HUDM 5123 - Linear Models and Experimental Design
09 - Logistic Regression

1 Logistic Regression for Dichotomous Outcomes

In regression analysis, we model the conditional expectation (i.e., conditional average) of the outcome, given the predictors. In multiple linear regression with continuous outcome variable a linear model is specified as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i.$$

Taking expected values on both sides and using the fact that the error term is assumed to have an expected value of 0 over all values of the predictors gives

$$E[Y_i | X_{i1}, X_{i2}, \ldots, X_{ip}] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}.$$

Assume, without loss of generality, that a dichotomous outcome variable is coded as 0/1. Recall that the expected value (average) of a dichotomous random variable is the probability that the variable takes on the value 1. That is, if Y_i is dichotomous 0/1, then

$$E[Y_i | X_{i1}, X_{i2}, \ldots, X_{ip}] = p[Y_i = 1 | X_{i1}, X_{i2}, \ldots, X_{ip}].$$

As we saw last class, we may use the usual linear model specification by modeling Y_i via a linear combination of the predictors, and, in some cases this may produce satisfactory results. But, in other cases, this approach might produce nonsensical predicted values and necessarily violates assumptions required for valid inferences about regression coefficients. Importantly, the linear specification will produce a range equal to $(-\infty, \infty)$ even though only values between 0 and 1 are allowed. To restrict the range to $(0, 1)$, we need a function $f : (-\infty, \infty) \rightarrow (0, 1)$. Ideally, the function will be simple to write, continuous, differentiable, and monotonic. The *logistic function*, which has all these properties, is a good candidate and may be expressed as follows:

$$f(t) = \frac{\exp(t)}{1 + \exp(t)}.$$

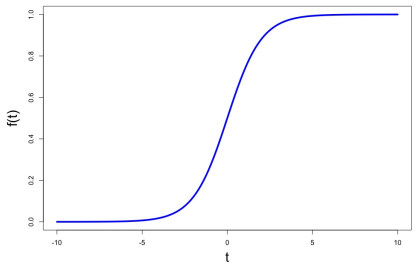

Figure 1: The logistic function f.

Imagine a case of simple regression (i.e., with only a single predictor, X). Applying the logistic function to the linear expression based on the covariates gives the following three ways of expressing the equation.

$$\Pr(Y = 1|X) = \frac{\exp(\beta_0 + \beta_1 X}{1 + \exp(\beta_0 + \beta_1 X} \qquad \text{Probability}$$

$$\frac{\Pr(Y = 1|X)}{1 - \Pr(Y = 1|X)} = \exp(\beta_0 + \beta_1 X) \qquad \text{Odds}$$

$$\log\left(\frac{\Pr(Y = 1|X)}{1 - \Pr(Y = 1|X)}\right) = \beta_0 + \beta_1 X \qquad \text{Log-odds or Logit}$$

Note that the logit is linear in X.

