Physician Income Expectations and Specialty Choice*

Sean Nicholson
The Wharton School
University of Pennsylvania
and NBER
nicholss@wharton.upenn.edu

Nicholas S. Souleles
The Wharton School
University of Pennsylvania
and NBER
souleles@wharton.upenn.edu

April 2002

Abstract:

In spite of the important role of income expectations in economics, little is known about how people actually form these expectations. We use a unique data set that contains explicit income expectations of almost 30 cohorts of medical students to examine how people form income expectations. We find that medical students significantly condition their income expectations on personal characteristics such as gender and ability. For instance, female students expect to earn substantially less than male students, even controlling for differences in the hours they expect to work. Income expectations also increase with the contemporaneous income of physicians currently practicing in the specialty a students plan to enter, which is consistent with learning models. Nonetheless, expectations are also significantly forward-looking: after students report relatively high income expectations for a given specialty, physicians practicing in that specialty subsequently tend to experience high income growth relative to physicians in other specialties. We also find that explicit income expectations are useful for predicting behavior. A specialty-choice model that uses the students' explicit income expectations has a better fit than a model that assumes income expectations are formed statically, and even a model that assumes that students have perfect foresight regarding their future income.

JEL codes: D84, I11, J24, J44, D83.
Key words: income expectations, learning, occupational choice.

---

* The University of Pennsylvania Research Foundation, Jefferson Medical College, and the Rodney L. White Center for Financial Research provided funding for this study. Joseph S. Gonnella, Jon Veloski, Mary Robeson and other members of the Center for Research in Medical Education and Health Care at Jefferson Medical College provided access to and assistance with the Jefferson Longitudinal Study. Patricia Danzon, Mark Pauly, Rob Lemke, Dan Polsky, and members of the Applied Economics workshop at The Wharton School and other seminars have provided helpful suggestions. Valuable contributions were provided by our research assistants: Josh Beer, Matt Cinque, Hanli Mangun, and Peter Margetis.

I. Introduction

Income expectations play a central role in many economic models, particularly models of schooling decisions and occupational choice. However, economists cannot usually observe peoples' income expectations and therefore must make assumptions about how people form these expectations. There is little agreement about whether income expectations are formed fully rationally or not, and what information people use in forming their expectations (Manski, 1993). In his "cobweb" model of occupational labor supply, Freeman (1971) assumes that income expectations are static (or adaptive): college students expect to earn the contemporaneous mean income of people who are already in the contemplated profession. In his study of physician specialty choice Sloan (1970) also assumes expectations are static, deriving medical students' expected income associated with each specialty based on contemporaneous physician income. By contrast, Willis and Rosen (1979) assume that income expectations are rational: high school students understand the process that will generate their future income. Students estimate the additional earnings associated with attending college, and select the utility-maximizing schooling level. Hay (1991), who also assumes that medical students have rational income expectations, extends the Willis and Rosen (1979) model to allow medical students to choose between different medical specialties. Willis and Rosen (1979) and Hay (1991) test their models by instrumenting for people's ex-post income with information known at the time of the schooling decision, such as gender and ability.[1]

In all four of the studies cited above, the authors find that students are more likely to choose a particular occupation or schooling level if their expected income in that occupation/schooling level is greater than in its alternative. However, as Manski (1993) points out, without any evidence on how

---

[1] Some papers indirectly test the rationality of occupational choice without observing peoples' subjective income expectations. Zarkin (1985) examines whether prospective teachers incorporate forecastable demand conditions into their decision to enter the occupation. He finds that future student enrollment rationally affects the occupational decisions of secondary school teachers, but not of elementary school teachers. Siow (1984) assumes that prospective lawyers form income expectations rationally and expect future cohorts of students to arbitrage away any rents that would otherwise occur from a wage shock. He finds evidence consistent with his model.

people form income expectations, "the most that one can do is infer the decision rule conditional on maintained assumptions on expectations." Manski (1993) shows that misspecifying how income expectations are formed can lead to incorrect parameter estimates, such as the responsiveness of enrollment to the rate of return to schooling.

This paper directly examines how medical students form income expectations. Specifically, to what extent does a medical student's income expectations vary with his individual characteristics such as ability, gender, and race? For instance, do female students expect to earn less than male students, and if so, is this because they intend to work fewer hours, perhaps in order to raise a family? Further, does a student's income expectations primarily reflect the contemporaneous income of physicians who are currently practicing in his specialty, as opposed to the actual income in that specialty in the future? That is, are students' expectations forward-looking, in that they correctly anticipate future trends in specialty income? Medical students are some of the brightest people in the country, so if we find that their income expectations are not forward-looking, it would seem unlikely that other students have forward-looking expectations. We then test whether students' explicit income expectations help predict their behavior, in particular their specialty choices -- decisions for which expectations of future returns should be critical. We separately compare the fit of a specialty choice model that uses the students' explicit income expectations with a model that assumes students form income expectations statically, and a model that assumes students have perfect foresight regarding their future income.

We use a unique data set that contains direct measures of medical students' subjective income expectations and a rich set of demographic and ability measures. The Jefferson Longitudinal Study contains information on all medical students who graduated from Jefferson Medical College, a large medical school in Philadelphia, since 1970. The students in every cohort were surveyed in their fourth year of medical school and asked to predict the following: the specialty in which they will practice, their income 5, 10, and 20 years after completing residency training (i.e., their income with 5, 10, and 20 years of post-residency experience), their peak career income, and other characteristics of their medical practice.

There have been a few studies of subjective income expectations data available in other surveys. Most of these studies focus on testing the rationality of the expectations (e.g., Dominitz, 1998; Das and van Soest, 1999; Nicholson and Souleles, 2002). There has been little analysis, however, of the determinants of income expectations and of whether subjective expectations data help predict people's behavior.[2] Souleles (2002) analyzes the time series and cross-sectional variation in expectations in the Michigan Surveys of Consumer Attitudes and Behavior, regarding variables like household income and financial position, inflation and aggregate economic activity. He shows that such expectations data help predict households' consumption and precautionary saving, as well as their portfolio allocations (Souleles, 2000).

Most surveys of expectations like the Michigan Surveys ask about expectations over a one-year horizon, whereas life-cycle decisions like occupational choice depend on longer-horizon expectations. Also, the answers to the expectation questions are often constrained to be discrete (e.g., Will your income increase, decrease, or stay the same over the next year?). The Jefferson Longitudinal Study is better suited to examine how students form income expectations and whether explicit income expectations help predict life-cycle decisions like specialty choice. The Jefferson survey solicits continuous-valued (not discrete) income expectations, over long horizons (e.g., 5, 10, and 20 years ahead). With long-horizon income expectations we can analyze people's occupational choice, a decision with substantial welfare implications. The sample spans a long period of time, 27 years, during which there were tremendous changes in the health care market. This provides a large amount of variation in factors that should affect income expectations. Furthermore, the database also includes information on student performance during medical school, including scores on the national board exams.

To summarize the results, we find that medical students significantly condition their expected

[2] Dominitz and Manski (1996) examine how high school students' expected returns to schooling vary by gender, and Dominitz (1998) and Das and van Soest (1999) examine how earnings expectations vary by personal characteristics

income on their individual characteristics, such as their gender and ability. For instance, female students expect to earn substantially less than male students, controlling for differences in the specialties they choose and even in the hours they expect to work. Students also condition their expectations on the *contemporaneous* income of physicians currently practicing in the specialty they intend to enter. Indeed, misinformation about contemporaneous physician income affects students' expectations for their own income almost dollar for dollar. These results are consistent with a learning model in which students learn by observing the practicing physicians with whom they interact. Hence a considerable amount of the variation in income expectations appears to be due to heterogeneity in information in addition to other individual characteristics. Nonetheless, expectations are also significantly forward-looking: students incorporate future trends in specialty income into their own expectations. In particular, after students report relatively high income expectations for a given specialty, that specialty subsequently tends to experience high income growth relative to other specialties. We also find that the students' explicit income expectations are useful for predicting their behavior, specifically their specialty choice. A $10,000 increase in a student's expected income in non-primary care specialties (e.g., surgery) relative to primary care specialties is associated with an increase of 0.057 in the probability of entering a non-primary care specialty. Furthermore, a specialty choice model that uses students' explicit income expectations fits the data better than models that assume students form income expectations statically or that they have perfect foresight regarding their future income (with expectations coinciding with ex-post income).

The paper is organized as follows. Section II describes the two data sets used, the Jefferson Longitudinal Study and the American Medical Association annual survey on physician income. Section III presents the empirical model we use to analyze how medical students form income expectations and how those expectations affect their choice of specialty. Results are reported in Section IV and concluding

---

such as gender and age, current earnings, and employment status.

comments in Section V.

II. Data

In 1970 Jefferson Medical College began surveying its medical students in their fourth year of medical school, asking students to predict various aspects of their future medical practice, such as the specialty in which they will practice and the number of hours per week they will work. Between 1970 and 1979 the students were asked to state the income, after medical expenses and before taxes, that they expect to receive 5, 10, and 20 years after completing residency training, and the peak income they expect to receive during their career. Students who graduated after 1979 were asked to predict their peak income only. Students were told to report their income in real terms rather than try to guess the inflation rate.[3] The Jefferson Longitudinal Study contains complete information on 3,025 students who graduated between 1971 and 1998, most of whom are now practicing physicians. The sample used below begins in 1971 because data for some key variables such as debt and the accuracy of the students' income information were not collected in 1970. Table 1 reports sample means of the key variables used below.

