Appendix A: Methodology

The methodological approach of this report builds on an extensive literature on the analysis of faculty pay, including Johnson and Stafford (1975), Hoffman (1976), Barbezat (1987), Ranson and Megdal (1993), Ornstein and Stewart (1996), Ginther and Hayes (2003) and Becker and Toutkoushian (2003).[1] These studies, in turn, are closely related to a much broader literature on the interpretation of gender and racial differentials in individual wages (e.g., Oaxaca, 1973; Malkiel and Malkiel, 1973; Cain, 1986; Altonji and Blank, 1999; Blank, Dabady and Citro, 2004).

a. Conceptual Model

The starting point for an analysis of faculty salaries is the assumption that observed salaries are determined by a combination of four factors:

(1) observable pay factors (like time since completion of degree or academic department)

(2) unobserved but legitimate pay factors (like measures of research productivity or teaching effectiveness)

(3) discriminatory factors (reflecting the decisions of current and previous department chairs, committees within the university, and outside actors like award committees and previous employers)

(4) purely random factors.

Let X denote the set of observed pay determinants for a particular faculty member, let u_1 denote the unobserved but legitimate pay factors, let u_2 denote the discriminatory factors, let v denote purely random factors, and let Y represent the salary of the faculty member. It is assumed that $Y=f(X, u_1, u_2, v; \beta)$ where f is some function with unknown parameters β representing the "pay determination system" of the university. It is also typically assumed that the pay system can be approximated by a linear function of the form:

$$(1) \qquad g(Y) = X\beta + u_1 + u_2 + v$$

where $g(Y)$ is most often the logarithm of salary.

[1] Becker and Toutkoushian (2003, Table 1) present a useful summary of the earlier literature, distinguishing between studies of salaries at individual institutions and salaries in national samples of faculty. The studies in Haignere (2002) provide some practical guidance for the conduct of salary equity studies.

Let D represent a set of demographic variables, including indicators for gender and key ethnic groups, and possibly their interactions. The next step is to consider how u_1 and u_2 are related to the observed salary factors X and D. In general, this relationship can be written as:

(2a) $u_1 = X\alpha_1 + D\gamma_1 + e_1$,

(2b) $u_2 = X\alpha_2 + D\gamma_2 + e_2$,

where the remainder terms e_1 and e_2 are uncorrelated with X or D.

Both the legitimate/unobserved pay factors u_1 and the discriminatory pay factors u_2 are potentially related to X and D. For example, suppose that the set of observed covariates X includes an indicator for being a member of the law school faculty. Then the term $X\alpha_1$ in equation (2a) includes a component representing the net difference in the legitimate/unobservable pay factors between faculty in the law school and faculty in other departments. Likewise the term $X\alpha_2$ in equation (2b) includes a component that represents the net difference in discriminatory pay factors between faculty in the law school and faculty in other departments. Similarly, the gender component of the term $D\gamma_1$ in equation (2a) represents the net difference in the legitimate/unobservable pay factors between female faculty and male faculty, while the gender component of the term $D\gamma_2$ in equation (2b) represents the net difference in discriminatory pay factors between female and male faculty.

Combining equations (1), (2a) and (2b) leads to an estimable salary regression model of the form:

(3) $g(Y) = X(\beta+\alpha_1 + \alpha_2) + D(\gamma_1+ \gamma_2) + \varepsilon$

where X and D are both observable, and the combined error term $\varepsilon = e_1 + e_2 + v$ consists of components that are uncorrelated with X and D (and is therefore a "clean" residual).

