sociological science

Supplement to:

Kim, Lanu, Christopher Adolph, Jevin D. West, and Katherine Stovel. 2020. "The Influence of Changing Marginals on Measures of Inequality in Scholarly Citations: Evidence of Bias and a Resampling Correction." Sociological Science 7: 314-341.

S 1 Aggregation of journals to disciplines and disciplines to fields

Across the Web of Science, journals are classified into one or more disciplines. If either of the first two disciplines listed fell into one of our broad categories, we include the journal in that field. We categorize broad disciplinary fields following National Science Foundation's taxonomy of disciplines created by the Integrated Postsecondary Education Data System survey. Out of 14 available categories, we use four broad categories that are more or less comparable to the analysis of Larivière et al. (2009).

S 1.1 Health

Allergy; Andrology; Anesthesiology; Audiology & Speech-Language Pathology; Cardiac & Cardiovascular Systems; Clinical Neurology; Critical Care Medicine; Dentistry, Oral Surgery & Medicine; Dermatology; Emergency Medicine; Endocrinology & Metabolism; Gastroenterology & Hepatology; Geriatrics & Gerontology; Health Care Sciences & Services; Health Policy & Services; Hematology; Infectious Diseases; Integrative & Complementary Medicine; Medical Ethics; Medicine, General & Internal; Medicine, Legal; Medicine, Research & Experimental; Neuroimaging; Nursing; Obstetrics & Gynecology; Oncology; Ophthalmology; Orthopedics; Pathology; Pediatrics; Peripheral Vascular Disease; Primary Health Care; Psychiatry; Public, Environmental & Occupational Health; Radiology, Nuclear Medicine & Medical Imagin; Radiology, Nuclear Medicine & Medical Imaging; Respiratory System; Rheumatology; Transplantation; Tropical Medicine; Urology & Nephrology; Veterinary Sciences.

S 1.2 Humanities

Art; Classics; Dance; Ethics; Film, Radio, Television; Folklore; History; Humanities, Multidisciplinary; Literary Reviews; Literary Theory & Criticism; Literature; Literature, African, Australian, Canadian; Literature, American; Literature, British Isles; Literature, German, Dutch, Scandinavian; Literature, Romance; Literature, Slavic; Logic; Medieval & Renaissance. Studies; Music; Philosophy; Poetry; Religion; Theater.

S 1.3 Mathematics and computer sciences

Computer Science, Artificial Intelligence; Computer Science, Cybernetics; Computer Science, Hardware & Architecture; Computer Science, Information Systems; Computer Science, Interdisciplinary Applications; Computer Science, Software Engineering; Computer Science, Theory & Methods; Information Science & Library Science; Mathematical & Computational Biology; Mathematics; Mathematics, Applied; Mathematics, Interdisciplinary Applications; Statistics & Probability.

S 1.4 Social sciences

Agricultural Economics & Policy; Anthropology; Archaeology; Area Studies; Asian Studies; Behavioral Sciences; Criminology & Penology; Cultural Studies; Demography; Economics; Ethnic Studies; Family Studies; Geography; Geography, Physical; Gerontology; History & Philosophy Of Science; History Of Social Sciences; International Relations; Language & Linguistics; Linguistics; Political Science; Public Administration; Social Issues; Social Sciences, Biomedical; Social Sciences, Interdisciplinary; Social Sciences, Mathematical Methods; Social Work; Sociology; Urban Studies; Women's Studies.

S 2 Robustness checks

S 2.1 Analysis with a fixed set of journals

In this section, we test the robustness of the analysis using different restrictions on the data. First, in order to assess whether changes in coverage of journals in the Web of Science database influenced the observed time trends, we limit the analysis to journals that published at least one paper continuously between 1996 and 2014 and were included in the Web of Science database during this entire period. Figure S1 shows the total number of papers published in the restricted set of journals by year and field, and the total number of citations to those papers from that restricted set of journals in the following two years. It is worth noting that small declines in total citations are more common using a restricted set of journals, which means we cannot perform corrections for marginals bias in the health field from 2013–2014, in the social sciences from 1997–1998, or in the humanities from 1997–2007.

When we repeat our analyses of citation concentration on this restricted dataset, our results are qualitatively similar to what we report in the main text. Nevertheless, there are some noteworthy new findings. First we review the results that hold in common for the Gini coefficient, the percentage of papers ever cited, the percentage of papers accounting for 20% and 80% of citations, and the Herfindahl-Hirschman Index (HHI). Looking at the unadjusted observed levels of citation inequality (marked *Obs* in the lower half of Figures S2–S6), we note the ostensible trend in the humanities towards greater equality has disappeared, suggesting that it was a function of new journals in conjunction with increasing total publications and citations. Focusing on only continuously published journals also reveals a slight uptick in the concentration in the broad area of health across all inequality metrics, whether adjustment is applied or not. Across all measures, the social sciences and mathematics and computational sciences remain strongly affected by marginals bias: although the unadjusted data suggests declining citation concentrations for these fields on each metric, the adjusted results (marked *Adj*) show that the inequality of citations has not changed for either discpline once marginals bias is removed. On balance, then, focusing on continuously published journals suggests no tendency towards greater equality in any field, and perhaps a movement in the opposite direction for health publications.

Finally, we note that all metrics still seem affected by marginals bias to the extent noted in the main text, with the partial exception of HHI, which is somewhat less biased – but still unreliable without adjustment – when a fixed set of journals is compared over time.

S 2.2 Analysis with longer time windows: citations over four years

In order to assess the sensitivity of our results to the use of a relatively short two-year citation window, we repeat our analysis using four-year citation windows. Because our analysis includes citations made between 1996 and 2016, the four-year citation window only includes papers published up to 2012 (four years before 2016.) Figure S7 shows the total number of papers published in each of these years, by field, and the total number of citations to those papers in the following four years. Total citations increased in every field except the humanities from 1997–2000, which are the only cases that could not be adjusted for marginals bias.

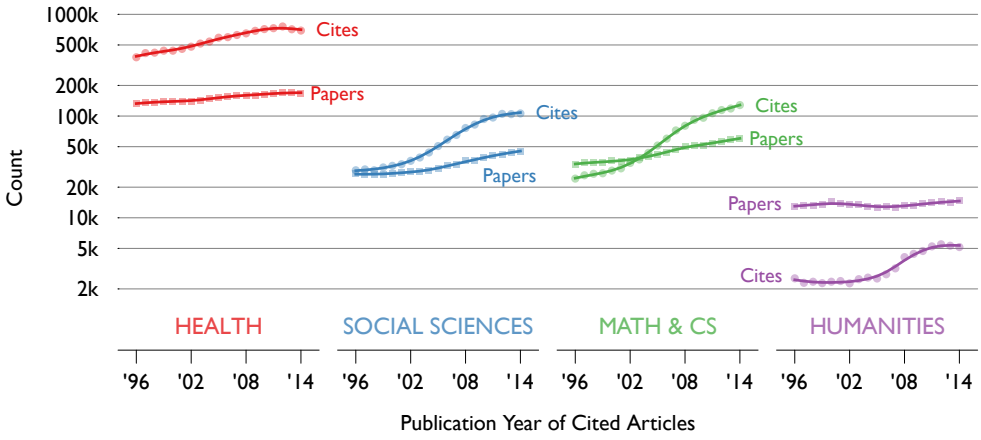

Figure S1. *Number of journal articles published 1996–2014 and citations to those articles within two years of publication with the fixed set of journals continuously published between 1996–2014.* Compiled from the Web of Science (Clarivate Analytics). Trend lines estimated by robust-and-resistant regression to minimize the influence of outliers. All curves are smoothing splines with span of 0.5.

The changes in citation concentration over time observed follow patterns similar to those produced using the shorter window in analyses presented in the main text (Figures S8–S12). The degree of marginals bias grows (almost imperceptibly) smaller as the citation window grows longer – and fewer published papers thus remain close to the zero-lower bound for citations – but does not disappear, suggesting adjustment is still necessary for longer windows. Part of the apparent reduction of marginals bias is also a visual artifact of comparisons across different citation windows. Due to a lack of data past 2016, we cannot report results for four-year windows for papers published after 2012, but these were the publication periods most affected by marginals bias when compared to papers published in 1996. Their omission makes these figures appear less biased because only earlier periods can be compared.)

S 2.3 Analysis with longer time windows: citations over six years

Again, we assess the sensitivity of our results presented in the main text with the two-year citation window, we repeat our analysis using six-year citation windows. Because

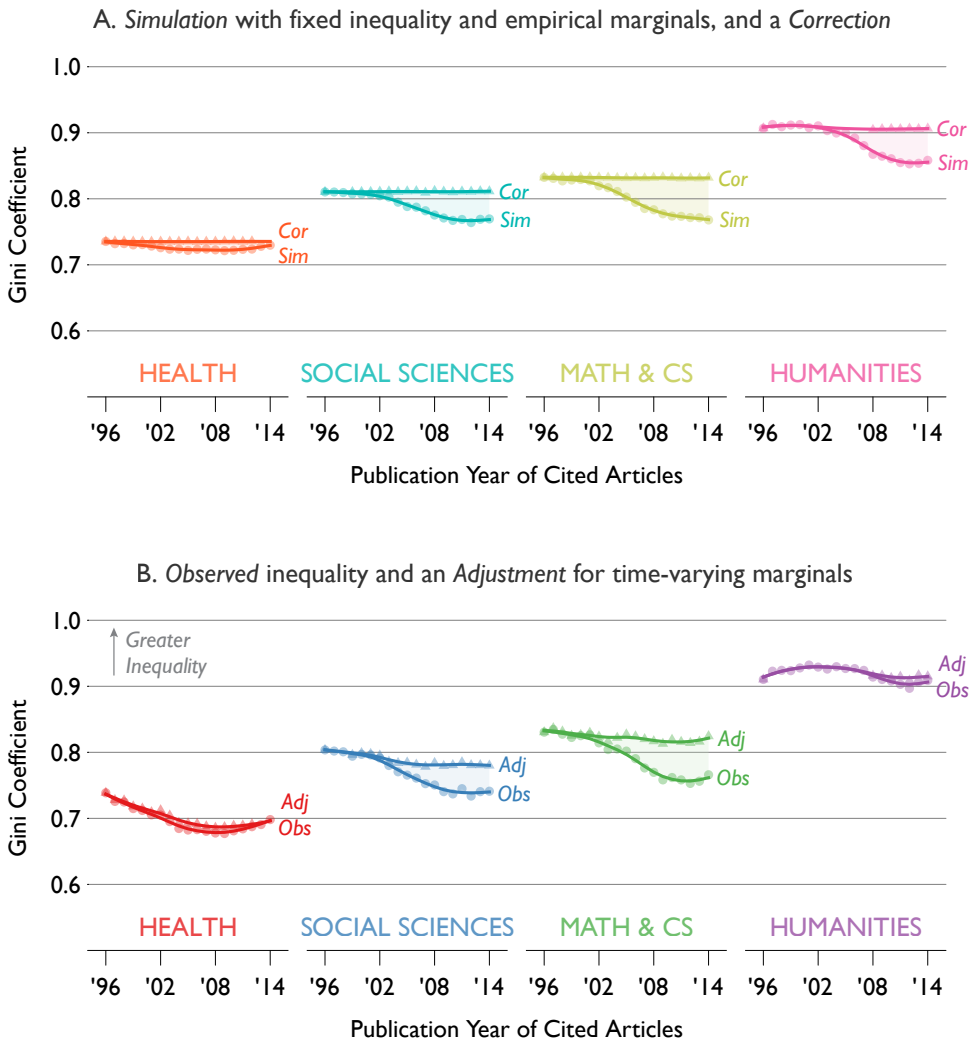

Figure S2. *Gini coefficient for citations within two years of publication with the fixed set of journals continuously published between 1996–2014: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show Gini coefficients of citation distribution from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in the Gini coefficient using a resampling correction. The lines marked *Obs* in the bottom panel shows the Gini coefficients over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the health in 2013–2014, in the social sciences in 1997–1998, and in the humanities in 1997–2007. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

