Appendix G. Analysis of Publication Counts in the Fixed Journal Set

G.1 Background Statistics on Fractional Publications Counts in a Fixed Journal Set

For publications measured by fractional counts in a fixed journal set, 95% of the observations by institution-year are between 27 and 1,942, the mean number per institution-year is 546, the median is 368, and the standard deviation is 529. Average publications counts vary by field. The average number of publications per institution-year by field group is as follows: biology-life sciences-agricultural sciences (2,931), medical sciences (1,723), engineering-math-physical sciences (1,493), social sciences-psychology (256), and computer science (13).

G.2 Analyses of Fractional Citation Counts in a Fixed Journal Set

A series of regressions were performed to determine, for each set of personnel and financial variables, which were most highly correlated with fractional citation counts in a fixed journal set. The data set was aggregated to the institution-year level. Financial variables were deflated using the GDP deflator; personnel variables were lagged by one year and financial variables by two years. Analyses were conducted using stepwise linear regression. We typically retained any variable that, when entered into the regression, increased r-squared by at least 0.01. All such variables were highly statistically significant when entered.

The set of faculty counts variables (all lagged by 1 year) in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) Full Profs (0.369), 2) Assoc Profs (0.455), and 3) Total Fac (0.490). These are the same set of variables, and order of entry, as for fractional publication counts in an expanding journal set, and the amount of variance explained is almost identical. We retained these variables for inclusion in later regressions. When only these three independent variables were included, the coefficients were 1.10 for full professors, -3.23 for associate professors, and 0.88 for total faculty. Since addition of an associate professor also increases total faculty, the net effect of adding an associate professor would be -1.25. A negative effect for associate professors was also found in the expanding journal set.

The set of Carnegie classifications and type of control variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) R-1 (0.489), and 2) Medical (0.500). These are the same set of variables, and order of entry, as for fractional publication counts in an expanding journal set, and the amount of variance explained is almost identical.

The set of academic R&D expenditures variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) Total academic R&D expenditures (0.804), 2) other funding sources of academic R&D expenditures (0.816), and 3) federally financed academic R&D expenditures (0.825). These are the same variables as were found in the fractional publication counts analysis in an expanding journal set and the amount of variance explained is almost identical.

The set of enrollment variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) number of S&E grad students (0.463) and 2) fall enrollment of undergraduates (0.527). These are the same variables as were found in the fractional publication counts analysis in an expanding journal set and the amount of variance explained is almost identical. All fall enrollment variables were institution-wide rather than S&E specific. The coefficient for undergraduates was negative (as it was in the analysis of the expanding journal set), perhaps reflecting less publication activity in some institutions with more undergraduate students.

The set of postdoctoral count variables in their order of entry into the regression and cumulative r-squared value after entry were as follows: 1) postdocs without M.D.s supported by federal research grants (0.756), 2) postdocs supported by federal traineeships (.770), and 3) postdocs supported by federal fellowship (0.780). These variables are somewhat different from those found in the analysis of fractional counts in an expanding journal set. However, there is very high correlation among the entire set of postdoctoral counts (with a single principal component containing 82% of the variability among the measures). When we substituted the first three postdoctoral count variables found in the analysis of fractional counts in an expanding journal set (i.e., total postdocs, postdocs supported by federal research grant, and postdocs with M.D.s supported by federal research grant) into the regression of fractional counts in a fixed data set, we found that they account for almost the same amount of variance (0.761). In addition, this is approximately the amount (0.803) found in the analysis of the expanding journal set. We retained the latter three variables to allow easier comparison between the analyses of the fixed and expanding journal set.

In the set of non-faculty research staff count variables, only non-faculty research staff (0.479) entered the regression, mirroring the result in the expanding journal set.

The set of degrees-awarded variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) S&E doctorates from the NSF Survey of Earned Doctorates (0.701), and 2) S&E BA/BS degrees (0.712). These results are essentially the same as that for fractional publication counts in an expanding journal set. The coefficient for BA/BS degrees was negative (as it was in the analysis of the expanding journal set), consistent with the finding for number of undergraduate students.

