Exhibit C

REPLY EXPERT REPORT OF PETER S. ARCIDIACONO

Students for Fair Admissions, Inc. v. University of North Carolina
No. 14-cv-954 (M.D.N.C.)

TABLE OF CONTENTS

1 Executive Summary 1

1.1 Overview 1

1.2 Professor Hoxby's criticisms of my model are unfounded 3

1.3 Professor Hoxby's criticisms of my conclusion that race is a dominant factor in UNC admissions are also unfounded. 6

2 Professor Hoxby's criticisms of my model are incorrect or misleading 8

2.1 Professor Hoxby's claims that my models are overfit are demonstrably wrong; my models have much greater predictive power than her preferred model. 10

2.1.1 Overfitting generally 11

2.1.2 Professor Hoxby focuses only on a few years of admissions data, improperly disregarding half the data set produced by UNC. 14

2.1.3 Contrary to standard practice, Professor Hoxby's tests focus on the predictions from models estimated using only one year of data. 16

2.1.4 Professor Hoxby is wrong to focus on changes in mean-squared error in her overfit tests. 17

2.1.5 Even using Professor Hoxby's incorrect approach, my models are not overfit; in fact, my models predict better out of sample than Professor Hoxby's models predict in sample 21

2.2 Standard statistical methods show that my models are more accurate and reliable than Professor Hoxby's models 23

2.3 Professor Hoxby's other criticisms of my models are incorrect or misleading. 25

2.3.1 Professor Hoxby's arguments for excluding UNC's ratings are misleading or incomplete. 25

2.3.2 Including applicants in special recruiting categories does not change my findings. 28

2.3.3 The method I use to account for missing performance variables is appropriate. 33

2.3.4 Alternative ways of accounting for missing performance variables yield similar findings on the magnitude of racial preferences. 35

2.3.5 My use of the information on SAT scores is appropriate, and my results are not sensitive to this choice. 39

2.3.6 Using more finely detailed geographical information would put the model at risk of being overfit 40

2.3.7 Professor Hoxby's criticism of how I incorporate capacity constraints is wrong 41

3 UNC provides substantial racial preferences to African-American and Hispanic applicants. 46

3.1 Race is a determinative factor for many African-American and Hispanic admits. 47

3.2 Professor Hoxby's use of the Pseudo R-square (and corresponding Shapley decomposition) is inappropriate for measuring whether race is determinative in admissions. 51

3.3 Professor Hoxby's use of capacity constraints is incorrect for measuring whether race is a determinative factor. 53

3.4 Professor Hoxby's decile analysis actually confirms that UNC's racial preferences for URMs are enormous. 55

3.5 Professor Hoxby's criticism of Mr. Kahlenberg's SES preferences inadvertently show that UNC's racial preferences for URMs are enormous. 62

3.6 Standard assumptions in the economics literature imply that my models underestimate the magnitude of UNC's racial preferences. 65

1 Executive Summary

1.1 Overview

In my opening report, I concluded that non-underrepresented minority (URM) applicants to UNC have significantly stronger academic qualifications than URM applicants; that UNC's admissions process is guided by an implicit formula; and that race plays a dominant role in individual admissions decisions (especially for out-of-state applicants).

More specifically, I illustrated the magnitude of UNC's racial preferences with an example of a hypothetical male, non-first-generation-college (FGC), Asian-American applicant whose observed characteristics would imply a 25% chance of admission. If he were an in-state applicant, his probability of admission would increase to more than 63% (i.e., more than double) had he been treated like a Hispanic applicant, and to more than 88% (i.e., more than triple) had he been treated like an African-American applicant. If he were an out-of-state applicant, his probability of admission would increase to more than 85% had he been treated like a Hispanic applicant, and to more than 99% had he been treated like an African-American applicant. Simply changing this hypothetical Asian-American applicant's race to either Hispanic or African American (with all other characteristics remaining the same) thus would transform him from an unlikely admit to an almost certain admit.

In addition, I concluded that holding fixed the number of admission slots and removing racial preferences would significantly change Asian-American and white representation at UNC. Fixing the number of in-state admits at the levels observed in the data and removing racial preferences would result in 1,219 additional non-URM admits over the six-year period, a 5.5% increase. Fixing the number of out-of-state admits at the levels observed in the data and removing racial preferences would result in 2,482 more non-URM admits from that pool, a 25.7% increase.

By the same token, the impact of racial preferences on URM admissions is enormous, particularly for out-of-state applicants. For in-state applicants, racial preferences account for nearly one quarter of Hispanic admissions and 42% of African-American admissions. The marginal effects are even larger for out-of-state applicants notwithstanding the fact that the base admission rates are much lower. For out-of-state applicants, *racial preferences account for 70% of Hispanic admissions and 91% of African-American admissions*.

Professor Caroline Hoxby, who was hired by UNC to opine on these issues, submitted a rebuttal report in response to my opening report. In her rebuttal report, Professor Hoxby claims that my admissions models are unreliable, and she attacks my conclusion that race is a predominant factor in UNC's admissions decisions.

1.2 Professor Hoxby's criticisms of my model are unfounded

Professor Hoxby makes several claims regarding my models. First, she claims that my models are overfit—that is, she claims that my models "work only in the data used in the estimation [there]of" and that for any "alternative sample of applicants or … hypothetical applicants (that is, any applicants outside the data used in estimation), [my] models cannot explain admissions decisions reliably." Hoxby Rebuttal 4. Second, Professor Hoxby claims that I erred by including in my models UNC's ratings, which she calls "unverifiable measures," because in her view they "can only be assessed through holistic, subjective review of an actual application file." Hoxby Rebuttal 3. Third, Professor Hoxby claims that I am wrong to exclude from my model the applicants in special recruiting categories. Last, she claims that I do not properly account for missing performance variables, SAT scores, and capacity constraints.

Each of Professor Hoxby's criticisms is incorrect. Professor Hoxby's claims that my models are overfit serve only to demonstrate a misunderstanding of the concept of overfitting. While purporting to evaluate the fit of my model, she disregards half the relevant data set, improperly focuses on the predictions from models estimated using only one year of data, and uses a faulty and misleading metric. Even then, if I were to accept Professor Hoxby's incorrect approach, my models still are not overfit; in fact,

my models predict better *out of sample* than Professor Hoxby's models predict *in sample*.

Likewise, Professor Hoxby's argument against controlling for UNC's ratings is also incorrect. In her view, UNC's ratings should be excluded from the model because they are assigned by UNC's readers and may therefore be influenced by race. Hoxby Rebuttal 42-44. But as I explained in my rebuttal report, it is entirely appropriate to include UNC's ratings in the model, notwithstanding that they include some subjective elements, when determining how formulaic UNC admissions is. And to the extent that UNC's ratings are affected by race, controlling for them would *understate* the impact of UNC's racial preferences on admissions. In any event, the evidence suggests that race does not substantially affect UNC's ratings, except for the personal rating (which favors URM applicants); and removing the personal rating from my model does nothing to change my conclusions.

Professor Hoxby also misses the mark in attacking the removal of applicants in special recruiting categories from my models. She argues that, since they are part of the admissions process, they should be included in the estimation. But applicants in these categories clearly are considered differently by UNC; indeed, they are virtually guaranteed admission to UNC. Simply put, they are not similarly situated to other applicants; excluding them from the model is thus appropriate. Nevertheless, I show below that

including these applicants in my model has no impact on my findings and conclusions *when one accounts for their special talents*.

Professor Hoxby also claims that I mishandle missing performance variables, SAT scores, and capacity constraints. These responses all fall flat. Professor Hoxby and I largely handle missing performance variables in the same way, in that we both control for them in our models. The only difference is that I interact these indicators—such as missing high school GPA—with race. But it is important to do so because high school GPA varies substantially with race. In any event, I show that accounting for missing performance variables through alternative means has no impact on my findings and conclusions regarding the magnitude of UNC's racial preferences. Similarly, my method of accounting for standardized test scores is entirely appropriate. But even if I were to adopt Professor Hoxby's preferred method for handling standardized test scores (as I did in my rebuttal report), it would have no impact on my findings and conclusions.

Finally, Professor Hoxby's claim that I do not handle capacity constraints is demonstrably false. I account for capacity constraints in a manner that ensures that the model's predictions of the number of admits matches *exactly* the number of applicants UNC actually admits in each admissions cycle. And the adjustments I make when evaluating the impact of removing racial preferences is entirely consistent with the underlying admissions model. Professor Hoxby not only fails to offer any alternative

method to account for capacity constraints; *she fails to account for them in her own models.*

1.3 Professor Hoxby's criticisms of my conclusion that race is a dominant factor in UNC admissions are also unfounded.

As I demonstrated in both of my previous reports, UNC gives substantial racial preferences to URM applicants; indeed, UNC's racial preferences account for approximately one third of admissions for in-state URM applicants and a huge majority of admissions for out-of-state URM applicants.

And as outlined more fully below, URM admission probabilities would drop substantially if racial preferences were turned off (i.e., UNC did not employ racial preferences). If UNC's racial preferences were turned off, Hispanic admits would see an average probability of admission of 75.8%—a 24.2 percentage point decrease (because an actual admit's probability of admission is 100%). For in-state African-American admits, removing racial preferences would result in a much larger change—a 42.2 percentage point decrease in the probability of admission. The results are even more striking for out-of-state applicants. Removing UNC's racial preferences would cause sharp drops in the average probability of admission for Hispanic and African-American admits. Hispanic admits would have an average admission probability of 29.2%—a *70.8 percentage point decrease in the probability of admission*; and African-Americans admits would have an average probability of admission of *only 8.7%, a 91.3 percentage point decrease.*

Professor Hoxby's position is that racial preferences do not have a dominant effect on admission decisions, but her arguments lack support. First, she continues to misinterpret the Pseudo R-square metric; her reliance on it in attempting to evaluate the share of admissions impacted by race is thus incorrect. Moreover, she continues to insist that race must have a substantial impact on every applicant in order for it to be a determining factor in admissions.

Perhaps more importantly, Professor Hoxby inadvertently shows that UNC's racial preferences for URM applicants are enormous. She does two things that demonstrate this effect. The first relates to Professor Hoxby equalizing admission rates within deciles of the "admissions index". The admission index is derived from the estimated admissions model and shows how UNC weighs the various factors used in their admissions decisions. Recognizing that URM admit rates are higher than non-URM admit rates in each decile, Professor Hoxby purports to measure the effect of race by equalizing the admit rates for URM and non-URM applicants within each decile and then observing the resulting increase on the number of non-URM admits. But this is misleading; because there are significantly more non-URM applicants than URM applicants to UNC, focusing on the impact on non-URM admits ignores the much larger effect that racial preferences have on the number of URM applicants admitted to UNC. Moreover, the data underlying Professor Hoxby's analysis reveals that, under her approach,

racial preferences account for *70% of out-of-state URM admits*. In terms of absolute numbers, this equates to 2,399 URM admits over the six-year period for which UNC produced admissions data.

Second, Professor Hoxby's criticism of Mr. Kahlenberg's race-neutral alternatives analysis highlights the enormous magnitude of UNC's racial preferences. Mr. Kahlenberg provides multiple simulations of admissions using socioeconomic-status (SES) preferences. Professor Hoxby attacks these SES preferences as being unreasonably large. For example, she claims that Mr. Kahlenberg's SES preference can equate to an SAT bump of 278 points. But using Professor Hoxby's methods reveals that out-of-state African-American applicants receive a racial preference that equates to nearly 400 SAT points (and more than 400 points for FGC male African-American applicants). If Mr. Kahlenberg's SES preferences are unreasonably large (as Professor Hoxby claims), then so are UNC's racial preferences.

In short, nothing in Professor Hoxby's rebuttal report undermines my models or my conclusion that UNC's racial preferences play an outsized role in admissions decisions. If anything, Professor Hoxby's work supports my analysis.[1]

2 Professor Hoxby's criticisms of my model are incorrect or misleading.

[1] In formulating my reply report, I have not relied upon any data or material other than the material produced with Professor Hoxby's reports, the material cited in this report and my rebuttal report, and the data and materials identified in my opening report.

