). Woittiez and Kapteyn (1998) apply the approach of interdependence to an empirical model of labor supply, of close econometric structure to that of double-hurdle models of consumption.

In this latter approach, consumption decisions are based on mean expenditures in the reference group of the consumer. The central idea in this approach is the notion of a social group, as a set of people who share certain characteristics -education, age, job type, etc.- and to use the social group to which an individual belongs as a proxy for endogenous characteristics. Kapteyn et al. (1997) set the assumptions under what such a procedure is justified and what kind of error is involved, characterizing the probability distribution of the mean consumption within each social group.

In this paper, seafood consumption in Spain with social interactions is studied using cross sectional survey data. First, a latent variable model is estimated in the spirit of the female labor supply model in Woittiez and Kapteyn (1998). The estimated latent variables collect information to the 'true' but unobserved reference groups, and are

used as explanatory in the double hurdle model of seafood consumption.

2. SEAFOOD CONSUMPTION IN SPAIN

There are two main sources of information on seafood consumption, based on consumption surveys (Spanish Bureau of Statistics: INE, and Spanish Ministry for Agriculture, Fisheries and Food: MAPA). INE collects quarterly data of several thousands of households on a rotating panel basis. This Consumption Survey begins in mid 80's, but both the methodology and the items examined have changed recently in 1997.

The Consumption Panel of MAPA is not a task of MAPA itself, but of private companies. However there is an effort by MAPA in showing homogeneous information as if there is not any break in series. The results are available monthly with a small delay of a few months. The main data are severely aggregated in four groups: fresh fish, frozen fish, processed -cured, canned, etc.-, and crustaceans and molluscs. An important point with this survey is the detailed demographic information. Regrettably, microdata is not available.

Most of the descriptive studies on seafood consumption in Spain have used the Consumption Panel of MAPA. In mid-90s, there is a global decreasing of seafood products, at a rate of 0.5% annual in the period 87-97, although not at a regular pace. While at the first 90s, consumption rises (1.7% annual) there is a remarkable descent in the period 1995 to 1997 (4% annual).

The decline of seafood consumption applies mainly to fish, and mainly to frozen fish (average growth rate of -5.3% by year), while not so pronounced for fresh fish (-0.6% by year). The consumption of preserved fish rises significantly over the period at an average rate of 7.8% annual. However, the growth rate is not so high in the more recent years. Consumption of crustaceans and molluscs is very stable with a growth rate of 0.1% by year in the period 1987 to 1997. Total consumption of seafood at home goes down at a higher rate (-1.2% annual) than total seafood consumption. Decline is particularly important for frozen fish (-7.2% annual). The different evolution is represented in the consumption of preserves with an impressive growth rate of 8.1% annual.

The previous literature, based on the Consumption Panel of MAPA, highlights some interesting sociodemographic factors explaining characteristics of household consumption of seafood: Fish consumption is negatively related with woman work. Seafood consumption is lesser in households with children. Fish consumption is

negatively related to household size and to income level. Fish consumption is larger in cities and metropolitan areas. There are regional differences in seafood consumption. The present study investigates these relationships on microdata belonging to the fourth quarter of 1996 wave of the *Encuesta Permanente de Presupuestos Familiares* of INE (ECPF).

3. DATA

The data we use for the estimation of the demand system is microdata belonging to the fourth quarter of 1996 wave of the *'Permanent Survey of Family Budgets'* (ECPF), organized by the Spanish INE. The whole sample is random and is extracted from a stratification process that makes it representative of the whole of Spain. The sample of the ECPF is rotated in 12.5% every quarter. This means that a balanced panel of 400 households should be obtained from every group of eight quarters. In practice, however, the rate of attrition is quite high and only about 10% of the households complete the eighth quarter term of cooperation. Attrition leads to substitution of the household in a way such that the representability of the sample is preserved. There is the possibility of using two waves (as 4th quarter of 1995 and 4th quarter of 1996) to study habit formation. However, given that for the selection of 'reference groups' a set of well-defined households is needed, the resulting sample using two waves would be very small. In fact, under the selected criteria the sample used is about 25% of the whole sample.

