Reducing Frictions in College Admissions: Evidence from the Common Application*

Brian Knight† Nathan Schiff‡

December 23, 2020

Abstract

College admissions in the U.S. is decentralized, creating frictions that limit student choice. We study the Common Application (CA) platform, under which students submit a single application to member schools, potentially reducing frictions and increasing student choice. The CA increases the number of applications received by schools, reflecting a reduction in frictions, and reduces the yield on accepted students, reflecting increased choice. The CA increases out-of-state enrollment, especially from other CA states, consistent with network effects. CA entry changes the composition of students, with evidence of more racial diversity, more high-income students, and imprecise evidence of increases in SAT scores.

*We thank seminar participants at American University, Brown University, the National University of Singapore, the University of Utah, the National Tax Association Fall 2018 Meetings, and the Urban Economics Association Fall 2018 Meeting. Peter Blair, John Friedman, and Josh Goodman provided helpful comments, and Juan Pedro Ronconi provided research assistance.

†Brown University

‡Shanghai University of Finance and Economics

1 Introduction

College admissions in the United States has traditionally followed a decentralized process, with students completing separate applications for each school and colleges independently making admissions offers.[1] There are advantages and disadvantages to the frictions associated with this decentralized system. By limiting the number of schools to which students apply, decentralized systems might ultimately limit the degree of student choice and create a less integrated and competitive market. At the same time, a more centralized system might increase stratification. In particular, with heterogeneous students and institutions, students with high test scores might be more likely to apply to and ultimately attend elite out-of-state institutions, rather than local institutions, under a more centralized system, and a similar logic applies to sorting according to any student attribute that is valued by institutions. In this paper, we investigate a movement towards greater centralization created by the Common Application (CA), a consortium of colleges that accept a single application. The CA began in 1975 with 15 liberal arts colleges but grew rapidly thereafter, with a significant acceleration of membership starting around 2000 and roughly 700 members by 2016 (Figure 1). Member institutions are disproportionately selective, as documented below, with nearly all elite private universities currently members and potential implications for stratification.

In this paper, we ask a series of research questions. Has the CA reduced frictions, resulting in more college applications at member institutions and increased student choice? Has the CA led to a more geographically integrated market, with more students attending CA institutions far from home? If so, by increasing student choice and integrating the market geographically, has the CA contributed to stratification, a widening of the selectivity gap between more selective and less selective institutions? Related to this, has the CA altered racial and socioeconomic diversity in higher education?

We address these questions using panel data from the College Board covering 1990-2016. We estimate fixed effects regression models, comparing outcomes for schools before and after joining the CA, and also provide event studies, investigating the timing of any effects associated with entry into the CA. Overall, we find that the CA increases applications, consistent with a reduction in frictions, and reduces yield, consistent with enhanced student choice. Schools respond to this

[1] These time-based frictions are in addition to any information frictions facing students as a result of decentralized college admissions.

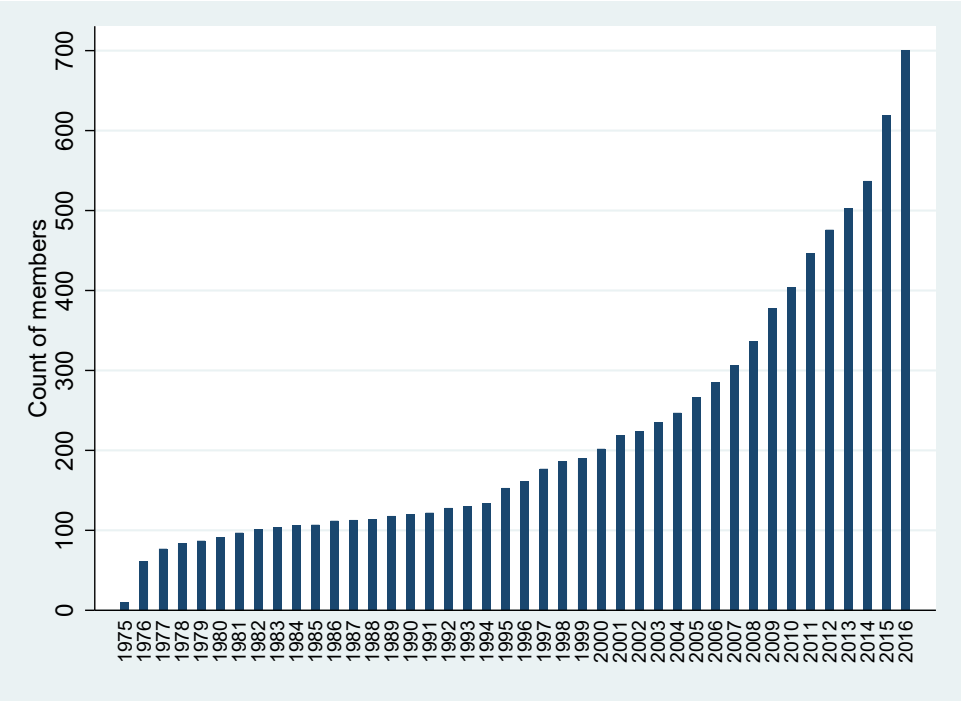


Figure 1: Common Application Membership by Year

Notes: Figure plots total count of schools using the Common Application in each year. School membership is based on most recent entry year, provided by The Common Application organization.

reduced yield by admitting more students. Turning to the composition of students, we find that the CA has accelerated geographic integration: entry into the CA is associated with an increase in the fraction of out-of-state students, especially from other states with significant CA penetration, consistent with network effects. Finally, we investigate three measures of student heterogeneity. We provide some evidence that entry into the CA is associated with an increase in SAT scores. While these results are imprecise, we provide stronger evidence that CA adoption increases the fraction of non-white students and reduces the fraction of low-income students. Given that, prior to joining, CA colleges tend to have fewer non-white students, fewer low-income students, and higher test scores, the CA has increased stratification according to income and tests scores but has reduced stratification according to race.

The most closely related paper in the literature is Liu et al. (2007), who also use panel data from the College Board to study how CA membership affects admissions outcomes and the composition of enrollees at public institutions over the period 1975-2005. Our paper focuses on a more recent time period, 1990-2016, allowing us to study both the rapid growth in the platform in recent years

and the role of private institutions, which were first allowed to join in 2002.[2][3] In addition, we develop new identification strategies focused on addressing pre-trends.[4] Two other studies have examined the impact of the CA on college admissions. Smith (2013) studies the effect of the number of applications on enrollment probabilities using variation induced by adoption of the CA by nearby colleges. He finds that increasing the number of applications, when induced by the CA, significantly increases enrollment probabilities. Smith et al. (2015) analyze various frictions in the application process, finding some evidence that the CA increases applications. Fees and essay requirements, by contrast, decrease applications. Our paper addresses different research questions, and, as such, the effect of the CA on college admissions outcomes remains up for debate.

More broadly, our paper relates to four other literatures on college admissions. The first examines specific policies that make it easier to apply to college.[5] In this paper, we focus on a different change in the college admissions process, namely a reduction in the complexity and time associated with applying to multiple colleges. A second literature focuses on information frictions associated with decentralized admissions.[6] Given that the CA reduces the time cost associated with submitting applications to multiple CA schools, we interpret the CA as reducing time-based

[2]As noted above, the CA grew from under 300 members in 2005 to roughly 700 in 2016,

[3]Perhaps due to these changes in the CA over time, our effects regarding admissions outcomes tend to be larger in magnitude than Liu et al. (2007). In particular, we find that applications rise 12 percent and that yield falls 8-9 percent, whereas Liu et al. (2007) find that applications rise 6 percent and that yield falls by 3 percent. We also find different results for some outcomes related to student composition, with some evidence of increases in SAT scores and reductions in the fraction of low-income students. An additional contribution involves our focus on geographic integration. In particular, we provide novel findings documenting robust increases in out-of-state enrollment, and we also examine the role of network effects in terms of out-of-state enrollment increases being driven by students from source states with significant CA penetration. This contributes to a literature on trends in geographic integration in higher education, as described below.

[4]In particular, we present event study figures, allowing readers to see both pre-trends and the dynamic effects of CA adoption, and also provide results from an identification strategy that compares outcomes for new CA members to outcomes for schools that will join the CA in the next few years. We argue that joiners are more comparable to this comparison group than to broader comparison groups that include never joiners and schools that join in the more distant future.

[5]Bettinger et al. (2012) show that assistance filling out the FAFSA increases aid receipt, college attendance, and persistence. Bond et al. (2018) documents that the opening or closing of nearby SAT testing centers changes college attendance and graduation.Goodman et al. (2018) documents that many low-income students do not re-take the SAT, even though retakes are both free and associated with increases in test scores.

[6]Hoxby and Avery (2013) document that many low-income but high-ability students do not apply to selective colleges, despite generous financial aid. In an experiment, Hoxby and Turner (2014) and Hoxby and Turner (2015) provide these types of students with information about college admissions and financial aid, increasing applications to and attendance at selective institutions. Gurantz et al. (2019) provided information about selective colleges to low-income and middle-income students but find little evidence of changes in college attendance patterns. Bird et al. (2019) conduct a field experiment using the CA platform, finding that information about financial aid does not increase college attendance.

frictions associated with applications. Yet it is also possible that the CA provides information to applicants through the platform. A third literature has examined how college admissions, and in particular the recent increase in the number of applications per student, has changed both university and student strategic behavior.[7] Our paper contributes towards this literature by examining the contribution of the CA to these recent trends. Finally, a literature has examined the causes and consequences of recent trends towards geographic integration in higher education.[8] We identify the CA as a new potential cause of this trend and examine its consequences for higher education.

The paper proceeds as follows. Following some background information on the CA, we present a theoretical model of the college admissions process and generate our key hypotheses. We then describe the data, empirical approach, and our key empirical results. The final section concludes.

2 Background

The CA was founded by 15 liberal arts colleges in the Northeast but has since expanded to a wide range of public and private institutions, especially more selective institutions. In particular, as shown in Figure 2, membership among the top 50 liberal arts colleges was already very high, over 80 percent, in 1990, the beginning of our analysis, and was universal in this group by the late 1990s. During our sample period, membership among top 50 private institutions increased rapidly, from under 40 percent in 1990 to roughly 90 percent by 2016. Taken together, the CA currently receives approximately 4 million applications from 1 million students annually.[9]

Membership among less selective liberal arts colleges and other private institutions also increased during our sample period but remained below 50 percent in 2016. The CA was originally

[7]Bound et al. (2009) document increases over time in the number of students applying to college, increases in applications per student, and reductions in acceptance rates at selective institutions. Blair and Smetters (2019) investigate why elite colleges haven't responded to this increase in applications by expanding capacity, arguing that colleges compete on prestige. Avery and Levin (2010) document that early applicants are more likely to be admitted and argue that this finding is consistent with an early application serving as a signal of student enthusiasm. Avery et al. (2013) argue that standard methods of ranking colleges provide incentives for institutions to manipulate admissions to reduce acceptance rates and to increase yield.

[8]Historically, the US market for higher education was highly localized, with most students attending universities close to their residence (Hoxby, 2000). During the subsequent decades, the market for higher education became more national, leading to both higher tuition and greater student sorting (Hoxby, 1997, 2000, 2009). Despite this shift, the US market may still be considered local in nature, with roughly 80 percent attending in-state institutions, and Knight and Schiff (2019) and Cohodes and Goodman (2014) identify financial incentives, in the form of in-state tuition discounts and financial aid for in-state institutions, as contributing factors.

