**SUPPORTING "THE BEST AND BRIGHTEST" IN SCIENCE AND ENGINEERING:
NSF GRADUATE RESEARCH FELLOWSHIPS***

Richard Freeman, Tanwin Chang, Hanley Chiang
The National Bureau of Economic Research and Harvard University

freeman@nber.org

Revised March 2006

Abstract

The National Science Foundation's (NSF) Graduate Research Fellowship (GRF) is a highly prestigious award for science and engineering (S&E) graduate students. This paper uses data from 1952 to 2004 on the population of over 200,000 applicants to the GRF to examine the determinants of the number and characteristics of applicants and the characteristics of awardees.   In the early years of the program, GRF awards went largely to physical science and mathematics students and disproportionately to white men, but as the composition of S&E students has changed, larger shares have gone to biological sciences, social sciences, and engineering, and to women and minorities. The absolute number of awards has varied over time, with no trend. Because the number of new S&E college graduates has risen, the result is a sharp decline in the number of awards per S&E bachelor's graduate. In the 2000s approximately 1/3rd as many NSF Fellowships were granted per S&E baccalaureate than in the 1950s-1970s. The dollar value of the awards relative to the earnings of college graduates has also varied greatly over time. Our analysis of the variation in the number and value of awards and of the characteristics of applicants and awardees finds that:

   1. The primary determinant of winning a GRF are academic skills, which greatly impact panel ratings of applicants. Consistent with efforts to increase S&E diversity, women and minorities have higher changes of winning an award than white men with similar attributes.

2. The size of the applicant pool varies with the relative value of the stipend, the number of S&E bachelor's graduates, and the lagged number of awards per graduate. We estimate that for every 10% increase in the stipend value, the number of applications goes up by 8 to 10 percent.

3. The average measured skill of awardees falls when the number of awards are increased and rises with the value of fellowships.

4. The supply of applicants contains enough qualified candidates to allow for a sizeable increase in the number of awards without greatly reducing measured skills.

5. The supply of highly skilled applicants is sufficiently responsive to the value of awards that increases in the value of stipends could attract some potentially outstanding science and engineering students who would otherwise choose other careers.

CONTENTS

INTRODUCTION

List of tables

List of figures

Stipends – payments to students as part of their studies – are a major form of income for graduate students in science and engineering (S&E) and a potential determinant of decisions to undertake graduate study. Broadly defined to include fellowships, research assistantships, teaching assistantships, and post-doc awards, stipends account for upwards of 1/4th of the lifetime incomes of S&E PhDs, depending on field, time to degree, and prevalence and length of post-docs.[1] Fellowships are potentially the most attractive stipend because they signal to students that the granting institution views them as top prospects for success in graduate study and is willing to support their studies rather than requiring them to work to cover graduate school expenses. In 2001 thirteen percent of full-time graduate students reported that a fellowship or traineeship was their primary source of graduate support [4]. Slightly more than one-third of those, or approximately 4% of full-time graduate students, obtained awards from federal government sources [3].

Since its establishment in 1952, the National Science Foundation's (NSF) Graduate Research Fellowship (GRF) has been the US's premier award for science and engineering graduate students. NSF's stated purpose for the Graduate Research Fellowship is to "ensure the vitality of the scientific and technological workforce in the United States and to reinforce its diversity." Because NSF fellowships are limited to citizens or permanent resident aliens, they are a potentially important policy tool for inducing citizens and residents, including those from underrepresented groups, to study S&E and thus to raise their representation in the S&E workforce.

What determines the likelihood that an applicant wins a GRF award? How, if at all, has the composition of applicants and awardees changed over time? What is the supply response of students

---

[1] In the 1990s, about two-thirds of S&E grad students received stipend support of some form and thus did not have to rely on loans, personal finances, or money from family and friends as their primary means of support [3].

to increases in the number or the amount of fellowships? How might alternative fellowship policies affect the numbers and measured attributes of awardees? Can we generalize from the response of students to GRFs to the possible impact of stipends on the supply of students to S&E more broadly?

This paper uses data from the NSF's 1952-1993 Cumulative Index (CI) file, its ensuing updates, and other sources to try to answer these questions. The Cumulative Index plus updated data provides information on the over 200,000 individuals who applied for NSF graduate fellowships from 1952 through 2004, including the student's graduate record exam scores (GREs), grade point average (GPA), ratings of reference letters, and from 1976 on, the ratings of applicants by review panels. We supplement these data with information on the earnings of college graduates, unemployment rates and the sizes of graduating bachelor's cohorts. Much of our analysis uses time series fluctuations in the number of GRFs and the amount of awards to provide the exogenous variation necessary to identify student supply responses.[2]

Section 1 describes the GRF program and how it has changed over time. It finds that the ratio of GRFs to S&E bachelor's degrees fell by nearly 75% from the 1950s to the 1990s/2000s; that the distribution of awards shifted from physics and chemistry to biological sciences, social sciences, and engineering, and that women and minorities have gained an increased share of awards.

Section 2 uses a linear probability model to estimate the characteristics that help an applicant

---

[2]   The 2002 report on the GRF by WestEd looks at what happens to fellows rather than what attracts students to the fellowships. It shows that in the time period leading up to1993 around 70% of fellows received their baccalaureate from RU1 institutions (as defined by the Carnegie Classification), and more than 90% went on to perform graduate work at RU1 institutions. Some programs receive a disproportionately high numbers of fellows: 34% of Life Sciences Fellows enrolled in 5 institutions. The report shows that 11-years after beginning their PhD program NSF Fellows had completion rates of around 70%, which exceeds the overall completion rate for entering students; that fellows pursued academic careers more than their peers, and that a slightly larger percentage reported having achieved traditional measures of accomplishment during graduate school, such as presentations, refereed articles, edited book part, etc.

win a GRF. It shows that the chance of winning depends greatly on GREs, GPAs, and reference

letters operating through panel ratings, and that consistent with NSF's diversity goal, women and

minorities have higher chances of winning an award than majority men with similar attributes.

Section 3 assesses the supply response of applicants to changes in the value of the fellowship

It finds that an increase in stipend value increases the number of applicants, including those with

very high skills that NSF is able to raise the average skill of awardees. By contrast, increases in the

number of awards are associated with modest declines in the average quality of awardees.

Section 4 explores new ways to determine the number and value of GRF awards: increasing

the number and/or value of awards; setting the number and value by indexing them to measures of

the supply of potential applicants and to the monetary attraction of alternatives; and giving awards

by the measured academic and personal characteristics of applicants, irrespective of field.

I. THE GRF PROGRAM OVER TIME

Congress established the National Science Foundation's Graduate Research Fellowship

program in 1952 to support the vitality of the scientific and technological workforce in the United

States. The program was open to US citizens only. It gave a $1600 stipend and covered normal

tuition and fees at the institution of choice. The NSF typically supported students for 3 years, but

this was not automatic; students had to reapply each year. The GRF has changed since its inception.

In 1963 it capped the cost-of-education, and in 1972 students were no longer required to reapply

[6]. Today awardees may choose 3 years of support over a 5 year period, and must be pursuing a

research-based graduate degree in an NSF supported field.   The fellowship may be used outside the

United States and is portable across institutions. In 1997 the NSF introduced new selection criteria

which put greater weight in its awarding fellowships to "broader impacts" such as integration of

research and education, diversity, benefit to society, and enhancement of scientific and technical understanding than in the past.

The number of awards and the dollar value of stipends has changed over time with NSF budgets and decisions. Figure 1 shows the absolute number of awards and constant dollar value of the stipends from 1952 to 2004. In the 1950s the NSF gave around 500-600 awards. In the 1960s it gave around 1000 awards, with a peak of 1373 in 1966. In the 1970s and 1980s the number of awards fell to around 500, then rose in the late 1980s/early 1990s toward 900-1000. The big jumps in the value of awards reflect NSF decisions to raise nominal stipend amounts. NSF raised the value of fellowships in the early 1980s to the late 1980s and again in 1999 to 2004.[3] In 1999, the Committee of Visitor's (COV) report evaluating the program noted that "the GRF awards are no longer as attractive as they once were" [2] at $15K per year for 3 years – significantly lower than comparable fellowships. The COV speculated that this could (at least partially) explain the steady decline in applicants in the years leading up to 1999 and recommended that the stipend value be raised to $18K. In the subsequent four years, the value of the stipend rose 83% to $27,500 prompting the next COV (convened in 2003) to worry that the stipend was now too large. Universities complained about the negative effects of a support gap between NSF fellows and other students. The 2003 COV recommended that the "NSF thoroughly examine the external consequences of the $27,500 stipend and what intended and unintended consequences this might engender."[1] For the 2005 fiscal year, the stipend has been set at $30,000 with an accompanying $10,500 cost of education allowance. The roller-coaster behavior of the award amount has

---

[3]The gradual declines in the stipend values in years when the NSF has not raised them reflect the impact of inflation. The NSF has never reduced the nominal value of awards.

prompted some in the graduate education community to seek a consistent basis for the determination of stipend levels.

The absolute number of GRF awards of about 900-1000 in recent years appears comparable to the levels in the 1950s. Relative to the increased supply of S&E bachelor's graduates, however, the number of awards has fallen sharply. In 1952-56 NSF awarded 5.4 GRFs per thousand S&E bachelor's degrees. In 2000-2004 NSF awarded 2.2 GRFs per thousand S&E bachelor's degrees – a 60% drop. Figure 2, which graphs the number of awards per S&E bachelor's, shows that the drop occurred discontinuously around 1970.

Field and demographic distribution of awards

In the early years of the program, GRF awards went largely to physical sciences and mathematics and disproportionately to white men. Today, larger shares go to biological sciences, social sciences, and engineering; and also to women and minorities. Figure 3 shows the huge change in the distribution of awards among fields. From 1952 to 1956 the physical sciences obtained 49% of the awards, math received 9% and life sciences obtained 25% compared to 4% in psychology and social sciences and 13% in engineering. In 2004, by contrast, 15% of awards were given to the physical sciences, 11% to mathematics and computer sciences, and 27% to life sciences; 30% were given in engineering and 17% in social sciences and psychology.

Given NSF policy to award approximately the same proportion of applicants among fields, the shift in awards among disciplines largely reflects a change in the distribution of applications among disciplines. Consistent with this, figure 4 shows that before 1990 the proportion of applicants who won awards in the major discipline categories were within a few percentage points of each other, although Social Science and Psychology were slightly disfavored. This is still the case

after 1990, except that applicants in Engineering and Computer Science have a higher chance of getting an award due to the process by which the NSF seeks to attract women to these disciplines.

Turning to the demographic characteristics of awardees, figure 5 shows that the proportion of women and underrepresented minorities[4] winning the GRF has risen sharply since the program was initiated. The female proportion of awardees climbed from about 5% in the 1950s to 55% of awards in 2004, with considerable variation among fields. Women make up roughly three quarters of the awardees in Life Sciences, Social Science, or Psychology compared to 20% of awardees in Physical Science or Computer Science. Around 1960, the female proportion in Life Science was about 25% while in Physical Science it was around 6%, while there were barely any female engineering fellows[5]. To increase the female representation in these fields, NSF set up the Women in Engineering Program (WENG) in 1990, which became the Women in Engineering and Computer Science (WECS) program in 1993. This progarm boosts the number of females by awarding fellowships to all or nearly all Computer Science or Engineering women in NSF's quality group 2, and by making women in quality group 3 also eligible for awards if enough slots are available.[6] The data in figure 4 show that in 2001 applicants in Computer Science and Engineering had a higher ratio of awards to applicants than other disciplines as a result of this program.

The data in figure 5 on the proportion of underrepresented minorities (URM) among GRF fellows starts in 1976 due to lack of ethnicity data before then. The GRF data is further complicated

---

[4] Underrepresented Minorities in S&E include African Americans, Hispanics, and Native Americans. Asian Americans are overrepresented in science and engineering.

[5] In 1960 there were few computer science fellows, either male or female.

[6] As described below, NSF categorizes applicants by group depending on their measured scholastic skills. All quality group 1 applicants are offered GRFs, and a sizable proportion of group 2 applicants are also given GRFs.

7

by the development in 1978 of the Minority Graduate Fellowship Program (MGF), which gave

awards solely to underrepresented minorities, and with its dissolution in 1998. The MGF was

officially a separate program, although executed in parallel with the GRF. In 1999, the first program

year since 1975 without the MGF, the NSF emphasized procedural changes in its review panel

process in the hope of maintaining a diverse awardee pool, partly by attracting more applicants

through expanded outreach, and partly via other measures such as insuring a diverse group of

panelists [1]. Combining the fellows from the GRF and the MGF, the URM proportion is around

10% for the early 1980s, and climbed to 30% in 1998, the last year of the program. After that, the

percentage dropped to the GRF-only proportion of around 10%, which is still much larger than the

1% to 2% two decades earlier. The disciplinary variation for URM fellows shows that a large

proportion are in Social Science and Psychology compared to other disciplines.