Professor Hoxby criticizes my modeling choices on a number of dimensions. I show in this section that each of her criticisms is unfounded. Before doing so, it is important to understand the key differences between how Professor Hoxby and I model UNC's admissions process. I have already outlined those differences in my rebuttal report but, for ease of understanding this report, I recap the major differences:

1) I believe it is appropriate to include UNC's ratings; Professor Hoxby does not.

2) I allow racial preferences to vary by gender, first generation college student, and—most importantly—residency status; Professor Hoxby does not.

3) I control for indicators for each admissions cycle; Professor Hoxby does not.

Against this backdrop, Professor Hoxby raises a number of concerns about my modeling choices. One of these is that my models are "overfit." That is, by controlling for too many factors, UNC's admissions decisions may appear more formulaic than they actually are; I will not have uncovered UNC's actual underlying admissions formula but rather a statistical artifact.

If true, this would be an important criticism of my conclusion that UNC's admissions process, especially for in-state applicants, is highly formulaic (see pages 22-28 of my rebuttal report). However, this criticism is entirely unfounded. Her "tests" for whether a model is overfit are incorrectly implemented. Properly testing for overfitting shows that my models are not overfit at all; in fact, my models have much greater predictive value than Professor Hoxby's preferred model.

Professor Hoxby's other criticisms of my approach are similarly misguided, as I show in Section 2.3. Her criticism of the inclusion of subjective ratings is ill-founded for questions regarding the size of racial preferences; her criticism of not properly taking into account of capacity constraints are factually wrong and actually apply to her models, not to mine; and her criticisms of my handling of missing performance measures (such as high school GPA) are also incorrect. Finally, Professor Hoxby's criticism of how I handle SAT scores and census tracts has no effect on the magnitude of racial preferences. Indeed, in the latter case, responding to her critique results in even larger estimates of racial preferences.

2.1 Professor Hoxby's claims that my models are overfit are demonstrably wrong; my models have much greater predictive power than her preferred model.

Overfitting is a phenomenon that occurs if one controls for too many factors in a statistical model. In simple terms, an overfit model is too complicated for the data set; the model ends up tailored to fit the noise in a specific sample rather than reflecting the overall data population, which means that it would not fit data that was not used in the estimation as well.

Professor Hoxby attacks my model for being "overfit." But her position on overfitting is fundamentally flawed. As explained further below, Professor Hoxby focuses on only three years of admissions data, improperly disregarding half of the data set produced by UNC. Contrary to standard practice in the field of statistical modeling, she focuses on the predictions

from models estimated using only one year of data; and her measures of overfit are incorrect. Even setting aside all these flaws, Professor Hoxby misses the mark. My models fit the data much better than her models; my models fit the *out-of-sample* data better than her models fit the *in-sample* data.

2.1.1 Overfitting generally.

The goal of the researcher in building a predictive model is to maximize its explanatory power (its ability to accurately explain the data set being analyzed) while minimizing out-of-sample error (making the model a reliable predictor for other data sets). There are often concerns that estimated models may find relationships that are artifacts of the specific data being analyzed—which would yield a model that accurately explains the data set being analyzed but that may be unreliable as a predictor for other data sets. That is, the model may have high explanatory power, but should the model be applied to another data set, those same relationships may not hold. A model that is overfit sacrifices out-of-sample predictive accuracy for in-sample explanatory power. This is what is meant by overfitting.

Overfitting occurs when a researcher controls for too many variables. As a researcher increases the complexity of a model by controlling for more factors, in-sample error will always decrease. Out-of-sample error will also decrease, but only up to a certain point. At some point, adding further controls detracts from the model's ability to make reliable predictions as to

out-of-sample data. Increasing model complexity that results in an increase in out-of-sample error is a classic sign of overfitting.

Figure 1: Model error in and out of sample

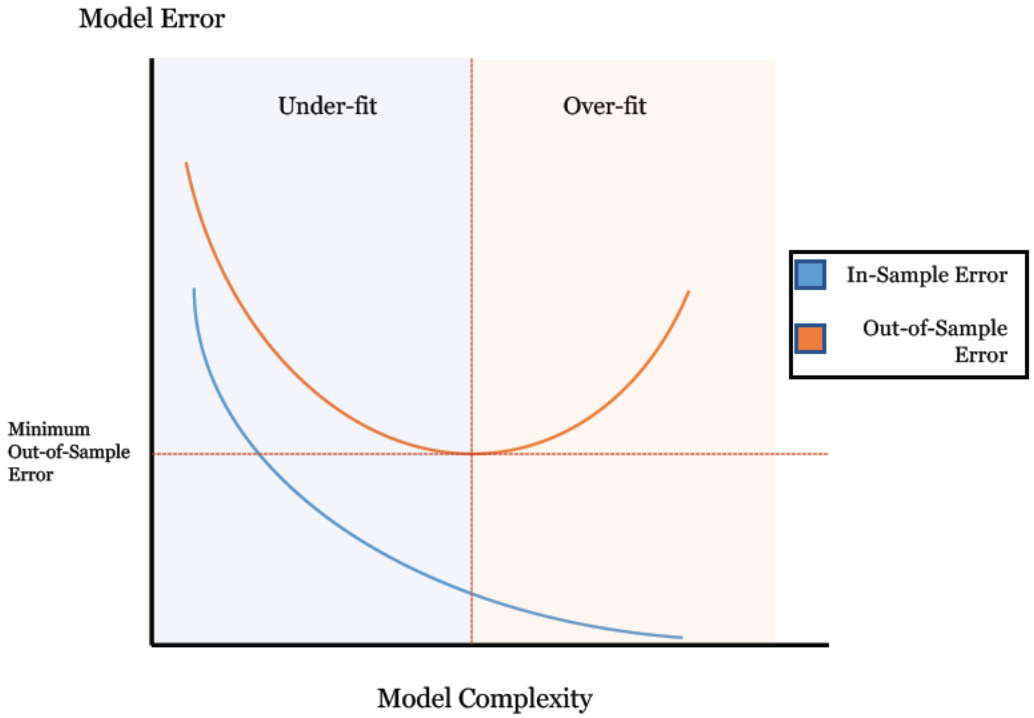

Figure 1 depicts the relationship between model complexity and model error in a way that illustrates what is meant by overfitting. As controls are added to a model, this increases model complexity; as model complexity increases (moves to the right), this decreases in-sample error (as shown by the blue curve). But at some point, the increased complexity starts to increase out-of-sample error (when the red curve turns back up). The area to the right of the vertical line is where the model is overfit.

This widely agreed upon definition of overfitting has the benefit of being intuitive and easy to identify.[2] In fact, much of the field of machine learning is based entirely on this principle: to find the best model for prediction, choose the model that has the best out-of-sample fit.[3]

In spite of these widely held principles, Professor Hoxby chooses not to use this easily accessible framework for assessing overfitting. Consider, in particular, her Exhibit 3. She estimates each of the admissions models separately on each year of data. She then examines how well the particular admission model in one year fits the data in that year as well as how well that same model fits the data in the other two years. It should be expected that if the model is estimated on data from one admissions cycle, the model will fit the data for that admissions cycle better than other two admission cycles. She uses a particular measure of fit—the average mean-squared error—and calculates this both for the admissions cycle that was used to estimate the model (referred to as in sample) as well as for the other two admissions cycles (referred to as out of sample). She then examines how the *percentage change* in mean-squared error varies between the in-sample and out-of-sample data. For example, in Exhibit 3 Figure 1, Professor Hoxby

---

[2] See Gareth James, et al., *An Introduction to Statistical Learning, with Applications in R*, at 29-33 (2015).

[3] Machine learning is concerned with finding patterns in the data and then using those patterns to predict future data. See Gareth James, et al., *An Introduction to Statistical Learning, with Applications in R*, at 29-42 (2015); see also Kevin P. Murphy, *Machine Learning, a Probabilistic Perspective,* MIT Press, at 23 (2012).

states that the increase in mean-squared error out of sample relative to in sample is 5% for her preferred model as compared to 34% for my preferred model (model 4 of my opening report) concluding that my models are overfit.

As I will show, there are a number of issues with Professor Hoxby's approach, including her choice of data and measures of fit. But most importantly, it is not the *percentage change* that is relevant for whether a model is overfit. Among other reasons, reporting the percentage change is misleading, as it will tend to be larger for models that are highly accurate in sample. Indeed, consider two models—A and B—with the same mean-squared error out of sample. If model A is more accurate in sample, it will have the higher percentage change in mean-squared error, notwithstanding that it is just as accurate out of sample and more accurate in sample than model B.

The key question, then, for assessing overfitting is this: which model predicts best out of sample? As I will show, on this dimension, my preferred model—as well as almost all of my models—performs substantially better than Professor Hoxby's preferred model. Indeed, it performs so much better that my model fits the *out-of-sample* data better than Professor Hoxby's model fits the *in-sample* data.

2.1.2 Professor Hoxby focuses only on a few years of admissions data, improperly disregarding half the data set produced by UNC.

UNC produced six full years' worth of admissions data. Yet Professor Hoxby focuses only on the last three admissions cycles for her tests of

"overfitting." This is in part because her model ignores the first two years of data—a decision she never explains. And, in testing for overfit, Professor Hoxby disregards the third year of data. In other words, her tests for overfit are based only on half the data set, notwithstanding that my model uses the entire data set. Arguably, the three years of data Professor Hoxby ignores are the most relevant for evaluating UNC's admissions process, because they cover the three admissions cycles prior to the filing of the complaint in this case.

A natural question, then, is why? The only rationale Professor Hoxby offers for removing the third year of data is in the footnote to Exhibit 3 Figure 1 that states, "Due to a change in the recording methodology of parent/guardian education level after 2013-14, which is used in Hoxby Report Exhibit 1 Table 1 row (9), the calculation of MSE starts in 2014-15." As my rebuttal report showed (see pages 19-20), this variable is problematic because the coding of it changes over time. What Professor Hoxby has done is to remove the data that are least consistent with her claims.

This is problematic in itself, of course; but it is also facilitates another misleading line of criticism Professor Hoxby undertakes about the strength of my model. With six years of data, it is possible for me to introduce more controls into my model; the more observations (in this case, applicants) in a data set, the more it is possible to uncover more relationships in the data

with a reasonable degree of certainty. Using only half the data makes it easier for Professor Hoxby to argue that my models are too complex.[4]

2.1.3 Contrary to standard practice, Professor Hoxby's tests focus on the predictions from models estimated using only one year of data.

Professor Hoxby compounds the problem of using too few years of data by using unorthodox and unsound methods to estimate out-of-sample error. She estimates the models on a single year of data and then predicts the outcomes for the other two years of data. Her approach works to the detriment of models, like mine, that have many controls. In testing for whether my models are overfit, she takes a model that was originally estimated on *six years* of data and then claims to show it is overfit by estimating the same model on *one year* of data. This section explains why this methodology is unsound.

Economists understand, and widely agree, that while it is important to test models with out-of-sample data, this involves a tradeoff. To assess out-of-sample data, one must leave some data out of one's "in-sample" estimation. But, of course, it is desirable to use as much of the data as possible in one's original "in-sample" model.

[4] This is *especially* true for my models that account for the high school or geography. Accurately measuring the effects of particular high schools or locations on admissions decisions requires a large amount of data; removing half the available the data will naturally weaken the reliability of the estimates.

To address this tension, researchers frequently turn to "cross-validation." I discuss some of the details associated with cross-validation—and in particular what is referred to as k-fold cross-validation—in Appendix B.1. This form of cross-validation, which is widely discussed in introductory and advanced textbooks for statistical and machine learning, has many advantages. First, it is easily to implement and very intuitive. Second, it allows models to be trained on all available data, albeit at different points in the estimation. Third, it provides a reliable estimate of the true out-of-sample mean-squared error. It typically uses 80 to 90 percent of the data to estimate the model and then sees how the estimated model fits on the remaining data.[5]

In spite of these clear advantages and widespread acceptance, Professor Hoxby instead employs a test for overfitting that is not supported by any literature or textbooks, and has none of the advantages described above. After throwing out the first three years of data, Professor Hoxby's method only uses 33% of the remaining data to estimate her model. Estimating the model on only one-third of the data has no theoretical justification and is contrary to standard practice.