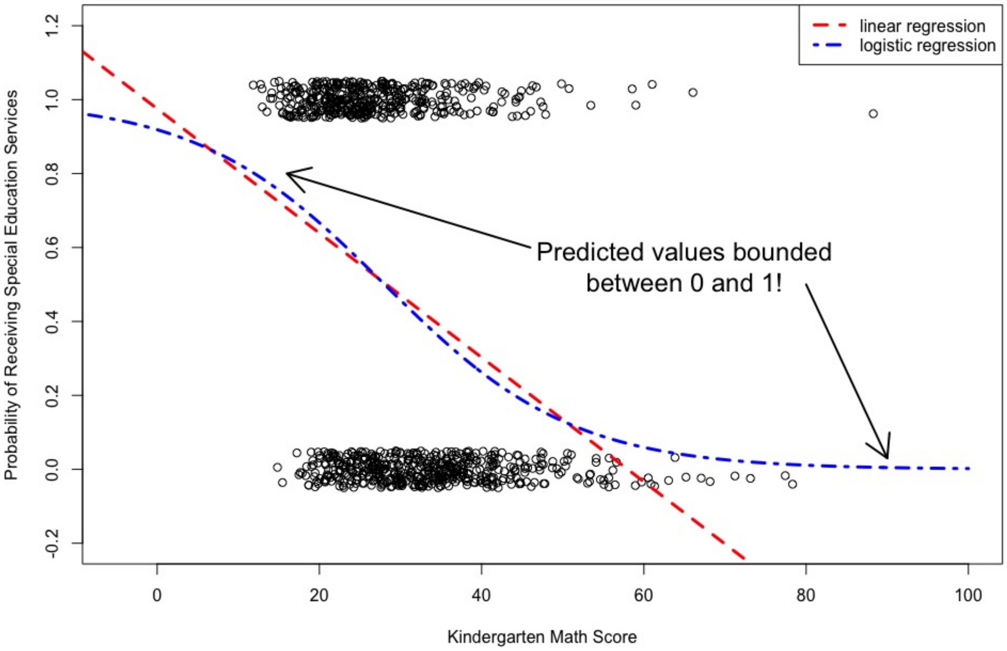

Figure 2: Linear and logistic regression of a dichotomous outcome on a continuous predictor.

The idea is the same with multiple predictors.

$$\Pr(Y = 1|\mathbf{X}) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)} \qquad \text{Probability}$$

$$\frac{\Pr(Y = 1|\mathbf{X})}{1 - \Pr(Y = 1|\mathbf{X} = \mathbf{x})} = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p) \qquad \text{Odds}$$

$$\log\left(\frac{\Pr(Y = 1|\mathbf{X})}{1 - \Pr(Y = 1|\mathbf{X} = \mathbf{x})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \qquad \text{Log-odds or Logit}$$

1.1 Two Examples

1.1.1 Dichotomous Covariate

Consider the relationship between child gender (0 = female, 1 = male) and assignment to special education services (0 = no special education, 1 = special education) in first grade based on the ECLS-K data set. Both variables are dichotomous. Special education services is the outcome variable and child gender is the predictor of interest. Frequency counts and proportions (conditional on gender) are displayed in the following tables based on R output:

```
(tab2 <- table(specEd, gend)) # FREQUENCY
          gend
specEd     Female Male
  NoSpecED   3565 3368
  SpecEd      140  289

(prop2 <- prop.table(x = tab2, margin = 2)) # PROPORTIONS
          gend
specEd     Female Male
  NoSpecED   0.96 0.92
  SpecEd     0.04 0.08
```

Next, we run a logistic regression of the following form:

$$\log\left(\frac{\Pr(S_i = 1|G_i = g_i)}{1 - \Pr(S_i = 1|G_i = g_i)}\right) = \beta_0 + \beta_1 G_i.$$

In words, the log of the odds of a student being assigned to special education are modeled by a linear transformation of student gender. The output from R provides maximum likelihood estimates for the coefficients in the linear part of the model. Here, those are β_0 and β_1.

```
Call:
glm(formula = specEd ~ gend, family = binomial)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.23728    0.08616 -37.573  < 2e-16 ***
gendMale     0.78163    0.10574   7.392 1.44e-13 ***
```

This tells us that for these data, we may write the (estimated) prediction equation as follows:

$$\log\left(\frac{\Pr(S_i = 1|G_i = g_i)}{1 - \Pr(S_i = 1|G_i = g_i)}\right) = -3.24 + 0.78 G_i.$$

First of all, note that the coefficient on the gender variable is significantly different from zero, and estimated to have a value of 0.78. What does that mean about the relationship between student gender and the probability of assignment to special education services in first grade? For that, we gain insight by transforming to the odds scale.

$$\begin{aligned}
\log\left(\frac{\Pr(S_i = 1|G_i = g_i)}{1 - \Pr(S_i = 1|G_i = g_i)}\right) &= -3.24 + 0.78G_i \\
\exp\left[\log\left(\frac{\Pr(S_i = 1|G_i = g_i)}{1 - \Pr(S_i = 1|G_i = g_i)}\right)\right] &= \exp\left[-3.24 + 0.78G_i\right] \\
\frac{\Pr(S_i = 1|G_i = g_i)}{1 - \Pr(S_i = 1|G_i = g_i)} &= \exp(-3.24)\exp(0.78G_i) \\
\frac{\Pr(S_i = 1|G_i = g_i)}{1 - \Pr(S_i = 1|G_i = g_i)} &= \exp(-3.24)\exp(0.78)^{G_i} \\
\frac{\Pr(S_i = 1|G_i = g_i)}{1 - \Pr(S_i = 1|G_i = g_i)} &= 0.04 \cdot 2.18^{G_i}
\end{aligned}$$