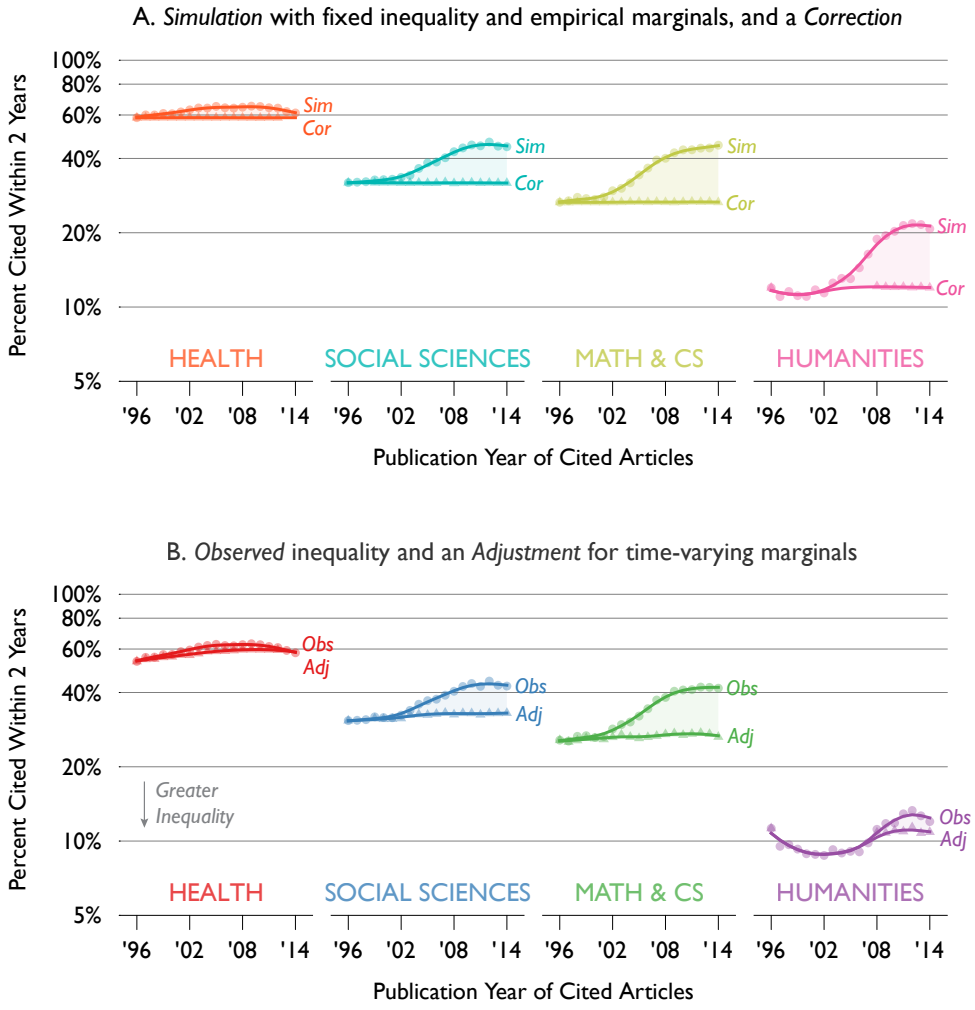

Figure S3. *Percent of papers with any citations within two years of publication with the fixed set of journals continuously published between 1996–2014: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers ever cited from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in percent-ever-cited using a resampling correction. The lines marked *Obs* in the bottom panel shows percent-ever-cited over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the health in 2013–2014, in the social sciences in 1997–1998, and in the humanities in 1997–2007. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

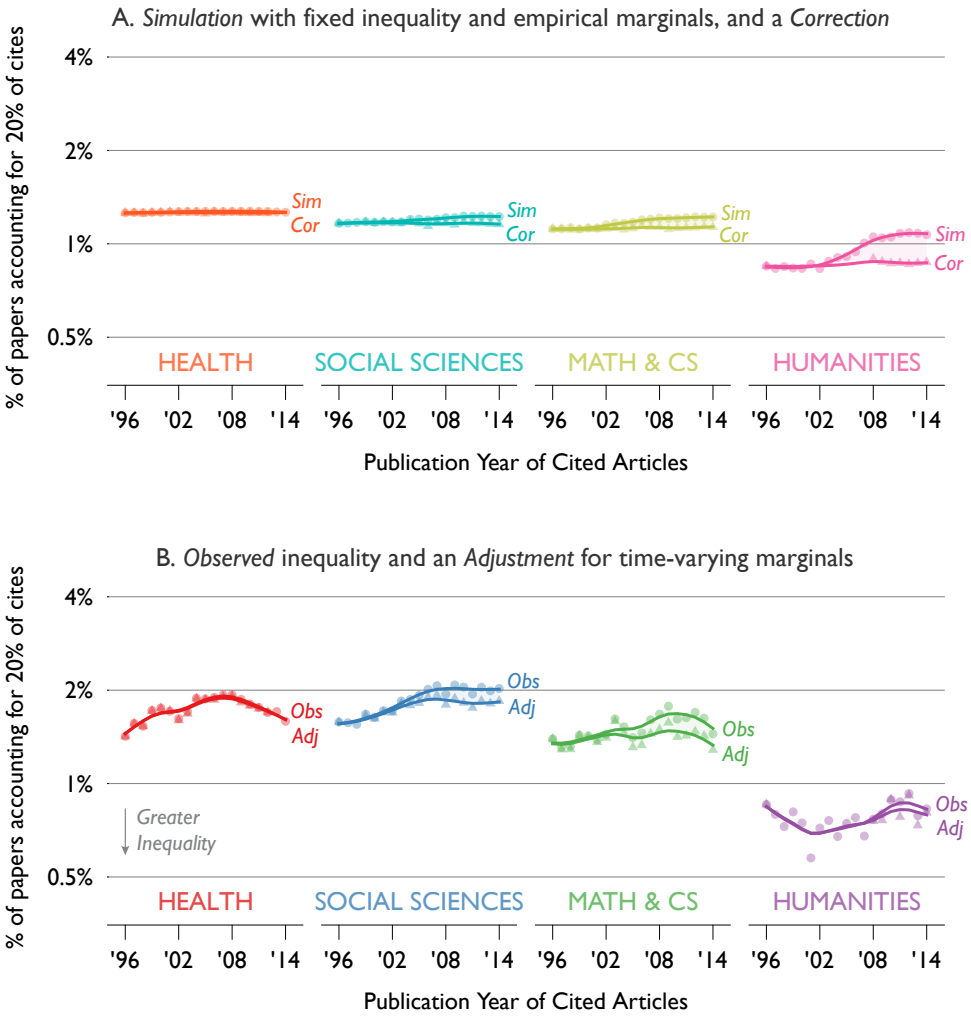

Figure S4. *Percent of papers accounting for 20% of all citations within two years of publication with the fixed set of journals continuously published between 1996–2014: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 20% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows percent of papers accounting for 20% of citations over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the health in 2013–2014, in the social sciences in 1997–1998, and in the humanities in 1997–2007. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

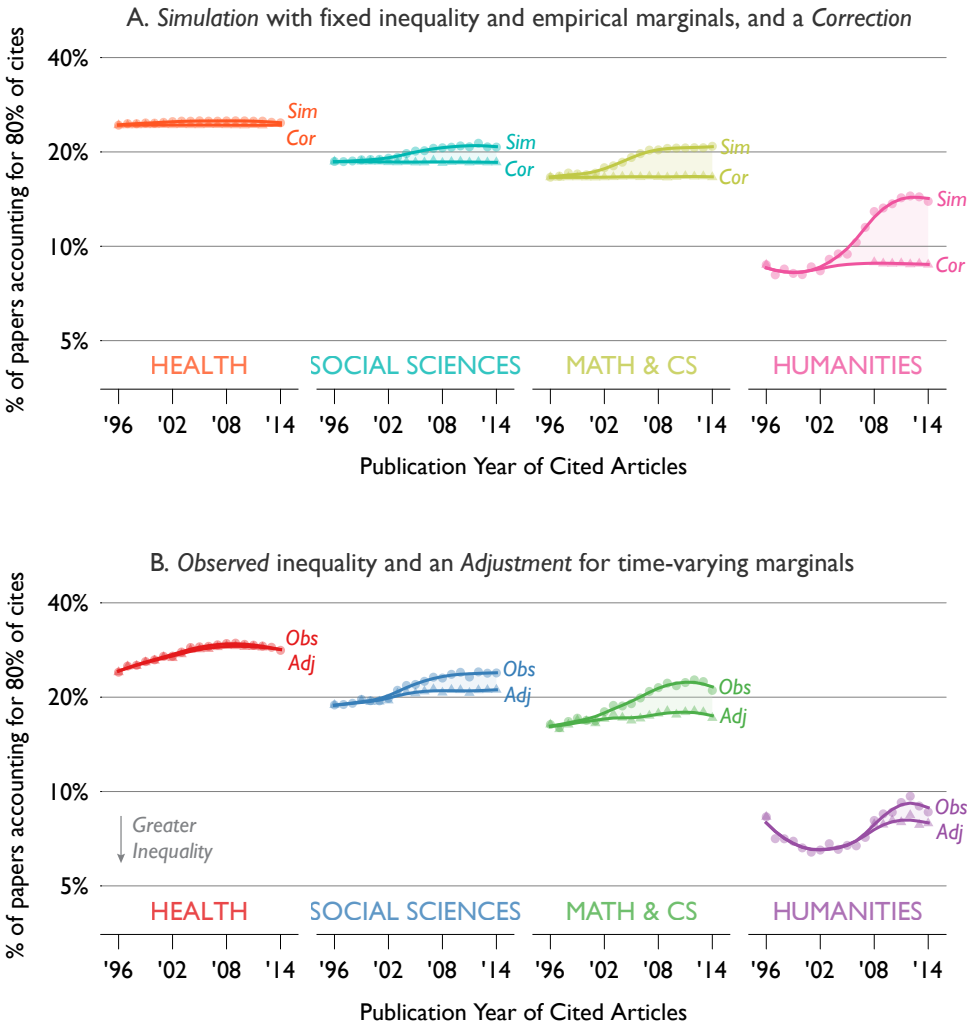

Figure S5. *Percent of papers accounting for 80% of all citations within two years of publication with the fixed set of journals continuously published between 1996–2014: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 80% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows percent of papers accounting for 80% of citations over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the health in 2013–2014, in the social sciences in 1997–1998, and in the humanities in 1997–2007. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

