Chapter 16 Discussion Problems

Binary Choice, Ordered Probit, Multinomial Logit

Jake Anderson

ECON 104 – UCLA

Outline

Problem 16.18: Setup

Predict mortgage delinquency (90+ days late) for 1,000 Las Vegas mortgages in 2008.

Problem 16.18: Setup

Predict mortgage delinquency (90+ days late) for 1,000 Las Vegas mortgages in 2008.

Variable	**Description**	**Type**
DELINQUENT	1 if 90+ days late	Dependent
LVR	Loan-to-value ratio (%)	Continuous
REF	1 if refinance	Binary
INSUR	1 if mortgage insurance	Binary
RATE	Initial interest rate (%)	Continuous
CREDIT	Credit score	Continuous
AMOUNT	Loan amount ($100K)	Continuous
TERM	Loan term (years)	Continuous
ARM	1 if adjustable rate	Binary

(a) Linear Probability Model

Estimate the LPM with White robust standard errors. Are the signs reasonable?

(a) Linear Probability Model

Estimate the LPM with White robust standard errors. Are the signs reasonable?

Solution:

	Estimate	**Robust SE**	t	
(Intercept)	0.6885	0.2285	3.01	**
LVR	0.0016	0.0007	2.41	*
REF	−0.0593	0.0240	−2.47	*
INSUR	−0.4816	0.0304	−15.86	***
RATE	0.0344	0.0098	3.50	***
AMOUNT	0.0238	0.0145	1.65	
CREDIT	−0.0004	0.0002	−2.13	*
TERM	−0.0126	0.0036	−3.55	***
ARM	0.1283	0.0277	4.63	***

(a) Linear Probability Model

Estimate the LPM with White robust standard errors. Are the signs reasonable?

Solution:

	Estimate	**Robust SE**	t	
(Intercept)	0.6885	0.2285	3.01	**
LVR	0.0016	0.0007	2.41	*
REF	−0.0593	0.0240	−2.47	*
INSUR	−0.4816	0.0304	−15.86	***
RATE	0.0344	0.0098	3.50	***
AMOUNT	0.0238	0.0145	1.65	
CREDIT	−0.0004	0.0002	−2.13	*
TERM	−0.0126	0.0036	−3.55	***
ARM	0.1283	0.0277	4.63	***

All signs are economically sensible. Only *AMOUNT* is not significant at 5%.

$N = 1{,}000$. White heteroskedasticity-robust standard errors.

(b) Logit Model

Estimate using logit. Compare signs and significance with the LPM.

(b) Logit Model

Estimate using logit. Compare signs and significance with the LPM.

	LPM				Logit			
	$\hat{\beta}$	SE	t/z		$\hat{\beta}$	SE	z	
LVR	0.002	0.001	2.41	*	0.014	0.008	1.65	.
REF	−0.059	0.024	−2.47	*	−0.525	0.230	−2.28	*
INSUR	−0.482	0.030	−15.9	***	−3.122	0.217	−14.4	***
RATE	0.034	0.010	3.50	***	0.309	0.083	3.73	***
AMOUNT	0.024	0.014	1.65		0.218	0.114	1.91	.
CREDIT	−0.000	0.000	−2.13	*	−0.004	0.002	−1.84	.
TERM	−0.013	0.004	−3.55	***	−0.133	0.035	−3.85	***
ARM	0.128	0.028	4.63	***	1.406	0.371	3.79	***

(b) Logit Model

Estimate using logit. Compare signs and significance with the LPM.

	LPM				Logit			
	$\hat{\beta}$	SE	t/z		$\hat{\beta}$	SE	z	
LVR	0.002	0.001	2.41	*	0.014	0.008	1.65	.
REF	−0.059	0.024	−2.47	*	−0.525	0.230	−2.28	*
INSUR	−0.482	0.030	−15.9	***	−3.122	0.217	−14.4	***
RATE	0.034	0.010	3.50	***	0.309	0.083	3.73	***
AMOUNT	0.024	0.014	1.65		0.218	0.114	1.91	.
CREDIT	−0.000	0.000	−2.13	*	−0.004	0.002	−1.84	.
TERM	−0.013	0.004	−3.55	***	−0.133	0.035	−3.85	***
ARM	0.128	0.028	4.63	***	1.406	0.371	3.79	***

Same signs throughout. *AMOUNT* is now marginally significant in logit. Logit coefficients are **not** marginal effects.

