Applied Economics Letters, 1997, **4**, 181–185

Calculating marginal effects in dichotomous–continuous models

ATANU SAHA, ORAL CAPPS* and PATRICK J. BYRNE‡§

*Micronomics Inc., 400 South Hope Street, Los Angeles, CA 90071, USA; *Department of Agricultural Economics, Texas A&M University, College Station, TX 77843, USA;* and ‡§*Food and Resource Economics Department, University of Florida, Gainsville, FL 32611, USA*

Received 13 March 1996

In many economic settings, individual decisions can be viewed as a sequential process where a dichotomous choice is followed by a continuous choice. These processes are frequently encountered in consumption demand studies, where the decision of whether or not to consume a particular commodity is followed by the choice of how much to consume. The Heckman two-step approach has been extensively used in estimating these models. Expressions are derived for calculating marginal effects of regressors in dichotomous–continuous models. It is proposed that the marginal effect expressions are incomplete in almost all consumption demand studies that use the Heckman approach. In dichotomous–continuous models, a change in an explanatory variable that is common to both stages of the decision process has two effects: (1) it affects the likelihood of whether the commodity will be consumed; and (2) if the commodity is consumed, it affects the expenditure on that commodity. The first effect has so far been omitted from applied demand studies. The correct marginal effect expressions are derived for single-commodity and multiple-commodity demand models. An application to consumption survey data on 12 food commodities shows that erroneous marginal effect expressions can introduce substantial bias in demand elasticity estimates.

I. INTRODUCTION

In many economic settings, individual decisions can be viewed as a sequential process where a dichotomous choice is followed by a continuous choice. Consider the adoption of a divisible technology. A producer's binary decision of whether or not to adopt is followed by the continuous choice of how much to adopt. Similar choices are frequently encountered in consumption demand studies (Haines *et al.*, 1988). For example, the choice of how much to spend in eating out is preceded by the decision to eat out (Yen, 1993). An analogous decision framework has also been used to examine consumption patterns of various commodity groups such as tobacco (Blisard and Blaylock, 1993), wine (Pompelli and Heien, 1992) and fish (Cheng and Capps, 1988).

The Heckman two-step approach (Heckman, 1976) has been extensively used in estimating dichotomous–continuous models. In this procedure, a probit model, reflecting the dichotomous decision, is estimated in the first stage; a regression equation with a continuous, non-zero dependent variable is estimated in the second. In this equation the estimated inverse Mills ratio from the first stage is included as an additional repressor.

We derive expressions for calculating marginal effects of regressors in Heckman two-step models. We contend, in almost all consumption demand studies, the marginal effect expressions omit the terms related to the change in the inverse Mills ratio. To illustrate, consider the marginal effect of income on the expected value of expenditure on food away from home. Higher income not only affects this expenditure

§ Senior authorship is equally shared, but all correspondence should be sent to Atanu Saha at Micronomics.

directly – wealthier families spend more when dining out – but also indirectly by increasing the likelihood of dining out. The latter effect is reflected by the income-induced change in the inverse Mills ratio. It is this effect that has remained omitted in applied consumption demand studies.

The exposition is framed in terms of consumption demand models, although the results are relevant to any application of the Heckman two-step estimation procedure. We derive the correct marginal effect expressions for single-commodity and multiple-commodity demand models. An application to consumption survey data on 12 food commodities illustrates the results.

II. MARGINAL EFFECTS FOR SINGLE-COMMODITY DEMAND MODELS

Let Z_k denote an indicator variable that equals 1 if the kth household decides to consume the commodity in question and 0 otherwise in the probit analysis of the Heckman procedure's first stage. Also, let the normal cumulative distribution and the probability density functions be denoted by Φ and ϕ, respectively. The estimated inverse Mills ratio is

$$\hat{\lambda}_k = \frac{\phi(\boldsymbol{W}_k\hat{\gamma})}{\Phi(\boldsymbol{W}_k\hat{\gamma})}$$

where $\mathbf{W}_k$ is the vector of regressors explaining the binary choice in the first-stage and γ is a conformable parameter vector. In the second-stage estimation equation, the conditional expectation of the dependent variable is