[9]From https://www.commonapp.org/about-us, accessed April 2018.

closed to public institutions but that ban was lifted in 2002, leading to a rapid increase in membership among the top 50 public institutions. Less selective public institutions, by contrast, joined at a slower rate, with membership rates still below 20 percent by the end of our sample period.

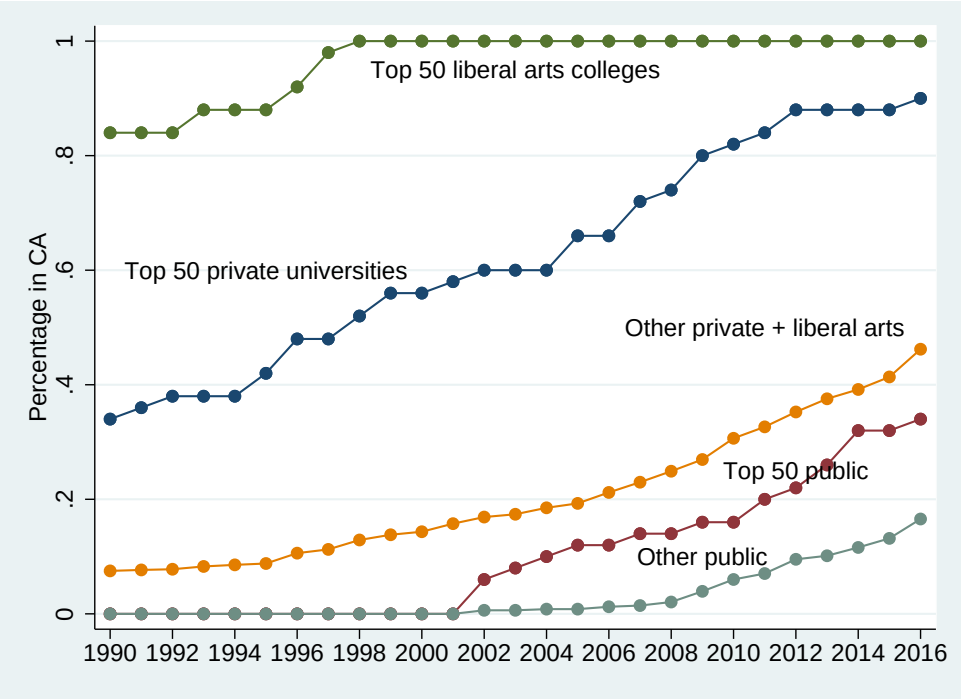


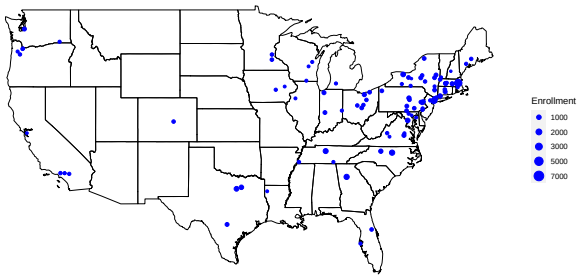
Figure 2: Common Application Membership Rates by Year and Type

Notes: Figure plots percentage of schools in each category using the Common Application. Categorization of schools comes from the data appendix of Bound et al. (2010).

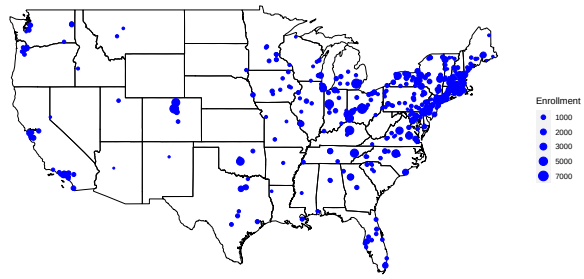
In addition to this rapid entry overall, the CA has also become more diverse from a geographic perspective. That is, the CA started in the Northeast but is now accepted by colleges in many different states. In particular, Figure 3 plots the locations of CA members in 1986 and 2014, documenting a much wider geographic distribution in the latter year, with significant new penetration in states such as California, Oregon, Colorado, Indiana, and Florida. Given this substantial adoption over time and a diverse set of members at current, it is natural that the CA may have led to significant changes in college admissions.

Figure 3: Common Application Membership by State

(a) CA in 1986



(b) CA in 2014



Notes: Each circle shows the location of a school using the Common Application, where the size of the circle depends on total enrollment in that year.

3 Theoretical Model

Following decentralized college admissions, our model has three stages: 1) students decide which colleges to apply to, 2) colleges set admissions rates, and 3) given admissions offers, students decide which college to attend. There are two colleges: $c = 1$ and $c = 2$. Students pay a fee for the first application (F) and a potentially lower fee ($f \leq F$) when applying to a second college. The CA can be interpreted as a reduction in f since the application developed for the first college can also be used for the second college, and we examine how the CA changes student application activity and college admissions rates. In extensions of the model, we consider more than two colleges, college preferences over heterogeneous students, and two types of application costs, fees and time.

Each college has a fixed capacity, with total capacity across the two colleges serving a fraction $\kappa < 1$ of the student population. Each college sets an admissions rate, given the number of applicants, in order to satisfy their capacity $\kappa/2$. In our baseline model, colleges do not have preferences over students, simply setting an admissions probability Q_c that applies to all applicants. In extensions below, we consider more realistic college objective functions in which admissions decisions reflect college preferences over student characteristics, such as test scores and race.

Students receive a payoff $V_c = U_c + \varepsilon_c$ from attending college c.[10] The first term (U_c) repre-

[10]Not attending college is also an option, with a payoff normalized to zero. This outside option could also be interpreted as attendance at non-selective institutions, such as community colleges, without an application process.

sents pre-application information and might include residency status, reflecting the fact that over 80 percent of U.S. students attend in-state institutions.[11] The second term (ε_c), assumed to be distributed type-1 extreme value, represents post-application information, is revealed after applying but before choosing a college, and could reflect, for example, scholarship and financial aid offers or impressions from campus visits. There are two types of students: one-half prefer college 1 over college 2 ex-ante, knowing pre-application information but not post-application information, and one-half prefer college 2 over college 1.[12] Given all of this, and under symmetric admission rates ($Q_1 = Q_2 = Q$), a student preferring college 1 ex-ante will find it optimal to apply to both colleges over only applying to their first choice under the following condition:

$$\underbrace{Q^2(C_{12} - C_1)}_{\text{option-value}} + \underbrace{(1-Q)QC_2}_{\text{safety-value}} \geq f \tag{1}$$

where C_{12} is the expected value of having a choice set of both colleges, prior to knowing post-application information, and C_1 and C_2 are the corresponding expected values of having a choice set of only college 1 or college 2.[13] As shown on the left-hand side of the equation, there are two benefits of applying to a second college. According to the option value, the ex-ante second choice might be preferred ex-post following the realization of the post-application information, and this is relevant when students are accepted to both colleges, an event that occurs with probability Q^2. According to the safety value, students might be accepted to only their second choice, an event that occurs with probability $(1-Q)Q$. The right hand side is the cost of applying to a second college; if this cost is sufficiently high, students prefer to apply to only one college. In equilibrium, equation 1 is binding and students are indifferent between applying to only one college and applying to both, with a symmetric fraction of students (b) applying to both colleges and a fraction $1-b$ applying to only their first-choice college.

Given application behavior, colleges set admissions rates in order to equate the number of student acceptances of admissions offers to university capacity. For college 1, for example, total

[11] https://nces.ed.gov/programs/digest/d19/tables/dt19_309.10.asp?current=yes, accessed August 7, 2020). This preference for in-state institutions could be due to either a preference for proximity or lower in-state tuition at public institutions, an issue examined in Knight and Schiff (2019).

[12] We assume symmetry in the expected utility gains from attending the first choice college. That is, $U_1 - U_2 = \delta > 0$ for the first type and $U_2 - U_1 = \delta > 0$ for the second type.

[13] Given the type-1 extreme value assumption, these are equal to $C_{12} = ln[exp(U_1) + exp(U_2) + 1]$, $C_1 = ln[exp(U_1) + 1]$, and $C_2 = ln[exp(U_2) + 1]$, with $C_{12} > C_1 > C_2$.

student acceptances equals the yield on first-choice students who are admitted to college 1 plus the yield on second-choice students who both apply to and are admitted to college 1. This must then equal capacity, as expressed in the college capacity condition below:

$$0.5Q\underbrace{[(1-b)Y_1 + bQY_{12} + b(1-Q)Y_1]}_{\text{yield on first-choice}} + 0.5Qb\underbrace{[Qy_{12} + (1-Q)Y_2]}_{\text{yield second-choice}} = \kappa/2 \tag{2}$$

Among first-choice students, a fraction $1-b$ apply to only their first choice, with yield of Y_1, and a fraction b also apply to their second choice. In the latter case, a fraction Q are also admitted to their second choice, with yield of Y_{12} at college 1, and a fraction $1-Q$ are denied admission to their second choice, with yield for college 1 thus equal to Y_1. The second term represents yield on second-choice students, with a fraction b applying to both colleges. Among these, a fraction Q are also admitted to their first choice and yield equals y_{12}. The remaining fraction $(1-Q)$ are not admitted to their first choice and yield on these students equals Y_2.[14]

In equilibrium, universities set an admissions rate Q^*, and a fraction of students b^* apply to both colleges, as determined by the student indifference condition (equation 1 when binding) and the college capacity condition (equation 2), with closed form solutions in the Appendix.[15] As shown in Figure 4, the horizontal line at Q^* plots the binding student indifference condition prior to the CA. This is the admissions rate at which students are indifferent between applying to one and two schools. If the admissions rate were less than Q^*, then all students would apply to only one school $(b = 0)$, and if the admissions rate were higher than Q^*, then all students would apply to both schools $(b = 1)$. Against this condition we plot the college capacity condition, which shows that colleges must reduce their admissions rate as the share of students applying to both schools increases so that enrollment does not exceed capacity. The intersection of these conditions (b^*, Q^*) represents an interior equilibrium at which a fraction of students apply to both colleges and the number of students accepting each college's admissions offers exactly equals their fixed capacity.

Under the CA, which reduces f, students are no longer indifferent between applying to one institution and applying to both institutions, and the fraction applying to both increases accordingly,

[14] Given the type-1 extreme value assumption, these yields are equal to $Y_1 = exp(U_1)/[1+exp(U_1)], Y_2 = exp(U_2)/[1+exp(U_2)], Y_{12} = exp(U_1)/[1+exp(U_1)+exp(U_2)]$ and $y_{12} = exp(U_2)/[1+exp(U_1)+exp(U_2)]$.

[15] In the Appendix, we also develop a set of conditions guaranteeing the existence of a unique interior equilibrium.

