The results for the second model, shown in Column 3 of Table 3.2, using SAS with the ascending option, simply model the probability that a child has a response in proficiency category 0 or 1, rather than the probability that a child has a response in category 5. Notice that the signs on the intercept and the effect for *gender* are reversed from those in Column 2, yet they are of the same magnitude. Also note that the sum of the probability estimates for boys in Columns 2 and 3 is equal to 1.0, and similarly for girls. SAS with the descending option (Column 2) models the complement of the event from the default approach (ascending, in Column 3). Thus, the probabilities derived from the ascending approach are the complementary probabilities to those found when using SAS with the descending option.

The model parameter estimates using SPSS PLUM look very different from those obtained using the earlier approaches, but in fact the probability estimates are identical to those in Column 3 (and therefore, by the rule of complements, can be used to find the probability estimates in Column 2). SPSS PLUM is a program specifically designed for analyzing ordinal response variables, and the resulting parameter estimates will not exactly correspond to those found under SPSS Logistic Regression. In particular, the probability being estimated in SPSS PLUM is the probability of a response being *at or below* a particular outcome value, that is, the lower category codes; in contrast, SPSS LOGISTIC models the probability of the category with the higher outcome value. Additionally, whereas SAS handles both dichotomous and ordinal responses through its LOGISTIC procedure, the SPSS PLUM procedure uses a slightly different formulation of the generalized linear model that looks like: $\ln(Y_j') = \theta_j - \beta_1 X_1$. In this expression, the subscript j refers to the response category, and X_1 refers to the single independent variable, *gender*. The estimate for the effect of *gender* is *subtracted* from the intercept. Another important distinction between PLUM results and those from logistic regression programs under SPSS or SAS is that PLUM internally sets up the coding for categorical predictors. In Column 4, the estimate provided for the *gender* effect corresponds to when *gender* = 0, that is, for females. The coding system used is clearly displayed on the printout (examples of PLUM and SAS printouts for ordinal models will be included in the next chapters). To find the estimated probability for a girl being (at most) in proficiency categories 0 or 1—that is, $P(Y \leq 0)$—which is equivalent in this case to $P(Y = 0)$ because there are no responses less than 0, we use the estimates to find the predicted logit for girls (.335 – .765 = –.43), exponentiate the result to find the odds for girls of being at (or below) $Y = 0$ (exp(–.43) = .65), and then solve for the estimated probability (.65/(1 + .65) = .394). The same process is used to find the estimated probability for boys of being at (or below) categories 0 or 1, or $P(Y = 0)$.

SPSS PLUM provides R^2_L, referred to as *McFadden's pseudo* R^2 (Long, 1997; Menard, 2000), in addition to R^2_{CS} and R^2_N. In order to obtain the necessary values for the –2LL deviance statistics, the "*kernel*" option must be specified in the SPSS PLUM "*/print*" statement, as shown in syntax A4 in Appendix A.

The previous discussion and simple comparison of how SAS and SPSS treat binary outcomes illustrate that although model parameter estimates may vary on the surface, the resulting predicted probabilities computed from the model estimates, as well as model fit statistics, are consistent across packages and approaches. These simple examples also illustrate that it is important for an analyst to be aware of the outcome being predicted as well as how categorical independent variables are incorporated into the models, once a statistical package is selected. Distinctions across approaches and packages become even more critical as the number of categories for an ordinal response variable increases beyond the binary case.

4. THE CUMULATIVE (PROPORTIONAL) ODDS MODEL FOR ORDINAL OUTCOMES

Overview of the Cumulative Odds Model

With only two categories for an outcome variable, logistic regression is used to model the likelihood of one of the outcomes, usually termed the "success," as a function of a set of independent variables. The estimated probabilities for the response of interest, *P*(success), as well as for its complement, 1 – *P*(success), can be determined using the prediction model for the logits, as shown in the example in Chapter 3. When the possible responses for an outcome variable consist of more than two categories and are ordinal in nature, the notion of "success" can be conceived of in many different ways. Regression models for ordinal response variables are designed for just this situation and are extensions of the logistic regression model for dichotomous data. The complexity in fitting ordinal regression models arises in part because there are so many different possibilities for how "success," and the consequent probability of "success," might be modeled.

For example, given a K-level ordinal response variable, such as proficiency in early literacy with $K = 6$ as in the ECLS-K study (Table 2.1), we could derive several different representations of "success" depending on how we view the data. In general, K-level ordinal data can be partitioned by $K - 1$ "success" cutpoints (Fox, 1997; McCullagh & Nelder, 1983). Success is, of course, a relative term; generally, it designates an event of interest. For example, "success" might be defined as having a child score in category 0 on the mastery test, that is, those children who were not able to recognize upper- and/or lowercase letters. Under this partitioning of the data, our interest would be in identifying factors associated with increased likelihood of being in this lowest category, rather than being beyond category 0, in categories 1 through 5. Perhaps there are harmful child, family, or school characteristics associated with increased probability of being in this lowest category. For these explanatory variables, we would calculate the odds of being at (or below) category 0.

We could next conceive of "success" as being at or below category 1; our interest in this partitioning of the data would be in identifying factors associated with greater likelihood of being in categories 0 or 1 relative to the likelihood of being beyond the lowest stages, in categories 2 through 5. We could continue to describe the data in this cumulative fashion, with the final conceptualization of "success" as being at or below the Kth category, which of course will always occur. Hence, the last split or partitioning of the data becomes unnecessary. Using this cumulative progression, we would have $K - 1$, or 5, distinct possible "success" characterizations of the data, given $K = 6$ ordinal response categories.

The analysis that mimics this method of dichotomizing the outcome, in which the successive dichotomizations form cumulative "splits" to the data, is referred to as proportional or cumulative odds (CO) (Agresti, 1996; Armstrong & Sloan, 1989; Long, 1997; McCullagh, 1980; McCullagh & Nelder, 1983). It is one way to conceptualize how the data might be sequentially partitioned into dichotomous groups, while still taking advantage of the order of the response categories. The ordinal nature of this approach is so appealing because of its similarity to logistic regression. If a single model could be used to estimate the odds of being *at or below* a given category across all cumulative splits, that model would offer far greater parsimony over the fitting of $K - 1$ different logistic regression models corresponding to the sequential partitioning of the data, as described above. The goal of the cumulative odds model is to simultaneously consider the effects of a set of independent variables across these possible consecutive cumulative splits to the data. There are other approaches, however, to defining "success." Each different method for performing ordinal regression characterizes the partitioning of the data in a very distinct way, and therefore they address very different research questions. The conceptualizations of how the data may be split to correspond to the cumulative odds (CO) model, as well as for the two other methods to fitting ordinal regression models that I will discuss in this book, the continuation ratio (CR) model and the adjacent categories (AC) model, are provided in the indicated columns of Table 4.1. The latter two approaches will be discussed fully in later chapters. This chapter focuses on the CO model.

A simplifying assumption is made of the data when applying ordinal regression models, and that is the assumption of proportional, or parallel, odds. This assumption implies that the explanatory variables have the same effect on the odds, regardless of the different consecutive splits to the data, for each category of model (CO, CR, AC), as shown in Table 4.1. For example, if the set of separate binary logistic regressions corresponding to the CO model described above were fit to the data, the assumption of parallelism implies that a common odds ratio (or effect) for a variable would be observed across all the regressions; the effect of an IV on the odds is assumed to be invariant across the corresponding splits (Agresti, 1989; Brant, 1990; Menard, 1995; Peterson & Harrell, 1990). Thus, one model would be sufficient to describe the relationship between the ordinal response variable and a set of predictors.

Both SAS and SPSS provide a score test for the proportional odds assumption within their ordinal regression procedures, but this omnibus test for proportionality is not a powerful test and is anticonservative (Peterson & Harrell, 1990); the test nearly always results in very small p values, particularly when the number of explanatory variables is large (Brant, 1990), the sample size is large (Allison, 1999; Clogg & Shihadeh, 1994), or continuous explanatory variables are included in the model (Allison, 1999). Therefore, conclusions about rejecting the null hypothesis of proportionality of the odds based solely on the score test should be made cautiously. Rejection of the assumption of parallelism (proportional odds) for the particular ordinal model being investigated implies that at least one of the explanatory variables may be having a differential effect across the outcome levels, that is, that there is an interaction between one or more of the independent variables and the derived splits to the data (Armstrong & Sloan, 1989; Peterson & Harrell, 1990). The key is to be able to identify which variable(s) may be contributing to rejection of this overall test.

