Econ 4220 Assignment 2

Dhillon, Shilling, and Sirmans (“Choosing between Fixed and Adjustable Rate Mortgages,” Journal of Money, Credit and Banking, 19(1), 1987, 260–267) estimate a probit model designed to explain the choice by homebuyers of fixed versus adjustable rate mortgages. They use 78 observations from a bank in Baton Rouge, Louisiana, taken over the period January 1983 to February 1984. These data are contained in the file sirmans.dat.

$ADJUST = A = 1$ 1 if an adjustable mortgage is chosen. The explanatory variables, and their anticipated signs, are:

1. $FIXRATE(+) = F =$ fixed interest rate
2. $MARGIN(-) = -MRN =$ the variable rate - the fixed rate
3. $YIELD(-) = -Y =$ the ten-year Treasury rate less the one-year rate;
4. $MATURITY(-) = -MTY$ ratio of maturities on adjustable to fixed rates
5. $POINTS(-) = -P =$ ratio of points paid on an adjustable mortgage to those paid on a fixed rate mortgage
6. $NETWORTH(+) = NW =$ borrower’s net worth.

(a) Obtain the least squares estimates of the linear probability model explaining the choice of an adjustable mortgage, using the explanatory variables listed above. Obtain the predicted values from this estimation. Are the signs consistent with expectations? Are the predicted values between zero and one?

linear probability model with robust se

The REG Procedure
Model: MODEL1
Dependent Variable: adjust

Number of Observations Read	78
Number of Observations Used	78

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	5.94768	0.99128	5.45	0.0001
Error	71	12.92411	0.18203		
Corrected Total	77	18.87179			

Root MSE	0.42665	R-Square	0.3152
Dependent Mean	0.41026	Adj R-Sq	0.2573
Coeff Var	103.99581		

1. We can see that our F Value is 5.45 and therefore it is statistically significant at all levels ($99.9\% CL(5.45 > 3.291)$)
2. $R^2 = 0.3152$ and $ADJR^2 = 0.2573$. Both are relatively low but not terrible.
3. $RMSE$ is pretty high which indicated not the greatest of fits as we have high errors.
4. The dependent mean is in between 0 and 1 which fits with expectations (0 < A < 1)

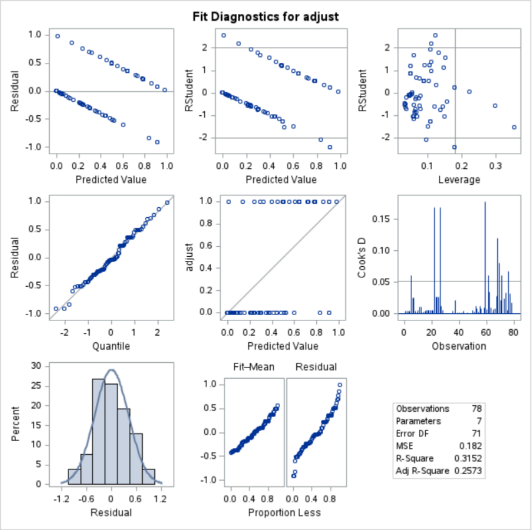
Our least squares estimate for the variables are:

$$A = -0.07077 + 0.16039F - 0.1318MRN - 0.7932Y - 0.08871MTY - 0.03414P + 0.02289NW$$

Overall **all of signs are consistent with expectations**, and all of the **predicted values are between zero and one**.