Thus, the study is based on the 4th quarter of 1996 only. The total number of households is 3145. For defining reference groups, only a limited set of categories is retained. The resulting sample consists of 767 households. In Table 1, the categories of several explanatory variables are presented. The reference groups are formed combining the eight regions and the four municipality sizes to define a total of 26 locations. This is because only region one, which includes Barcelona, and region four, which includes Madrid, have municipalities in size group four. Each location is combined with the following criteria to define four classes of reference groups: type of household, woman working, education and labor activity. Only types of household easily identifiable are included. Household with non active head is excluded because its inclusion over represent without education, one person household

Region	
1 Northeast	150
2 East	113
3 Andalucia	96
4 Center-South	134
5 Castilla-León	33
6 North	112
7 Northwest	70
8 Canary Islands	59
City size	
1 Population < 5000	194
2 5000 < Population < 50000	171
3 Population > 5000 & capitals	331
4 Madrid and Barcelona	71
Type of household	
1 One person	76
2 Couple without children	188
3 Couple, 1 child (< 14 years old)	214
4 Couple, 2 children	245
5 Couple, 3 or more children	44
Woman working	
1 Yes, rural habitat.	180
2 Yes, urban habitat	180
3 Yes, Madrid and Barcelona	44
4 No, rural habitat.	185
5 No, urban habitat	151
6 No, Madrid and Barcelona	27
Education	
1 Without	48
2 1st degree	218
3 2nd degree, 1st cycle	194
4 2nd degree, 2 cycle	196
5 University, 1st cycle	40
6 University, 2nd cycle or more	71
Activity	
1 Unemployed	60
2 Entrepreneurs & professionals	36
3 Agriculture	52
4 Own employed & executive	167
5 Nonproduction & skilled workers	358
6 Non skilled workers	94

Table 1. Socio-demographic characteristics.

The descriptive statistics of other variables used in the empirical analysis are shown in Table 2. The share of households purchasing seafood products is 74.7%. The share of seafood in food consumption is slightly lower than in data from the Food Consumption Panel of MAPA, perhaps due to infrequency of purchase. With respect to seafood consumption, all food purchases in a week are recorded.

In the empirical analysis, the dependent variable of the consumption equation SEAFOOD expenditures is in log form. The selected explanatory variables are among the more usual in food demand analysis with cross section data, perhaps after a transformation. AGE of the household head is included normally in square form looking for life cycle effects. It is expected a negative relationship between seafood expenditure and household size, measured in the number of persons. It is usual to consider income as an explanatory variable. However, due to the complex structure of income measures in ECPF, expenditure is selected as a proxy. Moreover, because expenditure elasticity for seafood is estimated usually in conditional food or meat and seafood systems, a decomposition of expenditure in two elements is used. First it is FOOD expenditures (in log) - significant positive relationship is expected- and (log) NONFOOD expenditure- where not a priori relationship is expected. The discrete variables in Table 1, with slight adjustments, are also included into the set of explanatory variables.

767 households				
	number	%		
Fish consumption	573	74.7		
Woman working	404	52.7		
	mean	std.dev.	min.	max
AGE (years)	38.7	9.9	20	84
Household SIZE	2.999	1.11	1	6
	mean	std.dev.	min.	max
Expenditure (pta)				
FOOD	16296	10034	390	76670
NONFOOD	60182	44412	3148	558492
SEAFOOD	1706	2368	0	31008
Share of seafood in food expenditure				
WPESC	0.097	0.104	0.000	0.717

Table 2. Statistical summary.

4. EMPIRICAL ANALYSIS

The first stage in the analysis is the characterization of the reference group. Woittiez and Kaptein (1998) construct a set of indicators assigning positive weight to other households close to the reference group of the household. A variety of the approach is as follows: For each household a set of indicators is constructed based of the mean -the household is excluded- of consumption in other households in the same cell according to a set of characteristics. There is the possibility of the construction

of four set of indicators according to type of household, woman working, education and labor activity, and dependent on each location. By the other side, we can double the number of indicators selecting average participation or average consumption. Thus, there are eight indicators to be included as the 'reference group'. These indicators reflect many other characteristics than seafood consumption preference within group. A measure of the 'true' preference for seafood consumption is constructed by a factor analysis.

Woittiez and Kapteyn use a maximum likelihood procedure to extract the factors representing preferences. This procedure can be considered only as an approximation particularly in the case for the Only the first factor is retained because is the factor that explain seafood consumption. The commonality with the participation variables is higher than with the consumption variables. In the female labor supply model of Woittiez and Kapteyn the results were reversed. However, the degree of fit of both the participation and the consumption factors is not very good in our model.

