MEASURING THE EFFECT OF SEXUAL ORIENTATION ON INCOME: EVIDENCE OF DISCRIMINATION?

NATHAN BERG and DONALD LIEN*

The effect of nonheterosexuality on individual income is estimated using 1991–1996 General Social Survey data. Other researchers have concluded that homosexuals earn less than similarly qualified workers, in contrast to the popular perception that homosexuals are more affluent than nonhomosexuals. Using improved statistical techniques, this article finds noticeable earnings effects that go in opposite directions across genders. Nonheterosexual men earn 22% less than heterosexual men, and nonheterosexual women earn 30% more than heterosexual women. These findings, viewed together with previous empirical work on this topic, help narrow the field of theories that can explain the sexual-orientation earnings gaps present in the data. (*JEL* J78, J31, J11, C24)

I. INTRODUCTION

One of the obstacles to coming out is the fear that you're gonna lose your job.
—Representative Barney Frank (D-MA)[1]

Legislative initiatives to outlaw workplace discrimination based on sexual orientation demonstrate that there are policy makers, activists, and voters who perceive such discrimination to be a significant problem in the United States. After all, if proponents of those initiatives did not perceive workplace discrimination to be a significant problem, they presumably would not spend their resources attempting to enact legal sanctions against it.[2] Among opponents, it is important to distinguish two very different attitudes toward discrimination. On one hand, some opponents actually support discrimination. For example, opponents who object to homosexuality based on moral or religious principles probably would approve of managers discriminating against nonheterosexuals. For these opponents, the question of whether or not homosexuals face discrimination is not a key point of contention. On the other hand, there is a larger bloc that opposes antidiscrimination laws based on the claim that workplace discrimination against those with same-sex sexual orientation is not a

*This is a revised version of a paper presented at the Midwest Econometrics Group meeting, March 2000, in Chicago, Illinois, as part of a session organized by Mark Jensen, Brigham Young University. The authors are grateful to Lawrence Marsh, Joseph Sicilian, Ted Juhl, Joshua Rosenbloom, Robert Michaels, and three anonymous referees for their comments and suggestions.

Berg: Cecil and Ida Green Assistant Professor of Economics, University of Texas at Dallas, GR 31 211300, Box 830688, Richardson, TX 75083. Phone 1-972-883-2088, Fax 1-972-883-2735, E-mail nberg@utdallas.edu

Lien: Professor, Department of Economics, University of Texas at San Antonio, San Antonio, TX 78249. Phone 1-210-458-7312, Fax 1-210-458-5837, E-mail dlien@utsa.edu

1. This quote comes from a 1994 interview with Rep. Frank about the Employment Non-Discrimination Act, a proposed extension of the 1964 Civil Rights Act to cover discrimination based on sexual orientation. A transcript of the interview is posted online at http://members.aol.com/barneyenda/transcri.html.

2. Other motives for supporting antidiscrimination policies are, of course, possible. For instance, those in any protected class may enjoy better-than-average treatment from firms that fear lawsuits. Thus, supporting new protections may be interpreted (cynically) as a strategic move intended to procure rents induced by asymmetries in the legal code. This seems unlikely, however, because the returns from exploiting new laws—consisting of tort case payouts, easier promotion, and softer handling by management—seem very small relative to the resources that activists have poured into numerous political battles involving this issue since the 1970s.

ABBREVIATIONS

ENDA: Employment Non-Discrimination Act
GSS: General Social Survey
MLE: Maximum Likelihood Estimation
OLS: Ordinary Least Squares

Contemporary Economic Policy
(ISSN 1074-3529)
Vol. 20, No. 4, October 2002, 394–414

significant problem. For these opponents, the question of whether or not there exists widespread discrimination at U.S. firms is important. Resting on the belief that there exists little or no discrimination, these opponents feel that the costs of enacting new laws are not justified by the benefits.

Among the first legal initiatives aimed at protecting gay workers was the Employment Non-Discrimination Act (ENDA). Proposed in the mid-1970s and defeated in the U.S. Congress, it would have extended existing civil rights protection (against discrimination based on race, religion, gender, national origin, and disability) to cover cases of discrimination based on sexual orientation.[3]

ENDA reappeared in various forms after the 1970s, including three different legislative proposals in 1994, 1995, and 1996, all of which were defeated. President Clinton supported ENDA, commending it in 1995 as a safeguard helping guarantee that "all Americans, regardless of their sexual orientation, can find and keep their jobs based on their ability to work and the quality of their work."[4] A rather long list of *Fortune* 500 firms also supported ENDA, as did a variety of religious groups.

After failing at the national level, laws similar to ENDA eventually succeeded in passing at the state level and currently stand in 10 states as well as the District of Columbia.[5] In addition, some 165 cities and municipalities have local antidiscrimination laws on the books concerning sexual orientation. These state and local rules are, however, rather limited in scope compared to the national proposals. For example, some of the state versions of ENDA cover only government jobs. Throughout most of the 40 U.S. states without laws explicitly prohibiting employers from treating homosexuals differently from heterosexuals, it is in fact legal to differentially compensate, fire, harass, not hire, or not promote a worker because of sexual orientation.[6]

At the core of this protracted policy battle is, among other issues, a factual dispute about the relative affluence of homosexuals. Opponents of ENDA who hold that antihomosexual discrimination is not a significant problem in American workplaces often point to the "affluence of homosexuals" as evidence that there is no discrimination and that such laws are not needed.[7] Apparently, it is a well-established stereotype, held by many marketing and advertising sales professionals, that homosexuals are unusually affluent.[8]

Thus, there is an important empirical question of earnings differentials based on sexual orientation. The primary purpose of this article is to uncover the statistical relationship between sexual orientation and income and interpret its meaning.

In identifying the statistical effect of sexual orientation on income, we aim to clarify this factual dispute and provide a richer empirical backdrop for further analysis of wage gaps and antidiscrimination policy. With the statistical facts of sexual orientation and income in hand, we attempt to discard those theories of income determination and discrimination that do not match the facts. By this process of elimination, knowing merely the marginal effect of regressors on the probability distribution of individual income can illuminate economic structure, providing evidence important for the analysis of policy.

II. PAST EVIDENCE ON THE EARNINGS OF HOMOSEXUALS

There are two main empirical studies of sexual orientation and earnings to date.

3. This law would have exempted the armed forces and firms with fewer than 15 employees. Religious nonprofits were also exempted in some versions.

4. The quote from President Clinton and a list of corporate sponsors that support ENDA is posted online at www.religious tolerance.org/hom_empl.htm.

5. The American Civil Liberties Union Web site posts a list of the states and municipalities with laws prohibiting discrimination based on sexual orientation at www.aclu.org/issues/gay/gaylaws.html.

6. This point is made on the Web site of the organization Religious Tolerance, found online at www.religioustolerance.org/hom_empl.htm.

7. This logic is, of course, faulty. If homosexuals disproportionately work in high-paying fields or are more productive for some reason, then uncontrolled average income may be higher for homosexuals even though they are discriminated against and paid less than heterosexual workers with similar characteristics. The real question is whether, after controlling for all determinants of income, homosexuals earn more or less than similarly productive workers.

8. The *Wall Street Journal* article "Are Gay People More Affluent than Others?" (Alsop, 1999) reports on a divergence of opinion between marketing experts and civil rights activists on the question of affluence. One activist explicitly refutes the claim that homosexuals tend to be affluent, stating that "gay wealth is a myth."

The first attempts to estimate a single wage equation in which an individual's expected wage depends on sexual orientation. The second seeks to measure the effect of enacted antidiscrimination laws on the earnings of same-sex couples.

Using 1989–91 General Social Survey (GSS) data, Badgett (1995) estimates a wage regression that includes a dummy variable for sexual orientation. She finds the estimated effect of homosexuality on earnings to be negative (using a variety of model specifications and definitions of the "homosexuality" dummy) and argues that this constitutes evidence of discrimination against homosexuals. The earnings effect is most dramatic for homosexual men, who earned 11% to 27% less than heterosexual men. Badgett appears to be the first economist to use a sequence of questions from the GSS survey regarding the gender of past sexual partners to study this issue.

Klawitter and Flatt (1998) use 1990 census data (public use micro-data sample) to analyze the impact of city and state antidiscrimination laws (similar to ENDA) on the earnings of couples. Their analysis attempts to identify distinct effects for three specific types of couples: married (heterosexual), unmarried heterosexual, and same-sex. Comparing these effects, Klawitter and Flatt find little evidence that the enactment of antidiscrimination laws affects the earnings of same-sex couples. However, they report another interesting set of statistics that point to a possible disparity between comparisons of individual versus household income among homosexuals and nonhomosexuals. As couples, homosexual men appear to do better than married heterosexual couples; married heterosexual couples do better than lesbian couples; and unmarried heterosexual couples fare the worst:[9]

$$\begin{array}{c}\text{male same-sex}\\\text{couples}\end{array} > \begin{array}{c}\text{(heterosexual) married}\\\text{couples}\end{array} > \begin{array}{c}\text{female same-sex}\\\text{couples}\end{array} > \begin{array}{c}\text{heterosexual unmarried}\\\text{couples}\end{array}$$

As individuals (rather than as couples), married heterosexual men do better than both homosexual men and unmarried men, with earnings ordered as follows:

$$\begin{array}{c}\text{married (heterosexual)}\\\text{individual men}\end{array} > \begin{array}{c}\text{unmarried heterosexual}\\\text{individual men}\end{array} > \begin{array}{c}\text{same-sex household}\\\text{individual men}\end{array}$$

In contrast to individual men, homosexual women do considerably better than other women, as individual earners:

$$\begin{array}{c}\text{same-sex household}\\\text{individual women}\end{array} > \begin{array}{c}\text{unmarried heterosexual}\\\text{individual women}\end{array} > \begin{array}{c}\text{married (heterosexual)}\\\text{individual women}\end{array}$$

This finding is suggestive of the possibility that differential anticipated household income (connected to the gender of one's partner) may lead otherwise identical workers to make different labor/leisure trade-offs, a point we return to later.