$N = 1{,}000$. Estimated by MLE; logit coefficients are not directly comparable to LPM.

(c) Predicted Probabilities

Compute predicted probabilities for observations 500 and 1000.

(c) Predicted Probabilities

Compute predicted probabilities for observations 500 and 1000.

	Obs. 500	**Obs. 1000**
LPM	18.3%	57.9%
Logit	13.4%	62.7%

(c) Predicted Probabilities

Compute predicted probabilities for observations 500 and 1000.

	Obs. 500	Obs. 1000
LPM	18.3%	57.9%
Logit	13.4%	62.7%

Worked example (Obs. 500):

$$\hat{P} = \frac{e^{X'\hat{\beta}}}{1 + e^{X'\hat{\beta}}} = \frac{e^{-1.866}}{1 + e^{-1.866}} = \frac{0.155}{1.155} = 0.134$$

(c) Predicted Probabilities

Compute predicted probabilities for observations 500 and 1000.

	Obs. 500	**Obs. 1000**
LPM	18.3%	57.9%
Logit	13.4%	62.7%

Worked example (Obs. 500):

$$\hat{P} = \frac{e^{X'\hat{\beta}}}{1 + e^{X'\hat{\beta}}} = \frac{e^{-1.866}}{1 + e^{-1.866}} = \frac{0.155}{1.155} = 0.134$$

$\Longrightarrow$ Logit pulls low-risk predictions closer to 0 and high-risk closer to 1. The S-curve compresses toward the boundaries.

(d)–(e) Predicted Probabilities by Credit Score

Delinquency probability at $CREDIT = 500, 600, 700$.

($LVR = 80\%$, $RATE = 8\%$, $AMOUNT = \$250\text{K}$, $TERM = 30$, all indicators $= 0$)

(d)–(e) Predicted Probabilities by Credit Score

Delinquency probability at $CREDIT = 500, 600, 700$.

($LVR = 80\%$, $RATE = 8\%$, $AMOUNT = \$250K$, $TERM = 30$, all indicators $= 0$)

	$CREDIT = 500$	$CREDIT = 600$	$CREDIT = 700$
LPM $\hat{P}$	55.3%	50.9%	46.5%
Logit $\hat{P}$	49.7%	40.9%	32.6%

(d)–(e) Predicted Probabilities by Credit Score

Delinquency probability at $CREDIT = 500, 600, 700$.

($LVR = 80\%$, $RATE = 8\%$, $AMOUNT = \$250K$, $TERM = 30$, all indicators $= 0$)

	$CREDIT = 500$	$CREDIT = 600$	$CREDIT = 700$
LPM $\hat{P}$	55.3%	50.9%	46.5%
Logit $\hat{P}$	49.7%	40.9%	32.6%

Marginal effects of moving 100 points:

	$500 \to 600$	$600 \to 700$
LPM	-4.42 pp	-4.42 pp
Logit	-8.81 pp	-8.26 pp

(d)–(e) Predicted Probabilities by Credit Score

Delinquency probability at $CREDIT = 500, 600, 700$.

($LVR = 80\%$, $RATE = 8\%$, $AMOUNT = \$250K$, $TERM = 30$, all indicators $= 0$)

	$CREDIT = 500$	$CREDIT = 600$	$CREDIT = 700$
LPM $\hat{P}$	55.3%	50.9%	46.5%
Logit $\hat{P}$	49.7%	40.9%	32.6%

Marginal effects of moving 100 points:

	$500 \to 600$	$600 \to 700$
LPM	-4.42 pp	-4.42 pp
Logit	-8.81 pp	-8.26 pp

$\Longrightarrow$ LPM: constant ME. Logit: $\text{ME} = \hat{\beta} \cdot \hat{P}(1 - \hat{P})$, larger near $P = 0.5$.

(f) LVR Predictions

Compare predictions for $LVR = 20\%$ vs. $LVR = 80\%$ (with $CREDIT = 600$, other variables as before).

(f) LVR Predictions

Compare predictions for $LVR = 20\%$ vs. $LVR = 80\%$ (with $CREDIT = 600$, other variables as before).

	$LVR = 20\%$	$LVR = 80\%$
LPM	41.2%	50.9%
Logit	23.5%	40.9%

(f) LVR Predictions

Compare predictions for $LVR = 20\%$ vs. $LVR = 80\%$ (with $CREDIT = 600$, other variables as before).