But, not all of the parameter estimates are statistically significant, nor at the same level:

1. The intercept is **not** statistically significant. We do not reject the null hypothesis $H_0 : \beta_1 = 0$
2. $FIXRATE$ is **very close** to being statistically significant at the $95\% CL(1.95 < 1.96)$ but **is** statistically significant at the $90\% CL(1.95 > 1.65)$
3. $MARGIN$ is statistically significant at the $99\% CL(2.64 > 2.576)$.
4. $YIELD$ is statistically significant at the $95\% CL(2.45 > 1.96)$ but **not** the $99\% CL(2.45 < 2.576)$
5. $MATURITY$ is **not** statistically significant. We do not reject the null hypothesis that $H_0 : \beta_5 = 0$
6. $POINTS$ is also **not** statistically significant, we do not reject the null hypothesis that $H_0 : \beta_6 = 0$
7. $NETWORTH$ is statistically significant at the $95\% CL(2.45 > 1.96)$ but **not** the $99\% CL(2.45 < 2.576)$



Lastly we look at our fit diagnostics:

1. Looking at our Q-Q Plot (row 2 column 1) we see that our distribution is approximately normal with a slight skew towards the lower end. This is verified by our histogram below it where we see a slight left skew between 0 and -0.6.
2. We also have about 7 outliers from looking at both our Cook's D and RStudent plots.

Overall, not a great fit, but not the worst regression out there. (We can do better e.x Probit)

(b) Estimate the model of mortgage choice using probit. Are the signs consistent with expectations? Are the estimated coefficients statistically significant?

probit

The QLIM Procedure

Discrete Response Profile of adjust		
Index	Value	Total Frequency
1	0	46
2	1	32

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	adjust
Number of Observations	78
Log Likelihood	-39.20713
Maximum Absolute Gradient	1.02701E-6
Number of Iterations	21
Optimization Method	Quasi-Newton
AIC	92.41426
Schwarz Criterion	108.91122

Goodness-of-Fit Measures		
Measure	Value	Formula
Likelihood Ratio (R)	27.19	2 * (LogL - LogL0)
Upper Bound of R (U)	105.6	- 2 * LogL0
Aldrich-Nelson	0.2585	R / (R+N)
Cragg-Uhler 1	0.2943	1 - exp(-R/N)
Cragg-Uhler 2	0.3968	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.3317	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.1653	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.2575	R / U
Veall-Zimmermann	0.4494	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.4205	
N = # of observations, K = # of regressors		

Algorithm converged.

Our "Likelihood Ratio (R)" is analogous to the F test in which it tests that all the variable coefficients are zero (except for the intercept). In this case our Likelihood Ratio is 27.19 which is a great indicator of fit.

The other ratios are analogous to the R^2 of our first model and provide varying goodness-of-fit measures.

Our lowest is the Adjusted Estrella - which compared to the regular Estrella accounts for overfitting.
- A significant reduction from our Estrella of 0.3317 to 0.1653 could indicate overfitting especially with our statistically insignificant parameter estimates where we do not reject the null hypothesis $H_0 : \beta = 0$

Results: Probit.sas

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-1.877265	4.120715	-0.46	0.6487
fixrate	1	0.498728	0.262498	1.90	0.0574
margin	1	-0.430951	0.173907	-2.48	0.0132
yield	1	-2.383963	1.083032	-2.20	0.0277
maturity	1	-0.299914	0.241388	-1.24	0.2141
points	1	-0.059185	0.622583	-0.10	0.9243
networth	1	0.083829	0.037853	2.21	0.0268

For our parameter estimates, again all of the signs are consistent with expectations.

As for statistical significance:

1. The intercept is not statistically significant and we do not reject the null hypothesis of $H_0 : \beta_1 = 0$
2. $FIXRATE$ is statistically significant at the $90\% CL (1.90 > 1.65)$
3. $MARGIN$ is statistically significant at the $95\% CL (2.48 > 1.96)$
4. $YIELD$ is statistically significant at the $95\% CL (2.20 > 1.96)$
5. $MATURITY$ is not statistically significant and we do not reject the null hypothesis of $H_0 : \beta_5 = 0$.