This development shows that the coefficients of the demographic variables in the salary regression model capture two conceptually distinct effects: (i) components of salary attributable to unobserved but legitimate pay factors that are correlated with D and not directly explained by the observed covariates X; (ii) components of salary attributable to discriminatory pay factors that are correlated with D and not directly explained by X. Debate over the interpretation of the measured demographic salary gaps from a regression model like (3) is, in essence, a debate over the relative size of the γ_1 and γ_1 terms, and in the extreme, whether one or the other term is zero, so all of the observed salary differences between demographic groups can be attributed to productivity differences (i.e., γ_2=0) or discrimination (i.e., γ_1=0). A purely statistical analysis can never fully resolve this debate. By comparing results from models with different sets of control variables, however, it may be possible to discern a plausible range for the magnitude of any discriminatory wage differences.

This development also illustrates the problem in interpreting the results from salary models that include X variables that are potentially correlated with discriminatory factors. One view of faculty rank indicators, for example, is that these variables would be "tainted" by any discrimination in the promotion process (see e.g., Ginther and Kahn, forthcoming, for a critical analysis of the faculty promotion process). In that case, when rank variables are included in the set of *X's* in the salary model they would soak up some of the differences across groups that are actually attributable to the discriminatory factors u_2, leading to an under-statement of the true magnitude of the discriminatory wages gaps by gender or ethnicity.

b. Model Specification

The models presented in this report are versions of equation (3), estimated by ordinary least squares (OLS) with varying sets of variables included as elements of X. We consider three broad classes of models:

(a) models for the logarithm of salary, fit to the overall population of faculty

(b) models for the level of salary, fit to the overall population of faculty

(c) models for the level of salary, fit to white male faculty only, and then extrapolated to derive predicted salaries for other groups.

The use of logarithmic models is widespread in the faculty salary literature and in the broader literature on wage differences across gender, race, and ethnicity groups.[2] This reflects the empirical observation that salaries are approximately log-normally distributed, and that observed pay determination factors (e.g., years of career experience or type of degree) generate approximately proportional pay differentials, rather than absolute pay differences.[3] Despite these advantages, some previous studies use the level of salaries, in part because the coefficient estimates may be more easily interpreted by non-specialists (see Haignere, 2002). As discussed in the report, however, the implied differentials from the alternative specifications are very similar.

The choice between fitting wage models for the entire faculty and fitting a model to white males only is more complex. The advantage of using only white male faculty to estimate the model is that the estimated effects of the control variables will be unaffected by discriminatory factors that may impact the salary determination of other groups.[4] This is offset by the disadvantage that only about one-half of UC Berkeley faculty are white men. Models fit to white males therefore yield relatively imprecise estimates of the effects of key control variables like length of career

[2] For example, the early studies by Oaxaca (1973), Malkeil and Malkeil (1973) and Johnson and Stafford (1974) all model the logarithm of salaries.
[3] In the UCB salary data the 90th and 10th percentiles of log salary are approximately equidistant from the median of log salary, and the 90-10 difference, divided by 2.56, is approximately equal to the standard deviation of log salaries, as would be expected if the log of salary were normally distributed.
[4] For example, if biases in the salary setting process lead to slower wage growth for female faculty, the estimated career profile of wages for the overall faculty will be "too flat". The net effect on the estimated average wage differences between male and female faculty, however, may still be relatively small.

service or department, leading to excess sampling variability in the demographic wage differentials from models that control for these factors. In practice, the implied wage differentials from models fit to the overall faculty and to white males only are typically very similar. An exception arises when the exclusion of female and non-white faculty prevents the estimation of separate department effects for smaller departments that have only a handful of white male faculty.

c. Specification of Demographic Effects

The demographic wage differentials presented in the main body of this report are based on models with four indicator variables: (i) an indicator for female faculty; (ii) an indicator for faculty members of Asian ethnicity; (iii) an indicator for faculty members of under-represented minority (URM) groups; (iv) an indicator for faculty of unknown ethnicity. Implicit in this specification is the assumption that the effects of gender and ethnicity are additive. For example, in a log salary model the implied percentage salary differential between Asian females and white males is the sum of the female percentage differential and the Asian percentage differential, while in a salary levels model the implied absolute wage gap is the sum of the absolute female differential and the absolute Asian differential.