	$LVR = 20\%$	$LVR = 80\%$
LPM	41.2%	50.9%
Logit	23.5%	40.9%

- Logit predicts much **lower** delinquency at low LVR
- Logit has a **larger marginal effect** (+17.4 pp vs. +9.7 pp)

(f) LVR Predictions

Compare predictions for $LVR = 20\%$ vs. $LVR = 80\%$ (with $CREDIT = 600$, other variables as before).

	$LVR = 20\%$	$LVR = 80\%$
LPM	41.2%	50.9%
Logit	23.5%	40.9%

- Logit predicts much **lower** delinquency at low LVR
- Logit has a **larger marginal effect** ($+17.4$ pp vs. $+9.7$ pp)

$\Longrightarrow$ At low LVR, we are on the flat part of the S-curve: probabilities are compressed near zero. The LPM extrapolates linearly and overpredicts.

(g)–(h) Classification and Thresholds

(g) Compare % correct predictions using a threshold of 0.5.

(g)–(h) Classification and Thresholds

(g) Compare % correct predictions using a threshold of 0.5.

Model	**Correct (out of 1000)**
LPM	858
Logit	854

(g)–(h) Classification and Thresholds

(g) Compare % correct predictions using a threshold of 0.5.

Model	**Correct (out of 1000)**
LPM	858
Logit	854

(h) Is 0.5 the best threshold for a loan officer?

(g)–(h) Classification and Thresholds

(g) Compare % correct predictions using a threshold of 0.5.

Model	**Correct (out of 1000)**
LPM	858
Logit	854

(h) Is 0.5 the best threshold for a loan officer?

Optimal threshold by accuracy: ≈ 0.49 (442 vs. 440 correct on the test set). But consider **asymmetric costs**:

- **False positive** (deny a good loan): lost profit
- **False negative** (approve a default): large loss

(g)–(h) Classification and Thresholds

(g) Compare % correct predictions using a threshold of 0.5.

Model	Correct (out of 1000)
LPM	858
Logit	854

(h) Is 0.5 the best threshold for a loan officer?

Optimal threshold by accuracy: ≈ 0.49 (442 vs. 440 correct on the test set). But consider **asymmetric costs**:

- **False positive** (deny a good loan): lost profit
- **False negative** (approve a default): large loss

$\Longrightarrow$ A risk-averse loan officer should use a **lower threshold** (e.g., 0.3) to minimize costly defaults.

Outline

Problem 16.23: Setup

Predict self-reported health satisfaction using personal characteristics.

Problem 16.23: Setup

Predict self-reported health satisfaction using personal characteristics.

Outcome: $HSAT3 \in \{1, 2, 3\}$, where $3 =$ highest satisfaction.

Problem 16.23: Setup

Predict self-reported health satisfaction using personal characteristics.

Outcome: $HSAT3 \in \{1, 2, 3\}$, where $3 =$ highest satisfaction.

(a) Distribution:

$HSAT3$	1	2	3
Count	1,777	1,792	914
Percent	39.6%	40.0%	20.4%

Problem 16.23: Setup

Predict self-reported health satisfaction using personal characteristics.

Outcome: $HSAT3 \in \{1, 2, 3\}$, where $3 =$ highest satisfaction.

(a) Distribution:

$HSAT3$	1	2	3
Count	1,777	1,792	914
Percent	39.6%	40.0%	20.4%

Why ordered probit? The outcome has a natural ranking ($1 < 2 < 3$) but the distance between categories is unknown. We cannot just run OLS on 1, 2, 3.

(b) Ordered Probit Estimation

Latent variable: $y_i^* = \beta_1 AGE + \beta_2 AGE^2 + \beta_3 EDUC2 + \beta_4 WORKING + \varepsilon_i$

(b) Ordered Probit Estimation

Latent variable: $y_i^* = \beta_1 AGE + \beta_2 AGE^2 + \beta_3 EDUC2 + \beta_4 WORKING + \varepsilon_i$

	Estimate	**SE**	t
AGE	-0.0557	0.0051	-10.96
AGE^2	0.000408	0.0001	5.62
$EDUC2$	0.0478	0.0073	6.51
$WORKING$	0.0911	0.0382	2.38
$\hat{\mu}_1$ (cutpoint $1\|2$)	-1.253	0.001	
$\hat{\mu}_2$ (cutpoint $2\|3$)	-0.119	0.023	