6. $POINTS$ is not statistically significant and we do not reject the null hypothesis of $H_0 : \beta_6 = 0$.
7. $NETWORTH$ is statistically significant at the $95\% CL (2.21 > 1.96)$

Overall the MLE probit model seems to be a decent fit, slightly better than our OLS regression. As said with our adjusted Estrella, $MATURITY$ and $POINTS$ could be causing overfitting for the model.

(c) Using the probit estimates from part (b), estimate the probability $\hat{p}$ of choosing an adjustable rate mortgage for each sample observation. What percentage of the outcomes do we successfully predict, using the rule that if $\hat{p} \geq 0.5$, we predict that an adjustable rate mortgage will be chosen?

Our P_Adjust for each entry x_i is equation (16.10) from the textbook:

$$\hat{p}(y_i = 1|x_i) = p(x_i) = P[Z \leq \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}] = \phi(\beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}) \tag{16.10}$$

and if $P\_Adjust \geq 0.5$ then $\hat{p} = 1$, else $\hat{p} = 0$

Thankfully we don't need to calculate it manually for all 78 entries that would be a nightmare, SAS does it for us.

In layman's terms, we look at our last 2 columns and if our second last column is greater than 0.5, we put a 1 in our last column to signify that the person is deciding an adjustable rate mortgage, if not we put a zero.

probit predictions

Obs	adjust	fixrate	margin	yield	maturity	points	networth	Xbeta_adjust	p_adjust	phat
1	1	13.62	1.50	1.38	2.33	1.50	7.558	0.82512	0.79535	1
2	1	13.62	1.50	1.38	2.33	1.50	7.558	0.82512	0.79535	1
3	1	13.62	1.50	1.38	2.33	1.50	7.558	0.82512	0.79535	1
4	1	13.62	1.50	1.38	2.33	1.50	7.556	0.82495	0.79530	1
5	1	14.00	5.50	1.38	1.75	1.00	7.821	-0.48358	0.31434	0
6	1	14.00	4.75	1.38	1.75	1.00	8.014	-0.14419	0.44268	0
7	1	14.00	4.75	1.38	1.75	1.00	8.014	-0.14419	0.44268	0
8	1	13.62	1.50	1.38	2.33	1.50	7.558	0.82512	0.79535	1
9	1	13.50	2.40	1.59	1.00	1.00	17.860	1.16886	0.87877	1
10	1	13.75	2.44	1.45	2.00	0.67	9.100	0.59534	0.72419	1
11	1	14.00	2.45	1.64	1.00	1.00	2.419	-0.01692	0.49325	0
12	1	14.00	2.45	1.64	1.00	1.00	2.419	-0.01692	0.49325	0
13	1	13.50	2.40	1.59	1.00	1.00	17.860	1.16886	0.87877	1
14	1	14.00	0.35	1.64	1.25	0.67	5.620	1.10097	0.86454	1
15	1	13.90	3.04	1.50	2.03	1.00	12.404	0.54082	0.70568	1
16	1	13.75	2.33	1.45	2.50	1.00	7.558	0.34399	0.63457	1
17	1	13.75	2.33	1.45	2.50	1.00	7.558	0.34399	0.63457	1
18	1	13.75	2.33	1.45	2.50	1.00	7.558	0.34399	0.63457	1
19	1	13.75	2.33	1.45	2.50	1.00	7.558	0.34399	0.63457	1
20	1	13.75	2.33	1.45	2.50	1.00	7.558	0.34399	0.63457	1
21	1	13.75	2.33	1.45	2.50	1.00	7.558	0.34399	0.63457	1