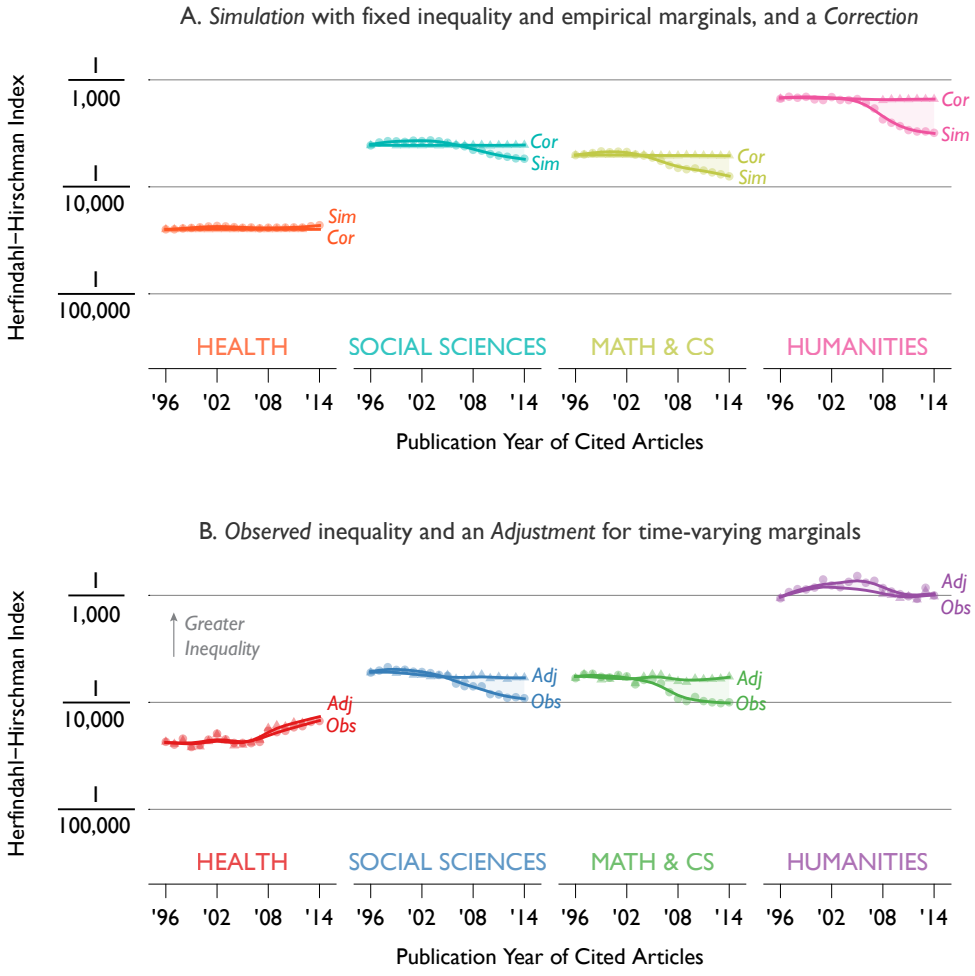

Figure S6. *Herfindahl-Hirschman Index of citations within two years of publication with the fixed set of journals continuously published between 1996–2014: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show the Herfindahl-Hirschman Index (HHI) of citation concentration from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in HHI using a resampling correction. The lines marked *Obs* in the bottom panel shows HHI over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the health in 2013–2014, in the social sciences in 1997–1998, and in the humanities in 1997–2007. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

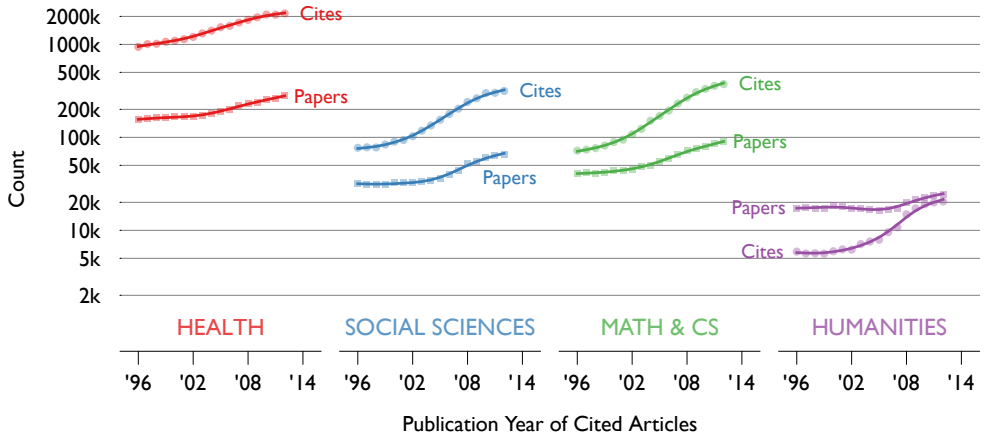

Figure S7. *Number of journal articles published 1996–2012 and citations to those articles within four years of publication.* Compiled from the Web of Science (Clarivate Analytics). Trend lines estimated by robust-and-resistant regression to minimize the influence of outliers. All curves are smoothing splines with span of 0.5.

our analysis includes citations made between 1996 and 2016, the six-year citation window only includes papers published up to 2010 (six years before 2016.) Figure S13 shows the total number of papers published in each of these years, by field, and the total number of citations to those papers in the following six years. Total citations increased in every field except the humanities from 1997–2000, which are the only cases that could not be adjusted for marginals bias.

The changes in citation concentration over time observed follow patterns similar to those produced using the shorter window in analyses presented in the main text (Figures S14–S18). The degree of marginals bias grows smaller as the citation window grows longer – and fewer published papers thus remain close to the zero-lower bound for citations – but does not disappear, suggesting adjustment is still necessary for longer windows.

Part of the apparent reduction of marginals bias is also a visual artifact of comparisons across different citation windows. Due to a lack of data past 2016, we cannot report results for six-year windows for papers published after 2010, but these were the publication periods most affected by marginals bias when compared to papers published in

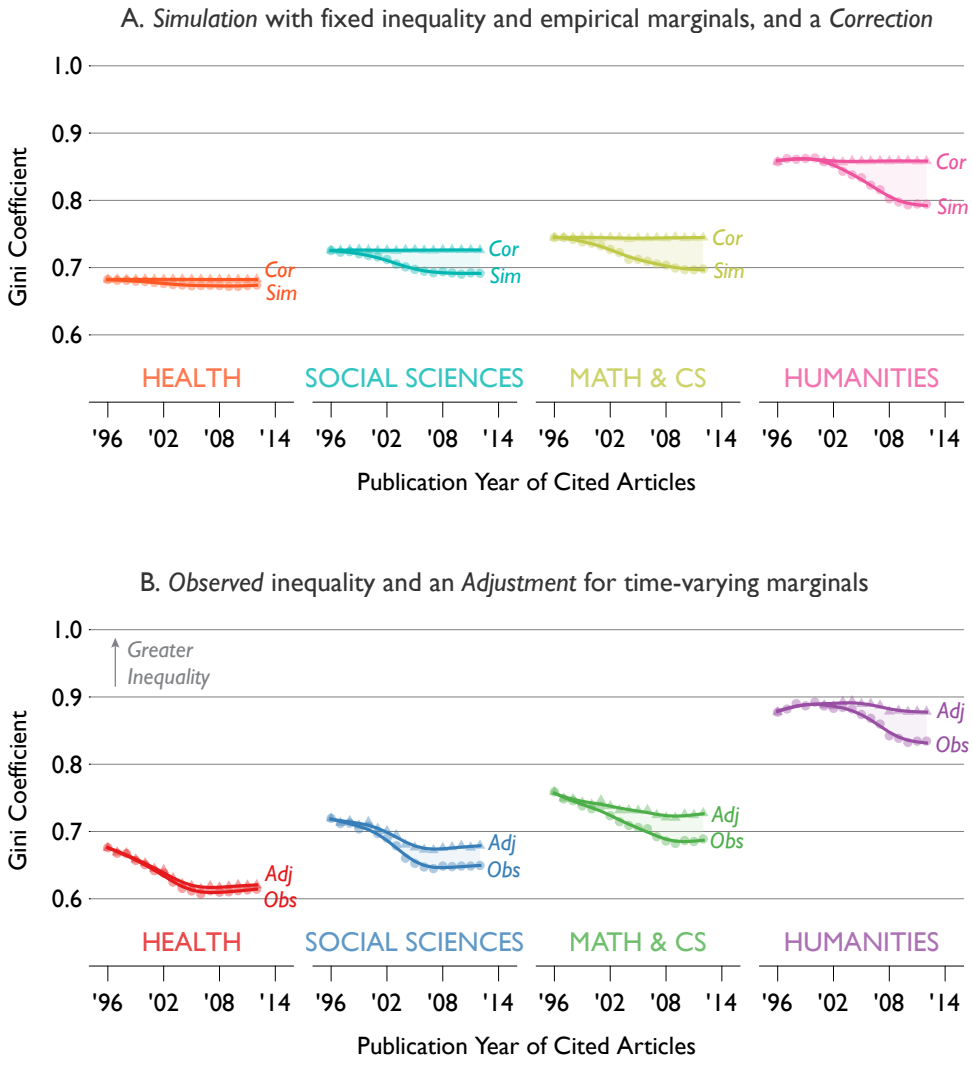

Figure S8. *Gini coefficient for citations within four years of publication, 1996–2012: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show Gini coefficients of citation distribution from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in the Gini coefficient using a resampling correction. The lines marked *Obs* in the bottom panel shows the Gini coefficients over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