2.1.4 Professor Hoxby is wrong to focus on changes in mean-squared error in her overfit tests.

[5] See Gareth James, et al., *An Introduction to Statistical Learning, with Applications in R*, at 184-86 (2015).

Professor Hoxby's use of only one year for rerunning her model is, as we have shown, unsound. But her outcome measure is also incorrect. In particular, she focuses on average mean-squared error. To calculate this, she uses the estimates of the model for one year and then calculates the predicted probability of admission for each applicant in the other two years. She then subtracts this predicted probability from the actual admissions decision, which is a 1 if the person was admitted and a 0 otherwise. She then squares this difference to get the mean-squared error for a particular applicant. A lower mean-squared error implies a better model fit because the predicted probabilities are then closer to the actual admission decisions. Finally, she calculates, the *percentage increase* in mean-squared error by moving from the estimation sample to each of the two test samples on which the model was not estimated. Then, she takes the average of these percentage increases.

There are a number of issues with this approach. First, like the R-squared, the mean-squared error makes sense in the context of a linear model. But when the variable is discrete and the model is non-linear (as is the case here), mean-squared error is not the appropriate measure.[6] A more standard measure for a discrete model is to evaluate the accuracy of the model, which I discussed at length in my previous report.[7]

---

[6] See Gareth James, et al., *An Introduction to Statistical Learning, with Applications in R*, at 127-37 (2015).

[7] I show my model is more accurate both in sample and out of sample than Professor Hoxby's model in the next section.

But even if we use the mean-squared error, it is clearly wrong to evaluate out-of-sample fit by measuring the percentage increase in the average mean-squared error. This tells us *nothing* about which model fits better out of sample. To see this, consider estimating a model with no controls except for an intercept. The intercept assures that the predicted admission rate from the model matches UNC's actual admission rate. If the overall admission rate in the estimation sample was the same as in the holdout sample, then the average mean-squared error would be exactly the same, implying a percentage increase of *zero*. But this model would fit the data horribly. As controls are added, it is likely the percentage increase in the average mean-squared error will rise. It is only when adding controls that result in the model fitting worse out of sample that overfitting becomes a concern.

Note that Professor Hoxby only shows her measure of overfit for her preferred model, model 9. In the first column of Table 2.1, I show Professor Hoxby's measures of overfit for each of her models where each model has more controls than the previous model.[8] According to Professor Hoxby's criteria, model 1 is the least overfit. This is consistent with what I just

[8] These refer to her additive models from Exhibit 1 Table 1 of her original report where, as in Professor Hoxby's rebuttal report, I only use the last three admissions cycles.

described: for models that explain little of the variation in the data there is not much scope for differences between the in-sample and out-of-sample fit.[9]

Table 2.1: Professor Hoxby's overfit measures for all of her additive models

Hoxby Model	Increase in MSE, Out-of-Sample Relative to In-Sample	In-Sample MSE	Out-of-Sample MSE
1	1.51%	0.169	0.171
2	3.58%	0.167	0.172
3	3.40%	0.135	0.139
4	3.40%	0.135	0.139
5	4.53%	0.109	0.114
6	4.92%	0.106	0.111
7	5.04%	0.105	0.110
8	5.50%	0.105	0.110
9	5.01%	0.101	0.106

Note: Following Hoxby's rebuttal report, Exhibit 3, Hoxby models are only estimated on the last three admissions cycles.

The right metric is then *not* the percentage increase in the average mean-squared error but the average mean-squared error in the out-of-sample data. That is, how well does the model fit on the data *not* used in the estimation? The second and third columns of Table 2.1 show the in-sample and out-of-sample average mean-squared error for each of Professor Hoxby's models. Smaller values in these columns mean that the models fit the data better. Of Professor Hoxby's models, model 9 fits the data best both in and out of sample, despite scoring worse than many of her other models on her

[9] Professor Hoxby's other test—shown in Exhibit 4 of her rebuttal report—suffers from the same problem. Namely, based on her criteria all of her additive models besides model 2 are less overfit than her preferred model.

measure of overfit (column 1). In the next section, I show that, according to how well the model fits out of sample, my models substantially outperform Professor Hoxby's preferred model.

2.1.5 Even using Professor Hoxby's incorrect approach, my models are not overfit; in fact, my models predict better out of sample than Professor Hoxby's models predict in sample.

I now show that—despite the fact that Professor Hoxby's method uses too little of the data and that mean-squared error is not the appropriate outcome measure—my models still perform better than her models both in sample *and* out of sample.

Table 2.2 shows the average mean-squared error both in sample and out of sample for the models Professor Hoxby displays in Exhibit 3, Figures 1 and 2. The first columns are for the in-state applicants.[10] There are three things to note. First, as expected given that my models progressively add controls, the measure of in-sample fit steadily improves as the errors become smaller and smaller. Second, as more controls are added to my models, the out-of-sample fit continues to improve up to my preferred model (Model 4). Using this measure shows some evidence of overfitting when high school or census tract fixed effects are included.[11] Finally, all of my in-state models fit

[10] Note that Professor Hoxby's models apply to the data as a whole; she does not distinguish between in-state and out-of-state applicants when calculating the average mean-squared error for her model.

[11] Recall that Professor Hoxby's method uses much less data to estimate the model than is standard in overfitting tests. This is especially relevant when I am controlling for detailed information on, for example, high schools. Estimating high school fixed effects requires having many observations per

the data better than Professor Hoxby's preferred model both in sample *and* out of sample. Indeed, my models generally have lower average mean-squared errors *out of sample* than Professor Hoxby's models have *in sample*. This point is striking; it means that my model does a better job of making predictions for new data sets than Hoxby's model predicts within the data set being analyzed. In short, my models have much greater predictive power than Professor Hoxby's models.

Table 2.2: My preferred model is not overfit and my models fit the data better than Professor Hoxby's preferred model.

	Average mean-squared error	
	In-Sample	Out-of-Sample
Hoxby		
Model 9	0.101	0.106
Arcidiacono in-state		
Model 2	0.092	0.102
Model 3	0.056	0.074
Model 4	0.055	0.074
Model 5	0.056	0.075
Model 6	0.035	0.088
Model 7	0.028	0.093
Arcidiacono out-of-state		
Model 2	0.072	0.084
Model 3	0.046	0.063
Model 4	0.045	0.063
Model 5	0.061	0.077
Model 6	0.037	0.104

Note: Following Hoxby's rebuttal report, Exhibit 3, Hoxby model is only estimated on the last three admissions cycles. Arcidiacono models refers to my opening report.

---

high school, which will not be the case when only one-sixth of the data is used in estimation.

The same patterns hold for the out-of-state applicants as shown in the second set of columns. Here again, adding controls reduces in-sample error, increasing the explanatory power of the model. And, as is the case for in-state applicants, the fit worsens out of sample beginning with model 5. But once again, *all* my models fit the data better both in sample and, more importantly, out of sample than Professor Hoxby's models. Professor Hoxby's claim that my models are unreliable thus is baseless.

2.2 Standard statistical methods show that my models are more accurate and reliable than Professor Hoxby's models.

In the previous section, I showed that my models are more accurate than Professor Hoxby's models—both in sample and out of sample—even using her flawed method that uses too little of the data and focuses on mean-squared error rather than accuracy.[12] Here, I show that my model also outperforms Professor Hoxby's when we use generally accepted evaluation techniques. While in the previous section, I focused on the model used in my first report to be consistent with Professor Hoxby's rebuttal report, I now use my preferred model from my rebuttal report (again model 4), though my conclusions are not sensitive to this choice.

[12] Accuracy requires setting a threshold for the admissions probabilities. As discussed by many authors (see, e.g., Kenneth E. Train, *Discrete Choice Methods with Simulation*, at 69 (2d ed. 2009); J.S. Cramer, *Logit Models from Economics and Other Fields*, at 66-67 (2003)), setting the threshold to 0.5 is often not appropriate. I set the threshold such that the same fraction of applicants is admitted as what is seen in the data.

In particular, I use five-fold cross validation to evaluate whether my models are overfit.[13] Results for my models as well as Professor Hoxby's model 9 are shown in Table 2.3.[14] Similar to what I showed in my rebuttal report, Professor Hoxby's preferred model is reasonably accurate, just significantly less accurate than my preferred model. Consider my preferred model—that is, model 4 from my rebuttal report. My accuracy for in-state admits and rejects is 92%, implying an overall accuracy also of 92%. As the out-of-state admit rate is substantially lower, the accuracy of out-of-state admissions is also lower but still quite high at 75%, and the overall accuracy of my preferred model on the out-of-state data is 93%. The overall accuracy of Professor Hoxby's preferred model is 85%.

The only evidence of overfitting for models on the in-state data is when census tracts are added (model 7). For the out-of-state data, accuracy slightly decreases when high school fixed effects are added (model 6). But in both cases, these models are still more accurate overall than Professor Hoxby's preferred model.

---

[13] Per the explanation above and in Appendix B.1, this means that the data is (randomly) divided into five groups. The model is then estimated on the first four groups and then the accuracy of the model is calculated based on the data from the last group. Hence, 80% of the data is used to estimate the model and 20% is used to assess accuracy. This process is repeated four additional times resulting in each data group appearing in estimation four times and being used to test the accuracy of the model once. The accuracy of the model—both for admits and rejects—is then calculated five times and averaged.

[14] Consistent with Professor Hoxby's rebuttal report, I use her estimates of model 9 from the last three years of data rather than the four that are in her original report.

Table 2.3: My models are more accurate than Professor Hoxby's preferred model.

	Accuracy for Admits	Accuracy for Rejects	Overall Accuracy
Hoxby			
Model 9	71.6%	89.9%	85.1%
Arcidiacono in-state			
Model 2	86.9%	87.9%	87.4%
Model 3	91.6%	92.2%	91.9%
Model 4	91.6%	92.2%	91.9%
Model 5	91.9%	91.9%	91.9%
Model 6	92.3%	92.4%	92.3%
Model 7	92.1%	91.8%	92.0%
Arcidiacono out-of-state			
Model 2	62.2%	94.1%	89.8%
Model 3	74.9%	96.0%	93.2%
Model 4	75.1%	96.0%	93.2%
Model 5	76.2%	95.3%	92.2%
Model 6	74.9%	93.6%	90.6%

Note: Following Hoxby's rebuttal report, Exhibit 3, Hoxby model is only estimated on the last three admissions cycles. Arcidiacono models refer to my rebuttal report, though results are similar using the models from my opening report.

2.3 Professor Hoxby's other criticisms of my models are incorrect or misleading.

2.3.1 Professor Hoxby's arguments for excluding UNC's ratings are misleading or incomplete.

Professor Hoxby criticizes my use of UNC's ratings in my admissions models for three reasons.[15] First, she argues that they are not "verifiable." Second, she argues that they are *endogenous,* meaning that they are a

[15] This is ironic given that she later accuses my models of putting "too much weight on observable student characteristics—especially test scores, grades, and race—relative to UNC's true admissions process." Hoxby Rebuttal 47. Her models (which ignore UNC's ratings) are much more susceptible to this criticism.

function of race themselves. Third, she argues that since they are not in the NCERDC data—and therefore cannot be used in counterfactuals where the applicant pool changes—they should not be included in the model.

The arguments for the inclusion or exclusion of these (or any) variables in a model depend on the question being addressed by the model. To address Professor Hoxby's criticism of my inclusion of UNC's ratings requires identifying and distinguishing between the different questions addressed by my models. There are at least three sets of questions that an admissions model might be designed to address regarding UNC's admissions process:

1) To what extent is UNC's use of race (and its admissions processes more generally) formulaic?

2) How large are the racial preferences that UNC gives to URM applicants?
3) If racial preferences were removed and replaced with race-neutral strategies, what would be the effect of these alternative strategies?

Professor Hoxby's first argument is related to the first question. I have already addressed this issue in my rebuttal report where I discuss how subjective factors can clearly be part of a formula and therefore should be included in the model (see pages 5-6).