(b) Ordered Probit Estimation

Latent variable: $y_i^* = \beta_1 AGE + \beta_2 AGE^2 + \beta_3 EDUC2 + \beta_4 WORKING + \varepsilon_i$

	Estimate	**SE**	t
AGE	-0.0557	0.0051	-10.96
AGE^2	0.000408	0.0001	5.62
$EDUC2$	0.0478	0.0073	6.51
$WORKING$	0.0911	0.0382	2.38
$\hat{\mu}_1$ (cutpoint 1|2)	-1.253	0.001	
$\hat{\mu}_2$ (cutpoint 2|3)	-0.119	0.023	

All coefficients significant at 5%. The AGE/AGE^2 pair implies an inverted-U relationship. Cutpoints $\hat{\mu}_1, \hat{\mu}_2$ are estimated along with β.

$N = 4{,}483$. Log-lik $= -4{,}603$. AIC $= 9{,}218$.

Ordered Probit: How It Works

The cutpoints map the latent variable to observed categories:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq \mu_1 \\ 2 & \text{if } \mu_1 < y_i^* \leq \mu_2 \\ 3 & \text{if } y_i^* > \mu_2 \end{cases}$$

Ordered Probit: How It Works

The cutpoints map the latent variable to observed categories:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq \mu_1 \\ 2 & \text{if } \mu_1 < y_i^* \leq \mu_2 \\ 3 & \text{if } y_i^* > \mu_2 \end{cases}$$

Probabilities:

$$P(y = 1) = \Phi(\mu_1 - X'\beta)$$
$$P(y = 2) = \Phi(\mu_2 - X'\beta) - \Phi(\mu_1 - X'\beta)$$
$$P(y = 3) = 1 - \Phi(\mu_2 - X'\beta)$$

Ordered Probit: How It Works

The cutpoints map the latent variable to observed categories:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq \mu_1 \\ 2 & \text{if } \mu_1 < y_i^* \leq \mu_2 \\ 3 & \text{if } y_i^* > \mu_2 \end{cases}$$

Probabilities:

$$P(y=1) = \Phi(\mu_1 - X'\beta)$$
$$P(y=2) = \Phi(\mu_2 - X'\beta) - \Phi(\mu_1 - X'\beta)$$
$$P(y=3) = 1 - \Phi(\mu_2 - X'\beta)$$

Marginal effects (continuous variable):

$$\frac{\partial P(y=1)}{\partial x_k} = -\phi(\mu_1 - X'\beta) \cdot \beta_k$$
$$\frac{\partial P(y=3)}{\partial x_k} = \quad \phi(\mu_2 - X'\beta) \cdot \beta_k$$

(c) Marginal Effects at Age 40

For a 40-year-old with 16 years of education, working:

(c) Marginal Effects at Age 40

For a 40-year-old with 16 years of education, working:

The marginal effect of AGE involves the quadratic: $\beta_1 + 2\beta_2 \cdot AGE$.

(c) Marginal Effects at Age 40

For a 40-year-old with 16 years of education, working:

The marginal effect of AGE involves the quadratic: $\beta_1 + 2\beta_2 \cdot AGE$.

Category	**ME of Age at 40**
$\partial P(HSAT3 = 3)/\partial AGE$	-0.770%
$\partial P(HSAT3 = 2)/\partial AGE$	-0.030%
$\partial P(HSAT3 = 1)/\partial AGE$	$+0.800\%$
Sum	0.000%

(c) Marginal Effects at Age 40

For a 40-year-old with 16 years of education, working:

The marginal effect of AGE involves the quadratic: $\beta_1 + 2\beta_2 \cdot AGE$.

Category	**ME of Age at 40**
$\partial P(HSAT3 = 3)/\partial AGE$	-0.770%
$\partial P(HSAT3 = 2)/\partial AGE$	-0.030%
$\partial P(HSAT3 = 1)/\partial AGE$	$+0.800\%$
Sum	0.000%

$\Longrightarrow$ At 40, an additional year of age **decreases** the probability of high satisfaction. Marginal effects always sum to zero across categories (since $\sum_j P(y = j) = 1$).