A reasonable strategy for investigating whether the effects of the independent variables are relatively stable or not across the cumulative logits is through comparison of variable effects across the separate logistic regression models that correspond to the ordinal model being considered, as in Table 4.1. Although the simplifying assumption of proportionality may be useful in terms of fitting an overall model to the data, it has been recommended that researchers examine the underlying binary models in order to supplement decisions about the aptness of an ordinal approach (Brant, 1990; Clogg & Shihadeh, 1994; Long, 1997; O'Connell, 2000). Informal comparison of the slopes across the corresponding separate logistic fits for a model can provide supportive information regarding the plausibility of parallelism for the data. Later in this chapter, an approach that relaxes the

TABLE 4.1
Category Comparisons Associated With
Three Different Ordinal Regression Model Approaches,
Based on a 6-Level Ordinal Outcome (j = 0, 1, 2, 3, 4, 5)

Cumulative Odds (ascending) P(Y ≤ j)	*Cumulative Odds (descending)* P(Y ≥ j)	*Continuation Ratio* P(Y > j|Y ≥ j)	*Adjacent Categories* P(Y = j + *1*|Y = j *or* Y = j + *1*)
Category 0 versus all above	Category 5 versus all below	Categories 1 through 5 versus category 0	Category 1 versus category 0
Categories 0 and 1 combined versus all above	Categories 5 and 4 versus all below	Categories 2 through 5 versus category 1	Category 2 versus category 1
Categories 0, 1, and 2 combined versus all above	Categories 5, 4, and 3 versus all below	Categories 3 through 5 versus category 2	Category 3 versus category 2
Categories 0, 1, 2, and 3 combined versus all above	Categories 5, 4, 3, and 2 versus all below	Categories 4 and 5 versus category 3	Category 4 versus category 3
Categories 0, 1, 2, 3 and 4 combined versus category 5	Categories 5, 4, 3, 2, and 1 versus category 0	Category 5 versus category 4	Category 5 versus category 4

proportional odds assumption for some explanatory variables, the *partial proportional odds* (PPO) model (Ananth & Kleinbaum, 1997; Koch, Amara, & Singer, 1985; Peterson & Harrell, 1990), is presented.

Example 4.1: Cumulative Odds Model With a Single Explanatory Variable

To illustrate the use of the cumulative odds model, I begin by fitting a simple model with just one categorical explanatory variable: *gender*. Table 4.2 provides the frequency of each of the five early-reading proficiency categories for boys and girls. The data are unbalanced across proficiency categories, with most children, regardless of *gender*, falling into proficiency category 3. This characteristic of the data can be an important consideration when deciding among models (CO, CR, AC, or others) that might best represent the data; however, for pedagogical purposes we will ignore this characteristic of the data for now, then reexamine its impact after the different ordinal models have been presented.

The cumulative odds model is used to predict the odds of being *at or below* a particular category. Because there are K possible ordinal outcomes, the model actually makes $K - 1$ predictions, each corresponding to the accumulation of probability across successive categories. If we let $\pi(Y \leq j|x_1, x_2, \ldots x_p) = \pi_j(\underline{x})$ represent the probability that a response falls in a category less than or equal to the jth category ($j = 1, 2, \ldots K - 1$), then we have a collection of cumulative probabilities for each case. The final category will always have a cumulative probability of 1.0. (Note that in the ECLS-K data, I use category 0 to refer to the first category, and the K = 6th category is proficiency category 5.) With an extension from the general logistic regression model, the predictions are logits for the *cumulative* probabilities, which are referred to as *cumulative logits*:

$$\ln(Y_j') = \ln\left(\frac{\pi_j(\underline{x})}{1 - \pi_j(\underline{x})}\right) = \alpha_j + (\beta_1 X_1 + \beta_2 X_2 + \ldots \beta_p X_p).$$

The cumulative logits associated with being at or below a particular category j can be exponentiated to arrive at the estimated cumulative odds and then used to find the estimated cumulative probabilities associated with being at or below category j.

Table 4.2 also contains the cross-tabulation of the ECLS-K data in terms of actual probabilities (p), cumulative probabilities (cp), and cumulative odds (co) for boys and girls of being in category j or below. The **bold** row contains the associated odds ratios (boys:girls) for these data. The last two rows of the table provide the category totals and the cumulative proportion ($P(Y_i \leq$ category $j)$) regardless of *gender*. From the table, we see that the odds of being at or below any specific category increases as the response value increases, for both boys and girls. This makes intuitive sense, as within the sample there are fewer children who are in the highest categories; children are more likely to be at or below a given category than beyond that category. In general, the odds for boys are always greater than the odds for girls, as proportionately fewer boys than girls in the sample reached the higher proficiency categories when tested at the beginning of first grade. The odds ratios make this pattern clear. The odds that boys are *at or below* a specific category are about 1.72 (on average) times the odds for girls of being *at or below* that category. The likelihood is that girls tend to exceed boys on this ordinal measure of proficiency at the beginning of first grade.

Similar to the example in Chapter 3, I am going to present results for this simple one-variable CO model using three different approaches: SAS PROC LOGISTIC, SAS PROC LOGISTIC with a "descending" option, and SPSS PLUM (syntax for all models is provided in the

TABLE 4.2

Observed Data Cross-Classification of Gender by Five Proficiency Categories: Frequency (*f*), Proportion (*p*), Cumulative Proportion (*cp*), Cumulative Odds[a] (*co*), and Odds Ratios (OR)

Category	*0*	*1*	*2*	*3*	*4*	*5*	*Totals* (f)
Males							
f	48	163	320	735	256	151	1673
p	.0278	.0974	.1913	.4393	.1530	.0903	1.000
cp	.0278	.1261	.3174	.7567	.9097	1.000	—
co	.0295	.1443	.4650	3.110	10.074	—	—
Females							
f	19	115	274	747	331	206	1692
p	.0112	.0680	.1619	.4415	.1956	.1217	1.000
cp	.0112	.0792	.2411	.6826	.8782	.9999	—
co	.0113	.0860	.3177	2.1506	7.210	—	—
OR	**2.6106**	**1.6779**	**1.4636**	**1.446**	**1.3972**	—	—
Totals (*f*)	67	278	594	1482	587	357	3,365
cp_{total}	.0199	.1025	.2790	.7195	.8939	1.000	—

a. Cumulative odds = Odds($Y_i \leq$ category *j*).

Appendix, section B). Figure 4.1 displays the SAS output (with the default "ascending" approach) for this simple one-variable cumulative odds model. The appropriate link function for the cumulative odds model is the logit link. To run this model, I used the SAS syntax in section B1 of the appendix. The syntax for the other two approaches to the CO model is in sections B2 and B3. Although these approaches are essentially identical in terms of prediction when the CO model is desired, such is not necessarily the case with the CR and AC ordinal regression models. It is important to be clear on the similarities and differences among programs and approaches, beginning with the simplest case of the CO model.

Using SAS (ascending), the odds are accumulated over the lower-ordered categories. That is, the associated predicted cumulative probabilities correspond to the pattern shown in the first column of Table 4.1. SAS is estimating the $P(Y \leq$ category $j)$, which for these data are $P(Y \leq 0)$, $P(Y \leq 1)$, $P(Y \leq 2)$, $P(Y \leq 3)$, $P(Y \leq 4)$, and of course $P(Y \leq 5) = 1.0$ for the final category (which typically is not included on printouts of these analyses). A reliable CO model would reproduce the cumulative odds and cumulative probabilities found from the data in Table 4.2.

In the models presented here, *gender* is coded as 0 for girls and 1 for boys. Reviewing the output provided in Figure 4.1, we see that the proportional odds assumption is upheld for these data ("Score Test for the


```
The LOGISTIC Procedure

                    Model Information

     Data Set                          WORK.GONOMISS
     Response Variable                 PROFREAD
     Number of Response Levels         6
     Number of Observations            3365
     Model                             cumulative logit
     Optimization Technique            Fisher's scoring

                    Response Profile

        Ordered                          Total
          Value        PROFREAD      Frequency
              1            0.00             67
              2            1.00            278
              3            2.00            594
              4            3.00           1482
              5            4.00            587
              6            5.00            357

           Probabilities modeled are cumulated
            over the lower Ordered Values.

                Model Convergence Status

       Convergence criterion (GCONV=1E-8) satisfied.

      Score Test for the Proportional Odds Assumption

         Chi-Square          DF        Pr > ChiSq
             5.3956           4            0.2491

                 Model Fit Statistics

                            Intercept          Intercept
      Criterion                  Only     and Covariates
      AIC                   10063.980          10028.591
      SC                    10094.586          10065.319
      -2 Log L              10053.980          10016.591

                 The LOGISTIC Procedure

 R-Square      0.0110       Max-rescaled R-Square      0.0116

           Testing Global Null Hypothesis: BETA=0

 Test                    Chi-Square       DF     Pr > ChiSq

 Likelihood Ratio           37.3884        1         <.0001
 Score                      37.2553        1         <.0001
 Wald                       37.2060        1         <.0001
```