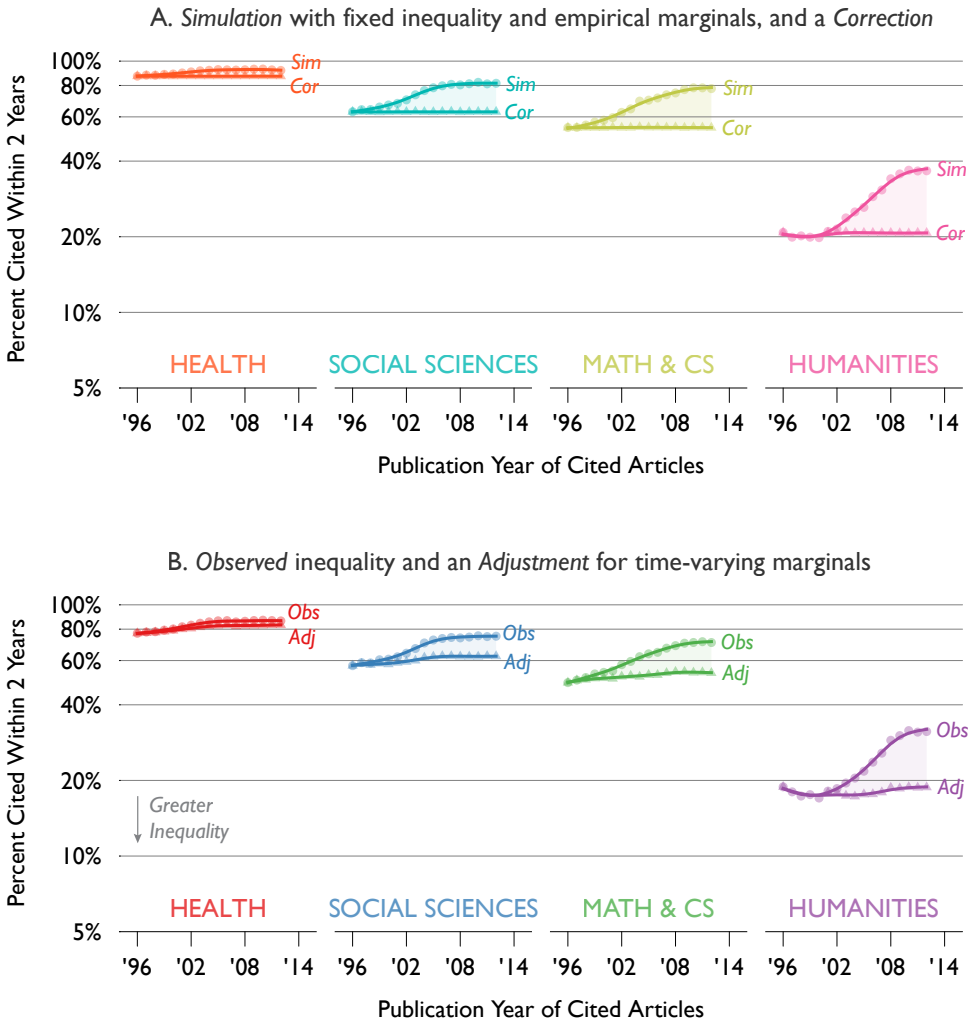

Figure S9. *Percent of papers with any citations within four years of publication, 1996–2012: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers ever cited from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in percent-ever-cited using a resampling correction. The lines marked *Obs* in the bottom panel shows percent-ever-cited over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

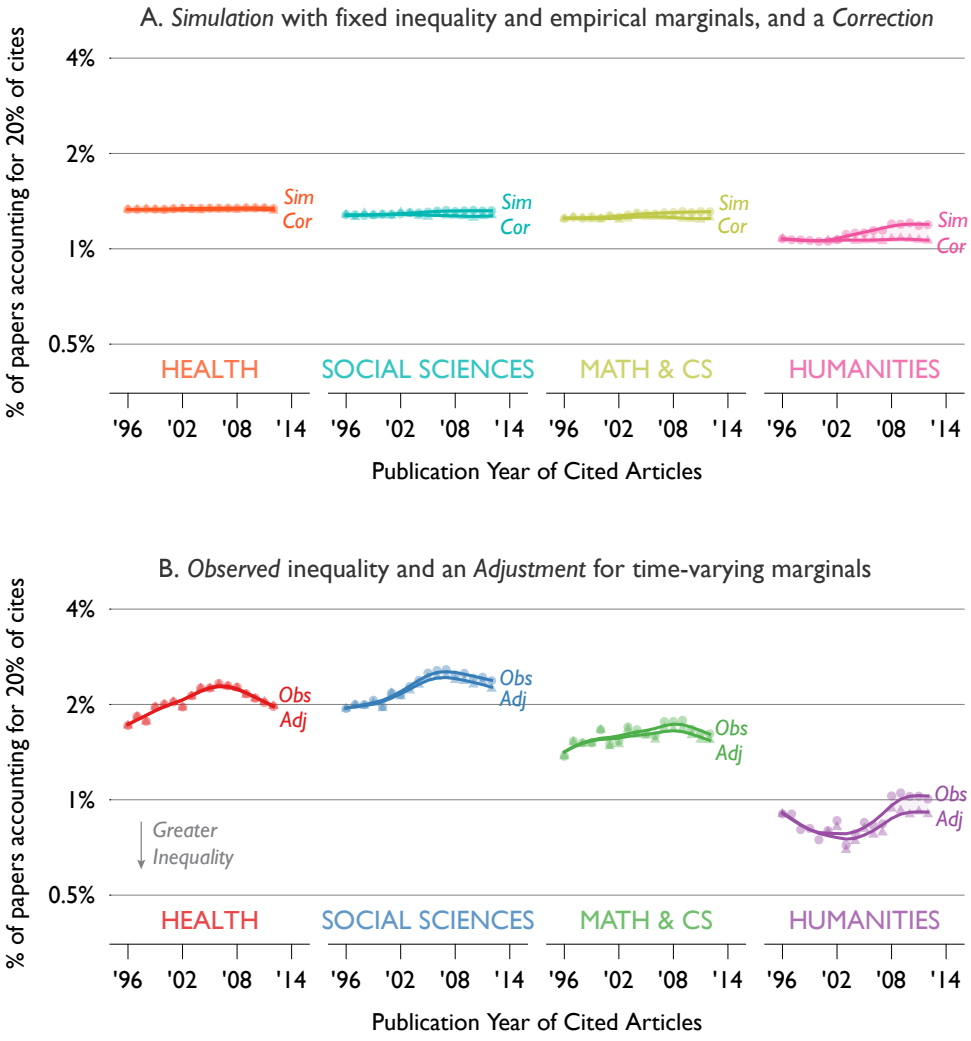

Figure S10. *Percent of papers accounting for 20% of all citations within four years of publication, 1996–2012: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 20% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows percent of papers accounting for 20% of citations over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

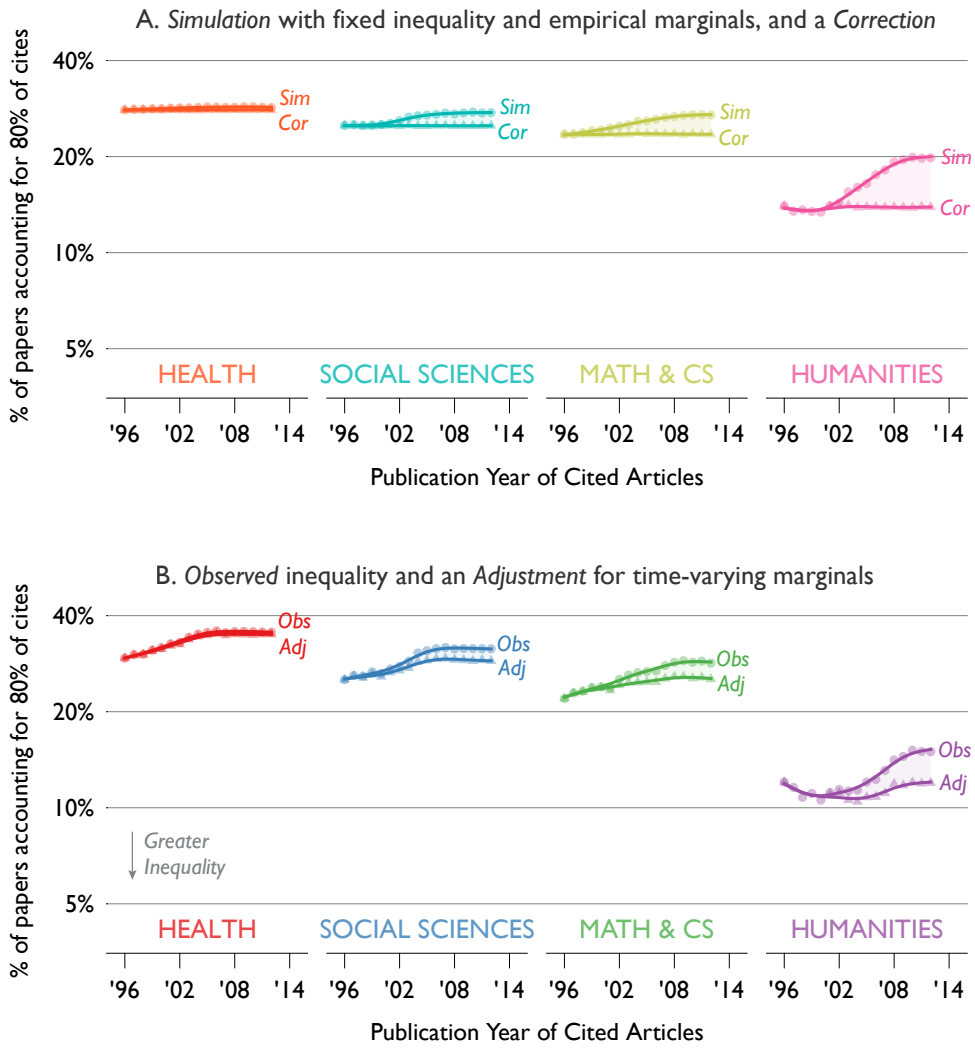

Figure S11. *Percent of papers accounting for 80% of all citations within four years of publication, 1996–2012: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 80% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows percent of papers accounting for 80% of citations over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

