THE PRODUCTIVITY OF U.S. POSTSECONDARY INSTITUTIONS

Caroline M. Hoxby†
July 2017 version

Abstract

The productivity of a postsecondary institution is its causal effect a student's lifetime outcomes ("value-added") divided by the lifetime cost of producing this effect. This study estimates the productivity of approximately 6,700 programs, which account for vast majority of undergraduate education in the U.S. The study remedies the selection problem by comparing outcomes for students who not only have the same measured incoming achievement but who also apply to the same colleges, thus revealing similar interests and motivation. It uses longitudinal data to compute lifetime costs, and it computes benefits based not only on earnings but also on public service and innovation. The study's most striking finding is that, when earnings are used to measure benefits, the productivity of a dollar is fairly similar across a wide array of selective postsecondary institutions. This is noteworthy because the most selective schools spend several times as much per student as those that are only modestly selective. The result suggests that market forces compel an allocation of educational resources that is roughly efficient among selective institutions. However, compared to selective institutions, non-selective postsecondary institutions are less productive on average and vary greatly in their productivity. This result may indicate that market forces exert little discipline upon non-selective schools, allowing non-productive ones to attract students even when located side-by-side with more productive ones. The estimates of productivity are sufficiently high to rationalize, under certain economic logic, the tax deductibility of gifts to non-profit postsecondary schools.

---

† Stanford University and National Bureau of Economic Research. The opinions expressed in this paper are those of the author alone and do not necessarily represent the views of the Internal Revenue Service or the U.S. Treasury Department. This work is a component of a larger project examining the effects of federal tax and other expenditures that affect higher education. Selected, de-identified data were accessed through contracts TIR-NO-12-P-00378 TIR-NO-15-P-00059 with the Statistics of Income (SOI) Division at the U.S. Internal Revenue Service. The author gratefully acknowledges the help of Barry W. Johnson, Michael Weber, and Brian Raub of the Statistics of Income Division, Internal Revenue Service. James Archsmith and Steven E. Stern provided important help with paired comparison methods. The author is grateful for comments from Joseph Altonji, David Autor, Sandra Black, Scott Carrell, John Cawley, Paul Courant, David Deming, Maria Fitzpatrick, Brian Jacob, Michael Lovenheim, Paco Martorell, Jordan Matsudeira, Jonathan Meer, Douglas Miller, Marianne Page, Juan Saavedra, Judith Scott-Clayton, Douglas Staiger, Kevin Stange, Sarah Turner, Miguel Urquiola, and Seth Zimmerman.

1 Introduction

This paper proposes procedures for measuring the productivity of U.S. postsecondary institutions. It also implements these procedures for most undergraduate programs. The evidence has interesting implications. For instance, at least at selective institutions, a dollar spent on a student's education appears to generate multiple dollars of value-added based on earnings over her lifetime. Productivity is also stable across a wide range of selective institutions, suggesting that market forces are sufficiently strong to maintain regularity in how these institutions' resources scale up with the capacity of their students to convert educational resources into value-added. Compared to selective institutions, non-selective institutions have productivity that is lower on average but also much more dispersed. This suggests that market forces may be too weak to discipline productivity among these schools.

This study also examines institutions' productivity based on their producing public service and innovation. Public service productivity varies substantially even among selective schools. Innovation productivity is distinctly higher at very selective schools than all other schools.

The study attempts to cover considerable ground, and is in a "glass half full" mode in the sense that it attempts to answer key questions about productivity in higher education while acknowledging that it cannot answer them all or answer them perfectly. In the first part of the study, I define what is meant by productivity in higher education and explain why measuring it is a useful exercise even if an imperfect one. I outline the key issues that plague measurements of productivity: (i) the multiplicity of outcomes that schools might affect and the difficulty of measuring some of them; (ii) the fact that a student may enroll in several schools before her education is complete (the "attribution problem"); (iii) selection.

Since vertical selection (students who differ in aptitude enrolling at different institutions) and horizontal selection (similarly apt students who differ on grounds like geography enrolling at different institutions) are arguably the most serious issues for measuring productivity, I especially discuss methods for addressing these problems. The proposed remedy for the selection problem is based on comparing the outcomes of students who are the same on measured incoming achievement (test scores, grades) and who also apply to the same postsecondary institutions, thus demonstrating similar interests and motivation. This approach employs all the possible quasi - experiments in which a student "flips a coin" between very similar schools or in which admission staff members "flip a coin" between very similar students. Put another way, schools are compared solely on the basis of students who are in the common support (likely to attend either one) and who are quasi-randomly allocated among them. See below for detail about the method.

Using this method to account for selection, the study computes productivity for approximately 6,700 undergraduate programs. I show productivity for three outcomes: earnings, a measure of public service based on the earnings a person forgoes by being employed in the non-profit or public sector (think of a talented attorney employed as a judicial clerk), and a measure of participation in innovation. The first measure is intended to reflect private returns; the second social returns; the third spillovers to economic growth.

In the next section, I define productivity for the purposes of this paper. Section 3 explains why productivity measures would be useful to policy making but also for numerous other reasons. The key challenges we face in measuring productivity are described in Section 4. Because selection is so important, Section 5 is dedicated to the method used to address it. Section 6 presents the main productivity results. A discussion of the broad implications of the results closes the paper in Section 7.

2. Defining productivity

For the purposes of this study, the productivity of an institution of higher education is the value to society of its causal effect on outcomes ("value-added") divided by the cost to society of educating its students ("social investment"). This productivity is what a policy maker or higher education leader would need in order to make any returns-on-investment argument (see below).[1]

The main causal effects of a postsecondary institution are likely to be on its own students' private outcomes, such as earnings. However, some effects, such as its students' contributions via public service, may not reflected in private outcomes. Also, some effects, like its students' contributions to innovation that raises economic growth, may spill over onto people who were not the students of the institution.

Social investment is the total cost of a student's education, not just costs funded by tuition. For instance, taxpayers fund some social investment through government appropriations and tax expenditures. Social investment is also funded by current and past donors to postsecondary institutions.

Thus, the productivity of an institution is not in general equal to the private return on private investment that an individual could expect if she were to enroll in the school. Such private calculations are interesting for individuals but less so for policy makers. We employ them in other related studies (for instance, studies of how students choose colleges), but they are not the object of interest here.

3. Why measuring productivity is useful

Higher education in the U.S. has survived and even thrived for many years without reasonably credible or comprehensive measures of institutions' productivity. Why, then, should we attempt to produce measures now? There are at least four reasons.

First, as government intervention in higher education has grown, it is reasonable for the public to ask for productivity

[1] For some economic applications, we might instead be interested in marginal productivity: the increase in value-added produced by a marginal increase in social investment. Although most of the analysis in this study applies equally to average and marginal productivity, I compute only average productivity because, to compute the marginal productivity, one would need a comprehensive set of policy experiments in which each school's spending was raised for an exogenous reason that was uncoordinated with other schools' spending. This is an extremely demanding requirement. We would not merely require "lightning to strike twice" in the same place. We would need it to strike many times.

measures. Most government interventions are based on returns-on-investment logic that requires the education to be productive. Policy makers, for instance, often argue that appropriations that support higher education institution pay for themselves by generating benefits that are more than equal to the social investment. They make a similar argument for tax credits, grants and scholarships, and subsidized student loans. Leaders of postsecondary institutions also make the returns-on-investment argument: donations to their school will more than repay themselves by delivering benefits to society.[2]

Second, the U.S. contains unusual (by international standards) and varied environments for higher education. We cannot know whether these environments promote a productive postsecondary sector if productivity is never measured. For instance, should institutions compete with one another for students and faculty or should these people be allocated through centralized rules as they are in many countries and a few U.S. public systems? What autonomy (in wage-setting, admissions, etc.) and governance structures (trustees, legislative budget approval, for instance) promote an institution's productivity? Does the information available to students when they are choosing schools affect the productivity of these schools? Is productivity affected by an institution's dependance on tuition-paying students versus students funded by grants or third-parties? While this study does not attempt to answer questions like those posed above, its does attempt to provide the dependent variable (productivity) needed for such analyses.[3]

Third, highly developed economies like that of the U.S. have comparative advantage in industries that are intensive in the advanced skills produced by postsecondary education. These industries tend to contribute disproportionately to such countries' economic growth and exports. Advanced-skill-intensive industries also have some appealing features such as paying high

---

[2] In a related paper, I show how the productivity estimates computed in this paper can be used to evaluate policies such as the tax deductibility of gifts to non-profit postsecondary institutions.