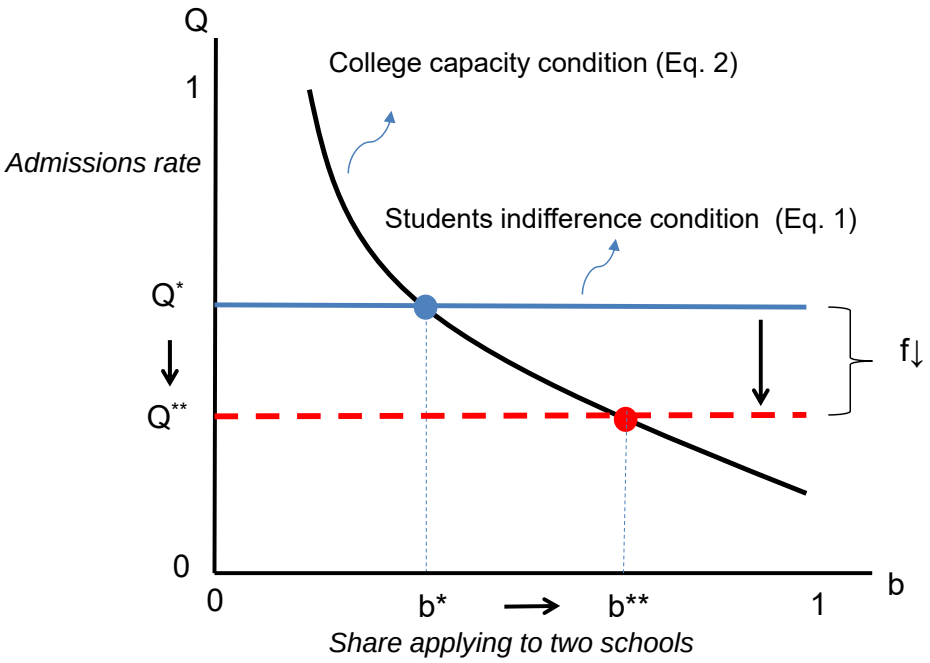


Figure 4: Theoretical Effects of Common Application

Notes: Initial equilibrium is at (b^*, Q^*), in which a fraction of students apply to both colleges and enrollment is exactly equal to each college's fixed capacity. After the second university joins the Common Application, the cost of applying to a second school decreases from F to f. As a result, the share of students applying to both institutions increases and universities must reduce the admissions rate so that enrollment doesn't exceed capacity. In the new equilibrium (b^{**}, Q^{**}), the share applying to both universities is higher and admissions rates are lower.

from b^* to b^{**}. Given this, admissions rates fall in order to satisfy capacity, from Q^* to Q^{**}. Despite the reduction in admissions rates, universities make more admissions offers in total in order to satisfy their capacity. This is driven by the reduction in yield on admitted students, who tend to have larger choice sets under the CA.[16] Finally, interpreting first-choice colleges as in-state and second-choice colleges as out-of-state, as described above, the CA leads to geographic integration, resulting from more students applying to out-of-state institutions. These results are summarized below, with a Proof in the Appendix.

Proposition: Consider the introduction of the CA, with a marginal reduction in the cost of applying to a second college. There are four effects: 1) application activity increases, 2) admissions rates fall, 3) yield falls and, despite increasing selectivity, universities accept a larger number of applications, 4) students are more likely to attend out-of-state institutions.

3.1 Extensions and Welfare

The Appendix develops three model extensions, all of which have three colleges, with only two in the CA. There is now also state 3, in which college 3 is located. The first extension considers network effects. The CA, with colleges 1 and 2 as members, increases the number of students applying to both colleges 1 and 2 but reduces second application activity for students whose ex-ante first choice is college 3, which is outside of the CA. Given this, the CA leads to network effects, with more student migration between states connected by the CA and less migration between states not connected by the CA.[17]

The second extension considers heterogeneous students, with low and high test scores. Colleges want to attract students with high scores and admit them with probability one. Students with low scores are then admitted at an admissions rate that fills remaining capacity. In equilibrium, students with high scores disproportionately attend CA schools, and students with low scores disproportionately attend schools outside of the CA, resulting from increased application activity among students with high scores at CA schools. Thus, the CA increases stratification, with high test score students disproportionately attending CA schools. This extension can handle any student

[16] Yield is defined as the probability that a student accepts an admissions offer.

[17] While we have not formally extended the model beyond three colleges, it is natural to conjecture that network effects might intensify as the size of the network grows beyond two. See the Appendix for more details.

attribute that is valued by colleges. If colleges value racial diversity, they can enroll a higher fraction of non-white students via an increase in the size of this applicant pool upon joining the CA. Given these predictions, our empirical analysis investigates the effect of joining the CA on both test scores and student race.

The third extension considers student income. There are now two types of application costs, financial and time. The financial cost, which includes the application fee, is unchanged under the CA. The time cost, by contrast, of applying to a second CA college falls under the CA. If financial costs are more salient than time costs for low-income students, relative to high-income students, then low-income students will be less responsive to the CA in terms of applying to multiple colleges.[18] In this case, CA schools might disproportionately attract high-income applicants and thus ultimately enroll more high-income students, relative to schools not in the CA. We investigate this issue below with data on Pell grants.

Does the CA increase student welfare? While the CA reduces frictions and increases student choice, it does not necessarily increase student welfare. In fact, the CA strictly reduces welfare in our baseline model.[19] Since the cost of applying to one college is unchanged and the probability of admission falls, the value of applying to only one college declines under the CA . Given that, in equilibrium, students are indifferent between submitting one and two applications, all students are worse off ex-ante. In our extensions with heterogeneous students, there are winners and losers under the CA. Students with high test scores are strictly better off since they are guaranteed admission and have a larger choice set under the CA. Students with low test scores, by contrast, might be worse off due to CA colleges becoming more selective. Likewise, low-income students are strictly worse off while high-income students might be better off.

[18]The salience of financial costs for low-income students could be driven by, for example, credit constraints. That is, if low-income students have a fixed budget for college applications, then a reduction in time costs will not induce additional applications. In a series of papers, Hoxby and Avery (2013), Hoxby and Turner (2014), and Hoxby and Turner (2015) argue that application fees deter applications from low-income students, who are often not aware of fee waivers. In an experiment, students provided fee waivers substantially increase their applications to selective colleges. Along these lines, Pallais (2015) shows that an increase in the number of free ACT score reports, which cost only $6 once the free reports have been sent, leads to a large increase in scores sent by ACT takers, and low-income ACT takers subsequently attended more selective colleges.

[19]Note that our model does not incorporate several possible benefits of the CA, such as a reduction in information frictions and an increase in overall college attendance.

3.2 Summary

In our model, the CA increases applications, reduces admissions rates, reduces yield, increases the number of admitted students, and increases out-of-state attendance, especially from other CA states. If schools have preferences for students with high test scores and for racial diversity, then schools joining the CA will have more enrollees with these characteristics. The reduction in time costs could also increase enrollment from high-income students. The CA strictly reduces ex-ante welfare in our baseline model but can lead to winners and losers under our extensions.

4 Empirical Analysis

To test the model predictions, we next consider how admissions outcomes and student demographics change when an institution joins the CA. We first describe the data and our empirical approach. We then present the key empirical results with respect to aggregate admissions outcomes, including applications, yield, admits, and selectivity. After addressing issues of geographic integration, we then turn to the question of whether and how the CA might change the composition of students.

4.1 Data

Our primary data source is the College Board's Annual Survey of Colleges, covering the years 1990-2016. These data include information on both admissions outcomes (the number of applications, admits, and subsequent freshman enrollment) and demographics of the entering student body (fraction out-of-state, SAT scores at the 25th and 75th percent level, and percent non-white). We focus on four-year institutions, both public and private, and the unit of observation is an institution-year pair. Summary statistics are provided in Appendix Table A1. As shown, institutions receive 4,462 applications on average, admit 2,672 students, and enroll 931 students. Acceptance rates average 70 percent across institutions and years, and yield, the fraction of admissions offers accepted by students, averages 41 percent. The fraction of out-of-state enrollment averages 31 percent, and the fraction of non-white students in the freshmen class averages 32 percent. The College Board panel is unbalanced since institutions do not respond to the survey every year. The median institution is included in 24 out of 27 surveys, and two-thirds of institutions are included in at least 21

surveys.

To examine the effects of the CA on admissions and enrollment outcomes, these data are combined with the year in which each university became a member of the CA.[20] We also use two other data sources. First, we downloaded data on the number of students receiving Pell grants by institution and year from the Department of Education.[21] Second, the Integrated Postsecondary Education Data System (IPEDS) has information on state-to-state student migration conducted biennially from 1986 to 2014, and we use these data to study network effects. In this case, the unit of observation is an institution by source state by year.

4.2 Identification Strategy

We use two estimating equations in most of our analysis: a regression model with a constant coefficient and an event study specification. The regression model relates an outcome y_{ct} (e.g., applications) to an indicator for CA membership in year t (CA_t) as follows:

$$ln(y_{ct}) = \beta CA_{ct} + \mu_c + \mu_t + \varepsilon_{ct} \tag{3}$$

where c indexes colleges, μ_c is a college fixed effect, and μ_t is a year fixed effect. Then, given the log specification, the parameter β captures the percent change in outcomes when joining the CA, after controlling for time effects and university effects.

Our event study specification is designed to measure the timing of any effects of entry and is given by:

$$ln(y_{ct}) = \underbrace{\sum_{k=-K}^{k=-1} \beta_{t+k} 1(t - J_c = k)}_{\text{pre-entry}} + \underbrace{\sum_{k=0}^{k=K} \beta_{t+k} 1(t - J_c = k)}_{\text{post-entry}} + \mu_c + \mu_t + \varepsilon_{ct} \tag{4}$$

where J_c is the year college c joined the Common App and $1(t - J_c = k)$ indicates that college c

[20] The entry year for current CA members was provided to us by The Common Application organization. The organization could not provide us with entry information for previous members and also only provided us with the most recent entry year for schools that left and then re-joined. They noted to us that it is uncommon for schools to leave the CA and even more rare for schools to then re-join at a later time.

[21] These data were downloaded from https://www2.ed.gov/finaid/prof/resources/data/pell-institution.html (accessed March 1, 2020).

joined the CA k years ago (or will join in the future when k is negative). We normalize β_{t-1} to zero and hence the key parameter β_{t+k} captures the effect of joining the CA at time t on outcomes at time $t+k$, relative to outcomes at time $t-1$.

A key threat to identification in our analysis is that joiners might have different pre-trends relative to the comparison group, which includes both schools that never join during our sample period and schools that will join in the future but before the end of our sample period. To address this concern, we also implement an alternative identification strategy in which, similarly to Deshpande and Li (2019), we compare outcomes for joiners to outcomes for colleges that will join the CA in the near future.[22] Under this approach, only colleges that eventually join the CA are included in the estimation and thus identification relies only on variation in the join year within a small observation window. More specifically, for each school that joins, we construct a comparison group that includes colleges that will join three to five years into the future. To ensure that the comparison group does not join during the relevant window, we analyze outcomes over a eight-year window, including the five years before joining, the join year, and the two years after joining. For example, for a school joining in 2000, the comparison group includes colleges that join in 2003, 2004, and 2005, and we analyze outcomes over the 1995-2002 period. We present both regression estimates with a single coefficient as well as event studies based upon this restricted comparison group of schools that will join in the near future. More details are provided in the Online Appendix Section A.12.

The idea behind this approach is that schools that join in the near future are more comparable than schools joining in the more distant future and schools that never join the CA during our sample period. The event study plots presented in the following sections test this idea directly and show that joiners and future joiners do indeed have similar pre-join time trends. To provide additional evidence on the comparability of joiners and joiners in the near future, Appendix Figure A8 compares the pre-join levels of each variable between joiners and the comparison group, separately for each identification strategy.[23] The first set of bars in each graph compares schools that join the CA during our sample period to schools that never join the CA, using the 1990 values of each variable,

[22] Deshpande and Li (2019) estimate the effects of Social Security Administration field office closings on local disability recipients by comparing areas where a field office closed to areas where an office closed several years later.