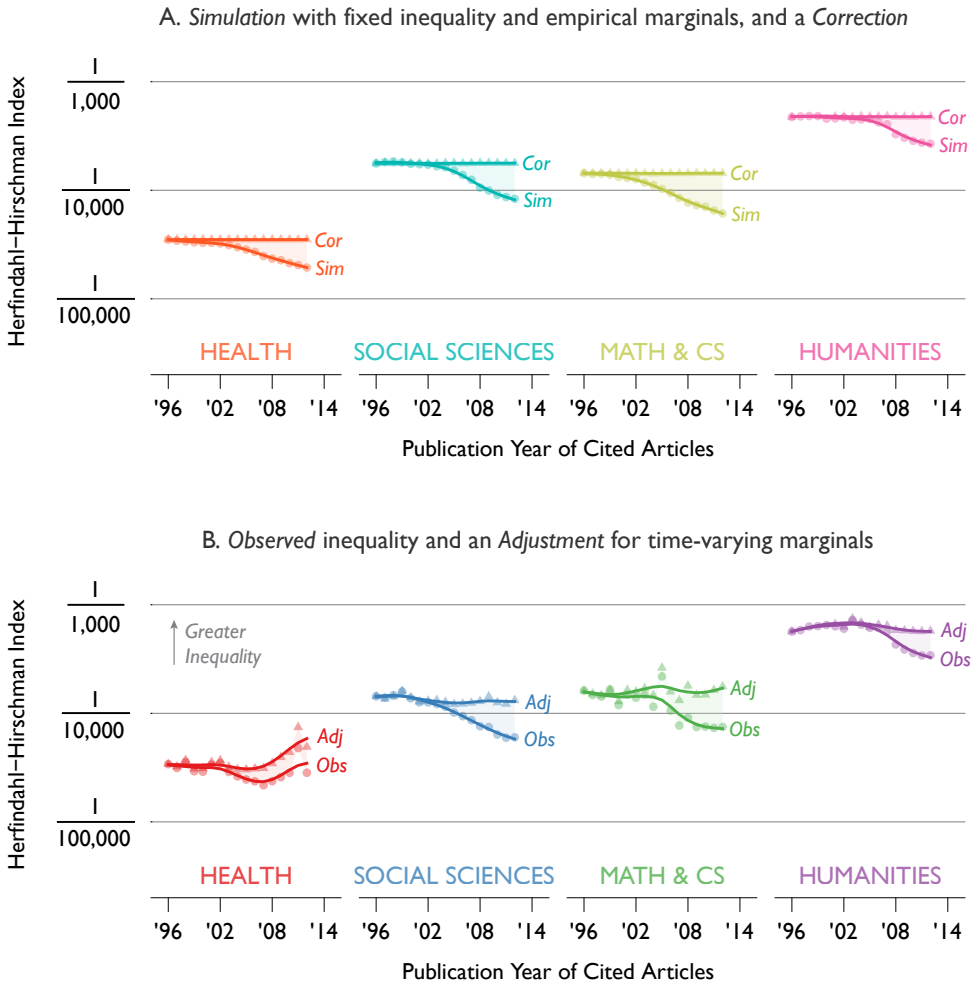

Figure S12. *Herfindahl-Hirschman Index of citations within four years of publication, 1996–2012: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show the Herfindahl-Hirschman Index (HHI) of citation concentration from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in HHI using a resampling correction. The lines marked *Obs* in the bottom panel shows HHI over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

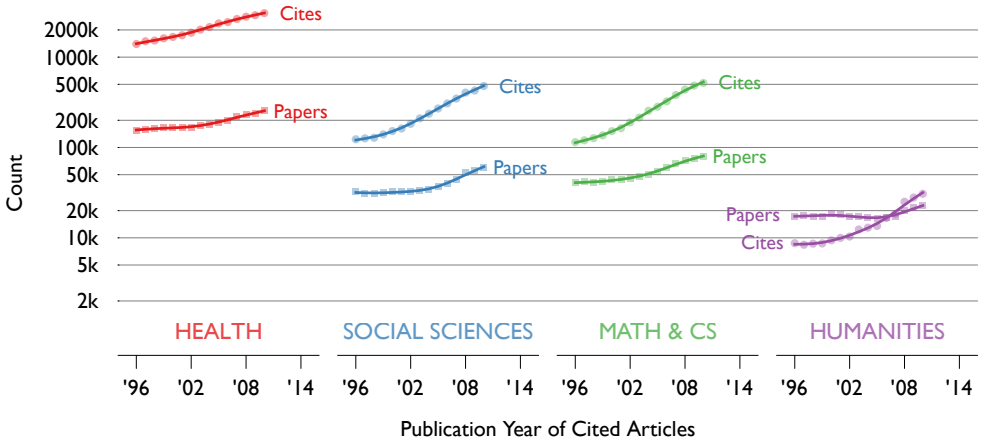

Figure S13. *Number of journal articles published 1996–2010 and citations to those articles within six years of publication.* Compiled from the Web of Science (Clarivate Analytics). Trend lines estimated by robust-and-resistant regression to minimize the influence of outliers. All curves are smoothing splines with span of 0.5.

1996. Their omission makes these figures appear less biased because only earlier periods can be compared.

S 2.4 Analysis including a single outlier paper in mathematics and computer sciences

In our results in the main text, we omit a single unusually highly-cited paper in mathematics and computer sciences. This section shows what happens to our main 2-year window results when we include this paper. With the exception of the Herfindahl-Hirschman Index, including the outlier makes little or no discernable difference (compare the Math & CS plots in the bottom row of main text Figures 3–6 with the corresponding plots in Figures S19-S22 below). However, because Herfindahl-Hirschman indexes are particularly sensitive to extreme cases of concentration, including this single paper produces a strong outlier in the HHI results (compare Figure 7 in the main text to Figure S23 below). The degree to which HHI is influenced by this single outlier is unaffected by our resampling correction.

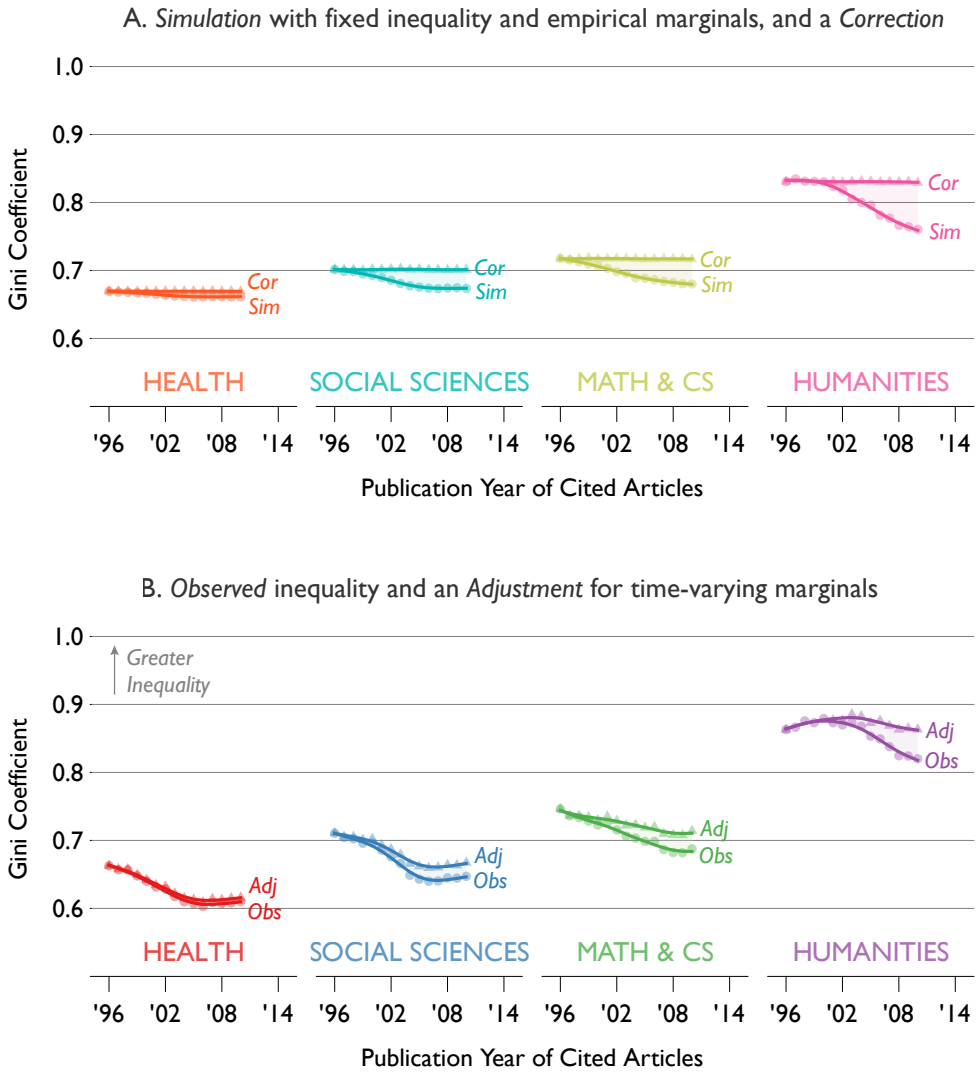

Figure S14. *Gini coefficient for citations within six years of publication, 1996–2010: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show Gini coefficients of citation distribution from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in the Gini coefficient using a resampling correction. The lines marked *Obs* in the bottom panel shows the Gini coefficients over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