Her second argument relates to the second question. Professor Hoxby argues against controlling for UNC's ratings because they are assigned by UNC's readers and may therefore be influenced by race. See Hoxby Rebuttal 42-44. It is related to the second question because by controlling for these ratings I may be missing out on the fact that race affects these ratings.

There are a number of reasons, however, why this argument is not relevant here or operates in a way in which, by including these ratings, I am *underestimating* the extent of racial preferences. First, I have shown that there are substantial racial preferences *even after* controlling for UNC's ratings. We would not expect UNC to, for example, penalize African-American applicants in one of the ratings and then give them a bonus later for admission. Hence controlling for these ratings should not result in an overestimation of the role of race. To the extent that racial preferences *also* affect the ratings, my estimates of racial preferences are understated because they do not take into account racial preferences in the ratings themselves.

Second, deposition testimony as well my own statistical analysis support the conclusion that race plays little role in the ratings themselves with the exception of the personal rating, in which URM applicants receive a significant advantage.[16] As I now show in Appendix Table A.1, removing from the analysis the personal rating—the one rating in which racial preferences are likely a factor—slightly increases my estimates of the marginal effects of race.

Third, not controlling for these ratings would produce *omitted variables bias*. I discussed this issue at length in my second report (see pages 7-8 of my rebuttal report).

[16] See, e.g., Rosenberg Depo. 250:13-51:8.

Professor Hoxby's final argument for the exclusion of these ratings is that, because these ratings are not available in the NCERDC data, we have no way of knowing how students who did not apply would be rated on these dimensions. Professor Hoxby argues that this is important because under counterfactual scenarios—such as the race-neutral alternatives Mr. Kahlenberg presents—this will likely change who applies to UNC. Note that this argument has no bearing on the first two questions (which are the questions germane to my report); whether UNC's ratings are excluded from the model is not relevant for whether UNC's use of race is formulaic or the magnitude of the racial preferences UNC gives to Hispanic and African-American applicants.

Nonetheless, Professor Hoxby's argument here is, in any case, rather odd because *Professor Hoxby's own admissions models have a number of variables that are not present in the NCERDC data.* Examples include: whether the applicant is eligible for a fee waiver, whether the applicant applied early action,[17] and the applicant's intended major.

2.3.2 Including applicants in special recruiting categories does not change my findings.

Professor Hoxby criticizes the removal from my models of applicants in special recruiting categories. Professor Hoxby argues that, since they are part

[17] Although Professor Hoxby purports to control for early action in her models, she in fact failed to do so—both in her opening report (see Arcidiacono Rebuttal 18) and in her rebuttal report.

of the admissions process, they should be included in the estimation.[18] But UNC's applicant data reveals that applicants in these special recruiting categories were treated much differently by UNC. At least *97%* of applicants in each of these categories were admitted. In other words, these applicants were virtually guaranteed admission. This necessarily implies that the defining feature of these students' applications is their inclusion in a special category (i.e., as an athlete). Further, the factors that affect the admissions decisions of other applicants may apply to these students in a very different way. This is what informed my decision to remove them from the estimation.

Professor Hoxby claims that when she adds these observations back into the sample, the estimated racial preference for African Americans goes down (as measured by the logit coefficient).[19] But it is important to remember that these applicants have certain qualities that place them in UNC's special recruiting categories and afford them significant (non-racial) preferences in UNC's admissions process. Remarkably, when she adds special recruiting categories to my models 2 and 3 (see Hoxby Rebuttal Exhibit 5), Professor Hoxby *does not* control for whether applicants are in one of these special

[18] Note that Professor Hoxby does not make the same objection regarding foreign applications, which I also remove from my models.

[19] Here it is important that the logit coefficients themselves do not have a natural interpretation beyond that higher values imply higher admission probabilities. The magnitude of the coefficient is not informative by itself, except as a comparison to other coefficients. For example, the coefficient on African American could be compared to the coefficient on FGC student to say which preference is *relatively* higher. What the coefficients *can* be used for is to calculate marginal effects; that is, how admissions probabilities change as a result of having that particular characteristic.

recruiting categories. This is the case even though in Professor Hoxby's own model she controls for athletes and athletes are one of the special recruiting categories. So despite recognizing the necessity of controlling for athletes, she does not control for them when she adds them to my model in Exhibit 5. It is therefore no surprise that the model would fit the data worse when those in special recruiting categories are added without controlling for them at the same time.

In Table 2.4, I replicate Exhibit 5 of Professor Hoxby's rebuttal report but add another row for each model that includes one additional variable: an indicator for whether the applicant belonged to one these special recruiting categories. That is, I add a control for whether the applicant had some quality or qualities that resulted in the applicant being assigned to one of these special recruiting categories.

Table 2.4: Including special recruiting categories has no impact on my model provided they are accounted for in estimation.

	Psuedo R^2	Coefficient on African American	Coefficient on Hispanic
In-State			
Model 3	0.715	2.85	1.81
Model 3 including specials	0.688	2.41	1.53
...plus special Indicator	0.732	2.82	1.79
Model 2	0.565	1.84	1.27
Model 2 including specials	0.556	1.72	1.17
...plus special Indicator	0.593	1.83	1.26
Out-of-State			
Model 3	0.584	5.85	3.01
Model 3 including specials	0.522	4.32	2.30
...plus special Indicator	0.640	5.66	2.90
Model 2	0.416	4.68	2.43
Model 2 including specials	0.352	3.75	1.95
...plus special Indicator	0.496	4.54	2.34

Note: This table replicates Exhibit 5 of Professor Hoxby's rebuttal report but adds a row in which, when special recruiting categories are included in the estimation, an indicator variable for being in a special recruiting category is also included.

Simply adding this one variable produces results that are notable in two ways. First, the coefficients on African American and Hispanic are very close to what they were when those in the special recruiting categories were not included in estimation. Consider, for example, in-state African-American applicants. My estimated coefficient for this group is 2.85 using model 3 without special recruiting categories. When Professor Hoxby adds special recruiting categories to the estimation, the coefficient falls to 2.41, informing her opinion that my omission is significant. But when I simply add an indicator variable for being in a special recruiting category—which reflects the special talents of the applicant—the coefficient rises to 2.82, very close to

the estimate when these applicants were removed from the estimation. True, the coefficient is lower, but only by a trivial amount *when one accounts for the qualities that placed them in these special recruiting categories.*

The second change is that the Pseudo R-square in each case is higher with this variable included, showing that this variable improves the fit of the model. Indeed, the Pseudo R-square is now higher than when these applicants were removed. This stands in direct contrast to Professor Hoxby's claim that including these applicants worsens the fit of the model. That the fit improves makes sense: it is easy to predict admission outcomes for these applicants because (as noted above) they are virtually guaranteed admission. In Appendix Table A.2, I show that these results also hold when I use my preferred measure of model fit—that is, by measuring the accuracy of my predictions, consistent with these applicants' admissions decisions being easier to predict.

Professor Hoxby's criticism is thus entirely unfounded. The estimates of racial preferences as she calculates them are unchanged when accounting for special recruiting status. And the fit of the model is also improved when these applicants are included, showing that UNC's admissions decisions are, if anything, even more formulaic than what I showed in my rebuttal report (see pages 23-28).

2.3.3 The method I use to account for missing performance variables is appropriate.

In the data produced by UNC, there are some applicants for whom there is no value indicated for certain performance variables (i.e., high school GPA and class rank) and others for whom the values for certain variables are implausible. Professor Hoxby criticizes how I handle these missing (or implausible) values. Both Professor Hoxby and I include controls for these missing performance variables in our models. I further interact these variables with race. Professor Hoxby claims this results in overfitting. But as I have shown earlier, my models are not overfit; even using Professor Hoxby's method my models predict better *out of sample* than her models predict *in sample*.

Professor Hoxby further argues that I should have turned off these interactions in my counterfactuals. For example, when I simulate the removal of racial preferences, I leave the coefficients on these variables unchanged.

But this was a deliberate choice and an appropriate one. To understand why requires an understanding of what these interactions between missing performance indicators and race are accomplishing in the model. Consider applicants who are missing a score for high school GPA. What both my models and Professor Hoxby's models do when we control for an indicator for missing high school GPA is let the observed admissions outcome "assign" a GPA for these individuals. For example, if these

applicants as a whole were as likely to be admitted as an applicant with a 3.0 (all else equal), then the "assigned" GPA would be a 3.0.

Professor Hoxby and I take the same general approach to replacing these missing performance variables; we both include indicators in our models for whether a particular variable was missing. However, there is an important difference in how we accomplish this replacement: I interact race with missing high school GPA; Professor Hoxby does not. The reason it is important to have this interaction is simple: average GPAs vary substantially by race. Therefore, we would suspect—or at least allow for the possibility—that the actual GPAs of those for whom GPAs are missing also vary with race. Interacting race with missing high school GPA allows the assignment of missing GPAs to vary by race.[20] Hence, in any counterfactual we would *not* want to turn off these interactions because this would also change the applicants' assigned GPAs. The counterfactuals involve changing race or removing racial preferences *holding all other characteristics fixed*, including their assigned GPAs.

Professor Hoxby's complaint boils down to her preferring an alternative theory: that UNC utilizes racial preferences differently for those who are missing high school GPA. Hence, when the missing interaction between African American and missing GPA is negative, it is not because

[20] This technique is commonly used in statistical analysis. See, e.g., Valentino Dardanoni, et al., *Regression With Imputed Covariates: A Generalized Missing-Indicator Approach*, Journal of Econometrics, Vol. 162, at 362-68 (2011).

African-American applicants with missing GPAs actually had lower GPAs than their white counterparts but rather that the preferences are as not strong for African Americans with missing GPAs as for those African Americans for whom GPA is not missing. It is only under this explanation that one would want to turn off these interactions in a counterfactual.

But there is no evidence in the record supporting Professor Hoxby's alternative theory that racial preferences differ for those applicants who are missing high school GPA. In fact, the data actually proves that this theory is wrong. As I illustrate in Appendix Table A.3, African-American applicants for whom high school GPA is observed have substantially lower GPAs than their white and Asian-American counterparts (more than 0.3 grade points for both in-state applicants and out-of-state applicants). Further, African-American applicants have lower performance scores (which are essentially a function of unweighted GPA) both when GPA is observed *and* when it is missing than their white and Asian-American counterparts.[21] Hence, the logical conclusion is that there *are* differences in GPAs across races for those for whom GPA is missing.

2.3.4 Alternative ways of accounting for missing performance variables yield similar findings on the magnitude of racial preferences.

Alternative ways of accounting for missing performance variables yield similar findings on the magnitude of racial preferences, demonstrating that

[21] This too holds for both in-state and out-of-state applicants.

my findings are not dependent on the method by which I account for missing variables. Instead of interacting the missing performance variables with race, I assign values for those who are missing SAT and/or high school GPA variables using two alternative methods and then control for an indicator for whether the performance variable is missing.[22] I then show that utilizing these alternative means of replacing these missing variables yields similar findings regarding the magnitude of UNC's racial preferences.

For the first method, I calculate the race-specific means for high school GPA and class rank, as these are the two performance variables for which I interacted race with the missing variables.[23] This is done separately for in-state and out-of-state applicants. For those who are missing one or both of these values, I then assign the race-specific mean.

For the second method, I assign values for those who are missing high school GPA or class rank but do so in a way that does not use race but does use other measures that are correlated with these variables. For example, consider high school GPA. I perform a linear regression of high school GPA on indicators for each of UNC's program and performance ratings, indicators for

[22] These methods are also used in the literature. See, for example, Roderick Little, Regression with Missing X's: A Review, Journal of the American Statistical Association, Vol. 87, No. 420, at 1227-36 (Dec. 1992).

[23] This calculation is made only for those who are not missing these variables. Further, I focus only on the official class rank as my estimation results show that unofficial class rank has little effect on admissions decisions.

each year, and the applicant's SAT math and verbal scores.[24] I do this separately for the in-state and out-of-state applicants. I then use the results of this regression to predict the high school GPAs of those who are missing that variable.

