NBER WORKING PAPER SERIES

HOW MUCH DO MEDICAL STUDENTS KNOW ABOUT PHYSICIAN INCOME?

Sean Nicholson

Working Paper 10542
http://www.nber.org/papers/w10542

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2004

I am grateful to Jon Veloski and Mary Robeson from Jefferson Medical College for providing the data. Rachel Croson, Robert Lemke, two anonymous referees, and members of The Wharton School's Decision Processes workshop provided helpful comments. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

How Much Do Medical Students Know About Physician Income?
Sean Nicholson
NBER Working Paper No. 10542
May 2004
JEL No. J24, J44, I11

ABSTRACT

Twenty-five cohorts of medical students were asked in their first and fourth year of school to estimate contemporaneous physician income in six different specialties. The students' income estimation errors varied systematically over time and cross-sectionally by specialty and type of student. The median student underestimated physician income by 15 percent, and the median absolute value of the estimation errors was 26 percent of actual income. Students were 35 percent more accurate when estimating market income in their fourth relative to their first year, which indicates medical students learn a considerable amount before choosing a specialty.

Sean Nicholson
The Wharton School
University of Pennsylvania
3641 Locust Walk
Philadelphia, PA 19104
and NBER
nicholss@wharton.upenn.edu

I. Introduction

Most labor economists assume that the formation of income expectations is homogeneous and people have unbiased expectations: people have access to the same information, process information in the same way when forming income expectations, and the mean ex ante prediction error of those expectations is zero (Manski 1993). However, most studies that compare people's realized incomes to their subjective expectations find that the mean ex post prediction error is nonzero (Das and van Soest 1997; Dominitz and Manski 1997; Dominitz 1998; and Nicholson and Souleles 2003). As these authors point out, one explanation is that expectations were rational ex ante, but positive or negative shocks occurred after expectations were formed to create the observed bias. Another explanation for nonzero prediction errors is that income expectations are biased ex ante because people demonstrate bounded rationality and systematically misinterpret information.

In this paper I examine how accurately medical students estimate contemporaneous physician income, whether they make systematic errors, the determinants of income estimation errors, and how much students learn about market income during medical school. If people systematically misinterpret information that affects income expectations, they may choose occupations and schooling levels other than those they would choose if they were perfectly informed.[1] Furthermore, if misinformation is systematically related to personal characteristics (e.g., older students underestimate contemporaneous earnings relative to younger students), then these characteristics might be good instruments for education or occupation. Concerned with selection bias, economists often instrument for education with variables that are assumed to

---

[1] Using the same data set as in this paper, Nicholson and Souleles (2002) find that a student's misinformation about market income is incorporated almost dollar for dollar into a student's own subjective income expectation. This implies that misinformation may lead to incorrect specialty choices.

affect a student's likelihood of receiving a certain level or type of education, but have no affect on a person's actual income conditional on his education (Manski 1993).[2] People with positively biased assessments regarding the mean income of a particular educational degree (for example, lawyers) should be more like to receive the degree, but should not necessarily earn more than others once they have the degree.

There are only a few studies that analyze how much people know about contemporaneous income. Betts (1996) surveyed 1,300 undergraduates at the University of California, San Diego in 1992 and asked them to estimate starting and average salaries for people with various educational degrees, majors, and experience levels (for example, MBA graduates). Students underestimated the actual mean income by six percent, on average, and the median absolute value of the estimation error was 20 percent of actual income. College seniors were 31 percent more accurate than freshmen.

Dominitz and Manski (1996) elicited beliefs about contemporaneous earnings from 110 high school and college students. Males had accurate assessments of the earnings of male high school and college graduates, whereas the median female overestimated the earnings of female high school and college graduates by 22 percent and 42 percent, respectively. Both Betts (1996) and Dominitz and Manski (1996) found considerable heterogeneity in students' assessments of earnings.

The above studies use cross-sectional data sets. I use a panel data set with information on 25 cohorts of medical students (3,807 students) between 1974 and 1998, which allows me to examine issues previous authors were unable to. Each student estimated, in both his first and fourth year of school, the contemporaneous average physician income in six different specialties

[2] For example, Willis and Rosen (1979) assume that family characteristics such as the parents' education affect the

and indicated his preferred specialty. Betts' conclusion that upper classmen have a relatively accurate assessment of market earnings may be due to both learning and a cohort effect, whereas my learning results control for cohort effects. The health care market changed considerably during this 25-year time period, which allows me to examine how quickly students incorporate market shocks into their information sets. Observing specialty preferences also allows me to examine whether students who should receive the greatest benefits from searching for income information – students who are undecided about a specialty – provide relatively accurate estimates.