[3] Some of these questions are addressed in Hoxby (2016).

wages and being relatively non-polluting. However, economic logic indicates that a country cannot maintain comparative advantage in advanced-skill-intensive industries if it is not unusually productive in generating those skills. A country cannot maintain comparative advantage in equilibrium simply by generously funding a low productivity higher education sector.

Finally, once we have measures of institutions' productivity, we may be better able to understand how advanced human capital is produced. What the "education production function" is a long-standing and complex question. While productivity measures, by themselves, would not answer that question, it is hard to make progress on it in the absence of productivity measures. For instance, some results presented in this paper strongly suggest that the production function for selective higher education exhibits single-crossing: a higher aptitude student is likely to derive more value from attending an institution with a higher level of social investment than a lower aptitude student would derive. While many economists and higher education experts have long suspected the existence of single-crossing and assumed it in their analyses, evidence for or against single-crossing is scanty. If true, single-crossing has important implications, a point to which I return toward the end of the paper.

4. The key issues for measuring productivity

A. Selection

As previously stated, vertical selection (students who differ in aptitude enrolling at different institutions) and horizontal selection (similarly apt students who differ on grounds like geography enrolling at different institutions) are probably the most serious issues for measuring productivity. A naive comparison of, for instance, earnings differences between Harvard University's former students and a non-selective college's former students would be largely uninformative about Harvard's value-added. A naive comparison of earnings differences of community college students in San Francisco (where costs of living are high) and rural Mississippi (where costs of living are low) would also be largely uninformative about their relative value-added. Addressing selection is sufficiently challenging that I devote the next section entirely to it.

B. The attribution problem

The second problem for evaluating a postsecondary institution's productivity is attribution. Suppose that we have mastered selection and can credibly say that we are comparing outcomes and social investment of students at schools A and B who are as good as randomly assigned. Even under random assignment, we would have the issue that school A might induce students to enroll in more education (a graduate program at school C, for instance) than school B induces. When we eventually compare the lifetime outcomes of school A students to school B students, therefore, the A students will have more education and not just at school A. There is no way for us to identify the part of the A students' outcomes that they would have had if they had attended only school A for as long as they would have attended school B. Part of school A's causal effect on outcomes flows through its inducing students to attend school C.

Another example of the attribution problem arises because, even when pursuing a single degree, students may take classes at multiple institutions. For instance, part of the effect of a two-year college flows through its inducing students to transfer to four-year colleges. One-third of students transfer institutions at least once before receiving a bachelor's degree and nearly one-sixth transfer more than once.[4]

Consider two examples. Suppose that one two-year college tends to induce students to finish associates degrees that have a vocational or terminal (not leading to a four-year degree) character. Suppose that another two-year college tends to induce students to transfer to four-year college and earn their degrees there. If we were not to credit the second college with the further education (and outcomes and social investment) it induced, the second college would appear to be very unproductive compared to the first college. Much of its actual productivity would be

[4] This is a quotation from Staiger, in this volume, who is quoting from National Student Clearinghouse (2015).

attributed to the four-year colleges to which its students transferred. Consider Swarthmore College. It is a liberal arts college and, as such, does not train doctoral students. However, it tends to induce students to attend Ph.D. programs in academic subjects. If they go on to become leading researchers, then Swarthmore is productive in generating research and should be credited with this effect.

In short, part of the outcomes and thus the productivity of any school are due to the educational trajectory it induces. This attribution issue cannot be evaded: it is a reality with which we must deal. I would argue that the best approach is to assess the productivity of a school using lifetime outcomes as the numerator and all the social investments induced by it as the denominator.[5]

C. Multiple outcomes

Postsecondary institutions may causally affect many outcomes. To name but a few: earnings, public service, civic participation, research, innovation, cultural knowledge, tolerance and open-mindedness, marriage, child-rearing.[6] These outcomes

---

[5] In theory, one could identify the contribution of each institution to a person's lifetime outcomes. To see this, consider teacher value-added research in elementary and secondary education. Some researchers have been able to identify the effect on long-term outcomes of each teacher whom a student encounters in succession. Identification can occur if teacher successions overlap in a way that generates information about their individual contributions. What is ideal is for each possible pair of students to have some teachers in common and some not in common. By combining the results of all pairs of students, one can back out each teacher's contribution. The Carrell and Kurlaender study in this volume has some of this flavor. Identification works, in their case, because California students who attend community colleges tend have overlapping experiences as they move into the four-year California State Universities. However, identification is often impossible in higher education because students' experiences do not overlap in a manner that generates sufficient information. Postsecondary students are not channeled neatly through a series of grades: they can exit, get labor market experience between periods of enrollment, choose multiple degree paths, take courses in the summer at school A and then return each fall to school B, and so on. In other words, there are so many factors that can differ between pairs of students that identifying the contribution of each institution is very challenging outside of somewhat special cases like the California Community College example.

[6] All of the papers in this volume deal with the multiple outcomes problem, but see especially those by Staiger; Minaya and Scott-Clayton; and Riehl, Saavedra, and Urquiola.

may affect individuals' utility, sometimes in ways that would be invisible to the econometrician. Also, some of these outcomes may have spillover effects or general equilibrium effects. For instance, higher education might generate civic participation and societies with greater civic participation might have institutions that are less corrupt and less corrupt institutions might support a better climate for business. Then higher education might affect the economy through this indirect channel. Researchers encounter severe empirical challenges when trying to evaluate such spillover and general equilibrium effects.

Even if we could accurately observe every outcome, there is no correct or scientific way to sum them into a single index that could be used as the universal numerator of productivity. Summing across multiple outcomes is an inherently value-laden exercise in which preferences and subjective judgements matter.[7] It is fundamentally misguided to attach a weight to each outcome, compute some weighted average, and thereafter neglect the underlying, multiple outcomes. To make matters worse, the choice of weights in such exercises is not merely arbitrary but sometimes designed to serve the ends of some interest group.

I would argue that researchers ought to make available credible estimates of all the outcomes for there appears to be a demand and for which reliable measures can be constructed. This would at least allow an individual student or policy-maker to evaluate each postsecondary institution on the grounds that matter to her. Accordingly, in this paper, I show evidence on multiple outcomes which—though far from comprehensive—are intended to represent the three basic types: private (earnings), social (public service), and spillover-inducing through non-social means (innovation).

This being said, lifetime earnings have a certain priority as an outcome because they determine whether social investments in higher education are sufficiently productive to generate societal earnings that can support social investments in higher education

[7] Staiger (in this volume) makes the same point, referring to multiple outcomes affected by health care providers.

for the next generation of students. However, even if we accept the priority of earnings, we are left with the problem that many of the outcomes listed above affect societal earnings, but do so in such an indirect way that the earnings could not plausibly be connected to the institution that produced them. For instance, consider the problem of attributing to specific schools the societal earnings that arise through the indirect civic participation channel described above.

How bad is it to focus on the private earnings of a school's own students as the basis of productivity measures? The answer depends, to some extent, on how the measures are used. If the data are used to evaluate individual institutions, the focus is problematic. Certain institutions are at an obvious disadvantage because they disproportionately produce outcomes whose effects on society run disproportionately through externalities or general equilibrium channels: seminaries, womens' colleges, schools that induce students to become future researchers, and so on. However, if we are looking not at individual schools but at more aggregate statistics (schools grouped by selectivity, for instance), these concerns are somewhat reduced. For instance, within the group of very selective schools, some may be more research-oriented, others more public service-oriented, and yet others more business-oriented.

5. Selection

For measuring productivity, the problems associated with selection are the "elephant in the room." Vertical selection occurs when students whose ability, preparation, and motivation is stronger enroll in different colleges than students whose ability, preparation, and motivation is weaker. If not addressed, vertical selection will cause us to overestimate the value-added of colleges whose students are positively selected and to underestimate the value-added of colleges whose students are negatively selected. This leads to the legitimate question that plagues college comparisons: Are the outcomes of students from very selective colleges strong because the colleges add value or because their students are so able that they would attain strong outcomes regardless of the college they attended?

However, colleges' student bodies are not only vertically differentiated: they are also horizontally differentiated. That is, they differ on dimensions like geography and curricular emphasis. For instance, suppose that earnings differ across areas of the country owing, in part, to differences in the cost of living. Then, two colleges that enroll equally able students and generate equal value-added may have alumni with different earnings. We could easily mistake such earning differences for differences in value-added. As another example, consider two colleges that are equally selective but whose students who, despite having the same test scores and grades, differ in preferring, on the one hand, an life rich in outdoor activities and, on the other hand, a life rich in cultural events (concerts, theater, etc.). These incoming differences in preferences are likely to play out in later earnings regardless of what the colleges do, and we do not wish to mistake differences in preferences for differences in value-added.