The increased diversity in GRFs reflects two factors. First is that as a result of these

programs and efforts, the likelihood that a woman or minority applicant would get an award has

risen. In the first decade of the GRF, the overall award rate[7] was 17% while the award rate for

women was 12%, as shown in figure 6. By the mid-1990s the female award rate exceeded the

overall rate. The award rate for URMs in the GRF was low at 3% during the run of the MGF. The

the combined rate of URMs in both the GRF + MGF topped 18% from 1994 - 1998, which is

actually greater than the MGF-only award rate, 15.3%, for the same time period[8]. After the end of

that program, the number of URM awardees per applicants approached the overall rate of awardees

---

[7] "Award rate" is synonymous with "awardee to applicant ratio".

[8] The explanation for this result lies in the high incidence of duplicate applications. Data was available from 1994 -1998 to match applicants in the two programs, and we found that approximately 90% of those who applied for the MGF also submitted an application for the GRF.

per applicant, though it is still noticeably lower by several percentage points.

The second and more important reason for the increased diversity of GRFs is that an increasing share of bachelor's degrees in science and engineering have gone to women and minorities. This has raised the proportion of applicants from these groups, which in turn produces a increased share of awards, coupled with the NSF's efforts to improve diversity.

II. DETERMINANTS OF GRF AWARDS

The GRF is a merit-based award. Each year, the NSF assembles a panel of experts to evaluate and rate applications and arrange them into quality groups. Quality Group 1 (QG1) is the highest quality group, and all applicants deemed fit for this group are offered the fellowship. Quality Group 2 (QG2) applicants are regarded as sufficiently meritorious for the fellowship, but not all in this category will receive the award. Applicants in other quality groups rarely receive the fellowship, but may receive an "honorable mention" which carries some non-monetary prestige. As with other aspects of the program, the procedures for determining awardees have evolved. Presently, panelists meet in groups arranged by discipline for an intensive 2-3 day session, which includes a briefing on the GRF Program and a calibration exercise using applications from previous years. The categorization process consists of cycles during which the panelists score the applications on a scale from 1.00 to 4.99 (lower number indicating higher quality), and then rank them. In the first cycle, each application is scored by at least two panelists, and then any applicant ranked below the 64th percentile are put in QG4, and then no longer considered for the fellowship. Remaining applications are then scored by a 3rd panelist before being re-ranked. Approximately 55% of the awardees are from QG1, all of whom receive an award. Virtually all of the remaining awards are from QG2, with approximately 60% of QG2 applicants offered the fellowship [8]

From the 1998 program year to the present, NSF has instructed panelists to evaluate applications based on the two broad criteria common to NSF solicitations: *intellectual merit* and *broader impacts*. The panelists are given rating sheets upon which intellectual merit is broken down into sub-criteria: GPA, GRE scores, references, proposed research, and previous research experience. Broader impacts are based on: integration of research and education, diversity, benefit to society, and enhancement of scientific and technical understanding. The NSF emphasizes a comprehensive concept of merit, and advises panelists to avoid placing too much emphasis on the easily quantifiable quality measures such as GRE scores and GPA. In the latest program years, GRE scores are not required on the GRF application. Further, there is a new emphasis on the value of leadership skills in NSF fellows [1]. It is possible that future applicants will be asked to provide evidence of leadership potential, and future panelists will include this in their assessment of merit.

Since the NSF recognizes the value of a diverse workforce, the program implements measures to insure representation in geography, demographics, and discipline. Specifically, the applicants from Quality Group 2 who receive the fellowship are chosen based on an algorithm that includes their rank within QG2, but also considers other factors. The key policy with respect to discipline is that NSF allocates the number of QG1 and QG2 positions to review panels on the basis of the disciplinary distribution of all applicants and the number of awards NSF plans to offer.

Regression analysis

To see how the selection procedure uses the information on student skills to determine which applicants win awards, we estimated linear probability models in which the dependent variable is the 0/1 measure of obtaining an award and the explanatory variables consist of measures of student skills from the NSF's Cumulative Index dataset . The CI has recorded the scores that a panelist gives

to an application starting in 1976, but does not record assessments of the sub-criteria, which limits

our choice of measures largely to the quantitative and verbal GRE scores, the GPA, and the scoring

of reference letters. Table 1 lists the measures of scholastic skills in the data and gives their mean

values and standard deviations, as well as the overall mean of the probability of winning the GRF

award, and the scales used to measure each quantity. GRE scores are coded on a 200 to 800 scale;

the GPA uses a 4 point scale; Panel Ratings[9] were originally coded on a scale of 100 to 600, with

lower score representing higher quality; reference scores were originally scaled from 10 to 70, also

the lower the better. The panel ratings and reference scores changed scales at various times in GRF

history. To deal with this, we normalized the panel and reference data such that they have zero

mean and standard deviation of one in the applicant pool of each year. We also reverse-coded the

actual panel and reference scores, so that increases in all variables imply a more favorable situation.

Regression coefficients for GRE scores are sometimes reported in tables with the scores scaled as

GREQuant/100 or GREVerbal/100, which effectively puts the them on a 2 to 8 point scale.[10]

Table 2 records linear regressions of the impact of scholastic attributes, demographic factors,

and field dummies for nearly the full data set from 1955 to 1998, which lacks measures of minority

status, panel scores, and reference letters; and for the 1976-1998 period when we have data on those

measures as well as the others. The column 1 and 2 regressions do not include the panel rating,

while the column 3 regression includes the panel rating. The column 1 estimates show that over the

entire period, GRE Quantitative, GRE Verbal, and GPAs have sizable and significant effects on the

---

[9] In the original Cumulative Index dataset, each applicant may have more than one panel rating if more than one panelist reviewed the application. An *average panel rating* was trivially generated for each applicant. In this paper, "panel rating" means "average panel rating". Similarly, "reference score" is synonymous with "average reference score".

[10] This allows us to report coefficients concisely as 0.XX rather than 0.00XX.

probability of obtaining the GRF. A 100 point increase in Verbal GRE score raises the probability

of award by 6.4%. This might not seem large, but since just 13.5% of applicants get an award this

raises the probability of getting an award by nearly 50%. In addition the calculations show that

women and minorities have higher chances of getting an award than other applicants. The field

dummy variables show some variation in the probability of getting an award among fields.

Engineering , life science, and earth/atmospheric science have significantly higher probabilities of

getting an award while mathematics has a lower one compared to the omitted social science group.

The results in column 2, which include the measure of the quality of references and dummy

variables for gender and minority status for the period 1976 to 1998 are similar: GREs, the GPA,

and the reference scores are key determinants of the probability of getting an award, while women

and minorities have a higher probability of gaining awards. The dummy variables for fields show

moderate disciplinary variation. Column 3 adds the panel rating to the list of variables. Since the

NSF panels weigh heavily the effect of GREs, GPAs, and reference scores in deciding which

applicants should be given awards, inclusion of the panel rating alters the estimated coefficients

markedly. The coefficients on factors on which panels place most weight in their evaluation fall

relative to those that enter less highly in the panel rating. The GRE quantitative score and the GPA

obtain negative coefficients conditional on the panel rating, implying that the panels weigh those

factors especially heavily in their ratings. In addition, conditional on the panel ratings the field

dummies become large and significant relative to the omitted social science group. This presumably

reflects the initial allocation of the number of QG1 and QG2 positions on the basis of the

disciplinary distribution of applicants, which leads panels in the quantitative and physical sciences to

give lower ratings than those in the social sciences to applicants with comparable measured

scholastic skills. From this perspective, the estimated positive coefficients on the field dummies in table 1 are a "correction" that takes account of the attributes of applicants in these fields.

Consistent with this interpretation, table 3 shows that panel ratings are indeed lower for applicants with the same measured scholastic skills in physical sciences than in the social sciences. For instance, the regression shows that an applicant in chemistry obtains a -0.198 lower panel rating than an applicant in the omitted social science group *with the same GREs, GPA, reference score and demographics.* The rank correlation of the ordering of fields between the table 3 estimates of the panel ratings and the ordering of fields in the column 3 of table 2 is -0.92 (inclusive of the omitted group). Thus, the modest estimated impacts of fields in columns 1 and 2 of table 2 reflect two offsetting factors: lower panel ratings in physical sciences but an offsetting higher probability of getting an award conditional on panel ratings in those disciplines.

The effect of demographics

As noted, the table 2 regressions for obtaining a GRF show that women and minorities have a greater probability of getting an award than do majority men with similar measured scholastic skills. This is consistent with the goal of increasing the diversity of the S&E work force. The table 3 regression of the panel ratings on scholastic skills and demographic characteristics show, however, that panels give lower scores to women than to majority men with the same measured attributes while they give higher scores to minorities than to majority men with the same measured attributes. Since the probability of getting an award is higher for both women and minorities than for majority men, the difference in coefficients in panel ratings suggests that the process of taking account of diversity differs between the two groups.

To unpack this difference, we estimated a recursive model. First, we related the panel scores

and the probability of being assigned to groups 1 and 2 to demographic characteristics. Then we examined the probability of getting an award conditional on being in groups 1 and 2 and on the panel rating. Table 4 gives the results of this analysis for 1990-97 and 1998-2004. The data for 1990-97 includes measures of GRE scores, GPA, and a measure of the quality of reference letters. The 1998-2004 data do not include GPA or the quality of reference letters, and cover the period when NSF revised its merit criteria. Column 1 shows that in 1990-97 women received a slightly lower score than men from review panels while there is no noticeable difference in panel ratings between minorities and the majority group. Women also were less likely to be assigned to quality groups 1 and 2 than men, while by contrast minority applicants were more likely to be assigned to those groups. Among persons in groups 1 and 2, and conditional on panel ratings, women were much more likely to obtain awards than men while minority applicants had a modestly higher probability of obtaining awards. Thus, the higher probability of women getting awards occurs at the last stage, presumably in part because the WECS (WENG) program boosted the chances that women in quality group 2 would get GRFs, whereas the higher probability of minorities getting awards occurs because panels were more likely to assign them into groups 1 or 2.

The 1998-2004 regressions show a very different pattern, indicating that the change in NSF's policies in awarding GRFs had a substantial effect on the awards process. Column 4 shows that panels give both women and minorities higher ratings in 1998-2004 than in 1990-97. Column 5 shows that panels are more likely to assign them to quality groups 1 or 2 than in 1990-97. The coefficient on being female in column 6 shows that, conditional on being in groups 1 or 2, the boost given to women in the probability of getting an award was more moderate in 1998-2004 than in 1990-97, presumably because they were getting higher panel ratings and having a higher chance of

being assigned to quality groups 1 or 2. By contrast, the coefficient on being minority in column 6 is nearly identical to the coefficient in column 3. Finally, panel ratings have a stronger impact on awards in 1998-2004 than in 1990-1997. In sum, table 4 provides evidence for the change in NSF's emphasis that was enacted, at least partly, to increase diversity in S&E fields.

Finally, table 5 uses a regression model to examine the GRE scores of applicants with different demographic characteristics from 1976 to 2004 – a period in which the proportion of applicants who were women or minority increased substantially. In 2004 there were 8,939 applicants, of whom 46% were women and 10% underrepresented minorities. In 1976, there were 5366 applicants of whom 30% were women and 4% were underrepresented minorities To the extent that increased numbers of applicants draw persons with decreasing measured scholastic skills, one would expect that the increased proportion of women and minorities would reduce the average scores of those groups.

The coefficients and standard errors on the demographic variables in the table show the opposite pattern, particularly for GRE quantitative scores. The scores of both women and minority applicants rise relative to those of majority men, though both groups still score lower. The coefficient on the dummy for women rises from -41.7 in 1976-83 to -28.9 in 1999-2004. The coefficient on being black rises from -175.9 in 1976-84 to -100.1 in 1999-2004. The negative coefficient on being Hispanic changes less evenly but is lower in 1999-2004 than in 1976-1983, while that for other minorities shows little change. The regression coefficients for GRE verbal scores tell a similar story, with the exception that women applicants go from having the same verbal scores as men in 1976-83 to a modest 7.2 lower score in 1999-2004. The coefficient for being black, however, advances steadily from -168.7 to -100.4. Overall, the results show that the increased

proportion of female and minority GRFs was associated with generally improved measured skills. The implication is that the GRF program is attracting increasingly qualified women and minorities.