(d) Marginal Effects at Age 70: Sign Reversal

Repeat for a 70-year-old (same education and employment):

(d) Marginal Effects at Age 70: Sign Reversal

Repeat for a 70-year-old (same education and employment):

	Age 40	**Age 70**
$\partial P(y = 3)/\partial AGE$	-0.770%	$+0.035\%$
$\partial P(y = 2)/\partial AGE$	-0.030%	$+0.018\%$
$\partial P(y = 1)/\partial AGE$	$+0.800\%$	-0.052%

(d) Marginal Effects at Age 70: Sign Reversal

Repeat for a 70-year-old (same education and employment):

	Age 40	**Age 70**
$\partial P(y = 3)/\partial AGE$	-0.770%	$+0.035\%$
$\partial P(y = 2)/\partial AGE$	-0.030%	$+0.018\%$
$\partial P(y = 1)/\partial AGE$	$+0.800\%$	-0.052%

The signs flip! At 70, aging **increases** the probability of high satisfaction.

(d) Marginal Effects at Age 70: Sign Reversal

Repeat for a 70-year-old (same education and employment):

	Age 40	**Age 70**
$\partial P(y = 3)/\partial AGE$	-0.770%	$+0.035\%$
$\partial P(y = 2)/\partial AGE$	-0.030%	$+0.018\%$
$\partial P(y = 1)/\partial AGE$	$+0.800\%$	-0.052%

The signs flip! At 70, aging **increases** the probability of high satisfaction.

Why? The derivative $\beta_1 + 2\beta_2 \cdot AGE$ changes sign because $\beta_2 < 0$. At the turning point, the marginal effect of age passes through zero. This may reflect survivor bias or adjusted expectations among older respondents.

(e) Marginal Effect of Working

For a 40-year-old with 16 years of education: effect of being employed?

(e) Marginal Effect of Working

For a 40-year-old with 16 years of education: effect of being employed?

For a binary variable, compute the **discrete difference**:

$$\frac{\Delta P(y=j)}{\Delta WORKING} = P(y=j \mid WORKING=1) - P(y=j \mid WORKING=0)$$

(e) Marginal Effect of Working

For a 40-year-old with 16 years of education: effect of being employed?

For a binary variable, compute the **discrete difference**:

$$\frac{\Delta P(y=j)}{\Delta WORKING} = P(y=j \mid WORKING=1) - P(y=j \mid WORKING=0)$$

Category	**Effect of Working**
$\Delta P(HSAT3=3)$	$+2.948\%$
$\Delta P(HSAT3=2)$	$+0.277\%$
$\Delta P(HSAT3=1)$	-3.225%

(e) Marginal Effect of Working

For a 40-year-old with 16 years of education: effect of being employed?

For a binary variable, compute the **discrete difference**:

$$\frac{\Delta P(y=j)}{\Delta WORKING} = P(y=j \mid WORKING=1) - P(y=j \mid WORKING=0)$$

Category	**Effect of Working**
$\Delta P(HSAT3=3)$	$+2.948\%$
$\Delta P(HSAT3=2)$	$+0.277\%$
$\Delta P(HSAT3=1)$	-3.225%

$\implies$ Being employed increases the probability of high health satisfaction by ≈ 3 pp and decreases the probability of low satisfaction by the same amount.

Outline

Problem 16.30: Setup

Model insurance choice using personal characteristics.

Problem 16.30: Setup

Model insurance choice using personal characteristics.

(a) Create *INSURED* from *PUBLIC* and *ADDON*:

PUBLIC	*ADDON*	*INSURED*	**Description**
0	0	1	No public insurance, no add-on (base)
1	0	2	Public insurance only
1	1	3	Public insurance + add-on

Problem 16.30: Setup

Model insurance choice using personal characteristics.

(a) Create *INSURED* from *PUBLIC* and *ADDON*:

PUBLIC	*ADDON*	*INSURED*	**Description**
0	0	1	No public insurance, no add-on (base)
1	0	2	Public insurance only
1	1	3	Public insurance + add-on

Explanatory variables: *AGE*, *FEMALE*, *WORKING*, *HHNINC*2 (household income, DM).

Problem 16.30: Setup

Model insurance choice using personal characteristics.

(a) Create *INSURED* from *PUBLIC* and *ADDON*:

PUBLIC	*ADDON*	*INSURED*	**Description**
0	0	1	No public insurance, no add-on (base)
1	0	2	Public insurance only
1	1	3	Public insurance + add-on

Explanatory variables: *AGE*, *FEMALE*, *WORKING*, *HHNINC2* (household income, DM).

Distribution:

INSURED	1	2	3
Count	572	3,778	133
Percent	12.8%	84.3%	3.0%

Problem 16.30: Setup

Model insurance choice using personal characteristics.