Obs	adjust	fixrate	margin	yield	maturity	points	networth	Xbeta_adjust	p_adjust	phat
22	0	13.50	2.40	1.59	1.00	1.00	17.860	1.16886	0.87877	1
23	0	13.88	0.35	2.04	0.83	1.00	4.260	0.07996	0.53187	1
24	0	13.88	0.35	2.04	0.83	1.00	4.260	0.07996	0.53187	1
25	0	13.88	0.35	2.04	0.83	1.00	4.260	0.07996	0.53187	1
26	0	13.50	2.40	1.59	1.00	1.00	17.860	1.16886	0.87877	1
27	0	13.50	3.86	1.60	0.74	0.42	1.977	-0.70331	0.24093	0
28	0	12.38	2.73	1.40	1.66	0.85	1.110	-0.67217	0.25074	0
29	0	12.13	3.36	1.60	1.66	0.85	0.118	-1.62830	0.05173	0
30	0	12.25	3.36	1.60	1.66	0.85	0.885	-1.50416	0.06627	0
31	0	12.38	3.36	1.60	1.66	0.85	0.358	-1.48350	0.06897	0
32	0	12.38	3.36	1.60	1.66	0.85	0.457	-1.47520	0.07008	0
33	0	12.25	3.36	1.60	1.66	0.85	0.573	-1.53031	0.06297	0
34	0	12.40	3.36	1.60	1.66	0.85	0.352	-1.47403	0.07024	0
35	0	12.50	2.10	1.77	0.00	1.00	0.610	-0.77582	0.21893	0
36	0	13.00	3.61	1.69	1.81	1.00	0.733	-1.51901	0.06438	0
37	0	13.25	3.61	1.69	4.34	1.00	13.572	-1.07684	0.14078	0
38	0	12.25	2.60	1.59	2.55	0.93	0.481	-1.45832	0.07238	0
39	0	13.00	2.40	1.59	2.00	1.00	0.170	-0.86334	0.19397	0
40	0	12.50	2.60	1.59	1.27	0.93	0.462	-0.95134	0.17072	0
41	0	12.50	2.60	1.59	2.55	0.93	0.419	-1.33883	0.09031	0
42	0	12.50	2.60	1.59	1.27	0.93	3.198	-0.72199	0.23515	0
43	0	13.00	3.86	1.60	1.48	1.69	3.426	-1.12831	0.12959	0
44	0	12.50	2.60	1.59	2.55	0.93	1.676	-1.23346	0.10870	0
45	0	13.25	3.86	1.60	1.48	1.27	0.066	-1.26043	0.10376	0
46	0	12.50	2.60	1.59	1.09	0.93	0.186	-0.92049	0.17866	0
47	0	12.75	3.86	1.60	1.48	0.85	0.721	-1.43003	0.07635	0
48	0	12.13	3.36	1.60	1.66	0.85	0.369	-1.60726	0.05400	0
49	0	12.75	3.86	1.60	1.48	0.85	0.211	-1.47278	0.07040	0
50	0	12.25	2.73	1.40	1.24	0.85	0.420	-0.66888	0.25179	0
51	0	12.75	2.60	1.59	0.76	0.93	1.000	-0.62860	0.26480	0
52	0	13.25	2.08	1.50	0.97	1.42	0.792	-0.05001	0.48006	0
53	0	13.90	3.04	1.50	2.03	1.00	0.261	-0.47711	0.31664	0
54	0	12.25	2.60	1.59	0.69	0.93	0.745	-0.87835	0.18988	0
55	0	12.75	2.08	1.50	0.49	0.95	0.107	-0.18502	0.42661	0
56	0	13.90	3.04	1.50	2.03	1.00	0.884	-0.42488	0.33546	0
57	0	12.60	3.36	1.60	1.66	0.85	0.598	-1.35366	0.08792	0
58	0	14.00	2.45	1.64	1.00	1.00	0.443	-0.18256	0.42757	0
59	0	13.70	2.08	1.50	0.97	2.38	0.797	0.11802	0.54698	1
60	0	13.80	3.04	1.50	2.03	1.00	0.241	-0.52866	0.29852	0
61	0	13.75	1.04	1.45	0.67	1.00	2.662	1.03834	0.85044	1
62	0	13.62	1.50	1.38	2.33	1.50	1.237	0.29524	0.61609	1
63	0	14.00	2.40	1.59	1.50	1.00	0.322	-0.20192	0.41999	0
64	0	13.00	2.40	1.59	2.00	1.00	0.116	-0.86787	0.19273	0
65	0	13.37	0.35	2.04	1.67	1.00	0.405	-0.74948	0.22678	0
66	0	13.50	0.35	2.04	1.67	1.50	0.268	-0.72572	0.23401	0
67	0	14.00	0.35	2.04	1.67	1.50	3.534	-0.20257	0.41974	0
68	1	11.77	1.90	1.88	0.46	1.13	0.435	-1.47627	0.06994	0
69	1	11.76	1.75	1.74	0.45	1.11	0.314	-1.08882	0.13812	0
70	1	14.00	1.66	1.74	0.50	1.50	0.441	0.03969	0.51583	1