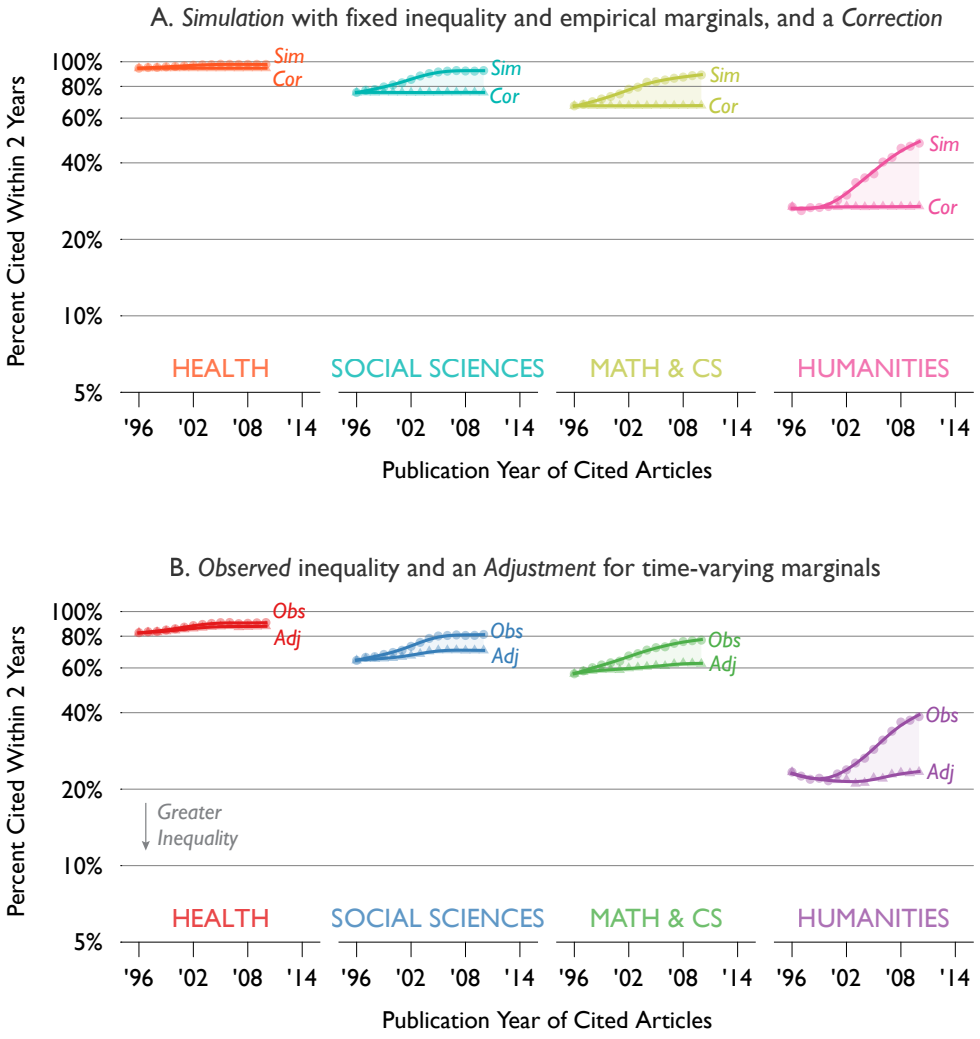

Figure S15. *Percent of papers with any citations within six years of publication, 1996–2010: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers ever cited from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in percent-ever-cited using a resampling correction. The lines marked *Obs* in the bottom panel shows percent-ever-cited over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

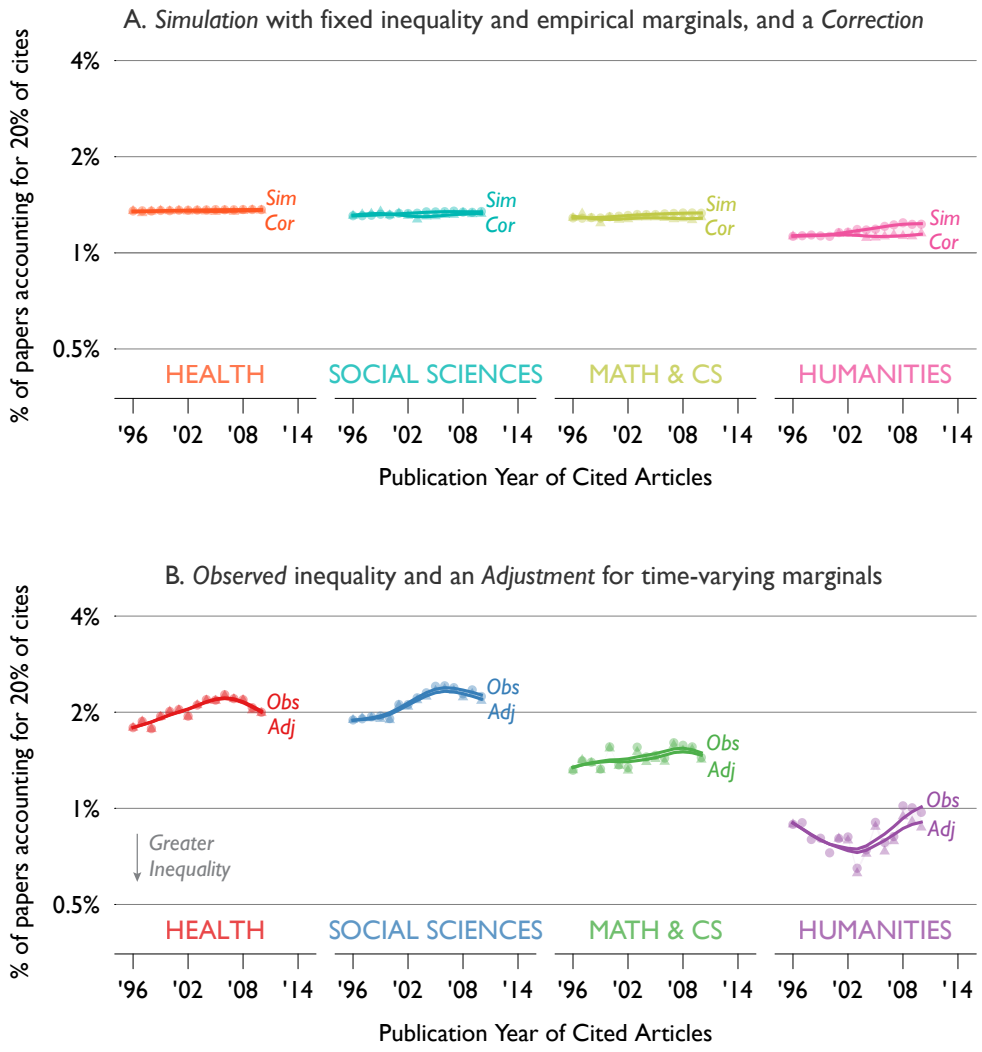

Figure S16. *Percent of papers accounting for 20% of all citations within six years of publication, 1996–2010: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 20% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows percent of papers accounting for 20% of citations over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

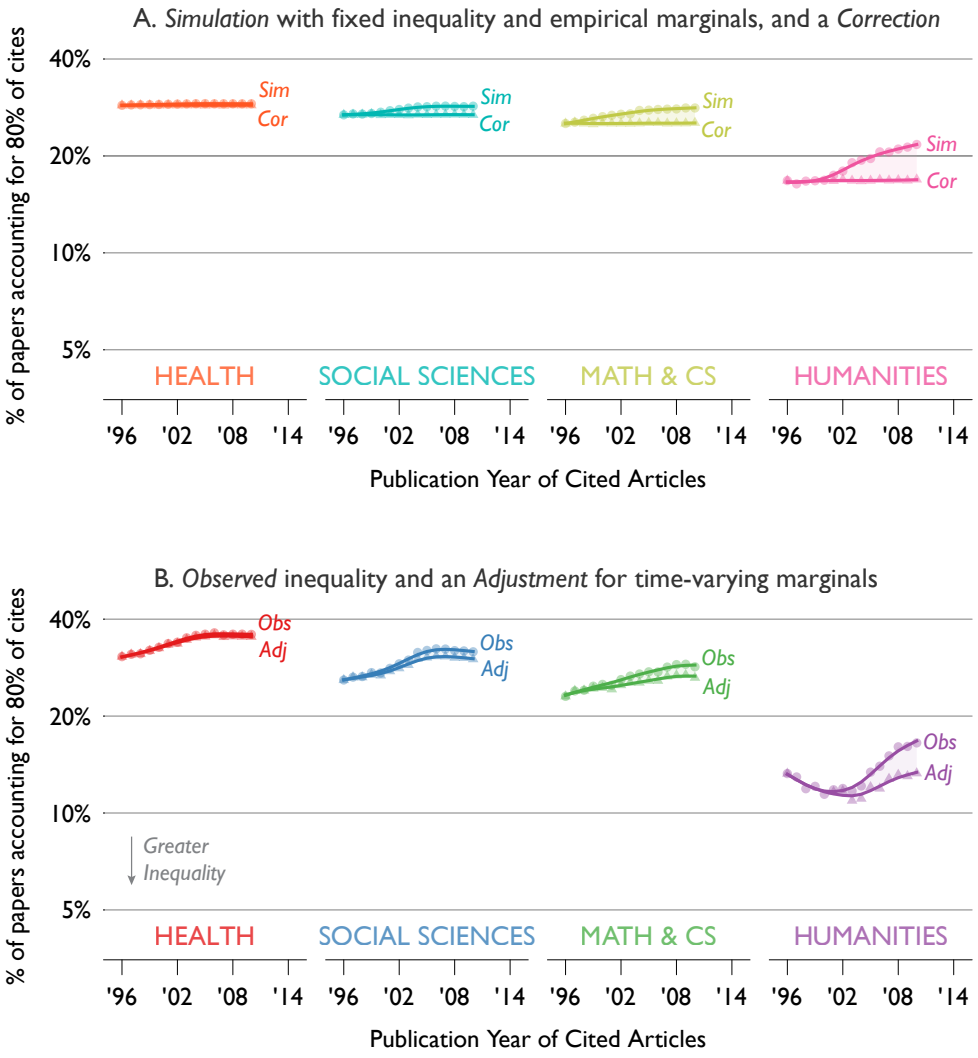

Figure S17. *Percent of papers accounting for 80% of all citations within six years of publication, 1996–2010: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 80% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows percent of papers accounting for 80% of citations over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