Building on these previous studies, this article aims to achieve two primary goals. First, it attempts to solidify our understanding of the statistical facts regarding the comovement of individual earnings and homosexuality by improving the econometric techniques used to measure the effect of sexual orientation on income. The maximum-likelihood estimation of a 21-category discrete-dependent-variables model presented herein, designed specifically to handle income data that is censored into income brackets, improves on a number of methodological drawbacks inherent in Badgett's approach and, in fact, overturns Badgett's finding on the direction of the earnings effect for homosexual women. The broad intent of introducing methodological improvements is, of course, to aid in settling the question—rich with policy implications—of whether sexual orientation has anything to do with how much typical workers earn. Because improved methodology in this case delivers distinct answers regarding the direction and significance of earnings effects, this article discusses the key econometric issues that, in this case anyway, clearly make a difference.

The second primary contribution of this article is theoretical. Unlike the earlier studies cited, which drew structural conclusions from statistical relationships without reference to an underlying economic model, this study attempts to tie the available empirical evidence to a short list of economic models, each of which specifies a mechanism that

9. These inequalities hold even after controlling for worker and location characteristics, according to Klawitter and Flatt (1998). However, their article reports only raw income averages with no controls.

may account for the observed wage gaps. Using improved measurements of individual earnings effects, together with Klawitter and Flatt's data on the earnings of couples, this study successfully eliminates several different explanations for the puzzling pattern of income inequalities based on sexual orientation visible in various sets of data.

III. THE DATA

The data under consideration are pooled 1991–96 responses from the GSS (Davis and Smith, 1996). That survey asks respondents the number and gender of their sexual partners. Although there are a variety of reasonable formulations of a *sexual orientation* variable based on the GSS survey responses, this article uses the broadest criterion available in the GSS.[10]

Same-sex sexual orientation is indicated by a dummy variable, *homosexual*, which is one if a respondent reports having at least one same-sex sexual partner in the last five years.[11] Although it may seem imprecise to some, we make no distinction among the terms *same-sex sexual orientation*, *homosexual*, *non-heterosexual*, and *gay*.

There is an important caveat regarding any attempt to link the *homosexual* variable to workplace discrimination, namely, that reported sexual orientation is almost assuredly not identical to managers' beliefs about the sexual orientation of workers. For the sake of measuring statistical regularities that pertain to discrimination, one would ideally want some measure of how managers actually perceive the sexual orientation of employees. Interpreting differential expected earnings based on sexual orientation as the result of managers who discriminate against homosexuals depends on a strong connection between the behavior reported in the survey and public perceptions of individual sexual orientation at work. Discrimination is only one among several scenarios considered herein, however.

Untruthful survey responses are another potential problem that must be acknowledged. It seems likely that some homosexuals might conceal their sexual orientation and that some heterosexuals might strategically misreport, perhaps to support homosexuals by making them appear more numerous. Both instances of untruthfulness, if uniformly distributed over all kinds of earners, would tend to bring the heterosexual and homosexual group averages closer together, biasing a test to detect differences against finding any.

There is a problem, however, if the truthfulness of survey responses is correlated with income. In this case, say, if only low-earning homosexuals were truthful, one might observe a spurious earnings gap. A priori, the sign of the correlation between income and the propensity to reveal one's sexual identity is not clear. Would being richer than average make one more secure in revealing homosexuality, or would it lead to concealment because there is more to lose? Hoping that the survey responses correspond closely to the truth and that the net effect of any dishonest responses is negligible, the authors interpreted the *homosexual* variable at face value.

We also invoke the standard missing-at-random assumption to justify discarding incomplete survey responses in which a respondent reports a mix of valid and invalid responses. If an individual's decision to enter an invalid response is correlated with other variables of interest, then estimation using only the valid responses would not be representative of the population. In such a case, techniques that explicitly adjust for the correlation between characteristics and missing responses should be considered. In the case of the GSS data, however, implementing such an adjustment is far from straightforward.

To begin with, the response rate for the GSS from 1991–96 ranges between 76% and 82%. Among respondents who agree to participate in the survey, the raw GSS data contain a considerable number of responses that are explicitly coded as "don't know" or "no response." In addition to these officially invalid responses that follow from the decision or utterance of survey respondents themselves, there are still other response

10. Badgett (1995) works with a variety of definitions for the variable *sexual orientation*. Her broadest definition is the same as ours. But more narrow definitions are possible by requiring the same-sex sexual activity to have occurred more recently or by eliminating bisexuals from the homosexual category.

11. Some authors discuss the question of whether sexual orientation is an objective classification, verifiable (in principle) via observed behavior, or whether it is a component of identity that resides on the interior of individual consciousness. Clearly, a behavioral definition applies to this article's analysis.

TABLE 1
Sample Sizes When Including or Excluding Non–Full-Time Workers

	All Men	All Women	Gay Men	Gay Women	Straight Men	Straight Women	Men and Women
Full-time and non–full-time	1719	1657	70	61	1649	1596	3376
Full-time only	1577	1310	64	52	1513	1258	2887
Non–full-time only	142	347	6	9	136	338	489
Percentage non–full-time	8	21	9	15	8	21	14

items that are missing entirely, because some questions were never asked of some respondents. Missing data of this variety most likely reflect a choice of the interviewer or survey designer rather than a decision of the survey respondent, or perhaps such cases result from exogenous mishaps during the data-collection process. Either way, it is probably safe to assume that these items, which are missing because the corresponding survey question was never asked, are indeed missing at random.

When we discard all respondents who were not asked one of the questions used to generate our variables, the pooled 1991–96 GSS sample size (of full-time workers) drops to 3061 from 4927. Among the remaining 3061, 108 refused to answer the question about the gender of sexual partners, and 45 refused to reveal how much they earn. Along with all other respondents who registered an invalid response to 1 of the 13 raw variables used in this study, 174 respondents among the 3061 are lost, leaving a sample of 2887 respondents.

Another important feature of the data is that the sample used includes full-time workers only.[12] Among those excluded are part-time workers, retired workers, unemployed workers, homemakers, and students. Table 1 breaks out the number of part-time workers by gender and sexual orientation. Excluding part-time workers implies discarding just over 20% of female observations and around 8% of male observations. Thus, deciding whether or not to include part-time workers in the estimation of an earnings equations could significantly affect the results. In fact, regression results including part-time workers reveal no qualitative differences.

So what is the justification for discarding non–full-time workers? We choose to study the population of full-time workers to sharpen the data's capacity to focus on the earnings differential associated with a relatively rare type, that is, homosexuals. We want the most stable population possible so that the sexual orientation effect, if any, may speak as clearly as possible. We feel that a wide variety of unobservable factors unrelated to sexual orientation underlie the labor supply decisions of non–full-time workers. We also suspect that employers treat part-time workers in a manner substantially different from full-time workers, again, for reasons unrelated to sexual orientation—and this further clouds any attempt to measure an earnings effect using a sample of pooled full- and part-time workers. We believe that the noise introduced by including part-time workers would cost more, in terms of precision, than would be gained by their inclusion, thus motivating the decision to exclude them.

IV. CONTROL VARIABLES

To isolate the effect of sexual orientation on earnings, one hopes to appropriately control for all other factors that affect expected earnings and are correlated with sexual orientation. These other factors include variables on years of education, potential experience (age − years of education − 5), the square of experience (to reflect arc-shaped earnings profiles as a function of workers' experience), dummies indicating the highest academic degree attained by an individual,

12. The sample includes those who reported they were currently employed full-time when interviewed by GSS examiners. However, the measure of individual income is actually previous-year income. Therefore, the income distribution of full-time workers includes some whose incomes are clearly below the incomes of most full-time workers. The presence of nonlabor income in this income measure further complicates the interpretation of the income variable. However, nonlabor income should be negligible for most workers, just as previous-year income should be strongly correlated with current income.

TABLE 2
Regressor Means, Pooled and Broken Out by Sexual Orientation

Variable	All Men	Homosexual Men	Nonhomosexual Men	All Women	Homosexual Women	Nonhomosexual Women
homosexual	0.04	1.00	0.00	0.04	1.00	0.00
white	0.86	0.77	0.87	0.83	0.87	0.82
highschool	0.53	0.47	0.53	0.54	0.46	0.54
junior college	0.08	0.06	0.08	0.08	0.08	0.08
college	0.20	0.22	0.20	0.23	0.17	0.23
graduate	0.10	0.17	0.10	0.08	0.15	0.08
experience	20.53	19.17	20.59	19.94	18.13	20.02
experience^2	544.52	473.30	547.53	516.18	468.60	518.15
union	0.17	0.20	0.17	0.16	0.21	0.16
executive	0.15	0.20	0.15	0.12	0.10	0.12
specialist	0.14	0.28	0.14	0.20	0.19	0.21
low-skill	0.28	0.19	0.28	0.22	0.37	0.22
newengland	0.05	0.03	0.05	0.05	0.04	0.05
pacific	0.15	0.28	0.15	0.16	0.23	0.15
south	0.07	0.02	0.07	0.07	0.08	0.07
urban	0.10	0.30	0.09	0.13	0.10	0.13
# of observations	1577	64	1513	1310	52	1258

Notes: The variable *experience* is [age − (years of education) − 5]. The variable *experience*^{2} is the square of [age − (years of education) − 5]. The variable *urban* is a dummy indicating that an individual resides in a metropolitan area with a population greater than 500,000. The academic degree variables indicate an individual's highest degree attained. For instance, the *college* entry under the heading All Men indicates that 20% of men finished college and did not go on to obtain a graduate degree. The total percentage of male college graduates is calculated by summing the *college* and *graduate* entries, just as the total percentage of high school graduates is calculated by summing over the percentages associated with each of the four academic degree variables. The regional variables were intentionally defined in a restrictive way so that the reference group would be a regionless middle American composite similar to but not limited to the Midwest. This in part accounts for the low frequencies associated with the variables labeled as *newengland* and *south*. The occupational category variables are described more fully in the body of the text.

as well as region of residence, urban versus nonurban status, and occupational category.[13]

The regressor means, broken out by gender and sexual orientation are presented in Table 2. Relative to average heterosexual survey respondents, homosexual men are less often white, less likely to work low-skill jobs or to live in the South, and more likely to hold a graduate degree. Relative to straight females, homosexual females are more likely to work low-skill jobs. They have both more graduate degrees and more high school dropouts per capita than do heterosexual females, indicated by summing the percentages holding each of the four (terminal) academic degree types and subtracting from 100. Like male homosexuals, female homosexuals reside disproportionately in the Pacific region, which includes northern California, Oregon, and the state of Washington. Unlike homosexual men, who appear to cluster in large urban areas, homosexual women exhibit no such tendency.