Information on student performance has been added to the Jefferson Longitudinal Study. Medical students must pass three national exams before they can receive a license to practice medicine in the United States. Part 1 of the National Board of Medical Examiners (NBME) test is administered after the second year of medical school and covers the classroom material taught during the first two years (e.g., anatomy, physiology, and pharmacology). We focus on the Part 1 score as a measure of student ability and performance because this exam took place before students stated their income expectations.[4] Jefferson students who graduated between 1996 and 1998 received an average score of 209.1 on this exam, only slightly below the national average of 210.8 for all students who graduated from U.S. medical

[3] For example, the 1970 survey question regarding expected income in 5 years was worded as follows: "In answering the following questions relating to income, please assume that dollars maintain their 1970 value. What do you think your own gross personal income (after professional expenses, but before income taxes) will be 5 years after completing residency training?"

[4] The second part of the NBME exam is administered in the fourth year of medical school and the third part is administered in the first year of the students' residency program.

schools during those three years.[5] Thus, Jefferson students appear to be generally representative of U.S. medical students.

Many economists who study occupational choice and schooling decisions assume that students base their income expectations on the contemporaneous mean income of earlier cohorts of students who are now employed in the occupation of interest. We use data from the annual American Medical Association (AMA) surveys, the most authoritative and widely known source of information about physician income, to characterize the information that was contemporaneously available to the Jefferson students. Each year the AMA surveys a nationally representative, random sample of about 4,000 practicing physicians. The AMA reports the mean and median medical practice income (income after professional expenses but before taxes) by specialty for physicians in the following age groups: under 36 years old, between 36 and 45, between 46 and 55, between 56 and 65, and over 65 years old. We assume that the mean income for an age group corresponds to physicians at the midpoint of the age range (e.g., the mean income of a 40-year-old physician is assumed to be the mean income reported for the 36-45 year old group). We linearly interpolate between age-specific observations to estimate the national, cross-sectional age-income profile for each specialty in each year.[6] Most students complete medical school between the ages of 26 and 28, so the age-income profile closely approximates the experience-income profile.

The income expectations of each Jefferson student are then matched with the corresponding national mean income of physicians who are currently practicing in the specialty in which the student intends to enter. For example, consider a student who graduates from medical school in 1977, plans to become a surgeon, and expects his income with five years of experience to be $185,000 (measured in

---

[5] The standard deviation of the board score among students who graduated from U.S. medical schools between 1996 and 1998 is 18.

[6] Specifically, we assume that physicians complete medical school at age 26, spend the number of years in residency training required by their specialty, and begin practicing medicine immediately after completing residency training.

1996 dollars).[7] According to the AMA survey, in 1976 the mean income of surgeons with five years of experience was $172,000 (1996 dollars). When analyzing the determinants of income expectations, we compare this student's expected income (EY_5 = $185,000, where the subscript refers to years of post-residency experience) with the mean contemporaneous income of surgeons in 1976 who had five years of experience (Y^N_5=$172,000, where the superscript "N" refers to the national AMA data). We lag the national data by one year because the AMA survey is published one year after the survey is conducted. The same procedure is used to assign a corresponding national mean income for students' expected income with 10 and 20 years of experience. The peak contemporaneous income in a specialty (Y^N_{peak}) is the maximum age-specific mean in the national cross-sectional experience-income profile, which for most specialties corresponds to physicians between the ages of 46 and 55.

By 1987 the student described above should have completed his five-year surgery residency program and have been practicing medicine for five years. In 1986 the mean income of surgeons with five years of experience from the AMA survey was $205,000 (also measured in 1996 dollars). The mean income of surgeons with five years of experience increased by $33,000 ($205,000 - $172,000) from the time the student formed his expectation to the time the student had five years of experience. In some of the analysis that follows, we examine whether students correctly anticipated the future change in income in their specialty and incorporated this change into their own expectations.

Betts (1996) and Nicholson (2002) have shown that there is substantial variation among college and medical students in how much they know about earnings. Similarly, Souleles (2002) found variation even in peoples' forecasts of aggregate variables like inflation and economic activity. One advantage of the Jefferson survey is that it includes questions that measure the quality of students' market information. Jefferson students were asked to estimate the average practice income *currently* being earned by physicians in six different specialties ($Y^{N,est}_i$): family practice, internal medicine, surgery, pediatrics,

[7] We convert all expected and realized income in this paper to 1996 dollars using the urban CPI.

obstetrics/gynecology (ob/gyn), and psychiatry. We calculate the accuracy of each student's market information by taking the difference ($Y^{N,est}_i$ - Y^N) between the student's estimate of the contemporaneous income in the specialty in which he plans to enter and the median income of physicians already practicing in that specialty, as reported in the AMA surveys.[8] Jefferson students under-predict the prevailing median income in their preferred specialty by $7,300, on average, as reported in the last row of Table 1.

III. Empirical Method

a. Income Expectations

We begin by examining how students form income expectations. Let $EY_{i,j,t=0}$ represent the income that student i in his fourth-year of medical school (t=0) expects to receive when he has j years of post-residency experience. We allow a student's information set to consist of personal characteristics ($\mathbf{X}_i$); the contemporaneous mean national income of physicians with j years of experience who are currently practicing (in year 0) in the specialty the student intends to enter ($Y^N_{j,t=0}$); indicator variables for the specialty the student expects to enter (S); time dummies for the student's graduation year (T); the ex-post *future* growth of physician income in the student's specialty ($Y^N_{j,t=j}$ - $Y^N_{j,t=0}$); and a measure of how accurately the student estimates the contemporaneous average income of physicians in the specialty he plans to enter ($Y^{N,est}_{i,t=0}$ - $Y^N_{t=0}$):

$$(1) \qquad EY_{i,j,t=0} = \alpha_1 X_i + \alpha_2 Y^N_{j,t=0} + \alpha_3 S + \alpha_4 T + \alpha_5 (Y^N_{j,t=j} - Y^N_{j,t=0}) + \alpha_6 (Y^{N,est}_{i,t=0} - Y^N_{t=0}) + u_1.$$

In specifications that pool across multiple years of experience we also include indicator variables for the number of years of experience (e.g., j=5, 10, and 20 years). Through a series of ordinary least squares regressions of form (1), we test whether and to what extent students condition their expected income on

---

[8] About three-quarters of the Jefferson students expected to enter one of these six specialties.

the above information.

Our measure of contemporaneous physician income ($Y^N_{j,t=0}$) is conditioned only on specialty and experience level. If medical students only observe the specialty and experience level of practicing physicians and expectations are static, α_2 will equal one and the other coefficients in equation (1) will be zero. If the α_2 coefficients are significantly different from one, this would be consistent with either rational expectations or static expectations where students observe characteristics other than just specialty and experience level (e.g., characteristics like ability, race, and gender). When the specialty and year indicators are excluded from the regression, α_2 is identified by variation in physician income between specialties at a point in time and variation over time in all specialty incomes. When the specialty and year indicators are included, α_2 is identified by within-specialty income variation over time. Therefore, a positive coefficient on α_2 in the latter specification indicates that students are knowledgeable about *relative* trends in contemporaneous specialty incomes. The coefficient α_6 measures the extent to which students' misinformation about contemporaneous physicians' income is incorporated into their own income expectations.

The variable ($Y^N_{j,t=j}$ - $Y^N_{j,t=0}$) in equation (1) measures the future change in physician income in a student's specialty. If students have perfect foresight, α_5 would be close to one.[9] More generally, α_5 measures the proportion of future income growth that is forecasted by students. For example, if students anticipate that the demand for surgical services will increase in the future, they might expect their future income in surgery to be higher than that of previous cohorts of surgeons. A positive coefficient on α_5 would represent compelling evidence that students are forward-looking when forming their income expectations. In a companion paper, Nicholson and Souleles (2002), we formally test whether income

[9] The physician's own income realization $Y_{i,t=j}$ is not in his information set at time t=0. By contrast, $Y^N_{j,t=j}$ is the average income in his specialty in year j. Because $Y^N_{j,t=j}$ can be interpreted as the projection of $Y_{i,t=j}$ on time and specialty dummies, it can be interpreted as a rational expectations forecast of $Y_{i,t=j}$, and so is a valid regressor in

expectations are rational (unbiased and efficient).[10]

b. Specialty Choice

Some economists question the value of survey questions like ours, which ask students to state their expected income. It is worth noting, however, that most variables in household data sets are based on self-reported information. Instead of repeating the well-known advantages and disadvantages of survey questions, we formally test whether subjective income expectations help predict people's behavior. As an application we analyze the specialty choice decisions of medical students, using three different assumptions about how students form income expectations. We compare the fit of a model that uses students' subjective income expectations to the fit of models in which medical students are assumed to have static income expectations (they expect to receive the contemporaneous mean income of physicians practicing in their specialty) or perfect foresight (they expect their income with, for example, 10 years of experience, to equal the mean income in their specialty that was actually received by their cohort 10 years later). If students' specialty choices can be predicted more accurately in the former model relative to the latter two models, this represents compelling evidence that subjective income expectations can help explain people's behavior.