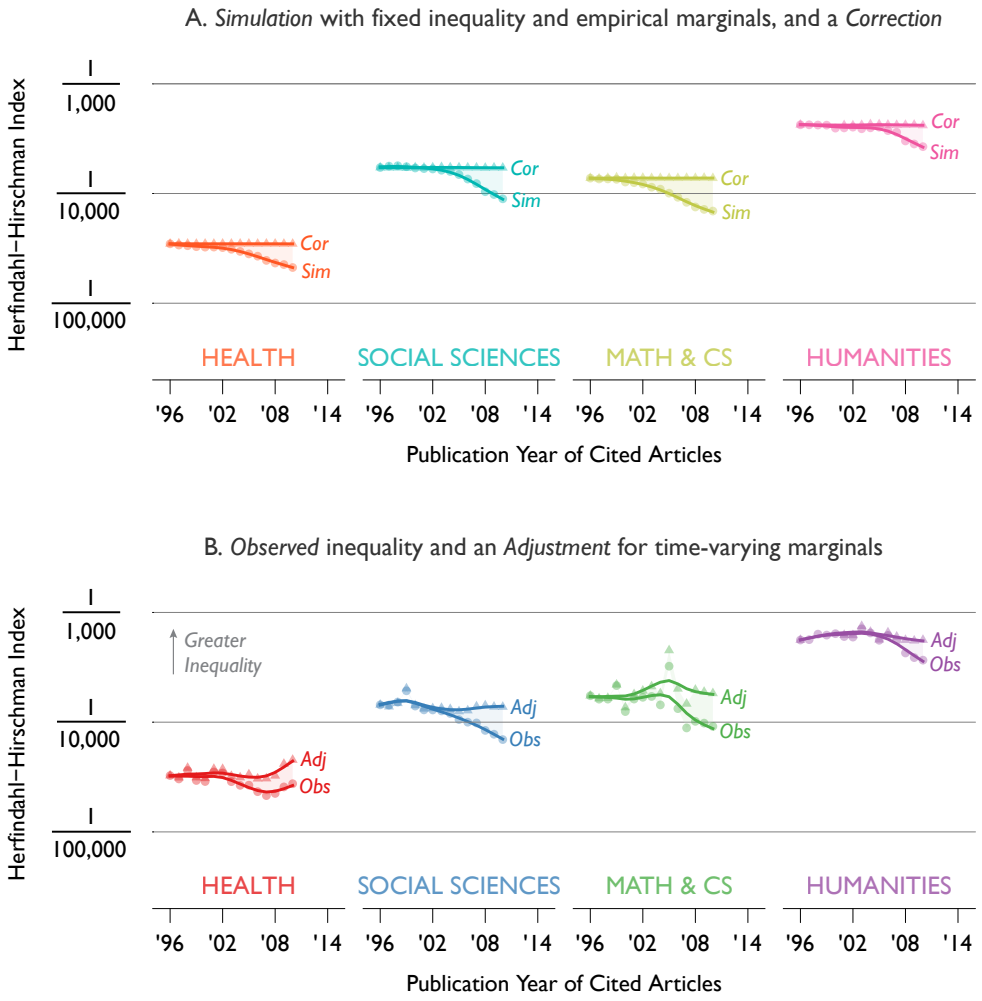

Figure S18. *Herfindahl-Hirschman Index of citations within six years of publication, 1996–2010: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show the Herfindahl-Hirschman Index (HHI) of citation concentration from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in HHI using a resampling correction. The lines marked *Obs* in the bottom panel shows HHI over fields and time using the empirical data from World of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2000. All curves are smoothing splines with span of 0.5. One exceptionally highly cited paper in Math & CS is omitted.

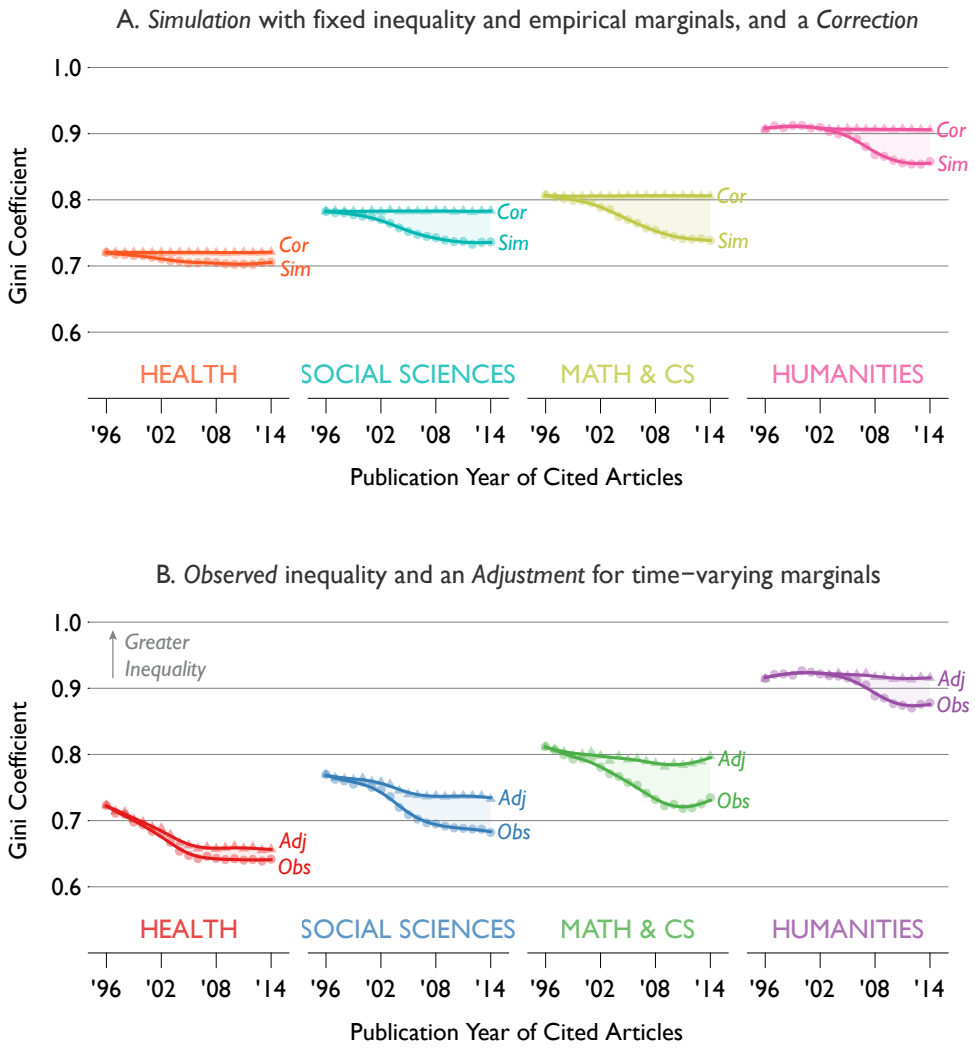

Figure S19. *Gini coefficient for citations within two years of publication, 1996–2014, including an outlier: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show Gini coefficients of citation distributions from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in the Gini coefficient using a resampling correction. The lines marked *Obs* in the bottom panel shows Gini coefficients over fields and time using the empirical data from Web of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2002. All curves are smoothing splines with span of 0.5. No cases are omitted.

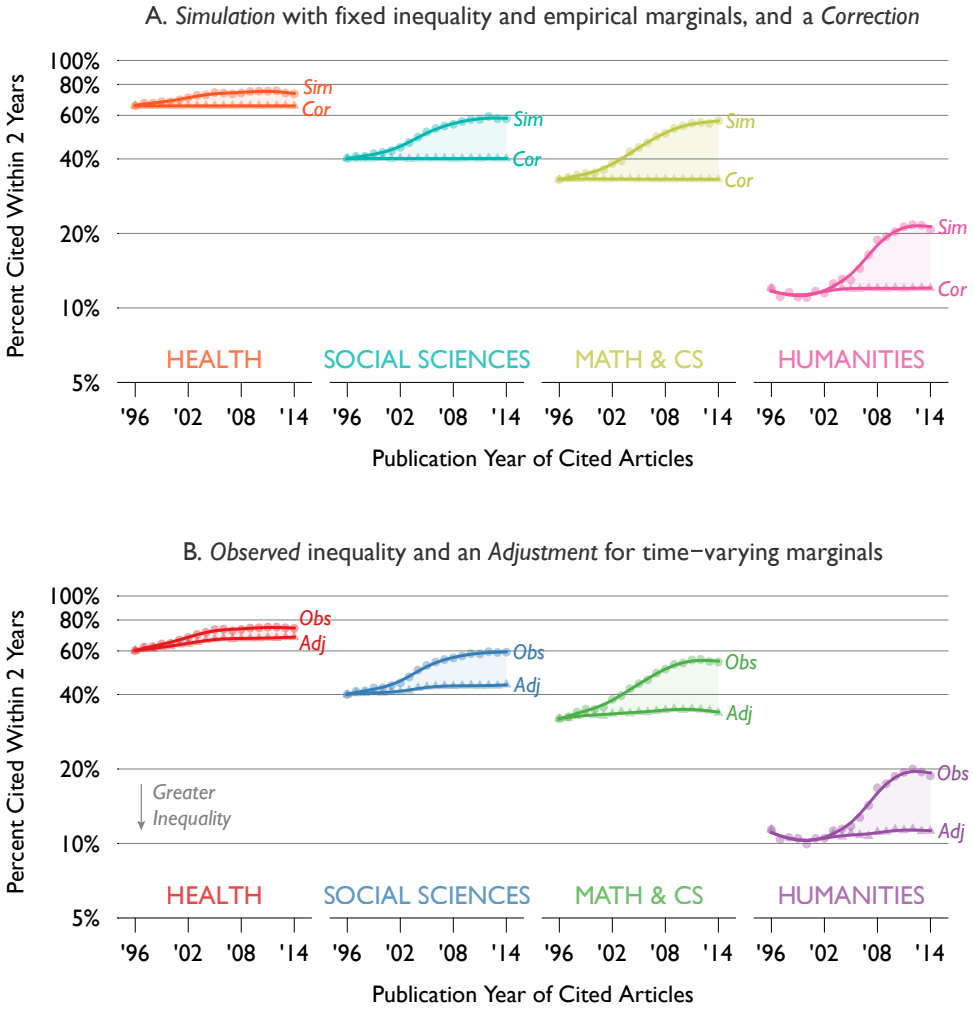

Figure S20. *Percent of papers with any citations two years after publication, 1996–2014, including an outlier: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers ever cited from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in percent-ever-cited using a resampling correction. The lines marked *Obs* in the bottom panel shows percent-ever-cited over fields and time using the empirical data from Web of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2002. All curves are smoothing splines with span of 0.5. No cases are omitted.