The simplified equation above is on the odds scale. For female students (i.e., $G_i = 0$), the model-predicted odds of assignment to special education are 0.04. For male students (i.e., $G_i = 1$), the model-predicted odds of assignment to special education are $0.04(2.18) = 0.087$. Thus, based on this analysis, the model predicts that for females students, the odds of assignment to special education are about 4%. For male students, on the other hand, the odds of assignment to special education are about 8 or 9%. The exponentiated estimate of the gender coefficient, 2.18, represents the estimated *odds ratio*; it is the multiplicative change in the odds of assignment to special education brought about by a one unit change in the predictor (i.e., going from female to male). In this case, being a male student increases the odds of assignment to special education by $(2.18 - 1.00) * 100 = 118\%$.

1.1.2 Continuous Covariate

Next we will use socioeconomic status as the predictor. Socioeconomic status was measured on a continuous scale ranging from about negative three to three. Just as in the last case, we will model the log of the odds as a linear function of the predictor. Also, like last time, the interpretability of the estimated coefficients is improved by using the odds scale. The output is as follows:

```
Call:
glm(formula = F5SPECS ~ WKSESL, family = "binomial", data = eclsk_c)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.75594    0.04990 -55.232  < 2e-16 ***
WKSESL      -0.37646    0.06441  -5.845 5.07e-09 ***
```

The same operations as above yield the following prediction equation on the odds scale:

$$\frac{\Pr(S_i = 1|SES_i = ses_i)}{1 - \Pr(S_i = 1|SES_i = ses_i)} = 0.06 \cdot 0.69^{SES_i}.$$

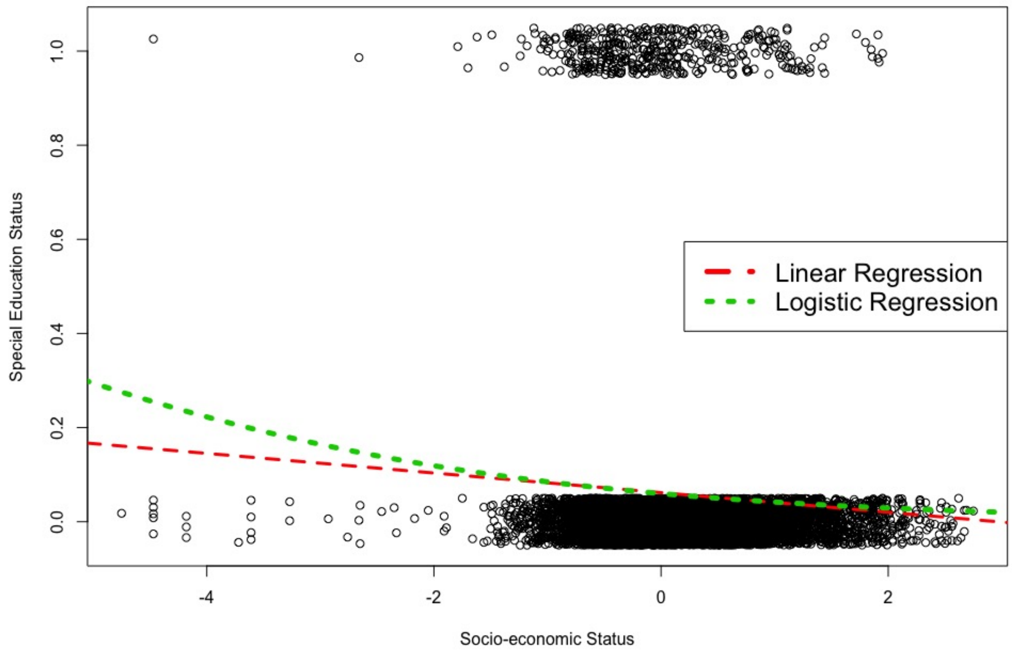

Figure 3: Linear and logistic regression of a dichotomous outcome on a continuous predictor.

The odds of special education assignment for a student with SES = 0 is about 6%. Each one unit increase in SES decreases the odds ratio by about 31%.