Figure 4.1 SAS Cumulative Odds Model Example: *Gender*

Figure 4.1 (Continued)

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 0.00	1	-4.1049	0.1284	1022.2632	<.0001
Intercept 1.00	1	-2.3739	0.0667	1266.5201	<.0001
Intercept 2.00	1	-1.1474	0.0510	505.4293	<.0001
Intercept 3.00	1	0.7590	0.0485	245.3247	<.0001
Intercept 4.00	1	1.9545	0.0627	971.9783	<.0001
GENDER	1	0.3859	0.0633	37.2060	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence	Limits
GENDER	1.471	1.299	1.665

Association of Predicted Probabilities and Observed Responses

Percent Concordant	29.0	Somers' D	0.079
Percent Discordant	21.1	Gamma	0.159
Percent Tied	49.9	Tau-a	0.058
Pairs	4110137	c	0.540

Proportional Odds Assumption"), $\chi^2_4 = 5.3956$, $p = .2491$. We can conclude that the effect of *gender* is not statistically different across the five cumulative splits for the data; this implies that if five separate binary logistic models were fit corresponding to the pattern in Table 4.1, the slopes (and odds ratios) for *gender* in each of these models would be similar. Thus, the *gender* ORs could be estimated simultaneously using only one model. Because *gender* is the only variable included here, this result also tells us that the five ORs in Table 4.2 are not statistically different, and that one common OR could be used to summarize the effect of *gender* on proficiency.

The pseudo R^2 statistics are found in the "Model Fit Statistics" section of the printout (Figure 4.1), in the line under "The LOGISTIC Procedure," with the Cox and Snell $R^2_{CS} = .0110$ and the Nagelkerke (which SAS refers to as Max-rescaled R-Square) $R^2_N = .0116$. The likelihood ratio $R^2_L = .0037$ can be calculated using the –2loglikelihood statistics for the intercepts-only model and the intercepts plus covariates model information contained in the "Model Fit Statistics" summary table. Collectively, these R^2 statistics suggest that the relationship between the response and predictor variables is a

weak one. However, the tests for overall model fit ("Testing Global Null Hypothesis"), which assess whether the fitted model improves predictions over those presented by the null (intercepts-only) model, are all statistically significant, so we reject the null model in favor of the model that includes *gender* as a predictor. Despite the low pseudo R^2 values, the likelihood ratio test suggests that the pattern of cumulative proportions for boys and girls as predicted from the model (see Table 4.3; entries explained later) provides a better match to the actual cumulative proportions for boys and girls (shown in Table 4.2) than what would be expected disregarding *gender* (last row of Table 4.2). This simple CO model makes clear how these proportions are different for boys versus girls.

The next section of the printout (Figure 4.1) contains "Analysis of Maximum Likelihood Estimates," a table with five intercepts, referred to as threshold parameters: one for each of the $K - 1$ cutpoints. It is useful to think of these thresholds as marking the point (in terms of a logit) at which children might be predicted into the higher categories, but they are not usually interpreted individually, similar to how the intercept functions in an ordinary multiple regression model. However, with dummy coding for *gender* (*gender* = 0 for girls), these threshold estimates represent the predicted logits corresponding to $Y \le$ category j for girls. The effect of *gender* on the logit is .3859, with an associated odds ratio of 1.471 (exp(.3859) = 1.471). The model informs us that the odds for boys of being at or below category j are about 1.471 times the odds for girls, regardless of which cumulative split we are considering. This result can be compared with the pattern we saw using the observed data in Table 4.2, where the average OR across

TABLE 4.3

Predicted Cumulative Logits, Estimated Odds of Being at or Below Category *j* for Boys and Girls, Estimated Cumulative Probabilities (*cp*), and Estimated Odds Ratios From the CO Model (SAS With Ascending Option)

Comparison	*(Y ≤ 0)*	*(Y ≤ 1)*	*(Y ≤ 2)*	*(Y ≤ 3)*	*(Y ≤ 4)*
Boys					
Cumulative logit	−3.719	−1.988	−.7615	1.1449	2.3404
Cumulative odds	.02427	.13696	.4670	3.1421	10.385
$\hat{cp}_b$	.0237	.1205	.3183	.7586	.9122
Girls					
Cumulative logit	−4.1049	−2.3739	−1.1474	.7590	1.9545
Cumulative odds	.0165	.0931	.3175	2.1363	7.0604
$\hat{cp}_g$	.0162	.0852	.2410	.6811	.8760
OR	**1.4711**	**1.4711**	**1.4709**	**1.4708**	**1.4709**

categories was 1.72. According to the model, boys are less likely to be *beyond* a particular category relative to girls, which is consistent with the actual data. Recall that this model assumes that the effect of *gender* is constant across the separate cumulative splits. Because we did not reject the assumption of proportional odds when *gender* was included as a predictor, the CO model suggests that the separate ORs for the cumulative splits (Table 4.2) are not statistically different from the OR of 1.471 found for the CO model.

Turning to a direct interpretation of the parameter estimates for the model, the intercepts and the effect of *gender* can be used to estimate the cumulative odds, that is, the odds of being at or below a given category for boys and for girls. These also can be used to estimate the ORs at each split, although we already know from our analysis that this is set at 1.471. The cumulative odds estimated for boys and girls can be compared back to those derived from the original data (Table 4.2). Predictions for girls, when *gender* = 0, correspond to the intercepts for each cumulative category, which when exponentiated provide the odds for girls of having a response at or below category *j*. Predictions for boys are found by substituting the value of *gender* = 1 into the cumulative odds model for each respective equation and exponentiating to find the odds: $\ln(Y_j') = \alpha_j + .3859(gender)$. For example, for the logit representing $Y \leq 0$, the predicted logit for girls is –4.1049; for boys, the predicted logit is –3.719. Table 4.3 provides these *estimated cumulative logits* based on the model as well as the *estimated cumulative odds* (*co*) for boys and girls (exp(cum. logit)). From these predicted cumulative odds, odds ratios comparing boys to girls can be found easily for each category, and these are shown in the last row of Table 4.3 (e.g., co_{boys}/co_{girls}). Within rounding error, the ORs are all approximately 1.47. The estimated cumulative odds are transformed into the estimated cumulative probabilities (*cp*) using $cp = (co/[1 + co])$, which yields $P(Y \leq$ category $j)$. The results are shown in Table 4.3 and can be compared with the observed cumulative probabilities presented in Table 4.2. Overall, the estimates seem to match the data well; recall that the likelihood ratio test was statistically significant for this model.

The model predictions make it clear what the assumption of proportional odds means for these data. The OR is fixed, and therefore remains constant across all cumulative categories, implying that overall, the odds for boys of being at or below any category *j* are about 1.47 times the odds for girls of being at or below category *j*. For this sample, boys are more likely than girls to be *at or below* any given category; girls are more likely than boys to be in *higher* categories. *Gender* (male = 1) has a positive effect ($b = .3859$) on the cumulative logit, corresponding to larger odds of being *at or below* category *j* for boys relative to girls. This last interpretation is consistent with the transformed outcome being modeled in this approach (response *at or below* category *j*), and hence the interpretation of the direction of the logit and the effects of explanatory variables hinge on how the outcome is characterized.

Differences between estimated and actual cumulative probabilities are due to the fact that the CO model is imposing a very specific structure on the data. This structure is evidenced through the behavior of the ORs, and thus it affects the cumulative proportions estimated from the model as well. The estimates for the cumulative probabilities are derived under the assumption of proportional odds. Although we saw earlier that this assumption is valid for our data across *gender*, it is important to recognize that the model estimates and the predicted probabilities are driven by this assumption. In situations where the assumption does not hold or seems empirically or theoretically implausible, these predicted probabilities could be grossly inaccurate. Unfortunately, when models become more complex, such as those that include additional explanatory variables, either categorical or continuous, it can become quite challenging to have confidence in the assumption of proportional odds. I will return to this topic toward the end of this chapter.

Somers' D for this analysis was .079 (see last section of Figure 4.1), which is quite low. With only one predictor, we are getting very weak concordance for the ordinal direction of predicted probabilities among pairs of children. To construct the classification table for the measures of predictive efficiency, τ_p and λ_p, we can use the collection of cumulative predicted probabilities for each child to assess where individual probability of category membership is at its maximum. With the "*predprobs = cumulative*" option specified in the output subcommand, SAS creates a data file containing the cumulative probabilities for the *i*th child at each level, $cp0_i = P$(at or below level 0), $cp1_i = P$(at or below level 1), and so on. Thus, each child has *K* new observations in the data set, with $cp5_i = P$(at or below level 5) = 1.0 for all children. Category probabilities can be found by using the relationship $P(Y =$ category $j) = P(Y \leq$ category $j) - P(Y \leq$ category $[j - 1])$. That is, $P_i(Y = 0) = cp0_i$; $P_i(Y = 1) = cp1_i - cp0_i$; $P_i(Y = 2) = cp2_i - cp1_i$; and so on. The maximum category probability for an individual child corresponds to his or her best prediction for proficiency level. In the *gender*-only, ascending model, all children are predicted into category 3, which is not surprising given that the model has weak fit and category 3 represents 44% of the children in this sample. Following the methods outlined in Chapter 3, the classification estimates can be tabled to the observed proficiency categories to calculate measures of predictive efficiency. For this analysis, τ_p and λ_p are .23 and 0, respectively. These results underscore the need to consider several different measures of association in conjunction with the likelihood ratio tests when assessing the reasonableness of a model.

