ECONOMIC GROWTH CENTER

YALE UNIVERSITY

P.O. Box 208269
New Haven, CT 06520-8269
http://www.econ.yale.edu/~egcenter/

CENTER DISCUSSION PAPER NO. 873

AN EXTENSION OF THE BLINDER-OAXACA DECOMPOSITION TECHNIQUE TO LOGIT AND PROBIT MODELS

Robert W. Fairlie
Yale University

November 2003

Notes: Center Discussion Papers are preliminary materials circulated to stimulate discussions and critical comments.

I would like to thank Rebecca London and Lori Kletzer for helpful comments and suggestions. Garima Vasishtha provided excellent research assistance.

An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models

Robert W. Fairlie

Abstract

The Blinder-Oaxaca decomposition technique is widely used to identify and quantify the separate contributions of group differences in measurable characteristics, such as education, experience, marital status, and geographical differences to racial and gender gaps in outcomes. The technique cannot be used directly, however, if the outcome is binary and the coefficients are from a logit or probit model. I describe a relatively simple method of performing a decomposition that uses estimates from a logit or probit model. Expanding on the original application of the technique in Fairlie (1999), I provide a more thorough discussion of how to apply the technique, an analysis of the sensitivity of the decomposition estimates to different parameters, and the calculation of standard errors.

Keywords: Logit, Probit, Decomposition, Race, Gender, Discrimination

JEL Codes: C8, J7

1. Introduction

Identifying the underlying causes of racial and gender differences in educational, labor market, health, and other outcomes has been the goal of an enormous body of literature in the social sciences. Perhaps the most common approach used in the past few decades to identify and quantify these causes is the technique of decomposing inter-group differences in mean levels of an outcome into those due to different observable characteristics or "endowments" across groups and those due to different effects of characteristics or "coefficients" of groups. The technique is commonly attributed to Blinder (1973) and Oaxaca (1973). Attesting to the wide use of the Blinder-Oaxaca decomposition technique, more than 1000 citations to these two articles were found in the Social Sciences Citation Index as of June 2003.

The Blinder-Oaxaca decomposition technique is especially useful for identifying and quantifying the separate contributions of group differences in measurable characteristics, such as education, experience, marital status, and geographical location, to racial and gender gaps in outcomes.[1] The technique is easy to apply and only requires coefficient estimates from linear regressions for the outcome of interest and sample means of the independent variables used in the regressions. A problem arises, however, if the outcome is binary, such as employment, college attendance, or teenage pregnancy, and the coefficients are from a logit or probit model. These coefficient estimates cannot be used directly in the standard Blinder-Oaxaca decomposition equations.

A relatively simple method of performing a decomposition that uses estimates from a logit or probit model was first described in Fairlie's (1999) analysis of the causes of the black/white gap in self-employment rates. In this paper, I provide a more thorough discussion of how to apply the technique, an analysis of the sensitivity of the decomposition estimates to different parameters, and the calculation of standard errors. The non-linear decomposition

[1] Although not as commonly used, the technique is also useful for identifying the causes of geographical (e.g. urban/rural or cross-country), time period, or other categorical differences in outcomes.

technique described below may be useful for identifying the causes of racial, gender, geographical or other categorical differences in a binary outcome in which a logit or probit model is used. [2]

2. Non-Linear Decomposition Technique

For a linear regression, the standard Blinder-Oaxaca decomposition of the white/black gap (male/female, North//South, etc...) in the average value of the dependent variable, Y, can be expressed as:

$$(2.1)\quad \overline{Y}^W - \overline{Y}^B = \left[(\overline{X}^W - \overline{X}^B)\hat{\beta}^W\right] + \left[\overline{X}^B(\hat{\beta}^W - \hat{\beta}^B)\right],$$

where $\overline{X}^j$ is a row vector of average values of the independent variables and $\hat{\beta}^j$ is a vector of coefficient estimates for race j. Following Fairlie (1999), the decomposition for a nonlinear equation, such as $Y = F(X\hat{\beta})$, can be written as:

$$(2.2)\quad \overline{Y}^W - \overline{Y}^B = \left[\sum_{i=1}^{N^W}\frac{F(X_i^W\hat{\beta}^W)}{N^W} - \sum_{i=1}^{N^B}\frac{F(X_i^B\hat{\beta}^W)}{N^B}\right] + \left[\sum_{i=1}^{N^B}\frac{F(X_i^B\hat{\beta}^W)}{N^B} - \sum_{i=1}^{N^B}\frac{F(X_i^B\hat{\beta}^B)}{N^B}\right],$$

where N^j is the sample size for race j. This alternative expression for the decomposition is used because $\overline{Y}$ does not necessarily equal $F(\overline{X}\ \hat{\beta})$.[3] In both (2.1) and (2.2), the first term in brackets represents the part of the racial gap that is due to group differences in distributions of X, and the second term represents the part due to differences in the group processes determining levels of Y. The second term also captures the portion of the racial gap due to group differences in unmeasurable or unobserved endowments. Similar to most previous studies applying the decomposition technique, I do not focus on this "unexplained" portion of the gap because of the difficulty in interpreting results (see Jones 1983 and Cain 1986 for more discussion).

[2] SAS programs for calculating the decomposition are available by request.
[3] Note that the Blinder-Oaxaca decomposition is a special case of (2.2).

To calculate the decomposition, define $\overline{Y}^{j}$ as the average probability of the binary outcome of interest for race *j* and *F* as the cumulative distribution function from the logistic distribution.[4] Alternatively, for a probit model F would be defined as the cumulative distribution function from the standard normal distribution.

An equally valid expression for the decomposition is:

$$(2.3)\ \overline{Y}^{W} - \overline{Y}^{B} = \left[\sum_{i=1}^{N^{W}} \frac{F(X_i^{W}\hat{\beta}^{B})}{N^{W}} - \sum_{i=1}^{N^{B}} \frac{F(X_i^{B}\hat{\beta}^{B})}{N^{B}}\right] + \left[\sum_{i=1}^{N^{W}} \frac{F(X_i^{W}\hat{\beta}^{W})}{N^{W}} - \sum_{i=1}^{N^{W}} \frac{F(X_i^{W}\hat{\beta}^{B})}{N^{W}}\right],$$

In this case, the black coefficient estimates, $\hat{\beta}^{B}$ are used as weights for the first term in the decomposition, and the white distributions of the independent variables, $\overline{X}^{W}$ are used as weights for the second term. This alternative method of calculating the decomposition often provides different estimates, which is the familiar index problem with the Blinder-Oaxaca decomposition technique. A third alternative, used in Neumark (1988) and Oaxaca and Ransom (1994), is to weight the first term of the decomposition expression using coefficient estimates from a pooled sample of the two groups. Ultimately, the choice across these alternative methods of calculating the first term of the decomposition is difficult and depends on the application with many studies reporting results for more than one specification.

The first terms in (2.2) and (2.3) provide an estimate of the contribution of racial differences in the entire set of independent variables to the racial gap in computer ownership. Estimation of the total contribution is relatively simple as one only needs to calculate two sets of predicted probabilities and take the difference between the average values of the two. Identifying the contribution of group differences in specific variables to the racial gap, however, is not as straightforward. To simplify, first assume that $N_B=N_W$ and that there exists a natural one-to-one

[4] A useful property of the logit regression that includes a constant term is that the average of the predicted probabilities must equal the proportion of ones in the sample. In contrast, the predicted probability

matching of black and white observations. Using coefficient estimates from a logit regression for a pooled sample, $\hat{\beta}^*$, the independent contribution of X_1 to the racial gap can then be expressed as:

$$(2.4)\ \frac{1}{N^B}\sum_{i=1}^{N^B} F(\hat{\alpha}^* + X_{1i}^W \hat{\beta}_1^* + X_{2i}^W \hat{\beta}_2^*) - F(\hat{\alpha}^* + X_{1i}^B \hat{\beta}_1^* + X_{2i}^W \hat{\beta}_2^*).$$ [5]

Similarly, the contribution of X_2 can be expressed as:

$$(2.5)\ \frac{1}{N^B}\sum_{i=1}^{N^B} F(\hat{\alpha}^* + X_{1i}^B \hat{\beta}_1^* + X_{2i}^W \hat{\beta}_2^*) - F(\hat{\alpha}^* + X_{1i}^B \hat{\beta}_1^* + X_{2i}^B \hat{\beta}_2^*).$$

The contribution of each variable to the gap is thus equal to the change in the average predicted probability from replacing the black distribution with the white distribution of that variable while holding the distributions of the other variable constant.[6] A useful property of this technique is that the sum of the contributions from individual variables will be equal to the total contribution from all of the variables evaluated with the full sample.