Ninety-seven percent of U.S. medical school graduates enter a residency training program after completing medical school. Residency positions are available in 26 different specialties, which range in length from three years for primary care specialties (family practice, internal medicine, and pediatrics) to five years for surgical specialties (e.g., orthopedic surgery and general surgery). In 1997 the mean income of primary care physicians was $155,000, considerably lower than non-primary care physicians ($230,000). However, due to the apparently favorable non-monetary attributes of the primary care specialties and barriers to entry into some non-primary care specialties, each year a majority of graduating

---

equation (1).

[10] The companion paper uses a recently completed survey of the Jefferson alumni that elicits the income the students received as practicing physicians after graduating from medical school.

medical students enter primary care residency programs (Nicholson, 2001).[11]

Our model of specialty choice and income expectations is based on those of Hay (1991) and Willis and Rosen (1979). Medical students are assumed to consider the monetary and non-monetary attributes of each specialty and choose the specialty that maximizes their expected utility. For simplicity, we consider the choice between a primary care (S=0) and a non-primary care specialty (S=1).[12] The difference in the expected utility of entering a non-primary care versus a primary care specialty, I, is defined as a function of student i's characteristics, Z:

$$(2) \qquad I_i = \gamma Z_i + u_i$$
$$\Pr(\text{choose } S = 1) = \Pr(I \geq 0) = \Phi(\gamma Z),$$
$$\Pr(\text{choose } S = 0) = \Pr(I < 0) = 1 - \Phi(\gamma Z),$$

where u has a standard normal distribution. One component of the net benefit of choosing a non-primary care specialty is the income a student expects to receive in a non-primary care specialty relative to the income he expects to receive in a primary care specialty. Students who actually enter a non-primary care specialty expect to receive income EY_1, and students who actually enter a primary care specialty expect to receive income EY_0:

$$(3) \qquad EY_{i1} = \beta_1 X_i + \varepsilon_{i1}$$
$$EY_{i0} = \beta_0 X_i + \varepsilon_{i0},$$

[11] Many students who begin primary care residency programs sub-specialize after completing their initial program (e.g., internal medicine residents can elect to receive further training in cardiology), so that the majority of physicians in the U.S. practice in a non-primary care specialty.

[12] We group the specialties into these two types for tractability of the multi-step estimation procedure below. The main non-primary care specialties are surgery and obstetrics/gynecology, and the main primary care specialties are family practice, internal medicine, pediatrics, and psychiatry. We group psychiatry with the three traditional primary care specialties for this analysis because the mean income of psychiatrists and the length of psychiatric residency training programs are more similar to the primary than the non-primary care specialties.

where **X** is a subset of **Z**.

We assume that u and ε_1 have a bivariate normal distribution with means of zero, standard deviations of one and σ_1, respectively, and correlation ρ_1; u and ε_0 are assumed to have a bivariate normal distribution with means of zero, standard deviations of one and σ_0, respectively, and correlation ρ_0. In the empirical application of this model, to maximize sample size EY will either refer to a medical student's expected peak income over the course of his career, or his expected income with 10 years of experience.[13]

The Jefferson survey records a student's expected income in his chosen specialty only. That is, we observe EY_1 when I > 0 and EY_0 when I < 0. The expected income of students who enter a non-primary care specialty is truncated on a positive expected net benefit of choosing such a specialty, and the expected income of students who enter a primary care specialty is truncated on a negative expected net

$$\begin{aligned} (4) \quad & EY_1 | EY_1 \text{ observed } = EY_1 | I \geq 0 \\ & = \beta_1 X + E[\varepsilon_1 | I \geq 0] \\ & = \beta_1 X + \rho_1 \sigma_1 \lambda_1 \\ & = \beta_1 X + \beta_{\lambda 1} \lambda_1 , \\ & \text{where } \lambda_1 = (\frac{\phi(\gamma Z)}{\Phi(\gamma Z)}) \end{aligned}$$

$$\begin{aligned} (5) \quad & EY_0 | EY_0 \text{ observed } = EY_0 | I < 0 \\ & = \beta_0 X + E[\varepsilon_0 | I < 0] \\ & = \beta_0 X + \rho_0 \sigma_0 \lambda_0 \\ & = \beta_0 X + \beta_{\lambda 0} \lambda_0 , \\ & \text{where } \lambda_0 = -(\frac{\phi(\gamma Z)}{1 - \Phi(\gamma Z)}) \end{aligned}$$

benefit of choosing a non-primary care specialty:

If selection into specialties is non-random, an ordinary least squares regression of students' expected

[13] Expected income with 20 years of experience is omitted because the AMA data for physicians with 20 years of

income on student characteristics would yield inconsistent estimates of β_1 and β_0. Consistent estimates can be obtained if one includes the Mills ratios, or selection-correction terms, λ_1 and λ_0.

λ_1 is non-negative, so a positive coefficient $\beta_{\lambda 1}$ would indicate that the observed expected income of students who enter a non-primary care specialty is greater than the income that students who chose a primary care specialty would expect to earn if they instead chose a non-primary care specialty. Since λ_0 is non-positive, a negative coefficient $\beta_{\lambda 0}$ would indicate that the expected income of students who chose a primary care specialty is also biased upward relative to the population expected income, where the population includes students who enter both specialties. That is, students choosing primary care would have a relative advantage in primary care compared to students choosing non-primary care. If, on the other hand, both $\beta_{\lambda 1}$ and $\beta_{\lambda 0}$ are positive, students who enter non-primary care specialties would have an absolute advantage, in terms of expected income in either specialty, relative to students who enter primary care specialties.

The model can be identified by the non-linearities of the error distributions. We posit other exclusion restrictions in order to make the identification more compelling. Willis and Rosen (1979) argue that variables characterizing a student's family background (e.g., the father's education and occupation, the mother's work status, the family's religion, and number of siblings) affect a student's decision to attend college by affecting the family's ability to finance the investment in human capital. The family background variables were also assumed to be uncorrelated with the additional income a person gains if he attends college. We apply similar reasoning here. We assume that a student's debt and age at the conclusion of the fourth year of medical school, as well as his race, affect his choice of specialty but have little effect on the *difference* in his expected income between non-primary care and primary care specialties, especially after conditioning on his ability and the other control variables **X**. Debt, age, and race are likely to be correlated with a student's (or his family's) ability to finance the relatively lengthy

---

experience used below is not yet available for most of the Jefferson graduates.

residency training in non-primary care specialties. For example, students with substantial debt might prefer a primary care specialty because this allows them to begin paying off their loans more quickly, especially if liquidity constraints are binding. Non-white students are disproportionately likely to be liquidity constrained (Jappellli, Pischke, Souleles, 1998). Similarly, relatively old medical students will tend to avoid non-primary care specialties because they have fewer working years over which to recoup their educational investment.[14] It is also conceivable that the distribution of equalizing differences between primary and non-primary care differs between young and old medical students, and between white and non-white students.[15]

Following Hay (1991) and Willis and Rosen (1979), our estimation strategy consists of three steps. First, we estimate equation (2) with a probit model to obtain the reduced-form estimates of γ. We then compute λ_1 for students who actually enter a non-primary care specialty and λ_0 for students who actually enter a primary care specialty. In the second step we estimate β_1 and $\beta_{\lambda 1}$ in equation (4) by regressing the expected income of students who chose a non-primary care specialty on observed characteristics and λ_1. Likewise, we estimate β_0 and $\beta_{\lambda 0}$ in equation (5) by regressing the expected income of students who chose a primary care specialty on observed characteristics and λ_0.

In order to estimate the responsiveness of specialty choice to expected income, we need to estimate the income a student would expect to receive in the specialty that he decided *not* to enter. According to our model, a student who chose a primary care specialty would have the following expected income in a non-primary care specialty:

[14] Note that age at the conclusion of medical school is not correlated with experience, since the students have no medical experience. Hence a student's age at graduation should not directly affect his expected income.

[15] Bhattacharya (2000) also assumes that debt and age influence specialty choices but not expected income. To further bolster the identification assumption regarding race, we use the 1991 Practice Patterns of Young Physicians survey to explore whether income differs between white and non-white physicians *within* a specialty. We found no evidence of a statistically significant income difference by race in any of the nine specialties we examined: family practice, pediatrics, internal medicine, surgery, ob/gyn, radiology, anesthesiology, psychiatry, and pathology.

$$(6) \qquad EY_1 | EY_0 \ observed = EY_1^* = EY_1 | I < 0$$
$$= \beta_1 X + E[\varepsilon_1 | I < 0]$$
$$= \beta_1 X + \beta_{\lambda 1} \lambda_0$$

We use a similar approach to predict the counterfactual expected income (EY^*_0) in primary care for students who actually entered a non-primary care specialty. For each student we calculate the difference in expected income between a non-primary and primary care specialty as $EY_1 - EY_0^*$ for students who chose a non-primary care specialty, and $EY_1^* - EY_0$ for students who chose a primary care specialty.

In the third step we re-estimate the probit equation (2) after including for each student the difference in expected income between non-primary and primary care specialties. Variables that are assumed to influence expected income but not directly affect the non-monetary value of a specialty, such as a student's board score and his "information accuracy" ($Y^{N,est}_i - Y^N$ in equation 1), are excluded from the final specialty choice regression. We estimate the standard errors of this final probit by jointly bootstrapping all three steps.