I find that medical students are systematically misinformed about contemporaneous physician income but learn a considerable amount during school. The students' income estimation errors vary systematically over time as well as cross-sectionally by specialty, type of student, and students' specialty preferences. Medical students overestimated physician income in the 1970s but now underestimate income by 25 percent. Since the bias in assessments was not constant, any parameters identified by changes over time in income expectations could be biased. Women, older students, and students with relatively high MCAT scores underestimate physician income relative to their peers.

II. Data and Empirical Methodology

Jefferson Medical College, a large medical school in Philadelphia, matriculates about 200 students per year. In 1970 Jefferson Medical College began surveying its medical students at the conclusion of their first and fourth years. Students predict the specialty in which they will practice and the peak income they expect to receive during their career. Students also estimate

---

probability a child attends college but does not affect the child's income, conditional on observables.

the average annual income of physicians currently practicing in six different specialties: family practice, internal medicine, surgery, pediatrics, obstetrics/gynecology, and psychiatry.[3] The data set also contains information on each student's gender, age, race, and score on the Medical College Admission Test (MCAT), which is taken prior to medical school. I use the MCAT score to measure a student's cognitive ability, which may affect his cost and/or benefit of acquiring income information.

Most Jefferson students provided 12 unique estimates of physicians' medical practice income: six specialties in the first year of school and the same six specialties in the fourth year. I restrict the sample to the 25 cohorts who matriculated between 1970 and 1994 (and graduated between 1974 and 1998), because data on physicians' mean income by specialty are available for these years.[4] The response rate among first- and fourth-year students was quite high (90.5 percent and 91.1 percent, respectively). Nevertheless, I restrict the sample to the 3,807 students who provided income estimates in both years to avoid issues of non-response bias.

Sample means and standard deviations are presented in Table 1. Twenty-five percent of the students are female and 88 percent are white. Fifty-nine percent stated a preference in their first year for one of the six specialties for which they were asked to provide an estimate of contemporaneous physician income, 8 percent preferred a specialty other than those six (for example, radiology), and 33 percent were undecided. In their fourth year of school, all students had a preferred specialty and 74 percent preferred one of the specified six specialties.

To examine how accurately students estimate physician income; one needs to define the benchmark, "true" average income in each specialty and year. I use data from the American

[3] The income question was worded as follows: "Please estimate the average annual income (gross personal income: amount after professional expenses but before income taxes) for each of the following (six specialties)."

[4] I include the responses of the students who graduated between 1970 and 1973, and therefore were only surveyed in their fourth-year of school, in Figure 1 and Figure 2, but not in any of the regressions.

Medical Association's (AMA) Socioeconomic Monitoring System, which is an annual, national survey of approximately 4,000 physicians. The survey is designed to be representative of the population of non-federal physicians who spend the majority of their time in patient care activities. The sample is drawn randomly from the AMA's Physician Masterfile, which contains information on virtually every physician in the United States.[5] Each year the AMA publishes physicians' mean income from medical practice in nine specialties, including the six specialties the Jefferson students are asked to estimate.

I assume the "true" average income is the mean income as reported by the AMA. Since the students were asked to estimate average income rather than income conditional on their own characteristics, I use the AMA's unconditional average income as a benchmark. The AMA publishes the results one year after surveying physicians, so I lag the AMA data by one year when calculating the students' accuracy.

The Jefferson students were asked to estimate the average income in a specialty without specifying whether it should be the mean or median. The students' estimates are slightly more accurate when compared against the national median rather than the mean. The mean of the absolute value of the students' income estimation error using the AMA median income as the benchmark is 30.4 percent of actual income, versus 31.9 percent when the AMA mean is used as a benchmark. Since there appears to be little difference in aggregate, I use mean income as the benchmark because the AMA did not report the median income for 1970-1973.[6]

In year t student i is asked to estimate the average income of physicians practicing in specialty j ($Y^e_{i,j,t}$). The superscript "e" refers to a student's estimate. I calculate two estimation

[5] The AMA survey is conducted by telephone and has a response rate of about 60 percent.

[6] The qualitative results of the paper do not change if I use the median rather than the mean income as the benchmark.

errors for each student in each specialty. The signed estimation error is the difference between a student's estimate and the national mean income in that specialty in the prior year ($Y_{j,t-1}$), divided by the national mean income:[7]

(1) $\varepsilon_{ij,t} = (Y^e_{ij,t} - Y_{j,t-1})/ Y_{j,t-1}$

$\varepsilon_{ij,t}$ is positive when a student overestimates and negative when he underestimates market income. If students do not make systematic errors when interpreting information the mean signed income estimation error will be zero; students will be equally likely to under or overestimate contemporaneous income. Since physician income has changed substantially between 1970 and 1998, I divide $\varepsilon_{ij,t}$ by $Y_{j,t-1}$ to express the signed estimation error as a percentage of national mean income, which facilitates comparisons of estimation errors over time.