A potentially important difference between homosexuals and nonhomosexuals is occupational category. As evident from Table 2, there are three special classes of occupations we attempt to control for. The construction of these categories from the raw GSS data is not straightforward and warrants some explanation. The GSS variable *occ80* is coded to correspond to definitions of job types from the 1980 census, of which there are nearly 1000 distinct job types possible. From this, it is possible to partition the numerous job types into four broad categories: executives (high-level managers

13. A referee pointed out that the gap between potential experience and actual experience might be systematically larger for homosexuals. If heterosexual men and homosexual women both experience fewer interruptions in their professional careers, then the homosexual variable may be picking up differences in actual experience. This could lead to a spurious pattern by which gay men and straight women appear to earn significantly less because of their sexual orientation but are actually paid less because they have less on-the-job experience. This explanation is intriguing and clearly worth pursuing with a data set that has better information about workers' career history, although it seems to depend on a delicate combination of assumptions about who is more likely to interrupt their careers.

and those who have risen to the tops of organizations), specialists (e.g., artisans, artists, and other highly autonomous jobs that are not typically attached to a firm), low-skill workers, and "everyone else."

One can, of course, imagine many alternative schemes for partitioning the numerous job types from the census definitions into a manageable list of broader categories. The overarching goal in constructing variables on occupational category is to control for the different types of workplace experiences that might lead homosexuals to choose certain lines of work over others.

The occupation variable in this article is constructed to control for differences in life circumstances and preferences that account for individual decisions to accept careers in low-skill jobs or, at the other end of the spectrum, to aggressively pursue income by becoming an executive or high-level manager. This view reflects the concern that without these controls, one could be misled by unknown patterns of correlation between homosexuality and these broad differences in approach to work life. The executive and low-skill variables proxy for underlying variation in preferences and for unobserved life circumstances that influence individual career aspirations.

A second goal motivating the construction of the occupation variable is to control for workplace autonomy in some way, in order to study the possibility that persecuted individuals may disproportionately choose certain (relatively autonomous or homosexual-friendly) occupations. The specialist variable proxies for variation in individual preferences for autonomy and facilitates the investigation of whether homosexuals have an unusual propensity to avoid working under direct managerial oversight.

The entries in the row corresponding to the specialist variable in Table 2 suggest that male homosexuals have a strong propensity to seek out specialist positions. Male homosexuals also hold executive positions more frequently and work low-skill jobs less often than heterosexuals do. It is conjectured that both occupational categories toward which male homosexuals gravitate—specialist and executive—would afford protection against discrimination by managers in workplaces offering less autonomy. This intriguing pattern in regressor means suggests the possibility that income and occupational choice feed back and affect one another.

To pursue this possibility, we estimate a two-equation model in which occupational choice and income are simultaneously determined (see Appendix A for details). The two dependent variables are income category and the decision to become a specialist. Although it is straightforward to explain why being a specialist affects a worker's earnings, the question at issue is whether a worker's expected income affects the decision to become a specialist. If so, then one of the key control variables is endogenous and the single-equation approach must be modified. As reported in Appendix A, the data support the hypothesis that the decision to become a specialist is independent of income. This in turn justifies including *specialist* as an independent variable and relying on it as a control for variation in preferences regarding workplace autonomy.

V. INCOME BRACKET EARNINGS DATA AND DISCRETE DEPENDENT VARIABLE METHODOLOGY

A particular challenge in using the GSS data is the manner in which income is recorded. Respondents are asked to select one of 21 ranges into which their individual pretax income falls. The percentage-wise income data, broken down by gender and sexual orientation, is presented in Table 3.

The GSS income variable simply indicates the income bracket into which each worker falls but provides no additional information on the actual level of income within each bracket. Thus, one faces the challenge of estimating the effect of sexual orientation on income without ever directly observing income. A common approach to working with income-bracket data is to impute an income level for each individual, often the midpoint of each interior bracket.

There are at least three drawbacks to this methodological choice. First, the imputed income estimator is inconsistent (Hsiao, 1983; Stewart, 1983). In particular, all the slope parameters depend in a nontrivial manner on the arbitrary choice of imputed income for the unbounded brackets—the income levels assigned to respondents in the highest and lowest income categories. Inconsistency means that this sensitivity to the

TABLE 3
Income Data—Percentage of Responses in Category $j \in \{1, 2, \ldots, 21\}$

Category Number (j)	Income Thresholds	Homosexual Male	Heterosexual Male	Homosexual Female	Heterosexual Female
1	<$1000	0	0	0	1
2	$1000–2999	2	0	2	1
3	$3000–3999	0	1	2	2
4	$4000–4999	5	1	0	2
5	$5000–5999	0	1	0	2
6	$6000–6999	0	1	2	1
7	$7000–7999	0	1	0	1
8	$8000–9999	0	2	4	4
9	$10,000–12,499	9	4	6	7
10	$12,500–14,999	2	4	12	7
11	$15,000–17,499	3	5	10	10
12	$17,500–19,999	6	5	0	8
13	$20,000–22,499	9	6	6	8
14	$22,500–24,999	9	7	4	8
15	$25,000–29,999	13	11	13	12
16	$30,000–34,999	5	12	10	10
17	$35,000–39,999	14	9	8	5
18	$40,000–49,999	8	11	12	6
19	$50,000–59,999	9	8	2	3
20	$60,000–74,999	5	5	6	1
21	$75,000+	2	7	4	1

Notes: The GSS records pretax earnings for the year prior to the year in which the interview is conducted, and it asks specifically for earnings in the same occupation that the respondent currently has. The sample tabulated here includes only those respondents who say they are employed full-time at the time of the interview. Full-time workers who report incomes below the annual full-time minimum wage level may have switched occupations or just recently (re-)joined the labor force. Since these data were pooled over a six-year span in which the U.S. urban consumer price index increased 15%, inflationary effects might also be cause for concern. One can argue, however, that this is a relatively minor problem in that the marginal effect of sexual orientation on income-bracket probabilities should be relatively stable for the six years under study as long as the homosexual responses are evenly distributed through time, as they appear to be.

analyst's imputation of representative income values persists no matter how large a sample is drawn. Second, the standard errors overstate the precision of estimation when using imputation because within-bracket variation is suppressed and because the error from imputation is not taken into account. Thus, the effects of regressors can appear significant when they are not. A third problem with imputed income estimation in a discrete dependent variable setting is that logical errors, such as negative probabilities or probabilities not summing to one, can arise due to the imposition of a linear relationship between regressors and income categories, for example, when applying ordinary least squares (OLS).

Badgett (1995) uses a sophisticated imputation technique that restores some within-bracket income variation by linking a worker's occupation category to the average income for that category in a second data set. Although this may be better than the conventional midpoint imputation technique, it still suffers from the disadvantages mentioned. To investigate whether these methodological issues actually matter or not in measuring the effect of sexual orientation on income, this article pursues a different measurement approach—one that fits into the maximum-likelihood framework so well accepted by econometricians. In particular, we specify a probability model in which the probability that an individual belongs to a given income bracket is a function of that individual's characteristics. Although this is close in spirit to the so-called ordered probit or multinomial logit models, there is an important difference. In those standard probability models, the thresholds that define the categories (the income levels that define the income brackets) are estimated along with the parameters of interest, that is, the slope parameters which appear as coefficients on the regressors. In the GSS, however, the income levels that define income brackets are known. They are given as part of the experimental design; it is, of course, foolish to ask the data to estimate what is already known. As a practical matter, implementing an ordered probit model would increase the number of parameters to be estimated from 18 in our model to 38, seriously compromising the precision of the estimates.

Using actual income-bracket thresholds turns out to introduce new complications. The primary complication stems from the variance of the error term, which is not identified in the probit and logit models and is therefore set equal to one. In contrast, this parameter is identified when using actual income-bracket thresholds, and therefore must be estimated together with the slope parameters. This can lead to unstable performance of the numerical algorithms used to maximize the likelihood function. But fortunately, this instability can be overcome with the aid of an EM-type algorithm from Stewart (1983). This is the approach taken herein, yielding maximum-likelihood estimates of the sexual orientation effect, estimates that enjoy the nice asymptotic features of maximum likelihood estimators. Of course, these methodological issues become interesting only if they lead to substantively different conclusions. In the present context, the methodology does appear to make a difference.