We assess the usefulness of the subjective income expectations in three ways. First, we compare the log likelihood of the model in which students are assumed to make specialty choices based on their own subjective income expectations to the log likelihood of the models where students' income expectations are assumed to be static or to match ex-post income. Second, for each of the three models we predict the specialty that each student will choose, and then compare the percentage of choices correctly predicted by each of the models. The coefficients from the specialty choice probits (equation 2) yield a predicted latent utility for each student, which can be translated into a predicted probability of choosing non-primary care. If the predicted probability for a particular student is greater than the proportion of students in the sample that actually chose a non-primary care specialty, we predict that this particular student will choose a non-primary care specialty. Third, we estimate another specialty choice probit that includes both the subjective income expectation and a new variable defined as the difference between the student's static income expectation and their subjective income expectation ($Y^N_{j,t=0} - EY_{i,j,t=0}$). If the coefficient on latter coefficient is insignificant, then information on contemporaneous physicians' income does not provide *incremental* predictive power for specialty choice, beyond the information available in the expectations variables. We perform the same procedure for ex-post income as well (using $Y^N_{j,t=j} - EY_{i,j,t=0}$).

IV. Results

a. Income Expectations

We begin with an analysis of how medical students form expectations of their peak income. Every cohort of Jefferson students has been asked to predict the peak income they will receive during their career, so the resulting sample is large and covers a time period when the health care market has undergone profound change.[16] The mean expected peak income for each cohort of Jefferson students between 1974 and 1998, measured in 1996 dollars, is depicted in Figure 1 for students entering family practice, and in Figure 2 for students entering surgery. These are two of the most popular specialties among the Jefferson students and therefore provide large sample sizes. The corresponding contemporaneous national peak incomes of practicing physicians, from the AMA cross-sectional surveys, are also depicted in these two figures (in 1996 dollars).[17] For instance, for fourth-year medical students graduating in 1987 and entering family practice, the contemporaneous national peak income is the peak income of family practice physicians from the 1986 AMA survey.

In Figure 1, between 1974 and 1987 the expected peak income of students entering family practice corresponded very closely with the contemporaneous peak incomes of physicians in family

[16] We focus on expected peak income because it has been recorded in all survey years. Recall that expected income with 5, 10, and 20 years of experience was asked only before 1980.
[17] We focus our analysis on the 1974-1998 cohorts because age-specific contemporaneous income data are available from the AMA surveys only for these years. The AMA did not conduct a physician survey in 1977, 1980, and 1981, so contemporaneous peak income by specialty is not available for those years as well.

practice. At the start of this time period the income of family practitioners fell substantially and the Jefferson students adjusted their own expectations accordingly. The two lines in Figure 1 diverge in 1987; family practice income has increased in real terms while the students' expectations have remained fairly constant. Nonetheless, the two time series remain correlated. In Figure 2 a similar divergence between the students' expected income and contemporaneous income is evident for students entering surgery. As with family practice, the two time series remain correlated; as surgeons' incomes increased in the late 1980s and early 1990s, the students' expectations likewise increased. In 1993 the students' mean expected peak income for surgery decreased by 25 percent from the previous year, whereas the expectations for family practice were more stable. This difference is probably due to the Clinton administration's proposals to reform the health care system, which were expected to be disproportionately detrimental to specialists. After the proposals failed, the income expectations for surgery sharply recovered. Note that these changes in expected income occurred even though the actual contemporaneous income of practicing surgeons changed very little. This dramatic episode suggests that the income expectations are not purely static, but are in fact responsive to information about future income.

One possible explanation for the divergence of expected and contemporaneous income in the mid 1980s is that Jefferson Medical College, like most medical schools, began accepting more female students in the 1980s. The population of practicing physicians, however, was still predominantly male until the 1990s. Since, as we demonstrate below, female medical students have lower income expectations than their male colleagues, an increase in female students will reduce the mean expected income, all else equal. However, if we restrict our analysis to male medical students only, their expectations also diverge from contemporaneous income in the mid 1980s, not much less than in Figure 1 and Figure 2. An alternative explanation is that over time the ability of the Jefferson students might have fallen relative to the national average; data on the board exam provides some support for this hypothesis.[18] In most specialties

[18] Nationally, the failure rate on Part 1 of the NBME has fallen from about 16 percent to 5 percent between 1990 and 1997, while the percentage of Jefferson students failing the exam has remained fairly constant at about 3-5 percent

physician income increased in the 1980s and was flat or declining in the 1990s. The students' expected income might have diverged from contemporaneous income because medical students anticipated the impending reduction in the growth rate of physician income. The analysis below will formally examine the extent to which the students' expectations were forward-looking. Focusing instead on the late 1970's, in both figures expected peak income decreased substantially between 1978 and 1981. It might have taken students several years to incorporate the unusually high inflation rates at the time into their own forecasts of real income.[19]

We perform a series of regressions of the form described by equation (1) in order to examine the determinants of income expectations. In the first column of Table 2, a student's expected peak income is regressed on personal characteristics only ($\mathbf{X}_i$ in equation 1) in order to examine the role of ability and other individual characteristics. To measure ability, we assign students who received a score on Part 1 of the NBME exam in the top quartile among Jefferson students a value of one for the high board score variable; students who scored in the bottom quartile are assigned a value of one for the low board score variable. Ability can affect expected income in two ways. A person of relatively high ability might be more likely to choose and be admitted into a high-paying specialty; and/or there might be returns to ability within a specialty. Since the first column does not include specialty dummies, the results for ability capture both of these effects. The coefficient on the high board score variable is significantly positive but small in magnitude; ability has a relatively modest effect on income expectations. Students who perform well on the board exam expect their peak career income to be about $10,000 higher than students with average performance. This represents a 5.7 percent premium relative to average income. Women expect their peak income to be $50,000 less than men, a substantial difference.

Female physicians generally work fewer hours than male physicians, however. Another advantage of the Jefferson data set is that some cohorts of students (those graduating before 1980) were

---

during this period.

[19] Souleles (2002) finds that inflation expectations in the Michigan surveys were substantially lower than actual

asked to report the number of hours they expect to work. In unreported regressions (available upon request), when we control for the number of hours students expect to work, the coefficient on the female indicator decreases only slightly (from –50.5 when we omit the expected hours variable in the sub-sample of students for whom we have information on expected hours, to –48.3 when we control for expected hours) in magnitude in the corresponding sample, and remains negative and statistically significant.[20] Thus the lower income expectations of females only partially reflect their expectations of working fewer hours; some of the difference appears to reflect an actual "gender gap".[21]

We next consider whether students condition their expected income on the contemporaneous income of physicians in different specialties. This might happen, for example, if students learn about their future income by observing the incomes of practicing physicians with whom they interact, such as their attending physicians. In the second regression of Table 2 we include the contemporaneous national peak income of practicing physicians ($Y^{N}_{peak,t=0}$ in equation 1) in the specialty the student intends to enter, where t=0 is the year in which the student states his expectation. The estimated coefficient on this variable is 0.67: a one-dollar increase in the peak income of physicians in a particular specialty is associated with a 0.67 dollar increase in the expected peak income of a medical student who plans to enter that specialty. The significance of this coefficient and the increase in R^2 from 0.07 to 0.25 suggests that students do condition their expectations on the contemporaneous income in the specialty they plan to enter. These results are consistent with the static learning model discussed above. Nonetheless, the results are not inconsistent with rational expectations. If income is serially correlated across cohorts, rational expectations should be partly correlated with contemporaneous income. Also, since the coefficient is less than one, expectations are not strictly static; they do not depend only on contemporaneous income. The female coefficient decreases in absolute value because women are more

inflation at this time.

[20] Female medical students in our sample expect to work 57.8 hours per week, on average, versus 62.3 for males. The coefficient on the number of hours a student expects to work is positive and significant.

[21] This conclusion persists when we control for a student's expected specialty. It is possible, for example, that

likely than men to enter a low-paying specialty.

The third regression of Table 2 adds indicator variables for the specialty a medical student intends to enter (omitting family practice) and time dummies for his graduation year (S and T from equation 1). The specialty coefficients measure the average difference in expected income, relative to family practice, over the entire sample period. The coefficient on the national peak income ($Y^N_{peak,t=0}$) of 0.22 is still statistically different from zero although smaller than in the previous specification. This coefficient is now identified by income variation within a specialty over time. Figure 3 plots the peak incomes of physicians in the six different specialties between 1973 and 1997, the years for which AMA surveys are available (reported in 1974-1998). Specialty incomes did not always move together during this time period. For example, the incomes of surgeons and obstetricians increased in the late 1980s relative to the other four specialties, and decreased relative to the other specialties in the early 1990s. This pattern could reflect the introduction of new technologies in the 1980s that were used primarily by specialists, and the advent of managed care in the 1990s, which favored non-specialists.

The results in column 3 suggest that students do incorporate such *relative* changes in specialty income into their own expectations, but not on a dollar for dollar basis. Four of the five specialty coefficients are positive and statistically significant in the third regression, and the coefficients for surgery and obstetrics are large. Thus, students condition their expectations on more than just the current income in their intended specialty. Consider the coefficient of $72,000 on the surgery indicator variable. Conditioning on the contemporaneous income of physicians in different specialties, students who intend to become surgeons expect their future peak income to be $72,000 higher than students who intend to become family practitioners, on average. This is consistent with the possibility that prospective surgeons were forward-looking; they expected income to grow in the future relative to family practice. In fact, on average over the 1973 to 1997 period, the real peak income of surgeons and obstetricians increased by 18

---

female medical students anticipate discrimination even within a given specialty.

percent and 6 percent, respectively, while the peak income of family practitioners decreased by 18 percent (see Figure 3).