The signed estimation error indicates whether medical students' have an upward or downward bias in their assessment of physician income, but does not necessarily measure the accuracy of students' information. If one-half of the students overestimate income by 50 percent and one-half underestimate it by 50 percent, the mean signed error will be zero. It would be incorrect; however, to conclude that students have perfect information. Therefore, I take the absolute value of $\varepsilon_{ij,t}$ from equation (1) to measure the accuracy of a student's estimate. I refer to this second measure as the absolute value of the estimation error, measured in the empirical analysis as a percentage of the mean income in that specialty ($Y_{j,t-1}$).

---

[7] Betts (1996) also refers to this measure as the *signed* estimation error.

III. Results

Bias of Students Income Information

I pool responses across all six specialties and plot in Figure 1 the mean signed income estimation error between 1970 and 1998, separately for first-year and fourth-year students. Students substantially overestimated physician income throughout the 1970s. This changed abruptly in the early 1980s, as both sets of students underestimated physician income. The magnitude of the underestimate has increased in absolute terms since the early 1980s. The mean estimation error for first- and fourth-year students in the 1990s was -21.8 percent of actual, national income. The persistent negative bias in assessments could be due, in part, to an overreaction by medical students to rising malpractice premiums and the growing influence of managed care insurance (for example, HMOs) on physicians' incomes.

Since medical students first overestimated and then underestimated contemporaneous physician income, their expected income probably increased less over time than physicians' actual income. An economist who uses longitudinal data and assumes prospective physicians expect to earn what practicing physicians currently earn, may derive a biased coefficient on the income elasticity of supply to the occupation.

To explore why the bias of students' assessments of physician income changed over time, I calculate the percentage change in national physician income, aggregated across the six specialties, for each year between 1971 and 1998.[8] I likewise calculate the percentage change in fourth-year students' mean estimate of physician income. Apparently it takes students several years to understand how physician income is changing and to incorporate this information into their own estimates. The correlation between the change in actual physician income between year t-1 and t, and the change in the students' estimate of income between the same years is close to zero (-0.13). However, the correlation between the change in actual physician income between t-3 and t-2 and the change in students' estimates between t-1 and t is 0.34. If physician income grew relatively rapidly two years ago, medical students tend to increase their estimates of contemporaneous physician income.

I now examine whether students' income assessments are biased overall and by specialty. The third column of Table 2 reports the mean signed income estimation error for first-year students (Panel A) and fourth-year students (Panel B). If students do not make systematic errors when interpreting information, they would be equally likely to under or overestimate income. In fact, first-year students significantly underestimated the income of family practitioners by 5.9 percent, internists by 5.6 percent, and obstetricians by 14.3 percent, and overestimated the income of pediatricians by 7.6 percent and psychiatrists by 23.4 percent, on average.[9] Fourth-year students underestimated income in psychiatry, family practice, pediatrics, and internal medicine by between 8.5 percent and 17.9 percent, on average, and underestimated physician income by 10.4 percent across all six specialties.

Contrary to the implicit assumption in most school and occupational choice models, medical students do not have homogeneous information sets. I present data on the distribution of signed income estimation errors in the first three columns of Table 2. The median signed error (column two of Table 2) is less than the mean signed error in all six specialties, which indicates that the distribution is skewed to the right. This is particularly true among first-year students (Panel A). More than 10 percent believe that pediatricians, surgeons, and psychiatrists earn 60

---

[8] Each of the six specialties is given an equal weight in deriving the overall percentage change in national physician income.

[9] First-year students' estimation error in surgery was not significantly different from zero, nor was fourth-year students' estimation error in surgery and ob/gyn.

percent more than they actually do (column 4), whereas students who substantially underestimate physician income tend to be more accurate.

The signed income estimation errors among fourth-year students (first four columns of Panel B in Table 2) are also quite heterogeneous, but the distribution is tighter relative to first-year students. This indicates that students learn about physician income during school. Fourth-year students are much less likely than first-year students to substantially overestimate physician income, as indicated by smaller positive errors at the 90^{th} percentile. Nevertheless, the substantial heterogeneity among fourth-year students implies that students' income expectations are likely to vary considerably in a cross-section due to differences in the type of information people use, or differences in how they interpret the same information.