[23] Note that, due to the inclusion of college fixed effects, similar levels are not required for identification so long as trends are similar.

and thus approximates pre-treatment differences in our baseline approach. The second set of bars corresponds to the future joiners approach and compares joiners to their comparison group. In nearly every graph, the second set of bars are closer to each other than the first set of bars. This provides evidence on the comparability of joiners and joiners in the near future.

Comparing the two approaches, our baseline approach has two key advantages. First, it is based upon a larger sample size, both in terms of the number of institutions and the time span analyzed per institution. Second, our baseline strategy is better able to detect any long-run effects of joining the CA given that the analysis focuses on a longer time span. The key advantage of the future joiners strategy is that the comparison group is more similar. In addition, the comparison group in this approach is fixed, whereas the comparison group in our baseline analysis changes over time for each school that joins during our sample period.

4.3 Admissions Outcomes

We begin our investigation of the effect of joining the CA by examining the number of applications using the College Board data.[24] As shown in the first column of Table 1, we find that applications are 12 percent higher after a college joins the CA, relative to the period before they joined the CA. This economically and statistically significant result is consistent with the CA reducing frictions in college admissions via a reduction in the cost of applying to multiple universities that use the CA. Results are similar when restricting the comparison group to future joiners, as shown in the first column of Table 2.

To investigate the role of pre-trends and to consider any dynamic effects of joining the CA, we next present results from the corresponding event study specification.[25] As shown in the left panel of Figure 5, which includes 95 percent confidence interval bars, there is a slight downward trend in applications just before a school joins the CA. After joining the CA, by contrast, there is a discontinuous 10 percent increase in the number of applications received. Moreover, the effect grows over time, rising to roughly 23 percent after eight years in the CA. There are at least two

[24] Standard errors are clustered at the institution level.

[25] The left-hand panel of all event study plots shows results from estimating equation 4 using our full sample, including coefficients β_{t+k} for a few schools as early as 27 years before joining or as late as 40 years after joining. Schools that never join the CA are also included in the sample. In the plots we show a window of eight pre-join years and nine post-join years, which captures a large percentage of our pre and post join observations. As noted above, we normalize the coefficient one year before joining to zero ($\beta_{t-1} = 0$).

possible reasons why the effects might increase over time. First, the effect could be increasing over time due to the design of the platform, with, for example, the internet playing a large role in the success of the CA today, relative to the early days of the CA, when applications were still submitted on paper. Second, there could be network effects, with larger effects associated with joining the CA as the number of other CA members increases over time. We investigate this possibility of network effects in more detail below. When restricting the comparison group to future joiners, as shown in the right panel of Figure 5, there are no pre-trends in the number of applications received. Applications again spike by 10 percent in the join year, with some evidence again that the effect of joining the CA grows over time, reaching 15 percent just two years after schools join the CA.

We next investigate whether entry into the CA has led to a decrease in yield. As shown in column 2 of Table 1, there is a 9 percent reduction in yield after a college joins the CA, relative to the period before they joined the CA, and this effect is statistically significant at conventional levels. In the context of our model, this finding is consistent with the CA increasing student choice via a reduction in frictions associated with submitting college applications to multiple CA schools. Results are similar when restricting the comparison group to future joiners, as shown in column 2 of Table 2. Figure 6 shows the event study specifications for yield. As shown in the left panel of Figure 6, there is an immediate and discontinuous drop after a college joins the CA, with yield falling by roughly 7 percent. This effect again becomes more pronounced over time, with a 13 percent reduction in yield eight years after joining the CA. This dynamic effect could again be driven by either the CA becoming more powerful over time or by network effects associated with an increase in the number of CA members. Finally, as shown in the right panel of Figure 6, results are similar when restricting the comparison group to future joiners, with an immediate 7 percent reduction in the join year and a 10 percent reduction just two years after joining.

Given this reduction in yield, colleges might need to increase the number of admitted students in order to satisfy their capacity, as discussed in the theoretical model. As shown in column 3 of Table 1, we indeed find a large 11 percent increase in the number of admitted students in our baseline regression. When restricting the comparison group to future joiners, as shown in column 3 of Table 2, we document a somewhat smaller effect. But this 8 percent increase remains economically and statistically significant. As shown in the event study for the full sample (left panel of Figure 7), there is a discontinuous 10 percent increase in admits upon joining, and the effect again

increases over time, rising to 17 percent after 8 years. When restricting the comparison group to future joiners, as shown in the right panel, we again document a discontinuous increase, with further evidence of an increasing effect after just two years, from roughly 7 percent to 11 percent.

Finally, we investigate whether selectivity has changed, as measured via acceptance rates. Here the results are more mixed. When analyzing the full sample via a regression, we find no evidence of a change in selectivity, as shown in column 4 of Table 1. When restricting the comparison group to future joiners, by contrast, we document a 3 percent reduction in acceptance rates (column 4 of Table 2), and this difference is statistically significant at conventional levels. The event study for the full sample, in the left panel of Figure 8, documents a small drop in the acceptance rate after joining and larger effects in the years following CA adoption, with a 5 percent reduction in acceptance rates 8 years after joining the CA. When restricting the comparison group to future joiners, as shown in the right panel of Figure 8, we document a sharp 2 percent reduction in acceptance rates upon joining, and the effect is stable over time in this case.

To summarize, we find strong evidence that CA entry increased the number of applications, consistent with reduced frictions, and reduced yield, consistent with large student choice sets. We find mixed evidence regarding a hypothesized fall in acceptance rates but strong evidence that the number of admitted students increased.

Table 1: CA Entry and Admissions Outcomes

	(1)	(2)	(3)	(4)
	Log Applications	Log Yield	Log Admits	Log Selectivity
CA member	0.1204***	-0.0868***	0.1138***	-0.0067
	(0.0201)	(0.0130)	(0.0204)	(0.0110)
Observations	34519	34360	34556	34468
Clusters	1632	1631	1632	1632

Notes: Results from constant coefficient specification (Eq. 3) on full sample. All specifications include institution and year fixed effects, standard errors clustered by institution in parentheses.

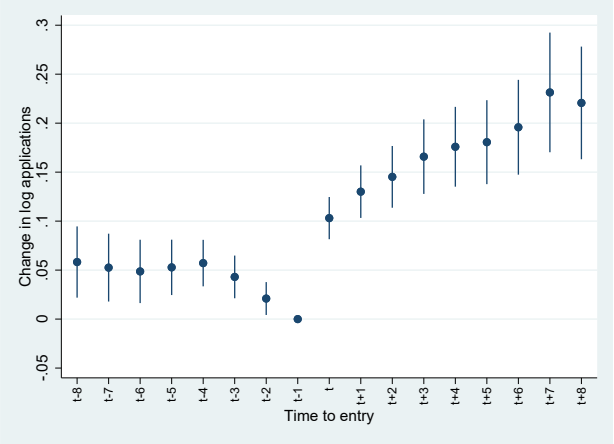
Table 2: CA Entry and Admissions Outcomes: Future Joiners Comparison

	(1) Log Applications	(2) Log Yield	(3) Log Admits	(4) Log Selectivity
CA member	0.1097***	-0.0735***	0.0818***	-0.0276***
	(0.0150)	(0.0115)	(0.0137)	(0.0077)
Observations	12182	12134	12184	12179
Clusters	488	488	488	488

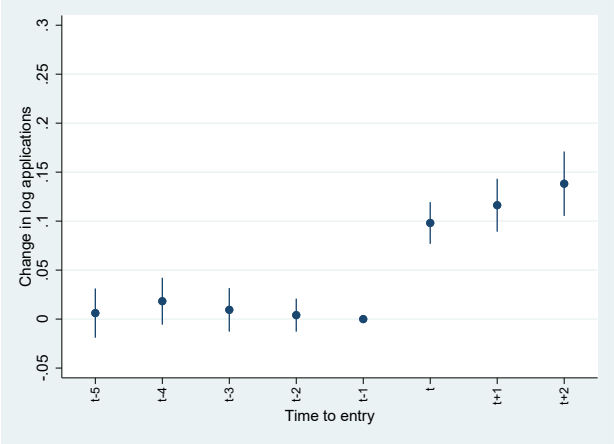
Notes: Results from constant coefficient specification (Eq. 3), restricting comparison group to future joiners. All specifications include institution and year fixed effects, and an indicator for joiner versus comparison group (see Appendix for details). Standard errors clustered by institution in parentheses.

Figure 5: CA Entry and Applications

(a) Full Sample



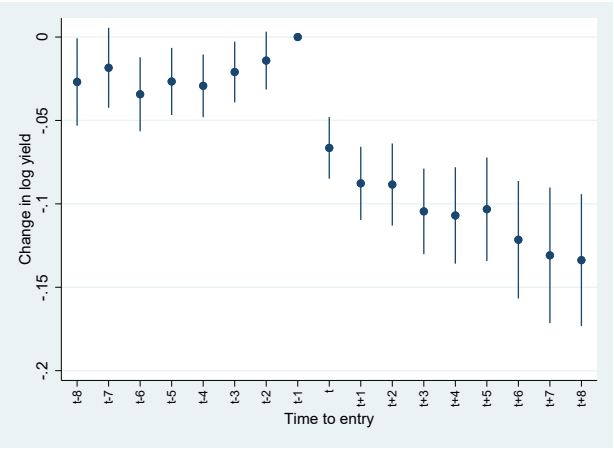
(b) Future Joiners Comparison



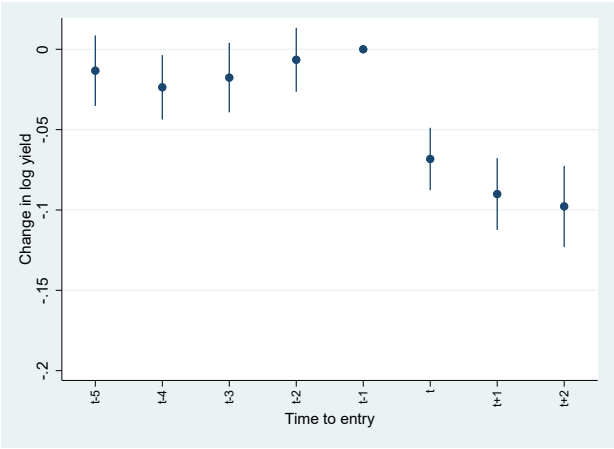
Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

Figure 6: CA Entry and Yield

(a) Full Sample



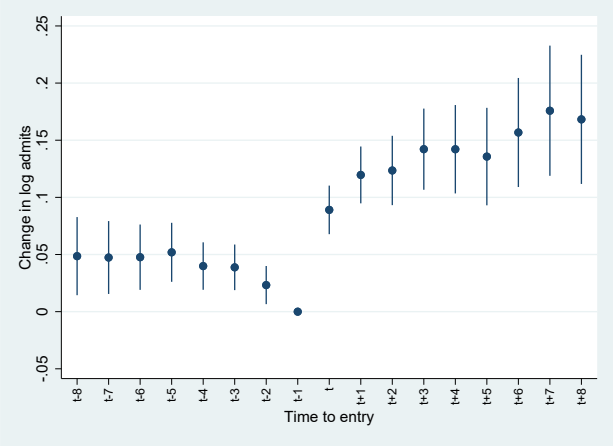
(b) Future Joiners Comparison



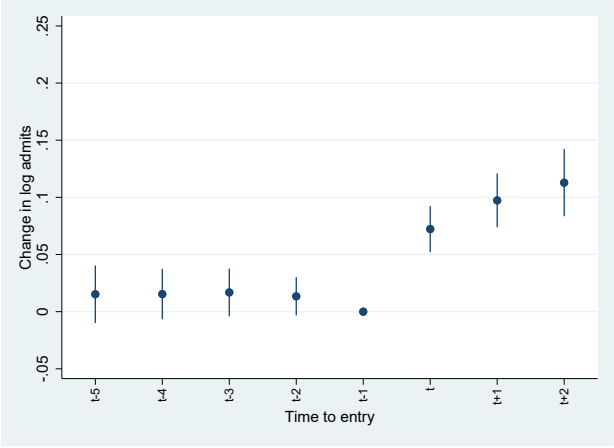
Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). Dependent variable is defined as log(enrollment/admits). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