III. SUPPLY RESPONSES OF GRF APPLICANTS

The NSF has two policy levers that can affect the supply of applicants to the GRF program and, given the process of granting awards, that can determine the measured skills of awardees: the dollar amount of the awards and the number of awards.

Increasing the dollar value of awards should increase the number of applicants. It could increase or decrease the average measured skill of *applicants* depending on whether they come largely from persons with the highest skills or from those with lower skills. But given the process of selecting awardees, increases in the value of awards should increase the average skill of *awardees* as long as the increased value of awards attract some persons with high skills. In this case, review panels would be expected to select high skill candidates from the pool of newly attracted persons over less skilled candidates that they would have selected from the smaller initial pool of applicants.

Increasing the number of awards, by contrast, is likely to reduce the measured quality of awardees since it means accepting applicants who would otherwise be rejected. The key issue here is the extent to which the increase in numbers reduces quality, which depends on whether the pool of similarly skilled applicants on the margin of acceptance is large or small. If there is a large pool, increased numbers of awards will reduce measured skills modestly. If theres is a small pool, increased number of rewards could reduce measured skills substantially.

Number of applicants

To examine the link between the value of stipends and the pool of GRF applicants, we calculated a *relative value of stipends* by taking the ratio of the dollar value of GRF awards to the

average annual earnings of college graduates aged 21-25[11] and contrasted this with the *relative number of applicants* by dividing the number of applicants by the number of S&E bachelor graduates. As figure 7 shows the two series track each other closely. For example, the sharp increase in the stipend value in the 1980s, when NSF increased stipend amounts from $3900 (1979) to $14,000 (1991) is matched by a large rise in the number of applicants relative to S&E bachelor's graduates. By contrast, from 1991 to 1999 the relative value of stipends fell, as the nominal stipend amount rose by $1000 while college graduate earnings rose by about $7600; and the relative number of applicants fell. Since 1999 the nominal amount of the stipend has doubled, and the number of applications has nearly doubled as well.[12]

Another factor that might affect the decision to seek a GRF and go on in graduate school is the ease of finding work with a bachelors degree, which varies with the business cycle. Figure 8 compares the relative number of applicants to the unemployment rate of college graduates[13]. Since the late 1980s, the size of the GRF applicant pool has moved countercyclically: a weak labor market for college graduates generated a higher number of applicants as a share of bachelors degrees. But in the early to mid-1980s, the unemployment rate fell while the number of applicants rose, presumably in part because of the increase in stipend amounts shown in figure 7. Thus, figure 8 suggests that unavailability of jobs for recent college graduates is a secondary factor influencing the

[11] We estimated the college graduate earnings from the Integrated Public Use Microdata Series (IPUMS) of the March Current Population Survey. College graduates include those who obtain degrees higher than a bachelor's degree, but the age range is restrictive enough to exclude most doctorate recipients.

[12] Bachelor's degree data for 2002-2004 are extrapolated because actual figures are not available. The key to the change in the ratio of applicants to bachelor's graduates is that the number of applicants jumped from 4852 in year 1998 to 8939 in year 2004.

[13] Estimated from the annual Current Population Survey (CPS) Outgoing Rotation Group

decision to apply for a GRF stipend.

Taking the analysis a step further, we amalgamated the data on GRF fellowships into nine major fields[14] by year from 1952 to 2004, to obtain a cross-field time series panel with 477 observations. Using this sample, we did a multivariate regression analysis of the determinants of the number of GRF applicants in each field and year. Since the NSF changes the stipend value in particular years by policy decision rather than changing it annually depending on the previous years' labor market, we take these changes as exogenous and use least squares to estimate the effect of relative stipend values on the relative number of applicants. To control for field differences and changes in market conditions by field over time, the regressions include field dummy variables and separate time trends for the nine fields. In addition, the regressions include the number of bachelor's degrees earned in the nine fields in the previous year.[15]

Table 6 displays the regression results. The basic specification in column 1 shows that a larger number of bachelor graduates in a field generates a larger number of applicants from that field; and that stipends in the previous year have a strongly positive effect on applications, with a near unitary elasticity. Column 2 adds to the set of explanatory variables the unemployment rate among college graduates 21-25 years of age, as an indicator of the availability of alternative jobs for those contemplating going directly to the job market. The sample size shrinks because the unemployment variable is only available starting in 1979. Consistent with figure 8, the unemployment rate has a positive effect on the number of applications in column 2. In column 3 we

---

[14] The nine fields are: Chemistry, Computer Science, Earth/Atmospheric Science, Engineering, Life Science, Mathematics, Physics, Psychology, and Social Science. The availability of earnings data for college graduates and degree data by field limits the sample to 1968-2001.

[15] We recognize that students may apply for a GRF in a different field than the field in which they earned their undergraduate degree, but still regard the number of undergraduate degrees as a reasonable indicator of interest in the field.

replace the unemployment rate of 21-25 year olds with the unemployment rate for all college graduates regardless of age, on the assumption that this variable has less measurement error since it is calculated for a larger sample. The effect of unemployment nearly doubles: a one percentage point rise in the unemployment rate raises the GRFP applications by just over 10%. Finally, in column 4 we include a measure of the relative availability of awards by field -- the number of awards divided by the number of bachelors graduates in the field lagged one period. This shows that more students apply when the probability of obtaining an award is perceived to be higher.

In sum, the regression specifications in Table 6 show that the number of GRF applications responds substantially to the relative value of the stipend and to the perceived availability of awards. For every 10% increase in the stipend value, the number of applications goes up by 8 to 10 percent.

Measured skills of awardees

To see how the value of stipends affects the measured achievement of GRF awardees, we compare the GRE scores and GPAs of awardees over time and across fields. Figures 9A and 9B provide a first indication that stipend values affect the qualifications of the awardees. Figure 9A shows that the average GRE quantitative score of awardees has been higher when the relative stipend value is higher. Figure 9B plots shows a similar relation between stipend values and average GRE verbal scores of awardees. The time series of raw verbal scores trended downward beginning in 1987, so we examined the time series of GRE verbal scores adjusted for trend, as well as the raw time series. The trend adjusted series shows a rise in scores with the large increases in relative stipend values in the first half of the decade, and a subsequent fall in verbal scores throughout the 1990s. However, the continuing decline in GRE verbal scores through the late 1990s/ early 2000s cannot be explained by the evolution of stipend values, which are rapidly rising during this time.

To quantify the pictures in the figures, we regressed measures of the scholastic qualifications of awardees on the number and value of GRF awards, using a cross-field time series regression framework similar to that in Table 6, with field fixed effects and field-specific time trends included in all regressions. Table 7 gives the results of this analysis. The first row of coefficients shows that the number of awardees has a significant but quantitatively modest negative effect on the measures of quality. A 10% increase in the number of awards is associated with a decline in average quantitative GRE scores by 1.1 points and a decline in verbal GRE scores by 2.2 points on the 200 to 800 GRE scale, and a drop in average GPA by 0.009 points (on a 1 to 4 scale). These findings are consistent with the notion that as the number of fellowships increases, the review panels move down the measured scholastic skill distribution to give stipends to marginal awardees who are less outstanding than the inframarginal winners. The second row of coefficients in Table 7 shows that, with the number of awards fixed, increases in the relative value of fellowships increases two of our measures of the skill of awardees, but has no discernible effect on the third measure. An increase in the relative value of the stipend by 10% raises the average quantitative GRE score by 2.9 points and raises the average verbal GRE score by 3.5 points. However, the average GPA of awardees appears invariant to the stipend values.

Overall, the table 7 results on the GRE scores suggest that at least some of the students on the margin of applying for a GRF fellowship are high-achieving students who can be drawn away from other career or academic opportunities through the prospect of larger awards. In the context of the occupational choice model of A.D. Roy (1951)[16], these results imply a positive correlation

---

[16]A.D Roy "Some Thoughts on the Distribution of Earnings," Oxford Economic Papers Vol. 3, 1951, pp. 135-14

between ability in science and engineering and ability in alternative careers: the able S&E students

who are drawn into the GRF Program by higher stipends perform well on general measures of

achievement (i.e., GRE scores) that are associated with success in other occupations.

To test this interpretation, we regressed the number of applicants with relatively high

measured skills (defined as those scoring 770 or over on the GRE quantitative exam, 680 on the

verbal GRE, and those with GPAs above 3.88 on a 4 point scale) on the relative value of stipends

and the number of awards given. By controlling for the number of awards, we get a measure of the

relative number of applicants with high skills, which should largely determine the measured skills of

awardees. Column 1 of Table 8 shows that a 10% increase in the stipend value raises the number of

applicants with 770 or over on the quantitative GRE by 12.2%. Column 2 shows a comparable

increase in the number of applicants with 680 or over on the verbal GRE. The column 3 regression

gives a 0.448 impact of stipend values on the number of students with very high GPAs[17].

In sum, higher stipend values incentivize enough high-achieving students to seek GRF offers

to enable the NSF selection procedure to increase the measured skills of awardees when stipend

values are higher.

The Economic Supply Link

Overall, our results show that the resources NSF allocates to stipends, through the number of

awards and their value, affects the supply of applicants, both in terms of numbers and measured

scholastic skills. As a summary of the economic supply link, we have graphed in figure 10 our

---

[17] We also estimated models with the log of the number of applicants in the regression. This
gave smaller positive coefficients to the relative stipend value in the regressions for the number
of applicants with high math and verbal GREs, because the number of applicants is dependent on
stipend values and associated with the number of applicants with high scores. It also raised the
number with high GPAs, but conditional on the number of applicants, the relative value of
stipends were modestly negatively associated the number of applicants with high GPAs.

measure of the supply of students seeking GRFs – the number of applicants relative to the number of S&E baccalaureates – and a measure of NSF's stipend budget (the number of awards offered in a particular year multiplied by the stipend amount) relative to GDP, which we take as the broadest possible measure of other economic activity. The stipend budget is an indicator of national resources spent to make graduate S&E study more financially attractive, whereas GDP represents national resources broadly. The tight link between the two series gives the bottom line message: if the US wants more applicants for S&E fellowships and thus potentially more graduate students and PhDs in these fields, the country has to spend more money on this objective.

IV. Alternative Policy Scenarios

The analyses thus far have shown that the two policy levers – numbers of GRFs and the dollar value of the awards -- affect the number and measured skills of persons seeking GRF awards and that coupled with NSF selection procedures, they determine the measured skills of awardees. In this section, we consider three scenarios that assess the way changes in the mode of determining stipend policy could affect the supply of students seeking and obtaining GRFs.

1. Agency determined changes in the number and value of awards.

2. Granting awards and setting values through fixed rules rather than policy discretion.

3. Allocating the number of quality group 1 and 2 positions among disciplines on the basis of quantifiable academic skills of applicants rather than on the number of applicants.

1. Agency Determined changes in stipend numbers and values

For this scenario, we assume that the goal of fellowship policy is to increase the number of the "best and brightest" pursuing S&E graduate studies, and that both the numbers and qualifications of GRF awardees contribute to the goal, as described in the appendix model. The agency is assumed

to face a fixed budget constraint, to reflect alternative uses of resources. Our finding that increasing the number of awards reduces average quality while increasing the value of awards raises quality and the budget constraint yields an interior solution that depends on the supply response of students to numbers of awards and to the stipend amount. The critical factor that determines whether it is better to give more awards or raise stipends are the characteristics of applicants who are on the margin of winning awards.

There are two ways to estimate the impact of changes in *awards on the margin* to the measured attributes of awardees. The first is to regress average GREs or other indicators of academic skills on the number of awardees, as we did in table 7, where the estimates suggested that average skills were not that sensitive to increased numbers of awards. The second way to assess skills on the margin is to go back to the micro records on individual attributes and rank applicants by the probability that they would get an award and then to examine the characteristics of those who would be given awards on the margin. To do this, we estimated the probability of getting awards in 2004 and then computed the quantitative and verbal GRE scores and GPA for awardees with the lowest probabilities in bins of fifty persons. We also estimated the average GRE and GPA for non-awardees with the highest probabilities of having gained an NSF, again in groups of 50. The difference in attributes between these non-awardees and the awardees in the lowest group reflects the potential change in quality from increasing the number of awards.