To examine whether there are systematic differences in income estimation errors between certain types of students, I pool the signed estimation errors across all years and specialties. I regress the estimation errors, expressed as percentage point deviations from actual income, on personal characteristics ($\mathbf{X}_i$), a separate set of indicator variables for each specialty (S) a student estimates in his first and fourth year, and indicator variables for the year the survey was administered (T):

(2) $$\varepsilon_{ij,t} = \gamma_0 + \boldsymbol{\gamma_1 X_i} + \boldsymbol{\gamma_2 S} + \boldsymbol{\gamma_3 T} + u_1$$

Recall that each student provided 12 unique income estimates: six specialties in the first year and six in the fourth year.[10]

A positive coefficient on an **X** variable indicates that students with a particular characteristic overestimate physician income relative to other students. The **X** vector includes an indicator variable for students estimating income in their preferred specialty and an indicator for students who are undecided about a specialty in their first year. The γ_2 coefficients measure the mean income estimation error in a particular specialty relative to the omitted specialty. I use family practice as the reference specialty because students assess income in this specialty more accurately than in the others, as I show below.

The coefficient estimates reported in the first two columns of Table 3 provide evidence that medical students' perceptions of physician income differ systematically by type of person. Women underestimate physician income by 4.7 percentage points relative, on average, relative to men. Older students, fourth-year students, and those with relatively high MCAT scores also underestimate physician income relative to their peers. The magnitude of the MCAT coefficient is small. A student with an MCAT score at the 90th percentile is predicted to provide an income estimate 2.2 percent lower than the actual income relative to a student with the median MCAT. The negative coefficient on the preferred specialty dummies indicate that first- and fourth-year students underestimate physician income in their preferred specialty by 6.7 percent and 1.0 percent, respectively, relative to the other five specialties. If students choose a specialty in part because they overestimate income in that specialty, the coefficient on this variable would instead be positive.

Accuracy of Students' Income Information

To examine accuracy I focus on the absolute value of the signed income estimation error, $|\varepsilon_{ij}|$, measured as a percentage of actual income. The mean of the absolute value of the income estimation error between 1970 and 1998 is plotted in Figure 2, separately for first- and fourth-

---

[10] The standard errors are corrected to allow for correlations in the error terms between the 12 observations for each

year students. I group the data by cohort, so the responses for first-year students line up with their responses three years later. The accuracy of fourth-year students' has remained fairly stable between 1970 and 1998, ranging from a 25 percent to a 35 percent error. First-year students were quite inaccurate in the late 1970s and early 1980s when physician income was particularly volatile, but their accuracy has improved since then. Students do learn about physician earnings before they formally choose a specialty: they provided more accurate estimates of in their fourth-year than their first-year, on average, in all but one year.

I report data on the distribution of the absolute value of income estimation errors in the right-hand side of Table 2. The median of the absolute value of the estimation error ranges from 24.6 percent to 30.3 percent across the 6 specialties among first-year students, and from 19.3 percent to 27.7 percent among fourth-year students. The median absolute error among undergraduates at UC-San Diego in 1992 was 19.6 percent of actual income (Betts 1996). If I restrict my sample to medical students who surveyed between 1990 and 1994, the median absolute error is 29.6 percent of actual income, so medical students are considerably less accurate. One explanation for this surprising result is that it may be more difficult to estimate average earnings (across all experience levels) than starting salaries. Alternatively, the equalizing differences between specialties/occupations might be larger for medical students than undergraduates. Medical students who do not expect to switch specialties based on changes in their perception of physician income would place little value on searching for income information.

The final three columns of Table 2 provide strong evidence that there is considerable heterogeneity regarding how much medical students know about contemporaneous physician

---

person.

income. Ten percent of first-year students provided an income estimate within 5.2 percentage points of the actual income, but 10 percent believe physician income is 62.6 percent higher or lower than it actually is. As with the signed error, the distribution of the absolute value of the signed error is tighter among fourth-year students than among first-year students. For example, the 90th percentile of the absolute value of the estimation error among fourth-year students is 48 percent of the actual income, versus 63 percent for first-year students.

Another way to measure accuracy is by the proportion of students who correctly rank the specialties according to their actual income. In many years the incomes of the six specialties were closely clustered in two groups: surgery and ob/gyn with relatively high incomes, and family practice, pediatrics, psychiatry, and internal medicine with substantially lower incomes.[11] Therefore, I calculate the proportion of students who report higher income estimates for surgery and ob/gyn than the other four specialties. Only 23 percent of first-year students reported higher income estimates for both surgeons and obstetricians than the other four specialties. Among fourth-year students, on the other hand, 75 percent correctly ordered the two specialty groupings. Betts (1996) found that 26 percent of college students were able to correctly rank salaries of four occupations, although the salary differences between the four occupations he examined were smaller than those that I examine.