Figure 7: CA Entry and the Number of Admits

(a) Full Sample



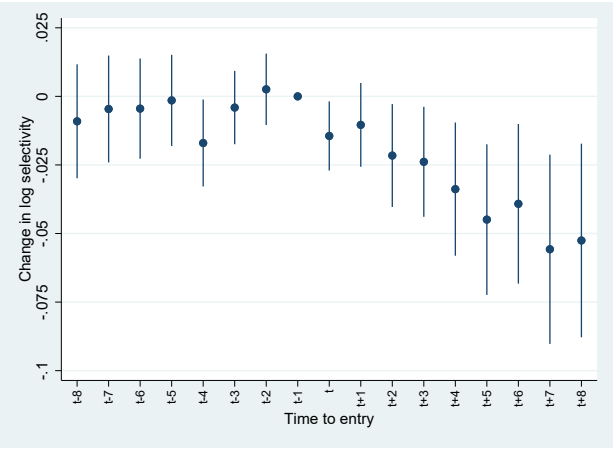
(b) Future Joiners Comparison



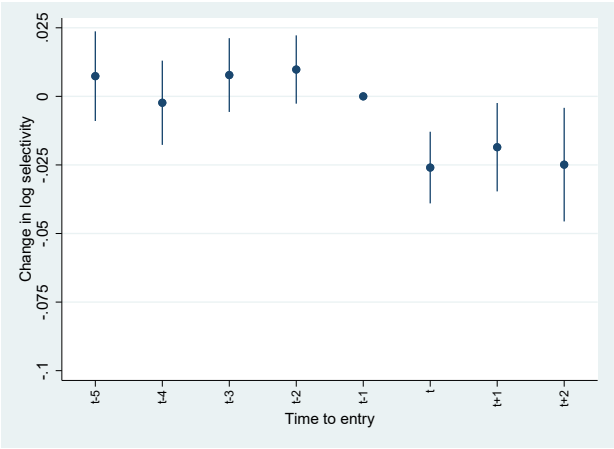
Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

Figure 8: CA Entry and Selectivity

(a) Full Sample



(b) Future Joiners Comparison



Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

4.4 Geographic Integration

Given the documented reduction in frictions and increased student choice sets, we next examine the role of the CA in contributing towards recent trends in geographic integration. In the Appendix, we first provide evidence that the geographic integration documented by Hoxby (2000), covering the period 1949-1994, has continued into our sample period. In particular, we find an increase over time in the average distance traveled by students and an increase in the fraction of out-of-state students.

Using College Board data, we measure the extent to which the CA has contributed to these trends in geographic integration. As shown in column 1 of Table 3, the fraction of out-of-state students rises by 1.4 percentage points in the years after joining, a roughly 5 percent increase relative to the sample average of 30 percent out-of-state. When restricting the comparison group to future joiners, we document an increase of roughly 0.6 percentage points, as shown in column 1 of Table 4. The corresponding event study for the full sample, as reported in the left panel of Figure 9, documents an immediate increase in the fraction of out-of-state enrollment of roughly 0.7 percentage points following a school joining the CA, and this effect roughly doubles, to over 1.2 percentage points, 8 years after joining the CA. When restricting the comparison group to future

joiners, we find smaller increases of roughly 0.5 percentage points, and these estimates are less precise due to the smaller sample size.

In the Appendix, we document similar results using our data on student migration from IPEDS. In particular, CA entry leads to an increase in out-of-state students. In addition, IPEDS includes information on state-to-state migration of college students, and we use this information to measure the average distance that students travel to attend college. We find that entry into CA increases distance traveled, and this effect largely comes from an increase in attendance from nearby states.

To summarize, we find that the CA has contributed towards geographic integration, with an increase in the fraction of out-of-state students when a school joins the CA. This is consistent with the predictions of our theoretical model, under which the CA induces more students to apply to and ultimately attend out-of-state institutions following a reduction in the costs of applying to multiple institutions. Below we consider the source of these out-of-state students via an investigation of network effects associated with the CA.

Table 3: CA Entry and Student Profiles

	(1) Out-of-State%	(2) SAT 25th Pctile	(3) SAT 75th Pctile	(4) Enroll % non-White	(5) Ugrad Enroll % Pell
CA member	0.0136***	4.4189	9.8235***	0.0082**	-0.0242***
	(0.0045)	(2.7833)	(2.5404)	(0.0039)	(0.0048)
Observations	37621	28504	28510	27494	26782
Clusters	1597	1428	1428	1567	1762

Notes: Results from constant coefficient specification (Eq. 3) on full sample. Specifications in columns 1-4 use the College Board data while column 5 uses separate Pell and IPEDS data. All specifications include institution and year fixed effects, standard errors clustered by institution in parentheses.

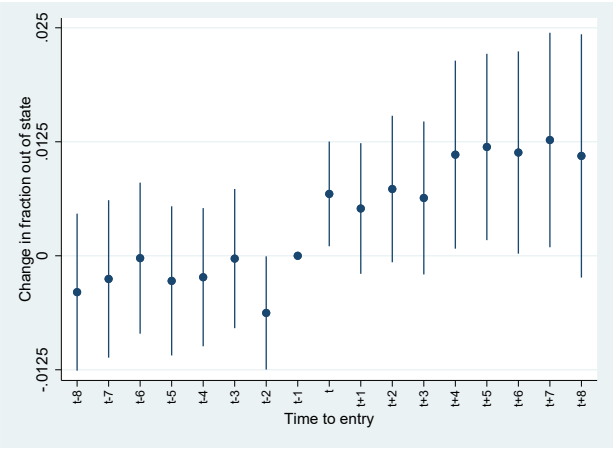
Table 4: CA Entry and Student Profiles: Future Joiners Comparison

	(1) Out-of-State%	(2) SAT 25th Pctile	(3) SAT 75th Pctile	(4) Enroll % non-White	(5) Ugrad Enroll % Pell
CA member	0.0064**	2.9630	2.4967	0.0193***	-0.0104***
	(0.0033)	(2.0422)	(1.9184)	(0.0047)	(0.0039)
Observations	12297	11475	11475	10303	10521
Clusters	487	474	474	480	469

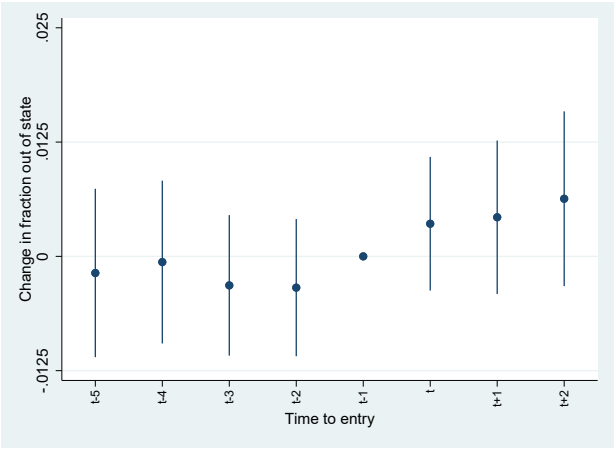
Notes: Results from constant coefficient specification (Eq. 3), restricting the comparison group to future joiners. Specifications in columns 1-4 use the College Board data while column 5 uses separate Pell and IPEDS data. All specifications include institution and year fixed effects, and an indicator for joiner versus comparison group (see Appendix for details). Standard errors clustered by institution in parentheses.

Figure 9: CA Entry and Out-of-State Enrollment

(a) Full Sample



(b) Future Joiners Comparison



Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

4.5 Network Effects

Given the network effects predicted by an extension of the theoretical model, we hypothesize that the effects of the CA should be increasing in the size of the network.[26] We investigate these issues by examining the outcomes described above but by also including in our regressions an interaction term between CA membership and network size, defined as the number of CA members in year t.

As shown in the first column of Table 5, we find that increasing the size of the network by 100 members increases the effect of CA membership on applications by 1.3 percent. For example, joining the CA at the beginning of our sample period, with roughly 100 members, increases applications by 8.5 percent. By the end of our sample period, by contrast, when the CA had 700 members, joining the CA increases applications by over 16 percent. We do not find any evidence of network effects when studying yield, as shown in column 2 of Table 5, and we find some evidence of reverse network effects when examining the number of admits, as shown in column 3 of Table 5. When measuring selectivity, by contrast, we find strong evidence of network effects, as shown in the final column of Table 5, with larger reductions in acceptance rates as the number of CA members grows.

[26]For an overview of network effects in two-sided markets, see Rysman (2009).

We next study network effects in the context of geographic integration. As shown in column 1 of Table 6, we find that the effect of CA membership is increasing in network size. For example, joining the CA at the beginning of our sample period, with roughly 100 members, increases the fraction of out-of-state students by only 0.5 percentage points. By the end of our sample period, by contrast, when the CA had 700 members, joining the CA increases out-of-state enrollment by approximately 2.5 percentage points.

Table 5: CA Entry, Network Size, and Admissions Outcomes

	(1) Log Applications	(2) Log Yield	(3) Log Admits	(4) Log Selectivity
CA member	0.0726**	-0.1155***	0.1933***	0.1220***
	(0.0304)	(0.0262)	(0.0342)	(0.0240)
CA X network	0.0131*	0.0079	-0.0217**	-0.0352***
	(0.0077)	(0.0067)	(0.0086)	(0.0058)
Observations	34519	34360	34556	34468
Clusters	1632	1631	1632	1632

Notes: Results from constant coefficient specification (Eq. 3) on full sample, but also including an interaction term between CA membership and network size. Network size is defined as the number of CA members in a year and is measured in hundreds; the average network size across all institution-years is 325 schools. All specifications include institution and year fixed effects, standard errors clustered by institution in parentheses.

Table 6: CA Entry, Network Size, and Student Profiles

	(1) Out-of-State%	(2) SAT 25th Pctile	(3) SAT 75th Pctile	(4) Enroll % non-White	(5) Ugrad Enroll % Pell
CA member	0.0015	-6.0308	-21.7286***	0.0044	0.0480***
	(0.0074)	(5.2642)	(4.8148)	(0.0065)	(0.0075)
CA X network	0.0033*	2.9767**	8.9885***	0.0010	-0.0184***
	(0.0019)	(1.2973)	(1.2274)	(0.0016)	(0.0018)
Observations	37621	28504	28510	27494	26782
Clusters	1597	1428	1428	1567	1762

Notes: Results from constant coefficient specification (Eq. 3) on full sample, but also including an interaction term between CA membership and network size. Network size is defined as the number of CA members in a year and is measured in hundreds. Specifications in columns 1-4 use the College Board data while column 5 uses separate Pell and IPEDS data. The average network size across all institution-years is 325 schools for the College Board data and 350 schools for the IPEDS data. All specifications include institution and year fixed effects, standard errors clustered by institution in parentheses.