Table 4.4 provides a comparison of the results for the SAS model just described, based on the "ascending" default option in the ordering of the ordinal dependent variable, with the results from SAS "descending," SPSS PLUM, and a multiple regression model with *gender* as the only predictor. Although the CO models are essentially the same and provide the same *interpretation* of the effect of *gender*, some important similarities and differences in the *presentation* of the results for these models should be pointed out.

First, similar to the results using the ascending and descending options in SAS PROC LOGISTIC with a dichotomous outcome, the estimates for the threshold (intercept) parameters are reversed in sign but not in magnitude; they also appear in reverse order on the printout. This is simply due to the fact that the descending and ascending options predict complementary events. With the descending option in place, the model is estimating the (reversed) cumulative odds, that is, $P(Y \geq 5)$, $P(Y \geq 4)$, $P(Y \geq 3)$, $P(Y \geq 2)$, and $P(Y \geq 1)$, and of course $P(Y \geq 0)$ will always equal 1.0.

TABLE 4.4
Results for Cumulative Odds Model Using SAS (Ascending), SAS (Descending), SPSS PLUM, and Multiple Linear Regression on an Ordinal Response Scale: Proficiency ($j = 0, 1, 2, 3, 4, 5$) by Gender, $N = 3{,}365$

	SAS (ascending)	*SAS (descending)*	*SPSS PLUM*	*SPSS REGRESSION*
Model estimates	$P(Y \leq \text{cat. } j)$	$P(Y \geq \text{cat. } j)$	$P(Y \leq \text{cat. } j)$	$E(Y\|X)$
Intercept α				3.108
Thresholds	$\alpha_0 = -4.105$	$\alpha_5 = -1.955$	$\theta_0 = -3.719$	
	$\alpha_1 = -2.374$	$\alpha_4 = -0.759$	$\theta_1 = -1.988$	
	$\alpha_2 = -1.147$	$\alpha_3 = 1.147$	$\theta_2 = -.762$	
	$\alpha_3 = 0.759$	$\alpha_2 = -2.374$	$\theta_3 = 1.145$	
	$\alpha_4 = 1.955$	$\alpha_1 = 4.105$	$\theta_4 = 2.340$	
gender = 1 (male)	.386**	–.386**	0	–.246**
gender = 0 (female)			.386**	
R^2	.004[a]	.004[a]	.004[a]	.012
Score test[b]	$\chi^2_4 = 5.3956$ ($p = .2491$)	$\chi^2_4 = 5.3956$ ($p = .2491$)	$\chi^2_4 = 5.590$ ($p = .232$)	
Model fit[c]	$\chi^2_1 = 37.388$ ($p < .001$)	$\chi^2_1 = 37.388$ ($p < .001$)	$\chi^2_1 = 37.388$ ($p < .001$)	$F_{1,3363} = 40.151$ ($p < .001$)

a. R^2_L = likelihood ratio R^2.
b. For the proportional odds assumption.
c. Likelihood ratio test for ordinal models; F test for ordinary least squares (OLS) regression.
**$p < .01$.

Second, the score test for the proportional odds assumption indicates that the assumption of proportionality is upheld across the analyses, $\chi^2_4 = 5.3956$, $p > .05$, as would be expected, although SPSS refers to this as the "Test of Parallel Lines."[9] For all three models, the omnibus likelihood ratio tests indicate that the ordinal gender model fits better than the null, $\chi^2_1 = 37.388$, $p < .001$.

Third, predictions of the cumulative odds and cumulative proportions using SAS ascending and SPSS PLUM are exactly the same; and the predictions for the cumulative odds for SAS descending yield the complements of these probabilities. Recall that for SPSS PLUM, the model predictions are found by subtracting the effect of *gender* from the threshold estimates. SPSS PLUM also uses an internal coding system for the categorical predictors. For example, to estimate the cumulative probability for a girl having a proficiency response less than or equal to 2 using the PLUM model, we would (a) find $\ln(\text{odds}(Y \leq 2) = \theta_2 - \beta_{(gender = 0)} = -.762 - (.386) = -1.148$; (b) exponentiate to find the odds, $\exp(-1.148) = .3173$; and (c) use these odds to find the cumulative probability for a girl, $P(Y \leq 2) = .3173/(1 + .3173) = .2409$, consistent with the SAS ascending results used for Table 4.3. To clarify the approach of SAS with the descending option, consider the complement of $Y \leq 2_{\text{girls}}$, that is, $Y \geq 3_{\text{girls}}$. Using the parameter estimates for the descending model in Table 4.4, we have cumulative $\log_{\text{girls}, Y \geq 3} = \alpha_3 + (-.386) \times gender = +1.147$ (because $gender = 0$ for girls). Then the cumulative $\text{odds}_{\text{girls}, Y \geq 3} = \exp(1.147) = 3.149$. The estimated probability is $P(Y \geq 3)_{\text{girls}} = 3.149/(1 + 3.149) = .759$. This is the complementary probability to $P(Y \leq 2)$ using either SAS ascending or SPSS PLUM; from Table 4.3, $1 - .2410 = .759$.

Fourth, as mentioned previously, all these programs can be asked to save estimated probabilities, which then can be compared easily (at least for models with a small number of predictors) with those for the original data. When running a CO model, SAS will calculate and save the *cumulative* probabilities, according to how you requested them (ascending or descending). SPSS PLUM, however, does not make the cumulative probabilities available directly, but instead calculates and saves the individual's category membership probabilities. As shown just above, the cumulative probabilities can be determined readily from the parameter estimate information provided for any of the three models.

Interpretations across the three models are identical, although the actual values of the thresholds and slopes are not similar between SPSS PLUM and SAS (ascending or descending). This is simply due to how the two packages parameterize the model being fit. Differences between the SAS ascending and descending approaches are seen readily in the reversal of signs and subscripts marking the thresholds. The cumulative odds for the *descending* approach are the odds of being *at or beyond* a particular proficiency level; the cumulative odds for the *ascending* approach and for PLUM are the odds of being *at or below* a particular proficiency level. The thresholds appear in reverse order on the output between the two SAS approaches, but once the predicted logits are transformed to cumulative probabilities, the results are essentially equivalent. The effect of *gender* is reversed in sign for the two SAS models, and in PLUM the *gender* effect corresponds to the case *gender* = 0, but the interested reader is urged to use these simple models to verify equivalence in predicted probabilities once the characterization of the model and the cumulative probabilities being derived are accounted for. An example of the treatment of *gender* across the three models is provided in the paragraphs to follow.

In SPSS PLUM, the threshold estimates are for the case when *gender* = 1 (males), whereas in SAS, the threshold estimates are for the case when *gender* = 0 (females). Regardless of the analysis used, the effect of *gender* is constant across all cumulative splits, $b = \pm.386$. For example, using SAS (ascending), the logit prediction for boys being in proficiency level 2 or lower is $\alpha_2 + b_{\text{gender(boys)}} = -1.147 + .386 = -.761$. This is equivalent to the prediction for boys following the SPSS PLUM analysis: $\theta_2 - b_{\text{gender(boys)}} = -.762 - 0 = -.762$. Exponentiating to find the cumulative odds and transforming the results to find predicted probability for boys of being at or below proficiency level 2, we have $P(Y \leq 2) = .318$ (see Table 4.3). The odds ratios for boys:girls across all cumulative splits are assumed constant based on the proportional odds model and are equivalent for SAS (ascending) and SPSS PLUM: exp(.386) = 1.47; this indicates that the odds for boys of being at or below any category *j* are 1.47 times the odds for girls of being at or below any category *j*.

