An advantage of this measure of fit is that it is easy to understand. A disadvantage is that it does not reflect the quality of the prediction: If $Y_i = 1$, the observation is treated as correctly predicted whether the predicted probability is 51% or 90%.

The **pseudo-R^2** measures the fit of the model using the likelihood function. Because the MLE maximizes the likelihood function, adding another regressor to a probit or logit model increases the value of the maximized likelihood, just like adding a regressor necessarily reduces the sum of squared residuals in linear regression by OLS. This suggests measuring the quality of fit of a probit model by comparing values of the maximized likelihood function with all the regressors to the value of the likelihood with none. This is, in fact, what the pseudo-R^2 does. A formula for the pseudo-R^2 is given in Appendix 11.2.

11.4 Application to the Boston HMDA Data

The regressions of the previous two sections indicated that denial rates were higher for black than white applicants, holding constant their payment-to-income ratio. Loan officers, however, legitimately weigh many factors when deciding on a mortgage application, and if any of those other factors differ systematically by race, the estimators considered so far have omitted variable bias.

In this section, we take a closer look at whether there is statistical evidence of discrimination in the Boston HMDA data. Specifically, our objective is to estimate the effect of race on the probability of denial, holding constant those applicant characteristics that a loan officer might legally consider when deciding on a mortgage application.

The most important variables available to loan officers through the mortgage applications in the Boston HMDA data set are listed in Table 11.1; these are the variables we will focus on in our empirical models of loan decisions. The first two variables are direct measures of the financial burden the proposed loan would place on the applicant, measured in terms of his or her income. The first of these is the *P/I ratio*; the second is the ratio of housing-related expenses to income. The next variable is the size of the loan, relative to the assessed value of the home; if the loan-to-value ratio is nearly 1, the bank might have trouble recouping the full amount of the loan if the applicant defaults on the loan and the bank forecloses. The final three financial variables summarize the applicant's credit history. If an applicant has been unreliable paying off debts in the past, the loan officer legitimately might worry about the applicant's ability or desire to make mortgage payments in the future. The three variables measure different types of credit histories, which the loan officer might weigh differently. The first concerns consumer credit, such as credit card debt; the second is previous mortgage payment history; and the third measures credit problems so severe that they appeared in a public legal record, such as filing for bankruptcy.

TABLE 11.1 Variables Included in Regression Models of Mortgage Decisions

Variable	Definition	Sample Average
Financial Variables		
P/I ratio	Ratio of total monthly debt payments to total monthly income	0.331
housing expense-to-income ratio	Ratio of monthly housing expenses to total monthly income	0.255
loan-to-value ratio	Ratio of size of loan to assessed value of property	0.738
consumer credit score	1 if no "slow" payments or delinquencies 2 if one or two slow payments or delinquencies 3 if more than two slow payments 4 if insufficient credit history for determination 5 if delinquent credit history with payments 60 days overdue 6 if delinquent credit history with payments 90 days overdue	2.1
mortgage credit score	1 if no late mortgage payments 2 if no mortgage payment history 3 if one or two late mortgage payments 4 if more than two late mortgage payments	1.7
public bad credit record	1 if any public record of credit problems (bankruptcy, charge-offs, collection actions) 0 otherwise	0.074
Additional Applicant Characteristics		
denied mortgage insurance	1 if applicant applied for mortgage insurance and was denied, 0 otherwise	0.020
self-employed	1 if self-employed, 0 otherwise	0.116
single	1 if applicant reported being single, 0 otherwise	0.393
high school diploma	1 if applicant graduated from high school, 0 otherwise	0.984
unemployment rate	1989 Massachusetts unemployment rate in the applicant's industry	3.8
condominium	1 if unit is a condominium, 0 otherwise	0.288
black	1 if applicant is black, 0 if white	0.142
deny	1 if mortgage application denied, 0 otherwise	0.120

Table 11.1 also lists some other variables relevant to the loan officer's decision. Sometimes the applicant must apply for private mortgage insurance.[3] The loan officer knows whether that application was denied, and that denial would weigh negatively with the loan officer. The next four variables, which concern the applicant's employment status, marital status, and educational attainment, as well as the unemployment rate in the applicant's industry, relate to the prospective ability of the applicant to repay. In the event of foreclosure, characteristics of the property are relevant as well, and the next variable indicates whether the property is a condominium. The final two variables in Table 11.1 are whether the applicant is black or white and

[3]Mortgage insurance is an insurance policy under which the insurance company makes the monthly payment to the bank if the borrower defaults. During the period of this study, if the loan-to-value ratio exceeds 80%, the applicant typically was required to buy mortgage insurance.

whether the application was denied or accepted. In these data, 14.2% of applicants are black, and 12.0% of applications are denied.