Figure 11 displays the results of our calculations in terms of the attributes of awardees and non-awardees. Persons in group 1 are awardees with the highest probability of getting an award. Persons in group 2 are awardees with the next highest probabilty of getting an award, and so on. Group five consists of awardees in the lowest probability group of getting an award. Groups 6-10

consist of persons who did not get an award, with those whom we estimated as most likely to get an award placed in group 6, followed by persons with less likelihood of getting an award in groups 7 - 10.   The figure shows that there is very little difference in the GRE Quantitative measures around the cutoff point between groups 5 and 6. This indicates that the number of awards could have been increased along a substantial margin without greatly reducing quality, or could have been reduced along a substantial margin without greatly increasing quality. But thefigure does show a noticeable drop in GRE Verbal scores between the marginal awardees and the marginal non-awardees. In part, this reflects the fact that while our model does a good job of predicting who gets an award, the errors in the model are for persons along the margin.

Table 9 simulates what would happen if NSF has $10 million additional funding to spend and divided the spending in three different ways: by increasing the number of awards, by increasing the value of awards, and by spending half of the additional money on increasing the number and half on increasing the value. Spending all of the money on increasing the number of awards at current stipend amounts would add 333 awardees and would give awards to person with modestly lower GRE quantitative and verbal scores than current awardees – 4.9 points lower in the quantitative GRE and 29.0 points lower in the verbal GRE. Since NSF gave 1020 awards, the impact of the lower scores for the 333 awardees on the average for all awardees is considerably smaller. Spending all of the money on increasing stipend values would raise awards by 32.7% and would change the overall mean GRE scores of awardees modestly. If such a large increase in stipends were to attract more able persons, this improved quality does not readily show up in the GRE measures. Finally, dividing the $10 million dollars between additional awards and higher valued stipends would increase the number of awards by 146. The increased stipend attracts just enough higher scoring applicants to

offset the mean quality decrease associated with more awardees. The result is virtually no change in the average measures of awardee skills.

It is easy to use our analysis to simulate the outcomes of other hypothesized changes in the NSF stipend budget and its allocation between numbers of awards and stipend values.

2. Using Fixed Rules to Determine the Number and/or Value of Stipends

An alternative way to set the number and/or value of stipends is to use a fixed rule to set them. Monetary economists and officials have often discussed using fixed rules to set the growth of money supply. Some central banks target a range of rate of inflation as their goal while others focus on keeping interest rates within some band. Fixed rules increase social security payments with inflation and have been proposed to raise minimum wages with inflation rather than through legislative means. If policy-makers chose a given adjustment rule, the value or number of stipends would increase according to the rule.

There are diverse rules that policy-makers could use to set stipend values and the numbers of stipends. Taking the value of stipends, one possible rule would be to set stipends as a percentage of alternative earnings in the labor market, such as the earnings of young college graduates. Another would be to make stipends a given percentage of the earnings of doctorates. Yet another would be to adjust stipends to measures of employment opportunities, such as unemployment rates or, given low rates of joblessness among S&E doctorates, the length of time to obtaining a job upon completion of a PhD, or to the growth of R&D spending. Since any single indicator would be subject to measurement error, perhaps the most sensible target would be a weighted average of indicators: stipends would rise/fall as the average of the indicators increased or decreased.

To set the number of stipends by a fixed rule, policy-makers could simply make them a given

proportion of the number of BS graduates, thereby assuring students in different cohorts the same

probability of gaining a stipend. When the supply of S&E baccalaureates increased, so too would

stipend availability. Alternatively one could use the same set of weighted indicators for setting the

number of stipends as for setting the value of stipends.

To get some sense of how a fixed rule would set the value and number of GRF stipends, we

have simulated what would happen to the number of awards and measured scholastic achievement of

GRF awardees under two conditions: (1) that NSF fixed the number of awards at 0.41% of S&E

bachelor's degrees, the level in 1968, before the ratio of awards to bachelor's graduates began to

fall; and (2) that NSF set stipends at the 2004 ratio of the value of stipends to the earnings of young

college graduates from the CPS, 115.8%. By setting numbers using a high awards per S&E

baccalaureate and setting values at a high value to earnings of young college graduates, we get an

upper bound on how a set of fixed rules would affect the supply of applicants to the field. Table 10

shows what would have happened from the 1970s to the 2000s if these fixed rules had been

employed. The greater number of awards would have lowered measured scholastic skills, but the

higher value of awards would have attracted enough applicants with greater skills to allow NSF to

raise average quality. Bottom line, these rules would have given awards to a larger number of

persons and, given our supply equations, would have raised the average quality of awardees, as well.

Alternative sets of rules would, of course, give different outcomes.

3. Equilibrating quality across disciplines.

Our final policy scenario considers what might happen if NSF changed its charge to panel

committees from giving awards in approximately the same proportion to the number of applicants

among fields to giving awards in a pure "measured scholastic" achievement, irrespective of the

number of applicants from different fields. Thus, if physics received 100 applicants, all of whom

had higher GREs, GPAs, and so on than, say social science, which had 500 applicants, all 100

physics applicants would receive awards before any social science applicant would receive an

award. To see how awards would be given among disciplines under thos hypothetical system, we

estimated a linear probability regression of actual award receipt on GRE Quant, GRE Verbal, GPA,

reference score (normalized), female indicator, and minority indicator in 1997, the last year for

which we had a full set of scholastic qualification measures. In 1997, NSF awarded 850 GRFs, so

under the hypothetical system, the applicants with the highest 850 predicted probabilities would be

offered an award. Since physical science and mathematics and computer science students have the

skills to shift to other fields and do well while students in the social sciences, psychology, and life

sciences cannot easily shift into the physical and mathematical sciences, we anticipated that this

change in policy would benefit the "harder" sciences.

The calculations summarized in table 11 tell a different story. Column 1 gives the number

and percentage of applicants in each field in 1997. Column 2 gives the number and percentage of

awardees in each field. The percentages of awards by discipline differ moderately from the

percentages of applicants by discipline, with engineering gaining more awards and life sciences

gaining fewer awards, due in part to the WECS program. Column 3 gives our hypothetical

distribution of awards by discipline on the basis of the measured scholastic attributes. The table note

shows the coefficients of the equation used to generate the hypothetical distribution. It gives equal

weight to the quantitative GRE and verbal GRE and a relatively heavy weight on the measure of

references.

The hypothetical distribution shows some differences from the 1997 actual distribution of

awardees among natural science fields: mathematics, chemistry, and physics /astronomy gain a few

awards while earth and atmospheric sciences and life sciences get fewer awards. The biggest

change, however, is between engineering and the social sciences and psychology. Engineering loses

a substantial number of awards, where social sciences and psychology gain. Why? The social

sciences have the highest mean verbal scores whereas engineering has the lowest mean verbal

scores. The social sciences and psychology score best on the reference letters. The GRE

quantitative scores favor the physical and mathematical sciences, including engineering, but a larger

proportion of students obtain top scores in the GRE quantitative test than in the GRE verbal test,

which puts a bound on the extent to which our hypothetical allocation gives awards to those fields.

Conclusion

Stipends to US citizens/residents are a natural policy tool for increasing the incentive for

Americans to enter S&E fields without directly impacting the flow of talent from overseas. Analysis

of NSF Graduate Research Fellowships suggests that raising the value of awards increases the

number of applicants and quality of awardees while giving more awards increases the number of

awardees, by definition, with only a modest reduction in measured academic skills. Because the

analysis is limited to a single stipend program, it is uncertain whether the finding that the number of

applicants responds to the relative value and relative number of stipends can be extrapolated to the

market for graduate students as a whole. To the extent that changes in NSF fellowship policy induce

changes in the policies of other stipend-granting groups, we suspect that the qualitative results, at

least, can be extrapolated to the broader market. At the minimum, the results are consistent with the

notion that more and better-paying stipends could raise the number of native-born/residents

choosing S&E fields broadly. To go further than that would require a more extensive analysis of the

decision of students to pursue S&E graduate studies as opposed to other options (as opposed to simply applying for the GRF) and a careful study of the responses of universities, other government agencies, and non-profit groups to NSF stipend policies.

REFERENCES:

[1] "Committee of Visitors (COV) Report for the National Science Foundation(NSF) Graduate Research Fellowship Program (GRFP)", June 17-18, 2003; Members: Mason, Agonafer, Givens, Jones, Petrochenkov, & Vollmer.

[2] "Report of the Committee of Visitors, National Science Foundation, The Graduate Fellowship Program", June 17-18, 1999, Members: Taylor, Denson, Flores, Gern, Sullivan, Syverson.

[3] National Science Board, "Science and Engineering Indicators 2004". Two volumes. Arlington, VA: National Science Foundation, Appendix Table 2-15

[4] NSF, Division of Science Resource Statistics, "Graduate Students and Postdoctorates in Science and Engineering: Fall 2001", Tables 39, 40.

[5] NSF, Graduate Research Fellowship Program (GRFP) Program Solicitation, NSF 04-615.

[6] WestEd, "National Science Foundation Graduate Research Fellowship Program: Final Evaluation Report", September 2002

[7] Richard B. Freeman, "Fellowship Stipend Support and the Supply of S&E Students: NSF Graduate Research Fellowships", Annual Meeting of the American Economic Association, Philadelphia PA, January 7-9, 2005

[8] NSF Division of Graduate Education, "Guideline for Panelists. In the Review and Rating of National Science Foundation Graduate Research Fellowship Applications," Program Year 2002.

[9] National Science Foundation, Division of Science Resources Studies, *Science and Engineering Degrees: 1966-98*, NSF 01-325, Author, Susan T. Hill (Arlington, VA 2001).

[10] Sheppard, Rutledge, & Johnson, "Merit Criteria, Eligibility and Diversity in the NSF Graduate Fellowships", Proc. of the 2001 Amer. Soc. for Eng. Educ. Annual Conference and Exposition,

American Society for Engineering Education.

[11] Joan Burrelli, "A comparison of the early career characteristics of NSF fellows and non-fellows", (2002) National Science Foundation

[12] NSF, "Scientific and Technical Manpower Resources: Summary Information on Employment Characteristics, Supply and Training", (1964) NSF 64-28

[13] Ruggles,Sobek,Alexander, Fitch, Goeken, Kelly Hall, King, and Ronnander. Integrated Public Use Microdata Series: Version 3.0,. Minneapolis, MN: Minnesota Population Center (2004)

Appendix A: Model of Optimal Stipend Policy

As a starting point for analyzing stipend policy, we consider the choices and constraints faced by an abstract granting agency. We assume that the granting agency cares about the number of awards (N) that it grants each year and the average quality (Q) of the students who receive the awards. The real-world analogue is that the NSF would like to support as many students as possible for graduate study in S&E, but at the same time the NSF would like its fellowship recipients to be as high-achieving as possible. The agency thus seeks to maximize

(1) $\qquad$ $U = a\text{N} + (1 - a)\text{Q}$,

a weighted average of N and Q where a is the weight placed on numbers. The agency has two policy levers to do this: the numbers of stipends (N) and the annual value of stipends (V). The agency faces a budget constraint

(2) $\qquad$ $\text{NV} = \text{K}$

where K is the (fixed) total annual budget allocated to stipends.
The agency's choice of N and V affects the quality of awardees. Average recipient quality is modeled as

(3) $\qquad$ $Q = Q(\text{V, N, other factors})$.

where $\partial Q/\partial V > 0$ because higher-valued stipends are able to attract higher-achieving candidates who would otherwise have pursued alternative fields and careers or seek other fellowships. We also expect that $\partial Q/\partial N < 0$: average quality of the awardees declines with the number of awards because the highest-achieving students are generally the first to obtain awards. The optimal policy in this model, derived from maximizing the objective (1) subject to the constraints (2) and (3), is characterized by the following condition:

(4) $\qquad$ $a = (1-a)\left(-\dfrac{\partial Q}{\partial N}\right) + (1-a)\dfrac{\partial Q}{\partial V}\dfrac{V}{N}$

The optimal policy is to equate the marginal benefit and marginal cost of funding another student. The marginal benefit, given on the left hasd side of (4), is simply the subjective value (a) that the agency places on being able to support an additional student. The marginal cost, on the right hand side, consists of two terms. The first term denotes the marginal decrease in quality resulting directly from giving more awards. The second term reflects the fact that, in the face of a fixed budget, a larger number of awards necessitates a lower annual stipend value, which in turn lowers the average quality of awardees.
This model suggests that the relationship between stipend values and awardee quality is of central interest. How do higher stipend values bring higher-achieving students into the GRFP program? The main mechanism underlying this relationship is that higher stipend values attract a larger applicant pool, and at least some of the additional applicants are highly able candidates who are pulled away from alternative careers (or from immediate entry into the labor market). Essentially, larger applicant pools allow the agency to be more selective in granting awards. Thus, it is also of interest to examine the factors that determine the number of applicants to the GRFP program. We model the number of applicants (A) as

(5) $\qquad$ $A = A(\text{N}_{-1}, \text{V, Other factors})$

where N_{-1} is the number of awards granted in the previous year. We predict that $\partial A/\partial V > 0$ and $\partial A/\partial N_{-1} > 0$; that is, more students apply when stipend values increase and when the perceived probability of receiving an award rises. Other factors predicted to influence the applicant pool include the attractiveness of alternative careers, as discussed in the next section.