Using the SAS (descending) approach, we can say that the odds for boys being in category *j or beyond* relative to girls are constant across all cumulative splits: exp(–.386) = .680, which implies that boys are less likely than girls to be *at or beyond* a given proficiency level. Interpreted slightly differently, this result shows that the odds for boys are .68 times the odds for girls of being *at or beyond* any category *j*. Girls are more likely to be in higher proficiency categories. Note that the odds ratios for either approach (ascending or descending) are inverses of each other: 1/.68 = 1.47. Note also that the probability predictions for boys being at or below category *j*, for example, can be determined from the SAS (descending) model as well, because $P(Y \leq j) = 1 - P(Y \geq j + 1)$. As another example of this process, to find $P(Y \leq 2)$ for boys, we can use the descending model to find the cumulative logit for $Y \geq 3$ for boys, $\alpha_3 + b_{\text{gender(boys)}} = 1.147 + (-.386) = .761$; exponentiating and solving for cumulative probability, we find $P(Y \geq 3) = .682$; finally, $1 - .682 = P(Y \leq 2) = .318$, consistent with results shown in Table 4.3.

When results of these models are compared to the multiple regression (MR) analysis, we see a similar pattern in terms of boys being below girls in proficiency. The dependent variable of proficiency in this MR analysis is coded to be increasing in value from 0 to 5. The slope for the *gender* variable (boys = 1) is negative, –.246. On average, girls are predicted to be at a proficiency level of 3.109, whereas boys are predicted to be at a lower proficiency level of (3.109 – .246) = 2.863. Although globally there are similarities between the ordinal models and the MR model in terms of direction of the effect of *gender*, the predicted outcomes from the MR model are not consistent with the data we are analyzing. A mean proficiency score is not the value we wish to predict when our response values are strictly ordinal; furthermore, the MR model does not allow us to make classification statements where we might compare across the different proficiency levels.

Example 4.2: Full-Model Analysis of Cumulative Odds

The analyses thus far indicate that the one-variable model could be improved upon. The predicted probabilities for the *gender*-only model under the proportional odds assumption are very similar to the actual cumulative proportions, and the likelihood ratio test results indicate that the cumulative probabilities when *gender* is included in the model are more consistent with the actual data than the null model (without *gender*). The R^2 statistics were very small, as were Somers' D and the measures of predictive efficiency. We now turn to the derivation of a more complex cumulative odds model to determine the relationship between additional explanatory variables and the cumulative probabilities across the six proficiency levels. Table 4.5 provides a summary of results for the fitting of the CO model with eight explanatory variables from Table 2.2 (recall that *public* is a school-level variable and will not be used in these single-level models). The results in Table 4.5 were obtained using SAS with the descending option; the probabilities being modeled are $P(Y \geq \text{category } j)$. This approach was taken to facilitate later comparison with the CR and AC ordinal models. The syntax for the full CO model is contained in Appendix B, section B4.

The proportional odds assumption for this model is not upheld, as can be seen in the row of Table 4.5 labeled "score test." This suggests that the pattern of effects for one or more of the independent variables is likely to be different across separate binary models fit according to the pattern indicated earlier for the CO model in Table 4.1. Unfortunately, with continuous predictors and large sample sizes, the score test will nearly always indicate rejection of the assumption of proportional odds, and therefore should be interpreted cautiously (Allison, 1999; Greenland, 1994; Peterson & Harrell,

TABLE 4.5
Full-Model Analysis of Cumulative
Odds (CO), SAS (Descending) ($Y \geq$ cat. j), $N = 3{,}365$

Variable	*b (se(b))*	*OR*
α_5	−6.01 (.54)	
α_4	−4.73 (.53)	
α_3	−2.62 (.53)	
α_2	−1.30 (.53)	
α_1	.50 (.54)	
gender	−.50 (.06)**	**.607**
famrisk	−.26 (.08)**	**.771**
center	.09 (.08)	**1.089**
noreadbo	−.32 (.09)**	**.729**
minority	−.15 (.07)*	**.862**
halfdayK	−.17 (.07)*	**.847**
wksesl	.71 (.05)**	**2.042**
plageent	.06 (.01)**	**1.063**
R^2_L	.05	
Cox & Snell R^2	.14	
Nagelkerke R^2	.15	
Somers' D	.33	
τ_p	.21	
λ_p	.00	
Model fit[a]	$\chi^2_8 = 524.17$ ($p < .0001$)	
Score test[c]	$\chi^2_{32} = 75.47$ ($p < .0001$)	

a. Likelihood ratio test.
b. For the proportional odds assumption.
*$p < .05$; **$p < .01$.

1990). We will return to an examination of this assumption later; for now, let us interpret what the model estimates and fit statistics mean for this analysis.

The model fit chi-square indicates that this full model is performing better than the null model (no independent variables) at predicting cumulative probability for proficiency. We see some improvement in the likelihood ratio and pseudo R^2 statistics, but not much more than what was obtained using the *gender*-only model. Somers' D is .333, which is markedly better than what was obtained through the *gender*-only model.

Recall that proficiency was measured through six categories with outcomes as 0, 1, 2, 3, 4, or 5. With the descending option, the threshold estimates in Table 4.5 correspond to predictions of the cumulative logits for students who have a score of 0 on the complete set of independent variables; α_5 corresponds to the cumulative logit for $Y \geq 5$, α_4 corresponds to the cumulative logit for $Y \geq 4$, and so on, until α_1 corresponds to the cumulative logit for $Y \geq 1$. Because all students will have $Y \geq 0$, this first

threshold is not included in the descending cumulative logit model (note that the same is true for $Y \leq 5$ for the ascending cumulative logit model).

The effects of the independent variables within the full CO model shed some important light on how variables contribute to the probability of being at or beyond a particular category. Consistent with the earlier *gender*-only model, boys are less likely than girls to be beyond a particular category (OR = .607). The presence of any family risk factor (*famrisk*, OR = .771), having parents who do not read to their children (*noreadbo*, OR = .729), being in a minority category (*minority*, OR = .862), and attending half-day kindergarten rather than full-day kindergarten (*haldayK*, OR = .847) all have negative coefficients in the model and corresponding ORs that are significantly less than 1.0. These characteristics are associated with a child being in lower proficiency categories rather than in higher categories. On the other hand, age at kindergarten entry (*plageent*, OR = 1.063) and family SES (*wksesl*, OR = 2.042) are positively associated with higher proficiency categories. The slopes for both variables are positive and significantly different from zero in the multivariable model. Attending center-based day care prior to kindergarten (*center*) is not associated with proficiency in this model; the slope is small, and the OR is close to 1.0. These findings are consistent with the literature on factors affecting early literacy, and as such the full model provides a reasonable perspective of the way in which these selected variables affect proficiency in this domain.

In terms of predictive efficiency, neither τ_p or λ_p offers better category predictions relative to the *gender*-only model, which classified *all* children into category 3. For the full-model CO analysis, the cumulative probabilities can be used to determine individual category probabilities as described in the *gender*-only analysis, with the maximum category probability corresponding to the best proficiency level prediction for each child. Table 4.6 provides the results of the classification scheme based on the full CO model. Most of the children are still classified into proficiency level 3, and we can determine from the classification table (using the formulas presented in Chapter 3) that τ_p = .23 and λ_p = 0, indicating no overall improvement in predictions from the *gender*-only analysis. This would be discouraging if category prediction was the sole goal of the model. However, as mentioned in the binary logistic regression example, these measures tell us very little as to *how* the explanatory variables are affecting estimates of cumulative probability across the proficiency levels. Hosmer and Lemeshow (2000) remark that classification is very sensitive to group size and "always favors classification into the larger group, a fact that is independent of the fit of the model" (p. 157). For model fit, the results of the omnibus likelihood ratio test and the Wald tests for contribution of each IV in the model should be preferred. Nonetheless, in some research

TABLE 4.6
Classification Table for Full CO Model, $N = 3,365$

	Predcat0	*Predcat1*	*Predcat2*	*Predcat3*	*Predcat4*	*Predcat5*	*Totals*
profread							
0	**0**	1	9	57	0	0	67
1	1	**2**	12	262	0	1	278
2	0	3	**24**	565	0	2	594
3	1	3	24	**1,428**	0	26	1,482
4	0	1	1	577	**0**	8	587
5	0	0	0	332	0	**25**	357
Totals	2	10	70	3,221	0	62	3,365

situations, reliability in classification may be an important component of model selection criteria; this example demonstrates how these statistics are calculated, as well as how much they can be influenced by group sample size.

Assumption of Proportional Odds and Linearity in the Logit

Within an ordinal model, linearity in the logit cannot be assessed directly, and "only if linear relations between the logits and the covariates are established in the separate binary logistic models [is] a check of the proportional odds assumption . . . meaningful" (Bender & Grouven, 1998, p. 814). Thus, this assumption was investigated for each of the five binary models to provide support for the ordinal model. Linearity in the logit was examined for the continuous variables using the Box-Tidwell method (Hosmer & Lemeshow, 1989; Menard, 1995) and by graphical methods (Bender & Grouven, 1998). For Box-Tidwell, multiplicative terms of the form $X \times \ln(X)$ are created for the continuous explanatory variables and added to the main effects models. Statistically significant interaction terms are an indication that linearity may not be a reasonable assumption for that variable. To look at linearity graphically, deciles can be created for the continuous explanatory variables, then plotted against the proportion of children in the "success" category for each binary logit (at or beyond category *j*). Both approaches were taken for the two continuous variables in the models looked at here: age at kindergarten entry (*plageent*) and family SES (*wksesl*). The graphs revealed a linear trend, but the Box-Tidwell method indicated nonlinearity for the two continuous variables in all five binary logits. Given the graphical pattern, large sample size, and sensitivity of the statistical tests, linearity in the logit was assumed plausible for both continuous variables.