Results: Probit.sas

Obs	adjust	fixrate	margin	yield	maturity	points	networth	Xbeta_adjust	p_adjust	phat
71	1	12.84	0.85	2.03	0.00	1.20	0.360	-0.72019	0.23570	0
72	1	13.75	-0.90	1.45	1.00	1.00	-0.056	1.54756	0.93914	1
73	1	12.50	0.95	1.77	0.67	1.00	0.182	-0.51705	0.30256	0
74	1	12.50	-0.25	1.77	1.00	1.00	0.253	-0.09293	0.46298	0
75	1	13.75	1.04	1.45	0.67	1.00	0.707	0.87445	0.80906	1
76	1	13.75	0.35	2.04	1.67	1.00	0.122	-0.58368	0.27972	0
77	1	14.50	2.10	1.77	0.00	1.00	0.336	0.19866	0.57874	1
78	1	14.00	1.10	1.74	0.00	1.50	0.090	0.40155	0.65599	1
										29

$$\sum_{i=1}^{n} \hat{P} = 29$$

We find that out of 78 observations, only 29 are predicted to choose an adjustable rate mortgage.

Therefore the percentage of outcomes we successfully predict is:

$$\hat{p} = \frac{29}{78} = 0.37179487179 = 37.18\%$$

There is approximately a 37.18% chance that when given a random person deciding on a mortgage that they will decide on an adjustable rate mortgage instead of a fixed one.

We can also verify this using PROC means where we get the following output:

probit predictions

The MEANS Procedure

Analysis Variable : phat				
N	**Mean**	**Std Dev**	**Minimum**	**Maximum**
78	0.3717949	0.4864121	0	1.0000000

It is important to note a relatively large standard deviation of 0.4864121

(d) Estimate the marginal effect of an increase in the variable MARGIN, with all explanatory variables fixed at their sample means. Explain the meaning of this value.

If X_k is continuous (which $MARGIN$ is), then we can calculate the marginal effect by taking the derivative of (16.10):

$$\frac{\partial p(x_i)}{\partial x_{ik}} = \frac{\partial \Phi(t_i)}{\partial t_i} \cdot \frac{\partial_i}{\partial x_{ik}} = \Phi(\beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})\beta_k \tag{16.11}$$

There is also a formula for discrete/indicator variables:

$$\Delta p(x_i) = p(x_i | x_{ki} = 1) - p(x_i | x_{ki} - 0) \tag{16.12a}$$

Again, we have SAS to do everything for us (thankfully)

marginal effect of increase in margin at the mean values

Obs	mef
1	-0.16471

The marginal effect of increase in margin at the mean values is the slope of the probability curve $dp(x_i)/dx$: The change in probability given a unit change in x.

