Logit and Probit: Interpretation and Hypothesis Testing

Categorical and Limited Dependent Variables

Paul A. Jargowsky

Summary

- Interpretation
  - A: Predicted Values
  - B: Significance of the Coefficients
  - C: Marginal effects
  - D: Discrete changes
  - E: Odds (in Logit)
  - F: Goodness of Fit Measures
  - G: Hypothesis Testing

Review

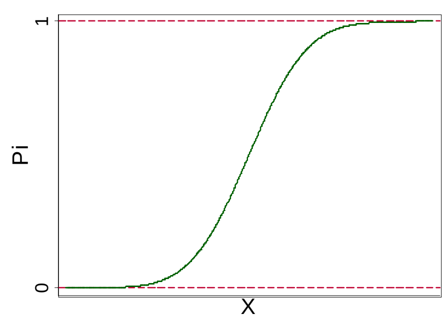
$$Y_i = \begin{cases} 1 & \text{event occurs} \\ 0 & \text{does not occur} \end{cases} \qquad Y_i = \mathrm{E}\left[Y_i \mid X_i\right] + u_i$$

E[Y|X] is the probability that Y=1 given X (or the Xs if more than one independent variable). It is the *conditional mean* of Y given **X**.

X has an effect on this probability, but it can't be linear, because probabilities are bounded by zero and one. So, P_i must be some non-linear function of **X** that is bounded by zero and one.

Convenient Functional Forms

$$P_i = F\left(\mathbf{x}_i\boldsymbol{\beta}\right)$$



Two functional forms that work well:

Logit $\qquad P_i = \dfrac{1}{1+e^{-\mathbf{x}_i\boldsymbol{\beta}}}, \quad e = 2.71828...$

Probit $\qquad P_i = \Phi\left(\mathbf{x}_i\boldsymbol{\beta}\right), \quad \Phi$ is cum. std. normal

Estimation by MLE

Since we know the probability of each point (either logit or probit) as a function of the data and the parameters to be estimated, we can use maximum likelihood.

If P_i is the probability that Y_i=1, then (1-P_i) is the probability that Y_i=0. Then the probability for a given point is:

$$\Pr\left(Y = Y_i \mid X_i\right) = P_i^{Y_i}\left(1-P_i\right)^{(1-Y_i)}$$

$$= \begin{cases} P_i^1\left(1-P_i\right)^0 = P_i & \text{if } Y_i = 1 \\ P_i^0\left(1-P_i\right)^1 = 1-P_i & \text{if } Y_i = 0 \end{cases}$$

Likelihood of the Sample

$$\mathcal{L} = \prod_{i=1}^{n}\left[P_i^{Y_i}\left(1-P_i\right)^{(1-Y_i)}\right]$$ Assuming independence

$$\ln \mathcal{L} = \sum_{i=1}^{n} Y_i \ln P_i + \sum_{i=1}^{n}\left(1-Y_i\right)\ln\left(1-P_i\right)$$

Plug in either $P_i = \frac{1}{1+e^{-\mathbf{x}_i\boldsymbol{\beta}}}$ or $P_i = \Phi\left(\mathbf{x}_i\boldsymbol{\beta}\right)$.

Send the computer in seach of estimators $\left(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k\right)$ that maximize the log of the likelihood function.

Example: Logit

```
. logit low age smoke

Iteration 0:   log likelihood =   -117.336
Iteration 1:   log likelihood = -113.66733
Iteration 2:   log likelihood = -113.63815
Iteration 3:   log likelihood = -113.63815

Logistic regression                               Number of obs   =        189
                                                  LR chi2(2)      =       7.40
                                                  Prob > chi2     =     0.0248
Log likelihood = -113.63815                       Pseudo R2       =     0.0315

------------------------------------------------------------------------------
         low |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0497792    .031972    -1.56   0.119    -.1124431    .0128846
       smoke |   .6918486   .3218061     2.15   0.032     .0611202    1.322577
       _cons |   .0609051   .7573199     0.08   0.936    -1.423415    1.545225
------------------------------------------------------------------------------
```

$$\mathbf{x}_i\hat{\boldsymbol{\beta}} = 0.0609 - 0.0498 age_i + 0.692 smoke_i$$

A. Predicted Values

$$\mathbf{x}_i\hat{\boldsymbol{\beta}} = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_K X_{Ki} \qquad \hat{P}_i = \frac{1}{1+e^{-\mathbf{x}_i\hat{\boldsymbol{\beta}}}}$$