(a) Create *INSURED* from *PUBLIC* and *ADDON*:

PUBLIC	*ADDON*	*INSURED*	**Description**
0	0	1	No public insurance, no add-on (base)
1	0	2	Public insurance only
1	1	3	Public insurance + add-on

Explanatory variables: *AGE*, *FEMALE*, *WORKING*, *HHNINC2* (household income, DM).

Distribution:

INSURED	1	2	3
Count	572	3,778	133
Percent	12.8%	84.3%	3.0%

The multinomial logit model (with *INSURED* = 1 as base):

$$e^{X_i'\beta_j}$$

(b) Multinomial Logit: Full Estimation Results

Coefficients are log-odds **relative to the base** ($INSURED = 1$).

(b) Multinomial Logit: Full Estimation Results

Coefficients are log-odds **relative to the base** ($INSURED = 1$).

	$INSURED = 2$ **(public only)**				$INSURED = 3$ **(public + add-on)**			
	$\hat{\beta}$	SE	z		$\hat{\beta}$	SE	z	
(Intercept)	2.728	0.243	11.23	***	-1.993	0.516	-3.86	***
AGE	0.009	0.004	2.08	*	0.018	0.009	1.97	*
$FEMALE$	0.321	0.101	3.17	**	0.431	0.211	2.04	*
$WORKING$	-0.213	0.120	-1.77	.	-0.377	0.234	-1.61	
$HHNINC2(\times 10^3)$	-0.322	0.026	-12.35	***	-0.040	0.046	-0.88	

(b) Multinomial Logit: Full Estimation Results

Coefficients are log-odds **relative to the base** ($INSURED = 1$).

	$INSURED = 2$ **(public only)**				$INSURED = 3$ **(public + add-on)**			
	$\hat{\beta}$	SE	z		$\hat{\beta}$	SE	z	
(Intercept)	2.728	0.243	11.23	***	−1.993	0.516	−3.86	***
AGE	0.009	0.004	2.08	*	0.018	0.009	1.97	*
FEMALE	0.321	0.101	3.17	**	0.431	0.211	2.04	*
WORKING	−0.213	0.120	−1.77	.	−0.377	0.234	−1.61	
$HHNINC2 (\times 10^3)$	−0.322	0.026	−12.35	***	−0.040	0.046	−0.88	

$N = 4{,}483$. Log-lik $= -2{,}179$. McFadden $R^2 = 0.049$. $HHNINC2$ scaled $\times 10^3$ for readability.

(b) Multinomial Logit: Full Estimation Results

Coefficients are log-odds **relative to the base** ($INSURED = 1$).

	$INSURED = 2$ **(public only)**				$INSURED = 3$ **(public + add-on)**			
	$\hat{\beta}$	SE	z		$\hat{\beta}$	SE	z	
(Intercept)	2.728	0.243	11.23	***	−1.993	0.516	−3.86	***
AGE	0.009	0.004	2.08	*	0.018	0.009	1.97	*
FEMALE	0.321	0.101	3.17	**	0.431	0.211	2.04	*
WORKING	−0.213	0.120	−1.77	.	−0.377	0.234	−1.61	
$HHNINC2(\times 10^3)$	−0.322	0.026	−12.35	***	−0.040	0.046	−0.88	

$N = 4{,}483$. Log-lik $= -2{,}179$. McFadden $R^2 = 0.049$. $HHNINC2$ scaled $\times 10^3$ for readability.

$\Longrightarrow$ Older individuals and women are more likely to have public insurance. Income strongly decreases public-only coverage ($z = -12.35$) but has no significant effect on the add-on decision.

(b) Interpreting Multinomial Logit Coefficients

A positive $\hat{\beta}_{j,k}$ means variable k increases the log-odds of choosing category j over the base ($INSURED = 1$).

(b) Interpreting Multinomial Logit Coefficients

A positive $\hat{\beta}_{j,k}$ means variable k increases the log-odds of choosing category j over the base ($INSURED = 1$).

$INSURED = 2$ **(public only) vs. base:**

- AGE (+, sig): Older $\implies$ more likely on public insurance
- $FEMALE$ (+, sig): Women more likely on public insurance
- $WORKING$ (−, marginal): Working individuals slightly less likely
- $HHNINC2$ (−, highly sig): Higher income $\implies$ much less likely public-only

(b) Interpreting Multinomial Logit Coefficients

A positive $\hat{\beta}_{j,k}$ means variable k increases the log-odds of choosing category j over the base ($INSURED = 1$).