Thus, equations (1) through (5) capture the quality-quantity choice facing a stipend-granting agency: to spend its budget on funding many students versus funding a few "superstars." Since identification of stars is difficult in most disciplines, the choice is in fact more complex. The key behavioral relations to be empirically estimated are: the change in the quality of students as the number of awardees is increased; the change in quality in response to the value of stipends; and the change in the number of applicants as stipend values rise. If the agencies responsible for stipend policy optimizes as in this model, then the preceding model reflects the basic tradeoffs factoring into the agencies' decisions. In this case the number and monetary value of GRFP fellowships cannot be assumed to be exogenous for the purposes of estimating behavioral relationships. On the other hand, if the number and value of GRFP fellowships depend on vagaries of budgeting, then adjustments to the number and value of stipends constitute "natural experiments" for assessing the effect of stipend policy.
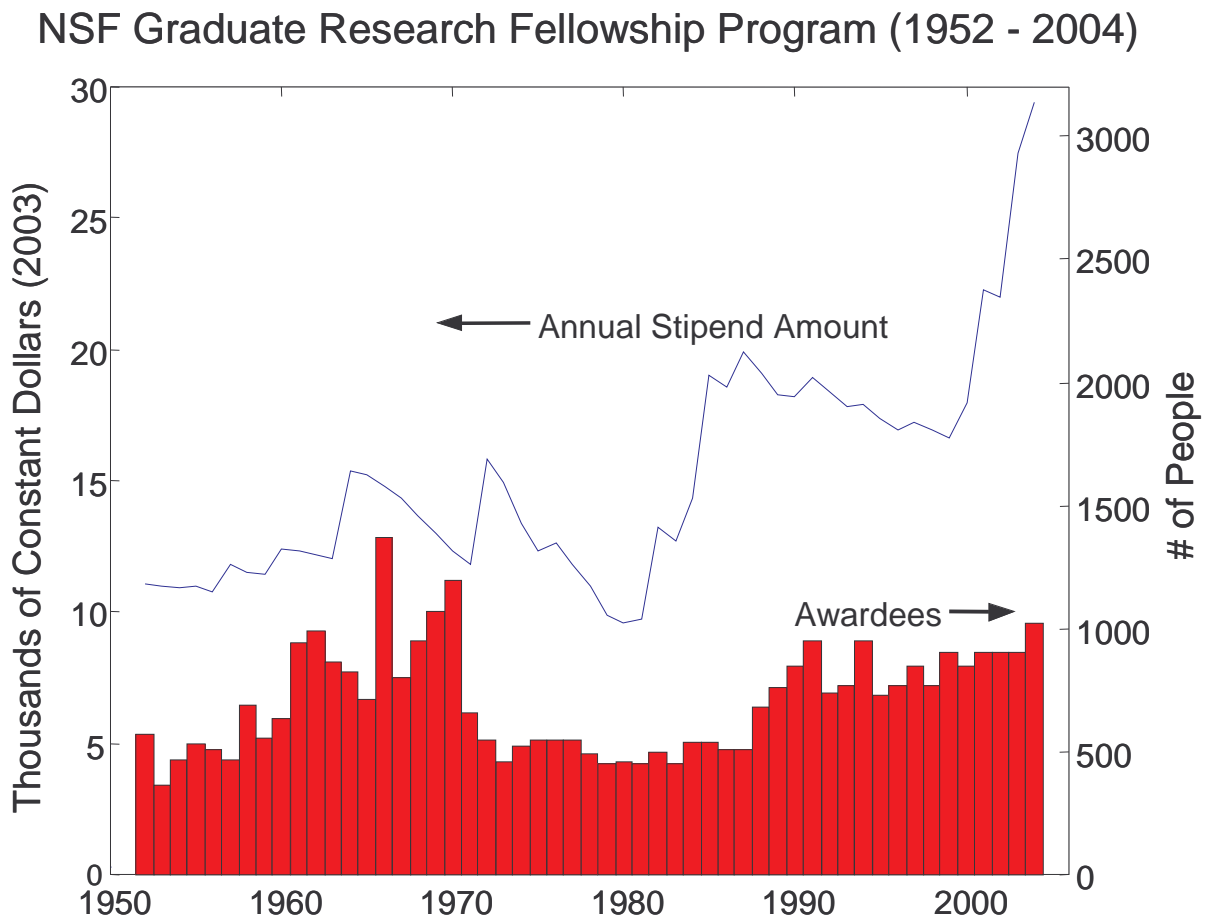
Figure 1. The number and value of GRF awards vary over time.
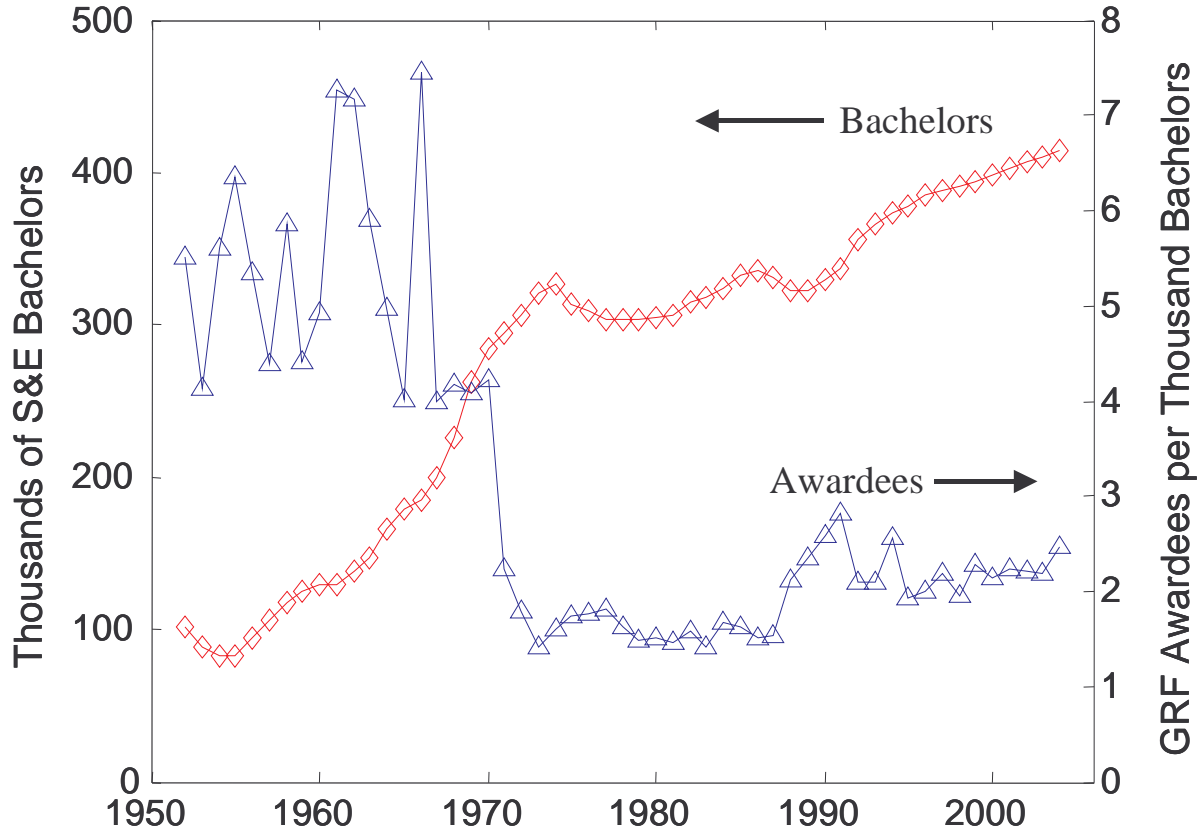
Figure 2. The Number of Awards per S&E Baccalaureate Has Shifted Downwards.

Source: NSF-DGE, Cumulative Index and related Datasets. Bachelor's degree data tabulated by National Science Foundation/Division of Science Resources Studies; data from Department of Education
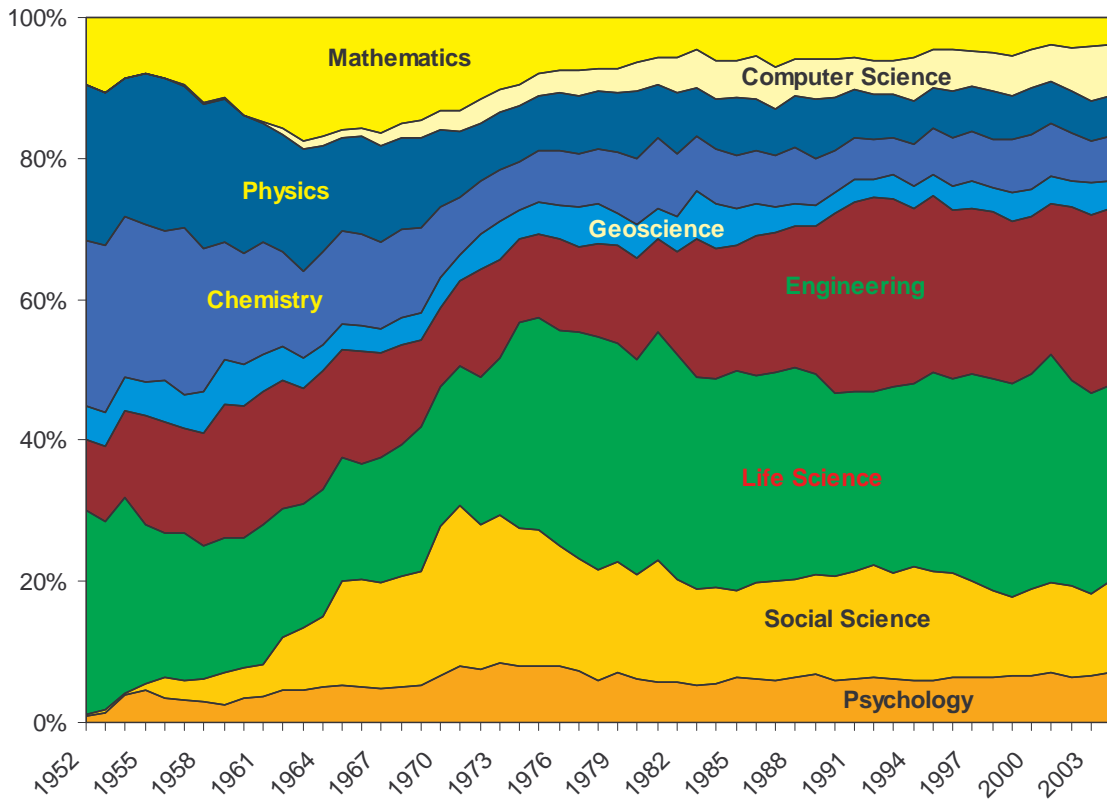
Figure 3. The GRF Disciplinary Distribution Over Time of Applicants.

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets.
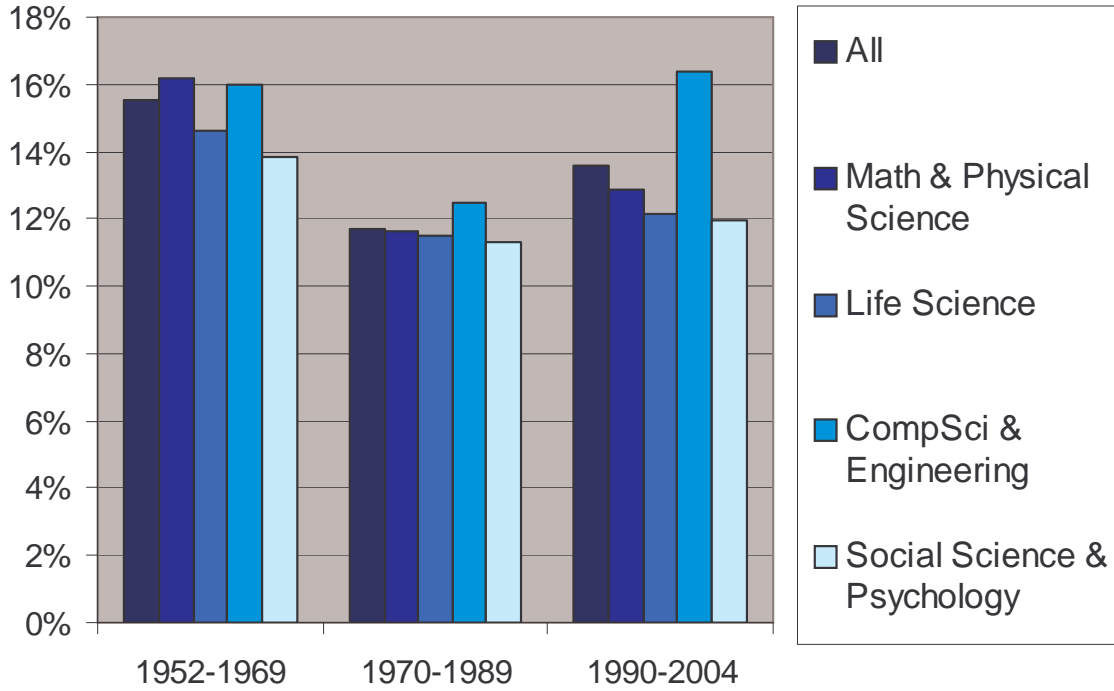
Figure 4. Percentage of Applicants Who Gain GRF Awards, Selected Years.