$$\begin{aligned} E[Y_k|Z_k = 1] &= \boldsymbol{X}_k\beta + \alpha\frac{\phi(\boldsymbol{W}_k\hat{\gamma})}{\Phi(\mathbf{W}_k\hat{\gamma})} \\ &= \boldsymbol{X}_k\beta + \alpha\,\hat{\lambda}_k \end{aligned} \tag{1}$$

where Y_k denotes the expenditure on the commodity, $\mathbf{X}_k$ is a vector of regressors explaining the magnitude of the expenditure, β is a conformable parameter vector and α denotes the parameter associated with the inverse Mills ratio. Importantly, only the non-zero observations on Y_k are used in the second-stage estimation.

Let X_{kj} denote the jth regressor that is common to both $\mathbf{W}_k$ and $\mathbf{X}_k$. The estimated marginal effect (ME) of a change in this regressor is given by

$$\widehat{ME}_{kj} = \frac{\delta E[Y_k|Z_k = 1]}{\delta \mathbf{X}_k} = \hat{\beta}_j + \hat{\alpha}\;\frac{\hat{\lambda}_k}{\delta X_{kj}}$$

Thus, ME is composed of two parts: a direct effect on expected expenditure, reflected by $\hat{\beta}_j$, and a change in the probability of consuming the commodity. The latter is denoted by the second term on the right-hand side of Equation 2. Only the first effect has been considered in consumption demand studies. The degree and direction of the attendant bias in the calculation of marginal effects depends on the magnitude and sign of the omitted term. After some simplification, the ME expression in Equation 2 can be rewritten as

$$\widehat{ME}_{kj} = \hat{\beta}_j - \hat{\alpha}\hat{\gamma}_j\left\{\boldsymbol{W}_k\hat{\gamma}\,\hat{\lambda}_k + (\,\hat{\lambda}_k)^2\right\} \tag{3}$$

Equation 3 represents the appropriate expression for ME in single-commodity models under the Heckman two-step estimation procedure. In general $\widehat{ME}_{kj} \neq \hat{\beta}_j$; the only case where $\widehat{ME}_{kj} = \hat{\beta}_j$ is where $\hat{\alpha} = 0$, which is true in the highly unlikely event that the errors of the first- and second-stage estimation equations have zero covariance.

Since the estimated ME is observation dependent, we propose to evaluate it at the sample means, so

$$\widehat{ME}_{kj}|_{\text{sample mean}} = \hat{\beta}_j - \hat{\lambda}\hat{\gamma}_j\left(\left(\overline{\boldsymbol{W}}\hat{\gamma}\right)\overline{\hat{\lambda}} + \overline{\hat{\lambda}}^2\right) \tag{4}$$

where $\bar{\mathbf{W}}$ denotes the vector of regressor sample means and

$$\overline{\hat{\lambda}} = \frac{\Phi(\overline{\boldsymbol{W}}\hat{\gamma})}{\Phi(\overline{\boldsymbol{W}}\hat{\gamma})}$$

is the inverse of the Mills ratio evaluated at those means.

III. MARGINAL EFFECTS FOR MULTIPLE-COMMODITY MODELS

Heien and Wessells (1990) and Heien and Durham (1991) have extended the Heckman single-equation approach to a system of equations. In this procedure all observations, corresponding to both zero and non-zero expenditures, are used in the second-stage estimation. All commodity equations are estimated as a system of seemingly unrelated regression equations, each having the same set of regressors, except for the inverse Mills ratios, that differ by commodity.

Denoting the ith commodity by the subscript i, and assuming m commodities, this mode may be written

$$\begin{aligned} E[Y_{ki}|Z_{ki} = 1] &= \boldsymbol{X}_k\beta_i + \alpha_i\frac{\phi(\overline{\mathbf{W}}_k\hat{\gamma}_i)}{\Phi(\overline{\mathbf{W}}_k\hat{\gamma}_i)} \\ E[Y_{ki}|Z_{ki} = 0] &= \boldsymbol{X}_k\beta_i + \alpha_i\frac{\phi(\overline{\mathbf{W}}_k\hat{\gamma}_i)}{1 - \Phi(\overline{\mathbf{W}}_k\hat{\gamma}_i)} \qquad i = 1, \ldots, m \end{aligned} \tag{5}$$

