Appendix B: Data

FACULTY SAMPLE: The study sample includes all 1519 UCB ladder-rank faculty with salary data, as of December 15, 2014. All professorial titles (Assistant, Associate, and Full Professor) in all salary scales are included, including all acting professorial titles. The sample includes job codes: 1100, 1104, 1110, 1143, 1144, 1180, 1180L, 1182L, 1200, 1210, 1243, 1300, 1307, 1343, and 1977. The data are derived from UCB's Human Capital Management (HCM) system, and they are merged with historical data extracted from the earlier academic personnel system (APS) that spans 1979-2002. They are also supplemented with data maintained by the Vice Provost for the Faculty on retention actions, dating back to AY 1998-1999; along with additional data on citations that have been pulled from Google Scholar by a graduate-student researcher working under the direction of Professor David Card during the summer of 2014.

SALARY or DEPENDENT VARIABLES: Two salary variables are used in this study: (1) annualized salary as of December, 15, 2014; and (2) the natural log of annualized salary as of December 15, 2014. Both variables are annualized at a 1.0 Full-Time Equivalency (FTE) rate to ensure that faculty salaries can be effectively compared across salary scales (academic year and fiscal year) and that part-time and full-time faculty salaries are treated similarly in the study.

INDEPENDENT VARIABLES:

Demographic Variables	
Women	1=women; 0=men.
Asian	1=Asian; 0=not Asian.
Underrepresented minority	1=URM (African American, Hispanic, Native American); 0=not URM.
Minority	1=Minority (Asian or URM); 0=not minority. Note: This variable is used at the unit-level when URM or Asian headcount is below 10.
Unknown ethnicity	1=Unknown ethnicity; 0=Ethnicity is known. As of the final data draw, 33 male faculty and 13 female faculty had not disclosed their race-ethnicity, with most relatively new to the campus. This is used in regression runs when 10 or more individuals have unknown ethnicity. If less than 10 individuals have unknown ethnicity, the categorical variable is not included and they are grouped in the below white reference category.
White	Exempted reference category for other race/ethnic variables.
Experience-Related Variables	
Time since highest degree	Time elapsed since highest degree (derived from degree dates listed in the personnel records, supplemented with web searches if needed).

Time since highest degree (squared)	The square of time elapsed since highest degree. Included and retained in regression models when the term is significant for the log salary model (or increases model fit in unit-analysis). As per Haignere (p. 38) and others, it allows time since highest degree to fit as a curvilinear or quadratic relationship.
Time since hired as a faculty member at UCB	Number of years since first hired at UCB as faculty (derived from electronic records; or paper files/records if hired before 1979).
Bachelor's/Master's degree	1=Bachelor's/Master's is highest degree; 0=Other degree is highest degree (a doctorate or a professional degree [e.g., JD, MD]). Included in the model if 10 or more individuals have a Bachelor's/Master's as their highest degree. This is more common in the arts & architecture.
Law degree	1=Law degree (JD, or similar law degree); 0=No law degree. If the faculty member has a law degree, this is coded 1, even if they also have another degree such as a PhD.
MD	1=Medical degree (MD); 0=No MD. If the faculty member has an MD, this is coded 1, even if they also have another degree such as a PhD.
Doctoral degree/other	Exempted reference category. The vast majority of UCB faculty have a PhD degree as their highest degree.
Field-Related Variables	
Departmental variables	All departments where there are 10 or more faculty with any level of appointment are coded (1 [a single appointment], .5 [a double joint appointment], or .3333 [a triple appointment]=yes, full or partial appointment in the department; 0=no, not in the department) for each data run. Both the log and total population models typically include more departmental variables because they include all faculty in the regression analysis; in contrast, the white-male only runs have fewer departmental variables. As referenced above, multiple appointment faculty with a double appointment are assigned a value of .5 for each of the two departments in which they hold a position; individuals with a triple appointment are assigned a value of 1/3 or .3333 (repeating) for each of their departments, etc. This fractional approach resulted in the best model fit of the options explored by the committee.
Exempted departments	Exempted fields are noted at the bottom of all regression tables.
Subfields in Haas Business	Subfields in Haas School of Business, as noted on their official webpage: http://facultybio.haas.berkeley.edu/faculty-group. These subfields are only used in Haas specific data runs. They are coded: 1=yes, in the subfield; 0=no, not in the subfield.