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets.
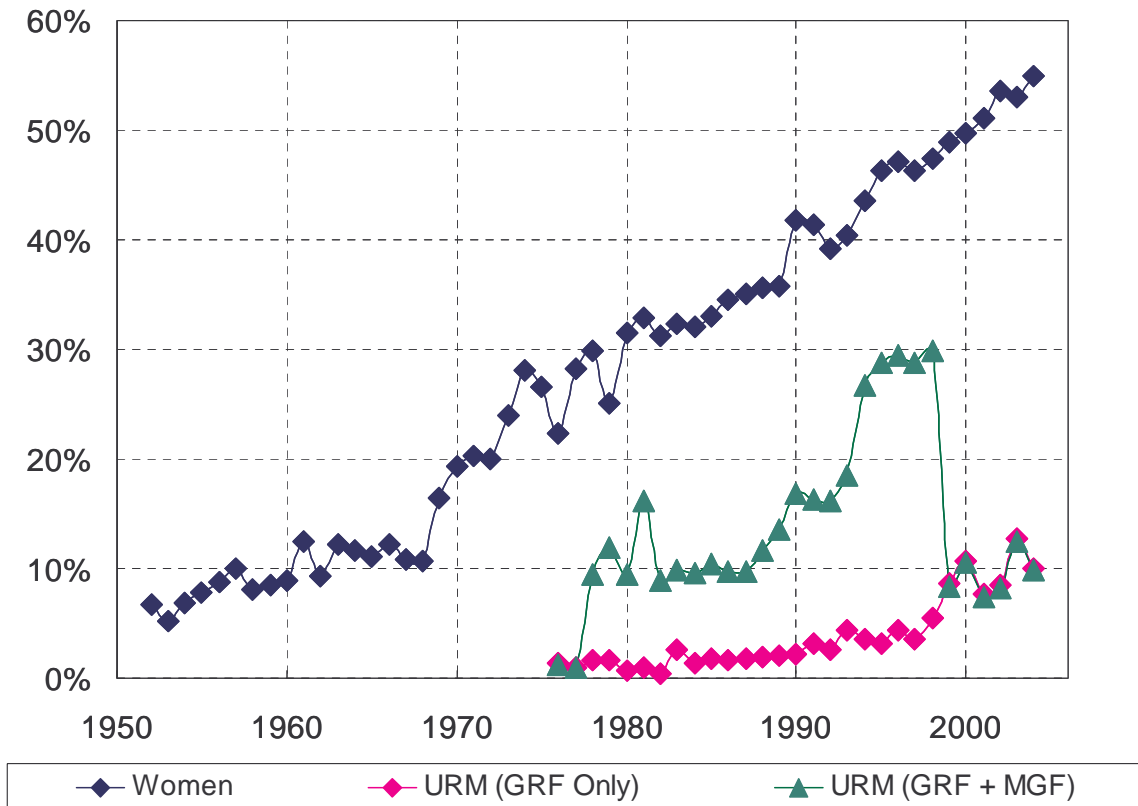
Figure 5. Percentages of Fellows who are women or URMs.

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets.
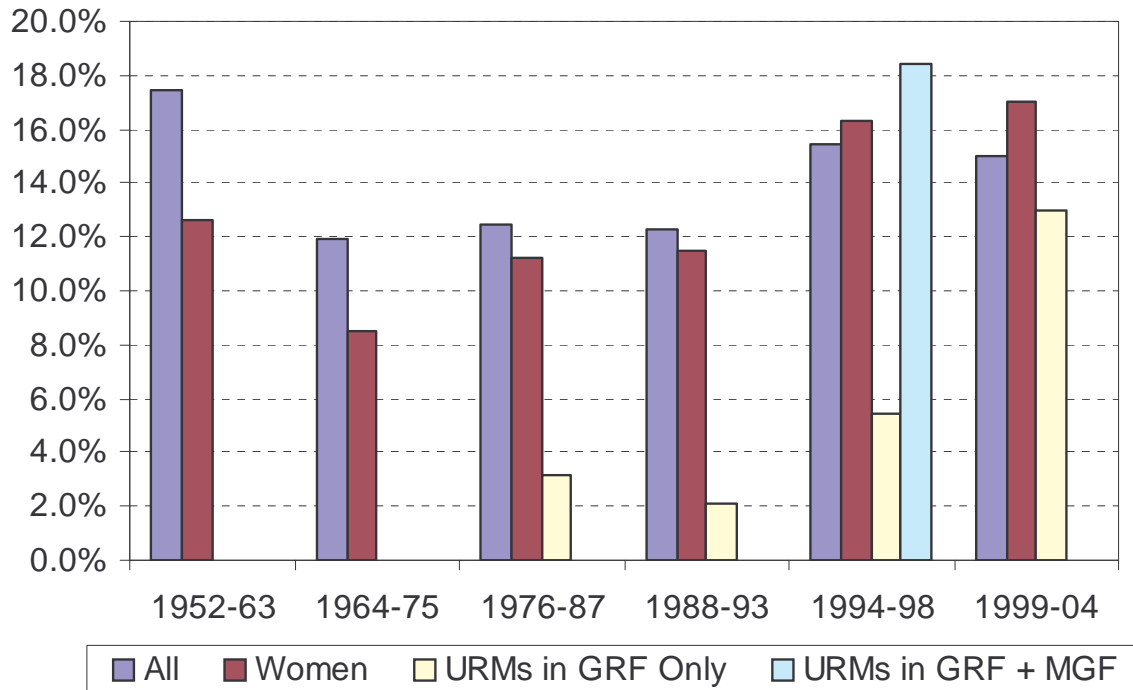
Figure 6. Average Percentage of Applicants Winning Awards in Selected Time Periods, by Broad Demographic Group.

Source and Note: Data on ethnic demography only available beginning in 1976. From 1994 - 1998 the available data allows the identification of duplicate applicants to both the MGF & GRF. To calculate "URMs in GRF + MGF" the total number of unique fellows in either program was divided by the total number of unique applicants. The final year of the MGF program was 1998. NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets.

Figure 7. GRFP Applications Relative to S&E Bachelor's Degrees And Relative Value of GRFP Stipend, 1968-2004

Figure 8. GRFP Applications Relative to S&E Bachelor's Degrees And College Graduate Unemployment Rate, 1979-2004
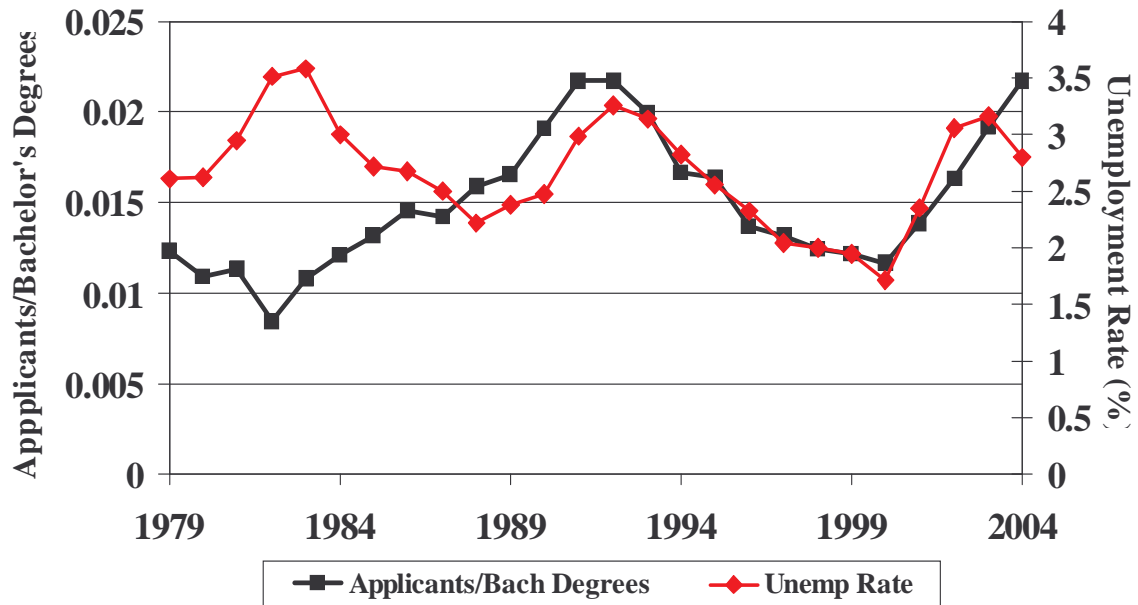
Source: NSF DGE, Cumulative Index of the GRF Program and related datasets. Bachelor's degree data tabulated by National Science Foundation/Division of Science Resources Studies; data from Department of Education. Unemployment data from estimated from the annual Current Population Survey (CPS) Outgoing Rotation Group, Bureau of Labor Statistics.

**Figure 9A. Mean GRE Quantitative Scores of Awardees
And Relative Stipend Value, 1968-2004**
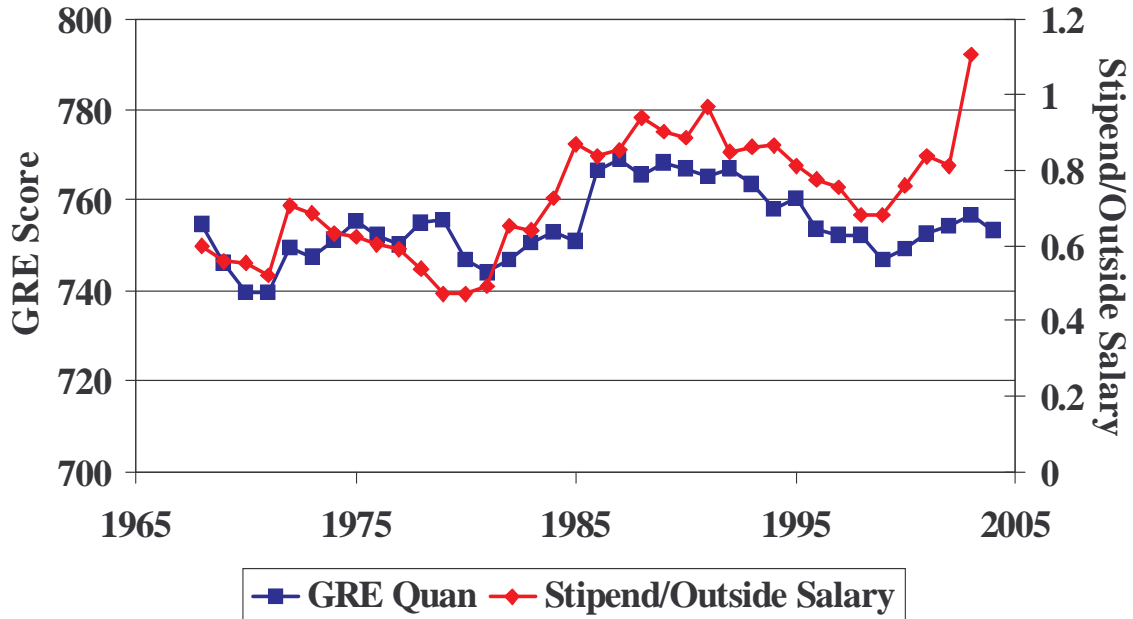
Source: NSF DGE, Cumulative Index of the GRF Program and related datasets. Salary data estimated from the Integrated Public Use Microdata Series (IPUMS) of the March Current Population Survey.