Given the assigned values from each of the methods, I then estimate my admissions models. I continue to include indicator variables for whether the applicant was missing, for example, high school GPA to take into account average differences in high school GPAs between those who were missing this variable and those who were not. I show in Table 2.5 the marginal effects of race for my preferred model (model 4) using my original way of handling missing race as well as the two imputation methods. The marginal effects are relatively unchanged.

[24] Note that the data used in the regressions only includes those who are not missing the particular outcome variable (for example, high school GPA). Regressions are also done separately for the two rank types I use in the analysis.

Table 2.5: Average marginal effects of race under alternative imputation procedures for missing performance measures.

		Average Admission Probability			
		with Racial Preferences	without Racial Preferences	Marginal Effect of Race	**Share due to Racial Preferences**
In-state					
African American	Preferred	30.5%	17.8%	12.7%	**41.7%**
	Impute 1	30.5%	17.4%	13.2%	**43.2%**
	Impute 2	30.5%	17.9%	12.6%	**41.3%**
Hispanic	Preferred	41.0%	31.2%	9.7%	**23.8%**
	Impute 1	41.0%	31.4%	9.6%	**23.4%**
	Impute 2	41.0%	31.9%	9.0%	**22.0%**
Out-of-state					
African American	Preferred	17.1%	1.5%	15.6%	**91.1%**
	Impute 1	17.1%	1.7%	15.4%	**90.3%**
	Impute 2	17.1%	2.0%	15.1%	**88.4%**
Hispanic	Preferred	20.3%	6.0%	14.2%	**70.2%**
	Impute 1	20.3%	5.9%	14.3%	**70.6%**
	Impute 2	20.3%	6.3%	13.9%	**68.8%**

Note: Preferred refers to model 4 of my rebuttal report. Impute 1 replaces missing high school GPA and missing official high school class rank with race-specific means; Impute 2 replaces these from regressions on UNC's program and performance ratings, year, and SAT scores.

This analysis supports that my original method of interacting missing performance indicators with race was appropriate. *Further,* it shows that these measures were capturing actual differences in high school GPA and class rank, *not* differential racial preferences based on whether, for example, high school GPA was missing as the marginal effects of race are relatively unchanged. This means that it was therefore appropriate to *not* turn off these interactions in my counterfactuals, either when evaluating how non-URM admit probabilities would change if they were treated like URM applicants or when evaluating various race-neutral alternatives.

2.3.5 My use of the information on SAT scores is appropriate, and my results are not sensitive to this choice.

Professor Hoxby criticizes my approach to handling SAT scores. First, she criticizes me for not taking into account changes from the old SAT to the new SAT. If this were a real concern, we would expect to see significant differences in SAT scores across the years where the different test scores were in place. Yet examining average test scores across years do not show patterns suggesting that the meaning of SAT scores has changed in any significant way. Further, to the extent that there are differences across years, these will be captured by my controls for the year of the admissions cycle. That is to say, one of the reasons to include year-by-year controls is to account for changes in how the means of the controls vary across years. For example, if new SAT scores are higher than old SAT scores then the estimated coefficient on the admissions cycle that used the new SAT score would be lower, capturing the average difference across the new and old SAT scores (if grade inflation in high school was increasing over time, for example, this would also be effectively controlled for by the year-to-year controls).

Professor Hoxby's second concern relates to the translation of ACT scores to SAT scores for those who only have the former. She employs ACT-to-SAT conversion tables that I do not use. There are a number of things to note here. First, at no point does she show that my results are sensitive to alternative ways of doing this translation. Second, the reason I did not use her preferred translation method is straightforward: the conversion tables

she refers to are for the *total* ACT score *not* the ACT component scores. As I control for the *separate* effects of SAT verbal and math on admissions outcomes, the conversion tool she suggests will not work. Finally—and most importantly—I showed in my rebuttal report that using her SAT score measures rather than mine had no effect on my findings. See Arcidiacono Rebuttal 34-36. These concerns are simply irrelevant.

2.3.6 Using more finely detailed geographical information would put the model at risk of being overfit.

Professor Hoxby also criticizes handling of geography in one of my models (in-state model 7). She claims that I "wrongly use a 9-digit FIPS code, rather than an 11-digit census tract." Hoxby Rebuttal 47. This is not correct; I am using census tracts. A FIPS code can have as many as 12 digits; the first two digits describe the state; the next three describe the county; and the next six digits describe the census tract; the twelfth digit (not included in any of the data produced in this case) describes "block groups" within census tracts. For the vast majority of census tracts, the last two digits of the census tract number are zeroes, which means there are only 9 digits in the code that have descriptively useful information. A minority of tracts have been subdivided at some point since the original census tract designation, and these subdivisions make use of the last two digits of the code. Thus, using an 11-digit code rather than a 9-digit code provides slightly more geographic detail, but only for a minority of tracts. When I run my models with the full 11-digit codes,

the results do not change—nor should we expect them to. The question of using 9-digit versus 11-digit codes is, in other words, a non-issue.

It should also be noted that Professor Hoxby is not arguing that I should be controlling for census tracts rather than basic census tracts in my models. Much of her rebuttal is premised on the claim that my models are overfit, a claim I have shown is false.[25] Adding finer geographical controls would create more, not less risk of overfit.

Partly because of concerns of overfitting and the desire to be consistent across the in-state and out-of-state counterfactuals, this model is only used as a robustness check. It provides complementary evidence that the estimates of my preferred model (model 4), if anything, understate the role of race in the admissions process.

2.3.7 Professor Hoxby's criticism of how I incorporate capacity constraints is wrong.

Finally, Professor Hoxby is critical of my model of capacity constraints, stating:[26]

> Prof. Arcidiacono's method of rescaling is ad hoc and does not apply capacity constraints in a manner that is accepted in the literature on discrete choice modeling. If Prof. Arcidiacono had accounted for "capacity constraints" in the estimation of his model, his predictions would not need this problematic ex post rescaling. The fact that his rescaling methodology has strong effects on the predictions indicates that his methodology is not well-suited for counterfactual exercises.

Hoxby Rebuttal 55.

---

[25] None of her models attempt to control for this sort of geography.

[26] Hoxby makes similar erroneous claims elsewhere. See Hoxby Rebuttal 28 n.57; Hoxby Rebuttal 54 n.124.

This statement is demonstrably wrong on multiple levels. First, capacity constraints are explicitly incorporated into my model through yearly intercepts. Namely, these yearly intercepts ensure that the model's predictions of the number of admits matches *exactly* the number of actual admits. Suppose, for example, that UNC won the NCAA basketball championship. This likely would lead to a rise in the number of applicants in the next admissions cycle. The data would then show a higher rejection rate for that year. Because there are more applicants, the threshold for admission would be higher, implying a lower intercept for that year as the probability of admission would be lower. The model adjusts these yearly intercepts in such a way that *guarantees* that the overall admit rate predicted by my model matches the overall admit rate in the data used for estimation *in every year*.

Second, although given any ranking system there is only one intercept that satisfies the capacity constraint, *changing* the ranking system *necessitates* changing the value of the intercept that satisfies the capacity constraint. A trivial example of this is that if I add 100 to the rankings of everyone then I will correspondingly have to lower the intercept by 100 to satisfy the capacity constraint.

What this means is that when I change the ranking of applicants by, for example, removing racial preferences, the intercept must also change to satisfy the capacity constraint. There is simply no way around this. This is

not an "ad hoc" adjustment to my original model—it is a step intrinsic to the analytical question being asked.

It is telling that Professor Hoxby says that my method "does not apply capacity constraints in a manner that is accepted in the literature" *yet she does not show an alternative method*. Nor does she provide any cites showing such a method. Nor do any of her models impose capacity constraints or discuss how capacity constraints apply to her models. Indeed, one of my criticisms in my rebuttal report was that Professor Hoxby's models did not incorporate capacity constraints because she did not control for yearly intercepts. See Arcidiacono Rebuttal 19.

To see that my models do incorporate capacity constraints while hers do not, Table 2.6 reports actual and predicted admits rates by year for each of our preferred models. As is clear from the table, my models of in-state and out-of-state admissions match the admit rates *exactly*. Because Professor Hoxby's models do not control for yearly intercepts, her models fail to match the admit rates in the data. In other words, Professor Hoxby's models fail to account for capacity constraints at all. Mine, on the other hand, do. Her criticism of the manner in which I incorporate capacity constraints thus is surprising.

Table 2.6: My model matches UNC admit rates in each year; Professor Hoxby's model does not.

Year	Applicants	Actual Admit Rate	Predicted Admit Rate	Difference
Hoxby model 9				
2018	31,331	0.285	0.278	-0.007
2019	31,956	0.297	0.270	-0.027
2020	35,875	0.262	0.268	0.007
2021	40,918	0.236	0.251	0.015
Arcidiacono in-state model 4				
2016	8,739	0.481	0.481	0
2017	8,804	0.496	0.496	0
2018	8,722	0.519	0.519	0
2019	8,964	0.490	0.490	0
2020	10,432	0.471	0.471	0
2021	11,564	0.434	0.434	0
Arcidiacono out-of-state model 4				
2016	15,563	0.121	0.121	0
2017	16,466	0.122	0.122	0
2018	16,617	0.155	0.155	0
2019	16,710	0.174	0.174	0
2020	18,712	0.135	0.135	0
2021	21,564	0.110	0.110	0

Note: Hoxby model follows her original report. Arcidiacono model refers to my rebuttal report, though, by construction the results would be the same using Arcidiacono models from my opening report.

Not only are Professor Hoxby's models unable to match the actual admit rate of the applicant pool in any given year, but they also fail to capture the differential way in which racial preferences operate. For example, Professor Hoxby's preferred model restricts racial preferences in such a way that they have the same magnitude for in-state and out-of-state applicants. This is an obvious modeling error given the clear evidence that racial preferences are substantially larger for out-of-state URM applicants.

I show how well both of our models predict admit rates for African-American and Hispanic applicants in Table 2.7. My models match these admit rates because I allow racial preferences to differ for in-state and out-of-state applicants. Because Professor Hoxby fails to do this, her models yield predicted admit rates that are too low for out-of-state applicants and too high for in-state applicants relative to the actual admit rates for those groups. For example, the actual admit rate for in-state African-American applicants is 32.4% but Professor Hoxby predicts an admit rate of 37.9%, 5.5 percentage points higher. This is counterbalanced for out-of-state African-American applicants. Here the actual admit rate is 18.2% but Professor Hoxby predicts an admit rate of 14.0%, more than four percentage points lower.

Table 2.7: My model matches UNC's differential treatment of minorities across residency status; Professor Hoxby's model does not.

	Race/ Ethnicity	Applicants	Actual Admit Rate	Predicted Admit Rate	Difference
Hoxby model 9					
In-State	African American	6,243	0.324	0.379	0.056
	Hispanic	3,042	0.430	0.464	0.034
Out-of-State	African American	8,325	0.182	0.140	-0.041
	Hispanic	8,173	0.222	0.209	-0.014
Arcidiacono in-state model 4					
In-State	African American	7,775	0.305	0.305	0
	Hispanic	3,589	0.410	0.410	0
Out-of-State	African American	9,585	0.167	0.167	0
	Hispanic	9,023	0.202	0.202	0

Note: Hoxby model follows her original report. Arcidiacono model refers to rebuttal report, though, by construction the results would be the same using Arcidiacono models from my opening report.

3 UNC provides substantial racial preferences to African-American and Hispanic applicants.

The opinion of both of my previous reports was that UNC's racial preferences are very large. My key findings were:

- Substantial racial preferences exist for in-state applicants. In-state Hispanic applicants are admitted at a rate of 41% and would experience a 9.7% drop in their admission rate if they were treated as white applicants. Racial preferences thus account for nearly a fourth of admissions for in-state Hispanic applicants. In-state African-American applicants are admitted at a rate of 30.5% with racial preferences. Removing racial preferences would result in a drop in admissions rate of 12.7%. In other words, racial preferences account for nearly 42% of African-American in-state admissions.