Table 11.2 presents regression results based on these variables. The base specifications, reported in columns (1) through (3), include the financial variables in Table 11.1 plus the variables indicating whether private mortgage insurance was denied and whether the applicant is self-employed. In the 1990s, loan officers commonly used thresholds, or cutoff values, for the loan-to-value ratio, so the base specification for that variable uses binary variables for whether the loan-to-value ratio is high (≥ 0.95), medium (between 0.8 and 0.95), or low (< 0.8; this case is omitted to avoid perfect multicollinearity). The regressors in the first three columns are similar to those in the base specification considered by the Federal Reserve Bank of Boston researchers in their original analysis of these data.[4] The regressions in columns (1) through (3) differ only in how the denial probability is modeled, using a linear probability model, a logit model, and a probit model, respectively.

Because the coefficients of the logit and probit models in columns (2)–(6) are not directly interpretable, the table reports standard errors but not confidence intervals. In addition, because the aim of these regressions is to approximate the loan officers' decision rule, it is of interest to know whether individual variables—especially the applicant's race—enter that decision rule. Thus the table reports, through asterisks, whether the test that the coefficient is 0 rejects at the 5% or 1% significance level.

Because the regression in column (1) is a linear probability model, its coefficients are estimated changes in predicted probabilities arising from a unit change in the independent variable. Accordingly, an increase in *P/I ratio* of 0.1 is estimated to increase the probability of denial by 4.5 percentage points (the coefficient on *P/Iratio* in column (1) is 0.449, and $0.449 \times 0.1 \cong 0.045$). Similarly, having a high loan-to-value ratio increases the probability of denial: A loan-to-value ratio exceeding 95% is associated with an 18.9 percentage point increase (the coefficient is 0.189) in the denial probability, relative to the omitted case of a loan-to-value ratio less than 80%, holding the other variables in column (1) constant. Applicants with a poor credit rating also have a more difficult time getting a loan, all else being constant, although interestingly the coefficient on consumer credit is statistically significant but the coefficient on mortgage credit is not. Applicants with a public record of credit problems, such as filing for bankruptcy, have much greater difficulty obtaining a loan: All else equal, a public bad credit record is estimated to increase the probability of denial by 0.197, or 19.7 percentage points. Being denied private mortgage insurance is estimated to be virtually decisive: The estimated coefficient of 0.702 means that being denied mortgage insurance increases your chance of being denied a mortgage by 70.2 percentage points, all else

[4]The difference between the regressors in columns (1) through (3) and those in Munnell et al. (1996), table 2 (1), is that Munnell et al. include additional indicators for the location of the home and the identity of the lender, data that are not publicly available; an indicator for a multifamily home, which is irrelevant here because our subset focuses on single-family homes; and net wealth, which we omit because this variable has a few very large positive and negative values and thus risks making the results sensitive to a few specific outlier observations.

TABLE 11.2 Mortgage Denial Regressions Using the Boston HMDA Data

Dependent variable: *deny* = 1 if mortgage application is denied, = 0 if accepted; 2380 observations.