Vertical selection is probably the more serious problem because social investment at the most and least selective colleges differs by about an order of magnitude. Also, some non-selective colleges' median student has a level of achievement that is similar to that attained by the 8th grade (or even earlier) by the median student at the most selective colleges.[8] One cannot give the most selective colleges credit for the four or so additional years of education that their incoming students have. Nor can one give them credit for the ability that allowed their students to acquire learning more readily than others. (Ability earns its own return, human capital apart.)

Solving selection problems is all about (i) randomization or something that mimics it, and (ii) overlap or "common support." Randomization solves selection problems because, with a sufficient number of people randomized into each treatment, the law of large numbers ensures that they are similar on unobservable characteristics as well as observable, measurable ones for which we might control.

---

[8] Author's calculations based on the National Educational Longitudinal Study, U.S. Department of Education (2000).

The point about common support is less obvious and is especially important for selection problems in higher education.

A. Common support

The requirement of common support means that it is highly implausible if not impossible to use comparisons between the outcomes of Harvard University students and students who would be extremely atypical at Harvard to judge Harvard's productivity. We need students who overlap or who are in the common support between Harvard and another institution. There are many such students but most end up attending another very selective institution, not a non-selective institution or a modestly selective institution. The common support requirement also exists horizontally. Two geographically proximate institutions that are similarly selective are far more likely to have common support than ones located thousands of miles away. Similarly, two similarly selective institutions that have the same curricular emphases (engineering or music) or campuses with similar amenities (opportunities for hiking versus opportunities for concert-going) are more likely to have common support.

We can analyze productivity while never moving outside the common support. In fact, the problem is almost exactly the same, as a statistical matter, as rating—say—tennis players. The top tennis players in the world rarely play matches against players who are much lower rated: the vertical problem. Also, apart from the top players whose matches are international, most players play most of their matches against other players from the same region: the horizontal problem. In tennis (as in many other sports that require ratings), the problem is solved by statistical Paired Comparison Methods (PCMs) that rely entirely on players who actually play one another (that is, players in the common support).

Sticking to the tennis analogy, the rating of a top player is built up by seeing how his outcomes compare to those of other fairly top players whom he plays often. Then their outcomes are compared to those of other slightly less apt players with whom they play often. And step by step, the distance in outcomes between most and least apt players is computed even if the most apt never play the least apt ones. Similarly, the rating of players who are geographically distant is built. A Portugese player might routinely play Spanish players who routinely play those from Southwest France, who play against Parisians, who play against Belgians, who play against Germans, who play against Danes. What is key is that PCMs never employ mere speculation of how one player would play someone whom in fact he never plays. Also PCMs are designed to incorporate the information generated when a lower rated player occasionally beats a much higher rated one. PCMs do not impose any functional form on the rating. There can be abrupt discontinuities: instance, the distance between players 2 and 3 could be small but the distance between players 4 and 5 could be very large. There can be ties.

If we compare the outcomes of and social investment in people who could attend either institution A or institution B but in fact divide fairly randomly between them, then we can apply PCMs to measure the relative productivity of the two schools. We derive the same benefits from common support: the measure of productivity is never based on mere speculation of how outcomes would compare among students who differ in aptitude or in the colleges they consider. An institution that is less productive on average is allowed to be more productive for some students. (This is the equivalent of the less apt player beating the more apt player sometimes.) There is no functional form imposed on institutions' productivity: institutions can be tied, very close, or very far in productivity. Interestingly, PCMs are much easier to apply to the productivity problem than to sports problems because outcomes like earnings are far more continuous than the score of a tennis match or other game. Also, small score differences that result in a win versus a loss matter in sports but not in the outcomes that matter for schools' productivity. No one would care if one school's students earned $50,000 on average and another school's earned $50,001.

B. Quasi-randomization

Applying PCMs to measuring productivity is straightforward if we can identify students whose attendance at any given pair of institutions is quasi-random. I do this with two procedures which correspond, respectively, to the vertical and horizontal selection problems.

The procedure for the horizontal problem is simpler. In it, we identify pairs of students who have equal observable application credentials, who apply to the same schools A and B that are equally selective, and who have a high probability of admission at both schools A and B. For instance, one might think of students who choose between equally selective branches of a state's university system. If the students knew for certain which school had the higher value-added for them, they might always choose it. But, in fact, they have an imperfect understanding and still must make a choice. Thus, horizontal college choices are often influenced by small factors that only affect the students' lifetime outcomes through the channel of which college they attend. While few students actually flip a coin, they choose among horizontally equal colleges based on the architecture, the weather on the day they visited, the off-hand suggestion of an acquaintance, and so on. This is quasi-randomization.

Once we have identified students who choose quasi-randomly among horizontally equal and proximate schools A and B, we can identify students who choose among horizontally equal and proximate B and C, C and D, etc. Thus, we can derive a measure of school's A productivity relative to D's even if students do not (or rarely) choose between A and D. The horizontal selection problem is solved. While geography is the most obvious source of differentiation among horizontal equals, the logic applies far more broadly. For instance, A and B might both have strong engineering programs, B and C might both have strong natural sciences programs, C and D might both be strong in the biological sciences, and so on.

Now consider the procedure for the vertical selection problem. Here, we identify pairs of students who have equal observable application credentials, who apply to the same schools A and B that are not necessarily equally selective, and who are "on-the-bubble" for admission at school A. I define students as being on-the-bubble if admissions staff are essentially flipping coins among them when making admissions decisions.

That is the definition, but why does this range exist and how can one learn where it is? A typical procedure for selective colleges is to group applicants, after an initial evaluation, into fairly obvious admits, fairly obvious rejects, and students who are on-the-bubble because they would be perfectly acceptable admits but are not obvious. The on-the-bubble group might contain two or three times as many students as the school has room to admit once the obvious admits are accounted for. (For instance, a school that plans to admit 1,000 students might have 800 obvious admits, and put 400 in the on-the-bubble group in order to admit 200 more.) The staff then look at the composition of the students whom they intend to admit and note deficiencies in the overall class composition. For instance, the prospective class may be missing students from some geographical area or with some curricular interest. Then, the staff conduct a final re-evaluation of the on-the-bubble students, keeping themselves attuned to these issues. Thus, an on-the-bubble student may be more likely to be admitted if she comes from a geographical area or plans to major in a field that was initially underrepresented. In another year, these same characteristics would not increase her probability of admission. Thus, admissions officers make decisions that, while not random, are arbitrary in the sense that they only make sense in the context of that particular school in that year.

How does one find the on-the-bubble range? It is the range where, as a statistical matter, there is a structural break in the relationship between the probability of admission and observable credentials. To clarify, the probability of admission above the bubble range is fairly high and fairly predictable. It increases smoothly and predictably in observable credentials such as test scores and grades. The probability of admission below the bubble range is low but also increases smoothly and predictably in observable credentials. (Student below the bubble range who gain admission usually have, in addition to their academic qualifications, some other observable characteristic such as athletic prowess.) In contrast, the probability of admission in the bubble range is very difficult to predict using observable credentials. This is because the on-the-bubble students all have perfectly acceptable credentials and the admissions decision, which occurs in the final re-evaluation, depends not on these credentials but on some characteristic that, in another year or similar school, would not predict a more favorable outcome.

Statistical methods that uncover structural breaks in a relationship are made precisely for situations like this: a relationship is smooth and predictable in range A; there is another range B in which the relationship is also smooth and predictable. Between ranges A and B, the relationship changes suddenly and cannot be predicted using data from either the A or B range. This is an issue into which I go into more detail in the companion methodological study, Hoxby (2015). The point is, however, that structural break methods are a statistical, objective way to find the on-the-bubble range. While structural break methods will find a strict credentials cut-off if one exists (for instance, if a school admits students who score above some threshold and rejects those who score below it), the methods will also find the on-the-bubble range for schools that practice more holistic admissions. It is worth noting that the on-the-bubble range does not typically coincide with the admits who have the lowest academic credentials. Rather, it contains students whose credentials usually place them only modestly below the median enrollee.[9]

Once one has located each school's on-the-bubble range, one can solve the vertical selection problem using chains of schools. One can compare schools A and B by comparing the outcomes of students who were on-the-bubble at school A. Some of them end up at school A; others end up at school B. Schools B and C may be compared using students on-the-bubble at school B. C and D may be compared using students on-the-bubble at school C. And so on. Thus, school A ends up being compared to school D through these connections even if few on-the-bubble students at A actually attend school D.