- What is predicted probability of low birthweight for a 25 year old *non-smoker*?

$$\begin{aligned}\mathbf{x}_i\hat{\boldsymbol{\beta}} &= 0.0609 - 0.0498\left(age_i\right) + 0.692\left(smoke_i\right) \\ &= 0.0609 - 0.0498(25) + 0.692(0) \\ &= -1.184\end{aligned}$$

$$\hat{P}_i = \frac{1}{1+e^{-(-1.184)}} = \frac{1}{1+e^{1.184}} = 0.234$$

B: Significance of Coefficients

$\hat{\beta}_{MLE}$ is consistent and asymptotically normally distributed. Thus, in large samples:

$$\hat{\beta}_{MLE} \sim N\left(\beta, \sigma_{\hat{\beta}}^{\ 2}\right) \rightarrow z = \frac{\hat{\beta}_{MLE} - \beta_0}{\sigma_{\hat{\beta}}} \sim N(0,1)$$

So, for the null hypothesis that $\beta_k = 0$, the decision rule is to reject the null if $|z| > 1.96$, $z = \frac{\hat{\beta}_k}{s_{\hat{\beta}_k}}$, similar to OLS.

$$z_{\hat{\beta}_{age}} = \frac{-0.0498}{0.0320} = -1.56 \qquad z_{\hat{\beta}_{smoke}} = \frac{0.692}{0.322} = 2.15$$

C: Marginal Effects

$\frac{\partial \hat{P}_i}{\partial X_k} = \hat{P}_i\left(1 - \hat{P}_i\right)\hat{\beta}_k$ which implies that....

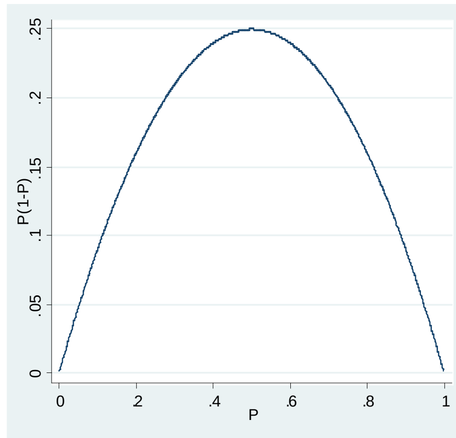
1. the **sign** of $\hat{\beta}_k$ tells you the **direction**
2. **maximum effect** is $0.25\hat{\beta}_k$ at $\hat{P}_i = 0.5$
3. As $\hat{P}_i$ approaches 1 or 0, the effect **diminishes**
4. the effect of X_k **depends on all the other X variables**, so you have to choose values

Valid for "small" changes in continuous variables

How the marginal effect varies

$$\frac{\partial \hat{P}_i}{\partial age} = \hat{P}_i\left(1-\hat{P}_i\right)\hat{\beta}_{age} = \hat{P}_i\left(1-\hat{P}_i\right)(-0.0498)$$

P	$1-P$	$P(1-P)$	$\frac{\partial \hat{P}_i}{\partial \exp}$
.1	.9	.09	−0.0045
.2	.8	.16	−0.0080
.3	.7	.21	−0.0105
.4	.6	.24	−0.0120
.5	.5	.25	−0.0125
.6	.4	.24	?
.7	.3	.21	?
.8	.2	.16	?
.9	.1	.09	?



D: Discrete Changes

$$\frac{\Delta \Pr(Y=1\mid \mathbf{x})}{\Delta X_k} = \Pr(Y=1\mid \mathbf{x}, X_k+\delta) - \Pr(Y=1\mid \mathbf{x}, X_k)$$

- The change in predicted probability when X_k changes by δ holding the other variables constant
- Also depends on all *Xs*.
- Use for change in dummy from 0 to 1.
- Use for larger changes in a single *X*.
- Use for changes in multiple *Xs* (e.g. comparing male dropout to female high school graduate).
- Never wrong, but sometimes tedious.