We further examine the role of network effects in geographic integration using data on the source state of enrollment. Recall that the extension of our theoretical model to three colleges

predicts that institutions joining the CA are likely to see a greater increase in applications from students in states that already have a significant number of CA colleges. For example, if New York has high CA penetration (i.e., many New York schools in the CA), then we might expect that UW-Madison will attract more New York students after joining the CA since these New York students are already using the platform to apply to CA colleges in New York.

To examine these issues around the CA and student migration from source to destination states, we use IPEDS biennial migration data. We provide two measures of CA penetration (P_{st}), one based upon the fraction of colleges in source state s at time t that are members of the CA and one that is similar but weighted by college enrollment, recognizing that large colleges naturally receive more applications and are thus more salient to applicants from source state s. We then add this penetration measure and an interaction with the CA entry indicator to our two-way fixed effects specification, where the dependent variable is the number of freshmen (N_{sct}) from source state s attending college c at time t. This interaction term provides a test of whether enrollment from high CA penetration states increases when college c joins the CA. In our specification, the unit of observation is now a college by source state by year, and we thus include college by source state fixed effects and source state by time fixed effects:

$$ln(N_{sct}) = \beta_1 CA_{ct} + \beta_2 P_{st} + \beta_3 CA_{ct} \times P_{st} + \mu_{sc} + \mu_{st} + \varepsilon_{sct} \tag{5}$$

The key parameter of interest, β_3, captures the increase in enrollment from states with high CA penetration when a college joins the CA, after controlling for differences across states according to CA penetration and overall differences across colleges in CA membership.

As shown in column 1 of Table 7, the coefficient on the interaction between CA membership and CA penetration is positive and statistically significant, suggesting that the increase in applications upon joining the CA is derived from students applying to other CA schools. In terms of the magnitude of the effect, schools joining the CA enroll 3 percent more students from source states with no CA penetration but over 20 percent more students from source states with complete CA penetration (i.e., $P_{st} = 1$). This positive interaction effect is robust to restricting the comparison group to future joiners (column 2), using the full sample but weighting CA penetration by enrollment (column 3), and both restricting the comparison group to future joiners and using the

weighted penetration measure (column 4).

Table 7: CA Entry, Network Size, and Source States

	(1) Log Enrollment	(2) Log Enrollment	(3) Log Enrollment	(4) Log Enrollment
CA member	0.0338***	0.0072	0.0371***	0.0150
	(0.0102)	(0.0099)	(0.0101)	(0.0096)
CA member x CA penetration	0.1767***	0.1888**		
	(0.0419)	(0.0662)		
CA member x CA penetration wtd			0.1138***	0.0909**
			(0.0223)	(0.0342)
Observations	1052079	539223	1052079	539223
Clusters	1652	486	1652	486

Notes: Results from specification 5, estimated using IPEDS data on the number of freshmen from each source state, for each institution. Columns 1 and 3 show results from the full sample, while columns 2 and 4 restrict the comparison group to future joiners. All specifications include institution-source and year-source fixed effects; specifications in columns 2 and 4 also include an indicator for joiner versus comparison group (see Appendix for details). Standard errors clustered by institution in parentheses.

4.6 Stratification

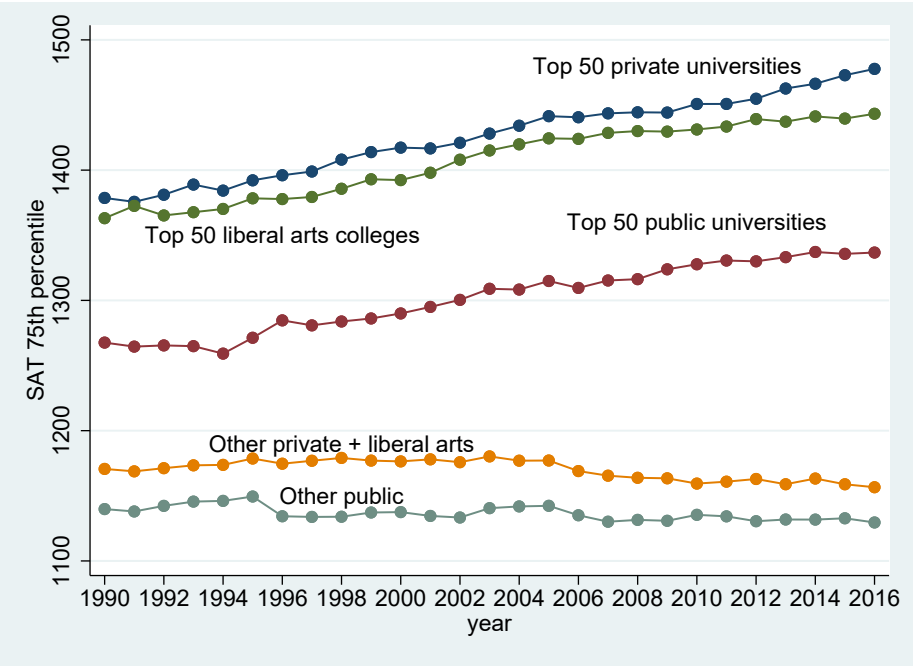
Our final research question involves whether the CA changed the types of students enrolling at institutions. We investigate heterogeneity along three dimensions: test scores, race, and income. First, given the reduction in frictions, the increased student choice sets, and geographic integration, we investigate whether the CA has contributed towards a widening of the gap between more selective and less selective institutions. Second, to the extent that colleges value racial diversity, they might be able to re-shape the racial composition of their student body due to the larger applicant pool after joining the CA. Third, motivated by our theoretical extension that considers income, we investigate whether joining the CA increases the fraction of high-income students enrolling at the university.

4.6.1 SAT scores

We begin by documenting general trends in SAT scores at different types of institutions. To do so, and in parallel with Figure 2, we classify schools into five categories: top 50 liberal arts, top 50 private, top 50 public, other private and liberal arts, and other public. As shown in Figure 10, there

is a large and increasing gap in SAT scores at the 75th percentile between selective schools (top 50 liberal arts, top 50 private, and top 50 public), and less selective institutions over our sample period. Thus, there is evidence of increasing stratification in general during our sample period.

Figure 10: Stratification in Higher Education



Notes: Figure plots average SAT scores at the 75th percentile in each year, weighted equally across schools within a category. Categorization of schools comes from the data appendix of Bound et al. (2010).

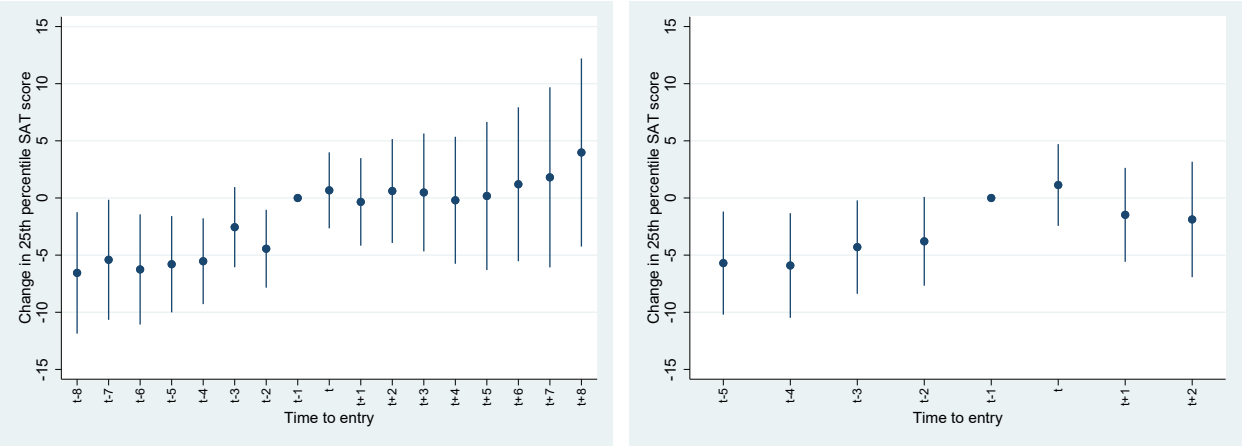
Given that the CA is disproportionately used by selective institutions, as documented above, we next investigate the degree to which the CA has contributed to this widening of the gap between more selective and less selective institutions. As shown in columns 2 and 3 of Table 3, there is a general increase in SAT scores of enrolled freshman following entry into the CA. In particular, SAT scores at the 25th percentile increase by 4.4 points and SAT scores at the 75th percentile increase by 9.8 points, although only the latter effect is statistically significant.[27] When restricting the comparison group to future joiners, the results remain positive but are now statistically insignificant, as shown in columns 2 and 3 of Table 4. Figure 11 (both panels) suggests that SAT scores do not increase at the 25th percent level. Figure 12, by contrast, documents increases in SAT scores at the

[27] One interpretation of the difference in effects between the 25th and 75th percentile, in the context of our theoretical extension to heterogeneous student ability, is that the fraction of high test score students is small. In this case, only the top of the distribution of SAT scores would change following entry into the CA, and the bottom of the distribution would be unaffected since it is composed of students with low test scores regardless of CA membership.

75th percentile upon CA entry, although there is evidence of pre-trends when using the full sample (left panel) and the results lack precision when restricting the comparison group to future joiners (right panel). When studying network effects in Table 6, we find strong evidence that any increases in SAT scores are stronger with a larger network; SAT scores at the 75th percentile, in particular, rise by over 40 points when a school joins the CA with 700 members.