1.2 Multiple Predictors

Suppose we include both SES and Head Start participation (0 = no, 1 = yes) as predictors in the model. Although the interaction is not significant, we will leave it in the model for illustrative purposes.

```
Call:
glm(formula = F5SPECS ~ WKSESL * P1HSEVER, family = "binomial", data = eclsk_c)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.79100    0.05546 -50.324  < 2e-16 ***
WKSESL          -0.39378    0.07850  -5.016 5.27e-07 ***
P1HSEVER         0.38471    0.15731   2.446   0.0145 *
WKSESL:P1HSEVER  0.29718    0.18388   1.616   0.1061
```

$$\begin{aligned}
&\text{Overall:}\\
\frac{\Pr(S_i = 1|SES_i = ses_i, HS_i = hs_i)}{1 - \Pr(S_i = 1|SES_i = ses_i, HS_i = hs_i)} &= 0.04 \cdot 0.67^{SES_i} \cdot 1.47^{HS_i} \cdot 1.34^{SES_i \times HS_i}\\
&\text{Not in HS Program:}\\
\frac{\Pr(S_i = 1|SES_i = ses_i, HS_i = 0)}{1 - \Pr(S_i = 1|SES_i = ses_i, HS_i = 0)} &= 0.04 \cdot 0.67^{SES_i}\\
&\text{In HS Program:}\\
\frac{\Pr(S_i = 1|SES_i = ses_i, HS_i = 1)}{1 - \Pr(S_i = 1|SES_i = ses_i, HS_i = 1)} &= 0.04 \cdot 0.67^{SES_i} \cdot 1.47 \cdot 1.34^{SES_i}\\
&= 0.04 \cdot 1.47 \cdot 0.67^{SES_i} \cdot 1.34^{SES_i}\\
&= 0.06 \cdot 0.67^{SES_i} \cdot 1.34^{SES_i}\\
&= 0.06 \cdot (0.67 \cdot 1.34)^{SES_i}\\
&= 0.06 \cdot 0.90^{SES_i}
\end{aligned}$$

Because of the interaction, the multiplicative effect of SES on the odds ratio differs by HS group. In the HS group, a one unit increase in SES decreases the odds of special education by only 10%. In the non-HS group, a one unit increase in SES decreases the odds of special education by $(.69 - 1.00) * 100 = 33\%$.

2 Multinomial Logistic Regression for Polytomous Outcomes

Here we will consider unordered categorical (i.e., nominal) variables with more than two categories. Some examples:

- Political party: D, I, R
- Marital status: married, divorced, never married, widowed
- Type of school: public/non-charter, public/charter, private/relig affil, private/no relig affil
- Blood type: A, B, AB, O

Suppse the polytomous outcome variable has M categories, and call them $1, 2, \ldots, M$, though note that the numbers do not imply a meaningful ordering of the categories; they are simply used for convenient indexing. Also assume that we have p predictors, $\mathbf{X} = X_1, X_2, \ldots, X_p$. Define $\pi_{ij}(\mathbf{X_i}) = \Pr(Y_i = j|\mathbf{X}_i)$ for $j = 1, \ldots, M$.

$$\begin{aligned}
\pi_{i1}(\mathbf{X_i}) =& \Pr(Y_i = 1|\mathbf{X}_i),\\
\pi_{i2}(\mathbf{X_i}) =& \Pr(Y_i = 2|\mathbf{X}_i),\\
&\vdots\\
\pi_{iM}(\mathbf{X_i}) =& \Pr(Y_i = M|\mathbf{X}_i).
\end{aligned}$$

Note that for any unit i, we assume the unit is in at least one of the M categories. In other words, the sum of the probabilities of category membership for each unit i is one. That is,

$$\sum_{j=1}^{M} \pi_{ij} = 1.$$

Thus, for every i, if we know $M-1$ category probabilities, we can find the third by adding them up and subtracting from one. The usual approach to handle the redundancy is to label one response category as the *reference* category. The reference category is typically chosen because (a) comparisons with it are meaningful or (b) it is the largest category. We assume the relationship between the outcome and the predictors can be modeled using the *multivariate logistic distribution*. If we assume the reference category is the last category M, then,

$$\pi_{ij} = \frac{\exp(\beta_{0j} + \beta_{1j}X_{i1} + \beta_{2j}X_{i2} + \cdots + \beta_{pj}X_{ip})}{1 + \sum_{k=1}^{M-1} \exp(\beta_{0k} + \beta_{1k}X_{i1} + \beta_{2k}X_{i2} + \cdots + \beta_{pk}X_{ip})}$$

for $j = 1, 2, \ldots, M-1$, and

$$\pi_{iM} = 1 - \sum_{j=1}^{M-1} \pi_{ij}$$

for category M.