In the fourth regression of Table 2 we also control for a student's knowledge of current physician income. The information accuracy variable ($Y^{N,est}_{i,t=0}$ - $Y^{N}_{t=0}$ in equation 1) is the difference between a student's own assessment of the current income of physicians in the specialty he plans to enter and the contemporaneous national median income of physicians in that specialty, as measured by the AMA surveys. The estimated coefficient on this variable is 0.84: a student's market misinformation is incorporated almost dollar for dollar into his own income expectation. If, for example, a student who plans to become a pediatrician believes that pediatricians currently make $10,000 more than they actually do, his expected peak income will be $8,400 higher than a similar person whose information regarding pediatricians' income is correct. Adding the information accuracy variable increases the R^2 from 0.31 to 0.45. Hence, a considerable amount of the variation in income expectations appears to be due to heterogeneity of information regarding the physician market. The coefficient on the contemporaneous national peak income is substantially larger in the fourth specification relative to the third specification, and the coefficients on the surgery and obstetrics indicator variables are substantially smaller. This latter result suggests that part of the reason students entering surgery and obstetrics expected their income to grow in the future relative to other specialties was that they overestimated the current income of physicians in those specialties.

One disadvantage of using the expected peak income variable is that we do not know when in his career a student expects his peak income to occur. By contrast, information on expected income for 5 and 10 years of experience is available for a smaller sample, the students who graduated before 1980. This more detailed information allows us formally to test whether students were able to anticipate future income changes in the specialty they intend to enter. In Table 3 we pool each student's expected income observations for 5 and 10 years of experience and include a variable that measures the future ex-post change in specialty income in the student's specialty (the variable ($Y^N_{j,t=j}$ - $Y^N_{j,t=0}$) in equation (1)).[22] Standard errors have been corrected to allow for correlation in the error terms between the multiple observations for given respondent, and an indicator variable is included for the observations with 10 years of experience (five years of experience is the omitted variable).

In the first regression in Table 3, the coefficient on the indicator variable for 10 years of experience implies that students expect their incomes to increase by about $24,000 on average between the 5th and 10th years of experience. The coefficient of 0.27 on the ex-post income growth variable indicates that students do incorporate future changes in specialty income into their own expectations. This is strong evidence that income expectations are forward-looking, and not entirely adaptive. Students were able to anticipate some of the relative changes in income across specialties that are illustrated in Figure 3. The second regression in Table 3 omits the income growth variable for purposes of comparison with the regressions in Table 2. Comparing the first and second regressions in Table 3, the coefficient estimates do not change substantially when the income growth variable is included.

b. Specialty Choice

In order to examine whether subjective income expectations data help predict behavior, we analyze the decision by medical students to enter a primary or non-primary care specialty after graduating from medical school. We estimate probit models of equation (2), in which the dependent variable is one if a student chooses a non-primary care specialty, and zero otherwise. We first analyze the specialty choice decision under the assumption that income expectations are static; i.e., students expect their peak lifetime income in primary and non-primary care to be equal to the contemporaneous peak incomes of

[22] For peak income, since we do not know the year j in which income is expected to peak, we cannot compute $Y^N_{peak,t=peak}$ in the AMA data. Hence we focus on the expectations for j=5 and j=10 years ahead. As noted above, expected income with 20 years of experience is omitted because the AMA data for physicians with 20 years of experience is not yet available for most of the Jefferson graduates.

practicing physicians in these two specialty groups, according to the AMA data.[23] Since the AMA does not release individual-level data, in this first model we do not control for non-random selection into the various specialties.

The results appear in the first column of Table 4.[24] Of greatest interest is the coefficient on the variable (labeled "B") measuring the difference between the contemporaneous peak income of non-primary care and primary care physicians, ($Y^{N,NPC}_{peak,t=0}$ - $Y^{N,PC}_{peak,t=0}$), where PC refers to primary care, NPC refers to non-primary care, and N refers to the national AMA data. The estimated coefficient is 0.00389 and is significant. To gauge the corresponding marginal effect, a $10,000 increase in the contemporaneous income of non-primary care relative to primary care physicians is associated with an increase of 0.014 (from 0.360 to 0.374) in the probability that a medical student will choose a non-primary care specialty. , In addition, female and white students are less likely to choose a non-primary care specialty relative to their peers. As indicated in the last row, this first model correctly predicts the specialty choices of 57.3 percent of the students.

We compare this model in which students are assumed to have static income expectations to a model that uses the students' explicit income expectations. The latter model is estimated in three steps to control for the possibility of non-random selection into the specialties. We first estimate a reduced-form probit model as specified by equation (2). Coefficient estimates from this model are reported in the second column of results in Table 4. The coefficients on individual characteristics, including gender and race, have the same sign and a similar magnitude as in the first specification that assumes static income expectations. Board scores and the students' misperceptions of contemporaneous physician income (income information accuracy) have been shown in the previous section to affect income expectations. Table 4 includes the board score and the difference in the students' misperceptions for non-primary care

[23] We weight the peak incomes in surgery and obstetrics to derive a contemporaneous non-primary care peak income. The weights are based on the number of practicing physicians in each of these two non-primary care specialties. Likewise, we weight the peak incomes in family practice, internal medicine, pediatrics, and psychiatry to derive a contemporaneous primary care peak income.

relative to primary care specialties. [25] We assume that these two variables affect specialty choice primarily through income expectations, and therefore they are included in the reduced-form probit but not the probits that directly include income expectations. The reduced form model correctly predicts 58.3 percent of specialty choices, slightly better than with the model with static income expectations. The coefficients from the reduced-form probit regression are used to derive a Mills ratio for each student, as specified in equations (4) and (5).

For students entering non-primary care and primary care specialties, we separately regress the students' subjective income expectations on individual characteristics and the Mills ratios, to control for selection. The results appear in Table 5. The coefficient on the Mills ratio λ_1 for non-primary care is positive and significant. This suggests that the students who actually chose non-primary care expect to earn more in that specialty than would students who actually chose primary care. The coefficient on λ_0 for primary care is negative and insignificant. These results are consistent with comparative rather than absolute advantage. Women expect to earn substantially less than men in both specialties, but particularly in the non-primary care specialties.

The coefficients from Table 5 are used to estimate each student's counterfactual expected income in the specialty not chosen, as outlined in equation (6). We then re-estimate the specialty choice probit equation (2) after including the difference in the expected peak income between non-primary and primary care for each student, $EY^{NPC}_{i,peak} - EY^{PC*}_{i,peak}$ or $EY^{NPC*}_{i,peak} - EY^{PC}_{i,peak}$ (labeled "A" in Table 4). Coefficient estimates for the selection-corrected probit model are reported in the third column of Table 4, with the standard errors jointly bootstrapped across all three estimation steps. The coefficient on the difference in students' income expectations is positive, and significant at the 10 percent level. It is about

[24] The sample size is slightly smaller than in Table 2 because of missing observations, mostly in the debt variable.
[25] The relative accuracy variable is the difference between a student's perception of the mean contemporaneous non-primary care income and the actual contemporaneous mean non-primary care income ($Y^{N,est,NPC}_{i,t=0} - Y^{N,NPC}_{t=0}$), minus the difference between their perception of the mean contemporaneous primary care income and the actual contemporaneous mean primary care income ($Y^{N,est,PC}_{i,t=0} - Y^{N,PC}_{t=0}$).

four-times greater in magnitude than the corresponding coefficient from the model with static income expectations (variable B). A $10,000 increase in a student's expected income in non-primary relative to primary care is associated with an increase of 0.057 in the probability of entering non-primary care (from 0.360 to 0.417), which is an economically significant effect.

Another way to measure the usefulness of subjective income expectations is to compare the log likelihood of the specification with subjective expectations to the model with static income expectations. The log likelihood of the model with subjective income expectations (-819) is considerably larger than the log likelihood of the model in which students expect to earn the contemporaneous peak income (-1,561). Furthermore, the model with subjective income expectations correctly predicts the specialty choice for 85.6 percent of the sample, compared to 57.3 percent using the model with static expectations. Students' explicit income expectations appear to be much more useful for predicting specialty choice than static expectations.

The previous section showed that income expectations are based, in part, on the contemporaneous income of practicing physicians. The final column of Table 4 separates the effects of contemporaneous income and subjective income expectations on specialty choice. This specification includes both the students' subjective income expectations (variable A) and a variable defined as the difference between the static and subjective income expectations (variable B - variable A). The coefficient on the former variable is still positive and significant, now at the 5 percent level. The coefficient on the latter variable, which represents information in contemporaneous physician incomes that is not present in the students' subjective income expectations, is insignificant. Also, including information on contemporaneous income hardly changes the fit of the model as measured by the log likelihood or the percent of specialty choices correctly predicted. Thus, conditional on subjective income expectations, contemporaneous income is no longer informative about specialty choices. That is, the income expectations variables appear to include all the information in contemporaneous income that is relevant for specialty choice.

We have also estimated a version of the specialty choice model that contains indicator variables for the year a student completed medical school in order to allow for changes over time in the non-monetary attributes of the specialties (results not shown). The coefficient for the difference in subjective expected income (variable A), which is now identified by variations between students within a cohort, is slightly larger in magnitude (0.0189 versus 0.0165) than the model without the year indicators, and the other results are essentially unchanged.