To the understanding of the value of the first factors as indicators of the 'true' preferences for seafood, the loadings are presented in Table 5. The first factor is retained because is the only factor that explain seafood consumption. The commonality with the participation variables is higher than with the consumption variables. In the female labor supply model of Woittiez and Kapteyn the results were reversed. However, the degree of fit of both the participation and the consumption factors are lesser in our model.

To the understanding of the value of the first factors as indicators of the 'true' preferences for seafood, the loadings are presented in Table 5.

Factor	Loadings	Factor	Loadings
	CIND1		PCIND1
IND1	0.658	PIND1	0.756
IND2	0.576	PIND2	0.753
IND3	0.638	PIND3	0.753
IND4	0.686	PIND4	0.804

Table 5. Loadings.

The loadings suggest the high content of information about seafood consumption, whatever the indicator chosen.

The next step is the estimation of a double hurdle model of seafood consumption. That is, the participation and the consumption equation are determined by two separate decision steps.

The double-hurdle model is a parametric generalization of the tobit model, in which the decision to consume and the level of consumption are determined by two separate stochastic processes. In some respects, parameterization of the double-hurdle model is similar to that of the Heckman procedure in that two separate sets of parameters are obtained in both cases.

The double-hurdle model has a participation equation or a decision to consume seafood:

$$P_i = 1 \text{ if } P_i^* > 0 \text{ and } 0 \text{ if } 0 \geq P_i^* \\ P_i^* = \alpha' Z_i + u_i \tag{1}$$

being Z a vector of household characteristics and α a vector of parameters; and a consumption equation, or a decision of the level of seafood consumption at home:

$$C_i = 1 \text{ if } C_i^* > 0 \text{ and } I_i^* > 0 \text{ and } 0 \text{ else} \\ C_i^* = \beta' X_i + v_i \tag{2}$$

being X a vector of household characteristics and β a vector of parameters.

The model above is a general double hurdle model of consumption. In this paper it is assumed that the household has no taste for consuming seafood. In this case,

$$C_i = 1 \text{ if } P_i^* > 0 \text{ and } 0 \text{ else} \tag{3}$$

whivh is a sample selection model. This mean that households with null consumption do not impose restrictions on the parameters of the decision equation. The model is a dependent model because there may be a relationship between the decision to consume and the decision on how much is consumed:

$$\rho = \frac{\text{cov}(u_i, v_i)}{(\text{var}(u_i) \text{ var}(v_i))^{1/2}} \tag{4}$$

Being $\rho=0$, the model is of complete independence and it decomposes in a probit equation and a standard regression on the truncated sample of positive values of consumption C.

In this paper, the problems of infrequency of purchase, as in Blundell and Meghir (1986), Gould (1992) and Blisard and Blaylock (1993), cannot be addressed. The dependent variable in the empirical model, is the value of all fresh, frozen, canned and cured seafood purchased, including shellfish (pta. per household). The demographic variables specified in the empirical model include type of household, activity of the household head,

education of the household head, municipality size, age, age squared (looking for life cycle effects), woman working, log of household size, log of food expenditure, log of nonfood expenditure, and the latent variable obtained as first common factor for participation and consumption. There is no reason for selecting particular variables in the participation and consumption decisions - excepting common factors. The same almost complete set of variables is chosen as explanatory in both equations. However, because of convergence problems, region dummies are excluded from the participation equation and type of household dummies are excluded from the consumption equation. Three models are estimated, and the results are shown in Appendix A. The first model includes all variables excepting in the construction of preference groups and the latent variables. The second model includes variables excepting in the construction of preference groups and not the latent variables. The third model includes all variables.

Given the large number of parameters, the model without dummies is accepted at $p=0.05$ in a likelihood ratio against the full model. The model without latent variables is rejected at $p=0.05$ but not rejected at $p=0.01$. The standard error in the consumption equation - σ - is lower in the model without dummies.

There is a problem about the coefficient of food consumption in both the participation and the consumption equation, mainly in the latter. Other unclear results are the impact of the latent variable in the consumption equation and the change in the value of the correlation coefficient between the disturbances of the participation and the consumption equation ρ .