It is possible to show that the regression coefficients for category j, when category M is held out as reference, represent the strength of relationship between the log-odds of membership in category j versus membership in the baseline category M. That is,

$$log\frac{\pi_{ij}}{\pi_{iM}} = \beta_{0j} + \beta_{1j}X_{i1} + \beta_{2j}X_{i2} + \cdots + \beta_{pj}X_{ip}.$$

This relationship also implies that we can determine the log-odds of membership in any pair of categories j and j' by taking the differences in regression coefficients for the two categories.

$$\begin{aligned} log\frac{\pi_{ij}}{\pi_{ij'}} &= log\frac{\pi_{ij}/\pi_{iM}}{\pi_{ij'}/\pi_{iM}} \\ &= log\frac{\pi_{ij}}{\pi_{iM}} - log\frac{\pi_{ij'}}{\pi_{iM}} \\ &= (\beta_{0j} - \beta_{0j'}) + (\beta_{1j} - \beta_{1j'})X_{i1} + \cdots + (\beta_{pj} - \beta_{pj'})X_{ip}. \end{aligned}$$

2.1 Multinomial Logistic Regression Example

An example from Fox deals with data from the British Election Panel Study (BEPS). There are 1525 cases on 10 variables. The outcome is the political party the participant identifies with. The three choices are conservative, labour, and liberal democrat. The other variables in the data frame are

- `Europe`; 11-point scale, attitude to European integration, high score is Eurosceptic
- `Leader_Cons`; Assessment of Conservative leader Hague, 1-5
- `Leader_Labour`; Assessment of Labour leader Blair, 1-5
- `Leader_Liberals`; Assessment of Liberals leader Kennedy, 1-5
- `Age`
- `Gender`
- `Political_Knowledge`; Knowledge of parties' positions on European integration, 0-3
- `National_Economy`; Assessment of current national economic conditions, 1-5
- `Household`; Assessment of current household economic conditions, 1-5

The research question for these data is related to the presence of an interaction between attitude toward European integration and knowledge about the party platforms. The other variables are included as controls. There are three categories, so $M = 3$, where $j = 1$ represents "Conservative", $j = 2$ represents "Labour", and $j = 3$ represents "Liberal Democrat." We will use the "Liberal Democrat" group (i.e., $j = 3$) as the reference group.

$$\begin{aligned} \pi_{i1} = P(Y_i = \text{Conservative}) &= \frac{\exp(\eta_{i1})}{1 + \exp(\eta_{i1}) + \exp(\eta_{i2})} \\ \pi_{i2} = P(Y_i = \text{Labour}) &= \frac{\exp(\eta_{i2})}{1 + \exp(\eta_{i1}) + \exp(\eta_{i2})} \\ \pi_{i3} = P(Y_i = \text{Liberal Democrat}) &= 1 - \pi_{i1} - \pi_{i2}, \end{aligned}$$

where

$$\begin{aligned}\eta_{i1} =& \beta_{01} + \beta_{11}\texttt{Europe}_i + \beta_{21}\texttt{Leader\_Cons}_i + \beta_{31}\texttt{Leader\_Labour}_i + \beta_{41}\texttt{Leader\_Liberals}_i + \\ & \beta_{51}\texttt{Age}_i + \beta_{61}\texttt{Gender}_i + \beta_{71}\texttt{Political\_Knowledge}_i + \beta_{81}\texttt{National\_Economy}_i + \\ & \beta_{91}\texttt{Household}_i + \beta_{10,1}\texttt{Europe} \times \texttt{Political\_Knowledge}_i\end{aligned}$$