Regression Model	*LPM*	*Logit*	*Probit*	*Probit*	*Probit*	*Probit*
Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
black	0.084** (0.023)	0.688** (0.182)	0.389** (0.098)	0.371** (0.099)	0.363** (0.100)	0.246 (0.448)
P/I ratio	0.449** (0.114)	4.76** (1.33)	2.44** (0.61)	2.46** (0.60)	2.62** (0.61)	2.57** (0.66)
housing expense-to-income ratio	−0.048 (0.110)	−0.11 (1.29)	−0.18 (0.68)	−0.30 (0.68)	−0.50 (0.70)	−0.54 (0.74)
medium loan-to-value ratio (0.80 ≤ *loan-value ratio* ≤ 0.95)	0.031* (0.013)	0.46** (0.16)	0.21** (0.08)	0.22** (0.08)	0.22** (0.08)	0.22** (0.08)
high loan-to-value ratio (*loan-value ratio* > 0.95)	0.189** (0.050)	1.49** (0.32)	0.79** (0.18)	0.79** (0.18)	0.84** (0.18)	0.79** (0.18)
consumer credit score	0.031** (0.005)	0.29** (0.04)	0.15** (0.02)	0.16** (0.02)	0.34** (0.11)	0.16** (0.02)
mortgage credit score	0.021 (0.011)	0.28* (0.14)	0.15* (0.07)	0.11 (0.08)	0.16 (0.10)	0.11 (0.08)
public bad credit record	0.197** (0.035)	1.23** (0.20)	0.70** (0.12)	0.70** (0.12)	0.72** (0.12)	0.70** (0.12)
denied mortgage insurance	0.702** (0.045)	4.55** (0.57)	2.56** (0.30)	2.59** (0.29)	2.59** (0.30)	2.59** (0.29)
self-employed	0.060** (0.021)	0.67** (0.21)	0.36** (0.11)	0.35** (0.11)	0.34** (0.11)	0.35** (0.11)
single				0.23** (0.08)	0.23** (0.08)	0.23** (0.08)
high school diploma				–0.61** (0.23)	–0.60* (0.24)	–0.62** (0.23)
unemployment rate				0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
condominium					–0.05 (0.09)	
black × *P/I ratio*						–0.58 (1.47)
black × *housing expense-to-income ratio*						1.23 (1.69)
additional credit rating indicator variables	no	no	no	no	yes	no
constant	−0.183** (0.028)	−5.71** (0.48)	−3.04** (0.23)	−2.57** (0.34)	−2.90** (0.39)	−2.54** (0.35)

(continued)

(Table 11.2 continued)

F-Statistics and p-Values Testing Exclusion of Groups of Variables

	(1)	(2)	(3)	(4)	(5)	(6)
applicant single; high school diploma; industry unemployment rate				5.85 (< 0.001)	5.22 (0.001)	5.79 (< 0.001)
additional credit rating indicator variables					1.22 (0.291)	
race interactions and black						4.96 (0.002)
race interactions only						0.27 (0.766)
difference in predicted probability of denial, white vs. black (percentage points)	8.4%	6.0%	7.1%	6.6%	6.3%	6.5%

These regressions were estimated using the $n = 2380$ observations in the Boston HMDA data set described in Appendix 11.1. The linear probability model was estimated by OLS, and probit and logit regressions were estimated by maximum likelihood. Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under the *F*-statistics. The change in predicted probability in the final row was computed for a hypothetical applicant whose values of the regressors, other than race, equal the sample mean. Individual coefficients are statistically significant at the *5% or **1% level.

equal. Of the nine variables (other than race) in the regression, the coefficients on all but two are statistically significant at the 5% level, which is consistent with loan officers' considering many factors when they make their decisions.

The coefficient on *black* in regression (1) is 0.084, indicating that the difference in denial probabilities for black and white applicants is 8.4 percentage points, holding constant the other variables in the regression. This is statistically significant at the 1% significance level $(t = 3.65)$.

The logit and probit estimates reported in columns (2) and (3) yield similar conclusions. In the logit and probit regressions, eight of the nine coefficients on variables other than race are individually statistically significantly different from 0 at the 5% level, and the coefficient on *black* is statistically significant at the 1% level. As discussed in Section 11.2, because these models are nonlinear, specific values of all the regressors must be chosen to compute the difference in predicted probabilities for white applicants and black applicants. A conventional way to make this choice is to consider an "average" applicant who has the sample average values of all the regressors other than race. The final row in Table 11.2 reports this estimated difference in probabilities, evaluated for this average applicant. The estimated racial differentials are similar to each other: 8.4 percentage points for the linear probability model [column (1)], 6.0 percentage points for the logit model [column (2)], and 7.1 percentage points for the probit model [column (3)]. These estimated race effects and the coefficients on *black* are less than in the regressions of the previous sections, in which the only regressors were *P/I ratio* and *black*, indicating that those earlier estimates had omitted variable bias.

The regressions in columns (4) through (6) investigate the sensitivity of the results in column (3) to changes in the regression specification. Column (4) modifies