Figure 9B. Mean GRE Verbal Scores of Awardees and Relative Stipend Value, 1968-2004
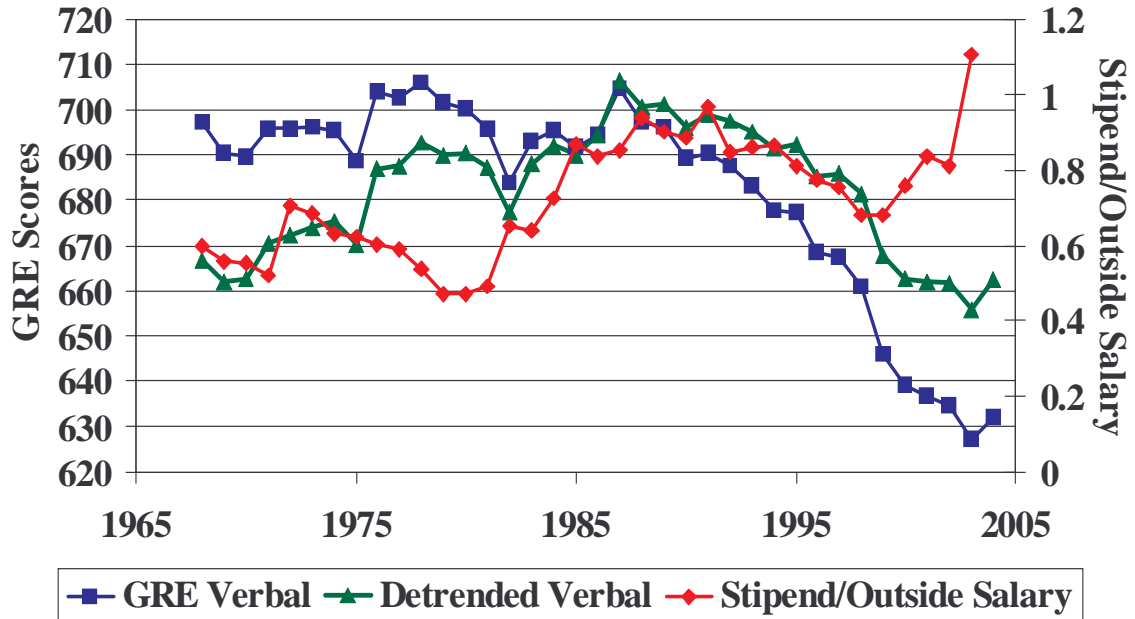
Source: NSF DGE, Cumulative Index of the GRF Program and related datasets. Salary data estimated from the Integrated Public Use Microdata Series (IPUMS) of the March Current Population Survey.
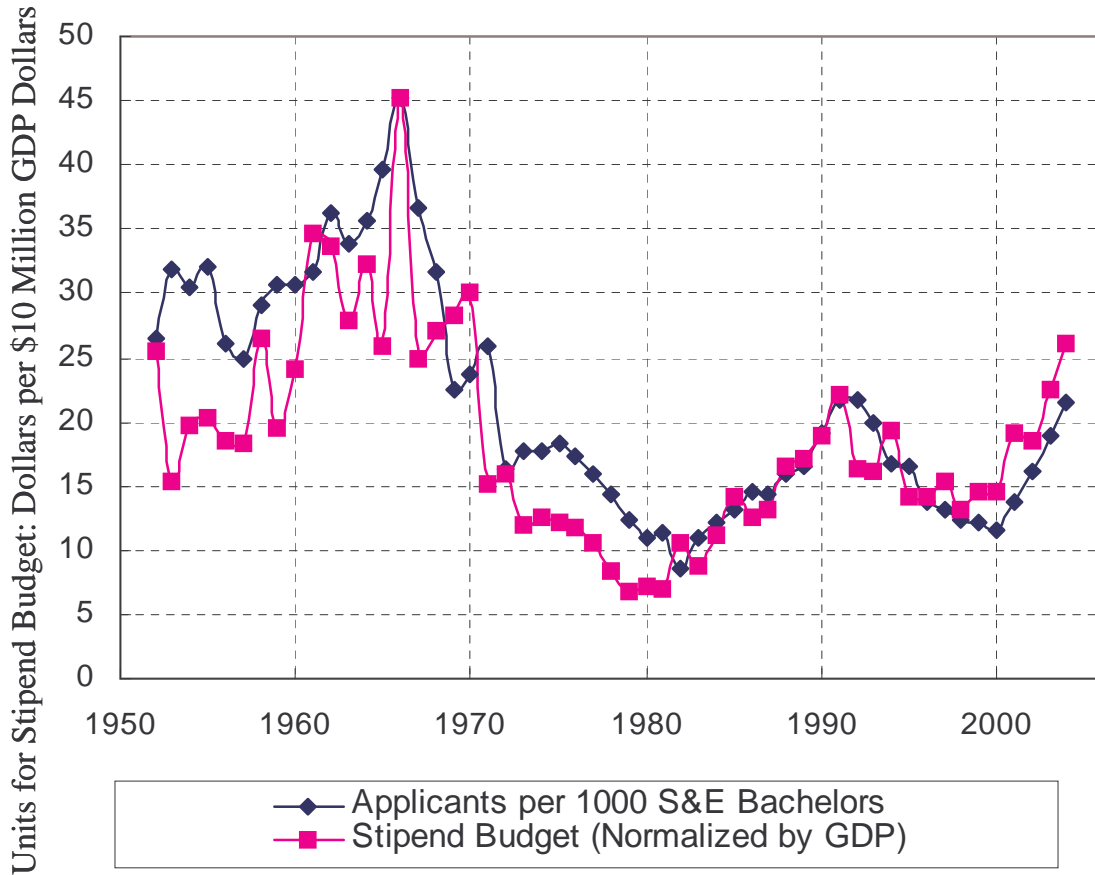
Figure 10. Fraction of Bachelors choosing to apply to GRF vs. total GRF stipend budget/GDP.

Source: NSF DGE, Cumulative Index of the GRF Program and related datasets. Data on the Gross Domestic Product (GDP) from the Bureau of Economic Analysis, an agency of the U.S. Department of Commerce.
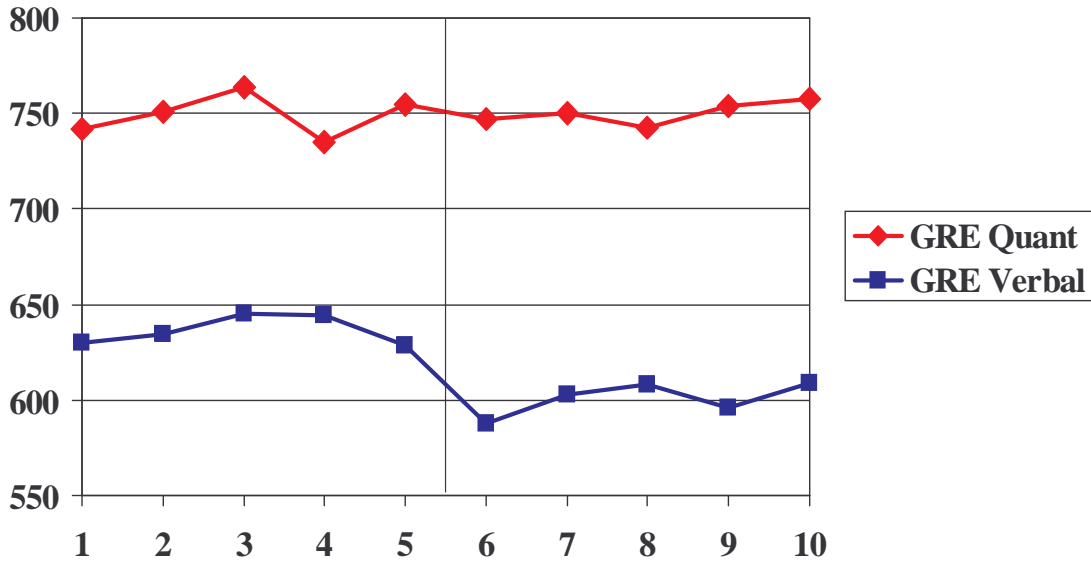
Figure 11. Quality of GRF Applicants on the Margin of Getting an Award.

Source and Notes: All persons to the left of the line were given awards. All persons to the right of the line did not get awards. The numbers relate to groups of 50 persons, ordered by the estimated probability they would win an NSF award. The 5th group consists of 50 awardees with the lowest probability of getting an award, the 4th group consists of 50 awardees with the next lowest probablity, and so on. The 6th group consists of the 50 non-awardees with the highest probablity of getting an award. The probabilities are predicted values from an OLS regression of an award receipt dummy variable on panel rating, female dummy, underrepresented minority dummy, and eight field dummies.

Table 1: Means, Standard Deviations, and Scaling of the Quantitative Measures of Applicant Scholastic Skills

Summary for Key Data in the GRF Datafiles	Total Time Period	Records	Overall Award Rate	
	1952 - 2004	271,391	13.50%	
	Data Available	Mean	STD	Full Range	Notes
Panel Ratings	1976 - 1993	328	115	100 to 600	For analysis, Panel Ratings and Reference Scores will be normalized such that they have mean zero and standard deviation one in the applicant pool of each year. These variables are also recoded such that higher values denote better ratings.
	1994 - 1998	315	118	100 to 600	
	1999	3.11	1.02	1.00 to 6.00	
	2000 - 2004	2.90	0.85	1.00 to 4.99	
Reference Score	1954 - 1993	24.2	8.7	10 to 70	
	1994 - 1998	1.91	0.65	1 to 7	
GRE Quant	1952 - 2004	691	100	200 to 800	Values in original dataset above 800 set to 800
GRE Verbal	1952 - 2004	606	104	200 to 800	
GPA	1955 - 1998	3.49	0.40	1 to 4	Full Range (0 to 3) before 1971, values shifted by +1
Ethnicity	1976 - 2004	(1, 2, 3, 4, 5)			
	1981 - 1987	(1, 2, 3, 4, 5, 6, 7, 8, 9)			
	1988 - 2000	(A, B, C, D, E, F, H, I, J)			
	2000 - 2004	(t, f) or (y,n,b)			a separate variable codes each ethnic category
Gender	1952 - 2000	(m, f)			
	2001 - 2004	(m, f, b)			value of "b" indicates "decline to answer"

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets.

Table 2. Linear Probability Model for the Impact of Scholastic and Demographic Variables on GRF Awards.

	Offered Award		
	1955-98	1976-98	1976-98
Panel Rating			0.170
			(0.001)
Reference Score	0.082	0.074	0.022
	(0.001)	(0.001)	(0.001)
GRE Quant/100	0.034	0.035	-0.010
	(0.001)	(0.001)	(0.001)
GRE Verbal/100	0.060	0.066	0.019
	(0.001)	(0.001)	(0.001)
GPA	0.089	0.095	-0.010
	(0.002)	(0.003)	(0.003)
Female		0.033	0.037
		(0.002)	(0.002)
Minority		0.087	0.077
		(0.004)	(0.003)
Field Effects:			
Chemistry	-0.003	0.011	0.045
	(0.003)	(0.004)	(0.004)
Computer Science	-0.007	0.009	0.115
	(0.004)	(0.005)	(0.005)
Engineering	0.022	0.048	0.093
	(0.003)	(0.004)	(0.003)
Earth/Atmospheric	0.033	0.034	0.049
	(0.004)	(0.005)	(0.005)
Life Science	0.037	0.034	0.040
	(0.002)	(0.003)	(0.003)
Math	-0.019	-0.012	0.051
	(0.003)	(0.005)	(0.005)
Physics/Astronomy	-0.030	-0.006	0.060
	(0.003)	(0.005)	(0.004)
Psychology	0.000	0.005	0.018
	(0.003)	(0.005)	(0.004)
Social Science			
Year Effects	Yes	Yes	Yes
Observations	207498	107658	107597
R-squared	0.198	0.1793	0.2731

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets.

Table 3 . Regression Coefficients and Standard Errors For Scholastic and
Demographic Determinants of Average Panel Rating, 1976-1998.

	Average Panel Rating
	1976 - 1998
Reference Score	0.304
	(0.002)
GRE Quant/100	0.268
	(0.003)
GRE Verbal/100	0.279
	(0.002)
GPA	0.614
	(0.006)
Female	-0.024
	(0.004)
Minority	0.057
	(0.007)
Field Effects:	
Chemistry	-0.198
	(0.009)
Computer Science	-0.629
	(0.010)
Engineering	-0.263
	(0.007)
Earth/Atmospheric	-0.092
	(0.011)
Life Science	-0.038
	(0.006)
Math	-0.367
	(0.010)
Physics/Astronomy	-0.386
	(0.009)
Psychology	-0.079
	(0.009)
Social Science	
Year Effects	Yes
Observations	107597
R-squared	0.6079

Table 4. Regression Coefficients for Determinants of Panel Rating, Quality Group, and Award Offer, 1990-2004.