is the linear predictor for the Conservative group, and

$$\begin{aligned}\eta_{i2} =& \beta_{02} + \beta_{12}\texttt{Europe}_i + \beta_{22}\texttt{Leader\_Cons}_i + \beta_{32}\texttt{Leader\_Labour}_i + \beta_{42}\texttt{Leader\_Liberals}_i + \\ & \beta_{52}\texttt{Age}_i + \beta_{62}\texttt{Gender}_i + \beta_{72}\texttt{Political\_Knowledge}_i + \beta_{82}\texttt{National\_Economy}_i + \\ & \beta_{92}\texttt{Household}_i + \beta_{10,2}\texttt{Europe} \times \texttt{Political\_Knowledge}_i\end{aligned}$$

is the linear predictor for the Labour group. See Table 1 for estimated regression coefficients. Predicted probabilities of group membership may be determined as follows.

$$\begin{aligned}\hat{\pi}_{i1} = P(Y_i = \text{Conservative}) &= \frac{\exp(\hat{\eta}_{i1})}{1 + \exp(\hat{\eta}_{i1}) + \exp(\hat{\eta}_{i2})} \\ \hat{\pi}_{i2} = P(Y_i = \text{Labour}) &= \frac{\exp(\hat{\eta}_{i2})}{1 + \exp(\hat{\eta}_{i1}) + \exp(\hat{\eta}_{i2})} \\ \hat{\pi}_{i3} = P(Y_i = \text{Liberal Democrat}) &= 1 - \hat{\pi}_{i1} - \hat{\pi}_{i2},\end{aligned}$$

where

$$\begin{aligned}\hat{\eta}_{i1} =& 0.72 - 0.07\texttt{Europe}_i + 0.78\texttt{Leader\_Cons}_i - 0.28\texttt{Leader\_Labour}_i - 0.66\texttt{Leader\_Liberals}_i + \\ & 0.02\texttt{Age}_i - 0.09\texttt{Gender}_i - 1.16\texttt{Political\_Knowledge}_i - 0.15\texttt{National\_Economy}_i - \\ & 0.01\texttt{Household}_i + 0.18\texttt{Europe} \times \texttt{Political\_Knowledge}_i\end{aligned}$$

is the estimated linear predictor for the Conservative group, and

$$\begin{aligned}\hat{\eta}_{i2} =& -0.16 - 0.07\texttt{Europe}_i - 0.09\texttt{Leader\_Cons}_i + 0.55\texttt{Leader\_Labour}_i - 0.42\texttt{Leader\_Liberals}_i - \\ & 0.01\texttt{Age}_i + 0.02\texttt{Gender}_i - 0.50\texttt{Political\_Knowledge}_i + 0.38\texttt{National\_Economy}_i + \\ & 0.17\texttt{Household}_i + 0.02\texttt{Europe} \times \texttt{Political\_Knowledge}_i\end{aligned}$$

is the estimated linear predictor for the Labour group. The significant interaction between Euro-skepticism and political knowledge may be easier to understand through a plot that shows estimated probabilites as a function of Euro-skepticism, broken out by the four levels of the political knowledge variable. Note that we only wish alter the levels of two variables for the plots: Euro-skepticism and political knowledge; we simply wish to control for the other variables. To do that, when making predictions, we will use the *average* (mean) values of the other predictors in the prediction equation.

The range of the Euro-skepticism variable ("Europe") is from 1 to 11. The range of the political knowledge variable is from 1 to 4. Begin by conditioning on political knowledge = 1 and Euro-skepticism = 1. The calculation of the linear predictor for the Conservative group:

$$\begin{aligned}\hat{\eta}_{i1} =& 0.72 - 0.07(1) + 0.78(2.75) - 0.28(3.33) - 0.66(3.14) + \\ & 0.02(54.2) - 0.09(.47) - 1.16(1) - 0.15(3.25) - \\ & 0.01(3.14) + 0.18(1) \times (1) = -0.667.\end{aligned}$$

The calculation of the linear predictor for the Labour group:

$$\begin{aligned}\hat{\eta}_{i2} = & -0.16 - 0.07(1) - 0.09(2.75) + 0.55(3.33) - 0.42(3.14) - \\ & 0.01(54.2) + 0.02(.47) - 0.50(1) + 0.38(3.25) + \\ & 0.17(3.14) + 0.02(1) \times (1) = 0.791.\end{aligned}$$