Multiple appointments	1=faculty with more than one departmental appointment; 0=faculty with a single departmental appointment. In the campus wide data runs, there were enough faculty to code separately for faculty with triple appointments and double appointments, e.g., see Appendix C, Table A1.
Multiple appointment in law, business, or economics	1=faculty with a multiple appointment in law, business, or economics; 0=no appointment in these high-paying departments. This is used in the L&S Social Science regressions to help account for these types of appointments that are associated with higher salaries. Other units did not have the requisite number of cases (10+) to include this variable.
Other multiple appointments	1=faculty with multiple appointments, not including law, business, and economics; 0=faculty without multiple appointments, or a multiple appointment in law, business, or economics. This is only used in the L&S Social Sciences regression runs when paired with the above referenced law/business/economics multiple appointment variable.
AAUDE market ratio	A market ratio derived from AAUDE peer university salaries; with specific departmental salaries compared to overall mean salaries by broad rank (e.g., mean salary of AAUDE peer full professors in Physics/mean salary of all AAUDE peer full professors). It is used when the small number of cases restricts the effective use of departmental variables. This is especially important when a few faculty might have an appointment or joint appointment in other high-paying departments, including law, business, and economics. It is used in unit-level runs for Haas School of Business (see Appendix C, Tables C8a-C8d, when Haas subfields were not included), Boalt School of Law (C9a-C9d), MCB (D1), Sociology (D2), and School of Public Health (F3). It is also used in detailed rank runs where small n's limited the use of departmental variable (E2-E6, only included in the final Submodel 3b or 4b). For more information, see detailed note on the construction of the market ratio following this table.*
Rank-Related Variables	
Assistant Professor	1=Assistant Professor; 0=not an Assistant Professor.
Associate Professor	1=Associate Professor; 0=not an Associate Professor.
Assistant/Associate Prof.	1=Assistant or Associate Professor; 0=not an Assistant or Associate.
Full Professor	Exempted reference category when above Assistant, Associate, or Assistant/Associate variables are included in the regression model—without additional rank-step designations as below.

Full Professor, steps 1-5	1=Full Professor, steps 1-5; 0=not a Full. Prof., Steps 1-5.
Full Professor, steps 6-9	1=Full Professor, Steps 6-9; 0=not a Full. Prof., Steps 6-9.
Full Professor, Above Scale	Exempted category when Assistant, Associate, Full Prof. 1-5, and Full Prof. 6-9 are included in the regression model. This is only used in the detailed regression runs related to rank, see Appendix C, Table E1.
Rank-step categorical variables	A large number of categorical variables specifying different ranks and steps. In all cases, 10 or more faculty are in each rank-step. See Appendix C, Table E1, for the use of these categorical variables in Submodels 5a and 5b. Full Professor, Above Scale, are the exempted reference category when rank-step categorical variables are used.
Years in rank	Years in current rank (dependent on which rank variables are used).
Years in step	Years in current step. This is only used in Appendix C, Table E1, Submodel 5b, when all rank-step categorical variables are used.
Citation-Related Variables	
Log of Google citations	Natural log of total Google citations of faculty in selected subfields (MCB, Sociology, Psychology, and Haas Business). The data were collected by a graduate student researcher working under the direction of Professor David Card, Summer 2014. Some faculty could not be effectively linked to listed Google Scholar citations, and faculty who were not present during academic year 2013-2014 are missing from these sub-studies. These data were used only in the regression runs included in Section D of Appendix C. The results from this sub-study are also shown in Figures 10A-10D in the Report.
H-index of Google citations	This is an H-index based on Google Scholar citations as of spring 2014 for a sub-study of the School of Public Health. It is only included in Appendix C, Table F3 and Figure 12B in the Report. Past salary studies suggest the log of citations and H-indexes perform similarly in salary regressions.
Retention-Related Variables	
Any job offer since 1998	1=Any outside job offer since 1998 that resulted in a retention case (as recorded by the Vice Provost for the Faculty [formerly the Vice Provost for Academic Affairs and Faculty Welfare]); 0=No outside job offers.
Any top-ten job offer	1=Any outside job offer since 1998 from a top-ten program (as per US News & World Reports); 0=no top-ten outside job offers. This is only included in the sub-study of School of Public Health; see Appendix C, Table F3; and Figure 12B in the Report.