Standard errors can also be calculated for these estimates. Following Oaxaca and Ransom (1998), I use the delta method to approximate standard errors. To simplify notation, rewrite (2.4) as:

$$(2.6)\ \hat{D}_1 = \frac{1}{N^B}\sum_{i=1}^{N^B} F(X_i^{WW}\hat{\beta}^*) - F(X_i^{BW}\hat{\beta}^*).$$

The variance of $\hat{D}_1$ can be approximated as:

$$(2.7)\ Var(\hat{D}_1) = \left(\frac{\delta \hat{D}_1}{\delta \hat{\beta}^*}\right)' Var(\hat{\beta}^*) \left(\frac{\delta \hat{D}_1}{\delta \hat{\beta}^*}\right).$$

---

evaluated at the means of the independent variables is not necessarily equal to the proportion of ones, and in the sample used below it is larger because the logit function is concave for values greater than 0.5.

[5] A black dummy variable is included in estimating the logit model with the pooled sample of blacks and whites, but is not used to calculate the decomposition.

where $\frac{\delta \hat{D}_1}{\delta \hat{\beta}^*} = \frac{1}{N^B} \sum_{i=1}^{N^B} f(X_i^{WW} \hat{\beta}^*) X_i^{WW} - f(X_i^{BW} \hat{\beta}^*) X_i^{BW}$ and f is the logistic probability density function.

In practice, the sample sizes of the two groups are rarely the same and a one-to-one matching of observations from the two samples is needed to calculate (2.4), (2.5), and (2.7). In this example, it is likely that the black sample size is substantially smaller than the white sample size. To address this problem, first use the pooled coefficient estimates to calculate predicted probabilities, $\hat{Y}_i$, for each black and white observation in the sample. Next, draw a random subsample of whites equal in size to the full black sample (N_B). Each observation in the white subsample and full black sample is then separately ranked by the predicted probabilities and matched by their respective rankings. This procedure matches whites who have characteristics, such as income and education, placing them at the bottom (top) of their distribution with blacks who have characteristics placing them at the bottom (top) of their distribution.[7]

The decomposition estimates obtained from this procedure depend on the randomly chosen subsample of whites. Ideally, the results from the decomposition should approximate those from matching the entire white sample to the black sample. A simple method of approximating this hypothetical decomposition is to draw a large number of random subsamples of whites, match each of these random subsamples of whites to the black sample, and calculate separate decomposition estimates. The mean value of estimates from the separate decompositions is calculated and used to approximate the results for the entire white sample. In

[6] Unlike in the linear case, the independent contributions of X_1 and X_2 depend on the value of the other variable. This implies that the choice of a variable as X_1 or X_2 (or the order of switching the distributions) is potentially important in calculating its contribution to the racial gap. I return to this issue below.
[7] The results presented below, however, are fairly similar when simply matching black and white observations randomly.

the decompositions reported below, I use 1000 random subsamples of whites to calculate these means.[8]

3. Results

For illustrative purposes, I employ the non-linear decomposition technique to identify the causes of racial differences in computer ownership (see Fairlie 2002 for more details). Estimates from the Computer and Internet Use Supplement to the August 2000 Current Population Survey (CPS) indicate that 70.9 percent of white, non-Latinos have access to a home computer, whereas only 41.3 percent of African-Americans have access to a home computer. Of particular interest is whether (and how much) group differences in the most likely "suspects" -- family income, education and family structure -- contribute to this racial disparity in computer ownership. The findings are useful for the policy debate over the causes and consequences of the "Digital Divide."

Table 1 reports estimates of the nonlinear decomposition technique for the black/white gap in home computer rates using four different set of coefficients. Logit regressions are estimated using four separate samples -- white only, black only, white and black pooled, and all races pooled.[9] The individual contributions from racial differences in sex and age, marital status and children, education, income, region, and central city status are reported. The contribution for a set of dummy variables, such as those for region, is calculated by simultaneously switching distributions of all dummy variables. The results are generally similar across specifications.