- Racial preferences are even larger for out-of-state applicants. For out-of-state Hispanic applicants, with racial preferences, their admit rates are 20.3%; without racial preferences, their admit rates would fall to 6.0%. Racial preferences thus account for 70% of out-of-state Hispanic admissions. Out-of-state admission rates for African-American applicants are 17%. *Their admissions rates would fall to 1.5% if they were treated as white applicants. In other words, racial preferences account for 91% of out-of-state African-American admissions.*

- Although UNC also gives preferences to FGC students, these preferences are much smaller than the racial preferences for URMs. Further, UNC gives FGC URM applicants a much smaller preference for their FGC status than it gives FGC non-URM applicants. Relative to non-URM applicants, then, FGC URM applicants are worse off than non-FGC URM applicants.

- Consider a male, non-FGC Asian-American applicant whose observed characteristics would imply a 25% chance of admission:

    - If he were an in-state applicant, his probability of admission would increase to over 67% (i.e., more than double) had he been treated like a Hispanic applicant, and to over 90% (i.e., more than triple) had he been treated like an African-American applicant.

    - If he were an out-of-state applicant, his probability of admission would increase to over 86% had he been treated like a Hispanic

applicant and over 99% had he been treated like an African-American applicant.

  - Simply changing this hypothetical Asian-American applicant's race to either Hispanic or African American thus would transform him from an unlikely admit to an almost certain admit.

- Holding fixed the number of admission slots and removing racial preferences would significantly change Asian-American and white representation at UNC. Fixing the number of in-state admits at the levels observed in the data and removing racial preferences would result in 1,171 additional non-URM admits over the six-year period, a 5.3% increase. Fixing the number of out-of-state admits at the levels observed in the data and removing racial preferences would result in 2,486 more non-URM admits from that pool, a 25.8% increase.

Professor Hoxby comes to a different conclusion regarding the importance of race in the admissions process. In this section, I show why we are coming to different conclusions and why her conclusions are not warranted. Before doing so, however, I show an additional illustration supporting my claim that racial preferences are determinative for many URM admits.

3.1 Race is a determinative factor for many African-American and Hispanic admits.

In my prior reports, I showed that race is a determinative factor for many URM admits using a variety of methods. Another way of showing the magnitude of UNC's racial preferences is to consider how many URM admits are admitted primarily because of racial preferences. That is, if racial preferences were removed, they would have been more likely to be rejected than admitted.

Because UNC admissions utilize, at the margins, some information or impressions that we do not directly observe in the data—and which we refer to as "unobserved factors"—we cannot perfectly predict which individuals would or would not be admitted if racial preferences were removed. However, we do know the range of unobserved factors the applicant must have had in order to be admitted when racial preferences were in place. With this information and the estimates of the model, it is possible to calculate the probability that the applicant would have been admitted absent racial preferences.[27] The strength of this approach is that, since it only focuses on admitted students, our method takes into account that the applicant's unobserved characteristics were strong enough to gain them admission when racial preferences are present.

I have used my preferred model—as well as Professor Hoxby's preferred model—to calculate the probability of admission with racial preferences "turned off" for those African-American and Hispanic applicants who were admitted when racial preferences were in place. Note that since all these students were in fact admitted, their probability of admission with racial preferences is "1" by definition (that is, 100%). If, without racial preferences, I estimate that a given person has, for example, a 40% probability of admission, then that equates to a 60 percentage point decline. Table 3.1 reports results from this analysis. First, I report the average

[27] I show the mathematical formulas for these calculations in Appendix B.2.

probability of admission (under both my model and Professor Hoxby's) for various cohorts, such as in-state African-American applicants. Second, I report the proportion of students in that cohort whose probability of admission falls by at least 50 percentage points (i.e., from "1" to less than "0.5") when racial preferences are turned off.

Table 3.1: How admissions probabilities would change for those admitted under racial preferences if racial preferences were turned off.

	In-State		Out-of-State	
	African American	Hispanic	African American	Hispanic
My preferred model				
Average admit probability	57.8%	75.8%	8.7%	29.2%
Share with greater than 50% drop	42.7%	21.9%	94.6%	78.4%
Hoxby preferred model				
Average admit probability	58.5%	73.9%	32.4%	49.8%
Share with greater than 50% drop	36.9%	12.7%	84.9%	56.7%

Note: Admit probabilties when racial preferences are present for this group is (definitionally) 1. Assumes race preferences are turned off for each applicant; capacity constraints are not imposed.

For in-state applicants, if racial preferences were removed, Hispanic admits would see an average probability of admission of 75.8%—a 24.2 percentage point decrease (because an actual admit's probability of admission is 100%). The share of Hispanic admits who would see a drop of more than 50% in their probability of admission is 21.9%. For in-state African-American admits, removing racial preferences results in a much larger change. The average probability of admission would be 57.8%, a 42.2 percentage point decrease. And the share of African-American admits who would see a drop of more than 50% in their probability of admission is 42.7%.

But especially striking are the results for out-of-state admits. For out-of-state Hispanic admits, removing racial preferences would result in an average probability of admission of 29.2%, a *70.8 percentage point decrease*. The share of Hispanic admits with a greater than 50% point decrease in their probability of admission is 78.4%. And the corresponding numbers for African Americans are even larger: being treated as white would result in an average probability of admission of only *8.7%, a 91.3 percentage point decrease* given that they were admitted when racial preferences were in place. And the share of African-American admits with at least a 50% drop in their probability of admission is *94.6%*.

Note that these numbers are entirely consistent with the results I showed for the average marginal effects of race. For example, in my rebuttal report, I showed that treating out-of-state African-American applicants as white applicants would decrease the African-American admit rate from 17% to 1.5%. See Arcidiacono Rebuttal 30, Table 3.3. These dramatic changes can only occur because the vast majority of out-of-state African-American admits would become likely rejects in the absence of racial preferences.

Finally, Professor Hoxby's models *also* show substantial effects of racial preferences despite her models suffering from omitted variable bias. Performing similar calculations using her preferred model shows that 85% of out-of-state African-American admits would see their admissions probabilities fall by more than 50 percentage points.

3.2 Professor Hoxby's use of the Pseudo R-square (and corresponding Shapley decomposition) is inappropriate for measuring whether race is determinative in admissions.

Professor Hoxby's preferred way of evaluating the magnitude of the effect of race in the admissions process is to take a measure of fit of the admissions model as a whole (the Pseudo R-square) and then calculate the share of the measure of fit that is due to race (the Shapely decomposition). Her argument is that if race explains a small share of admissions decisions, then race is not determinative.

This argument is incorrect for at least two reasons. First, although Professor Hoxby continues to refer to her measure of fit as the R-square, her actual measure of fit is the *Pseudo* R-square.[28] As I explained in my rebuttal report, the two are not equivalent, a fact that is recognized in introductory econometrics textbooks. It is simply incorrect to refer to the Pseudo R-square as the amount of variation explained by the model. Therefore, any decomposition of the Pseudo R-square to make a claim regarding the share of variation explained by race is also incorrect.

Indeed, there are actually *many* different Pseudo R-square measures in the economics literature. The one Professor Hoxby and I use is McFadden's Pseudo R-square. But there are many others, each of which has different properties and different relationships to the R-square.

[28] See, e.g., Hoxby Rebuttal 8 ¶ 17.

Veal and Zimmerman (1996) investigate the various Pseudo R-square measures in the literature and show how they relate back to the R-square used in standard linear models. Of the six measures they evaluate, McFadden's Pseudo R-square is actually the second least related to the R-square, with McFadden's Pseudo R-square always substantially lower than the corresponding R-square.[29] For example, the authors note that their experiments show a Pseudo R-square of 0.25 is associated with an R-square of 0.5, twice as high.[30]

Second, even assuming that her incorrect method was appropriate, the impact of race as a whole on every applicant in the admissions process is not the relevant metric for evaluating whether race is determinative in admissions. A simple example illustrates this point.

Suppose that a particular racial group makes up 1% of applicants. Suppose further that for this group *all* applicants are admitted but all other groups have a 50% chance of admission. Clearly, being a member of the racial group that is automatically admitted *is* determinative for admission. Yet,

---

[29] See Michael R. Veal, Klaus F. Zimmerman, *Pseudo-R Measures for Some Common Limited Dependent Variable Models*, Journal of Economic Surveys, Vol. 10, No. 3, at 249, 241-59 (1996). To link back to the R-square, they use simulation so that they know what the latent index is that dictated whether the outcome occurred. The R-square is then based on the latent index fitted to the observed covariates.

[30] See Michael R. Veal, Klaus F. Zimmerman, *Pseudo-R Measures for Some Common Limited Dependent Variable Models*, Journal of Economic Surveys, Vol. 10, No. 3, at 250 (1996).

what would the Pseudo R-square be for this model? The answer is less than 0.01.

Of course, had the size of the group been larger, the Pseudo R-square would have been higher. But evaluating whether race is determinative *for a particular racial group* should be based on how racial preferences affect *admit rates for that particular group*.

3.3 Professor Hoxby's use of capacity constraints is incorrect for measuring whether race is a determinative factor.

Professor Hoxby then proceeds to say that racial preferences are not as large as I claim because when I speak of the magnitude of racial preferences, I do so not taking into account capacity constraints. Professor Hoxby notes correctly that if all applicants were treated as African Americans (or, alternatively, Hispanics), UNC would have to lower its admit rates for everyone or its admitted class would be too big.[31] Professor Hoxby argues that, even though there are substantial average changes in admissions probabilities if an Asian-American applicant were treated as an African-American applicant, if *all* applicants were treated as African-American

[31] In fact, this is what I actually showed in my discussion of capacity constraints. Note further that, in the presence of capacity constraints it does not matter *which* race everyone is switched to. All members of all races will be treated the same with capacity constraints implying that whether we switch everyone to African American or everyone to Hispanic will make no difference. Professor Hoxby does not recognize this point, and actually shows in Exhibit 2 Figure 4a of her rebuttal report *different* estimates of admission rates scaled for capacity constraints depending on whether everyone's race is switched to African American or Hispanic. In fact, what this shows is that she has not imposed a tight enough criteria for the capacity constraint.

applicants then admission rates for Asian-American applicants would not change very much because of capacity constraints. That is, treating everyone as African American would result in admitted classes that are too large so UNC would need to admit fewer students.

But this argument is entirely misleading. To see this, consider again the example where a racial group that represents 1% of the applicant pool is admitted 100% of the time while all other applicants are admitted 50% of the time. Removing the racial preference in this case would result in everyone being admitted at a rate of 50.5%, only a 0.5% increase for non-preferred groups. But the admit rate for the preferred group would have fallen by *49.5%*. Professor Hoxby's argument completely misses how racial preferences affect admissions probabilities *for preferred groups,* which is the key consideration in assessing whether race is determinative.

As I discussed in Section 3.1, I do not believe imposing capacity constraints is appropriate in evaluating how *individual* applicants benefit or are hurt based on UNC's racial preferences. Nonetheless, my analysis with capacity constraints shows *substantial* effects of the removal of racial preferences on the admitted class. For in-state admissions, Table 4.4.R of my rebuttal report shows that the number of Hispanic admits are 21% higher as

a result of racial preferences; for African Americans the number of admits is 55% higher.[32]

As shown in Table 4.5.R of my rebuttal report, the effect of racial preferences on out-of state URM admits is much higher. Racial preferences result in the number of Hispanic admits being *2.5 times higher* than the number of admits absent racial preferences. And for African Americans, the number of admits with racial preferences is *7.5 times higher* than the number of admits absent racial preferences.

3.4 Professor Hoxby's decile analysis actually confirms that UNC's racial preferences for URMs are enormous.

But perhaps where Professor Hoxby is most misleading is in her decile analysis. As background, I constructed an academic index that was a weighted combination of an applicant's SAT scores and high school grades. Separately for out-of-state and in-state admits, I then split applicants into ten groups based on the values of their academic indexes. Note that, by construction, the size of the groups was exactly the same for each in-state decile (and for out-of-state deciles).

In Tables 3.3 and 3.4 of my opening report, I showed how admit rates varied across races by academic index decile for both in-state and out-of-state applicants. These tables showed substantial differences in admit rates across

[32] From Table 4.4.R, the number of Hispanic admits with and without racial preferences is 1,470 and 1,212 respectively. The percentage point increase resulting from racial preferences is calculated by (1470/1212)-1.

racial groups within the same academic index decile. To illustrate how the admit rates vary by race, consider the fourth decile.