Summarizing, I identify "horizontal experiments" among students who have equal admissions credentials and who apply to the same equally selective schools where they are obvious admits. They more or less flip coins among the colleges. I

[9] Note that students who have minimal academic credentials but some offsetting observable characteristic such as athletic prowess are not on-the-bubble. Their admission is predictable.

identify "vertical experiments" among students who are on-the-bubble at some college and who are therefore admitted based on the equivalent of coin flips by the admission staff. I combine all of these experiments using PCMs. This procedure generates a value-added and a social cost of investment for each institution and, notably, these measures comply to the maximum extent possible with the requirements of randomization and common support. More detail on the procedure can be found in the companion paper.

In this paper, I show the results of applying this procedure to undergraduate students. It could also be applied to graduate and professional schools where test scores (LSAT, GMAT, etc.) and undergraduate grades dominate an admissions process that is run by staff.[10]

6. Data and Selectivity

A. Data

I use administrative data on college assessment scores, score sending, postsecondary enrollment, and 2014 earnings from wages and salaries for people in the high school graduating classes of 1999 through 2003 who were aged 29 through 34 in 2014. That is, I employ data on students who graduated from high school at age 18 or 19, which are the dominant ages at high school graduation in the U.S. Earnings are from de-identified Form W-2 data, and these data are available for non-filers as well as tax filers. A student with no W-2 is assumed to have zero wage and salary earnings. Enrollment data come not from students' self-reports from institutions reporting to the National Student Clearinghouse and through Form 1098-T. Further details on this part of the dataset are in Hoxby (2015).

The data on social investment come from the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS), a source derived from institutions' official reports. For the purposes of this study, social investment is equal to the

[10] It would work less well for small doctoral programs where faculty meet with or read considerable material from the students with whom they may choose to work and whose admission they greatly influence.

amount spent on an undergraduate's education. This is called "core" spending by the U.S. Department of Education and is equal to the per-pupil sum of spending on instruction, academic support[11] (for instance, advising), and student support.[12]

6. The Results

Because differences in vertical selection among institutions are a dominant feature of U.S. higher education and because these differences are the primary threat to accurate calculations of productivity, I present the results using figures in which institutions are classified by their selectivity. Specifically, each figure puts the data for an institution in a "bin" according to the empirical combined math and verbal SAT (or translated ACT) score of its average student.[13] Note that it is institutions, rather than individual students, that are binned since we are interested in showing the productivity of institutions.

Although score-based bins are the probably the most objective way to organize the institutions by selectivity, it may help to provide an informal translation between the scores and the "competitiveness" language used in Barron's Profiles of American Colleges and Universities, familiar to higher education researchers and policy makers. Roughly, institutions with an average combined score of 800 are non-competitive. Indeed, they often explicitly practice "open admission" which means that they admit anyone with a high school diploma or passing score on a

[11] Academic support includes expenses for activities that support instruction. For instance, it includes libraries, audiovisual services, and academic administration. The source is National Center for Education Statistics (2016).

[12] Student support includes expenses for admissions, registrar activities, and activities whose primary purpose is to contribute to students' emotional and physical well-being and to their intellectual, cultural, and social development outside the context of the formal instructional program. Examples include student activities, cultural events, student newspapers, intramural athletics, student organizations, supplemental instruction outside the normal administration, and student records. The source is U.S. Department of Education (2016).

[13] The empirical average score is not necessarily the same as the SAT/ACT score that appears in college guides. Some schools submit scores to the college guided that reflect "management" of the (sub-population of) students for whom scores are reported.

high school equivalency test. Institutions with an average combined score of 1000 to 1050 are "competitive plus," 1050 to 1150 are "very competitive," 1150 to 1250 are "very competitive plus," 1250 to 1350 are "highly competitive" or "highly competitive plus," and 1350 and over are "most competitive." These classifications are approximate and some schools do not fit them well. There is an indeterminate area between non-selective and selective schools which corresponds roughly to the 800 to 1000 range. Toward the top of this range, schools tend to be selective but more reliant on high school recommendations and grades and less reliant on test scores. Toward the bottom of this range, schools tend to be non-selective. However, schools in this range can be hard to classify because information about them is often only partial. This is a point to which we return.

I show productivity for three key outcomes: wage and salary earnings (including zero earnings for people who have none), a measure of public service, and a measure of innovation produced. The construction of each measure is discussed more below. All of these are "lifetime" measures in which I compute the actual measure for ages 18 through 34 and then project the outcome for ages 35 through 65, the ages for which persons' outcomes cannot be linked to their postsecondary institutions. I show the present discounted value of earnings and public service using a real discount rate of 2.5 percent. Although the magnitudes of productivity vary with the discount rate selected, the basic pattern of results does not. Furthermore, although the magnitudes vary with the methods of projecting after age 34, the basic pattern of results does not. There is a simple reason for both forms of robustness: the pattern of results is already evident by age 34 after which earnings, public service, and research grow in a fairly stable way.[14]

I do not believe that there is a method of accounting for selection between no and some postsecondary education that is both credible and broadly applicable. That is, there are methods that credibly account for this selection-at-the-margin for a

[14] See below for more on how earnings are projected after age 34.

particular institution or set of students.[15] However, a method that works fairly ubiquitously does not exist for the simple reason that the decision to attend postsecondary school at all is not a decision that most people make lightly or quasi-randomly. It is a fairly momentous decision. Thus, it does not lend itself to selection control methods which require quasi-randomization and common support.

Because I do not have a method of accounting for selection between no and some postsecondary education, I normalize the productivity of the least selective institutions to zero. This does not mean that their productivity is actually zero. In consequence, readers should only interpret differences in institutions' productivity.

This normalization is more of an issue for some productivity measures than others. At one extreme, it seems fairly innocuous for research productivity because so few students from the normalized-to-zero institutions are ever employed in research-related jobs. Moreover, research is very disproportionately produced by people with at least some postsecondary education. Thus, one should not be concerned that, by enrolling in postsecondary education, many people who would otherwise have been employed in research-related jobs (with only a secondary education) are not doing so. At the other extreme, one wishes that one did not have to make a normalization for productivity measures based on earnings. Just because an institution belongs in the lowest selectivity bin does not mean that its value-added is zero. It might raise earnings substantially relative to no postsecondary education at all. But, it may also have productivity that near zero. If one takes opportunity costs into account (as one implicitly does when making comparisons among schools), a postsecondary institution's productivity can even be negative. For instance, if attending an institution is costly but keeps the student away from employment at which he would gained valuable skills and experience, the institution could easily have negative value-added relative to no postsecondary education

[15] See the recent review by Oreopoulos and Petronijevic (2013).

at all.

In short, we caution readers against interpreting the zeros for the lowest selectivity institutions.

A. Productivity measures based on earnings

Earnings through age 34 are simply taken from data that links outcomes to postsecondary institutions. But earnings must be projected to from age 34 to 65 because there are no data that provide similarly high quality links between postsecondary institutions and people of these ages.

To project earnings, we use empirical earnings dynamics. Specifically, we categorize each 34-year-old by his percentile within the income distribution for 34-year-olds. Then we compute a transition matrix between 34-year-olds' and 35-year-olds' income percentiles. For instance, a 34-year-old with income in the 75th percentile might have a 10 percent probability of moving to the 76th percentile when aged 35. We repeat this exercise for which subsequent pair of ages: 35 and 36, 36 and 37, and so on. In this way, we build up all probable income paths, always using observed longitudinal transitions that differ by age. (When a person is younger, she has a higher probability of transitioning to a percentile far from her current one. Incomes stabilize with age so off-diagonal transition probabilities fall.)[16]

Figure 1 shows life time wage and salary earnings and value-added for institutions of higher education. The earnings are "raw" because no attempt has been made to account for the effects of selection. Value-added, in contrast, is computed using

---

[16] I investigated two alternatives to this procedure. The first was to use empirical earnings paths that played out for the same person over a decade. Thus, one could use the longitudinal pattern for each 34-year-old, following him through age 43. This alternative has the advantage of allowing longer forms of serial correlation. However, it has the disadvantage that one is forced to use data from calendar years that are further away from the present. This alternative produced results similar to those shown. The second alternative was to keep a person at the same percentile in the earnings distribution as he was at age 34. For instance, a person at the 99th percentile at age 34 would be assigned 99th percentile earnings for all subsequent ages. This alternative produces outcomes similar to those shown for middling percentiles but not for low and high percentiles. Because this alternative does not allow for a realistic degree of reversion toward the mean, it produces some lifetime earnings that are extremely low and high compared to reality.

the method described above to account for selection.

The figure shows that both raw earnings and value-added are higher for institutions that are more selective. Indeed, both series rise almost monotonically. However, value-added rises more slowly than earnings. This is particularly obvious as we reach the most selective institutions where the slope of the relationship implies that about two-thirds of the earnings gains do not represent value-added but instead represent what their very apt students would have earned if they had attended less selective schools. (Because of the normalization, only the gain in value-added relative to the lowest selectivity schools is meaningful. The level is not.)