Effect of Smoking

Smoker (smoke=1) vs. Non-smoker (smoke=0), holding age constant at 25.

$$\begin{aligned}\mathbf{x}_1\hat{\boldsymbol{\beta}} &= 0.0609 - 0.0498(age_i) + 0.692(smoke_i)\\ &= 0.0609 - 0.0498(25) + 0.692(1)\\ &= -0.492\end{aligned}$$

$$P_1 = \frac{1}{1+e^{-(-0.492)}} = 0.379$$

$$\Delta P = P_1 - P_0 = 0.379 - 0.234 = 0.145$$

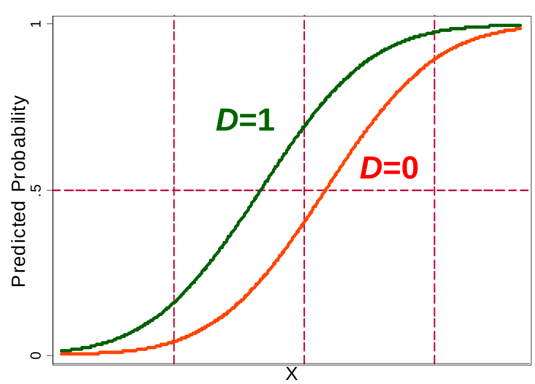
Depends on Age!

The effects vary with X

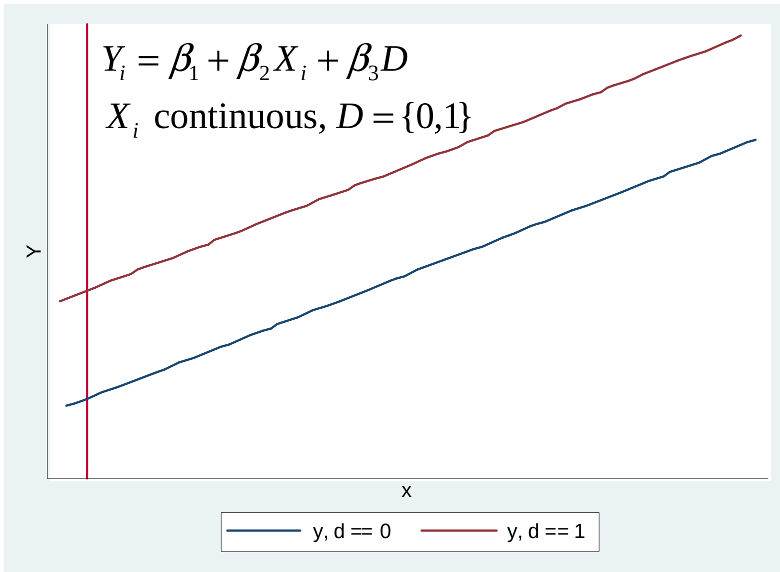
$$\mathbf{x}_i\boldsymbol{\beta} = \beta_1 + \beta_2 X_i + \beta_3 D \quad X_i \text{ continuous, } D = \{0,1\}$$

Slope varies for X, reaching a maximum when the P is closest to 0.5.

Difference (discrete change) in P between D=1 and D=0 holding X constant varies as well (β_3>0)



Compare to OLS



How to present marginal effects

- This is tricky. Depends on the problem.
- There are lots of options:
  - Maximum (theoretical)
  - Minimum and maximum effect *in the data*
  - Effect for mean person (means of all Xs)
  - The mean effect *in the data* ***
  - Discrete changes for interesting scenarios
- For dummies, use a discrete change from 0 to 1 holding others constant
- ****Stata's powerful* `margins` *command*

margins

```
After: logit low age i.smoke

  Average predicted means:
       margins                     // average pred
       margins, atmeans            // pred. at the means
       margins smoke               // averages for groups
       margins, at(age=(20 30 40 50))    // at values
       margins smoke, at(age=(20 30 40 50))

  Marginal effects (dydx option)
       margins, dydx(age) // average effect
       margins, atmeans  // effect at means
       margins smoke, dydx // average effects
       margins, at(age=(20 30 40 50)) dydx(age)
       margins smoke,at(age=(20 30 40 50)) dydx(age)

       ***Long and Freeze “mchange” command
```