VI. ESTIMATION PROCEDURE

The model of individual income determination to be estimated is

$$(1) \qquad y_i^* = \beta' x_i + \delta d_i + \epsilon_i,$$

where y_i^* is the natural logarithm of income, x_i is a vector of individual characteristics (including a constant), and d_i is an indicator for nonheterosexuality. Theoretical justifications as to why sexual orientation might determine income are discussed in a subsequent section. At the very least, the model facilitates a statistical test of whether sexual orientation helps predict income.

As discussed already, the income variable y_i^* is not continuously observed. Instead, because the GSS asks respondents to report income brackets, the observed income variable y_i is discrete, defined as

$$(2) \qquad y_i = j, \quad \text{if } a_{j-1} < y_i^* < a_j, \quad j \in \{1, 2, \ldots, 21\}.$$

The income-bracket thresholds $(a_j)_{j=1}^J$ are given as part of the GSS experimental design and partition the real line: $-\infty = a_0 < a_1 < \cdots < a_J = \infty$.

Instead of imputing a continuous-valued income variable and proceeding with OLS, a probability model is developed for the purpose of consistently estimating β using the censored income data available in the GSS. Equation (1) is estimated twice, when ϵ_i is distributed standard normal and again when ϵ_i is logistic, attempting to see how robust the estimates are across those specifications.[14] To get a feel for how much is gained, relative to OLS, by using maximum-likelihood estimation, the predicted frequency distribution of income categories from those two estimation techniques is compared with the empirical distribution. Appendix B contains a detailed description of the estimation procedure.

VII. ESTIMATION RESULTS

Referring to the estimated coefficients listed in Table 4, male homosexuals' expected income is significantly lower than heterosexuals'. Holding all other regressor values constant, the model estimates that nonheterosexual men earn 22% less than heterosexual men. To see this, recall that y^* is the natural log of income and is distributed $N(\beta' x + \delta d, \sigma^2)$, where, as before, $d = 1$ indicates nonheterosexual status. Define $Y^* = \exp(y^*)$ to be the level of income, meaning that Y^* is log normal. Then, the percentage change in expected income for homosexual men can be computed as

$$(3) \qquad \frac{E[Y^*|x, d=0] - E[Y^*|x, d=1]}{E[Y^*|x, d=0]} = \frac{e^{(\beta' x + \frac{\sigma^2}{2})} - e^{(\beta' x + \delta + \frac{\sigma^2}{2})}}{e^{(\beta' x + \frac{\sigma^2}{2})}} = 1 - e^{\delta}.$$

14. The estimated effects on earnings in the logit model are qualitatively similar to the probit results, as is often the case when working with the standard versions of the logit and probit models. Therefore, only probit model estimates are reported in the remainder of this article.

TABLE 4
Probit Estimates for Men

Variable	Estimated Coefficient	$\|t\|$ Hessian	$\|t\|$ Outer Product	$\|t\|$ Sandwich
const	8.99	105.61	111.24	90.47
homosexual	−0.25	3.30	3.69	2.91
white	0.06	1.33	1.36	1.27
highschool	0.33	5.82	6.46	4.94
juniorcollege	0.51	6.60	5.89	6.54
college	0.73	11.01	11.95	9.61
graddegree	0.92	11.43	12.05	10.36
experience	1.22	12.44	13.79	10.86
experience^2	−0.48	9.19	10.44	7.89
union	0.20	4.84	4.30	5.33
executive	0.18	3.70	3.52	3.82
specialist	0.07	1.25	1.17	1.30
lowskill	−0.16	4.34	4.40	4.17
newengland	0.15	2.14	1.84	2.49
pacific	0.03	0.66	0.71	0.59
south	−0.14	2.20	2.20	2.18
urban	0.16	3.20	2.84	3.49

Plugging in the estimate of δ, $\hat{\delta} = -0.2470$, one obtains an estimated expected percentage loss of 21.89%. Using the delta method, the asymptotic variance of $1 - e^{\hat{\delta}}$ is computed to be $e^{2\hat{\delta}}\widehat{\mathrm{Var}(\hat{\delta})}$. Substituting in the estimates $\hat{\delta}$ and $\widehat{\mathrm{Var}(\hat{\delta})}$ yields a standard error of 5.91 percentage points.

The sign pattern of the other variables conforms to our expectations. Being white, obtaining more education, being in a union, being an executive or specialist, and living in the New England region, the Pacific region, or in a city with more than 500,000 people increases expected income. Earnings are arc-shaped in the experience variable, increasing over low experience levels and decreasing over high experience levels; this is consonant with the idea that more experience helps only up to a point, past which depreciation of one's human capital dominates improvements in performance due to repetition.[15] Low-skill job status and residing in the South both decrease earnings relative to the Middle America reference group in this setup.

Three t-statistics are presented in Table 4 alongside the estimated slope parameters, computed using different estimates of the asymptotic variance matrix—the Hessian, outer product, and robust sandwich estimators.[16] Regardless of the variance matrix used, sexual orientation appears to be a significant component of income determination for men. As shown in Table 5, the likelihood ratio test soundly rejects the hypothesis that all coefficients except the constant are zero. Exponentiating the minimum and maximum estimated (log) income values yields very reasonable minimum and maximum predicted incomes of $8,842 and $91,620, respectively.

Figure 1 shows a histogram comparing the frequencies of probit, logit, and OLS predictions with the actual empirical distribution. It has been pointed out that this kind of frequency comparison can be misleading, because a model could make all wrong predictions at the individual level yet fit the frequency distribution perfectly. Figure 1 can nonetheless aid in making a qualitative assessment of model performance by comparing the shape of the predicted density with the actual empirical density function. The model's ability to correctly predict individual income categories is around 14%. That may not be such a bad performance when considering the precision required to correctly predict 21 discrete outcomes using a

15. Murphy and Welch (1990) challenge the quadratic specification of earnings in experience, suggesting instead that income is modeled better as a fourth-degree polynomial in experience. Including these two additional polynomial terms as regressors changes the estimates very little, however.

16. The robust sandwich estimator is the correct variance matrix even when the density F is misspecified. Its derivation and a discussion of some of its advantages are given in Cameron and Trivedi (1998).

TABLE 5
Summary Statistics (Probit Estimate for Men)

Percentage correctly predicted $100 \times \frac{\#\{\widehat{y_i^*} \in \text{category } y_i\}}{\text{sample size}}$	14.26
Expected percentage loss for homosexuals	22 ± 6
Log likelihood value	−4048.44
Likelihood ratio	$1448.24 > \chi(16)_{.005} = 34.27$
Minimum predicted income ($)	8,841.76
Maximum predicted income ($)	91,620.37
Estimated standard error (log units)	0.59
Sample size	1578

linear function. As the number of categories goes to infinity, the probability of any prediction rule getting a single correct prediction would of course be zero.

The predicted distribution in the probit case (Figure 1) nicely follows the shape of the empirical distribution. The logit predicts a mostly similar distribution of income. But the model estimated by OLS is noticeably worse at matching the empirical distribution and predicts with a lower rate of accuracy. In all cases, the predicted distributions miss what is going on at the tails, not predicting enough high incomes in particular.

Turning to the case of women, Table 6 indicates that homosexual women earn significantly more than heterosexual women. This stands in contrast to the (mostly) negative effects of homosexuality on income that Badgett found for both men and women. Aside from the different direction of the sexual orientation effect relative to men, the other signs are identical to the male case, encouraging us in thinking that we have a sensible

FIGURE 1
Men: Frequencies of Predicted Income

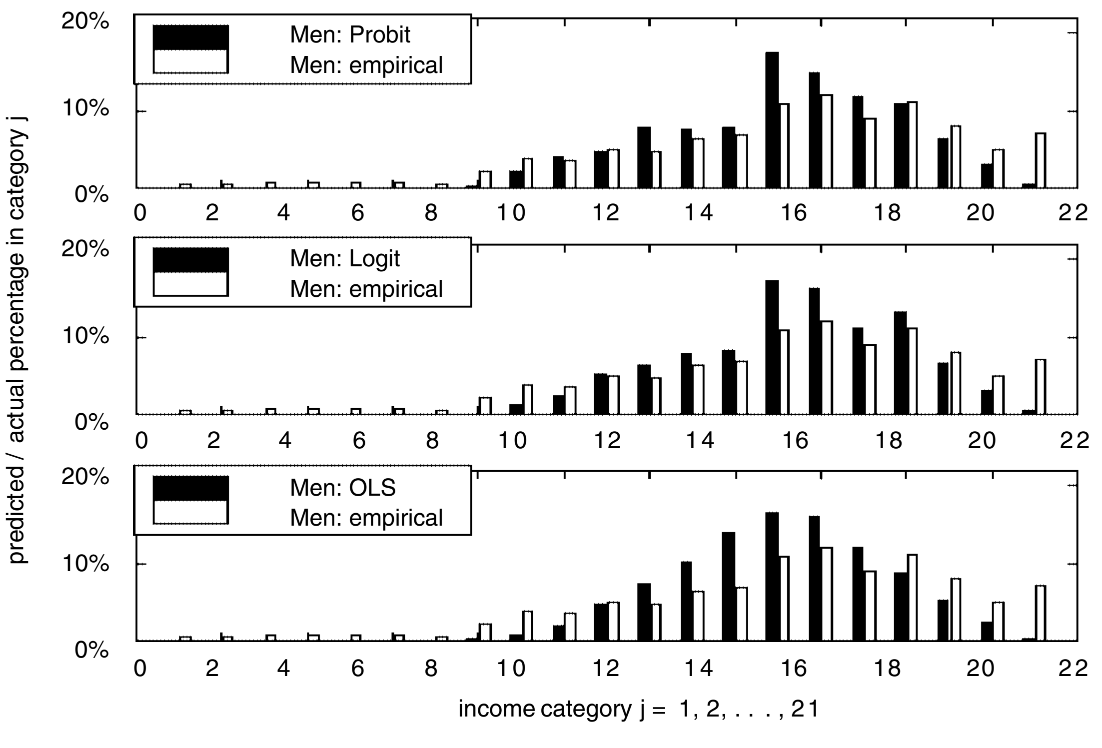


TABLE 6
Probit Estimates for Women

Variable	Estimated Coefficient	$\|t\|$ Hessian	$\|t\|$ Outer Product	$\|t\|$ Sandwich
const	8.82	86.36	106.86	66.25
homosexual	0.26	2.00	1.43	2.71
white	0.03	0.54	0.52	0.54
highschool	0.39	5.12	6.21	4.11
juniorcollege	0.65	6.70	6.25	6.15
college	0.78	8.86	10.18	7.27
graddegree	0.94	9.00	9.64	7.87
experience	0.81	7.29	8.18	6.01
experience^2	−0.34	5.53	6.64	4.30
union	0.11	2.20	2.04	2.31
executive	0.25	4.27	3.99	4.41
specialist	0.11	2.07	1.93	2.09
lowskill	−0.29	6.15	6.72	5.43
newengland	0.19	2.27	1.92	2.63
pacific	0.03	0.64	0.66	0.58
south	−0.09	1.34	1.09	1.59
urban	0.08	1.51	1.38	1.59