The complete set of retained variables identified above was entered into a stepwise regression. The variables in their order of entry into the regression and the cumulative r-squared values after entry were as follows: 1) Total academic R&D expenditures (0.825), 2) postdoctorates (0.891), and number of S&E doctoral recipients (0.910). All of the other variables increased r-squared to 0.930. These are the same three variables as found in the analysis of fractional counts in an expanding journal set, and very similar increments to r-squared (i.e., 0.816, 0.926, and 0.936). The coefficients for the number of postdocs and S&E doctoral recipients in these two sets of regressions were similar: postdoctorates (fixed = 0.727, expanding = 0.882), and S&E doctoral recipients (fixed = 1.24, expanding = 1.18). The coefficient for total academic R&D was slightly lower for the fixed journal set (2.49 per $1M) than for the expanding journal set (3.31 per $1M),

although some of this difference reflects the fact that there are approximately 15% more publications in the expanding journal set than in the fixed journal set.

G.3 Analyses of Whole Count Publications in an Fixed Journal Set

We repeated this process for publications as measured by whole counts in a fixed journal set. A set of four variables entered into the regression, including total academic R&D expenditures, the number of S&E doctoral recipients, and two postdoctoral counts — the number of postdoctorates without M.D.s supported by federal research grants and the number of postdoctorates with M.D.s. The total r-squared was 0.925. Essentially the same r-squared (0.919) was achieved by substituting the total number of postdoctorates for the two postdoctoral measures mentioned above. The coefficients for this model were 3.60 per $1M in academic R&D expenditures, 1.38 per postdoctorate, and 1.59 per S&E Ph.D. recipient, with corresponding standard errors of 0.14, 0.03, and 0.07, respectively.

Since both publications as measured by fractional and whole counts in an expanding journal set could be explained with the same three variables — total academic R&D expenditures, postdoctorates, and S&E Ph.D. recipients — we retained these three variables for further analysis.

G.4 Path Analytic Model for Publication Counts

We developed a path analytic model under the assumption that total academic R&D expenditures have a direct and indirect effect on publications, where the indirect effect is through funding of some postdoctorates and Ph.D. graduate students. We also assumed that postdoctorates and S&E Ph.D. recipients have a direct effect on number of publications. A path analytic model allows total academic R&D expenditures to directly influence the number of postdoctorates and S&E Ph.D. recipients, and for total academic R&D expenditures, the number of postdoctorates, and the number of S&E Ph.D. recipients to directly influence the number of publications.

For publications as measured by fractional counts in a fixed journal set, the regression coefficient for total academic R&D expenditures was 2.49 (i.e., 2.49 publications per each $1M in academic R&D expenditures), the coefficient for postdoctorates was 0.727, and the coefficient for S&E Ph.D. recipients was 1.24. We then performed a regression where the dependent variable was the number of postdoctorates and the independent variable was total academic R&D expenditures. We find that each $1M additional funding is associated with an increase of 2.53 additional postdoctorates (with a standard error of estimate of 0.05). A similar regression shows that each $1M additional funding is associated with an increase of 1.39 S&E Ph.D. recipients (with a standard error of estimate of 0.02). Thus, adding together the direct and indirect effects, we find that each $1M in academic R&D expenditures results in 6.05 (i.e., 2.49 + 0.727 x 2.53 + 1.24 x 1.39) publication counts (about $165K per fractional publication count). This is higher than the cost per fractional publication count in the expanding journal set because the fixed journal set contains fewer publications and the amount of academic R&D expenditures, postdoctorates, and S&E doctoral recipients has not been proportionately reduced.

For publication counts as measured by whole counts in a fixed journal set, the regression coefficient for total academic R&D expenditures was 3.60 per $1M, the coefficient for postdoctorates was 1.38, and the coefficient for S&E Ph.D. recipients was 1.59. The coefficient for total academic R&D expenditures regressed on the number of postdoctorates was 2.53 per $1M. The coefficient for total academic R&D expenditures regressed on the number of S&E Ph.D. recipients was 1.395 per $1M. Consequently, adding together the direct and indirect effects, we find that each $1M in additional funding is associated with an increase of 9.3 (i.e., 3.60 + 1.38 x 2.53 + 1.59 x 1.395) publications as measured by whole counts (about $107K per count).