E: Odds Ratios (Logit Only)

$$\ln\left(\frac{P_i}{1-P_i}\right) = \mathbf{x}_i\boldsymbol{\beta} \quad \rightarrow \quad \frac{P_i}{1-P_i} = e^{\mathbf{x}_i\boldsymbol{\beta}}$$

Smoker: $odds_1 = \frac{0.379}{1-0.379} = 0.610$ $\quad e^{-0.492} = 0.611$

Non-Smoker: $odds_0 = \frac{0.234}{1-0.234} = 0.305$ $\quad e^{-1.184} = 0.306$

Odds Ratio: $\Omega_{smoke} = \frac{0.610}{0.305} = 2.00$

What happens to the odds if X_2 increases by 1 unit?

$$\mathbf{x}_1\boldsymbol{\beta} = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki}$$

$$\begin{aligned}\mathbf{x}_2\boldsymbol{\beta} &= \beta_1 + \beta_2 \left(X_{2i} + 1\right) + \ldots + \beta_K X_{Ki} \\ &= \beta_1 + \beta_2 X_{2i} + \beta_2 + \ldots + \beta_K X_{Ki}\end{aligned}$$

$$\Omega_{\Delta X_2 = +1} = \frac{\left(\dfrac{P_2}{1-P_2}\right)}{\left(\dfrac{P_1}{1-P_1}\right)} = \frac{e^{\beta_1+\beta_2 X_{2i}+\beta_2 \ldots + \beta_K X_{Ki}}}{e^{\beta_1+\beta_2 X_{2i}+\ldots+\beta_K X_{Ki}}} = e^{\beta_2}$$

Note: this is constant!

$$\Omega_{smoke} = e^{\beta_{smoke}} = e^{0.692} = 2.00$$

Much easier to calculate!

Odds in General

$$\Omega_{X_k} = \frac{e^{\beta_1+\beta_2 X_{2i}+\ldots+\beta_k (X_{ki}+\delta)+\ldots+\beta_K X_{Ki}}}{e^{\beta_1+\beta_2 X_{2i}+\ldots+\beta_k X_{ki}+\ldots+\beta_K X_{Ki}}} = e^{\delta\beta_k}$$

How do the odds change for a one year increase in age?

$$e^{-0.0498} = 0.951$$

How do the odds change for a 10 year increase in exp?

$$e^{(10)(-0.0498)} = 0.608$$

```
. logit low age smoke, or
…
------------------------------------------------------------------------------
         low | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9514394   .0304194    -1.56   0.119     .8936482    1.012968
     1.smoke |   1.997405    .642777     2.15   0.032     1.063027    3.753081
       _cons |   1.062798   .8048781     0.08   0.936     .2408901    4.689025
------------------------------------------------------------------------------
```

Where does the initial log likelihood come from?

```
. tab low

birthweight |
     <2500g |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        130       68.78       68.78
          1 |         59       31.22      100.00
------------+-----------------------------------
      Total |        189      100.00
```

If no X's predict low birthweight, $P_i = 0.312$ for all i.

$$\begin{aligned}\ln L_0 &= \sum_{Y_i=1} \ln(0.312) + \sum_{Y_i=0} \ln(1-0.312) \\ &= 59\ln(0.312) + 130\ln(0.688) \\ &= -68.7 - 48.6 \\ &= -117.3\end{aligned}$$

The Likelihood of Perfect Prediction

If your model was perfect,

$$\text{then } \begin{cases} \hat{P}_i = 1 \text{ if } Y_i = 1 \\ \hat{P}_i = 0 \text{ if } Y_i = 0 \end{cases}$$

$$\begin{aligned}L &= \prod_{i=1}^{n} \left[\left(\hat{P}_i\right)^{Y_i} \left(1-\hat{P}_i\right)^{(1-Y_i)} \right] \\ &= \underbrace{(1)\ldots(1)}_{96 \text{ union members}} \; \underbrace{(1-0)\ldots(1-0)}_{438 \text{ non-members}} \\ &= 1\end{aligned}$$

$$\ln L = 96\ln(1) + 438\ln(1) = 0$$

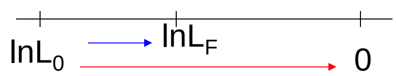
If we could perfect predict who was a union member, and who was not, then the likelihood would be 1 and the log likelihood would be 0.