As a further test of the usefulness of subjective income, we compare the results with a model that assumes medical students have perfect foresight regarding their income. Specifically, we assume that when fourth-year medical students were choosing a specialty, they expected their income with 10 years of experience to equal the mean income that was actually received in that specialty by their cohort 10 years later, according to AMA data ($Y^{N}_{10,t=10}$). We use 10 years of experience because it is sufficiently far into a physician's career that it should correlate closely with his peak income.[26]

For example, consider a student who is completing medical school in 1980 and forming expectations regarding her income in family practice. After completing a three-year family practice residency program, her 10th year of experience would occur in 1993. We assume that this student's expected income in family practice is equal to the actual mean income of family practitioners with 10 years of experience in 1993. As before, we weight expected incomes in the various specialties to derive an overall expected income for primary and non-primary care. Since the AMA data are aggregated, in the perfect foresight model we do not control for non-random selection into the various specialties. The sample for this specification consists of the students who graduated from Jefferson Medical College between 1971 and 1986, for whom the 10-year expectations variable is available.

In the first column of Table 6 we present the coefficient estimates from the specialty choice probit when expectations are based on ex-post income. The variable of greatest interest is the difference

---

[26] Recall that we use peak income in the subjective income expectation model. However since we don't know exactly when the peak is expected to occur, we cannot calculate the corresponding actual peak income realization for his cohort of medical students. We omit from the analysis students who graduated after 1986 because we do not observe their cohorts' actual income with 10 years of experience.

between the actual mean income of non-primary and primary care physicians with 10 years of experience for physicians in a student's cohort, $Y^{N,NPC}_{10,t=10} - Y^{N,PC}_{10,t=10}$ (variable B). Its estimated coefficient is negative and insignificant. This model correctly predicts the specialty choice for only 51.4 percent of the students.

For purposes of comparison, we re-estimate the subjective income expectation model with this smaller sample of students. The results are reported in the remaining columns of Table 6. In the third column of results, the coefficient on variable A, the difference in students' subjective income expectations with 10 years of experience (0.0112), is slightly smaller than in the previous specification (0.0165 in Table 4) and is not significantly different from zero. Our estimation method is identical to that of Table 4, so the reduced significance is presumably due to the smaller sample size. The model correctly predicts the specialty choices for 72.8 percent of the students.

In the fourth column of results in Table 6 we estimate a probit model that includes students' subjective income expectations (variable A) as well as the difference between the ex-post income of the students' cohorts and the students' explicit income expectations (variable B - variable A). The coefficients on both variables are positive, although neither is significant. Therefore, we reject the hypothesis that ex post income provides additional predictive power for specialty choice. That is, subjective income expectations summarize most of the explanatory power of ex-post income for specialty choice. Including ex-post income increases the log likelihood slightly and improves the percentage of correct specialty choice predictions by about one percentage point.

V. Conclusion

This paper analyzes the determinants of income expectations for a prominent group of students, medical students. Their expectations are found to vary significantly with their individual characteristics;

for instance, students who perform relatively well on the national medical board exam expect to earn more than their colleagues. Female medical students expect to earn substantially less than male medical students, even after controlling for the number of hours per week they expect to work. Furthermore, medical students appear to condition their income expectations on the contemporaneous income of physicians currently practicing in the specialty they plan to enter. Indeed, students' misinformation about physicians' contemporaneous income affects their expectations of their own income almost dollar for dollar. If, for example, a student who plans to become a pediatrician believes that pediatricians currently make more than they actually do, the student's expectations for his own income in pediatrics rises accordingly. These results are consistent with a model in which students learn by observing the practicing physicians with whom they interact. Therefore, a considerable amount of the variation in income expectations appears to be due to heterogeneity in information in addition to other individual characteristics.

Nonetheless, students' expectations are not strictly static; they anticipate future trends in specialty income. After students report relatively high income expectations for a given specialty, that specialty subsequently tends to experience high income growth relative to other specialties. For example, the students anticipated the fact that the income of specialists rose in the 1980s relative to the income of primary care physicians, but then decreased relative to primary care physicians in the 1990s.

We also find that subjective income expectations help explain an important life-cycle decision, the specialty choice of medical students. A $10,000 increase in a student's expected income in non-primary care relative to primary care specialties is associated with an increase of 0.057 in the probability of entering a non-primary care specialty. Subjective income expectations are more useful in predicting specialty choice than the static expectations often used in the literature, and than assuming that students have perfect foresight about their future income. That is, subjective income expectations summarize most of the information in either contemporaneous or ex-post income that is useful in explaining specialty choice. More generally, these results suggest that subjective expectation questions can help predict people’s behavior, not only their investment in human capital but also other forward-looking decisions.

References

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

Bhattacharya, Jayanta, 2000, "Lifetime Returns to Specialization in Medicine," mimeo.

Das, Marcel and Arthur van Soest, 1999, "A Panel Data Model for Subjective Information on Household Income Growth," *Journal of Economic Behavior & Organization* 40: 409-426.

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

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

Freeman, Richard, 1971. The Market for College-Trained Manpower. Cambridge, MA: Harvard University Press.

Jappelli, Tullio, Jorn-Steffen Pischke, and Nicholas S. Souleles, 1998, "Testing for Liquidity Constraints in Euler Equations with Complementary Data Sources," *The Review of Economics and Statistics*: 251-262.

Manski, Charles F., 1993, "Adolescent Econometricians: How do Youth Infer the Returns to Schooling?," in Studies of Supply and Demand in Higher Education, eds. C. Clotfelter and M. Rothschild, Chicago: University of Chicago Press: 43-57.

Nicholson, Sean and Nicholas S. Souleles, 2002, "Physician Income Prediction Errors: Sources and Implications for Behavior." NBER Working Paper 8907.

Nicholson, Sean, 2001, "Physician Specialty Choice Under Uncertainty," *Journal of Labor Economics, forthcoming*.

Nicholson, Sean, 2002, "How Much Do Medical Students Know About Physician Income: Accuracy, Bias, and Implications for Career Decisions," mimeo.

Siow, Aloysius, 1984, "Occupational Choice Under Uncertainty," *Econometrica* 52(3): 631-645.

Sloan, Frank A., 1970, "Lifetime Earnings and Physicians' Choice of Specialty," *Industrial and Labor Relations Review* 24: 47-56.

Souleles, Nicholas S., 2002, "Consumer Sentiment: Its Rationality and Usefulness in Forecasting Expenditure," *Journal of Money, Credit, and Banking*, forthcoming.

Souleles, Nicholas S., 2000, "Household Portfolio Choice, Hedging Motives, and Transactions Costs," mimeo.

Willis, Robert J. and Sherwin Rosen, 1979, "Education and Self-Selection," *Journal of Political Economy*

87(5): S7-S36.

Zarkin, Gary A., "Occupational Choice: An Application to the Market for Public School Teachers," *Quarterly Journal of Economics* 100(2): 409-446.

Table 1
Sample Summary Statistics

Variable	Mean	Standard Deviation
Age at graduation	26.8	2.88
Female	0.255	0.436
White	0.867	0.340
Part 1 NBME board exam score	204.7	17.7
Debt ($000)	39.7	40.4
Expected income, in 1996 dollars ($000)		
5 years of experience: EY_5	95.5	41.3
10 years of experience: EY_{10}	141.8	55.1
20 years of experience: EY_{20}	169.5	67.6
Peak income: EY_{peak}	170.9	86.8
Expected hours worked/week	63.2	11.5
Chosen specialty		
- internal medicine	0.307	0.461
- family practice	0.206	0.405
- pediatrics	0.091	0.288
- surgery	0.282	0.450
- ob/gyn	0.074	0.261
- psychiatry	0.041	0.198
Accuracy of income information ($000) ($Y^{N,est}_i$ - Y^N)	-7.33	45.8

Notes: N=3,025. Sample period covers 1971-1998. Students who graduated after 1979 were not asked to predict their income 5, 10, and 20 years after completing residency training, nor the number of hours they would work. Sample means for these variables reflect the responses from students who graduated before 1980. Income expected after j years of experience (EY_j) was recorded by the Jefferson surveys during the fourth year of medical school (t=0). Accuracy of income information is the difference between $Y^{N,est}_{i,t=0}$, the student's assessment of current mean physician income in the specialty he plans to enter, and $Y^N_{t=0}$, actual current income in that specialty, as measured by the national AMA surveys. Sample sizes in subsequent tables can differ due to missing variables. All income variables are expressed in thousands of 1996 dollars.