$INSURED = 2$ **(public only) vs. base:**

- AGE ($+$, sig): Older $\implies$ more likely on public insurance
- $FEMALE$ ($+$, sig): Women more likely on public insurance
- $WORKING$ ($-$, marginal): Working individuals slightly less likely
- $HHNINC2$ ($-$, highly sig): Higher income $\implies$ much less likely public-only

$INSURED = 3$ **(public + add-on) vs. base:**

- AGE ($+$, sig), $FEMALE$ ($+$, sig): similar pattern
- $WORKING$ ($-$, not sig), $HHNINC2$ ($-$, not sig): income does **not** significantly affect the add-on decision

(b) Interpreting Multinomial Logit Coefficients

A positive $\hat{\beta}_{j,k}$ means variable k increases the log-odds of choosing category j over the base ($INSURED = 1$).

$INSURED = 2$ **(public only) vs. base:**

- AGE (+, sig): Older $\Longrightarrow$ more likely on public insurance
- $FEMALE$ (+, sig): Women more likely on public insurance
- $WORKING$ (−, marginal): Working individuals slightly less likely
- $HHNINC2$ (−, highly sig): Higher income $\Longrightarrow$ much less likely public-only

$INSURED = 3$ **(public + add-on) vs. base:**

- AGE (+, sig), $FEMALE$ (+, sig): similar pattern
- $WORKING$ (−, not sig), $HHNINC2$ (−, not sig): income does **not** significantly affect the add-on decision

$\Longrightarrow$ The income effect is concentrated in the public-only vs. no-insurance margin, not in the add-on margin.

(c) Predicted Probabilities Match Sample Shares

Obtain predicted probabilities $\hat{P}_1, \hat{P}_2, \hat{P}_3$ for each person. Compare averages to actual shares.

(c) Predicted Probabilities Match Sample Shares

Obtain predicted probabilities $\hat{P}_1, \hat{P}_2, \hat{P}_3$ for each person. Compare averages to actual shares.

	$INSURED = 1$	$INSURED = 2$	$INSURED = 3$
Sample share	12.76%	84.27%	2.97%
$\bar{\hat{P}}_j$	12.76%	84.27%	2.97%

(c) Predicted Probabilities Match Sample Shares

Obtain predicted probabilities $\hat{P}_1, \hat{P}_2, \hat{P}_3$ for each person. Compare averages to actual shares.

	$INSURED = 1$	$INSURED = 2$	$INSURED = 3$
Sample share	12.76%	84.27%	2.97%
$\bar{\hat{P}}_j$	12.76%	84.27%	2.97%

They match **exactly**:

$$\bar{\hat{P}}_j = \frac{1}{N} \sum_{i=1}^{N} \hat{P}_{ij} = \frac{\#(INSURED = j)}{N}$$

(c) Predicted Probabilities Match Sample Shares

Obtain predicted probabilities $\hat{P}_1, \hat{P}_2, \hat{P}_3$ for each person. Compare averages to actual shares.

	$INSURED = 1$	$INSURED = 2$	$INSURED = 3$
Sample share	12.76%	84.27%	2.97%
$\bar{\hat{P}}_j$	12.76%	84.27%	2.97%

They match **exactly**:

$$\bar{\hat{P}}_j = \frac{1}{N}\sum_{i=1}^{N} \hat{P}_{ij} = \frac{\#(INSURED = j)}{N}$$

Why? The MLE first-order conditions for the intercept terms force the predicted probabilities to match the observed counts. This is analogous to $\bar{\hat{y}} = \bar{y}$ in OLS.

(d)–(e) Income and Insurance Choice

Predicted probabilities for a 50-year-old female, working:

(d)–(e) Income and Insurance Choice

Predicted probabilities for a 50-year-old female, working:

	$HHNINC2 = 2{,}400$ (25th percentile)	$HHNINC2 = 4{,}200$ (75th percentile)
$\hat{P}_1$ (no insurance)	7.3%	12.2%
$\hat{P}_2$ (public only)	90.4%	84.2%
$\hat{P}_3$ (public + add-on)	2.3%	3.6%

(d)–(e) Income and Insurance Choice

Predicted probabilities for a 50-year-old female, working:

	$HHNINC2 = 2{,}400$ (25th percentile)	$HHNINC2 = 4{,}200$ (75th percentile)
$\hat{P}_1$ (no insurance)	7.3%	12.2%
$\hat{P}_2$ (public only)	90.4%	84.2%
$\hat{P}_3$ (public + add-on)	2.3%	3.6%

Effect of moving from 25th to 75th percentile of income:

	$\Delta\hat{P}$
P_1 (no insurance)	+4.9 pp
P_2 (public only)	−6.2 pp
P_3 (public + add-on)	+1.3 pp

(d)–(e) Income and Insurance Choice

Predicted probabilities for a 50-year-old female, working:

	$HHNINC2 = 2{,}400$ (25th percentile)	$HHNINC2 = 4{,}200$ (75th percentile)
$\hat{P}_1$ (no insurance)	7.3%	12.2%
$\hat{P}_2$ (public only)	90.4%	84.2%
$\hat{P}_3$ (public + add-on)	2.3%	3.6%

Effect of moving from 25th to 75th percentile of income:

	$\Delta\hat{P}$
P_1 (no insurance)	$+4.9$ pp
P_2 (public only)	-6.2 pp
P_3 (public + add-on)	$+1.3$ pp

$\Longrightarrow$ As income rises, individuals shift away from public-only insurance. The effect on P_2 (-6.2 pp) is much larger than on P_3 ($+1.3$ pp).

IIA: The Limitation of Multinomial Logit

Multinomial logit assumes the **Independence of Irrelevant Alternatives (IIA)**: the ratio $P(j)/P(k)$ is unaffected by other alternatives.

IIA: The Limitation of Multinomial Logit

Multinomial logit assumes the **Independence of Irrelevant Alternatives (IIA)**: the ratio $P(j)/P(k)$ is unaffected by other alternatives.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

IIA: The Limitation of Multinomial Logit

Multinomial logit assumes the **Independence of Irrelevant Alternatives (IIA)**: the ratio $P(j)/P(k)$ is unaffected by other alternatives.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

IIA predicts: $P(\text{car}) = P(\text{red bus}) = P(\text{blue bus}) = 0.33$

IIA: The Limitation of Multinomial Logit

Multinomial logit assumes the **Independence of Irrelevant Alternatives (IIA)**: the ratio $P(j)/P(k)$ is unaffected by other alternatives.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

IIA predicts: $P(\text{car}) = P(\text{red bus}) = P(\text{blue bus}) = 0.33$
Realistically: $P(\text{car}) = 0.5$, $P(\text{red bus}) = P(\text{blue bus}) = 0.25$

IIA: The Limitation of Multinomial Logit

Multinomial logit assumes the **Independence of Irrelevant Alternatives (IIA)**: the ratio $P(j)/P(k)$ is unaffected by other alternatives.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

IIA predicts: $P(\text{car}) = P(\text{red bus}) = P(\text{blue bus}) = 0.33$
Realistically: $P(\text{car}) = 0.5$, $P(\text{red bus}) = P(\text{blue bus}) = 0.25$

$\Longrightarrow$ Test with the Hausman-McFadden test. If IIA fails, consider nested logit or mixed logit.

Outline

1. Problem 16.18: Mortgage Delinquency (Logit)
2. Problem 16.23: Health Satisfaction (Ordered Probit)
3. Problem 16.30: Insurance Choice (Multinomial Logit)
4. Summary

Summary: Three Models, Three Types of y

Problem	**Outcome**	**Model**	**Takeaway**
16.18	Binary (0/1)	LPM vs. Logit	MEs differ at the tails
16.23	Ordered (1/2/3)	Ordered probit	MEs sum to zero; sign can reverse
16.30	Unordered (1/2/3)	Multinomial logit	Heterogeneous effects across categories

Summary: Three Models, Three Types of y

Problem	Outcome	Model	Takeaway
16.18	Binary (0/1)	LPM vs. Logit	MEs differ at the tails
16.23	Ordered (1/2/3)	Ordered probit	MEs sum to zero; sign can reverse
16.30	Unordered (1/2/3)	Multinomial logit	Heterogeneous effects across categories

The common thread across all three:

1. **Raw coefficients are not marginal effects** in non-linear models
2. **Marginal effects vary** with x (unlike OLS)
3. Always report **marginal effects or predicted probabilities**, not just $\hat{\beta}$

Thank you!

jakeanderson@g.ucla.edu