For the full CO model, the score test for the assumption of proportional or parallel odds was rejected. This means that there are some independent variables for which the odds of being at or beyond category *j* are not stable across proficiency levels as *j* changes. Table 4.7 (values have been rounded to save space) provides the results of five separate binary logistic regressions, where the data were dichotomized and analyzed according to the pattern in the second CO column of Table 4.1. That is, each logistic model looks at the probability of being at or beyond proficiency level *j*. For these logistic models (using SPSS), the grouping of categories coded 1 corresponds to children who were at or beyond each successive category, and the code of 0 is used for children below each successive category.

Reviewing the results of the separate logistic models in Table 4.7, relative to the results of the CO model in Table 4.5, we see that all five binary models fit the data well. The model χ^2's are all statistically significant, indicating that each model fits better relative to its corresponding null model; and the H-L tests are all not statistically significant, indicating that observed to predicted probabilities are consistent.

Now let us look at the patterns of slopes and ORs for each explanatory variable across these five models. The effect of *gender*, after adjusting for the other independent variables, does seem to have a dissimilar pattern across the five separate logistic regression splits. Although the average *gender* slope for these five regressions is –.604, which is somewhat close to the *gender* slope from the multivariable CO model (–.500), the odds ratio for boys to girls of being at or beyond proficiency level 1 (.354) are somewhat lower relative to the other four comparisons (.552, .625, .631, and .635, respectively). Note, however, that if we compare the OR for the averaged *gender* slopes from these binary models, exp(–.604) = .547, to the single *gender* OR from the CO model of .607, we see little difference, *on average*. Directionally and on average, the effect of gender is similar across the five logistic regressions. This is true for all the explanatory variables in the model, with the exception of the effect of *minority*. Notice that the direction of the effect of *minority* changes between the first three analyses and the last two. In the first three analyses, the odds are less than 1.0, suggesting that minority children, relative to nonminority children, are more likely to be in the lower proficiency categories. However, there is no difference in the likelihood of being at or beyond proficiency category 4, because the OR is not statistically different from 1.0. The last analysis compares children in categories 0 through 4 with children in category 5. Here we see that minority children are *more likely* than nonminority children to be in category 5 ($b = .238$, OR = 1.268) after adjusting for the presence of the other explanatory variables in the model. This result was not apparent through the cumulative odds model. The CO model provides summary estimates of the effect

TABLE 4.7
Associated Cumulative Binary Models for the CO Analysis (Descending), Where $CUMSP_j$ Compares $Y <$ cat. j to $Y \geq$ cat. j, $N = 3{,}365$

Variable	*$CUMSP_1$* *b* *(se(b))* *OR*	*$CUMSP_2$* *b* *(se(b))* *OR*	*$CUMSP_3$* *b* *(se(b))* *OR*	*$CUMSP_4$* *b* *(se(b))* *OR*	*$CUMSP_5$* *b* *(se(b))* *OR*	*Score Test*[a] p *value*
Constant	3.53 (2.11)	–.55 (.99)	–2.15** (.68)	–4.67** (.68)	–7.34** (.99)	
gender	–1.04 (.28) **.35***	–.60 (.12) **.55***	–.47 (.08) **.63***	–.46 (.08) **.63***	–.46 (.12) **.64***	.249
famrisk	–.15 (.29) .87	–.25 (.13) .78	–.21 (.09) **.81***	–.33 (.10) **.72***	–.28 (.15) .76	.450
center	–.03 (.28) .97	–.10 (.14) .91	.10 (.09) 1.10	.10 (.10) 1.10	.26 (.16) 1.30	.219
noreadbo	–.65 (.27) **.52***	–.36 (.14) **.70***	–.28 (.10) **.75***	–.28 (.12) **.76***	–.50 (.21) **.61***	.095
minority	–.23 (.29) .80	–.42 (.13) **.66***	–.39 (.09) **.68***	.09 (.09) 1.09	.24 (.13) **1.27***	**.000**
halfdayK	.07 (.26) .93	–.00 (.12) 1.00	–.11 (.08) .89	–.26 (.08) **.77***	–.11 (.12) .89	.033
wksesl	1.00 (.17) **2.73***	.77 (.10) **2.17***	.73 (.07) **2.07***	.64 (.06) **1.89***	.87 (.08) **2.39***	**.000**
plageent	.02 (.03) 1.03	.05 (.02) **1.06***	.06 (.01) **1.06***	.06 (.01) 1.06*	.08 (.02) **1.08***	.645
R^2_L	.125	.097	.092	.070	.096	
R^2_N	.136	.128	.149	.115	.128	
Model χ^2_8	**82.11****	**215.49****	**366.40****	**280.39****	**217.92****	
H-L[b] χ^2_8	7.80	10.43	13.41	.74	9.16	

a. Score test for each IV, unadjusted (no other covariates in the model).
b. Hosmer-Lemeshow test, all n.s.
*$p < .05$; **$p < .01$.

of a variable across all cumulative proficiency-level dichotomizations or splits to the data. The imposition of the assumption of proportionality of the odds across these splits does not seem to be valid for the *minority* variable. For all other explanatory variables in the model, the direction and average magnitude of the slopes and the ORs corresponds well to the CO results.

Unfortunately, the score test for the proportional odds assumption is very sensitive to sample size and the number of different possible covariate patterns, which will always be very large when continuous explanatory variables are used. If the assumption is not rejected, the researcher should feel confident that the overall CO model represents the pattern of ORs across the separate cumulative splits very well. If the assumption is not upheld, however, good practice dictates that the separate models be fit and compared with the CO results to check for discrepancies or deviations from the general pattern suggested by the CO model (e.g., Allison, 1999; Bender & Grouven, 1998; Brant, 1990; Clogg & Shihadeh, 1994; Long, 1997; O'Connell, 2000).

To provide an additional check on the plausibility of the proportionality assumption, separate score tests unadjusted for the presence of the other covariates in the cumulative odds model can be reviewed for each of the explanatory variables. In light of the large sample size, a .01 level of significance was used to guide decisions regarding nonproportionality. For each of the single binary models, the score test for the assumption of proportional odds was upheld, with the exception of *minority* and family SES (*wksesl*). The *p* values for these unadjusted tests are presented in the final column of Table 4.7. Across the five binary logit models, the ORs for *wksesl* are approximately 1.9 or larger, indicating that higher-SES children are at least twice as likely as lower-SES children to be in the higher proficiency categories. Given the fact that SES is continuous, the magnitude of the difference in ORs across the binary splits seems to be negligible and as such, a common OR may be a reasonable assumption for this variable. As mentioned above, however, the pattern of change in the ORs for the *minority* variable may clearly be relevant to the study of proficiency, and the effects of this variable should be examined more closely. Although not provided here, follow-up analyses including interactions among the predictors or using a variable for separate categories of race/ethnicity rather than an overall assignment to a minority category could be used to better explain the effects seen in the five binary logit models.

Alternatives to the Cumulative Odds Model

Recall that the best use of the cumulative odds model is to provide for a single parsimonious prediction model for the data. However, if the restriction of equal slopes is not realistic, it is incumbent upon the researcher to work toward explaining how the data are behaving rather than forcing the data to conform to a particular model. There are several alternatives available if, after review of the separate logistic regression analyses and checks on linearity and proportionality, the overall assumption of proportionality in the multivariate ordinal model is deemed suspect.

If variable effects are of primary importance, the researcher may decide to work with the separate logistic regressions to explore and explain divergent explanatory variable patterns across the different cumulative models (Bender & Grouven, 1998). This decision depends on the researcher's overall goals for the analysis and clearly may not be appropriate for every situation or research question. If a parsimonious model or a single set of predicted probabilities is desired, these separate binary logits will not provide it. Alternatively, the researcher may decide to forfeit the ordinal nature of the DV altogether and to fit a multinomial model to the data. This approach may provide some meaningful information in terms of overall variable effects and classification, but it neglects the ordinal nature of the outcome and thus disregards an important aspect of the data. This option, too, may not be optimal for the researcher's goal, but it should be considered if the researcher believes that the majority of the explanatory variables are contributing to the violation of the proportional odds assumption. See Borooah (2002), Ishii-Kuntz (1994), and Agresti (1990, 1996) for examples and discussion of these alternative multinomial approaches.