- For in-state applicants, the admit rates for Asian-American and white applicants was 17.1% and 17.2%, respectively.
- The corresponding admit rates for in-state Hispanic and African-American applicants was 38.0% and 48.6%, respectively.
- For out-of-state applicants, the admit rates for Asian-American and white applicants was 0.9% and 1.4%, respectively.
- The corresponding admit rates for out-of-state Hispanic and African-American applicants was 15.6% and 39.2%, respectively.

So in the fourth decile, admit rates for in-state URM applicants are at least twice as high as the admit rates for non-URMs. And for out-of-state applicants the admit rate for Hispanics is more than 10 times as high as the admit rate for non-URMs; for African Americans the admit rate is more than *28 times higher* than for non-URMs.

Note that in my initial report, I was quite clear that there are other factors that matter in admissions besides this academic index (see pages 38-39 of my opening report). This academic-index analysis was designed simply to show how large those other factors must be in order to explain the large differences in academic qualifications across races.

Professor Hoxby attempts to counter this analysis by sleight of hand. She states as follows:

> In fact, within each in-state decile and each out-of-state decile, setting the URM and non-URM admissions rates to be equal would never cause the number of non-URM admits to rise *by even 1 percent* relative to all admits.

Hoxby Rebuttal 13. The comparison Professor Hoxby is making is ridiculous. Simply put, one would never expect large changes in overall admits from changing the admit rates in only one of twenty deciles (ten for in-state, ten for out-of-state)—no matter how determinative race was in that decile. Here she is equalizing admit rates in one of twenty deciles to say that the number of admits across *all* non-URM applicants would increase by less than one percent. To keep the math simple, suppose the number of applicants were the same in-state versus out-of-state. This means that 5% of applicants would be in each cell. Within a particular cell, the *maximum* effect of equalizing admit rates would occur if the overall non-URM admit rate was 100% in that cell and the non-URM admit rate was 0% *and* there were an equal number of URM and non-URM applicants in the decile. In this case, the admit rate for the decile would be 50%.[33] But there would only be 5% of non-URMs in this decile, implying that the largest possible effect on the overall admit rate would be 2.5% (50% times 5%, or 50% divided by twenty deciles).

But even this number is much too high because, by definition, there are fewer URM applicants than non-URM students. And this is the basic tactic Professor Hoxby takes: making any removal of racial preferences appear small by spreading the effect of that removal across a much larger group. Indeed, this is the tactic she takes in all her figures associated with

[33] Professor Hoxby equalizes admit rates within deciles by setting the admit rate for each racial group to be the same as the total admit rate for that decile.

Exhibit 2, attempting to mask the large changes for URM applicants by focusing on the changes for non-URM applicants.

Her next statement continues with this tactic:

> Considering all the in-state and out-of-state deciles together, and setting the URM and non-URM admissions rates to be equal would cause the number of non-URM admits to rise by only 7 percent of total admits. Even if we were to assume that the Arcidiacono Index should be the sole basis for admission, his evidence still does not demonstrate that race/ethnicity is a dominant factor in admissions.[34]

Hoxby Rebuttal 13.

There are two issues here. First, what she does not show is the effect on the number of URM admits. Again, consider the hypothetical in which one URM racial group is admitted 100% of the time but makes up only 1% of the applicant pool. Equalizing admit rates for URMs and non-URMs within a decile would have only a small effect on the number of non-URM admits. This is because within a decile non-URMs would substantially outnumber URMs. When admission rates are set to the total admit rate for the decile, the primary driver of the total admit rate is non-URMs. Yet, race was obviously determinative for this hypothetical URM group. And the change in admissions probabilities for URMs would be substantial.

So what is important is not what happens to the within-decile non-URM admit rate but happens to the within-decile *URM* admit rate. Using Professor Hoxby's calculations of equalizing admit rates within deciles results

[34] Professor Hoxby refers to my academic index as the "Arcidiacono Index."

in a 30% drop in URM in-state admits and a 57% drop in out-of-state admits.[35]

How can these effects be so big given that Professor Hoxby points out that 75% of in-state admits come from the top four deciles and racial preferences are smaller here (see Hoxby Rebuttal 15 ¶ 32)? The reason is that only 37% of in-state *URM* admits come from the top four deciles.

Second, the last line of Professor Hoxby's statement suggests that these are the changes in total admits if "the Arcidiacono index should be the sole basis for admission." But this is incorrect. Other factors also matter even when admit rates across racial groups are equalized within a decile. This is why some applicants are rejected in the top deciles and some are admitted in the bottom deciles. Strictly applying the academic index results in larger changes as I illustrated in my opening report. Using the numbers from Tables 3.6 and 3.7 of my opening report shows that the total number of in-state (out-of-state) URM admits would fall by 45% (63%) if admissions were based solely on the academic index.[36]

Because UNC admissions officers consider factors outside of the academic index, the academic-index-decile analysis is insufficient by itself for

[35] Professor Hoxby's 7% calculation is for the total number of non-URM admits. This masks differences between out-of-state and in-state applicants as the racial preferences are much larger in the out-of-state pool. Applying Professor Hoxby's calculations in-state results in a 5% increase in non-URM admits; for out-of-state applicants, the increase is 21%.

[36] Even these figures mask differences between African-American and Hispanic applicants. For example, the number of African-American out-of-state admits would fall by 86% under this criteria.

showing that differences in admissions rates are the result of racial preferences. However, it illustrates the potential for racial preferences to play a large role in the admissions decisions of many students. And this is exactly what the statistical modeling shows.

Indeed, the academic-index-decile analysis discussed above substantially understates the role of racial preferences. Professor Hoxby inadvertently shows this herself in Exhibit 2, Figures 3A and 3B. Now, rather than use an index based solely on test scores and grades, she uses the index that underlies my model 4. This index includes all the variables in model 4, including UNC's subjective ratings and demographic information. My estimates in model 4 allow me to rank applicants by how their characteristics translate into admissions chances. The weights placed on these characteristics come from the model estimates; UNC "reveals" what these weights are through their admissions decisions.[37] These weights are then used to form this admissions index.

Professor Hoxby's figures then show the status quo number of admits for URM and non-URM applicants by deciles of this index and then compares these to what the number of admits would look like if admit rates were equalized within the decile; that is, if the admit rates for each racial group were set to be the same as the total admit rate for the decile. In making this

[37] Note that there are still unobserved characteristics but, as I showed in my rebuttal report (see pages 32-34), the role of these characteristics is very small when compared to racial preferences.

comparison, Professor Hoxby employs the same tactic as before, which conceals the true effect that removing racial preferences has on admissions decisions; she improperly focuses solely on the effect on the number of non-URM admits in any given decile. Again, because there are many more non-URM applicants, the increase in admits is spread out across a larger pool of students.[38]

But what Professor Hoxby fails to report here is the large change in the number of admits for URM students. Similar to her previous decile analysis figures, Professor Hoxby does not report the actual number of admits in her figures. But these numbers are shown in her replication files. In Table 3.2, I show the numbers underlying her Exhibit 2, Figure 3b.

Table 3.2 Number of out-of-state admits and admit rate by race and admissions index decile, averaged across six admission cycles.

	Admission Rates			Actual Admits			Admits with Equal Admission Probabilities		
Decile	URM	non-URM	Total	URM	non-URM	Total	URM	non-URM	Total
1	0.0%	0.2%	0.1%	0	2	2	1	1	2
2	0.9%	0.2%	0.4%	5	2	6	2	4	6
3	3.9%	0.2%	1.0%	15	2	17	4	13	17
4	11.5%	0.3%	2.4%	35	4	39	7	32	39
5	21.8%	0.5%	3.8%	55	8	63	10	53	63
6	30.6%	0.6%	4.7%	68	9	77	10	66	77
7	45.1%	1.4%	6.6%	88	20	108	13	95	108
8	65.9%	4.9%	10.9%	106	72	177	18	160	177
9	83.1%	22.0%	27.2%	115	327	441	38	404	441
10	96.8%	77.2%	78.3%	86	1,164	1,250	69	1,180	1,250
Total	18.6%	12.1%	13.3%	571	1,609	2,180	171	2,008	2,180

Note: Admissions deciles from Arcidiacono model 4. These are the numbers underlying Exhibit 2 Figure 3b of Professor Hoxby's rebuttal.

[38] Note that this exercise should in no way be interpreted as the effect of removing racial preferences. Removing racial preferences would result in more admits from those non-URMs in the top deciles, not shifting admits from URMs in low deciles to non-URMs in low deciles.

Consider the first two columns that show the admissions rates in the different deciles for URM and non-URM students. What these columns show is the probability that whatever qualities are not accounted for by the observables will be so strong that UNC will admit them. In each of the bottom six admissions deciles, the chances of being so strong on these unobserved factors that UNC admits the applicant are less than 0.7% for non-URMs. Yet URM applicants in the sixth decile have a 30.6% chance of admission; the URM probability of admission in this decile is *50 times* higher than the non-URM probability of admission. So despite there being 6.4 times more non-URM applicants in this decile than URM applicants, there are 7.8 times more URM admits in this decile.

It is of course true that in the very top and very bottom deciles the effects are smaller. Certain applicants have characteristics that are so strong or so weak that their race is not relevant. That is why it is also important to look at the total effect of removing race. Under Professor Hoxby's approach of equalizing admissions rates within deciles the total effect on the number of out-of-state URM admits over the six-year period is a drop of 2,399 students, *a 70% decrease.*[39] There is no sense under which this effect is modest.

3.5 Professor Hoxby's criticism of Mr. Kahlenberg's SES preferences inadvertently shows that UNC's racial preferences for URMs are enormous.

[39] I say Professor Hoxby's approach because I never undertake the analysis she does here. That said, the underlying model used to generate these findings is model 4 of my opening report.

What is especially ironic about Professor Hoxby's claims regarding the size of racial preferences is that she criticizes Mr. Kahlenberg's SES preferences for being extremely large when, in fact, the SES preferences Mr. Kahlenberg proposed are *smaller* than some of the racial preferences UNC currently employs. With regard to the size of the SES preferences, Professor Hoxby states:

> Even a single Kahlenberg Bump is worth a great many SAT points: 278. For instance, this would change a student with a score of 1000, who would ordinarily not have much chance of admission, into a student with a score of 1278, well into UNC's normal admit range.

Hoxby Rebuttal 58, ¶ 140.

She further states:

> Translating the Kahlenberg Bump into GPA points (where A=4, B=3, etc.) shows even more dramatic results. In fact, the Kahlenberg Bumps are so large that they are equivalent to impossibly large increases in GPA. For example, just one Kahlenberg Bump is equivalent to 4.41 GPA points, which is essentially impossible to achieve (GPA points usually range from 1 to 5).

Hoxby Rebuttal 59, ¶ 141.[40]

Table 3.3 does the same translation of preferences to SAT score points, or alternatively GPA and percentile ranking points, that Professor Hoxby uses but this time for the preferences that UNC currently employs. Just like

[40] The GPA conversion is misleading because if an applicant's GPA was increased this would increase other factors that would also raise the applicant's chances of admission. For example, the applicant's percentile ranking in the class and the applicant's score on UNC's performance rating will both mechanically rise with increases in GPA. I make the conversion of racial preferences to GPA point solely to illustrate what Professor Hoxby's methodology implies about the size of racial preferences.

Professor Hoxby, I use the estimates from the 2019 admissions cycle. I take the estimated coefficient (or coefficients) from the 2019 cycle associated with a particular preference and then translate the coefficient to the corresponding points applying Professor Hoxby's conversion formula.

Table 3.3: Total preferences African-American applicants receive in admissions translated into SAT, GPA, and class rank.

		In-state			Out-of-state		
		SAT points	GPA points	Percentile	SAT points	GPA points	Percentile
Not FGC	Male	224	3.56	58.3	386	6.12	100.2
	Female	191	3.03	49.7	369	5.85	95.8
FGC	Male	249	3.94	64.6	414	6.56	107.4
	Female	216	3.42	56.0	397	6.29	103.0

Note: Percentile refers to high school class rank. Point conversions follow Professor Hoxby's Exhibit 7 using the 2019 admissions cycle in-state model. Using the out-of-state model produces similar findings except for the conversion of GPA that produces much larger effects.