Of course, this does not mean that the more selective an institution is, the greater its productivity. Value-added rises with selectivity but, as Figure 2 demonstrates, so does social investment in each student's education. Recall that social investment is the increase in educational spending triggered by attending one institution rather than another. Like value-added, this measures accounts for selection using the method described above. Also like value-added, it is a lifetime measure and discounted using a real rate of 2.5 percent. However, I do not attempt to project it after age 34 because the vast majority of people have completed their postsecondary education by age 34.[17]

Just for comparison, Figure 2 also shows the present discounted value of tuition paid. This is always lower than social investment because it does not include spending that is funded by taxpayers, donors, and so on.

Social investment in each student's education is higher for institutions that are more selective. It rises almost monotonically with the institution's average test score. Note also that social investment rises notably more steeply than tuition paid. This is partly because more selective institutions spend considerably more per curricular unit on each student's education. But, it is

[17] More precisely, most people have, by age 34, completed the postsecondary education that is induced by their initial enrollment. If people return to college after, say, a decade in the labor market, that second enrollment episode is likely triggered by a labor market experience and should be evaluated separately.

also partly because students who attend more selective institutions tend to enroll in more curricular units. They are less likely to drop out, more likely to attend full-time, more likely to continue onto graduate school, and so on. This is true even when we have accounted for selection.

Figures 1 and 2 suggest that the pattern of institutions' productivity across less and more selective institutions will be something of a race between value-added (the numerator), which is rising, and social investment (the denominator), which is also rising. Figure 3 shows the results of this "race." The pattern is striking: productivity is rather flat across a wide range of selectivity except that productivity is distinctly lower for institutions that are non-selective. That is, the main take-aways are that (i) within the selective institutions, productivity does not rise or fall dramatically by selectivity, (ii) within the non-selective institutions, productivity does not rise or fall dramatically by selectivity, (iii) the non-selective schools are less productive than the selective ones.

The first of these results—that productivity is rather flat in selectivity among selective schools—is very striking and has potentially important implications. It is striking because social investment and earnings both rise substantially as selectivity rises. Thus, the flatness comes from the numerator and denominator rising at a sufficiently similar rate that the value-added for a dollar of spending is not terribly different at a institution with an average score of 1000 and one with an average score of 1400. This means that spending is scaling up with student aptitude such that higher aptitude students get an increase in resources that is commensurate with their capacity to use them to create value. This seems unlikely to be a pure coincidence. Since students are actively choosing among the institutions throughout this range, the flatness may be the result of market forces: students choosing among schools and schools consequently competing for faculty and other resources. In other words, students who can benefit from greater resources may be willing to pay more for them, inducing an allocation of schools' resources that corresponds roughly to students' ability to benefit from them.

That market forces would have this effect would not be surprising if all students paid for their own education, the financing of such education was efficient, and students were well-informed about the value they could expect to derive from educational resources. But, clearly, these idealized conditions do not obtain: third party payers (taxpayers, donors) are the proximate funders of a considerable share of selective higher education, student loan volumes and interest rates are such that students can be liquidity constrained (on the one hand) or offered unduly generous terms (on the other); many students appear to be poorly informed when they choose a postsecondary school. Thus, what the result suggests is that, even with all these issues, market forces are sufficiently strong to maintain some regularity in how institutions' resources scale up with the aptitude of their students.

The empirical result does not imply that the educational resources are provided efficiently. It could be all the institutions provide resources in a similarly inefficient manner. However, unless the productivity of least productive institutions is substantially negative (so that the normalization-to-zero overstates their productivity a lot), a dollar spent on educational resources at a selective institution appears to generate multiple dollars of value over a person's lifetime.

The second of these results--that productivity is rather flat in selectivity among non-selective schools--is not terribly surprising. More specifically, among institutions whose average student has a combined score of 800 or below, productivity is rather flat. This may because the institutions do not actually differ much in student aptitude: Their average student's score may not be terribly meaningful because some of their students take no college assessment or take an assessment but only for low stakes.[18] Or, it could well be that the aptitudes that may matter for their students' success are poorly measured by tests. Finally,

[18] Many American students take a college assessment (or preliminary college assessment) solely to satisfy their state's accountability rules or for diagnosis/placement. Thus, many students who do not apply to any selective postsecondary school nevertheless have scores.

being non-selective, these institutions may differ mainly on horizontal grounds (geography, curriculum, how learning is organized) so that showing them vis-a-vis an axis based on the average student's score is just less informative.

The third of these results--that productivity is distinctly lower at non-selective institutions--is interesting and consistent with several possible explanations. First, non-selective institutions enroll students who have struggled in secondary school, and it may simply to harder to turn a dollar of investment into human capital for them. Simply put, they may arrive with learning deficits or study habits that make them harder to teach. Second, many students who enroll in non-selective schools do not choose among them actively or in an informed manner. They simply choose the most proximate or one that becomes salient to them for an arbitrary reason (an advertisement, for instance). Because these schools infrequently participate in national college guides, students may have a difficult time comparing them on objective grounds. For all these reasons, market forces may fail to discipline these institutions' productivity. Third, non-selective institutions disproportionately enroll students who do not pay for their own education but, instead, have it funded by a government grant, veterans' benefits, or the like. As in other third-party-payer situations, this may make the students less sensitive to the commensurability between cost and benefit than they would be if they were paying the bills themselves.

The patterns discussed so far are robust to several alternatives in computing productivity such as using discount rates anywhere within the plausible range of 2 to 3 percent real. They are robust to removing institutional support from social investment. (Social investment should certainly include instructional spending, academic support, and student support.) They are robust to excluding Extensive Research Universities whose accounting of how spending is allocated across undergraduates and other uses is most contestable.[19] All of these

[19] The 2000 edition of the Carnegie Commission on Higher Education classified postsecondary schools as Extensive Research Universities if they not only offer a full range of baccalaureate programs but also (i) are committed to graduate education

alternatives change the magnitudes of productivity but they do not change the three key patterns just discussed. They also do not change the fact that both earnings and social investment rise fairly monotonically in selectivity.

Among institutions of similar selectivity, is productivity similar? In other words, is the average productivity within each bin representative of all institutions or does it represent an average among schools whose productivity differs widely? This question is clearly important for interpretation, and Figure 4 shows the answer.

Figure 4 shows not just the average productivity in each selectivity bin but also shows the productivity of the 5th and 95th percentile institutions with each bin. It is immediately obvious that productivity differences among schools are wide among non-selective institutions but narrow as schools become more selective. Indeed, among the very selective schools, productivity differences are relatively small.

Given the results on the average levels of productivity, these results on the dispersion of productivity should not be too surprising. The level results suggested that market forces might be operative among selective institutions. The students who would likely maintain the most market pressure would be students who make active choices among schools (not merely choosing the most proximate), who are best informed, whose families pay for some or most of their education, and who are the least likely to be liquidity constrained. Such students will disproportionately be apt. Thus, the more selective an institution is, the more it is probably exposed to market forces that discipline its productivity—explaining why we see low dispersion.

If market forces weaken as students get less apt then the pressure for similarly selective schools to be similarly productive would fall as selectivity falls. This would be consistent with the pattern of dispersion in Figure 4. Market pressure might be very weak for non-selectives if the students who tend to enroll in th

---

through the doctorate, give high priority to research, award 50 or more doctoral degrees each year, and annually receive tens of millions of dollars in federal research funding.

em choose only among local schools, are poorly informed, and have their tuition paid by third parties. Indeed, for many non-selective schools, there is not much information available about students' outcomes. Thus, we should not be surprised that low productivity non-selective schools do not get eliminated even though some non-selective schools have much higher productivity.

B. Productivity measures based on public service

Conceptually, one wants to have a measure of public service that picks up contributions to society that earnings do not. This suggests that a good measure of public service is the percentage difference in earnings in a person's occupation if he works in the public or non-profit sectors versus the for-profit sector. This is a measure of his "donation" to society: the earnings he foregoes by not being in the for-profit sector. Two concrete examples may be helpful. Highly able lawyers usually work for for-profit law firms, but some work as judicial clerks, district attorneys, and public defenders. The latter people earn considerably less than they would in the for-profit sector. Similarly, executives and managers of non-profit organizations, such as foundations, usually earn considerably less than those in the for-profit sector. While a measure of public service based on "pay foregone" is certainly imperfect (in particular, the different sectors may draw people who have different levels of unobserved ability), it is at least an economics-based measure, not an ad hoc measure. It is also a continuous measure and a measure that can be specific to the schools in each selectivity bin, limiting the unobserved ability problems just mentioned.