To examine the determinants of accuracy, I regress the absolute value of a student's income estimation error, measured as percentage points of actual income, on the same regressors as in equation (2):

$$(3) \qquad |\varepsilon_{ij,t}| = \theta_0 + \boldsymbol{\theta_1 X_i} + \boldsymbol{\theta_2 S} + \boldsymbol{\theta_3 T} + u_2$$

---

[11] In 1998, for example, mean incomes were as follows: surgery ($258,000), ob/gyn ($206,000), internal medicine ($175,000), family practice ($137,000), pediatrics ($134,000), and psychiatry ($133,000). I do not attempt finer specialty orderings because the mean income of ob/gyn was higher than surgery in one year, and the mean income of internal medicine was similar to family practice, pediatrics, and psychiatry in the 1970s.

Negative coefficients are associated with more accurate information. The coefficient on the indicator variable for fourth-year students measures the amount of learning about physicians' incomes that occurs during medical school.

Coefficient estimates are reported in last two columns of Table 3.[12] Men and women do not differ in their accuracy. Students with relatively high MCAT scores provide more accurate income estimates, although the magnitude of this coefficient small. A student who receives an MCAT score at the 90th percentile is predicted to provide an income estimate 1.2 percentage points closer to the actual income (4.8 percent more accurate) relative to a student with the median MCAT score. This is consistent with high-ability students having a lower cost of acquiring information and/or a greater benefit from searching for information.

Students' estimates of physician income in their preferred specialty are 4.1 percentage points and 1.3 percentage points more accurate in their first and fourth year, respectively, relative to estimates for the other five specialties. This is intuitive; students should have more accurate information specialties they are actively considering. About one-third of first-year students had not decided on a specialty. The income estimates for these students were just as accurate as those with a stated preference, even though undecided students may spend more time searching for income information.

All the specialty coefficients are positive and statistically significant for both first- and fourth-year students. Students estimate the income of family practitioners (the omitted specialty) more accurately than every other specialty. Relative to family practice, fourth-year students provide income estimates in surgery, ob/gyn, and psychiatry that are over six percentage points farther from the actual income. Psychiatry and ob/gyn are the least popular specialties among the Jefferson medical students, being preferred by three and six percent of the fourth-year students, respectively. If many medical students were not actively considering these specialties, one would expect these estimates to be relatively inaccurate. Although surgery is a popular specialty, the actual income of surgeons increased sharply in real terms in the 1980s, and has been volatile in the 1990s. I showed earlier that students react to market income shocks with a lag, which implies students would estimate surgeons' incomes relatively inaccurately.

According to Table 3, students are 7.6 percentage points more accurate, on average, when estimating family practitioners' incomes (the omitted specialty) in their fourth year of school relative to their first year. Accuracy improves by an average of 9.4 percentage points, or 35 percent, across all six specialties. This result is consistent with Betts' (1996) finding that college seniors are 31 percent more accurate than freshmen when estimating starting salaries.[13] Although there are systematic differences in accuracy across types of students and specialties, these characteristics explain only three percent of the variation in students' assessments.

IV. Conclusions

I use a unique panel data set that contains medical students' estimates of contemporaneous physician income between 1974 and 1998 to examine how knowledgeable medical students are about market income. Although this group of intelligent students is systematically uninformed about incomes in their profession, they do learn a considerable amount during school. Income estimation errors vary systematically over time, and cross-sectionally by specialty, type of student, and students' specialty preferences.

---

[12] As before, I adjust the standard errors to allow for correlations in the error terms between the 12 observations for each respondent.

[13] If I restrict the sample to students who graduated between 1990 and 1994, fourth-year students were 38 percent more accurate than first-year students.

Medical students overestimated physician income in the 1970s but now underestimate income by about 25 percent, on average. Nicholson and Souleles (2002) show that students incorporate misperceptions of market income into their own income expectations almost dollar for dollar. This implies that parameters such as the income elasticity of supply to the medical profession could be biased if the parameter is identified by changes over time in income expectations, and an economist assumes prospective physicians base income expectations on contemporaneous earnings. Medical students either access different information or interpret it differently. For example, 10 percent of fourth-year medical students overestimated physician income by at least 43 percent, while 10 percent underestimated physician income by at least 44 percent. Thus, students' income expectations are likely to vary considerably in a cross-section due to differences in information.