The first row shows that, using Professor Hoxby's conversion, African-American males who are not FGC receive a preference equivalent to 386 points if they are out-of-state applicants.[41] In other words, they receive a preference that is 108 points *higher* than the "Kahlenberg Bump."

Converting this preference instead to GPA points further underscores the magnitude of the preferences African-American applicants receive. For example, out-of-state African-American males who are not FGC receive a racial preference equivalent to 6.12 GPA points.

[41] As shown in the second row, the numbers are slightly lower for females, likely due to the fact that African-American men are more underrepresented in the applicant pool than African-American women.

African American males who are FGC get a preference both for being African American and also for being FGC.[42] Hence, the equivalent SAT score and GPA points are even higher. In particular, converting the preferences for this group to SAT scores points results in 249 points for in-state applicants and 414 points for out-of-state applicants. Alternatively converting the preferences for this group to GPA points results in 3.94 points for in-state applicants and the 6.56 points for out-of-state applicants. Again, Professor Hoxby's criticism of Mr. Kahlenberg's analysis thus serves to underscore the massive racial preferences UNC gives to URM applicants.

3.6 Standard assumptions in the economics literature imply that my models underestimate the magnitude of UNC's racial preferences.

In my initial report, I stated that "my estimates of the magnitude of racial/ethnic preferences are, if anything, understated." Arcidiacono Report 56. My reasons were two-fold. First, non-URM applicants are stronger on observed characteristics associated with higher admissions probabilities than their URM counterparts. Consistent with standard economic assumptions, stronger observed characteristics suggest that non-URM applicants would likely be stronger on unobserved characteristics. As to this first point, Professor Hoxby argues in her rebuttal report that this is speculation. Hoxby Rebuttal 30-31.

[42] Note that the FGC bump African Americans receive from being FGC is smaller than the bump non-URM applicants receive for being FGC.

Drawing any conclusions regarding unobservables is to some extent speculative. That said, the standard assumption in economics is that observables run in the same direction as unobservables. Indeed, multiple econometric papers have been developed around this idea.[43] The case for it here is clear: as my models show, when more controls are added to the model, the effects of racial preferences become more important, not less. Hence, it stands to reason that the pattern would continue if additional controls (reflecting the unobservables) were also added. This is why, for example, racial preferences are estimated to continue to increase in magnitude above what my preferred model shows when high school and census tract fixed effects are added.

Professor Hoxby argues against this logic, stating that for someone to apply who was weak on observed characteristics it would be because they were strong on unobserved characteristics. First, this only holds for someone close to the margin of applying to UNC, not for the applicant pool as a whole. More importantly, the logic breaks down when substantial racial preferences are present. If URM applicants receive a preference in admissions then they will be willing to apply with worse observed *and* unobserved characteristics

[43] See, e.g., Joseph Altonji, et al., *An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling*, Journal of Human Resources (2005); Joseph Altonji, et al, *Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools*, Journal of Political Economy (2005); Emily Oster, *Unobservable Selection and Coefficient Stability: Theory and Evidence*, Journal of Business & Economic Statistics (2015); Brian Krauth, *Bounding a Linear Causal Effect Using Relative Correlation Restrictions*, Journal of Econometric Methods (Aug. 2011).

than their non-URM counterparts. Indeed, Professor Hoxby actually makes this very same point in criticizing Mr. Kahlenberg's race-neutral plans. See Hoxby Rebuttal 32-34. Namely, Professor Hoxby argues that increases in preferences for a group result in weaker members of the group applying. The same logic applies to racial preferences that induce weaker URM students to apply.

My second argument for why I am likely underestimating the magnitude of racial preferences is the evidence that URM students receive a preference in the personal quality rating. Indeed, deposition testimony suggests race is taken into account in forming this rating.[44] And Professor Hoxby herself cites additional evidence that race is one of the factors taken into account in the formation of the personal quality rating.[45]

My findings on personal quality are similar to what I find for admissions. Namely, as controls were added to the model of the personal quality rating, the coefficient on African American and Hispanic became larger and larger. Consistent with this, I showed that non-URM applicants were stronger on the observed characteristics associated with higher personal quality ratings. And within URM applicants, Hispanics had observed characteristics associated with higher personal quality ratings yet did not receive as large of a preference as African Americans.[46]

---

[44] See, e.g., Rosenberg Depo. 250:13-251:8; Perkins Depo. 42:7-43:10.

[45] See Hoxby Report 10 & n.37 (citing UNC0326625).

[46] Arcidiacono Rebuttal, Tables A.5.1.R, A.5.2.R, and A.5.3.R.

Professor Hoxby again makes the argument that this is speculation. Here it is important to be clear how large the estimated coefficients are on the race variables. For example, the coefficient on African American is 0.81 in the model of the personal rating for out-of-state applicants.[47] Translating this into standard deviation units means that African-American applicants would on average need to be 0.45 standard deviations stronger on these unobservable characteristics than white applicants. This is a very large effect. Given the deposition testimony and the fact that non-URM applicants are stronger on the *observed* characteristics associated with higher personal rating, it is difficult to reconcile this large effect outside of race playing a substantial role in the personal rating.

[47] Arcidiacono Rebuttal, Table A.5.2.R.

Dated: June 8, 2018 */s/Peter S. Arcidiacono*
Peter S. Arcidiacono

Appendix A

Table A.1: Average marginal effect of race when personal quality rating is removed

	Average Admission Probablity with Racial Preferences	without Racial Preferences	Marginal Effect of Race	Share due to Racial Preferences
In-State				
African American	30.5%	17.6%	12.9%	42.2%
Hispanic	41.0%	30.9%	10.1%	24.6%
Out-of-State				
African American	17.0%	1.4%	15.7%	91.9%
Hispanic	20.2%	5.8%	14.5%	71.5%

Note: Estimated using model 4 of Arcidiacono rebuttal report with the personal quality rating removed

Table A.2: Accuracy of models when special recruiting categories are included

	Accuracy on Admits	Accuracy on Rejects	Overall Accuracy
In-State			
Model 3	91.5%	92.2%	91.8%
Model 3 including specials	92.1%	91.5%	91.8%
…plus special Indicator	92.7%	92.1%	92.4%
Model 2	86.3%	87.4%	86.9%
Model 2 including specials	87.7%	86.9%	87.3%
…plus special Indicator	88.2%	87.4%	87.8%
Out-of-State			
Model 3	75.2%	96.1%	93.3%
Model 3 including specials	76.4%	95.3%	92.1%
…plus special Indicator	80.5%	96.1%	93.5%
Model 2	62.0%	94.1%	89.7%
Model 2 including specials	64.8%	93.0%	88.3%
…plus special Indicator	70.1%	94.0%	90.1%

Note: Analysis is for the models reported in Hoxby rebuttal Exhibit 5 plus an additional model that includes an indicator variable for special recruiting category. These models are from my opening report.

Table A.3: Comparison of High School GPA and Performance Rating by race and missing GPA status

	Grade Point Average		In-State Performance Rating		Out of State Performance Rating	
Ethnicity	In-State	Out-of-State	GPA Present	GPA Missing	GPA Present	GPA Missing
White	4.43	4.15	7.01	7.27	7.47	7.59
African American	4.08	3.84	5.42	5.59	5.70	5.85
Hispanic	4.26	4.16	6.10	6.03	6.71	7.01
Asian	4.47	4.17	6.56	7.54	7.17	7.33

Note: GPA and GPA missing are defined according to Arcidiacono opening and rebuttal reports. Data are from the 2016 to 2021 admission cycles Arcidiacono estimation samples.

Table A4: Model 7 estimates using 9-digit and 11-digit census tracts

Variable	9-Digit Census Tracts	11-Digit Census Tracts
African American	4.729	5.142
	(0.178)	(0.195)
Hispanic	2.605	2.822
	(0.205)	-0.223
Asian American	0.176	0.113
	(0.139)	(0.151)
female	0.172	0.185
	(0.064)	(0.069)
FGC	1.298	1.406
	(0.094)	(0.103)
Regular Admission	-0.353	-0.377
	(0.058)	(0.063)
alum	0.426	0.452
	(0.071)	(0.076)
waiver	0.2	0.234
	(0.100)	(0.111)
female * race		
African American	-0.624	-0.741
	(0.178)	(0.193)
Hispanic	-0.201	-0.359
	(0.219)	(0.238)
Asian American	-0.283	-0.248
	(0.163)	(0.175)
FGC * race		
African American	-1.434	-1.468
	(0.191)	(0.213)
Hispanic	-0.456	-0.367
	(0.232)	(0.253)
Asian American	-0.125	-0.094
	(0.202)	(0.223)
Academic variables	X	X
Ratings variables	X	X
Heterogeneity variables	X	X
HS fixed effects sample	X	X
HS fixed effects model	X	X
Census tract fixed effects	X	X
Observations	38,870	38,736
Pseudo R2	0.771	0.786

Note: First column gives the results for in-state model 7 from Arcidiacono rebuttal report. Second column uses indicators for 11-digit census tracts rather than 9-digit basic census tracts.

Appendix B

B.1 Overfitting

A fundamental problem with assessing out-of-sample error is that some data must not be used in the estimation of the model. Generally, more data is better, so leaving out data for testing purposes will yield a less accurate model, all else equal. However, without data "new" to a model, the researcher cannot assess out-of-sample error rates.

To address this tension, researchers most frequently turn to "cross-validation." One of the most common forms of cross-validation is "*k*-fold cross-validation." The basic idea with *k*-fold cross-validation is to randomly divide all of your data into *k* "folds," or partitions. Then, estimate the model on (*k*-1) of the folds and calculate out-of-sample mean-squared error on the 1 fold not used. Repeat this process *k* times, cycling through so each fold of the data is the test set once. The result will be *k* estimates of the out-of-sample mean-squared error that can then be averaged. Common choices of *k* are $k = 5$ and $k = 10$. These values of *k* have been shown to have the best properties for estimating out of sample error. These correspond to out-of-sample test sets that are 20% and 10% of the total amount of data, respectively, implying that 80% or 90% of the data is used to estimate the model.[1]

[1] See James, G., Witten, D., Hastie, T., and Tibshirani, R. (2015) "An Introduction to Statistical Learning, with Applications in R," Section 5.1, pg 184-186.

B.2 Derivation of admission probabilities for admitted URM students when race is turned off

In this appendix I consider URM applicants who were actually admitted to UNC and show how the admissions model can be used to calculate their probabilities of admissions had there been no racial preferences at UNC. I derive the formula for these admission probabilities which follows directly from Bayes rule. Let $y = 1$ if a URM applicant was admitted in the status quo environment (i.e. the environment with racial preferences). Let $y' = 1$ if the URM applicant would have been admitted without racial preferences. Let X denote the observed characteristics of the applicant. Since we do not see the applicant's unobserved characteristics, we can only form a probability that the applicant would be admitted in the absence of racial preferences. Let the conditional probability that the applicant would be admitted absent racial preferences given that the applicant was admitted in the status quo environment be given by $P(y' = 1|y = 1, X)$. Then, using Bayes' rule, we can express this as:

$$P(y' = 1|y = 1, X) = \frac{P(y = 1|y' = 1, X)P(y' = 1|X)}{P(y = 1|X)} \quad (1)$$

$$= \frac{P(y' = 1|X)}{P(y = 1|X)} \quad (2)$$

where the second line follows because if the URM applicant would have been admitted without racial preferences then the probability of being admitted with racial preferences is one. The reason it is one is because turning off racial preferences means it is harder for URM applicants to be admitted. Hence if the URM applicant would have been admitted without racial preferences then the applicant would have been admitted in an environment with racial preferences.

Both of the terms on the right hand side are known as we can calculate them using the logit formula. The term in the denominator effectively adjusts for the fact that the applicant had unobservables that were good enough to lead to admission in the status quo case. This term is the predicted probability of admission taken directly from the model estimates. The term in the numerator is the same predicted probability but calculated as though the applicant were white.