Concerning non dummy variables, the estimates of the model without dummies seem sensible. In fact, it is a very parsimonious model that concentrates on some interesting parameters. The latent variables are proxies of unobserved attributes in the utility function.

However, there is some information in the estimates of the full model and in the model without latent variables. As an example, the lesser consumption in rural areas than in urban areas is a result in the consumption equation. By the other side, purchase of seafood has a higher probability for the group of entrepreneurs with salaried workers and professionals. The problem is that some coefficients seem strange, particularly $\log FOOD$ in the consumption equation and ρ . Thus, although some information about groups coefficients seem interesting, the coefficients on common variables in the model without dummies seem better.

There is an additional problem about identification.

Manski (1993,1995) is critical on inferring the composition of reference groups. The main idea in Manski's argument is that the conditional expectation of a variable, say y , on other variable x and any specified function of x $g(x)$ adds nothing to our knowledge. Manski (1995, p.135) presents that $E(y/x, g(x)) = E(y/x)$ holds tautologically. However, there is a difference in our approach because groups coefficients are estimated without the particular realization of x_i . This is a main finding in Kapteyn et al. (1997).

By the other side, other problems remain with respect to the meaning of the interdependence effects. Manski distinguishes endogenous effects, 'wherein the propensity of an individual to behave in some way varies with the prevalence of that behavior in the group', contextual effects, 'wherein the propensity of an individual to behave in some way varies with the distribution of background characteristics in the group', and correlated effects, 'wherein individuals in the same group tend to behave similarly because they face similar institutional environments or have similar individual characteristics'.

Although the theoretical analysis is on endogenous preferences, the empirical analysis usually is carried out on contextual or correlated effects. The 'reflection problem' is too hard to be discussed here. However, the empirical results of the simplified model without dummies and with latent variables suggest that something can be gained with this approach, although the meaning of the particular coefficient or the latent variable is unclear (Manski, 1995).

5. CONCLUSION

This paper has presented a model of interdependence or, in a weak form, of correlated effects in seafood consumption in Spain, using cross sectional data. The empirical results are mixed. Although the meaning of some construct is difficult to understand, the coefficients of some parameters of interest seem better in a parsimonious double hurdle model with latent variables.

Models with a large set of sociodemographic variables present sensible estimates on these variables, although not always reinforcing the previous literature on seafood consumption in Spain. However, the estimates of some other very important coefficients, as expenditure elasticities or the relationship between errors in the decision of purchasing seafood and the decision on the level of seafood consumption, seem wrong intuitively.

Despite the problems for extracting robust inferences from an approach using such a large set of latent variable

constructs, it seems that the methodology is worth of further empirical research.

6. ACKNOWLEDGEMENTS

The author gratefully acknowledge financial support from Spains' DGICYT SEC98-484 and EU's FAIR PL-97-3500.