Thus, 0 is the *theoretical maximum attainable likelihood.*

F: Goodness of Fit

$$R^2_{McFadden} = \frac{\text{Actual Improvement}}{\text{Max Potential Improvement}}$$

$$= \frac{\ln L_F - \ln L_0}{0 - \ln L_0}$$



$$= \frac{\ln L_0 - \ln L_F}{\ln L_0} = 1 - \frac{\ln L_F}{\ln L_0}$$

For the Logit example (see earlier slide):

$$R^2_{McFadden} = 1 - \frac{\ln L_F}{\ln L_0} = 1 - \frac{-117.3}{-113.6} = 0.0326$$

Alternative Fit Measures

$$R^2_{Efron} = 1 - \frac{\sum \left(Y_i - \hat{P}_i\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2} \qquad R^2_{ML} = 1 - \left(\frac{\ln L_0}{\ln L_F}\right)^{\frac{2}{N}}$$

- Information Measures
  - AIC
  - BIC
- Generally, all measures of fit are low in binary models, not much to worry about.

G: Hypothesis Testing

Likelihood Ratio Test – Like the F test in OLS, you must identify unrestricted and restricted models.

$$-2\Delta \ln L \sim \chi^2_m \quad m = \text{number of restrictions}$$

$$\begin{aligned} G^2 &= -2\left(\ln L_R - \ln L_U\right) \\ &= 2\left(\ln L_U - \ln L_R\right) \\ &= 2\ln\left(\frac{L_U}{L_R}\right) \\ &= \ln\left(\frac{L_U{}^2}{L_R{}^2}\right) \end{aligned}$$

Why it's a ratio test.

Why it's distributed as Chi square.

G.1: Testing the Model as a Whole

$$-2\Delta \ln L \sim \chi^2_{K-1} \quad K = \text{ number of parameters}$$

$\ln L_0$: intercept only model, $\beta_2 = \beta_3 = ... = \beta_K = 0$

$\ln L_F$: full model (unrestricted)

$\chi^2 = -2\left(\ln L_0 - \ln L_F\right)$

Reject null if $\chi^2 > \chi^2_{k-1}$ (critical value)

Example:

$$\begin{aligned} \chi^2 &= -2\left(\ln L_0 - \ln L_F\right) \\ &= -2\left(-117.3 + 113.6\right) = 7.4 \qquad \chi^2_{2\text{ DOF}} = 5.99 \end{aligned}$$

G.2: Testing a Subset of Parameters

$-2\Delta \ln L \sim \chi^2_m \quad m =$ number of restrictions

$\ln L_R$: restricted model

$\ln L_F$: full model (unrestricted)

$\chi^2 = -2\left(\ln L_R - \ln L_U\right)$

Reject null if $\chi^2 > \chi^2_m$ (critical value)

The Likelihood Ratio, LaGrange Multiplier, and Wald Hypothesis tests are asymptotically equivalent, so use which ever is convenient.