Since both zero and non-zero expenditures are considered, ME for X_{kj}, evaluated at the sample means, is composed of two parts:

$$\widehat{ME}^A_{kj} = \frac{\partial E[Y_{kj}|Z_{ki} = 1]}{\partial X_{kj}}\bigg|_{\substack{\text{sample}\\ \text{mean}}} = \hat{\beta}_{ij} - \hat{\alpha}_i\hat{\gamma}_{ij}\left\{\overline{\mathbf{W}}\hat{\gamma}_i\,\overline{\hat{\lambda}}^A_i + \left(\overline{\hat{\lambda}}^A_i\right)^2\right\} \tag{6}$$

and

$$\widehat{ME}^B_{kj} = \frac{\partial E[Y_{kj}|Z_{ki}=0]}{\partial X_{kj}}\Bigg|_{\substack{= \\ \text{sample} \\ mean}} \hat{\beta}_{ij} - \hat{\alpha}_i\hat{\gamma}_{hj}\left\{\overline{\boldsymbol{W}}\hat{\gamma}_i\bar{\hat{\lambda}}^B_i + \left(\bar{\hat{\lambda}}^B_i\right)^2\right\} \tag{7}$$

where

$$\bar{\hat{\lambda}}^A_i = \frac{\phi(\overline{\boldsymbol{W}}\hat{\gamma}_i)}{\Phi(\overline{\boldsymbol{W}}\hat{\gamma}_i)} \quad \text{and} \quad \bar{\hat{\lambda}}^B_i = \frac{\phi(\overline{\boldsymbol{W}}\hat{\gamma}_i)}{1-\Phi(\overline{\boldsymbol{W}}\hat{\gamma}_i)} \tag{8}$$

To compute the overall ME of X_{kj}, we propose to take the weighted average of $\widehat{ME}^A_{kj}$ and $\widehat{ME}^B_{kj}$, giving

$$\widehat{ME}_{kj} = \hat{\beta}_{ij} - \hat{\alpha}_i\hat{\gamma}_{ij}\left[\theta_i\left(\overline{\boldsymbol{W}}\hat{\gamma}_i\bar{\hat{\lambda}}^A_i + \left(\bar{\hat{\lambda}}^A_i\right)^2\right) + (1-\theta_i)\left(\overline{\boldsymbol{W}}\hat{\gamma}_i\bar{\hat{\lambda}}^B_i - \left(\bar{\hat{\lambda}}^B_i\right)^2\right)\right] \tag{9}$$

where the weight, $0 < \theta_i < 1$, is the proportion of observations for which $Z_{ki} = 1$. Equation 9 represents the appropriate expression for calculating MEs in a system of equations using the Heckman approach.

IV. EMPIRICAL ANALYSIS

To empirically determine the extent to which the MEs change when the correct expressions are used, we examine expenditure on 12 food commodities. The data come from the 1987–88 US Nationwide Food Consumption Survey. The food commodities are (1) food away from home, (2) beef, (3) pork, (4) chicken, (5) fish, (6) cheese, (7) milk, (8) fruit, (9) vegetables, (10) breakfast cereals, (11) bread and (12) fats and oils. Information was collected on various socioeconomic and demographic characteristics of the household as well as detailed records on the money value, quantity and types of foods used by the household over one-week period. The socioeconomic and demographic characteristics considered here are region, urbanization, race, income and household

Table 1. *Descriptive statistics of the dependent variables used in the analysis: sample size 3896*

Dependent variable	Percentage of non-zero observations	MEAN ($)	STND DEV ($)	MIN ($)	MAX ($)
Food away from home	81.3	27.23	33.63	0	350
Beef	83.1	6.35	6.95	0	92
Pork	66.8	3.64	5.27	0	142
Chicken	64.4	2.40	3 50	0	100
Fish	51.6	2.73	5.65	0	70
Cheese	74.3	2.17	2.43	0	27
Milk	90.5	3.27	3.23	0	38
Fruit	77.2	3.03	3.61	0	57
Vegetables	82.5	3.29	3.35	0	31
Breakfast cereals	70.4	2.02	2.36	0	23
Bread	82.1	1.58	1.44	0	14
Fat and oils	75.0	1.57	1.57	0	19