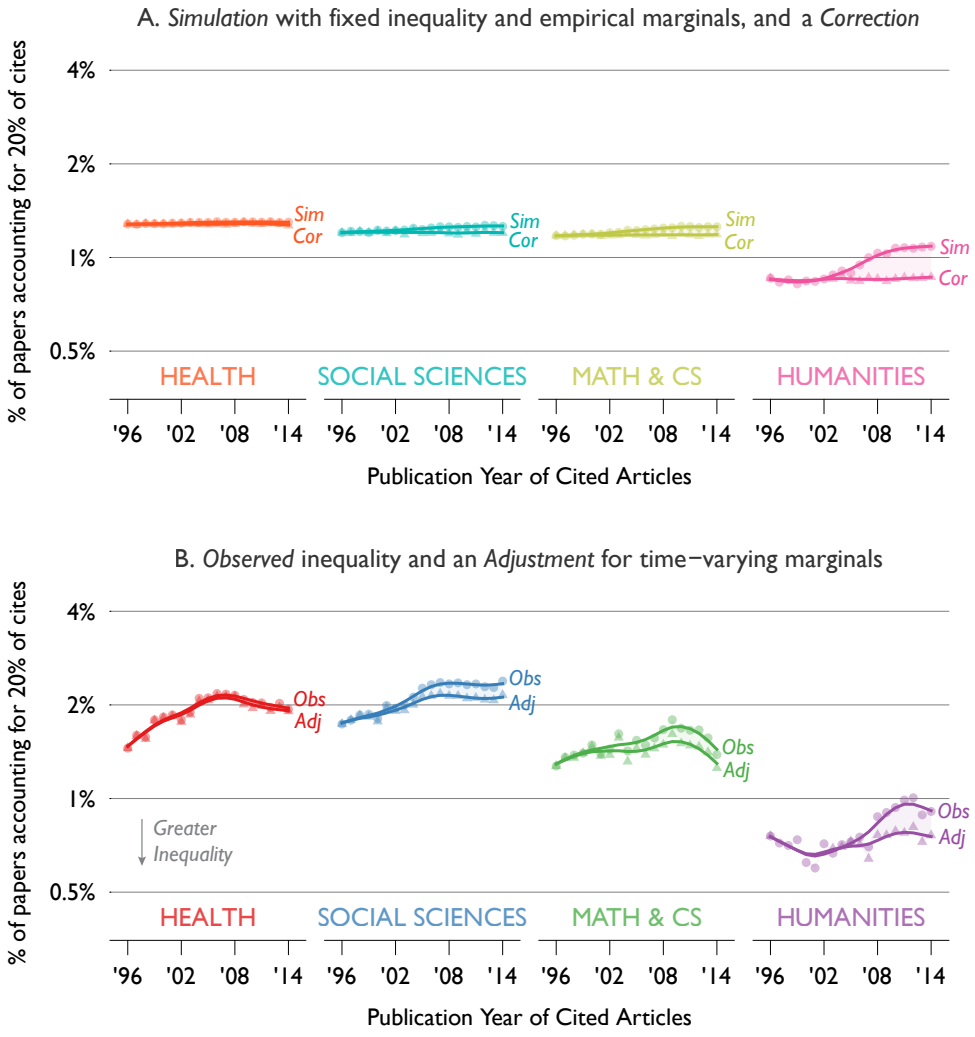

Figure S21. *Percent of papers accounting for 20% of all citations within two years of publication, 1996–2014, including an outlier: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 20% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows the percent of papers accounting for 20% of citations over fields and time using the empirical data from Web of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2002. All curves are smoothing splines with span of 0.5. No cases are omitted.

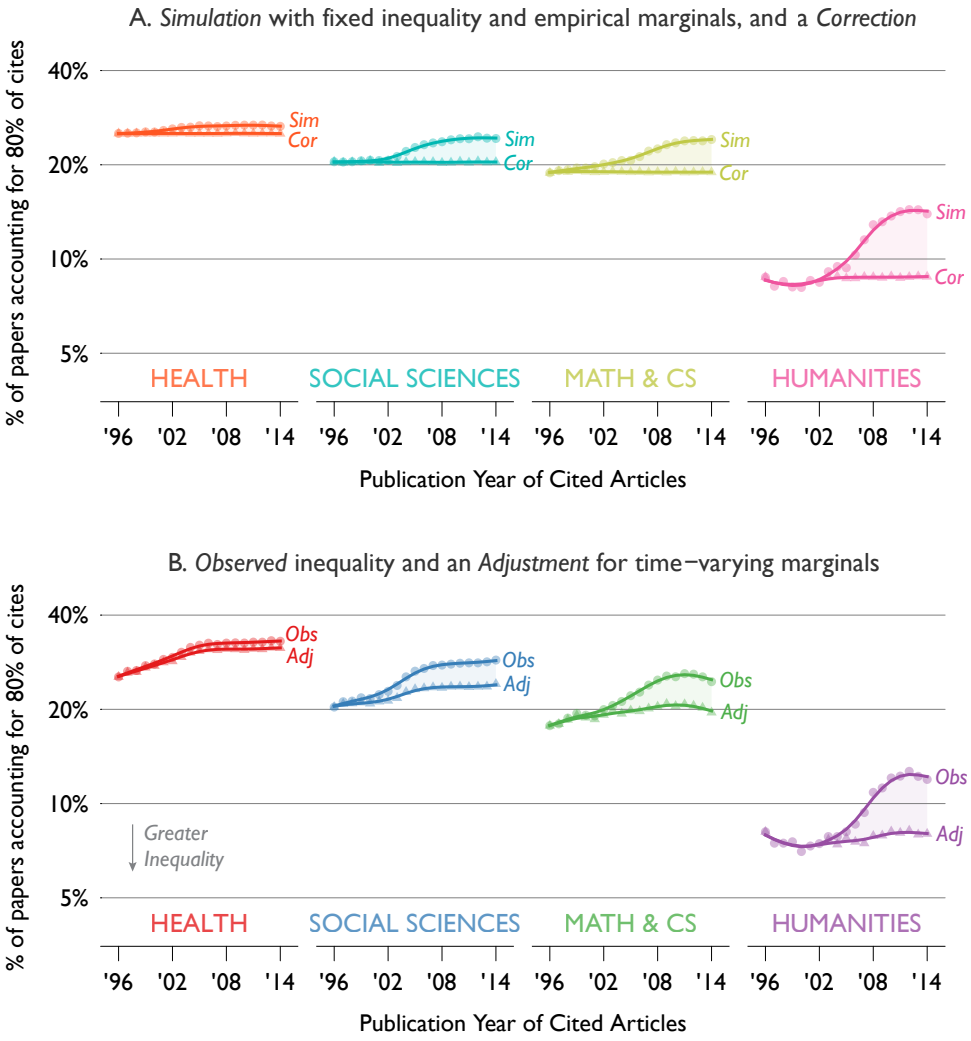

Figure S22. *Percent of papers accounting for 80% of all citations within two years of publication, 1996–2014, including an outlier: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show percent of papers accounting for 80% of all citations from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias using a resampling correction. The lines marked *Obs* in the bottom panel shows the percent of papers accounting for 80% of citations over fields and time using the empirical data from Web of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2002. All curves are smoothing splines with span of 0.5. No cases are omitted.

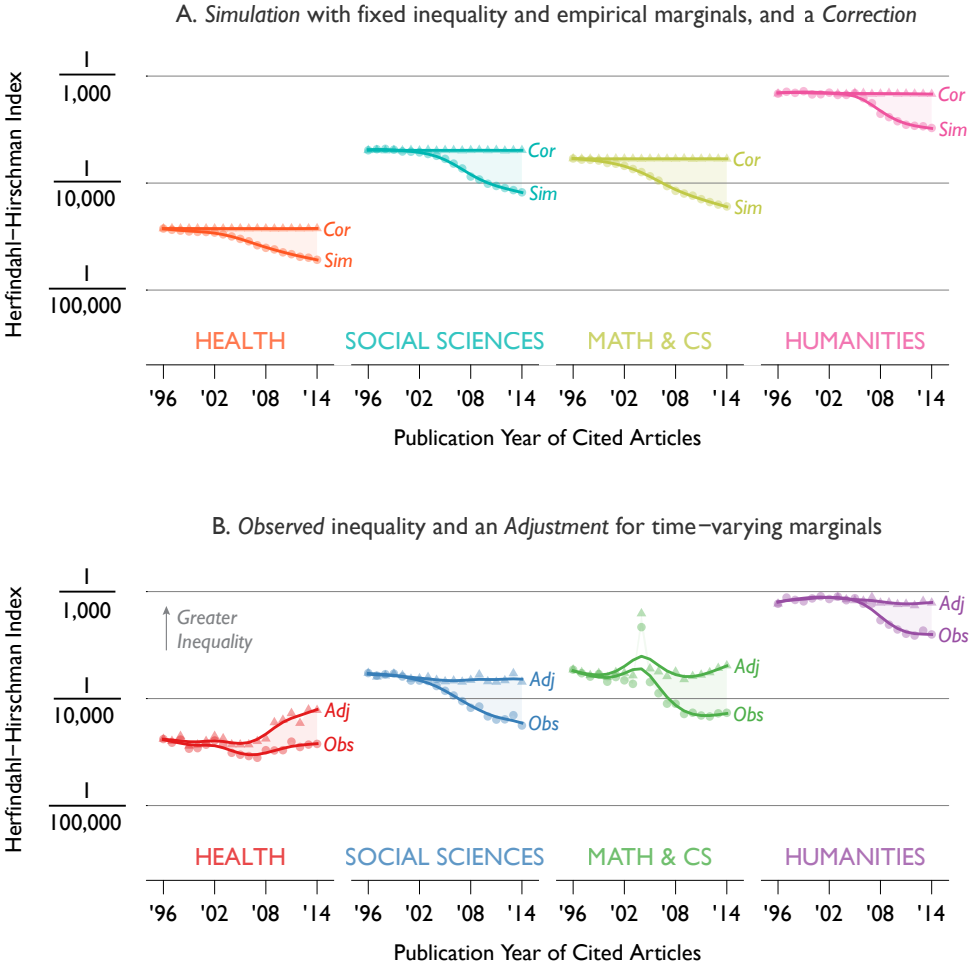

Figure S23. *Herfindahl-Hirschman Index of citations within two years of publication, 1996–2014, including an outlier: Monte Carlo simulation and empirical results.* The lines in the top panel marked *Sim* show the Herfindahl-Hirschman Index (HHI) of citation concentration from Monte Carlo results for hypothetical papers and citations designed to have a fixed pattern of inequality across years and fields, but total papers and citations matching the empirical marginals of those fields by year. The lines marked *Cor* remove the marginals bias in HHI using a resampling correction. The lines marked *Obs* in the bottom panel shows HHI over fields and time using the empirical data from Web of Science; these results are subject to marginals bias from differences in total papers and citations by field and year. Lines marked *Adj* adjust for marginals bias in the empirical data by resampling to the marginals in 1996 by field. Corrections and adjustments omitted for the humanities in 1997–2002. All curves are smoothing splines with span of 0.5. No cases are omitted.