model of income determination. Table 7 provides additional summary statistics and the estimated magnitude of the earnings premium for homosexual women of 30% plus or minus 17 percentage points. Figure 2 compares the predicted distributions with the empirical distribution for all three specifications and demonstrates patterns similar to the male case.

VIII. THEORY IN LIGHT OF THE DATA

Having demonstrated that homosexual earnings are different than heterosexual earnings, this section attempts to analyze the much trickier question of what causes those differences. When considering whether the asymmetric pattern of expected earnings—whereby homosexual men do worse than average, and homosexual women do better than average—is consistent with the idea that firms discriminate against homosexuals, an attempt is made to take account of the costs borne by those who might choose to discriminate and determine circumstances under which it could be worth it for employers to indulge discriminatory sentiments by paying a premium for straight workers. Other explanations that do not involve discrimination will also need to be considered to see whether those alternative interpretations can be ruled out. The following scenarios describe several distinct theories that purport to explain the pattern of individual earnings across sexual orientation. Strong assumptions are made in the interest of starkly delineating these different scenarios, which hopefully capture important aspects of the real economy.

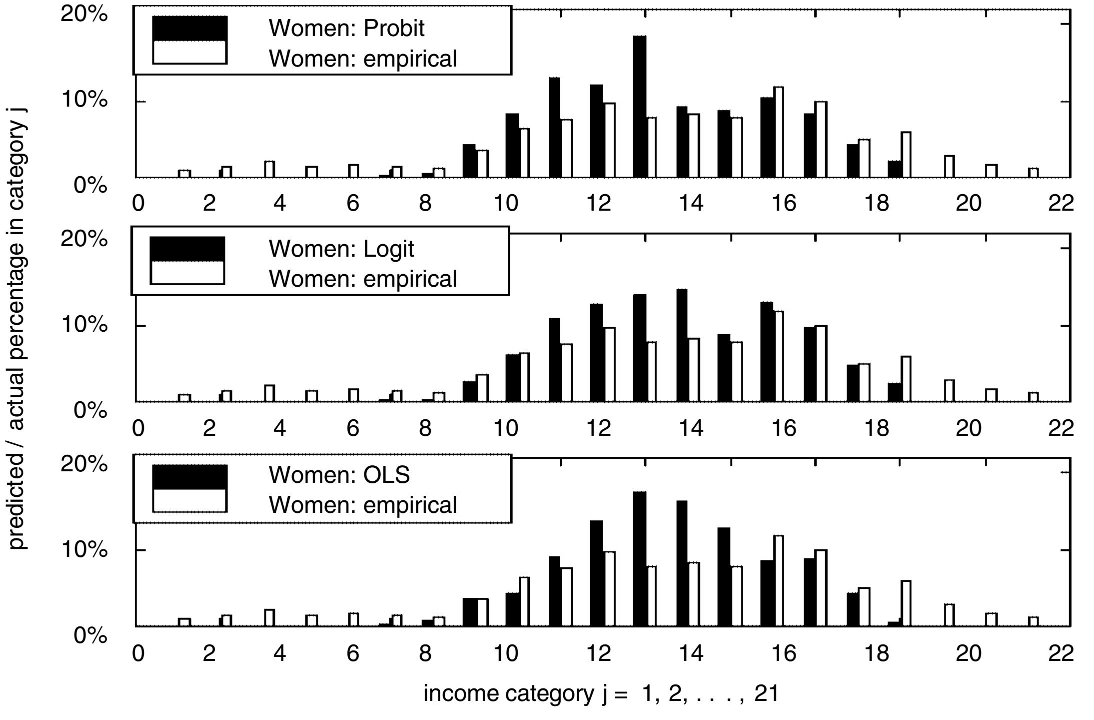
TABLE 7
Summary Statistics (Probit for Women)

Percentage correctly predicted	12.53
Homosexual coefficient (t-statistic)	0.2621 (2.0021)
Percentage gain for homosexuals	30 ± 17
Likelihood ratio	$926.45 > \chi(16)_{.005} = 34.27$
Log likelihood value	−3575.94
Minimum predicted income ($)	5,652.02
Maximum predicted income ($)	48,391.04

Scenario 1: Different Preferences over Leisure and Income, Not Discrimination, Explain Why Homosexual Men Earn Less and Homosexual Women Earn More

Attributing variation in any sort of observed behavior to different underlying preferences runs the risk of settling for a trivial explanation. If one suspects there to be systematic differences in preferences that correlate with sexual orientation, it is necessary to explain why. It is not immediately obvious why homosexual men would prefer income relative to leisure less than

FIGURE 2
Women: Frequencies of Predicted Income



other men, or why the reverse would hold for homosexual women. One possibility, however, relates to differences in individual plans about becoming a parent.

If there is a cultural or institutional norm that dictates that women stop working when they become parents, then it is easy to see that anticipating zero children raises the lifetime earnings payoff from investments in human capital. Although the estimated earnings model attempts to control for differences in human capital using a variety of education and experience variables, presumably there are other unobservable channels through which women anticipating uninterrupted careers can gain a productivity edge. This might explain higher earnings for female homosexuals, if it were true that they are less likely to interrupt their careers to care for children.

Another avenue through which plans about becoming a parent may affect labor/leisure trade-offs relates to saving money to pass on to future generations. If bequest motives are an important component in deciding how much lifetime earnings one needs relative to leisure, then anticipating zero children might lead to trading off labor income in favor of more leisure. This account, however, would seem to make identical predictions for both male and female homosexuals, tending to decrease the incomes of both. If one explains the male homosexual earnings gap of 22% as the result of different labor supply choices (arising from different plans about becoming a parent), then female homosexuals must start with a 52% earnings advantage over heterosexual women so that, after netting out the earnings reduction due to a smaller bequest motive, their earnings advantage matches the observed premium of 30%.

To test these ideas, one would like to measure the intentions of young homosexuals and young straights regarding family and career aspirations. Such an undertaking would require extensive survey work going well beyond the items in the GSS and is beyond the scope of the current study. The question of homosexuality and parental status can be pursued a bit further, however,

TABLE 8
Correlation of Parent-Status and Homosexuality

	Men	Women
Correlation coefficient	0.001	−.003
$\chi^2(1)$ test statistic $N \times \hat{\rho}^2$ [for $H_0 : \rho = 0$]	0.13	0.01

Note: The test statistics are well below even the 90% $\chi^2(1)$ critical value of 2.71.

using a GSS variable indicating whether an individual has children. Table 8 shows uncontrolled correlation coefficients for these two variables, broken out separately for male and female observations. Curiously, homosexuals do not appear any less likely to be parents, as one might have suspected.[17]

This correlation measure is not entirely satisfactory for at least two reasons. Even if we knew how being homosexual changes the probability of becoming a parent, we would also need to know how becoming a parent affects the probability of being a full-time worker, or striving for a high income, or being a more loyal employee so as to win an enhanced degree of job security from the employer, and so on. The second difficulty regards measuring the intention to become a parent, because the argument hinges on forward-looking behavior. In spite of these limitations, the quantitative measures in Table 8 are taken as tentative evidence against the notion that the decision to become a parent is the cause behind the sexual-orientation earnings gaps. After all, it would be surprising to find that family aspirations early in life are very different across sexual orientation and then, later in life, the two groups' propensities to be parents are identical.

17. Even if homosexuals are just as likely to be parents as heterosexuals, they may not be just as likely to interrupt their careers to care for children. In this case, the homosexuality variable may pick up differences in actual experience, because of a systematic tendency of female homosexuals, for example, to continue working even when children are present in the home. Explaining the odd earnings pattern in the GSS this way, however, requires asymmetric assumptions about labor supply decisions with respect to male and female homosexuals, an assumption for which there seems to be little evidence.

Scenario 2: Homosexuality, Accurately Perceived by Employers, Is Taken (Rightly or Wrongly) as an Informative Signal about an Employee's Future Career Trajectory and Interpreted Differentially According to Gender

A firm's decisions about setting wage schedules for females and investing in worker training may be a function of the firm's beliefs about the likelihood that individual female workers will become pregnant. In this case, making one's lesbian status public may be an especially credible signal of loyalty or workforce attachment, inducing differential investment in female human capital across sexual orientation.