Table 2. *Goodness-of-fit measures from the first-stage probit regressions*

Commodity	McFadden R^2 [a]	Percentage of correct predictions[b]
Food way from home	0.149	0.8172
Beef	0.0891	0.8311
Pork	0.0511	0.6855
Chicken	0.0479	0.6552
Fish	0.0230	0.5746
Cheese	0.0718	0.7494
Milk	0.1047	0.9058
Fruit	0.0419	0.7720
Vegetables	0.0470	0.8259
Breakfast cereals	0.0810	0.7166
Bread	0.1283	0.8236
Fats and oils	0.1009	0.7582

[a] −[(log of the likelihood function evaluated at the maximum likelihood estimates) / (log of the likelihood function when all coefficients, except the constant, are set to zero)].
[b] An observation is predicted to be 1 if the predicted probability is 0.5 or greater; otherwise the observation is predicted to be 0.

size. The data sample used in the estimation has 3896 observations.

The specification for the ith commodity estimation equation in the second stage of the Heckman procedure is assumed to be

$$Y_{ik} = \beta_{0i} + \beta_{1i}NE + \beta_{2i}MW + \beta_{3i}WEST + \beta_{4i}CC + \beta_{5i}SUB + \beta_{6i}BLACK + \beta_{7i}ASIAN + \beta_{8i}INCOME + \beta_{9i}HSIZE + \beta_{10i}\hat{\lambda}_{ik} + \epsilon_{ik} \quad (10)$$

where $NE = 1$ if the household resides in the north-east, 0 otherwise; $MW = 1$ if the household resides in the mid-west, 0 otherwise; $WEST = 1$ if the household resides in the west, 0 otherwise; $CC = 1$ if the household resides in a city which has a population of 50 000 or more, 0 otherwise; $SUB = 1$ if the household resides in a suburban area, 0 otherwise; $BLACK = 1$ if the race of the household is black, 0 otherwise; $ASIAN = 1$ if the race of the household is Asian or Pacific Islander, 0 otherwise; $INCOME =$ household income before taxes in dollars; $HSIZE =$ household size, i.e the number of persons in the household; and $\hat{\lambda}_{jk} =$ estimated inverse Mills ratio from a first-stage probit regression. The same regressors, excent $\hat{\lambda}_{ik}$, were used in the first-stage probit analysis.

Descriptive statistics of the dependent variables used in the analysis are presented in Table 1. Goodness-of-fit measures for the probit analysis are provided in Table 2. In the interest of brevity, the parameter estimates of the probit and second-stage regressions are not reported here. We focus on comparing the *ME*s of income and household size with and without taking into account changes in the inverse Mills ratio. The estimated *ME*s, converted into elasticities, are evaluated at the sample means and reported in Tables 3 and 4.

Table 3 shows that, for the single-equation applications, the income and household size elasticities change considerably after substituting the correct expression for *ME*. For example, without taking into account changes in the inverse of the Mills

Table 3. *Calculation of income and household size elasticities with and without taking into account the inverse of the Mills Ratio*[a]

Commodity	$\bar{R}^2$	Income elasticity without correction	Income elasticity with correction	Household size elasticity without correction	Household size elasticity with correction
Food away from home	0.1720	0.3398	0.6721	−0.0739	0.2059
Beef	0.1070	0.1302	0.1229	0.0943	0.6818
Pork	0.0896	0.0201	0.0448	0.1306	0.7770
Chicken	0.0632	0.2213	0.1953	0.8722	0.4984
Fish	0.0434	0.3404	0.4744	−0.0810	0.1738
Cheese	0.1050	0.0995	0.0999	0.6121	0.6141
Milk	0.2752	0.0144	0.0131	1.2148	0.9099
Fruit	0.0695	0.0793	0.2705	0.2121	0.3508
Vegetables	0.0733	0.0944	0.2207	0.0628	0.2406
Breakfast cereals	0.2114	−0.0277	−0.0119	1.4140	0.9418
Bread	0.1627	0.0596	0.0514	0.8404	0.5618
Fats and oils	0.0728	0.0264	0.0450	0.3760	0.4929