To test this assumption a series of alternative models were estimated that included a full set of interactions. Specifically, the alternative models included the following effects:

- Female = indicator for female faculty
- Asian = indicator for Asian faculty
- URM = indicator for URM faculty
- Unknown = indicator for unknown ethnicity
- Asian Female = indicator for Asian female faculty
- URM Female = indicator for URM female faculty
- Unknown Female = indicator for female faculty of unknown ethnicity

The implicit assumption of the "additive model" is that the last three of these effects are all 0.

Appendix Table A1 compares the estimated percentage wage differentials from the baseline specifications and the alternative specifications for submodels 2, 3 and 4, using the logarithmic salary model for all faculty members.

Appendix Table A1: Comparison of Additive Gender/Ethnicity Models with General Models

	Log Salary Models for All Faculty					
	Submodel 2		Submodel 3		Submodel 4	
	Additive	General	Additive	General	Additive	General
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient Estimates (Omitted group = White Men)[a]						
Female	-11.35	-10.82	-4.29	-4.25	-1.77	-1.63
	(1.51)	(1.52)	(1.11)	(1.28)	(0.96)	(1.10)
Asian	-3.68	-3.96	-1.66	-1.53	-1.75	-2.22
	(2.07)	(2.54)	(1.50)	(1.81)	(1.27)	(1.54)
URM	-7.16	-6.27	-1.20	-1.65	-0.99	-0.30
	(2.49)	(3.10)	(1.90)	(2.27)	(1.61)	(1.93)
Unknown Ethnicity	-9.04	-5.37	-2.38	-1.34	-2.84	-1.51
	(4.18)	(4.87)	(2.99)	(3.47)	(2.57)	(2.96)
Female × Asian	0.00	0.69	0.00	-0.36	0.00	1.30
	–	(4.35)	–	(3.09)	–	(2.63)
Female × URM	0.00	-2.56	0.00	1.29	0.00	-1.99
	–	(5.17)	–	(3.68)	–	(3.11)
Female × Unknown	0.00	-13.31	0.00	-3.76	0.00	-5.11
	–	(9.00)	–	(6.41)	–	(5.47)
R-squared of model	0.40	0.40	0.72	0.72	0.80	0.80
P-value: All interactions =0	–	0.48	–	0.92	–	0.66
Implied Percentage Gaps Relative to White Males (p-values below)[b]						
Asian Men	-3.68	-3.96	-1.66	-1.53	-1.75	-2.22
	p=.08	p=.12	p=.27	p=.40	p=.17	p=.15
URM Men	-7.16	-6.27	-1.20	-1.65	-0.99	-0.30
	p=.00	p=.04	p=.52	p=.47	p=.54	p=.88
Unknown Men	-9.04	-5.37	-2.38	-1.34	-2.84	-1.51
	p=.03	p=.27	p=.43	p=.70	p=.27	p=.61
White Women	-11.35	-10.82	-4.29	-4.25	-1.77	-1.63
	p=.00	p=.00	p=.00	p=.00	p=.06	p=.14
Asian Women	-15.03	-14.09	-5.95	-6.14	-3.52	-2.56
	p=.00	p=.00	p=.00	p=.01	p=.02	p=.22
URM Women	-18.51	-19.65	-5.49	-4.61	-2.76	-3.91
	p=.00	p=.00	p=.01	p=.13	p=.14	p=.13
Unknown Women	-20.39	-29.50	-6.67	-9.35	-4.61	-8.25
	p=.00	p=.00	p=.04	p=.09	p=.10	p=.08

Notes: Estimated standard errors in parentheses. See text for description of submodels. Sample size for all is 1,519. Mean and standard deviation of dependent variable for all models are 11.886 and 0.345.

[a]Coefficient estimates and standard errors are multiplied by 100.