I classify each school's former students by their 1-digit occupation at about age 34. Then, I compute, for each selectivity bin, the average earnings by occupation for those employed in the for-profit sector. Next, I compute each public or non-profit employee's contribution to public service as the difference between his occupation-by-selectivity bin's for-profit average earnings and his earnings. To make this akin to a lifetime measure, I multiply it by the person's ratio of projected lifetime earnings to his age 34 earnings. (The last is simply to make magnitudes analogous to those in the previous subsection.) Also, if the contribution calculated is negative, I set it to zero. (I return to this point below.) I assume that the contribution to public service is zero for for-profit employees. Clearly, they may make contributions through volunteering or other means, but most such contributions pale in comparison to those of someone who foregoes 15 percent of pay, say.

Once this contribution to public service is computed, it can be used to make productivity calculations in a manner that is exactly analogous to how earnings are used in the productivity calculations based on earnings. To be precise, productivity based on public service is value-added through public service contributions divided by social investment.

Figure 5 shows the results of this exercise. The relationship is fairly noisy and non-monotonic although, overall, productivity based on public service rises with selectivity. The bumpy relationship is the net result of two competing relationships. The percentage of former students who take up government employment falls as selectivity rises. This would tend to make public service productivity fall as selectivity rises. However, this fall is offset by the rise in earnings foregone as selectivity rises. A concrete example may help. For most of the selectivity range (above the non-selectives), the tendency of former students to become public school teachers is falling with selectivity. However, in the lower selectivity bins, public school teachers are relatively well paid compared to for-profit employees in their occupational category so their foregone earnings are little to none. Relatively few former students from the highest selectivity bins become public school teachers but those that do forego a large share of their for-profit counterparts' earnings. Similar phenomena hold for other local, state, and federal employees.

Figure 6 shows the dispersion of productivity based on public service. The pattern shown contrasts strikingly with that of Figure 4 which showed that the dispersion of productivity based on earnings fell steady with selectivity. The dispersion of productivity based on public service does not. It is noisy, but it rises with selectivity. This indicates that, among very selective schools, some are much more productive in public service contributions than others. Put another way, some very selective schools are much more likely to induce their students to enter public service than are other very selective schools. One might speculate that some schools have more of a service ethos or greater service opportunities available to students on or near campus. In any case, there is little indication that market or any other forces constrain similarly selective schools to have similar productivity based on public service.

C. Productivity measures based on innovation

Conceptually, one wants to have a measure of contributions to innovation that is broader than, say, a measure based on patenting would be. Many more people contribute to innovation than own patents. Similarly, a measure based on former students themselves becoming researchers seems too narrow. The software industry, for instance, fits the definition of innovative that many economists have in mind, and it is certainly a growing industry in which the U.S. has comparative advantage.[20] Yet, it does not have many employees who would describe themselves as researchers. For these reasons, I computed a measure of contributions to innovation based on the research and development (R&D) spending of each person's employer. Specifically, I took each employer's ratio of R&D spending to total expenses. Non-profit and public employers, especially universities, were included as much as possible. I then multiplied each employee's earnings at age 34 by this R&D ratio. Finally, I multiplied by the person's ratio of lifetime earnings to her age 34 earnings. (This final multiplication is simply to make the magnitude analogous to those in the previous subsections.)

Thus, a person who works for a firm that spends 10 percent of its budget on R&D would have 10 percent of her lifetime pay listed as her contribution to innovation. Of course, this is not meant to be a measure of her direct contributions. Rather, it is a way of forming an index that both reflects value (from earnings) and innovation (from the R&D ratio). This index permits people

[20] See Hecker (2005).

to contribute to innovation even if they do so in a supportive capacity, as most do, rather than as an investigator or patentee. For instance, a secretary or market researcher for a software firm would be counted because she indirectly supports the innovation occurring there.

Once this contribution to innovation is computed, it can be used to make productivity calculations in a manner that is exactly analogous to how earnings are used in the productivity calculations based on earnings. To be precise, productivity based on innovation is value-added through the innovation measure divided by social investment.

Figure 8 shows the results of this exercise. The pattern is mildly upward-sloping in selectivity until the most selective institutions are reached. At that point, the relationship becomes steeply upward sloping. This convex relationship indicates that very selective institutions are much more productive in contributions to innovation than all other institutions. There are at least two possible explanations. Most obviously, there is no reason to think that the relationship should be flat as it is for productivity based on earnings. In the latter case, market forces could plausibly generate a flat relationship. But, if much of the return to innovation spills over onto others or works through general equilibrium effects, there is no obvious mechanism that would ensure that social investment scales up with contributions to innovation. Alternatively, social investment, the denominator of productivity, could be understated in the most selective schools. Social investment does not include these schools' spending on research, and this research spending may have benefits for undergraduates. In fact, the channel could be subtle. It may be that research spending has no direct benefits for undergraduates but that it attracts a different type of faculty (research-oriented) who, even when they teach undergraduates, teach in a manner that is oriented toward developing knowledge at the frontier. Thus, the undergraduate program might be almost unintentionally research-oriented.

I do not show dispersion in productivity based on research because the focus would be on such a small number of schools.

7. Discussion

At selective institutions, a dollar in social investment appears to generate multiple dollars of value-added based on earnings over a person's lifetime. This conclusion is only unwarranted if non-selective schools have substantially negative productivity.

This is a simple but important result with broad implications for many government policies. For instance, the estimated productivity of selective institutions appears to be sufficiently high to justify taxpayer support and philanthropic support that is incentivized by the tax deductibility of gifts. I lay out such calculations in a companion paper.

For non-selective schools, it is less clear whether a dollar in social investment generates at least a dollar in value-added based on earnings. This is not to say that these schools' productivity is near zero. Rather, it is say that understanding their productivity is difficult because their students tend to be at the no-enrollment versus some-nonselective-enrollment margin where it is extremely difficult to account for selection. For instance, this study does not attempt to say how the productivity of non-selective schools compares to the productivity of on-the-job training.

The results for productivity based on earnings suggest that market forces are sufficiently strong to maintain some regularity in how institutions' resources scale up with the ability of their students to convert social investment into value-added. Without market forces as the explanation, the stability of productivity over a wide range of schools would be too much of a coincidence. (This does not necessarily imply that selective institutions provide educational resources with maximum efficiency: market forces might only compel them to provide resources in a similar but inefficient manner.) However, selective schools' efficiency is at least such social investment channeled through them generates multiple dollars of value-added.

Productivity based on earnings is much more dispersed among non-selective and less-selective schools than among very selective schools. This is a hint that market forces weaken as selectivity falls, perhaps because students become less informed and/or less responsive to productivity when choosing which school to attend. In any case, a student choosing among non-selective schools can make a much larger "mistake" on productivity than a student choosing among very selective schools.

The results for productivity based on public service suggest that market forces do not maintain regularity in how institutions' resources scale up with the ability of their students to convert social investment into public service. A plausible explanation is the lack of market rewards for public service. Without such market-based rewards, there may be no mechanism by which schools that are that more productive at public service generate more funds to support additional investments.

The results for productivity based on innovation suggest that highly selective schools are much more productive than all other schools. This is not surprising if the rewards to innovation run largely through spillovers or general equilibrium effects on the economy. In such circumstances, there would be no market forces to align social investment with contributions from innovation. Alternatively, social investment (the denominator of productivity) could be understated because it is does not include spending on research. Undergraduates may learn to be innovative from research spending or simply by being taught by faculty who spend part of their time on research, supported by research spending.

The three outcomes by which productivity is measured in this paper were chosen to represent private returns (earnings), social returns (public service), and likely sources of economic spillovers (innovation). But, there are of course, there are many other outcomes by which productivity of postsecondary institutions could be measured.

8. References

Avery, Christopher, Mark Glickman, Caroline Hoxby, and Andrew Metrick. 2013. “A Revealed Preference Ranking of U.S. Colleges and Universities,” Quarterly Journal of Economics, 128 (1): 1-45.

Barrow, Lisa, and Ofer Malamud. forthcoming. "Is College a Worthwhile Investment?," Annual Review of Economics.

Bulman, George B. and Caroline M. Hoxby. 2015. "The Returns to the Federal Tax Credits for Higher Education." Tax Policy and the Economy 29: 1-69.

Carrell, Scott, and Michal Kurlaender. forthcoming. "Estimating the Productivity of Community Colleges in Paving the Road to Four-year College Success," in Productivity in Higher Education, Caroline Hoxby and Kevin Stange, eds.