A third option, and the focus of later chapters in this book, is to consider other types of ordinal regression analyses, such as the continuation ratio method or the adjacent categories method, to try and obtain a single well-fitting and parsimonious model that would aid in our understanding of the data at hand. Chapter 5 demonstrates the use of the CR or continuation ratio model, and Chapter 6 presents the AC or adjacent categories model.

Before turning to a discussion of these additional strategies for analyzing ordered outcomes, one additional method will be presented. In situations where proportionality is questionable based on the behavior of only a subset of the explanatory variables, researchers may opt to fit what are called *partial proportional odds* (PPO) models (Ananth & Kleinbaum, 1997; Koch et al., 1985; Peterson & Harrell, 1990). In essence, PPO models allow for an interaction between an independent variable and the different logit comparisons, which clarifies how the odds for an IV may change across the levels of the outcomes being compared. SAS currently estimates PPO models using PROC GENMOD. The analysis requires data restructuring to reflect whether or not an individual is at or beyond a particular response level (Stokes, Davis, & Koch, 2000). In the restructured data set, a new binary response for each person for each ordered logit comparison is created to indicate whether or not that person is at or beyond each particular response level. For example, with a K-category ordinal response variable, each person would have $K - 1$ lines in the restructured data set. The new outcome variable of interest is derived to indicate, for each of the $K - 1$ logits, whether or not the person was at or beyond category K (excluding the lowest category ($Y = 0$), which all children are at or beyond). Because the data are now correlated (repeated observations among persons), generalized estimating equations (GEE) are used to fit the nonproportional model and then the partial proportional odds model. The use of the GEE approach (Liang & Zeger, 1986) is particularly well suited to the study of repeated measurements over time when the outcomes of interest are categorical (nominal or ordinal). It is based on large-sample properties, which means that the sample size has to be sufficient enough to produce reliable estimates. Stokes et al. (2000) suggest that two-way cross-classifications of the data should yield observed counts of at least five. With continuous explanatory variables, this typically will not be the case, so the sample size should be considered carefully.

Example 4.3: Partial Proportional Odds

Using the ECLS-K example to demonstrate, we can release the assumption of proportional odds for the *minority* variable and refit the model in an attempt to better reflect the pattern seen in Table 4.7. That is, the assumption of proportional odds is retained for all variables in the model except for *minority*. The syntax for the PPO model, including the restructuring of the data set, is included in Appendix B5, following the process outlined by Stokes et al. (2000). Figure 4.2 presents the (edited) printout for this analysis. GENMOD models the probability that a child is at or beyond category j, but because the odds ratios are kept constant across all splits for each variable except *minority*, the results include only one intercept parameter. The threshold values are found by adding the estimates for each corresponding split, which are included toward the middle of the "Analysis of GEE Parameter Estimates" table in Figure 4.2. When reviewing this table, note that the explanatory variable coding scheme uses the "0" category for the categorical variables as the referent. For example, the slope for *gender*, $b = .5002$, is provided for girls (*gender* = 0) rather than for boys (*gender* = 1).

The intercept (–6.9805) is the log-odds that a child would be at or beyond proficiency category 5 ($Y \geq 5$) if all his or her covariate scores were 1, or 0 if continuous; note that the coding of categorical variables follows an internally constructed pattern and that the estimate for split 5 is 0.00. To find the log-odds for $Y \geq 4$, the threshold would be the intercept plus the effect for split 4 (–6.9805 + 1.2670 = –5.7135). The other threshold estimates may be found similarly.

The GEE analysis provides a score statistic for testing the contribution of each explanatory variable to the model; these are found at the end of the output. Results of the score tests indicate that the effect of *minority* for the fifth logit comparison is just marginally statistically significant, $\chi^2_1 = 4.04$, $p = .0445$, yet its interaction with the split variable is strongly significant overall, $\chi^2_4 = 28.80$, $p < .0001$. This result suggests that there are reliable differences in the effect of *minority* depending on split. For the other explanatory variables in the model, all effects are statistically significant except attendance at a daycare center (*center*), consistent with what was found in the full cumulative odds model. GENMOD also provides z tests (the normal distribution version of the Wald statistic) for the contribution of explanatory variables in the model; these are found in the "Analysis of GEE Parameter Estimates" table of Figure 4.2. Results of the z tests are consistent with the score tests, with the exception of *minority*.

Given the interaction between *minority* and split, the effect for *minority* is interpreted via the score test that specifically examines its contribution for the fifth cumulative comparison. For each of the other splits, the z tests for the *minority* × split interactions contained in the model suggest that there is no difference in the odds for minority versus nonminority children for the first cumulative comparison ($Y \geq 1$), $b_{int.1} = .5077$, $p = .0677$, nor for the fourth ($Y \geq 4$), $b_{int.4} = .0282$, $p = .7884$. Substantively, these findings are consistent with those of the separate binary models in Table 4.7. There, *minority* had no statistical effect on the individual cumulative logits either for the first binary model ($p > .05$) or for the fourth ($p > .05$).

The *minority* × split interactions inform us as to how much change occurs in the effect of *minority* across the thresholds of the response variable. With the assumption of proportional odds relaxed for *minority*, the results shown in the printout tell us how much the log-odds are expected to change for nonminority children relative to minority children, across the different logistic regression splits. For example, after adjusting for the other covariates in the model, the odds ratio for a nonminority child relative to a minority child for a proficiency score at or beyond category 5 is exp(–.1560) = .855; the odds ratio for a minority child relative to a nonminority child for a proficiency score at or beyond category 5 is then exp(+.1560) = 1.169. This OR can be compared with Table 4.7 for the fifth cumulative logistic regression split (where OR = 1.268). Further, this OR is statistically different from 1.0 in the PPO model ($p = .0445$ in "Score Statistics For Type 3 GEE Analysis" table), as it is in the fifth cumulative comparison based on the separate binary models ($p < .05$ for last split in Table 4.7).


```
The GENMOD Procedure

                         Model Information

                 Data Set                  WORK.PPOM
                 Distribution              Binomial
                 Link Function             Logit
                 Dependent Variable        beyond
                 Observations Used         16825


                       Class Level Information

Class       Levels   Values

split           5    1 2 3 4 5
GENDER          2    0 1
FAMRISK         2    0.00 1.00
CENTER          2    0.00 1.00
NOREADBO        2    0.00 1.00
MINORITY        2    0.00 1.00
HALFDAYK        2    0.00 1.00
CHILDID      3365    0212014C 0294004C 3035008C 3042008C 3042023C
                     0044007C 0195025C 0243009C 0621012C 0748011C
                     0832023C 3041005C 0028009C 0028014C 0052003C
                     0052007C 0195020C 0196007C 0196016C 0196017C
                     0196018C 0212002C 0212012C 0216006C 0220005C
                     0220020C 0301002C 0301004C ...


                         Response Profile

                  Ordered                      Total
                   Value        beyond      Frequency

                     1            1            10045
                     2            0             6780

PROC GENMOD is modeling the probability that beyond='1'.


               Criteria For Assessing Goodness Of Fit

  Criterion                  DF           Value          Value/DF
  Deviance                 17E3      12003.9325            0.7142
  Scaled Deviance          17E3      12003.9325            0.7142
  Pearson Chi-Square       17E3      16074.6352            0.9564
  Scaled Pearson X2        17E3      16074.6352            0.9564
  Log Likelihood                     -6001.9662
```

Figure 4.2 Partial Proportional Odds for Minority: GEE Analysis

Figure 4.2 (Continued)