	1990 - 1997			1998 - 2004		
	Panel Rating	Assigned Quality Group 1 or 2	Award conditional on being group 1 or 2	Panel Rating	Assigned Quality Group 1 or 2	Award conditional on being group 1 or 2
	1	2	3	4	5	6
Female	-0.012	-0.021	0.38	0.135	0.024	0.25
	(0.006)	(0.003)	(0.014)	(0.009)	(0.004)	(0.012)
Minority	-0.003	0.076	0.072	0.161	0.091	0.073
	(0.009)	(0.004)	(0.041)	(0.015)	(0.006)	(0.021)
Panel Rating			0.275			0.546
			(0.039)			(0.028)
Quant/100	0.263	0.035		0.334	0.081	
	(0.004)	(0.002)		(0.007)	(0.003)	
Verbal/100	0.235	0.074		0.285	0.093	
	(0.003)	(0.002)		(0.005)	(0.002)	
GPA	0.634	0.118				
	(0.010)	(0.004)				
Reference	0.289	0.093				
	(0.003)	(0.002)				
Field Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47821	47851	4904	41404	41404	5695
R-squared	0.5882	0.2109	0.1445	0.215	0.0901	0.1574

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets

Note: For comparison with the 1998-2004 analyses, we estimated the 1990-97 equations excluding GPA and reference scores. The coefficients (standard error) on the dummy for female barely changed: it was -0.012 (.006) in the panel rating equation and -0.026 (.003) in the equation for being assigned to group 1 or 2. For minorities, the coefficient on the dummy for minority status in the panel rating beaome -0.241 (.011). The implication is that GPA and reference scores for minority applicants were much lower than for others. But the probability of being assigned to group 1 or 2 gave to minorities a positive, though smaller, coefficient than in the table, .018(.003)

Table 5. Demographic Correlates of Test Scores, Regression Coefficients and Standard Errors, 1976 - 2004

	GRE QUANTITATIVE TEST SCORES				GRE VERBAL TEST SCORES			
	1976-1983	1984-1991	1992-1998	1999-2004	1976-1983	1984-1991	1992-1998	1999-2004
Female	-41.7	-36.3	-35.8	-28.9	0.0	-10.6	-16.1	-7.2
	(1.0)	(0.9)	(0.8)	(0.8)	(1.2)	(1.0)	(1.0)	(1.0)
Black	-175.9	-166.5	-146.3	-100.1	-168.7	-160.1	-146.5	-100.4
	(3.2)	(1.8)	(1.4)	(1.9)	(3.8)	(2.1)	(1.7)	(2.4)
Hispanic	-77.5	-98.0	-87.1	-54.7	-80.5	-107.8	-95.9	-70.6
	(3.5)	(2.0)	(1.5)	(1.7)	(4.2)	(2.3)	(1.9)	(2.2)
Other Minority	-49.2	-73.9	-69.1	-44.7	-41.0	-69.8	-60.7	-38.7
	(8.1)	(4.1)	(3.1)	(3.4)	(9.5)	(4.9)	(3.8)	(4.4)
Field Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30213	41045	41804	36658	30213	41045	41804	36659
R-Squared	0.3256	0.3688	0.3778	0.2536	0.116	0.1868	0.2273	0.1279

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets.

Table 6 Determinants of the Number of GRFP Applicants, 1969-2004

	Dep Var: ln(applicants in academic field in current year)			
	(1)	(2)	(3)	(4)
Log(Bachelor's Degs) by Field in Current Yr	0.195	0.304	0.298	0.516
	(0.057)	(0.063)	(0.062)	(0.066)
Log(Stipend/Outside Salary) in Previous Yr	0.996	0.916	0.852	0.772
	(0.084)	(0.060)	(0.059)	(0.056)
Unemp Rate for College Grads Age 21-25		0.049		
		(0.013)		
Unemp Rate for All College Grads			0.104	0.094
			(0.024)	(0.022)
Log(Awards/Bachelor's Degs) by Field in Previous Yr				0.349
				(0.054)
Field Effects	Yes	Yes	Yes	Yes
Field x Time Trend	Yes	Yes	Yes	Yes
Observations	324	234	234	234
R-squared	0.8931	0.955	0.9561	0.9634

Source: NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets, as described in text. Outside salary are earnings of college graduates aged 21-25. Outside salaries and unemployment rates from Current Population Survey.

Table 7. Determinants of Awardee Achievement, 1969 - 2004

	GRE Quant	GRE Verbal	GPA
Log(Number of Awards) by	-10.6	-21.8	-0.087
Field in Current Yr	(1.80)	(3.30)	(0.01)
Log(Stipend/Outside Salary) in	29	35.4	-0.007
Previous Yr	3.3	6.1	0.017
Field Effects	Yes	Yes	Yes
Field x Time Trend	Yes	Yes	Yes
Observations	324	324	270
R-squared	0.8943	0.684	0.7354

Source: Tabulated from NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets, as described in text. Outside salary are earnings of college graduates aged 21-25, tabulated from Current Population Survey.

Table 8. Determinants of the Number of Applicants Scoring Above High Thresholds

	Log(Applicants with GRE Quant > 770)	Log(Applicants with GRE Verbal > 680)	Log(Applicants with GPA > 3.88)
	(1)	(2)	(3)
Log(Stipend/Outside Salary) in Previous Yr	1.221	1.276	0.448
	(0.104)	(0.086)	(0.100)
Log(Number of Awards) by Field in Current Yr	0.539	0.664	0.512
	(0.090)	(0.074)	(0.077)
Field Effects	Yes	Yes	Yes
Field x Time Trend	Yes	Yes	Yes
Observations	261	261	207
R-squared	0.9101	0.9097	0.9235

Source: Tabulated from NSF, Division of Graduate Education, Cumulative Index of the GRF Program and related datasets, as described in text. Outside salary are earnings of college graduates aged 21-25, tabulated from Current Population Survey.

Table 9. Impact of Hypothetical $10 Million Increase in Annual Funding On Selected Outcomes in 2004, By Alternative Uses of Funds

	CHANGE IN ANNUAL NUMBER OF AWARDEES	% CHANGE IN RELATIVE VALUE OF STIPEND	DIFFERENCE BETWEEN MEAN QUALITY OF EXTRA AWARDEES AND MEAN QUALITY OF ORIGINAL AWARDEES		CHANGE IN OVERALL MEAN QUALITY OF AWARDEES	
			GRE Quant	GRE Verbal	GRE Quant	GRE Verbal
	(1)	(2)	(3)	(4)	(5)	(6)
ALL EXTRA FUNDS TO INCREASING NUMBER OF AWARDS	333	0.00%	-4.9	-29	-1.2	-7.1
ALL EXTRA FUNDS TO INCREASING VALUE OF STIPENDS	0	32.70%	N.A.	N.A.	8.2	10
HALF TO ADDITIONAL AWARDS AND HALF TO HIGHER STIPEND	146	14.30%	N.A.	N.A.	3.3	0.7

Notes: Non-awardees in 2004 were ranked by their predicted probability of award receipt; see text for details. For a given increase of N in the number of awards, columns 3 and 4 are calculated by subtracting the mean quality of the actual 2004 awardees from the mean quality of the top N non-awardees. In column 5 (6), the change in the overall mean quality induced by more awardees is calculated by multiplying column 3 (4) by N/(1020+N), given that 1020 awards were actually granted in 2004. The change in the overall mean quality induced by higher stipend values is calculated from the coefficients in column 1 (3) of Table 7. The change in overall mean quality induced by both more awards and higher stipend values is the sum of the two preceding changes.

Table 10. Hypothetical GRFP Characteristics from Maintaining 1968 Relative Award Availability and 2004 Relative Stipend Value

Period	ACTUAL PROGRAM CHARACTERISTICS (Annual Averages)					HYPOTHETICAL PROGRAM CHARACTERISTICS						
	Awards	Awards As % of S&E Bach Degrees	Stipend as % of Alt. Wage	Average Quality of Awardees		Awards	Change in Awardee Quality from Actual Due to More Awards		Change in Awardee Quality from Actual Due to Higher Stipend		Predicted Quality of Awardees	
				Quant	Verbal		Quant	Verbal	Quant	Verbal	Quant	Verbal
1968-1971	921	0.35%	55.70%	745.2	692.8	1082	-1.7	-3.5	21.2	25.9	764.7	715.2
1972-1975	511	0.16%	66.00%	751	693.8	1285	-9.8	-20.1	16.3	19.9	751	693.8
1976-1979	504	0.17%	55.00%	753.1	703.6	1237	-9.5	-19.6	21.6	26.3	765.2	710.4
1980-1983	460	0.15%	56.30%	746.9	693.1	1261	-10.7	-22	20.9	25.6	757.1	696.6
1984-1987	520	0.16%	82.00%	759.6	696.5	1342	-10.1	-20.7	10	12.2	759.6	688
1988-1991	807	0.25%	92.20%	766.3	692.9	1330	-5.3	-10.9	6.6	8.1	767.7	690.1
1992-1995	794	0.22%	84.50%	762	681.2	1493	-6.7	-13.8	9.1	11.2	764.4	678.6
1996-1999	820	0.21%	72.10%	751.1	660.3	1580	-7	-14.3	13.8	16.8	757.9	662.8
2000-2004	915	0.23%	93.40%	753.2	633.9	1640	-6.2	-12.7	6.3	7.6	753.3	628.8

Notes: "Hypothetical" characteristics denote those predicted to prevail if awards as % of S&E bachelor's degrees had been maintained at 0.41% (its 1968 value) and stipend as % of alternative wage had been maintained at 115.8% (its 2004 value). Hypothetical changes in GRE Quant and GRE Verbal are based on coefficients in columns 1 and 3 of Table 7. In the calculations, actual numbers of bachelor's degrees in 2002-2004 were imputed from 2000-2001 field-specific growth rates in earned bachelor's degrees, and actual alternative wage in 2004 was imputed from 1983-2003 growth rate in salary of 21-25 yr olds with bachelor's degrees.

Table 11. Existing and Hypothetical Allocation of GRFP Awards across Fields in 1997.

Field	Number of Candidates (Percent in Parentheses)			Mean Applicant Quality			
	Applicants	Awardees Actual	Awardees Hypothetical	Quant	Verbal	GPA	Reference
Chemistry	358	57	69	709	586	3.66	-0.05
	(7.0)	(6.7)	(8.1)				
Comp Sci	261	47	46	730	600	3.66	-0.02
	(5.1)	(5.5)	(5.4)				
Engineering	1212	252	184	728	573	3.67	0.05
	(23.6)	(29.6)	(21.7)				
Earth/Atmos	194	32	25	690	603	3.57	-0.05
	(3.8)	(3.8)	(2.9)				
Life Science	1502	210	198	686	605	3.54	-0.09
	(29.3)	(24.7)	(23.3)				
Math	242	41	59	747	614	3.69	-0.12
	(4.7)	(4.8)	(6.9)				
Physics/Astro	330	58	62	749	632	3.72	-0.26
	(6.4)	(6.8)	(7.3)				
Psychology	331	48	63	653	615	3.57	0.27
	(6.4)	(5.6)	(7.4)				
Social Sci	698	105	144	667	642	3.60	0.19
	(13.6)	(12.4)	(16.9)				
Total	**5128**	**850**	**850**	**702**	**604**	**3.62**	**0.00**
	(100.0)	**(100.0)**	**(100.0)**				

Notes: Awards under the "hypothetical" system are allocated as follows. A linear probability regression of actual award receipt on GRE Quant, GRE Verbal, GPA, reference score (normalized), female indicator, and minority indicator was used to predict the probability of being offered an award in 1997. Under the hypothetical system, applicants with the highest 850 predicted probabilities are designated as being offered an award.

Regression with robust standard errors

Number of obs = 4566
R-squared = 0.1772

```
-------------------------------------------------------------------------------
      gotaward        |     Coef.      Std. Err.        t
      quant/100       |    .0541323    .0062852       8.61
      verbal/100      |    .0582354    .0056672      10.28
      gpa             |    .1265555    .0153695       8.23
      refscore        |    .0879098    .0050868      17.28
      female          |    .059188     .011009        5.38
      minority        |    .0319515    .0126827       2.52
      _cons           |   -1.04754     .0670683     -15.62
-------------------------------------------------------------------------------
```