[a] Heckman procedure: single equation application

Table 4. *Calculation of income and household size with and without taking into account the inverse of the Mills Ratio*[a]

Commodity	$\bar{R}^2$	Income elasticity without correction	Income elasticity with correction	Household size elasticity without correction	Household size elasticity with correction
Food away from home	0.2908	0.4160	0.4938	0.1005	0.1666
Beef	0.2513	0.1102	0.1088	0.4031	0.5244
Pork	0.2652	0.0343	0.0387	0.4526	0.5675
Chicken	0.2430	0.1357	0.1417	0.3368	0.4240
Fish	0.0785	0.3208	0.3241	0.2203	0.2263
Cheese	0.3285	0.0665	0.0909	0.4134	0.5133
Milk	0.3452	−0.0010	−0.0007	0.8471	0.9148
Fruit	0.2564	0.1498	0.1931	0.2497	0.2809
Vegetables	0.2548	0.1527	0.1811	0.1782	0.2181
Breakfast cereals	0.4249	−0.0176	−0.0228	0.6694	0.8224
Bread	0.3582	0.0352	0.0391	0.4465	0.5789
Fats and oils	0.3338	0.0258	0.0471	0.3252	0.4585

[a] Heckman procedure: system of equations.

ratio (described as 'without correction' in Tables 3 and 4), the income elasticity for food away from home is 0.3398. But with correction, i.e. taking into account the changes in the inverse of the Mills ratio, the income elasticity is 0.6721 – over 97% larger. This increase reflects the large and positive effect of higher income on the probability of eating out. Similar and sometimes more pronounced elasticity differences are evident for other commodities.

The household size elasticities for most commodities also appear to be extremely sensitive to the *ME* specification. Not only do the magnitudes differ, but in some cases the signs change after substituting the correct expression for *ME*. For food away from home and for fish, the household size elasticity changes from −0.0739 to 0.2059 and from −0.0810 to 0.1738, respectively. The household size elasticities rise in magnitude for beef, pork, fruit, vegetables, and fats and oils; whereas their magnitudes fall for chicken, milk, breakfast cereals and bread.

For the system of equations (Table 4), the pattern of the changes of income and household size elasticity is similar to those in the single-equation applications. In most cases the elasticities are larger after substituting the correct expression for *ME*. However, the differences between the correct and uncorrected elasticities are noticeably smaller than those in the single-equation approach.

Thus, the ultimate conclusion is that it is imperative to take into account the changes in the inverse of the Mills ratio when calculating the marginal effects of explantory variables common to both stages in the two-stage Heckman approach to estimating dichotomous–continuous models.

REFERENCES

Blaylock, J.R and Blisard W.N. (1992) US cigarette consumption: the case of low-income women, *American Journal of Agricultural Economics*, **74**, 698–705.

Cheng, H. and Capps, O. Jr (1988) Demand analysis of fresh and frozen fish and shellfish in the United States, *American Journal of Agricultural Economics*, **70**, 533–42.

Haines, P.S., Guilley, D.K. and Popkin, B.M. (1988) Modeling food consumption decisions as a two-step process, *American Journal of Agricultural Economics*, **70**, 543–52.

Heckman, J.J. (1976) The common structure of statistical models of truncation. Sample selection and limited dependent variables and a simple estimation for such models, *Annals of Economic and Social Measurements*, **5**, 475–92.

Heien, D. and Durham, C. (1991) A test of the habit formation hypothesis using household data, *Review of Economics and Statistics*, **73**, 189–99.

Heien, D. and Wessells, C.R. (1990) Demand systems estimation with microdata: a censored regression approach *Journal of Business and Economic Statistics*, **8**, 365–71.

Pompelli, G. and Heien, D. (1992) Discrete continuous consumer demand choices: an application to the US domestic and imported white wine markets, *European Review of Agricultural Economics*, **18**, 117–30.

Yen, S.T. (1993) Working wives and food away from home: the Box–Cox double hurdle models, *American Journal of Agricultural Economics*, **75**, 884–95.