Finally, the model-predicted probabilities for party membership given that Euro-skepticism is 1, political knowledge is 1, and other variables are held at their means are shown below. Compare these probabilities with the upper right panel of the plot given below at a value of Euro-skepticism = 1.

$$\begin{aligned}\hat{\pi}_{i1} = P(Y_i = \text{Conservative}) &= \frac{\exp(-0.667)}{1 + \exp(-0.667) + \exp(0.791)} = 0.14 \\ \hat{\pi}_{i2} = P(Y_i = \text{Labour}) &= \frac{\exp(0.791)}{1 + \exp(-0.667) + \exp(0.791)} = 0.59 \\ \hat{\pi}_{i3} = P(Y_i = \text{Liberal Democrat}) &= 1 - 0.14 - 0.59 = 0.27.\end{aligned}$$

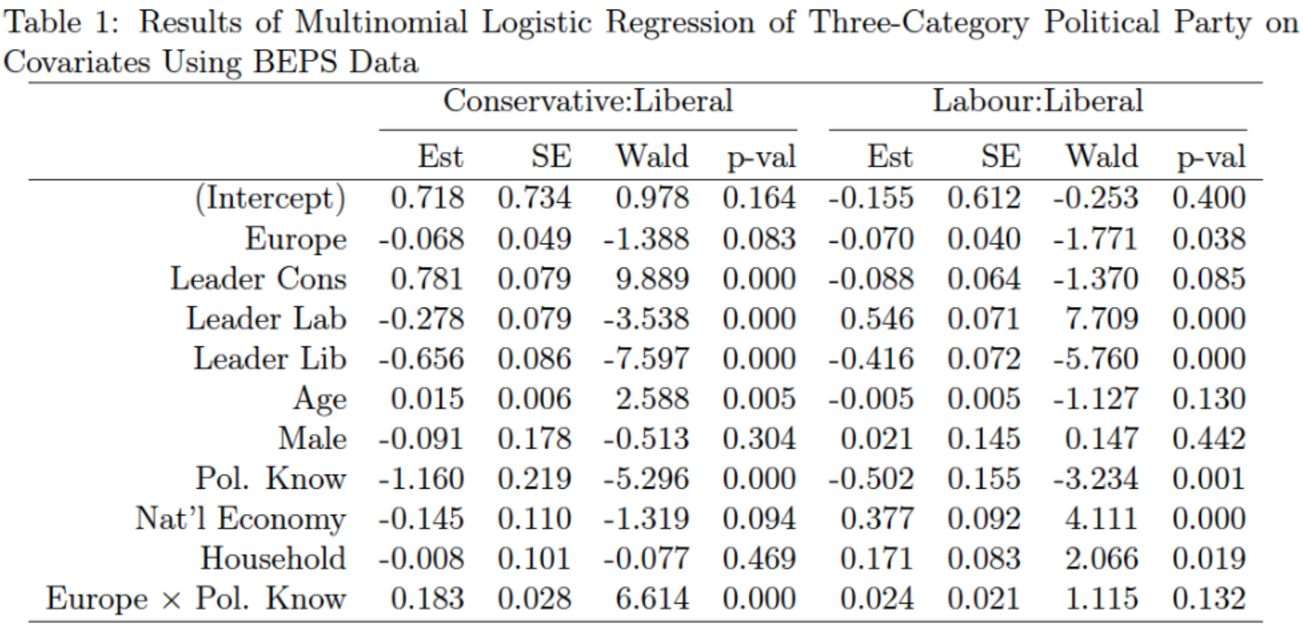

Table 1: Results of Multinomial Logistic Regression of Three-Category Political Party on Covariates Using BEPS Data