In this case x is $MARGIN$ and the value of its marginal effect is -0.16471.
-For every unit increase in $MARGIN$, the probability of P_ADJUST goes down by 0.16471.
Meaning if the difference between the variable rate and the fixed rate increases by 1 (either the variable rate increases or the fixed rate decreases) then the probability of people buying an adjustable mortgage goes down by 16.471%

> This intuitively makes sense! If variable rate mortgages suddenly become more expensive (higher rate) or fixed rate mortgages become cheaper (lower rate) then people on the margin are incentivized to substitute away from an adjustable rate mortgage!

e) Interpret the outcome of the LR test given in the last set of commands. What is the H_0 that is tested? How was it computed? Do you reject or not?

probit LR test of two coefficients jointly equal to 0

The QLIM Procedure

Discrete Response Profile of adjust		
Index	Value	Total Frequency
1	0	46
2	1	32

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	adjust
Number of Observations	78
Log Likelihood	-39.20713
Maximum Absolute Gradient	1.02701E-6
Number of Iterations	21
Optimization Method	Quasi-Newton
AIC	92.41426
Schwarz Criterion	108.91122

Goodness-of-Fit Measures		
Measure	Value	Formula
Likelihood Ratio (R)	27.19	2 * (LogL - LogL0)
Upper Bound of R (U)	105.6	- 2 * LogL0
N = # of observations, K = # of regressors		

Results: Probit.sas

Goodness-of-Fit Measures		
Measure	**Value**	**Formula**
Aldrich-Nelson	0.2585	R / (R+N)
Cragg-Uhler 1	0.2943	1 - exp(-R/N)
Cragg-Uhler 2	0.3968	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.3317	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.1653	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.2575	R / U
Veall-Zimmermann	0.4494	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.4205	
N = # of observations, K = # of regressors		

Algorithm converged.

Parameter Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	-1.877265	4.120715	-0.46	0.6487
fixrate	1	0.498728	0.262498	1.90	0.0574
margin	1	-0.430951	0.173907	-2.48	0.0132
yield	1	-2.383963	1.083032	-2.20	0.0277
maturity	1	-0.299914	0.241388	-1.24	0.2141
points	1	-0.059185	0.622583	-0.10	0.9243
networth	1	0.083829	0.037853	2.21	0.0268

Test Results				
Test	**Type**	**Statistic**	**Pr > ChiSq**	**Label**
Test0	L.R.	11.85	0.0027	yield = 0 , networth = 0

probit LR test of two coefficients jointly equal to 0

The QLIM Procedure

Our F Test is testing the null hypothesis that both $YIELD$ and $NETWORTH = 0$. Formally:

$$H_0 : \beta_{YIELD} = 0, \beta_{NETWORTH} = 0$$

$$H_A : \beta_{YIELD} \neq 0, \text{or}, \beta_{NETWORTH} \neq 0$$

The test statistic is calculated by:

1. Calculating a restricted model where β_{YIELD} and $\beta_{NETWORTH}$ are set to zero
2. Calculating an unrestricted model where all of the coefficients are estimated normally (we already did this)
3. Applying the formula:

$$LR = 2\big[\ln L_U - \ln L_R\big]$$

We find that our likelihood ratio test gives us an F statistic of 11.85 which is statistically significant across all measures. We reject our null hypothesis H_0 and do not reject the null hypothesis that either or both are not zero.

Conclusion:

Overall across the entire board the models aren't bad but could be better, and other tests could be ran to improve the statistical inference. If I had more time I'd do:

1. An LR testing on $MATURITY$ and $POINTS$ as they aren't statistically significant in either of our models. (OLS or PROBIT). I'd then look at the test statistic, and the goodness-of-fit measures in the restricted model to determine whether $MATURITY$ and $POINTS$ actually cause overfitting, or if there's just too much variance in the small sample size we have to conclude a statistically significant parameter estimate.
2. Running an alternate logit model instead of a probit and comparing the two: Seeing if random errors e_{ik} are statistically independent and identically distributed with an extreme value distribution.
    1. This is more of a personal interest rather than a research question, as a logit model wouldn't be good at all for our estimation at hand. There is no reason for us to assume that random utility errors are statistically independent nor identically distributed.