In Stata: *lrtest* = likelihood ratio test, *test* = Wald test

Example: Probit

```
. probit low age i.smoke

Iteration 0:   log likelihood =   -117.336
Iteration 1:   log likelihood = -113.57013
Iteration 2:   log likelihood = -113.55889
Iteration 3:   log likelihood = -113.55889

Probit regression                               Number of obs     =        189
                                                LR chi2(2)        =       7.55
                                                Prob > chi2       =     0.0229
Log likelihood = -113.55889                     Pseudo R2         =     0.0322

------------------------------------------------------------------------------
         low |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0312426   .0193475    -1.61   0.106     -.069163    .0066777
     1.smoke |   .4241072   .1952064     2.17   0.030     .0415098    .8067046
       _cons |    .051387   .4596968     0.11   0.911    -.8496023    .9523762
------------------------------------------------------------------------------
```

$$\mathbf{x}_i\hat{\boldsymbol{\beta}} = 0.0514 - 0.0312 age_i + 0.424 smoke_i$$

Interpretation: *Probit*

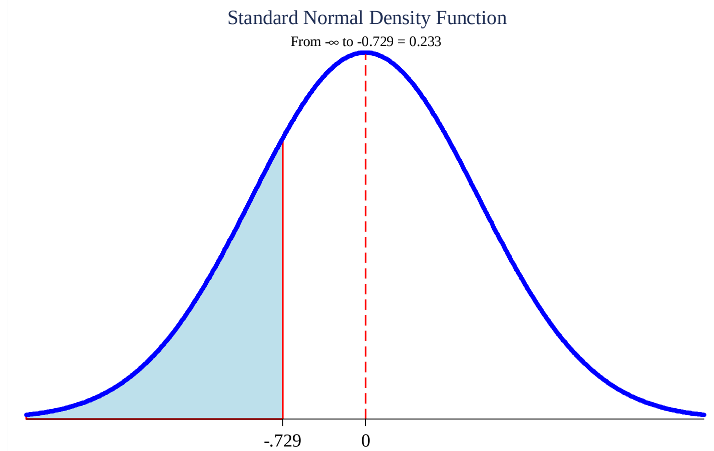
- *Different* from logit:
  - How you calculate the predicted probabilities (A)
  - How you calculate the marginal effects (C, but D is the same)
  - No simple form for Odds Ratios (E)
- *Same* as logit:
  - Sign of coefficient = direction of effect
  - Effect size varies, depends on other Xs
  - Alternative ways to present the effects
  - Goodness of fit measures (F)
  - Hypothesis testing (B, G)

A: Predicted Values in Probit

$$
\begin{aligned}
\mathbf{x}_i\hat{\boldsymbol{\beta}} &= 0.0514 - 0.0312 age_i + 0.4245 smoke_i \\
&= 0.0514 - 0.0312(25) + 0.4245(0) \\
&= -0.729
\end{aligned}
$$

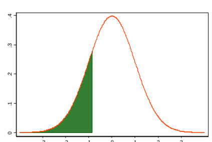
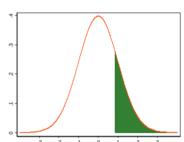
$$
\begin{aligned}
\hat{P}_i &= \Phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right) \\
&= \Phi(-0.729) \\
&= ?
\end{aligned}
$$

Predicted Value: Standard Normal



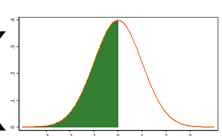
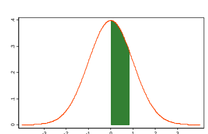

```
. disp normal(-0.729)
.23300083
```

Getting from Normal Table to the Predicted Probability

 = 

$$\Phi(z) = 1 - \Phi(-z)$$

$$\Phi(-0.729) = 1 - \Phi(0.729)$$

= 1 – ( + )

= 1-(0.5+0.267) = 0.233

C: Marginal Effects in Probit

$$\frac{\partial \hat{P}_i}{\partial X_k} = \phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\hat{\beta}_k = \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)^2}\right)\hat{\beta}_k$$

1. the **sign** of $\hat{\beta}_k$ tells you the **direction**
2. **maximum effect** is $0.4\hat{\beta}_k$ at $\mathbf{x}_i\boldsymbol{\beta} = 0$
3. As $\hat{P}_i$ approaches 1 or 0, the effect **diminishes**
4. the effect of X_k **depends on all the other X variables**, so you have to choose values

Valid for "small" changes in continuous variables

Maximum Marginal Effect in Probit

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(0)^2} = \frac{1}{\sqrt{2\pi}} \approx 0.4$$

What P_i does $X_i\beta=0$ correspond to?