	Conservative:Liberal				Labour:Liberal			
	Est	SE	Wald	p-val	Est	SE	Wald	p-val
(Intercept)	0.718	0.734	0.978	0.164	-0.155	0.612	-0.253	0.400
Europe	-0.068	0.049	-1.388	0.083	-0.070	0.040	-1.771	0.038
Leader Cons	0.781	0.079	9.889	0.000	-0.088	0.064	-1.370	0.085
Leader Lab	-0.278	0.079	-3.538	0.000	0.546	0.071	7.709	0.000
Leader Lib	-0.656	0.086	-7.597	0.000	-0.416	0.072	-5.760	0.000
Age	0.015	0.006	2.588	0.005	-0.005	0.005	-1.127	0.130
Male	-0.091	0.178	-0.513	0.304	0.021	0.145	0.147	0.442
Pol. Know	-1.160	0.219	-5.296	0.000	-0.502	0.155	-3.234	0.001
Nat'l Economy	-0.145	0.110	-1.319	0.094	0.377	0.092	4.111	0.000
Household	-0.008	0.101	-0.077	0.469	0.171	0.083	2.066	0.019
Europe × Pol. Know	0.183	0.028	6.614	0.000	0.024	0.021	1.115	0.132

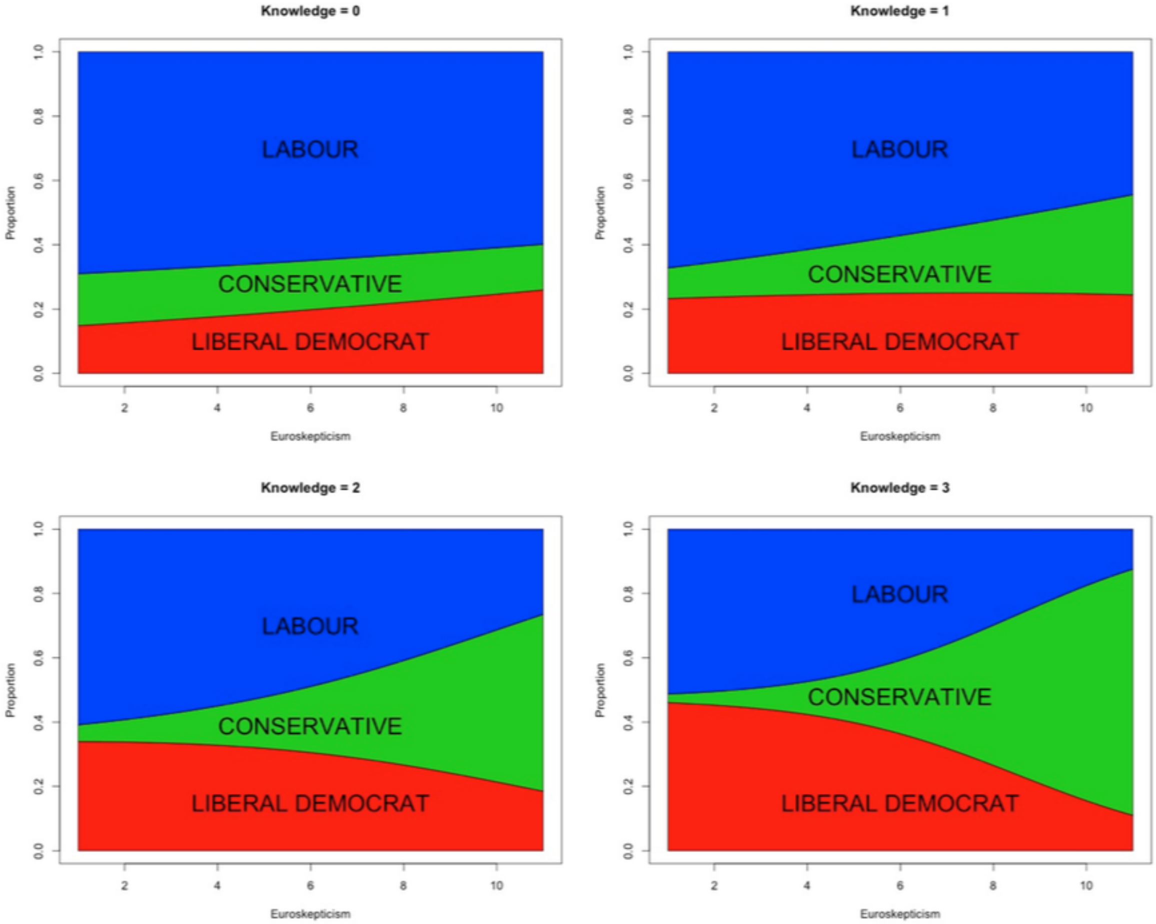

Figure 4: Plots of the Interaction between attitude toward European integration and political knowledge for the 2001 BEPS data