[8] Estimates for the main specification are identical to the 4th decimal place using 10,000 simulations for all contributions except two groups of variables (which were both less than 0.0001 different). In fact, using only 100 simulations provided contribution estimates that were identical to the 4th decimal place except for only two groups of variables (which were both less than 0.0002 different).

[9] Estimates are reported in the Appendix. All specifications include measures of sex, age, marital status, children, education, family income, region, and central city status. The coefficient estimates indicate that being married, the presence of children between the ages of 6 and 17, education, and family income increase the probability of computer ownership. The coefficient estimates are generally similar across specifications

The difference between white and black computer ownership rates is 0.3030. As expected, the largest factor explaining this large racial disparity in home computer ownership is income. Lower levels of income among blacks account for 0.0775 to 0.1031 (or 25.6 to 34.0 percent) of the white/black gap in the probability of having a home computer. In all specifications these contributions are statistically significant. Lower levels of education among blacks also contribute to the racial gap in computer ownership. The decomposition estimates for the contribution of racial differences in education range from 0.0340 to 0.0369 (or 11.2 to 12.2 percent). Group differences in family characteristics explain a similarly large portion of the gap (9.5 to 10.5 percent), whereas group differences in regional distributions explain a small portion of the gap (2.8 to 5.2 percent). Finally, racial differences in sex and age and racial differences in central city status explain virtually none of the gap. The decompositions reveal that group differences in all of the included characteristics explain roughly half of the black/white gap in computer ownership for the white and pooled specifications and 61.3 percent of the gap for the black specification.

COMPARISON TO BLINDER-OAXACA RESULTS

Perhaps the first question that comes to mind regarding use of the technique is how the results compare to those from a standard Blinder-Oaxaca decomposition using estimates from a linear probability model. Specification 2 of Table 2 reports estimates. Specification 1 reports estimates from the non-linear decomposition technique for comparison. For brevity, I focus on the results for the All Races Pooled Specification. The estimates from the Blinder-Oaxaca decomposition do not differ substantially from those from the non-linear decomposition technique. The largest difference is for the marital status and children contribution, which is 0.0064 smaller in the linear specification. For this application in which the coefficient estimates are well behaved and the decomposition is explaining a racial gap primarily located in the middle of the distribution (i.e. between 0.4257 and 0.7286) the Blinder-Oaxaca decomposition

approximates the nonlinear decomposition results. The Blinder-Oaxaca decomposition may not perform as well in cases in which the racial gap is located in the tails of the distribution or racial differences in independent variables (e.g. a continuous measure of wealth) are very large. In fact, an endowment effect may even be larger than 1 (or 100 percentage points) in the Blinder-Oaxaca decomposition, which is difficult to interpret in terms of probability.

A good example of when the Blinder-Oaxaca technique may be problematic is in Fairlie (1999). The contribution of racial differences in asset levels to the black/white gap in the transition rate out of self-employment is 0.0117 or 9.0 percent using black coefficients as reported in Specification 1 of Table 3 in the article. Using a linear probability model and the Blinder-Oaxaca decomposition, I find a contribution estimate of 0.1812 or 114.4 percent. The linear technique provides a much larger contribution because white asset levels are approximately 7 times larger than black asset levels as measured in the regression. The linear technique fails to limit the influence on the probability that the outcome occurs at increasingly high asset levels (even in this case when a quadratic term is included), and thus may be providing an implausible estimate of the impact of racial differences in asset levels on black/white differences in exit rates.

Related to this issue, I also estimate a model using coefficients from a probit regression. In this case, F in (2.4) and (2.5) is defined as the cumulative distribution function from the standard normal distribution. Estimates are reported in Specification 3 of Table 2. The contribution estimates are very similar to those using the logit coefficients. For this application, the decomposition estimates are not sensitive to whether the logit or probit model is used.