There are at least two problems with this account. First, the data do not imply that female homosexuals are less likely to become mothers, calling into question why female homosexuality serves as a signal in the first place. Second, this story does not explain why male homosexuality serves as a signal that employers believe to be correlated with labor-force attachment. If a compelling theoretical account could be given that would explain why male homosexuality logically implies that the employee is a less worthwhile candidate to be a recipient of training and promotion and why female homosexuality implies the opposite, then this scenario might be plausible.

To address the question of whether wage schedules among different types of workers are truly distinct, a pooled sample of both men and women is used to ask the following question: Do female homosexuals, who appear to enjoy a wage premium because of their sexual orientation, resemble heterosexual males (the best-paid workers)? That is, can the data reject the hypothesis that female homosexuals are treated by labor markets as heterosexual males are? Similarly, is it possible to test whether male homosexuals are treated by labor markets in a manner similar to heterosexual females, with earnings determined by an identical set of weights on their characteristics?

Table 9 presents a regression of earnings categories on all the controls in the main model, except that the homosexual indicator variable is now replaced by three separate indicators distinguishing homosexual men, heterosexual women, and homosexual

TABLE 9
Are Homosexual Women Distinguishable from Straight Men?

	Estimated Earnings Effect	t-Statistic
const	8.4	22.59
white	0.27	1.48
highschool	1.99	7.71
juniorcollege	3.27	9.68
college	4.26	14.22
graddegree	5.06	14.18
experience	0.3	14.59
experience^2	−0.005	−11.01
union	1.13	6.41
executive	1.19	5.74
specialist	0.62	2.93
low-skill	−1.35	−8.16
newengland	0.84	2.83
pacific	0.14	0.78
south	−0.73	−2.83
urban	0.67	3.22
Homosexual men	−1.42	−3.21
Straight women	−2.64	−19.81
Homosexual women	−0.82	−1.68

Note: These are estimated effects on earnings measured in categories 1 through 21.

women from the heterosexual male reference group. The difference between homosexual women and straight men falls just short of the 95% level on a one-sided test, with a t-statistic of -1.6796. A Wald test of the hypothesis that heterosexual females are like homosexual males, however, is rejected, with a test statistic of 7.49, well above the 1% critical value $\chi^2(1) = 6.63$. It is possible to interpret the first result to mean that, in employers' eyes, homosexual females are equivalent to heterosexual men.

To understand better what is going on with sexual orientation and income, one may compute an Oaxaca decomposition. The goal is to distinguish whether it is the characteristics of homosexuals, that is, their regressor values, or the way they are treated in the labor market, as summarized by the coefficients in their wage equation, that accounts for the gap between what they earn and what heterosexuals earn. For both men and women, the share of the respective earnings gaps accounted for by being treated differently, that is, by different coefficients rather than characteristics, exceeds one! That is, although gay males have better average characteristics than straight males (more education and better job categories), they suffer from an earnings equation in which bad coefficients appear to deprive them of income, on the order of one or two income brackets. The wage premium enjoyed by homosexual women, at least by the Oaxaca measure, is equivalent to nearly two income brackets. This is all the more noteworthy in light of the fact that female homosexuals' mean characteristics appear to be less valuable, including fewer high school graduates and more low-skill workers. The Oaxaca results are summarized in Table 10.

Scenario 3: Individuals Make Labor/Leisure Decisions with Respect to Household Choice Sets; Taking the Gender Gap in Personal Income as Given, Household Choice Sets Differ According to Whether the Two Heads of Household Are Both Women, Opposite Sex, or Both Men

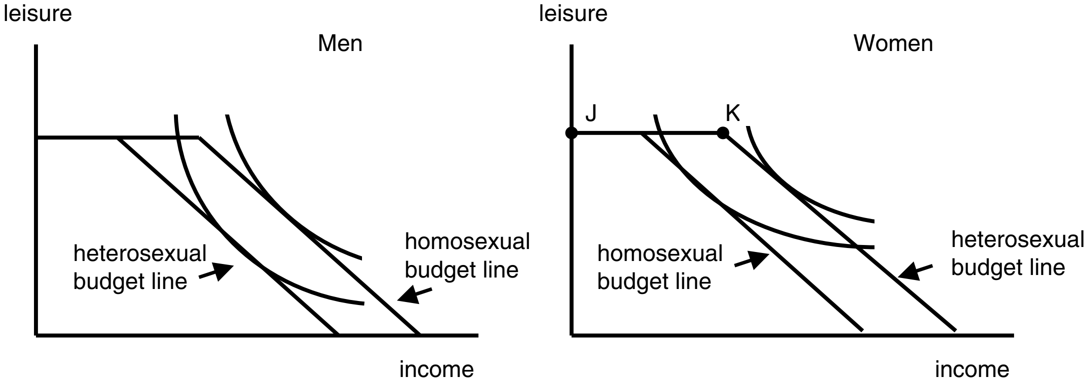
Unlike the previous two scenarios, the account offered in scenario 3 does not depend on different preferences among workers or among firms. Exogenous institutional norms prevail in which women earn less than men. This, in turn, leads to different levels of expected household income, depending on the gender of the earners in

TABLE 10
Oaxaca Decomposition Says Coefficients Matter Much More Than Characteristics

	Straight–Gay Wage Gap	Due to Coefficients	Due to Characteristics
Male level	0.79	1.37	−0.58
Male percentage	100	173	−73
Female level	−1.22	−1.88	0.66
Female percentage	100	154	−54

Note: Income is in units of income brackets, since this decomposition was computed by OLS directly on the categorical income bracket data.

FIGURE 3
Individual Income Differences Caused by Different Household Budget Sets



a particular household. As a result, one can observe identically productive workers of the same gender earning different levels of individual income, simply because they make leisure/labor decisions with respect to different anticipated household budgets.

Figure 3 depicts this scenario. The household budget sets differ depending on the gender of one's partner, not on sexual orientation per se. Because of the gender gap in income, a worker of either gender with a male partner has a higher expected household income. The budget sets include the option of zero work and positive income (points along the segment $\overline{JK}$), because an individual's partner has positive expected earnings regardless of whether that individual works. Because of a larger budget set, homosexual men choose more leisure; that is, they work less and therefore earn less as individuals. But as couples, homosexual males will enjoy higher household income because there are two male incomes. Female couples work more and therefore earn more as individuals, yet do worse as couples than women in married households do, without the benefit of a high male income in the household. This scenario matches both the individual income data presented in this article and the data in Klawitter and Flatt (1998) on the income of couples. The male–female gender gap explanation for the sexual-orientation gaps identified follow from a pure and simple income effect. To test this idea further, one would need to find an economy with no gender gap and test to see whether the sexual-orientation gap disappears, as predicted here.

Scenario 4: Discrimination

In contrast to the previous explanations of wage gaps, which are consistent with there being no discrimination, the case of bona fide discrimination is considered next. In light of recent antihomosexual hate crimes in the United States and the extreme rhetoric used by some opponents of civil rights laws like ENDA, it should not be hard to imagine that there are employers who discriminate. Employers who discriminate may dislike the idea of working with homosexuals themselves; they may fear that their customers dislike transacting with homosexuals; or they may worry that co-workers who dislike homosexuals will be less productive in their presence. Employers with such concerns might offer homosexuals a wage lower than the heterosexual wage to equalize the unit cost of labor, after psychic costs and perceived negative externalities are factored in.

What should be kept in mind, however, is the point made by Becker (1976) that discrimination—far from a strategic tool to advance the interests of a favored group—actually incurs costs that are borne in part by the group that discriminates. In his analysis, Becker provides a numerical heuristic demonstrating the range of possible income losses for both groups, bounded by the extreme cases of no discrimination on the one hand and complete segregation on the other. Becker provides inequalities guaranteeing that the group discriminated against suffers more than the group that discriminates. An analogous heuristic for the case of a society of homosexuals and heterosexuals in which all heterosexuals despise homosexuals can be studied to gain insight into the plausibility of the discrimination scenario.[18] The heuristics tell us that, because of the small number of homosexuals relative to heterosexuals, the majority group barely suffers at all from discrimination, even under the worst-case scenario of total segregation. Homosexuals themselves do not fare so badly either, at least in the Becker model which predicts that they lose less than 5% under total segregation. Thus, this theory cannot account for the sizable wage gaps present in the data. The discrimination account faces the further challenge of explaining why firms discriminate against homosexual men and in favor of homosexual women.

Scenario 5: The Labor Market Can Get Stuck in an Inefficient Nash Equilibrium Where Workers and Employers Condition Their Strategies on Sexual Orientation to Coordinate other Activities, Even Though No One Actually Cares about Sexual Orientation

Kaneko and Kimura (1992) model a society comprised of two groups, identical in all ways except for one. This one difference does not enter anyone's utility function, meaning that no one really cares about the trait. Nevertheless, Kaneko and Kimura show that if each group begins playing random strategies over a small strategy space (spanning a range of discriminatory and nondiscriminatory strategies), then the "everyone discriminates" outcome is a Nash equilibrium. Members of society wind up using the observable trait, about which everyone is indifferent, to coordinate their actions along dimensions where there are real stakes. The trait, which is intrinsically meaningless from an economic standpoint, comes to play an economically important role as a coordination device.

18. The analysis is less clear-cut when only some discriminate and others do not. In fact, the existence of major U.S. firms that publicly support homosexual organizations and advocate on behalf of ENDA makes it hard to understand why competition does not drive out discriminators, that is, how employers who discriminate could profitably coexist with nondiscriminating firms.

This fascinating result suggests that discriminatory attitudes might arise as a signaling strategy used by firms to coordinate unrelated activity, even in the absence of underlying intergroup hostility. Without denying the existence of real hatred and horrific violence against homosexuals in the United States, it is interesting to ponder whether the sexual orientation issue is a kind of short-hand signal conveying information about more important interpersonal differences that firms, workers, and consumers do care about. Thus, a firm's attitude toward sexual orientation may serve merely to coordinate the coming together of a harmonious pool of like-minded workers who are more productive when they work together.