Figure 11: CA Entry and SAT Scores at 25th Percentile

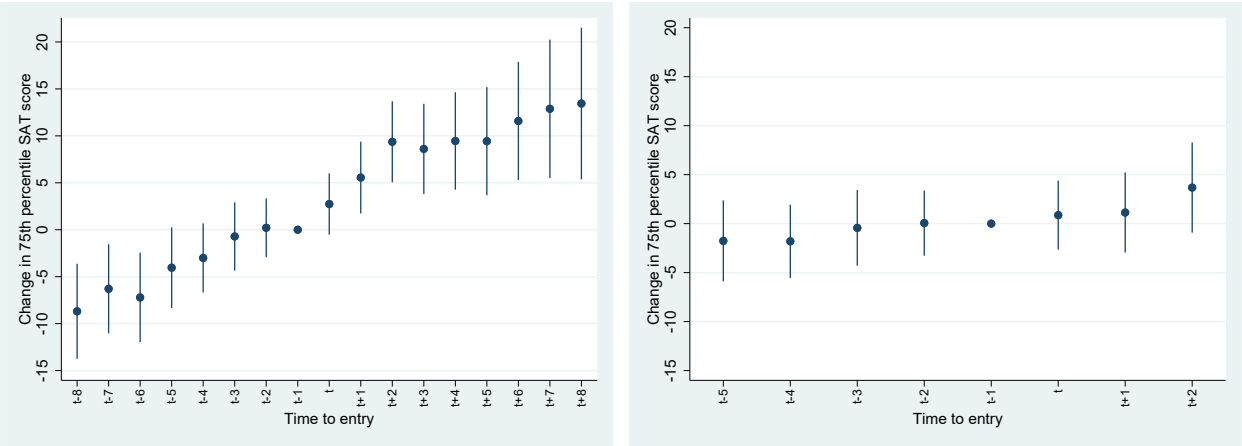
(a) Full Sample (b) Future Joiners Comparison



Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

Figure 12: CA Entry and SAT Scores at 75th Percentile

(a) Full Sample (b) Future Joiners Comparison



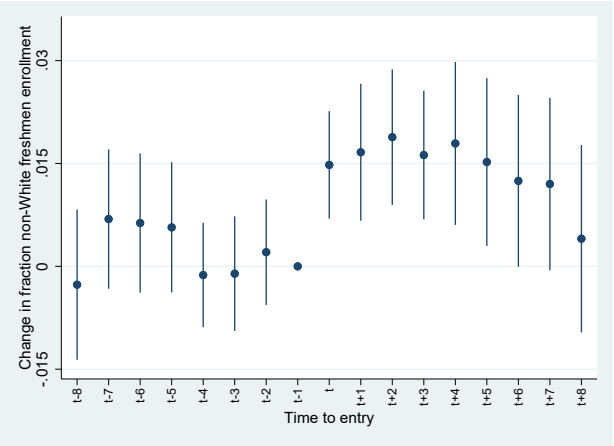
Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

4.6.2 Racial composition

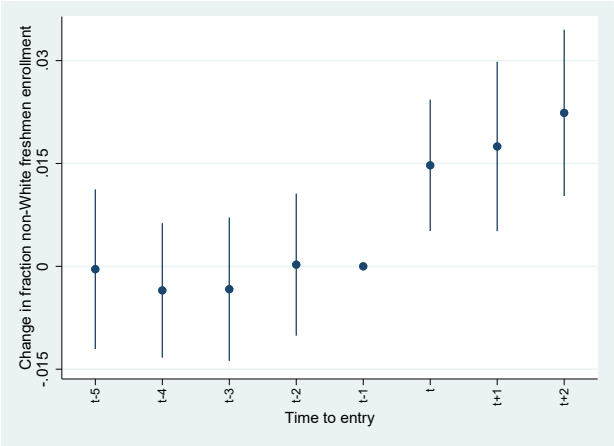
Given that universities tend to value racial diversity, they might be able to use the larger applicant pool after joining the CA to increase the fraction of non-white students.[28] While we do not have any data on the racial composition of the applicant pool, we can examine the racial composition of the entering class. As shown in column 4 of Table 3, we document an increase of nearly 1 percentage point in the fraction of non-white students following CA entry, relative to the sample average of 32 percent. Restricting the comparison group to future joiners, we document a 1.9 percent increase in the fraction non-white in the entering class, as shown in column 4 of Table 4. The full sample event study, as shown in the left panel of Figure 13, documents a discontinuous increase in the fraction non-white in the year of entry. The results are generally noisy, however, and statistically insignificant starting eight years after CA entry. When restricting the comparison group to future joiners (right panel of Figure 13), the results are cleaner, with no pre-trends, a discontinuous increase of 1.5 percentage points at time of entry, and stable or slightly increasing effects thereafter. We find no evidence of network effects in terms of racial composition, as shown in column 4 of Table 6.

Figure 13: CA Entry and Fraction Non-White

(a) Full Sample



(b) Future Joiners Comparison



Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

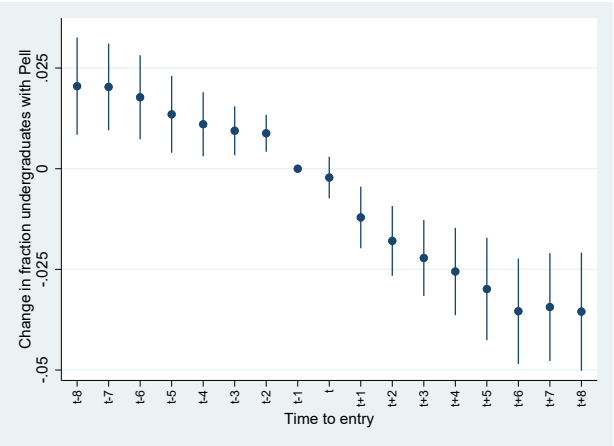
[28] Arcidiacono and Lovenheim (2016) review the literature measuring the degree of racial preferences in admissions and note that less selective institutions have less scope for such preferences given that they tend to admit a large fraction of applicants.

4.6.3 Income distribution

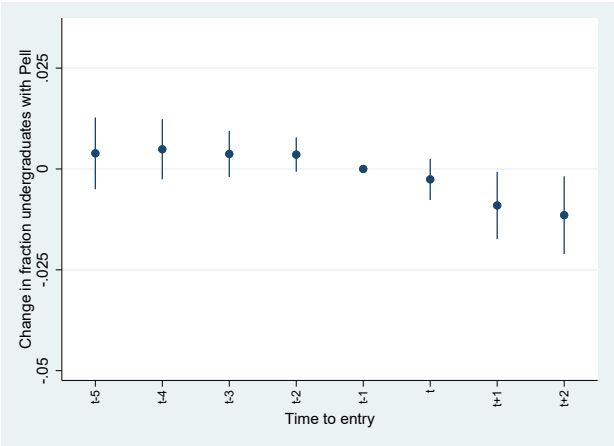
As predicted by our theoretical extension to income, higher-income students might be more responsive to the CA, relative to low-income students, for whom financial costs associated with applying are more salient. To measure the fraction of low-income students, we use data on the fraction of students with Pell grants. Importantly, while our previous measures are based upon the entering freshman class, these measures of Pell grants are based upon the entire student body. Given this, we do not expect to see discontinuous changes upon entry and instead expect to see more gradual changes in outcomes following CA entry. As shown in column 5 of Table 3, we find a reduction in percent Pell of 2.4 percentage points, a large effect relative to the baseline of 43 percent. When restricting the comparison group to future joiners, we find a smaller decrease of about 1 percentage point. Event studies using the full sample, as shown in the left panel Figure 14, document gradual declines in percent Pell, with a reduction of roughly 3.5 percentage points 8 years after joining the CA. Likewise, when restricting the comparison group to future joiners, we document a gradual decline, with a reduction of one percentage point two years after joining the CA. As noted above, gradual, rather than discontinuous, declines are consistent with the fact that the Pell data cover all enrollees and not just first-year students. As shown in the final column of Table 6, we again find strong evidence of network effects, with percent Pell falling by 8 percentage points when a school joins the CA with 700 members.

Figure 14: CA Entry and Fraction Receiving Pell Grants

(a) Full Sample



(b) Future Joiners Comparison



Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). Dependent variable is percentage of all undergraduates receiving Pell grants; data on Pell grants comes from the Department of Education. The left panel shows results from the full sample while the right panel restricts the comparison group to future joiners.

4.6.4 Summary

We find that the CA leads to changes in the degree of diversity on campus. Regarding SAT scores, CA colleges tend to be more selective with higher test scores at baseline. We find some evidence that SAT scores increase when joining the CA but those results are imprecise; if there is an effect of the CA on stratification by test scores, it is an increase. We find stronger evidence that the CA increases racial diversity, with a robust increase in the fraction of non-white students. Given that CA members tend to have fewer non-white students prior to joining the CA, as shown in Table 8, the CA has reduced racial stratification. Finally, we also find evidence that the CA reduces income diversity, with a documented reduction in the fraction of students receiving Pell grants. Given that CA members tend to have higher income students prior to joining the CA, as shown in Table 8, the CA has increased stratification according to income.

Table 8: CA Membership, Race, and Income

	Never joiners	Current CA members	Future CA members
Percent non-white (1990)	25.0	17.7	21.5
Percent Pell (1999)	44.1	19.3	33.2

Notes: Row one uses College Board data to show the percentage of non-white freshmen in 1990 for schools that never join the CA, schools that joined in 1990 or earlier, and schools that will join between 1991 and 2017. Row two uses the Pell data to show analogous results for the percentage of all undergraduates receiving Pell grants in 1999, the first year for which we have Pell grant data.

4.7 Other Outcomes

While we have attributed our results to the Common Application, it remains possible that the CA was adopted as part of a larger institutional strategy to increase applications and to change the composition of the student body. Thus, the effects that we have attributed to joining the CA might instead reflect other changes in institutional strategy adopted at the same time as CA entry.

To address this issue, we next explore changes in other university policies and outcomes. While we lack data on university recruiting and outreach, we do attempt to examine three other potential aspects of larger institutional strategy. First, it could be the case that universities want to expand their size and adopt the CA at the same time in order to increase the size of their applicant pool. Second, universities might have attempted to increase the quality of instruction at the same time as CA adoption. Finally, in an effort to increase the number of applications, admissions offices might have both joined the CA and reduced application fees. We examine these outcomes in Table 9 and show the event studies from the future joiners specification in Figure 15; the event studies from the full sample are shown in Appendix section A.9.

In column 1 of Table 9, we find some evidence of universities increasing their size when joining the CA, with an 3.7 percent increase in enrollment, when analyzing the full sample.[29] When restricting the comparison group to future joiners in column 4, however, we do not find any increases in the size of universities. Likewise, the full sample event study in Figure A3 shows an increase, but the left-most panel of Figure 15 shows only a small increase or no increase when restricting the comparison group to future joiners.

To investigate the quality of instruction, we use data on the number of PhD faculty. As shown

[29] It is also possible that this result reflects university size being below capacity prior to joining the CA.

in columns 2 of Table 9, we do not find any changes in the log number of PhD faculty upon CA entry, and results are similar when restricting the comparison group to future joiners (column 5). Event studies, the middle panel of Figure 15 and Appendix Figure A4, also do not document any changes in the number of PhD faculty when a school joins the CA.

Finally, we investigate whether the application fee changes upon CA entry. As shown in column 3 of Table 9, application fees, if anything, tend to increase upon CA entry, with a statistically significant increase in application fees of $1.64 when examining the full sample and a statistically insignificant increase of $0.83 when restricting the comparison group to future joiners. The event studies are inconclusive, with the right-most panel of Figure 15 showing at most a small increase, and the Appendix Figure A5 showing no stable pattern.

Taken together, we find little evidence that joining the CA is part of a larger institutional strategy to increase the number of applications. We do find some evidence that enrollment grows upon CA entry but no evidence of changes in instructional quality or application fees, which, if anything, tend to increase.

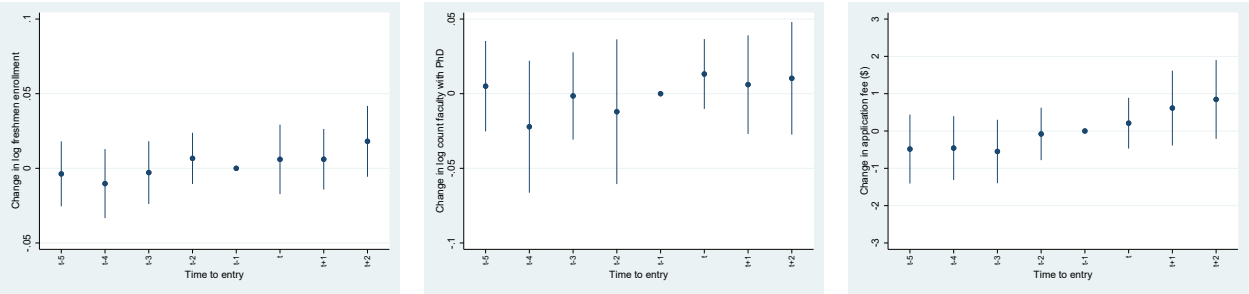
Table 9: CA Entry and Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Enrollment	Log PhD Faculty	Application Fee	Log Enrollment	Log PhD Faculty	Application Fee
CA member	0.0372**	0.0055	1.6358**	0.0113	0.0161	0.8290
	(0.0147)	(0.0161)	(0.6834)	(0.0102)	(0.0164)	(0.5199)
Observations	38359	33882	40881	12427	10821	12733
Clusters	1632	1602	1632	489	474	489

Notes: Results from estimating the constant coefficient specification (Eq. 3); columns 1-3 show results for the full sample while columns 4-6 show results restricting the comparison group to future joiners. All specifications include institution and year fixed effects, and specifications in columns 4-6 also include an indicator for joiner versus comparison group (see Appendix for details). Standard errors clustered by institution in parentheses.