Women, older students, and students with relatively high MCAT scores underestimate physicians' incomes relative to their peers. In this case, a student's age may be a valid instrument for the specialty he chooses. Age probably affects a student's expected income in a specialty (and thus the likelihood of entering) through its affect on information about market income, but should not affect income conditional on choosing that specialty. Thus having subjective data on people's income information can help economists address selection bias (Manski 1993).

The median absolute value of the estimation errors was 26 percent of actual income. One reason medical students provide inaccurate estimates is that they extrapolate recent income growth rates, and tend therefore to overshoot or undershoot actual income when there have been relatively large changes in the near past. Students who were undecided about a specialty in their first year, and therefore presumably place a relatively high value on information, provide relatively accurate income estimates. Students also provide more accurate estimates in the specialties they have the highest probability of choosing -- their own preferred specialty and the specialties that are relatively popular among the entire sample. Students learn a considerable amount about income during school. Relative to their first year of school, students are 35 percent more accurate when estimating physician income in their fourth year, and are three times more likely to correctly order specialties according to their actual market earnings.

References

Betts, Julian. 1996. "What Do Students Know About Wages?" Journal of Human Resources 31(1):27-56.

Das, Marcel, and Arthur van Soest. 1997. "Expected and Realized Income Changes: Evidence From the Dutch Socio-Economic Panel" Journal of Economic Behavior & Organization 32(1):137-154.

Dominitz, Jeff, and Charles Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling" Journal of Human Resources 31(1):1-26.

Dominitz Jeff and Charles Manski. 1997. "Using Expectations Data to Study Subjective Income Expectations" Journal of the American Statistical Association, 92(439):855-867.

Dominitz, Jeff. 1998. "Earnings Expectations, Revisions, and Realizations" The Review of Economics and Statistics 80(3):374-388.

Manski, Charles. 1993. "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" In Studies of Supply and Demand in Higher Education, eds. Michael Rothschild and Lawrence J. White, 291-312. Chicago: The University of Chicago Press.

Nicholson, Sean. 2002. "Physician Specialty Choice Under Uncertainty" Journal of Labor Economics 20(4):816-847.

Nicholson, Sean, and Nicholas Souleles. 2003. "Physician Income Prediction Errors: Sources and Implications for Behavior" National Bureau of Economic Research working paper 8907.

Nicholson, Sean, and Nicholas Souleles. 2002. "Physician Income Expectations and Specialty Choice" National Bureau of Economic Research working paper 8536.

Willis, Robert and Sherwin Rosen. 1979. "Education and Self-Selection" Journal of Political Economy 87(5):S7-S36.

Table 1
Sample Means and Standard Deviations
n = 3,807

	Mean	Standard Deviation
Female	0.252	0.434
Age in first year of medical school	22.8	2.87
White	0.878	0.327
MCAT score	18.8	2.77
Preferred specialty:	First-year Students	
- family practice	0.167	0.373
- internal medicine	0.122	0.328
- surgery	0.175	0.380
- pediatrics	0.085	0.278
- obstetrics/gynecology	0.023	0.149
- psychiatry	0.016	0.127
- other specialty	0.082	0.274
- undecided	0.330	0.470
	Fourth-year Students	
- family practice	0.172	0.383
- internal medicine	0.235	0.430
- surgery	0.182	0.391
- pediatrics	0.067	0.255
- obstetrics/gynecology	0.057	0.235
- psychiatry	0.030	0.174
- other specialty	0.258	0.438
- undecided	0.000	0.000

Table 2
Income Estimation Errors of First- and Fourth-Year Medical Students

Panel A: First-Year Medical Students (n = 3,807)

	Signed Income Estimation Error				Absolute Value of Income Estimation Error		
Specialty	10th Percentile	Median	Mean	90th Percentile	10th Percentile	Median	90th Percentile
Family practice	-42.3	-12.8	-5.9	37.4	4.7	24.6	51.3
Internal medicine	-47.2	-16.0	-5.6	44.0	4.6	26.8	58.2
Pediatrics	-38.9	-2.9	7.6	61.0	4.9	24.6	64.0
Surgery	-47.0	-10.7	2.8	60.7	4.9	28.6	65.6
Obstetrics	-53.7	-23.4	-14.3	29.7	7.1	30.3	58.8
Psychiatry	-35.6	7.5	23.4	94.7	6.0	28.9	94.7
First-year total	-45.4	-10.5	1.2	56.6	5.2	27.0	62.6

Panel B: Fourth-Year Medical Students (n = 3,807)