If in fact homosexuality as a workplace issue is used by firms to assemble pools of employees with compatible attitudes, we would expect to see a polarization among firms into gay-friendly and gay-hostile types. The participation of many high-profile firms in alliances with organizations that advocate on behalf of homosexuals is consistent with this idea. But the earnings differentials—negative for gay men and positive for homosexual women—do not find an easy explanation with this theory.

In total, two of the five scenarios (Scenarios 2 and 4) are directly contradicted by the data. The last scenario cannot explain the asymmetric earnings differentials. The first does well until one considers the lack of correlation between parental status and homosexuality. This leaves only the third scenario, which is the simplest, standing as an explanation for what is observed. The third scenario explains the sexual-orientation earnings gaps as the consequence of an income effect on homosexuals' demand for leisure (labor supply). This income effect is caused by an exogenously given gender wage gap. Because male workers expect to earn more than equally productive female workers, the gender of one's partner affects lifetime household income. The implied labor-supply responses match the earnings pattern in the GSS data, without assuming any kind of heterogeneity across gender or sexual orientation. Though it cannot be claimed that this is a complete explanation, it does appear to be the theory that best explains the data.

IX. CONCLUSION

This article has provided new evidence that there are significant income disparities between otherwise similar homosexual and nonhomosexual workers in the United States. Homosexual men earn 16% to 28% less than nonhomosexual men with similar demographic characteristics. Unlike homosexual men, homosexual women appear to benefit from their sexual orientation, with an expected earnings premium of 13% to 47% over the earnings of nonhomosexual women.

Of the five theories considered, only one is consistent with the empirical record. That theory, taking the well-established gender gap in personal income as given, attributes the sexual-orientation earnings gap to an income effect on labor supply caused by budget sets that differ across couples only because of the genders of the two heads of household. According to this story, male homosexual earnings decline due to an income effect resulting from membership in a household with two high (male) earnings schedules. Female homosexual earnings increase due to an income effect resulting from membership in a household with two low (female) earnings schedules.

Although there is some evidence that female homosexuals enjoy earnings functions that are indistinguishable from heterosexual male earnings, heterosexual females and homosexual males appear to be treated differently somehow, with distinct labor market experiences, unattributable to differences in characteristics. Surprisingly, being a parent and being homosexual are not correlated, weakening support for the theory that distinct intentions about becoming a parent can account for the earnings differentials. Far short of a resounding corroboration of a single dominant theory, this analysis does succeed in eliminating several theories that contradict the available data and brings us one step closer to understanding the curious role nonproductive factors appear to play in the determination of income.

By estimating a single earnings equation using categorical earnings data (corresponding to income brackets from the GSS experimental design) and a maximum-likelihood approach, this article helps solidify the empirical record on sexual orientation and individual earnings. The estimates provide strong evidence against the notion that most homosexuals are affluent. But the results are inconclusive on the question of whether homosexuals suffer from workplace discrimination. It is difficult to reject the notion that male homosexuals are treated differently than straight males. And this may in fact be attributable to discrimination, which would argue in favor of legislation, such as ENDA. The surprisingly successful labor market outcomes of homosexual women, however, should be kept in mind when debating the need for new sanctions against antihomosexual discrimination. Perhaps generalizations about the labor market experiences of all homosexuals should be resisted in favor of more gender specificity in analyzing the discrimination question as it concerns homosexuals.

The rhetoric of earnings differentials has played a notable role in the debate about policies such as ENDA, whether it be the myth of gay affluence perpetuated by some opponents of ENDA or the dogged assertion, made by some ENDA supporters, of widespread discrimination against both male and female homosexuals. At the very least, our empirical work should convince activists on both sides of the ENDA debate that there is no true generalization about the average earnings of male and female homosexuals taken together as a single bloc. Our finding should disarm the extreme rhetoricians on both sides and, hopefully, force the discussion toward more sensible criteria for deciding on an important policy question that continues to hang in the balance.

APPENDIX A: TESTING FOR ENDOGENEITY OF INCOME AND OCCUPATIONAL CHOICE

In the earnings equation, income category y depends on a constant and 15 control variables, comprising the vector x, as well as d, a dummy variable indicating homosexuality. In this appendix, the variable "specialist" (an element of x) is relabeled as a, which is 1 if an individual chooses to become a specialist, that is, if the individual seeks out *autonomy*. Aside from a, the remaining 14 controls in x are partitioned into two vectors of control variables: z, which affects the decision to pursue autonomy, and w, the controls that do not affect the decision about choosing an autonomous occupation. In this setup, w includes the following nine variables (those that affect income but not autonomy): *experience, experience*^{2}, *union, executive, low-skill, newengland, pacific, south*, and *urban*. The remaining six variables, including the constant, appear in both equations as z. Apart from reordering,

(4) $$x = \begin{bmatrix} z \\ w \\ a \end{bmatrix}.$$

A two-equation system consisting of an earnings equation and an autonomy equation can be written

(5) $$y = \lambda' z + \gamma' w + \alpha a + \eta,$$

(6) $$a = \xi' z + \rho' y + \nu.$$

In reduced form, the system becomes

(7) $$y = \frac{1}{1-\alpha\rho}[(\lambda + \alpha\xi)' z + \gamma' w] + \eta^*,$$

(8) $$a = \frac{1}{1-\alpha\rho}[(\xi + \rho\lambda)' z + \rho\gamma' w] + \nu^*.$$

The reduced-form equations are estimated and the 9×1 reduced form coefficient on w in the second equation, $\frac{1}{1-\alpha\rho}\rho\gamma$, is seen to be statistically indistinguishable from zero (a vector of zeros), while the 9×1 coefficient $\frac{1}{1-\alpha\rho}\gamma$ on w in the first equation is different from zero.

Specifically, we test the hypotheses $\frac{1}{1-\alpha\rho}\rho\gamma \equiv \gamma_a^* = 0$ and $\frac{1}{1-\alpha\rho}\gamma \equiv \gamma_y^* = 0$. Denoting the variance of those estimators as Σ_a and Σ_y, the corresponding test statistics are

(9) $$\frac{(e'\widehat{\gamma_a^*})^2}{e'\Sigma_{\hat{a}} e} = 16.17 \quad \text{and} \quad \frac{(e'\widehat{\gamma_y^*})^2}{e'\Sigma_y e} = 21.16,$$

where e is a 9×1 vector of ones. The 5% critical threshold $\chi^2(9) = 16.9$, meaning that we accept the first hypothesis and reject the second. Together, these results imply $\rho = 0$, which means that a is not a function of y and that there is no simultaneity problem to worry about, at least regarding the variables y and a.

APPENDIX B: ESTIMATION PROCEDURE IN DETAIL

We write the model of individual income determination in matrix form:

(10) $$y^* = X\beta + \epsilon,$$

where y^* is an $N \times 1$ vector, the ith element of which is the natural log of the ith individual's income, $i = 1, \ldots, N$. The N disturbances ϵ satisfy $E(\epsilon|X) = 0$, are homoscedastic $[E(\epsilon\epsilon') = \sigma^2 I_N]$, and $\frac{\epsilon}{\sigma}$ is distributed according to a known cdf, F. The $N \times K$ matrix X contains a constant as well as observations on individual characteristics thought to determine income, including sexual orientation.

To get a feel for why OLS does not make sense in estimating this model, suppose imputed income values $\hat{y}_1^*, \hat{y}_2^*, \ldots, \hat{y}_J^*$ could be assigned and then used as the dependent variable in OLS estimation of β.[19] Expanding the lefthand side of the linear regression $E(\widehat{y^*}|x) = \lambda' x$, one is led to the following restrictions:

(11) $$E(\widehat{y^*}|x) = \sum_{j=1}^{J} y_j^* \Pr(\widehat{y^*} = \widehat{y_j^*}|x)$$

(12) $$= \sum_{j=1}^{J} y_j^* \Pr(y = j|x)$$

(13) $$= \sum_{j=1}^{J} y^* \left[F\left(\frac{a_j - \beta' x}{\sigma}\right) - F\left(\frac{a_{j-1} - \beta' x}{\sigma}\right)\right]$$
$$= \lambda' x.$$

The last equality imposes linearity on F, ruling out that F is normal or logistic.

Instead, this article's approach begins with the log-likelihood function for β and σ,

(14) $$\log(L) = \sum_{i=1}^{N} \sum_{j=1}^{J} q_{ij} \log(F_{ij} - F_{ij-1}),$$

where

(15) $$F_{ij} \equiv F(Z_{ij}),$$

(16) $$Z_{ij} \equiv \frac{a_j - \beta' x_i}{\sigma},$$

(17) $q_{ij} = 1$ when $a_{j-1} < y_i^* < a_j$ and $q_{ij} = 0$ otherwise.

Although the log-likelihood function is globally concave in β, the Hessian with respect to the full parameter $[\beta'\sigma]'$ is indefinite for some parameter values. Newton's method–type numerical procedures are unstable and fail to converge for some initial values. In the case where F is normal, Stewart (1983) provides a two-step iterative procedure (of the EM algorithm type) that converges to the maximum likelihood estimation (MLE) yielding consistent estimates of β and σ. Stewart's procedure, rewritten here in a more convenient matrix form, is described later. When F is nonnormal, it is possible to perform line-search on σ, combined with Newton-Raphson, to obtain a corresponding estimate of β for each value of σ, ultimately producing an MLE estimate.