[b]Entries are the sum of the relevant percentage gaps for the specific gender/ethnicity group.

The upper panel reports the coefficient estimates from each specification, while the lower panel of the table shows the implied percentage wage differentials for all seven possible gender/ethnicity groups relative to white males. The odd-number columns present the results from the baseline specification (which only includes the Female, Asian, URM, and unknown ethnicity effects) while the even-numbered columns show the corresponding generalized models, which add the three gender/ethnicity interaction effects. We also show the p-value for the joint test that the three interaction effects are jointly equal to zero (in the last row of the upper panel). For all three submodels the test statistic is insignificant (p-value = 0.48 for submodel 2, 0.92 for submodel 3, and 0.66 for submodel 3).

Consistent with the lack of significant interaction effects, comparisons of the implied wage gaps relative to white men from the additive and general specifications for each submodel suggest that the implied salary differentials from the two specifications are quite similar. We therefore conclude that the additive specification provides an adequate model for the wage gaps for female faculty in each of the non-white ethnicity groups.

Professor David Card, Class of 1950 Professor of Economics
Dr. Marc Goulden, Director of Data Initiatives, Office of the Vice Provost for the Faculty

References

Altonji, Joseph G. and Rebecca M. Blank (1999). "Race and Gender in the Labor Market." In Orley Ashenfelter and David Card, Editors. *Handbook of Labor Economics* Volume 3A. New York: Elsevier, 1999: 3143-3259.

Barbezat, Debra A. (1987). "Salary Differentials by Sex in the Academic Labor Market." *Journal of Human Resources* 22 (Summer 1987): 422-428.

Becker, William E. and Robert K. Toutkoushian (2003). "Measuring Gender Bias in the Salaries of Tenured Faculty Members." *New Directions for Institutional Research* 117 (Spring 2003): 5-20.

Blank, Rebecca M., Marilyn Dabady, and Constance F. Citro, editors (2004). *Measuring Racial Discrimination*. Panel on Methods for Assessing Discrimination, Committee on National Statistics, National Academy of Science. Washington DC: National Academies Press, 2004.

Cain, Glen C. (1986). "The Economic Analysis of Labor Market Discrimination: A Survey." In Orley Ashenfelter and Richard Layard, editors. *Handbook of Labor Economics* volume 1. New York: Elsevier, 1986: 693-785.

Ginther Donna K. and Kathy J. Hayes (2003). "Gender Differences in Salary and Promotion for Faculty in the Humanities 1977-95." *Journal of Human Resources* 38 (Winter 2003): 34-73.

Ginther Donna K and Shulamit Kahn (forthcoming). "Women's Careers in Academic Social Science: Progress, Pitfalls and Plateaus." In Alessandro Lanteri and Jack Vromen, editors, *The Economics of Economists*. Cambridge: Cambridge University Press, forthcoming 2014.

Haignere, Lois (2002). *Paychecks: A Guide to Conducting Salary Equity Studies for Higher Education Faculty*. Albany NY: American Association of University Professors, 2002.

Hoffman, Emily P. (1976). "Faculty Salaries: Is there Discrimination by Sex, Race and Discipline?" *American Economic Review* 66 (March 1976): 196-198.

Johnson George P. and Frank P. Stafford (1975). "The Earnings and Promotion of Women Faculty." *American Economic Review* 64 (December 1974): 888-903.

Malkiel, Burton G. and Judith A. Malkiel (1973). "Male Female Pay Differentials in Professional Employment." *American Economic Review* 63 (September 1973): 693-705.

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

Ornstein, Michael and Penni Stewart (1996). "Gender and Faculty Pay in Canada." *Canadian Journal of Sociology* 21 (Autumn 1996): 461-481.

Ranson, Michael R. and Sharon B. Megdal (1993). "Sex Differences in the Academic Labor Market in the Affirmative Action Era." *Economics of Education Review* 12 (Spring 1993): 21-43.