Appendix A. Statistical Models

This appendix describes the statistical models used in this study. These include the multivariate logit analyses used in Phase I and the multivariate hazard analyses used in Phase II.

We adopted a 95 percent confidence level for rejecting the null hypothesis throughout this report. A 95 percent confidence level means that there is a 5 percent chance of incorrectly rejecting the null hypothesis and concluding that the differences observed in the sample are significant.

We followed the statistical convention of testing the null hypothesis. In this case, the null hypothesis states that, after controlling for other factors, the probability that an outcome occurs for female doctorate recipients is not different from the probability that the same outcome will be observed for male doctorate recipients. Rejecting the null hypothesis allows us to accept the alternative hypothesis that the probabilities are different. If we can accept the alternative hypothesis, we might infer an association between sex and the outcome (i.e., employment in a tenure-track position, earning tenure, or employment in different academic ranks).

Logit Analysis

The logit analyses allow us to determine whether sex is related to the likelihood that a given outcome will occur after accounting for the contributions of other controlling variables. We do this by comparing the likelihood that a given outcome will occur for female doctorates with the likelihood that the outcome will occur for male doctorate recipients, holding constant other factors that might be related to outcomes.

The general structure of the logit model is

$$\text{Prob}(Y_j) = e^{F_i\alpha_j + X_i\beta_j} / \sum_j e^{F_i\alpha_j + X_i\beta_j}$$

where

$\text{Prob}(Y_j)$ = the probability of observing an outcome;

F = a vector of female variables (i.e., FEMALE and the female interaction variables);

α = a vector of coefficients on F;

X = a vector of controlling variables;

β = a vector of coefficients on X;

i references individuals; and

j references outcomes.

The Phase I estimates presented in Sections 3 and 4 of this report are the marginal relations between the female variables and the likelihood of an outcome occurring and are not the estimates of the α and β. These marginal relations, typically referred to as "marginal effects," are given by the partial derivative of the probability of the outcome occurring with respect to the female variables. For example, the marginal effect of the k^{th} female variable is given by

$$\text{Marginal effect}_k = \partial \text{Prob}(Y_j) / \partial F_k$$

computed at the sample means of X and F. We computed the marginal effects of the controlling variables analogously.[1] We provide complete estimates of the logit models in Appendix C. Note that the estimates reported there are the marginal effects and are not estimates of the α and β. We provide an alphabetical list of variable acronyms in Appendix B.

Note that the logit models for the tenure track and tenure analyses are binomial in that only two outcomes are possible (e.g., the individual is either tenured or not tenured). However, the logit models for the academic rank analyses are multinomial in that several (three) outcomes are possible (i.e., junior ranks, associate professor rank, or full professor rank).

We generally included cases when observations were missing for independent variables and included missing dummy variables as controls. However, we excluded cases in which information required to define the dependent variable, (i.e., career outcomes) was missing.[2]

---

[1]The logit models were estimated using LIMDEP ver. 7.0. See Greene, 1995.

[2]We followed this convention in all the multivariate analyses conducted for this study.

HAZARD ANALYSIS

Hazard analysis allows us to estimate the likelihood that an outcome will be observed for an individual at any given time, conditional on the fact that the outcome has not occurred previously for that individual. We employed the Cox proportional hazard model in our Phase II analyses. The structure of the model is

$$h(t, F, X) = h(t,0,0)e^{\alpha' F + \beta' X}$$

where

t = time (years since doctorate);

h(t,F,X) = the hazard rate at time t, conditional on F and X;

h(t,0,0) = the baseline hazard rate; and

all else is as previously defined.

The probability of an outcome occurring for an individual at time T^* can be written

$$\text{Prob}(i, j \mid T^*) = e^{\alpha' F_i + \beta' X_i} / \sum_{i \in R_k} e^{a' F_i + \beta' X_i}$$

where R_k is the set of individuals with durations greater than or equal to T^*, and all else is as previously defined.

In Sections 3 and 4 of this report, we present Phase II estimates of the marginal relations between the female variables and the likelihood of outcomes occurring. These marginal relations, typically referred to as "relative risks" (e.g., the risk of moving from the untenured state to being tenured), give the ratio of the probability of an outcome for a surviving (e.g., untenured) female doctorate recipient to the probability for a similarly situated male doctorate recipient. For example, for the FEMALE[3] variable, the relative risk is given by

$$\text{Relative risk}_F = \text{Prob}(j \mid F = 1) / \text{Prob}(j \mid F = 0) = e^{\alpha_F}$$

Although we report relative risks in the tables presented in Sections 3 and 4 of this report, the tables in Appendix D report the estimated coefficients of the hazard function (i.e., the α and β). These can be converted to estimates of relative risks by exponentiation of the estimated coefficients.

[3] FEMALE = 1 if individual is female; FEMALE = 0 if not.