Define M_r to be an $N \times 1$ vector, the ith component of which is

(18) $$\sum_{j=1}^{J} q_{ij} \frac{Z_{ij-1}^r f_{ij-1} - Z_{ij}^r f_{ij}}{F_{ij} - F_{ij-1}}, \quad r = 0, 1, 2, 3$$

where f_{ij} is the pdf of $\frac{\epsilon}{\sigma}$ evaluated at Z_{ij}. Define T_r to be the $N \times N$ diagonal matrix with diagonal M_r. Following

19. Badgett used a second data set and some information from the regressors to impute more than J values for income. Still, her imputed income variable is discrete-valued and subject to similar logical consistency problems in the context of a linear model, in addition to being inconsistent and producing unreliable standard errors. Her approach may lead to genuine improvements over the simpler midpoint technique. But the approach taken here is even simpler and resolves all three shortcomings of imputed income estimation already mentioned.

Stewart, the first and second derivatives used to express first-order conditions and compute standard errors can be compactly expressed as

$$(19) \qquad (1\times 1)\ \frac{\partial \log(L)}{\partial\sigma} = e'M_1\sigma^{-1}$$

(e is an $N\times 1$ vector of ones),

$$(20) \qquad (J\times 1)\ \frac{\partial \log(L)}{\partial\beta} = X'M_0\sigma^{-1}$$

$$(21) \qquad (J\times J)\ \frac{\partial^2 \log(L)}{\partial\beta\partial\beta'} = X'(T_1 - T_0T_0)X\sigma^{-2},$$

$$(22) \qquad (J\times 1)\ \frac{\partial^2 \log(L)}{\partial\beta\partial\sigma} = X'(M_2 - M_0 - T_1M_0)X\sigma^{-2},$$

$$(23) \qquad (1\times 1)\ \frac{\partial^2 \log(L)}{\partial\sigma^2} = e'(M_3 - 2M_1 - T_1M_1)\sigma^{-2}.$$

When F is normal, the conditional mean of continuous income y^* given the realized category y and regressors X is

$$(24) \qquad m \equiv E(y^*|y, x = X\beta) + \sigma M_0.$$

Multiplying through by X' and substituting first-order condition (20), we have

$$(25) \qquad X'm = X'X\beta + \sigma X'M_0 = X'X\beta.$$

Also, the variance of y^* given category y and x is

$$(26) \qquad \mathrm{var}(y_i^*|y_i, x_i) = \sigma^2 v_i,$$

where

$$(27) \qquad (N\times 1)\ v \equiv [v_1, \ldots, v_N]' = M_1 - T_0M_0 + e.$$

Multiplying through by e' and using first-order condition $e'M_1 = 0$ and $m - X\beta = \sigma M_0$, one gets

$$(28) \qquad (1\times 1)\ \sigma^2 e'v = \sigma^2(e'M_1 - e'T_0M_0 + N) = \sigma^2(N - e'T_0M_0)$$

and

$$(29) \qquad \hat\sigma^2 = \frac{(m - X\beta)'(m - X\beta)}{N - e'v}.$$

The iteration is as follows. Begin with initial estimates $[\hat\beta'_{(0)}, \hat\sigma_{(0)}]$ (which need not be consistent for the numerical procedure to converge to the MLE). Given the results of the tth iteration $\hat\beta_{(t)}$ and $\hat\sigma_{(t)}$, compute

$$(30) \qquad \hat m_{(t)} = X\hat\beta_{(t)} + \hat\sigma_{(t)}\hat M_{1(t)}.$$

Then use $\hat m_{(t)}$ in

$$(31) \qquad \hat\beta_{(t+1)} = (X'X)^{-1}X'\hat m_{(t)} \quad \text{[from (25)]},$$

and

$$(32) \qquad \hat\sigma^2_{(t+1)} = \frac{(\hat m_{(t)} - X'\hat\beta_{(t)})'(\hat m_{(t)} - X'\hat\beta_{(t)})}{N - e'v_{(t)}}.$$

Iteration is repeated until a stopping criterion is satisfied, for example,

$$(33) \qquad \frac{\log[L(\hat\beta_{(t+1)}, \hat\sigma_{(t+1)})]}{\log[L(\hat\beta_{(t)}, \hat\sigma_{(t)})]} - 1 < 10^{-6} \quad \text{and}$$

$$\max\left\{\left(\left|\frac{\hat\beta^k_{(t+1)}}{\hat\beta^k_{(t)}} - 1\right|\right)_{k=1}^{K}, \left|\frac{\hat\sigma_{(t+1)}}{\hat\sigma_{(t)}} - 1\right|\right\} < 10^{-4}.$$

APPENDIX C: DETAILS ON THE BECKER MODEL

Assume homosexuals and heterosexuals both have the same technology f, but that heterosexuals hold more per capita capital and outnumber homosexuals. Normalizing the gay labor force and capital stock to unity ($L_g = K_g = 1$), suppose the straight labor force[20] is $L_s = 19$ and suppose each heterosexual holds twice the capital of homosexuals so that $K_s = 2\times 19 = 38$. Then, with intensive-form production function $f(\frac{K}{L}) \equiv (\frac{K}{L})^{\frac{1}{3}}$, competitive equilibrium with no discrimination implies that marginal factor returns are equal across production groups. With K_t defined as the level of capital traded from the (straight) s-group to the (gay) g-group, the no-discrimination-equilibrium level of K_t solves

$$(34) \qquad f'\left(\frac{K_s - K_t}{L_s}\right) = f'\left(\frac{K_g + K_t}{L_g}\right), \text{ or}$$

$$\frac{K_s - K_t}{L_s} = \frac{K_g + K_t}{L_g}.$$

Plugging in values for capital and labor stocks, $\frac{38-K_t}{19} = \frac{1+K_t}{1}$ so that $K_t = 19/20 = 0.95$. Aggregate income for gays and straights, Y_g and Y_s, respectively, depends on output as well as rental costs (revenue) from borrowing (loaning) capital. Substituting in marginal products for equilibrium prices, aggregate income for gays and straights with no discrimination is

$$(35) \qquad Y_g^0 = L_g\left(\frac{K_g + K_t}{L_g}\right)^{\frac{1}{3}} - K_t\frac{1}{3}\left(\frac{K_g + K_t}{L_g}\right)^{-\frac{2}{3}}$$

$$= 1\times\left(\frac{1.95}{1}\right)^{\frac{1}{3}} - \frac{0.95}{3}\left(\frac{1.95}{1}\right)^{-\frac{2}{3}}$$

$$= 1.0464,$$

20. Badgett (1995) reports that the demographic literature contains a fairly wide range of estimates for the percentage of Americans who are homosexual: from 2% to 10%. In the GSS data under consideration in this article, the fraction of homosexuals is 4%. In Becker's discrimination model, the more of a minority a group is, the more it loses from the presence of discrimination against it.

$$(36)\quad Y_s^0 = L_s\left(\frac{K_s - K_t}{L_s}\right)^{\frac{1}{3}} + K_t\frac{1}{3}\left(\frac{K_s - K_t}{L_s}\right)^{-\frac{2}{3}}$$

$$= 19 \times \left(\frac{38 - .95}{19}\right)^{\frac{1}{3}} + \frac{0.95}{3}\left(\frac{38 - .95}{19}\right)^{-\frac{2}{3}}$$

$$= 23.9402.$$

When the marginal revenue from the first unit of transferred capital exceeds the psychic cost of transacting with gays, no "trade" occurs, which is referred to as "total segregation." When total segregation prevails, traded capital $K_t = 0$, and aggregate income is

$$(37)\qquad Y_g' = L_g\left(\frac{K_g + 0}{L_g}\right)^{\frac{1}{3}} = 1,$$

$$(38)\qquad Y_s' = L_s\left(\frac{K_s - 0}{L_s}\right)^{\frac{1}{3}} = 23.9385.$$

The maximum loss from discrimination is 4.6% for gays but less than 0.01% for straights. Alternatively, we could work backward from the earnings gap that found for gay men of around 18% and calculate a lower bound for the ratio of capital per straight worker to capital per gay worker. It would have to be considerably more than two, as it was in the example.

Of course, this is only a heuristic and is not intended as a serious econometric finding. Yet it helps approximate a range of magnitudes for earnings losses that can be attributed to discrimination and therefore encourages us to rule out discrimination in the current setting, where the earnings differentials based on sexual orientation easily exceed 5%, which was the upper bound on gays' loss due to discrimination.

REFERENCES

Alsop, R. "Are Gay People More Affluent than Others?" *Wall Street Journal*, December 30, 1999, p. B1.

Badgett, M. V. L. "The Wage Effects of Sexual Orientation Discrimination." *Industrial and Labor Relations Review*, 48(4), 1995, 726–39.

Becker, G. S. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press, 1976.

Cameron, A. C., and P. K. Trivedi. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, 1998.

Davis, J. A., and T. W. Smith. *General Social Surveys, 1972–1996: Cumulative Codebook.* Chicago: NORC, 1996.

Kaneko, M., and T. Kimura. "Conventions, Social Prejudices and Discrimination: A Festival Game with Merrymakers." *Games and Economic Behavior*, 4, 1992, 511–27.

Klawitter, M. M., and V. Flatt. "The Effects of State and Local Antidiscrimination Policies on Earnings for Gays and Lesbians." *Journal of Policy Analysis and Management*, 17(4), 1998, 658–86.

Hsiao, C. "Regression Analysis with a Categorized Explanatory Variable," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, edited by S. Karlin, T. Amemiya, and L. Goodman. New York: Academic Press, 1983.

Murphy, K. M., and F. Welch. "Empirical Age-Earnings Profiles." *Journal of Labor Economics*, 8, 1990, 202–29.

Stewart, M. B. "On Least Squares Estimation when the Dependent Variable Is Grouped." *Review of Economic Studies*, 50, 1983, 737–53.