Table 2
Determinants of Expected Peak Income

	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Female	-49.8**	3.07	- 37.3**	2.74	-30.2**	2.92	-27.3**	2.67
White	-7.47	4.91	12.0**	4.57	3.65	4.58	3.00	4.12
High board score	10.3**	4.21	4.30	3.77	7.87**	3.70	7.66**	3.34
Low board score	-1.46	3.89	4.55	3.48	3.78	3.36	3.36	3.01
National peak income, $Y^N_{peak,t=0}$			0.667**	0.029	0.218**	0.085	0.590**	0.076
Accuracy of income information ($Y^{N,est}_{i,t=0}$ - $Y^N_{t=0}$)							0.835**	0.047
Expected specialty (family practice is omitted)								
- internal medicine					15.7**	5.19	5.58	4.61
- pediatrics					- 4.68	3.04	4.59	2.84
- surgery					72.4**	11.3	33.1**	9.90
- ob/gyn					63.7**	9.51	29.8**	8.16
- psychiatry					14.4**	7.14	17.2**	6.41
Constant	189**	5.12	36.1**	7.70	98.7**	15.1	55.8**	13.8
Indicator variables for year of graduation					Included		Included	
Observations	2,824		2,824		2,824		2,824	
R^2	0.07		0.25		0.31		0.45	

Notes: The dependent variable is medical students' expectations of their peak income, EY_{peak}, as reported in the Jefferson surveys during the fourth year of medical school (t=0). This variable, and all other income variables, are expressed in thousands of 1996 dollars. The omitted indicator variables are male, nonwhite, middle medical board score (between the 25th and 75th percentile) and family practice specialty. Following equation (1), $Y^{N}_{peak,\ t=0}$ is the contemporaneous peak income of physicians currently practicing in the specialty the student intends to enter, in year t=0, based on national AMA data. $Y^{N,est}_{i,t=0}$ is the student's assessment of current mean physician income in the specialty he plans to enter ($Y^{N}_{t=0}$). ** = significantly different from zero at the 5 percent level. * = significantly different from zero at the 10 percent level.

Table 3
Determinants of Expected Income with 5 & 10 Years of Experience

	Coefficient	S.E.	Coefficient	S.E.
Female	-19.0**	3.48	-18.8**	3.47
High board score	2.48	3.57	2.49	3.56
Low board score	0.319	3.88	0.269	3.87
National income: $Y^N_{j,t=0}$	0.509**	0.0944	0.366**	0.0924
Accuracy of income information: $Y^{N,est}_{i,t=0}$ - $Y^N_{,t=0}$	0.402**	0.0339	0.401**	0.0338
Growth in national income between year 0 and year j : $Y^N_{j,t=j}$ - $Y^N_{j,t=0}$	0.270**	0.0482		
Year of income expectation (5 yrs omitted):				
- 10 years experience	23.7**	3.80	35.1**	3.50
Expected specialty indicator				
- internal medicine	-10.5**	3.65	-3.72	3.49
- pediatrics	7.07	4.76	6.45	4.77
- surgery	-34.2**	8.15	-1.63	6.11
- ob/gyn	-21.6**	6.84	-1.19	5.82
- psychiatry	-6.66	5.38	-2.51	5.33
Constant	37.8**	11.9	43.3**	11.9
Observations	1,440		1,440	
R^2	0.41		0.40	

Notes: Dependent variable is the medical students' expected income with 5 and 10 years of experience, EY_5 and EY_{10}. Time dummies for a student's graduation year are included in each regression. The sample is limited to students who graduated before 1980 because expected income with 5 and 10 years of experience was not asked afterwards. Growth in national income measures the ex-post change in average income in the student's specialty, over the next j years, using the AMA data. Standard errors are adjusted to allow for within-student correlation. For other notes, see Table 2. ** = significantly different from zero at the 5 percent level. * = significantly different from zero at the 10 percent level.

Table 4
Probit Models of Specialty Choice: Subjective Income Expectations vs. Static Expectations

	Contemporaneous national income		Reduced form		Selection-corrected, explicit income expectations		Explicit vs. Contemporaneous expectations	
Variable	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
A: Diff. in expected income (non-primary - primary care): $EY^{NPC}_{i,peak,t=0} - EY^{PC}_{i,peak,t=0}$					0.0165*	0.00860	0.0134**	0.00533
B: Diff. in contemporaneous income of MDs $Y^{N,NPC}_{peak,t=0} - Y^{N,PC}_{peak,t=0}$	0.00389**	0.00103						
B – A							- 0.00332	0.00389
Female	- 0.468**	0.0637	- 0.467**	0.0648	- 0.125	0.0888	- 0.0941	0.0901
Age	- 0.0773	0.0847	- 0.0614	0.0855	0.0853	0.109	0.105	0.110
Age squared	- 0.00196	0.00165	- 0.00167	0.00166	- 0.00222	0.00211	- 0.00254	0.00213
White	- 0.272**	0.0771	- 0.304**	0.0807	- 0.430**	0.102	- 0.464**	0.103
Debt ($00,000)	- 0.278	0.203	- 0.336	0.209	- 0.522*	0.271	- 0.451*	0.272
Debt squared	-2.10×10^{-6}	1.54×10^{-6}	2.41×10^{-6}	1.64×10^{-6}	3.76×10^{-6} *	2.08×10^{-6}	3.67×10^{-6} *	2.08×10^{-6}
Board score	---	---	0.0434*	0.0262	---	---	---	---
Board score squared	---	---	- 0.00010	0.000064	---	---	---	---
Relative accuracy of income information	---	---	0.0833	0.425	---	---	---	---
Constant	-1.09	1.07	-4.80*	2.87	-1.28	1.38	-1.23	1.39
Observations	2,458		2,458		2,458		2,458	
Log likelihood	-1,561		-1,542		-819		-816	
Pseudo R^2	0.027		0.038		0.49		0.49	
Percent predicted correctly	57.3		58.3		85.6		85.5	

Notes: Dependent variable is one if the medical student chose a non-primary care specialty and zero if he chose a primary care specialty. See equation (2). The first column includes the difference in peak income between non-primary care (NPC) and primary care (PC) specialties, using contemporaneous physician income in each specialty according to the AMA surveys (variable B). The third column uses instead the difference in peak income each student expects to receive in non-primary care relative to primary care specialties (variable A), using the three-step procedure outlined in the text. The standard errors have been jointly bootstrapped across all three steps. Expected income has been corrected for non-random selection, based on the reduced form model in the second column and in Table 5. The fourth column includes both variable B and the difference between variables A and B. Relative accuracy is the difference between the students' perception of the mean contemporaneous non-primary care income and the actual mean contemporaneous non-primary care income ($Y^{N,est,NPC}_{i,t=0}$ - $Y^{N,NPC}_{t=0}$), minus the difference between their perception of the mean contemporaneous primary care income and the actual mean contemporaneous primary care income ($Y^{N,est,PC}_{i,t=0}$ - $Y^{N,PC}_{t=0}$). Age and debt refer to the 4th year of medical school. For race, non-white is the omitted variable. For additional notes, see Table 2. ** = significantly different from zero at the 5 percent level; * = significantly different from zero at the 10 percent level.

Table 5
Selection Corrected Models of Expected Peak Income

Variables	Non-primary Care	Primary Care
Female	-76.7**	-22.1**
	(14.8)	(4.83)
Board score	1.82	-0.916
	(4.15)	(1.28)
Board score squared	-0.00334	0.00264
	(0.0102)	(0.00312)
Relative accuracy of income information	0.655**	1.03**
	(0.123)	(0.122)
λ	72.3*	-20.0
	(37.0)	(18.1)
Constant	38.8	248*
	(431)	(127)
Observations	886	1,583
R^2	0.25	0.30

Notes: The dependent variable is medical students' expectations of their peak income, $EY_{i,peak,t=0}$. The Mills ratio λ is computed from the reduced form probit results in Table 4 (see equations 4 and 5). Including the Mills ratios controls for the fact that EY is observed only in the specialty each student actually chooses. Indicator variables are included for the year a student graduated from medical school. For additional notes, see Table 4. ** = significantly different from zero at the 5 percent level; * = significantly different from zero at the 10 percent level.

Table 6
Probit Models of Specialty Choice: Subjective Income Expectations vs. Perfect Foresight

Variable	Ex-Post Income in specialty Coefficient	S.E.	Reduced Form Coefficient	S.E.	Selection-Corrected, Explicit Income Expectations Coefficient	S.E.	Explicit vs. Ex-Post Expectations Coefficient	S.E.
A: Diff. in expected income (non-primary - primary care): $EY^{NPC}_{i,10,t=0} - EY^{PC}_{i,10,t=0}$					0.0112	0.0118	0.0205	0.0190
B: Ex post income diff. for student's cohort of MDs $Y^{N,NPC}_{10,t=10} - Y^{N,PC}_{10,t=10}$	-0.000506	0.00338						
B – A							0.00902	0.00854
Female	- 0.407**	0.111	- 0.451**	0.113	- 0.183	0.128	-0.191	0.129
Age	- 0.278**	0.138	- 0.336**	0.140	-0.282*	0.152	-0.264*	0.152
Age squared	0.00534*	0.00281	0.00641**	0.00285	0.00543*	0.00308	0.00505	0.00308
White	- 0.380**	0.154	- 0.392**	0.158	-0.582**	0.177	-0.602*	0.177
Debt ($00,000)	- 0.198	0.425	0.0635	0.453	-0.180	0.482	-0.0820	0.487
Debt squared	2.07×10^{-6}	5.05×10^{-6}	-2.90×10^{-6}	5.60×10^{-6}	2.05×10^{-6}	5.81×10^{-6}	8.37×10^{-7}	5.90×10^{-6}
Board score	---	---	0.0460	0.0391	---	---	---	---
Board score squared	---	---	- 0.00012	0.00010	---	---	---	---
Relative accuracy of income information	---	---	0.166	0.537	---	---	---	---
Constant	0.356**	0.174	-0.258	4.26	3.54*	1.87	2.39	1.93
Observations	1,149		1,149		1,149		1,149	
Log likelihood	- 719		- 707		- 549		- 546	
Pseudo R^2	0.015		0.032		0.25		0.25	
Percent predicted correctly	51.4		58.1		72.8		74.0	

Notes: Dependent variable is one if the medical student chose a non-primary care specialty and zero if he chose a primary care specialty. See equation (2). The first column includes the difference in ex-post income between non-primary care (NPC) and primary care (PC) specialties, using the mean physician income realized by the student's cohort 10 years later in each specialty according to the AMA surveys (variable B). For other notes, see Table 4. ** = significantly different from zero at the 5 percent level; * = significantly different from zero at the 10 percent level.