ORDERING OF VARIABLES

Another potentially important issue regarding use of the technique is the effect of ordering of variables in the decomposition. As noted above, because of the nonlinearity of the decomposition equation the results may be sensitive to the ordering of variables. To investigate this issue, Specification 4 of Table 2 reports estimates in which the order of switching

distributions of variables is reversed. Overall, the estimates are not substantially different than the original estimates, but the differences are worth noting. The main changes are that the family characteristics contribution declined from 0.0289 to 0.0175, and the education and income contributions increased slightly. The total contribution, however, remains unchanged because the sum of the individual contributions, regardless of their order, must equal the total contribution defined in (2.2) or (2.3).

The effects of reordering, however, depend on the application. The initial location in the logistic distribution and the total movement along the distribution from switching distributions of other variables contribute to how sensitive the results are to the ordering of variables. Perhaps the best solution to the problem is to experiment with different orders of variables to confirm the robustness of results.[10] If the results continue to vary substantially, one solution for coming up with a single point estimate is to randomize the ordering of variables. In fact, the ordering of switching distributions could be randomized at the same time as drawing the random subsample of whites. By using a large number of simulations the procedure approximates the average decomposition across all possible orderings of variables while preserving the summing up property. As a check, I estimate the decomposition using this procedure. All of the estimates lie in the intervals created by the estimates reported in Specifications 1 and 4, which represent the original "ad hoc" ordering of the variables and its reverse. In fact, I experimented with different initial orderings of the variables and their reverses, and found that the average contribution in each case approximated the estimate from the random ordering decomposition, suggesting a relatively easy method of checking the sensitivity of results.

THE USE OF SAMPLE WEIGHTS

All of the estimates previously reported are unweighted for simplicity. If sample weights are required, however, the decomposition technique needs to be modified slightly. If sample weights are used to estimate both the mean outcomes and the logit regressions, then each observation in (2.4) and (2.5) should be weighted. The only complication arises in choosing whether the white or black sample weight is used when switching distributions, which represents another index problem as there is no theoretical justification for preferring one over the other.

The other case is where sample weights are used to estimate mean outcomes, but not the regressions. In this case, each observation in (2.4) and (2.5) should be weighted and the interpretation of the contribution from group differences in observed characteristics remains unchanged. The "left-over" or unexplained portion of the decomposition, however, now also includes group differences in the discrepancy between the weighted mean of the outcome and the average predicted probability using the weighted distribution of characteristics and unweighted regression estimates.[11] Estimating decompositions using both white and black sample weights from the CPS, I find estimates that do not differ substantially from the unweighted estimates.

5. Summary

The non-linear decomposition technique discussed above is a relatively easy-to-implement alternative to the standard Blinder-Oaxaca decomposition. The main advantage is that the coefficient estimates from a logit or probit model can be used directly in the decomposition specification. The technique is thus useful for applications in which it is inappropriate to model the dependent variable as a linear function of the explanatory variables.

---

[10] Another solution is to estimate each contribution by switching the variable of interest first (i.e. use (2.4) to estimate the contribution for each variable). The sum of these contributions, however, may differ substantially from the total contribution defined in (2.2) or (2.3).

[11] Note that if the weighted mean outcomes are similar to the unweighted mean outcomes then this is unlikely to be large.

References

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Variables." Journal of Human Resources, 8, 436-455.

Cain, Glen G. 1986. "The Economic Analysis of Labor Market Discrimination: A Survey," Handbook of Labor Economics, Vol. 1, eds. O. Ashenfelter and R. Laynard, Elsevier Science Publishers BV.

Fairlie, Robert W. 1999. "The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment," Journal of Labor Economics, 17(1): 80-108.

Fairlie, Robert W. 2002. "Race and The Digital Divide," University of California, Santa Cruz Working Paper.

Jones, F.L. 1983. "On Decomposing the Wage Gap: A Critical Comment on Blinder's Method," Journal of Human Resources, 18(1): 126-130.

Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination," Journal of Human Resources, 23, 279-295.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets," International Economic Review, 14 (October), 693-709.

Oaxaca, Ronald, and Michael Ransom. 1994. "On Discrimination and the Decomposition of Wage Differentials," Journal of Econometrics, 61, 5-21.

Oaxaca, Ronald, and Michael Ransom. 1998. "Calculation of Approximate Variances for Wage Decomposition Differentials," Journal of Economic and Social Measurement, 24, 55-61.