Cohodes, Sarah, and Joshua Goodman. 2012. “First Degree Earns: The Impact of College Quality on College Completion Rates,” HKS Faculty Research Working Paper Series RWP12-033. http://web.hks.harvard.edu/publications/getFile.aspx?Id=836

Deming, David, Claudia Goldin, and Lawrence Katz. 2011. “The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?” NBER Working Paper 17710.

Duquette, Nicholas. 2016. "Do Tax Incentives Affect Charitable Contributions? Evidence from Public Charities' Reported Revenue," Journal of Public Economics, Volume 137: 51-69.

Goodman, Joshua, Michael Hurwitz, Jonathan Smith. 2015. "College Access, Initial College Choice and Degree Completion," NBER Working Paper 20996
http://www.nber.org/papers/w20996

Hastings, Justine, Christopher Neilson, and Seth Zimmerman. 2012. “Determinants of Causal Returns to Postsecondary Education in Chile: What’s Luck Got to Do With It?” NBER conference paper.

Hecker, Daniel. 2005. "High Technology Employment: A NAICS-Based Update," Monthly Labor Review, July: 57-72.

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," Review of Economics and Statistics, 91 (4): 717-724.

Hoxby, Caroline. 2009. "The Changing Selectivity of American Colleges," Journal of Economic Perspectives, 23(4): 95-118.

Hoxby, Caroline. 2015. "Estimating the Value-Added of U.S. Postsecondary Institutions. Internal Revenue Service Statistics of Income Division working paper.

Hoxby, Caroline. 2016. "The Dramatic Economics of the U.S. Market for Higher Education," The 8th Annual Martin Feldstein Lecture, National Bureau of Economic Research. The full lecture is available online at http://www.nber.org/feldstein_lecture_2016/feldsteinlecture_2016.html and is summarized in The NBER Reporter, 2016, no. 3.

Kaufmann, Katja Maria, Matthias Messner, and Alex Solis. 2012. "Returns to Elite Higher Education in the Marriage Market: Evidence from Chile," Bocconi University working paper. http://tinyurl.com/kaufmanncollrd

Langville, Amy N. and Carl D. Meyer. 2012. Who's #1? The Science of Rating and Ranking. Princeton: Princeton University Press.

National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education. Integrated Postsecondary Education Data System (data as of July 2015 nces.ed.gov website).

National Student Clearinghouse, 2015. "Signature Report: Transfer & Mobility: A National view of Student Movement in Postsecondary Institutions, Fall 2008 Cohort" downloaded from https://nscresearchcenter.org/wp-content/uploads/SignatureReport9.pdf

Oreopoulos, Philip and Uros Petronijevic. 2013. "Making college worth it: A review of research on the returns to higher education," in The Future of Children: Postsecondary Education in the United States, editors Lisa Barrow, Thomas Brock, Cecelia E. Rouse, 23(1): 41-65.

Saavedra, Juan Estaban. 2009. “The Learning and Early Labor Market Effects of College Quality: A Regression Discontinuity Analysis.” Rand Corporation
working paper. http://tinyurl.com/saavedracollrd-pdf

Staiger, Douglas. forthcoming. "What Healthcare Teaches Us About Measuring Productivity in Higher Education," in Productivity in Higher Education, Caroline Hoxby and Kevin Stange, eds.

Zimmerman, Seth D. 2014. "The Returns to College Admission for Academically Marginal Students," Journal of Labor Economics, 32(4): 711-754.

Figure 1
Postsecondary Institutions' Value-added based on Lifetime Wages
(institutions are binned by selectivity)

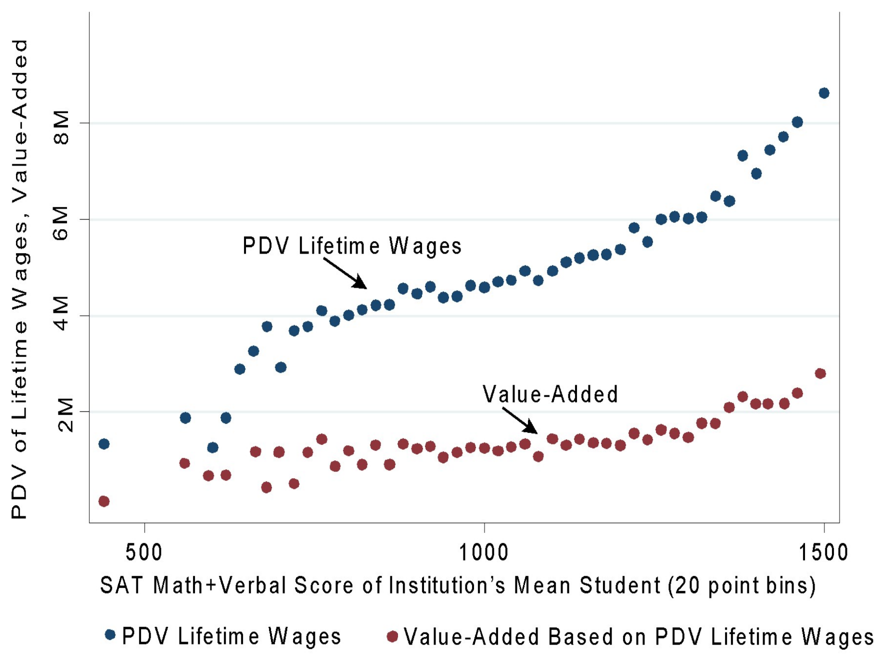

Figure 2
Social Investment and Tuition Paid
(postsecondary institutions binned by selectivity)

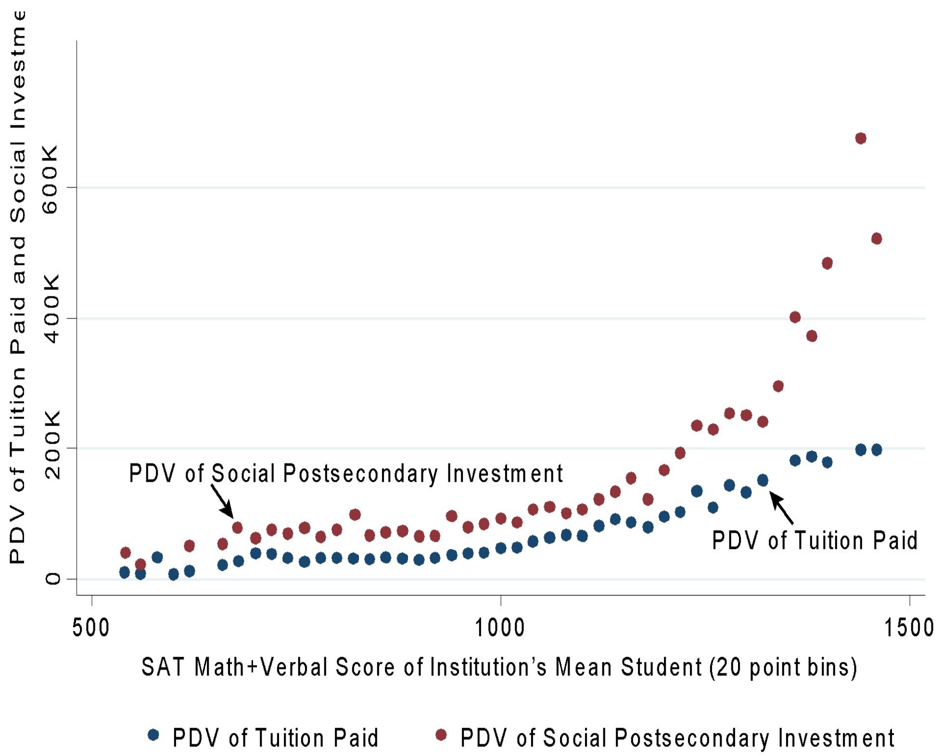

Figure 3
Productivity of Postsecondary Institutions (binned by selectivity)
productivity = (lifetime value-added)/(social investment)