7. REFERENCES

- Bertail, P. F. Cavaillet, and V. Nichele. A bootstrapped double-hurdle analysis: consumption of home produced food. *Applied Economics*, 31, 163-39, 1999.
- Blaylock, J.R. and W.N. Blisard. US cigarette consumption: the case of low-income women. *American Journal of Agricultural Economics*, 74(3), 698-705. 1992.
- Blisard, N. and J. Blaylock. Distinguishing between market participation and infrequency of purchase models of butter demand. *American Journal of Agricultural Economics*, 75, 314-20. 1993.
- Blundell, R.W. and C. Meghir. Bivariate alternatives to the univariate Tobit model. *Journal of Econometrics*, 33, 179-200. 1987.
- Cragg, J.G. Some statistical models for limited dependent variables with application for durable goods. *Econometrica*, 39, 829-44. 1971.
- Gould, B.W. At-home consumption of cheese: a purchase-infrequency model. *American Journal of Agricultural Economics*, 74, 543-52. 1992.
- Haines, P.S., D.K. Guilkey and B.M. Popkin. Modeling food consumption decisions as a two step process, *American Journal of Agricultural Economics*, 70, 543-552. 1988.
- Jensen, H.H. and S. T. Yen. Food expenditures away from home by type of meal. *Canadian Journal of Agricultural Economics*, 44, 67-80. 1996.
- Kapteyn, A., S.A. van de Geer, H. van de Stadt, and T.J. Wansbeek. Interdependence of preferences: an econometric analysis. *Journal of Applied Econometrics*, 12(6), 665-686, 1997
- Lázaro, N. M.L. Moltó, and R. Sánchez. Unemployment and consumption patterns. *Applied Economics*, 32, 367-379. 1998
- Lin, C.T.J. and J.W. Milon. Attribute and safety perceptions is a double-hurdle model of shellfish consumption. *American Journal of Agricultural Economics*, 75, 724-29. 1993.
- Manrique, J. and H.H. Jensen. Spanish household demand for seafood products. In Eide A. and T. Vassdal (eds.), *IIFET'98 Tromsø Proceedings*, 711-717. Corvallis OR: IIFET. 1998
- Manski, C.F. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60, 531-542. 1993
- Manski, C.F. *Identification problems in the Social Sciences*. Harvard University Press. Cambridge. 1995
- Nayga, R.M. and O. Capps, Factors affecting the probability of consuming fish and shellfish in the away from home and home markets. *Journal of Agricultural and Applied Economics*, 27, 161-71. 1995
- Øystein, M., T. Trondsen, R.S. Johnston, and E. Lund. Determinants of seafood consumption in Norway: Lifestyle, revealed preferences, and barriers to consumption. In Eide A. and T. Vassdal (eds.), *IIFET'98 Tromsø Proceedings*, 718-736 Corvallis OR: IIFET. 1998
- Woittiez, I. and A. Kapteyn. Social interactions and habit formation in a model of female labour supply. *Journal of Public Economics*, 70, 185-205. 1998
- Yen, S.T. Working wives and food away from home: the Box-Cox double hurdle model. *American Journal of Agricultural Economics*, 75, 884-95. 1993.
- Yen, S.T. and C.L. Huang. Household demand for finfish: a generalized double hurdle model. *Journal of Agricultural and Resource Economics*, 21(2), 220-234. 1996.

APPENDIX A. EMPIRICAL MODELS

DOMINANCE MODELS

	Without dummies		Without latent		Full	
Loglik	-1035		-1014		-1010	
PARTICIPATION						
C	-7.46	1.34	-6.55	1.50	-6.54	1.50
TYPE OF HOUSEHOLD Reference 1 person						
Non sig.			
ACTIVITY Reference: Non skilled						
Unemployed			0.07	0.25	0.03	0.25
Entrep.&prof.			1.35	0.49	1.23	0.47
Agriculture			-0.01	0.32	-0.11	0.33
Own emp.&exec.			0.15	0.20	0.10	0.20
Nonprod&skill			0.26	0.16	0.22	0.16
EDUCATION: Reference University 2nd cycle and more						
Non sig.			
POPULATION: Reference Metropolitan						
Non sig.			
AGE	0.05	0.04	0.08	0.04	0.08	0.04
AGE*AGE	-0.0003	0.0005	-0.0007	0.0005	-0.0007	0.0005
Woman working	0.01	0.11	0.07	0.11	0.10	0.12
logFOOD	1.02	0.12	0.81	0.11	0.82	0.11
logNONFOOD	-0.23	0.09	-0.26	0.11	-0.24	0.11
logSIZE	-0.27	0.15	1.94	3.47	2.38	3.69
Latent part.	0.23	0.05			0.15	0.06
CONSUMPTION						
C	-3.95	1.28	1.51	1.12	1.51	1.12
AGE	0.04	0.03	0.01	0.03	0.01	0.03
AGE*AGE	-0.0004	0.0003	0.0000	0.0004	0.0000	0.0004
Woman working	0.02	0.07	0.03	0.09	0.02	0.09
logFOOD	0.94	0.13	0.46	0.09	0.46	0.09
logNONFOOD	0.11	0.07	0.18	0.08	0.18	0.08
logSIZE	-0.16	0.10	-0.05	0.12	-0.03	0.12
ACTIVITY Reference: Non skilled						
Non sig.			
REGION Reference: Northwest						
Non sig.			
EDUCATION: Reference University 2nd cycle and more						
Non sig.			
POPULATION: Reference Metropolitan						
< 5000			-0.34	0.17	-0.38	0.17
5000<Pop<50000			0.31	0.16	-0.35	0.18
>50000 & capitals			0.22	0.07	-0.26	0.16
Latent cons.	0.08	0.03			-0.03	0.05
SIGMA	0.81	0.26	0.94	0.07	0.93	0.07
RHO	0.31	0.06	-0.90	0.07	-0.89	0.07