Other job offers	1=One or more job offers since 1998 from non-top-ten program (as per US News & World Reports); 0=no job offers or a job offer from a top-ten program. This is only included in the sub-study of School of Public Health; see Appendix C, Table F3 and Figure 12B in the Report. It is only used when the top-ten job offer variable is used.
Salary increase if outside job offer were to be matched	The calculated difference between the outside job offer's annualized salary and the individual's UCB annualized salary at the time of the outside job offer. This is only included in the sub-study of School of Public Health; see Appendix C, Table F3.

*The AAUDE market ratio is constructed using AAUDE peer private faculty salary data, 2012-2013. In a few departments, private and public AAUDE peers are combined to have the requisite number of institutions to assure reliability of findings (primarily in the College of Natural Resources). We received the most recent data from UCB Office of Planning and Analysis who works with AAUDE in the development of the data and develops reports on salary peer data for the UCB campus. The peer institutions have been chosen by the Deans of relevant units. For each UCB faculty member, we constructed a market ratio by comparing the AAUDE peer average salary data for their department/field to the average salary data for all AAUDE peer faculty at the same faculty rank (e.g., average AAUDE private full professor business faculty salaries/all full professor salaries at AAUDE privates; or $269,461/$192,887=1.397 market ratio for business faculty at full professor rank). Market ratios are typically constructed in this manner using external market data; and they are commonly included in faculty salary studies at other colleges and universities (Luna 2007; Corcoran, Courant, and Raymond 2001; Williford and Gray-Little 2002). In the case of multiple appointment UCB faculty, all of the relevant market ratios are captured for each of their departments and then averaged to create a single composite market ratio for the faculty member. This approach of averaging market ratios was arrived at by examining data for all UCB faculty and determining which approach fit best in regression models for the entire campus based on adjusted r-square values. The AAUDE market ratio is only used when the relatively small number of faculty in a particular regression run made it inappropriate to include departmental unit variables, or the number of exempted departments was so great that the specification of field in the models was less than desirable. This was important in cases where a small number of faculty (n<10) might hold an appointment or joint appointment in high-paying fields, including business, law, and economics; and the necessary categorical field variables could not be introduced into the model due to small n's.

References

Corcoran, Mary E., Paul N. Courant, and Pamela A. Raymond. 2001. University of Michigan gender salary study: Summary of initial findings. Ann Arbor, Michigan: University of Michigan. As retrieved from http://www.provost.umich.edu/reports/U-M_Gender_Salary_Study.pdf.

Haignere, Lois. 2002. Paychecks: A guide to conducting salary-equity studies for higher education faculty. Second edition. Washington, DC: American Association of University Professors.

Luna, Andrew L. 2007. Using a market ratio factor in faculty salary equity studies. AIR Professional File 103: 1-12.

Williford, Lynn, and Bernadette Gray-Little. 2002. Report on the 2002 faculty salary equity study: The University of North Carolina at Chapel Hill. Chapel Hill, North Carolina: University of North Carolina. As retrieved from http://oira.unc.edu/files/2013/04/2002-faculty-salary-equity-study.pdf.