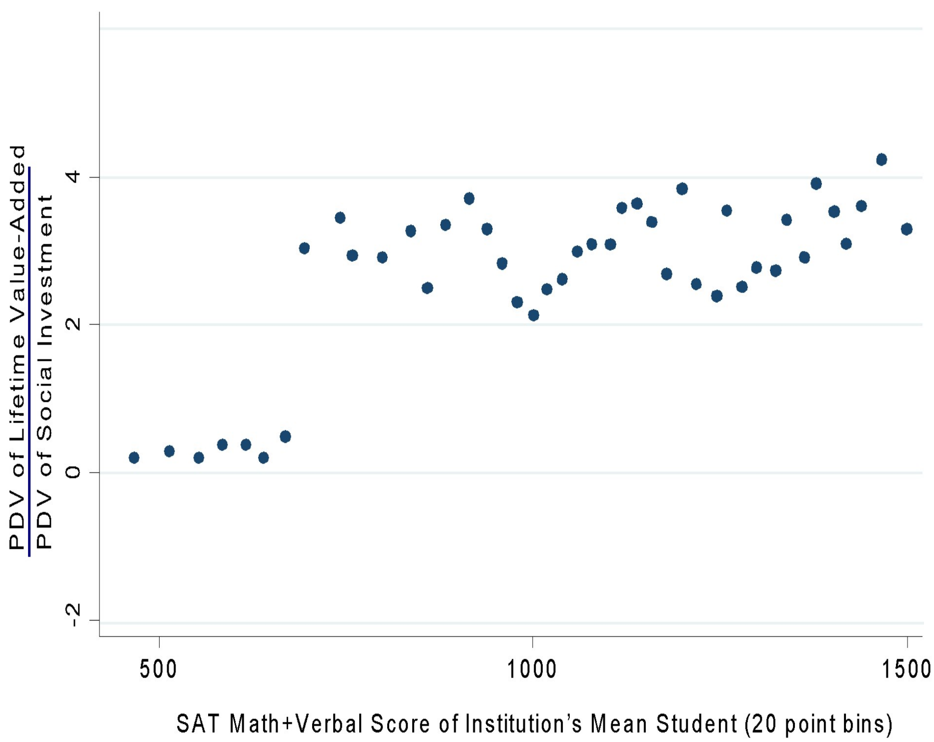

Figure 4
Dispersion of Productivity of Postsecondary Institutions (binned by selectivity)
in each bin, 95th %ile, mean, and 5th %ile productivity are shown

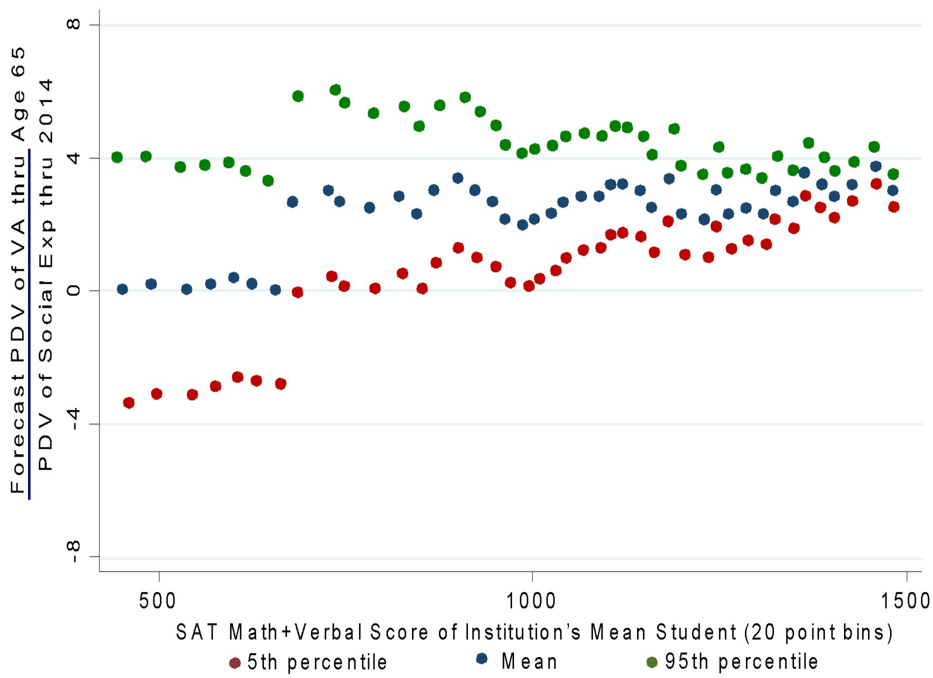

Figure 5
Public Service Productivity of Postsecondary Institutions
(institutions are binned by selectivity)

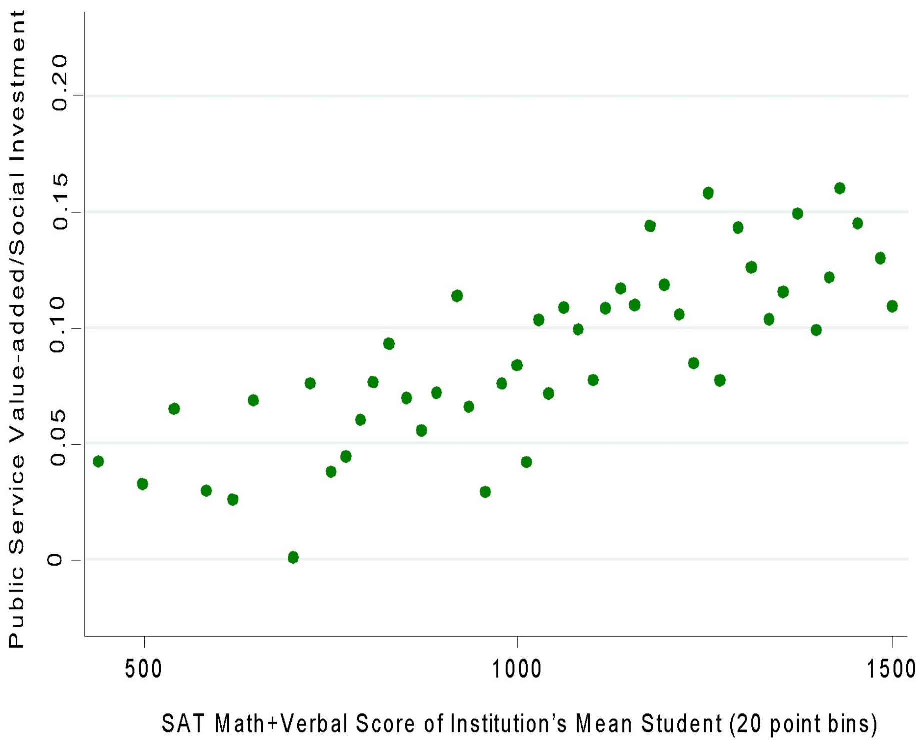

Figure 6
Dispersion of Public Service Productivity among Postsecondary Institutions
(institutions are binned by selectivity)

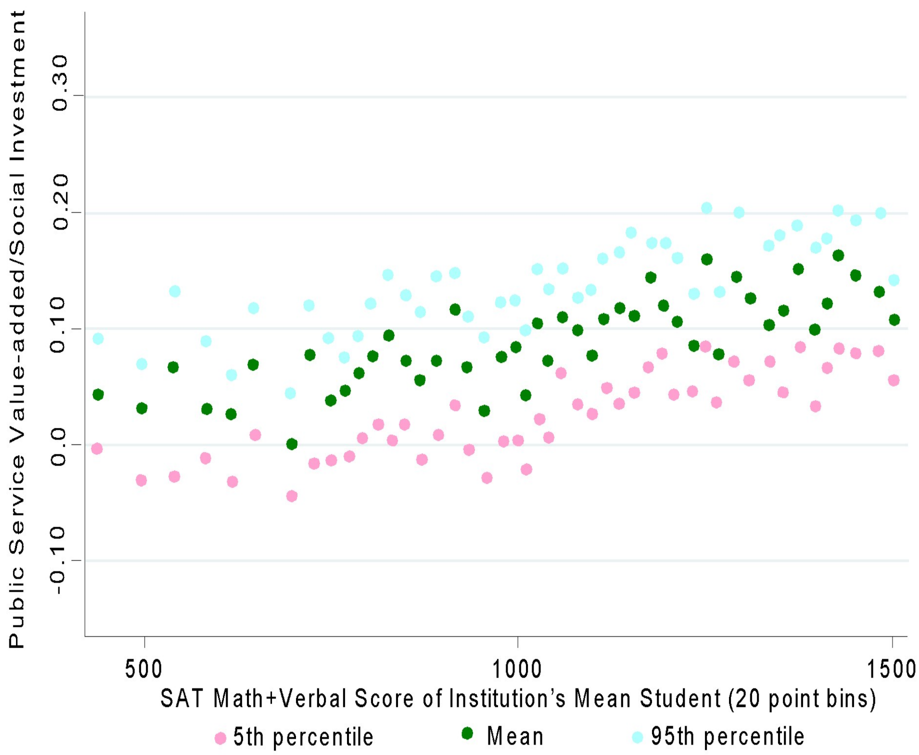

Figure 7
Innovation-based Productivity of Postsecondary Institutions
(institutions are binned by selectivity)