Table 1
Non-Linear Decompositions of Black/White Gaps in Home Computer Rates
Using Various Coefficient Estimates

	Specification			
	(1)	(2)	(3)	(4)
Sample used for coefficients	White	Black	Black/White Pooled	All Races Pooled
White computer ownership rate	0.7286	0.7286	0.7286	0.7286
Black computer ownership rate	0.4257	0.4257	0.4257	0.4257
Black/White gap	0.3030	0.3030	0.3030	0.3030
Contributions from racial differences in:				
Sex and age	-0.0004	0.0001	-0.0004	-0.0002
	(0.0003)	(0.0010)	(0.0003)	(0.0002)
	-0.1%	0.0%	-0.1%	-0.1%
Marital status and children	0.0315	0.0302	0.0317	0.0289
	(0.0017)	(0.0041)	(0.0015)	(0.0014)
	10.4%	10.0%	10.5%	9.5%
Education	0.0340	0.0367	0.0341	0.0369
	(0.0010)	(0.0028)	(0.0009)	(0.0008)
	11.2%	12.1%	11.2%	12.2%
Income	0.0775	0.1031	0.0797	0.0799
	(0.0020)	(0.0048)	(0.0019)	(0.0017)
	25.6%	34.0%	26.3%	26.4%
Region	0.0102	0.0157	0.0104	0.0085
	(0.0017)	(0.0055)	(0.0016)	(0.0015)
	3.4%	5.2%	3.4%	2.8%
Central city status	-0.0015	0.0000	-0.0007	-0.0019
	(0.0021)	(0.0041)	(0.0019)	(0.0017)
	-0.5%	0.0%	-0.2%	-0.6%
All included variables	0.1512	0.1859	0.1548	0.1519
	49.9%	61.3%	51.1%	50.2%

Notes: (1) The sample consists of adults ages 25-55 from the specified racial group/s. (2) Standard errors are reported in parantheses below contribution estimates. (3) The sample sizes used to estimate the coefficients in Specifications 1-4 are 34,386, 4,555, 38,941 and 46,322, respectively. (4) Contribution estimates are mean values of the decomposition using 1000 subsamples of whites. See text for more details.

Table 2
Non-Linear Decompositions of Black/White Gaps in Home Computer Rates
Linear Probabilty Model and Probit Estimates, and Reverse Ordering of Variables

	Specification			
	(1)	(2)	(3)	(4)
Modification to decomposition	All Races Pooled	Linear Model	Probit Model	Reverse Order
White computer ownership rate	0.7286	0.7286	0.7286	0.7286
Black computer ownership rate	0.4257	0.4257	0.4257	0.4257
Black/White gap	0.3030	0.3030	0.3030	0.3030
Contributions from racial differences in:				
Sex and age	-0.0002	-0.0008	-0.0002	-0.0003
	(0.0002)	(0.0003)	(0.0002)	(0.0003)
	-0.1%	-0.3%	-0.1%	-0.1%
Marital status and children	0.0289	0.0225	0.0279	0.0175
	(0.0014)	(0.0012)	(0.0014)	(0.0015)
	9.5%	7.4%	9.2%	5.8%
Education	0.0369	0.0386	0.0366	0.0420
	(0.0008)	(0.0008)	(0.0008)	(0.0009)
	12.2%	12.7%	12.1%	13.9%
Income	0.0799	0.0823	0.0806	0.0855
	(0.0017)	(0.0015)	(0.0017)	(0.0017)
	26.4%	27.2%	26.6%	28.2%
Region	0.0085	0.0078	0.0082	0.0090
	(0.0015)	(0.0013)	(0.0014)	(0.0013)
	2.8%	2.6%	2.7%	3.0%
Central city status	-0.0019	-0.0016	-0.0019	-0.0017
	(0.0017)	(0.0014)	(0.0016)	(0.0015)
	-0.6%	-0.5%	-0.6%	-0.6%
All included variables	0.1519	0.1488	0.1512	0.1519
	50.2%	49.1%	49.9%	50.2%

Notes: (1) The sample consists of adults ages 25-55. (2) All specifications use coefficient estimates from the full sample of all races. (3) Standard errors are reported in parantheses below contribution estimates. (4) Contribution estimates are mean values of the decomposition using 1000 subsamples of whites. See text for more details.