	Signed Estimation Error				Absolute Value of Estimation Error		
Specialty	10th Percentile	Median	Mean	90th Percentile	10th Percentile	Median	90th Percentile
Family practice	-36.1	-16.0	-10.4	19.3	4.7	19.3	40.1
Internal medicine	-45.4	-23.2	-17.9	16.3	4.9	26.8	47.1
Pediatrics	-41.4	-21.3	-16.2	12.9	5.9	26.0	43.6
Surgery	-44.3	-16.8	-3.7	45.3	5.3	27.7	55.5
Obstetrics	-43.4	-14.2	-5.8	42.9	5.8	25.1	52.2
Psychiatry	-41.2	-19.4	-8.5	36.2	6.0	26.1	51.2
Fourth-year total	-42.5	-18.2	-10.4	29.0	5.4	24.9	48.4
Overall	-43.5	-15.4	-4.7	43.3	5.3	26.0	55.7

Note: income estimation errors are represented as percentage points of actual income.

Table 3

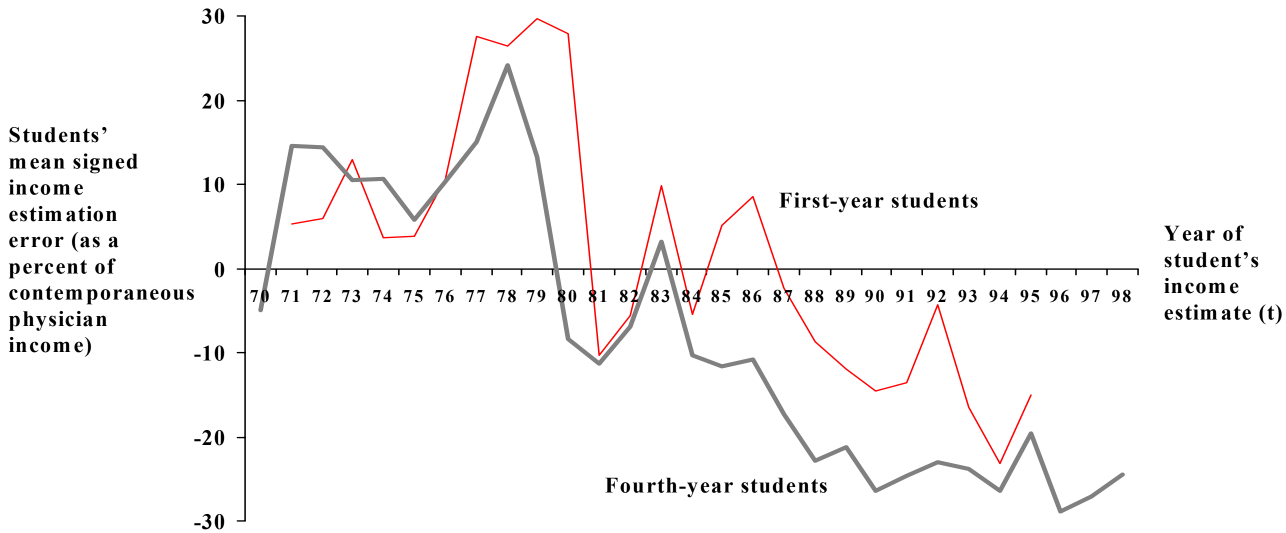
Determinants of Students' Income Estimation Errors

Variable	Signed Income Estimation Errors: Coefficient	Signed Income Estimation Errors: Standard Error	Absolute Value of Signed Estimation Errors: Coefficient	Absolute Value of Signed Estimation Errors: Standard Error
Female	-4.73**	0.949	0.480	0.637
White	1.25	1.40	-2.15**	1.00
Age as 1st-year student	-0.429**	0.138	-0.106	0.090
MCAT score	-0.742**	0.192	-0.413**	0.147
Fourth-year student	-2.96**	1.19	-7.60**	0.971
Student's preferred specialty: 1st year	-6.70**	0.809	-4.09**	0.709
Student's preferred specialty: 4th year	-1.04*	0.579	-1.26**	0.493
Undecided in 1st year: 1st-year responses	-0.033	1.49	-1.68	1.08
Undecided in 1st year: 4th year responses	-1.63*	0.851	-0.813	0.630
Specialty indicators for **first-year students**: (family practice is omitted)				
- internal medicine	0.043	0.511	3.80**	0.438
- surgery	8.79**	0.797	8.95**	0.701
- pediatrics	13.0**	0.530	4.12**	0.490
- ob/gyn	-9.27**	0.575	5.14**	0.507
- psychiatry	28.1**	0.936	15.0**	0.870
Specialty indicators for **fourth-year students**: (family practice is omitted)				
- internal medicine	-7.36**	0.470	4.89**	0.445
- surgery	7.05**	0.801	11.2**	0.686
- pediatrics	-5.83**	0.601	3.54**	0.550
- ob/gyn	4.67**	0.636	6.76**	0.539
- psychiatry	2.02**	0.754	6.94**	0.643
Constant	2.45	5.44	41.4**	3.96
Observations	44,503		44,503	
R^2	0.15		0.03	