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter			Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept			-6.9805	0.5753	-8.1081	-5.8528	-12.13	<.0001
GENDER	0		0.5002	0.0663	0.3703	0.6301	7.55	<.0001
GENDER	1		0.0000	0.0000	0.0000	0.0000	.	.
FAMRISK	0.00		0.2596	0.0778	0.1071	0.4120	3.34	0.0008
FAMRISK	1.00		0.0000	0.0000	0.0000	0.0000	.	.
CENTER	0.00		-0.0759	0.0770	-0.2268	0.0750	-0.99	0.3240
CENTER	1.00		0.0000	0.0000	0.0000	0.0000	.	.
NOREADBO	0.00		0.3366	0.0913	0.1575	0.5156	3.68	0.0002
NOREADBO	1.00		0.0000	0.0000	0.0000	0.0000		.
MINORITY	0.00		-0.1560	0.1240	-0.3989	0.0870	-1.26	0.2083
MINORITY	1.00		0.0000	0.0000	0.0000	0.0000	.	.
HALFDAYK	0.00		0.1451	0.0666	0.0145	0.2757	2.18	0.0295
HALFDAYK	1.00		0.0000	0.0000	0.0000	0.0000	.	.
WKSESL			0.7450	0.0514	0.6442	0.8457	14.49	<.0001
P1AGEENT			0.0588	0.0084	0.0423	0.0753	7.00	<.0001
split	1		6.2595	0.1851	5.8968	6.6223	33.82	<.0001
split	2		4.4067	0.1204	4.1707	4.6428	36.59	<.0001
split	3		3.0966	0.1043	2.8923	3.3010	29.70	<.0001
split	4		1.2670	0.0846	1.1012	1.4328	14.98	<.0001
split	5		0.0000	0.0000	0.0000	0.0000	.	.
split*MINORITY	1	0.00	0.5077	0.2779	-0.0370	1.0523	1.83	0.0677
split*MINORITY	1	1.00	0.0000	0.0000	0.0000	0.0000	.	.
split*MINORITY	2	0.00	0.6021	0.1621	0.2843	0.9198	3.71	0.0002
split*MINORITY	2	1.00	0.0000	0.0000	0.0000	0.0000	.	.
split*MINORITY	3	0.00	0.5230	0.1334	0.2615	0.7844	3.92	<.0001
split*MINORITY	3	1.00	0.0000	0.0000	0.0000	0.0000	.	.
split*MINORITY	4	0.00	0.0282	0.1050	-0.1777	0.2340	0.27	0.7884
split*MINORITY	4	1.00	0.0000	0.0000	0.0000	0.0000	.	.
split*MINORITY	5	0.00	0.0000	0.0000	0.0000	0.0000	.	.
split*MINORITY	5	1.00	0.0000	0.0000	0.0000	0.0000	.	.

Score Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
GENDER	1	57.02	<.0001
FAMRISK	1	11.10	0.0009
CENTER	1	0.97	0.3243
NOREADBO	1	13.30	0.0003
MINORITY	1	4.04	0.0445
HALFDAYK	1	4.75	0.0294
WKSESL	1	195.03	<.0001
P1AGEENT	1	49.02	<.0001
split	4	2447.16	<.0001
split*MINORITY	4	28.80	<.0001

To find the effect of *minority* for the fourth cumulative logit, $(Y \geq 4)$, the interaction terms are added to the main effect. That is, for the odds of a nonminority child being at or beyond category 4, exp(–.1560 + .0282) = exp(–.1278) = .880; for minority children, this corresponds to exp(+.1278) = 1.136. Minority children are 1.136 times as likely to be at or beyond category 4, although this effect is not statistically different from 1.0 (p = .7884). This effect is consistent with the OR for the fourth cumulative logit in Table 4.7, which also was not statistically significant (OR = 1.092, not significant). For the first logit, $(Y \geq 1)$, the effect for nonminority children is exp(–.1560 + .5077) = exp(.3517) = 1.42; for minority children, the effect is exp(–.3517) = .7035. According to the PPO model, this effect is not significant (p = .0677), consistent with the result for the effect of *minority* at this first split in Table 4.7 (OR = .796, not significant). Overall, minority children are less likely than their nonminority peers to advance beyond proficiency levels 2 and 3, but given that they have attained at least proficiency level 4, they are more likely than their nonminority peers to then achieve mastery in proficiency level 5.

To examine the effects of the explanatory variables for which the proportional odds assumption was retained, the slope estimates can be interpreted directly. For the effect of *gender*, girls (*gender* = 0) are exp(+.5002) = 1.65 times as likely as boys to be at or beyond level 1, after adjusting for other covariates in the model, and this OR remains constant across all underlying cumulative logits. Because the events for this explanatory variable with only two levels are complementary, we can easily interpret the effect for boys as well: boys are exp(–.5002) = .606 times as likely as girls to be at or beyond a given proficiency category j, after adjusting for other covariates. For all explanatory variables with the exception of *minority*, the effects are equivalent to those presented for the full CO model in Table 4.5, once the coding of IVs is taken into account. For example, in the full CO model the gender slope is –.500 with OR = .607. Variable effects in the PPO model for those variables for which the proportional odds assumption was retained are of the same magnitude and statistical significance as those in the CO model. The direction has changed because SAS PROC GENMOD provides the estimates for the values of the explanatory variable coded as 0 rather than 1. Note that the nature of the coding for the categorical IVs does not affect the results for the continuous variables in the model between the CO and PPO models.

To summarize the PPO analysis, this approach does resolve some of the issues surrounding the full proportional odds model, particularly for the *minority* variable. The GEE estimates correspond quite well with the separate effects for minority that were examined across the binary logit models in terms of both magnitude and statistical significance. In the study of early reading achievement, this result bears further investigation. Creating a variable that categorizes groupings of children based on race/ethnicity for inclusion in these models rather than including all "nonwhite" children together in a dichotomous arrangement should be further examined, but this is not the focus of the current demonstration. The effects for the variables that were constrained to follow the proportional odds assumption were found to be consistent with the earlier CO analysis. There is currently no overall summary measure of goodness of fit for a GEE analysis provided through GENMOD (Stokes et al., 2000), but the criteria included in the output under the "Criteria for Assessing Goodness of Fit" heading indicate that the deviance (found through a comparison between the fitted model and the perfect, or saturated, model) is less than its degrees of freedom (value/*df*) < 1.0), suggestive of adequate model fit (Allison, 1999). Recall that there is no reliable test of the model deviance when continuous variables are present. However, these statistics can be useful for comparisons of competing models. Overall, the PPO model seems to be more informative than the CO model, particularly with regard to the explanatory variable of *minority*.

5. THE CONTINUATION RATIO MODEL

Overview of the Continuation Ratio Model

As we saw in Chapter 4, the cumulative odds model uses all the data available to assess the effect of independent variables on the log-odds of being *at or beyond* (or the reverse, *at or below*) a particular category. The odds are found by considering the probability of being *at or beyond* a category relative to the probability of being *below* that category. A restrictive assumption made in the CO analysis is that across all cumulative logit comparisons, the effect of any independent variable is similar; that is, the odds of being in higher categories relative to being in *any* category below it remains constant across the categories. However, these logit comparisons for the cumulative odds may not be theoretically appropriate in every research situation. If interest lies in determining the effects of independent variables on the event of being in a higher stage or category, then a comparison group that includes *all* people who failed to make it to a category may not lead us to the best conclusions or understanding of the data in terms of differences between people at a low stage versus all higher stages. Rather than grouping together all people who failed to make it to a category at any point, an alternative ordinal approach involves comparisons between respondents in any given category versus all those who achieved a

higher category score. This approach forms the class of models known as continuation ratio (CR) models. The focus of a CR analysis is to understand the factors that distinguish between those persons who have reached a particular response level but do not move on from those persons who do advance to a higher level. Fox (1997) refers to this process as the analysis of a series of "nested dichotomies" (p. 472).

A continuation ratio is a *conditional* probability. The discussion to follow explains how these continuation ratios can be formed in different ways, depending on the researcher's goals. The examples presented are based on continuation ratios that take the form $\delta_j = P$(response beyond cat. j|response in at least cat. j), or its complement, $1 - \delta_j = P$(response in cat. j|response in at least cat. j).

Armstrong and Sloan (1989), McCullagh and Nelder (1983), Greenland (1994), and Agresti (1990, 1996) have discussed the continuation ratio model in depth and have highlighted the relationship between the CR model and the proportional hazards model proposed by D. R. Cox (1972). The proportional hazards model is a familiar one in epidemiological contexts and in the survival analysis research literature, but its value can be extended to other contexts as well.

The CR models can be fit using a suitably restructured data set with either a logit link function or a complementary log-log (clog-log) link function. The restructuring is explained in greater detail later, but essentially, a new data set is created from $K-1$ smaller data-sets, in which each person has as many data lines as his or her outcome score allows. The process is similar to how the concatenated data set was created for the partial proportional odds analysis, with the very important exception that inclusion in a data set is conditional on whether or not mastery was attained at a particular level. The resulting data sets then correspond to the specific comparisons contained in Table 4.1 for the continuation ratio analyses. Once the data set is concatenated, the outcome of interest is on whether or not a child advances beyond a particular category, given that at least mastery in that category was attained. The data sets formed in this fashion are conditionally independent (Armstrong & Sloan, 1989; Clogg & Shihadeh, 1994; Fox, 1997); thus, the restructured data set can be analyzed using statistical methods for binary outcomes.

The restructuring is necessary in order to derive the desired conditional probabilities, or, in the case of the clog-log link, the hazards. In the epidemiological literature, the hazard ratio is also known as relative risk; it is a ratio of two hazards, where the hazard is an explicit conditional probability. The odds ratio, on the other hand, is a ratio of two odds, where the odds are a quotient of complementary probabilities, $p/(1 - p)$. Of course, the probability of interest in a logit model could be a conditional probability, which clarifies the usefulness of the logit link for continuation ratio