Appendix

Logit Regressions for Probability of Having a Home Computer

	Specification			
Explanatory Variables	(1)	(2)	(3)	(4)
Sample	White	Black	White and Black	All Races
Female	0.0144	-0.0341	0.0050	0.0083
	(0.0276)	(0.0749)	(0.0258)	(0.0235)
Age	-0.0032	-0.0017	-0.0034	-0.0016
	(0.0018)	(0.0048)	(0.0017)	(0.0015)
Married	0.4799	0.5545	0.5032	0.4648
	(0.0424)	(0.0945)	(0.0383)	(0.0346)
Previously married	0.0172	0.0906	0.0408	0.0383
	(0.0474)	(0.1067)	(0.0430)	(0.0397)
Number of children	0.0455	-0.0522	0.0241	0.0077
	(0.0213)	(0.0478)	(0.0192)	(0.0166)
Children ages 6 to 17	0.7425	0.8049	0.7527	0.7410
	(0.0464)	(0.1133)	(0.0426)	(0.0380)
High school graduate	0.6545	0.4017	0.6098	0.7203
	(0.0535)	(0.1253)	(0.0492)	(0.0410)
Some college	1.2688	1.0271	1.2240	1.3490
	(0.0554)	(0.1285)	(0.0508)	(0.0426)
College graduate	1.6291	1.5359	1.6046	1.7265
	(0.0609)	(0.1555)	(0.0565)	(0.0485)
Graduate degree	1.9702	1.8084	1.9440	2.0742
	(0.0778)	(0.2105)	(0.0727)	(0.0642)
Family Income: $10,000 to $15,000	0.3497	-0.0632	0.2688	0.3230
	(0.0911)	(0.1840)	(0.0808)	(0.0710)
Family Income: $15,000 to $20,000	0.5734	0.2783	0.5166	0.4912
	(0.0892)	(0.1790)	(0.0792)	(0.0702)
Family Income: $20,000 to $25,000	0.4808	0.1933	0.4368	0.4576
	(0.0828)	(0.1751)	(0.0741)	(0.0662)
Family Income: $25,000 to $30,000	0.7448	0.5826	0.7208	0.7352
	(0.0808)	(0.1641)	(0.0720)	(0.0643)
Family Income: $30,000 to $35,000	1.0453	0.7098	0.9967	0.9843
	(0.0797)	(0.1701)	(0.0714)	(0.0638)
Family Income: $35,000 to $40,000	1.0767	1.0635	1.0736	1.0859
	(0.0805)	(0.1740)	(0.0724)	(0.0651)

(continued)

Appendix (continued)
Logit Regressions for Probability of Having a Home Computer

	Specification			
Explanatory Variables	(1)	(2)	(3)	(4)
Family Income: $40,000 to $50,000	1.3234	1.2009	1.3030	1.3644
	(0.0760)	(0.1598)	(0.0680)	(0.0613)
Family Income: $50,000 to $60,000	1.4889	1.2622	1.4597	1.4917
	(0.0774)	(0.1654)	(0.0694)	(0.0625)
Family Income: $60,000 to $75,000	1.6876	1.8948	1.7030	1.7088
	(0.0781)	(0.1763)	(0.0705)	(0.0637)
Family Income more than $75,000	2.1842	2.5912	2.2102	2.2049
	(0.0758)	(0.1776)	(0.0684)	(0.0615)
In MSA but not in central city	0.0526	0.0930	0.0662	0.0319
	(0.0409)	(0.0881)	(0.0363)	(0.0317)
Rural area	-0.1121	-0.1913	-0.1014	-0.1159
	(0.0439)	(0.1329)	(0.0403)	(0.0367)
Central city status not identified	0.0172	0.2772	0.0466	0.0425
	(0.0484)	(0.1166)	(0.0440)	(0.0399)
Region Controls	Yes	Yes	Yes	Yes
Race/Ethnicity Controls	No	No	Yes	Yes
Mean of Dependent Variable	0.7286	0.4257	0.6932	0.6590
Sample Size	34,386	4,555	38,941	46,322

Notes: (1) The sample consists of adults ages 25-55 from the specified racial group/s. (2) Standard errors are in parentheses below coefficient estimates.