Notes: The dependent variable for the first regression is the difference between a student's estimate of the average contemporaneous physician income in a specialty and the actual mean income in that specialty, as reported in the AMA survey, divided by the actual mean income, multiplied by 100. Coefficients are expressed as percentage points. The dependent variable in the second regression is the absolute value of the income estimation error: $|(Y^e_{ijt} - Y_{j,t-1})/Y_{j,t-1}|$, multiplied by 100. Indicator variables are included for the year in which a student was surveyed. ** = significantly different from zero at the 5 percent level. * = significantly different from zero at the 10 percent level. Standard errors are adjusted to allow the error terms to be correlated between specialties for an individual.

Figure 1

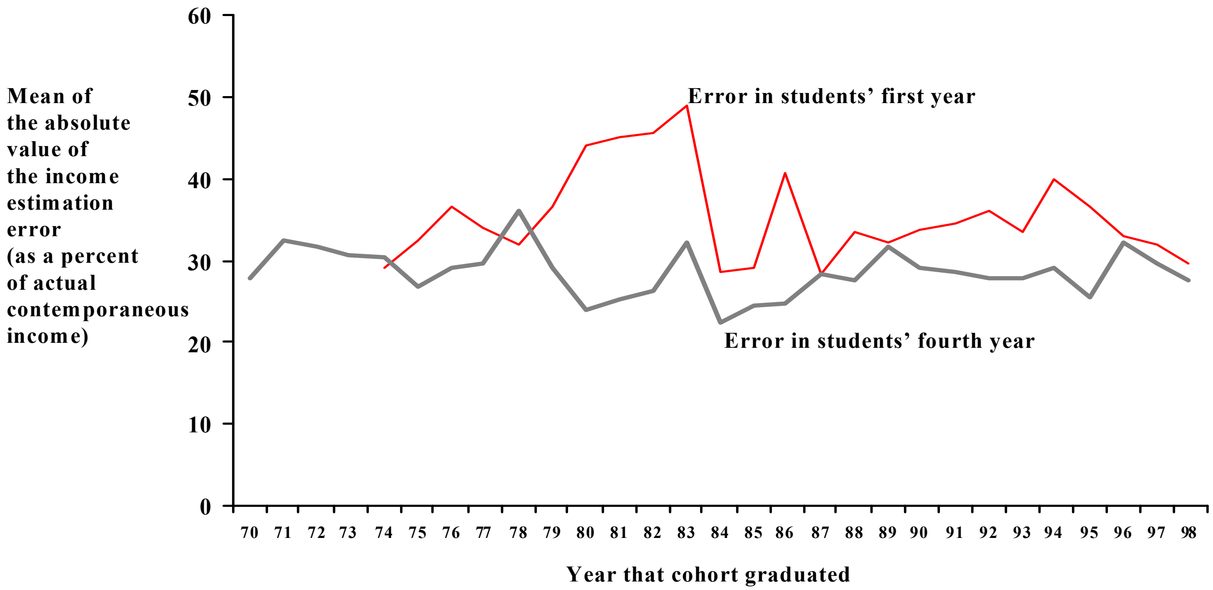
Mean Income Estimation Error For First- and Fourth-Year Students, 1970-1998



Notes: the figure displays the coefficients from a regression of the signed income estimation error (a student's estimate of physician income in a specialty minus the actual contemporaneous mean income, as reported by the American Medical Association survey, multiplied by 100) on indicator variables for the year of the survey. All of the coefficients in the fourth-year student regression are significantly different from zero at the 10-percent level except 1983; all of the coefficients in the first-year student regression are significantly different from zero at the 10-percent level except 1974, 1975, 1985, 1987, and 1992. Standard errors are adjusted to allow the error terms to be correlated between medical students who form their income estimate in the same year.

Figure 2

Absolute Value of Income Estimation Error by Student Cohort, 1974-1998



Notes: the figure displays the coefficients from a regression of the absolute value of the signed income estimation error (a student's estimate of physician income in a specialty minus the actual contemporaneous mean income, as reported by the American Medical Association survey, multiplied by 100) on indicator variables for the year the student graduated. All of the coefficients on the year indicators are significantly different from zero at the one-percent level. Standard errors are adjusted to allow the error terms to be correlated between medical students who form their income estimate in the same year.