Figure 1

Expected Peak Income in Family Practice, 1974-1998

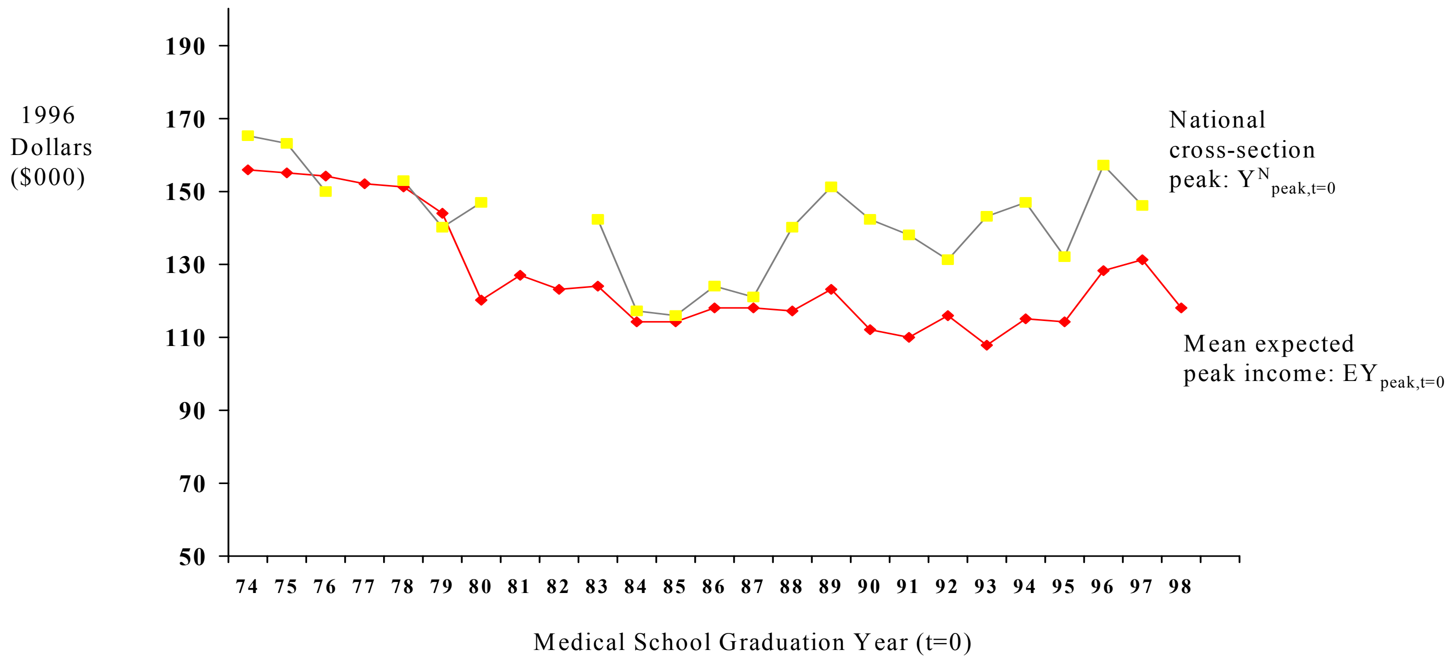


Figure 2

Expected Peak Income in Surgery, 1974-1998

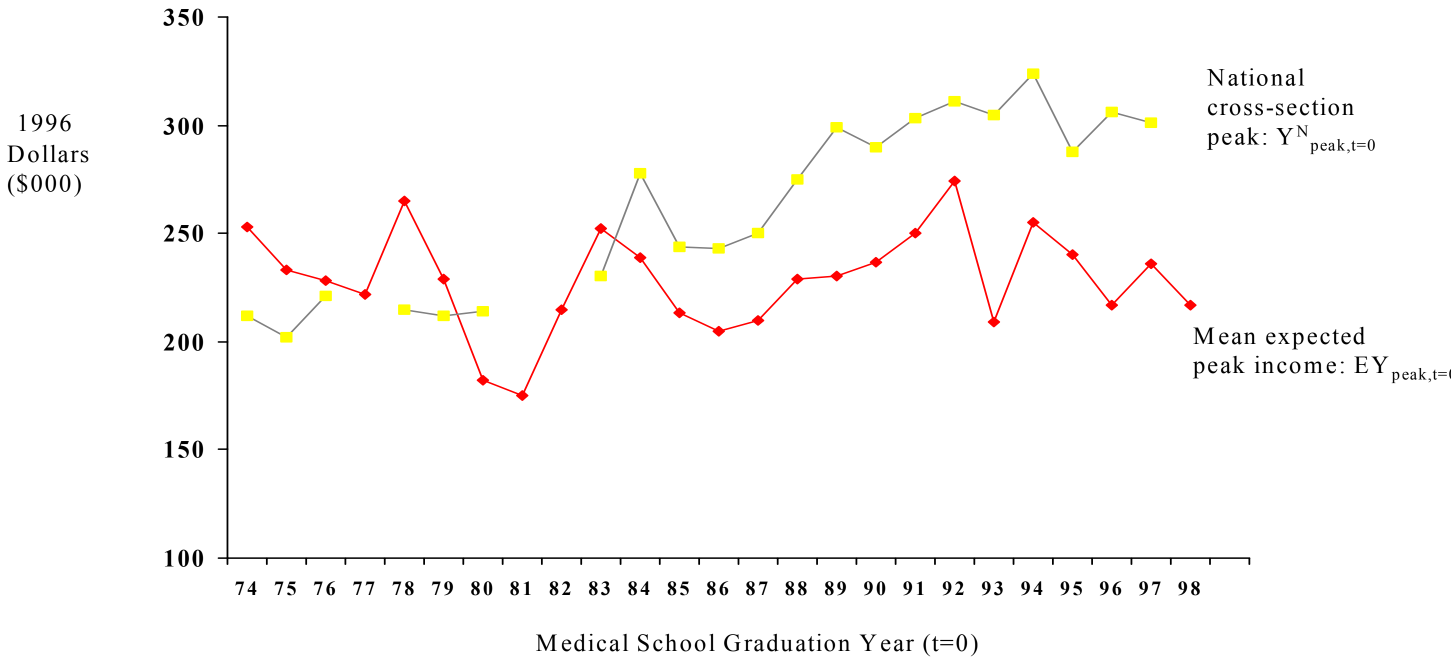
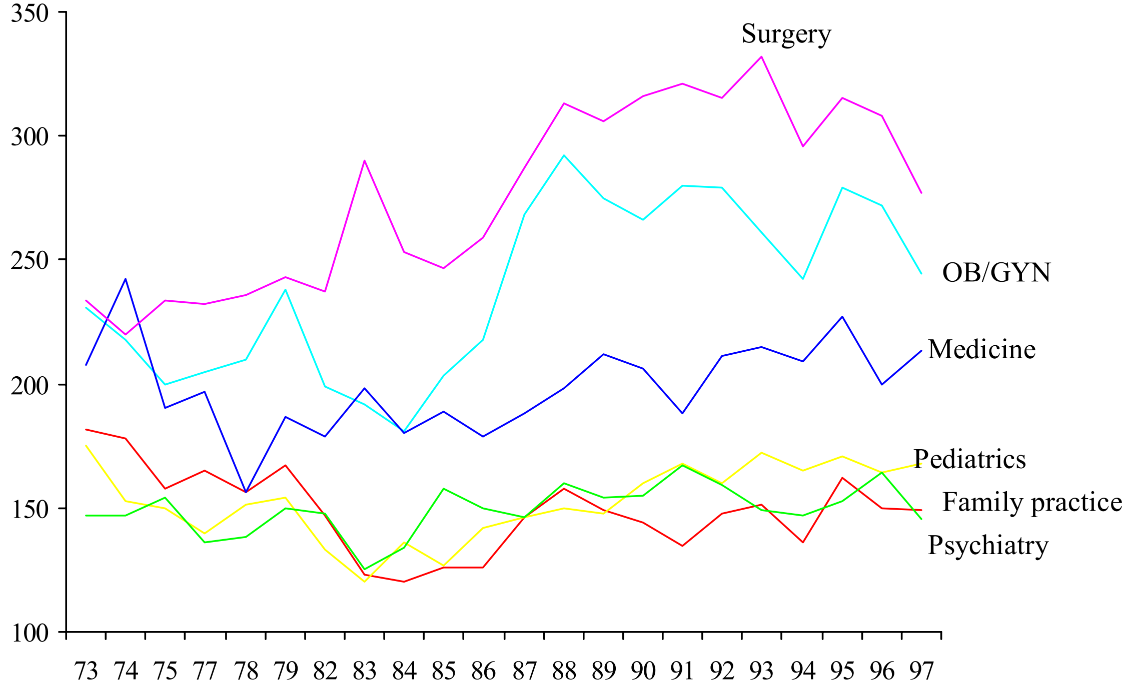


Figure 3

National Peak Physician Income by Specialty, 1973 - 1997
(1996 $000)



Source: AMA Socioeconomic Monitoring Study