Figure 15: CA Entry and Other Outcomes, Future Joiners Comparison

(a) Enrollment (b) PhD Faculty Count (c) Application Fee



Notes: Figures plot coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4). All three panels show results from restricting the comparison group to future joiners; results using the full sample are shown in the Appendix.

5 Conclusion

Consistent with model predictions, we find that the CA has significantly altered college admissions. In particular, after joining the CA, institutions experience an increase in the number of applications, consistent with a reduction in frictions. There is also a significant reduction in yield, consistent with increased student choice due to the CA. We also provide evidence that the CA has accelerated geographic integration, with more out-of-state students. Moreover, these out-of-state students tend to come from other states with significant CA penetration, patterns consistent with network effects in the CA. Taken together, these results suggest that the CA, by reducing application costs, has reduced frictions and increased student choice sets in college admissions, resulting in a more integrated market. Finally, we provide some evidence that CA entry is associated with changes in the composition of students, with increases in racial diversity, fewer low-income students, and weaker evidence of increases in SAT scores.

References

Arcidiacono, Peter and Michael Lovenheim, "Affirmative action and the quality-fit trade-off," *Journal of Economic Literature*, 2016, *54* (1), 3–51.

Avery, Christopher and Jonathan Levin, "Early admissions at selective colleges," *American Economic Review*, 2010, *100* (5), 2125–56.

Avery, Christopher N, Mark E Glickman, Caroline M Hoxby, and Andrew Metrick, "A revealed preference ranking of US colleges and universities," *The Quarterly Journal of Economics*, 2013, *128* (1), 425–467.

Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *The Quarterly Journal of Economics*, 2012, *127* (3), 1205–1242.

Bird, Kelli A, Benjamin L Castleman, Jeffrey T Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger, "Nudging at a national scale: Experimental evidence from a FAFSA completion campaign," 2019. NBER working paper 26158.

Blair, Peter and Kent Smetters, "Why Don't Elite Colleges Expand Supply?," 2019. University of Pennsylvania Working Paper.

Bond, Timothy N, George Bulman, Xiaoxiao Li, and Jonathan Smith, "Updating human capital decisions: Evidence from SAT score shocks and college applications," *Journal of Labor Economics*, 2018, *36* (3), 807–839.

Bound, John, Brad Hershbein, and Bridget Terry Long, "Playing the admissions game: Student reactions to increasing college competition," *Journal of Economic Perspectives*, 2009, *23* (4), 119–46.

_ , Michael F Lovenheim, and Sarah Turner, "Why have college completion rates declined? An analysis of changing student preparation and collegiate resources," *American Economic Journal: Applied Economics*, 2010, *2* (3), 129–57.

Cohodes, Sarah R and Joshua S Goodman, "Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy," *American Economic Journal: Applied Economics*, 2014, *6* (4), 251–285.

Deshpande, Manasi and Yue Li, "Who is screened out? Application costs and the targeting of disability programs," *American Economic Journal: Economic Policy*, 2019, *11* (4), 213–48.

Goodman, Joshua, Oded Gurantz, and Jonathan Smith, "Take two! SAT retaking and college enrollment gaps," Technical Report, National Bureau of Economic Research 2018.

Gurantz, Oded, Jessica Howell, Michael Hurwitz, Cassandra Larson, Matea Pender, and Brooke White, "Realizing Your College Potential: Impacts of College Board's RYCP Campaign on Postsecondary Enrollment," 2019. working paper.

Hoxby, Caroline and Christopher Avery, "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low-Income Students.," *Brookings Papers on Economic Activity*, 2013.

_ **and Sarah Turner**, "Expanding college opportunities for high-achieving, low income students (SIEPR Discussion Paper No. 12-014)," *Stanford Institute for Economic Policy Research, Stanford University*, 2014.

Hoxby, Caroline M, "How the changing market structure of US higher education explains college tuition," *NBER WP 6323*, 1997.

_ , "The effects of geographic integration and increasing competition in the market for college education," 2000.

_ , "The changing selectivity of American colleges," *Journal of Economic Perspectives*, 2009, *23* (4), 95–118.

_ **and Sarah Turner**, "What high-achieving low-income students know about college," *American Economic Review*, 2015, *105* (5), 514–17.

Knight, Brian and Nathan Schiff, "The Out-of-State Tuition Distortion," *American Economic Journal: Economic Policy*, 2019, *1* (11), 317–50.

Liu, Albert Yung-Hsu, Ronald G Ehrenberg, and Jesenka Mrdjenovic, "Diffusion of Common Application membership and admissions outcomes at American colleges and universities," Technical Report, National Bureau of Economic Research 2007.

Pallais, Amanda, "Small differences that matter: Mistakes in applying to college," *Journal of Labor Economics*, 2015, *33* (2), 493–520.

Rysman, Marc, "The economics of two-sided markets," *Journal of economic perspectives*, 2009, *23* (3), 125–43.

Smith, Jonathan, "The effect of college applications on enrollment," *The BE Journal of Economic Analysis & Policy*, 2013, *14* (1), 151–188.

_ , Michael Hurwitz, and Jessica Howell, "Screening mechanisms and student responses in the college market," *Economics of Education Review*, 2015, *44*, 17–28.

A Appendix (For Online Publication)

A.1 Additional notation

There is a unit mass of type-1 students and a unit-mass of type-2 students. Colleges have a fixed capacity $\kappa \in (0,1)$; thus, 2κ represents overall capacity across the two institutions.

We solve the model backwards. In the third stage, following student application decisions and university admissions offers, there are four possible student choice sets: choosing between both colleges, college 1 or 2 only, or only the outside option. The value of being accepted to both colleges equals $C_{12} = ln[exp(U_1) + exp(U_2) + 1]$.[30] Likewise, we denote $Y_{12} = exp(U_1)/[1 + exp(U_1) + exp(U_2)]$ as the yield for the first-choice college and y_{12} as the yield for the second-choice college for students accepted to both. The value of being accepted to only the first choice equals $C_1 = ln[exp(U_1) + 1]$, and the corresponding yield is $Y_1 = exp(U_1)/[1 + exp(U_1)]$. Similar expressions apply to the value of being accepted to only the second choice (C_2) and the corresponding yield is denoted by Y_2, with $Y_1 > Y_2$.

In the second stage, taking yield as given, schools set their admission rates in order to satisfy capacity. We focus on an equilibrium in which all students apply to their first choice and a fraction b of students also apply to their second-choice college. Then, admissions rates are set in order to equate the number of student acceptances of university admissions offers to university capacity. For college 1, for example, total students acceptances equal the yield on first-choice students who are admitted to college 1 plus the yield on second-choice students who both apply to and are admitted to college 1. This must then equal the overall university capacity, as expressed below:

$$0.5Q_1 \underbrace{[(1-b)Y_1 + bQ_2Y_{12} + b(1-Q_2)Y_1]}_{\text{yield on first-choice}} + 0.5Q_1 b \underbrace{[Q_2 y_{12} + (1-Q_2)Y_2]}_{\text{yield on second-choice}} = \kappa \tag{6}$$

Among first-choice students, a fraction $1-b$ apply to only their first choice, with yield of Y_1, and a fraction b also apply to their second choice. In the latter case, a fraction Q_2 are also admitted to their second choice, with yield of Y_{12}, and a fraction $1-Q_2$ are denied admission to their second choice, with yield for college 1 thus equal to Y_1. The second term represents yield on second-

[30] This follow the standard formula for consumer surplus in a logit model. Similar derivations apply for type 2 students, given the symmetry of the model.

choice students, with a fraction b applying to both colleges. Among these, a fraction Q_2 are also admitted to their first choice and yield thus equals y_{12}. The remaining fraction $(1-Q_2)$ are not admitted to their first choice and yield on these students equals Y_2.

Then, in the first stage, applying to both colleges yields a value of $A_{12} = Q_1Q_2C_{12} + Q_1(1-Q_2)C_1 + (1-Q_1)Q_2C_2 - F - f$ for type 1 students. That is, students are accepted to both colleges with probability Q_1Q_2, college 1 only with probability $Q_1(1-Q_2)$, college 2 only with probability $(1-Q_1)Q_2$ and face application costs of $F+f$. For type 1 students applying to only college 1, the value equals $A_1 = Q_1C_1 - F$. In equilibrium, the fraction of students applying to both colleges increases until the value from a second application equals the value of a single application $(A_{12} = A_1)$. This can be written as:

$$Q_2[\underbrace{Q_1(C_{12}-C_1)}_{\text{option-value}} + \underbrace{(1-Q_1)C_2}_{\text{safety-value}}] = f \tag{7}$$

The option value from a second application represents the benefit of being able to attend college 2 when the student has been accepted to both colleges, which occurs with probability Q_2Q_1. This captures the idea that students may learn that college 2 is actually preferred to college 1 throughout the admissions process, following the realization of ε_1 and ε_2. The safety value from a second application represents the benefit of being able to choose college 2 if not admitted to college 1, and this event occurs with probability $Q_2(1-Q_1)$.

A.2 Conditions for an Interior Solution

Regarding equation 7, the key condition for a unique solution is that the upper solution to the quadratic equation implies an admissions rate in excess of one. To ensure that only the lower solution is feasible requires that application costs be small, relative to the benefits of a larger choice set:

$$F < C_{12} - C_1 \tag{8}$$

That is, the cost of a second application must be less than the option value of also being admitted to one's second choice. The requirement that F is small also guarantees that a solution exists, in

the sense that the discriminant is positive.

Regarding equation 6, we require the following condition for an interior solution:

$$QY_1 < \kappa < Q^2(Y_{12} + y_{12}) + Q(1-Q)(Y_1 + Y_2) \tag{9}$$

where Q is set at its equilibrium value and is thus a function of model parameters. The left hand side of the inequality requires that college capacity is more than sufficient to accommodate accepted students when all students apply to only their first choice, given equilibrium admissions rates. The right hand side requires that the college capacity is not sufficient to accommodate the situation when all students apply to both colleges, given equilibrium admissions rates. Thus, capacity can be neither too small nor too large.

A.3 Equilibrium Solution

We first solve equation 7 for the equilibrium admissions rate. While this equation is quadratic in Q and thus has two solutions in principle, the upper solution implies an admissions rate in excess of 1, under the assumptions outlined in the Appendix above, and we thus focus on the lower solution:

$$Q^* = \frac{C_2 - \sqrt{C_2^2 - 4f(C_1 + C_2 - C_{12})}}{2(C_1 + C_2 - C_{12})} \tag{10}$$

Given this equilibrium admissions rate, one can then calculate the equilibrium fraction of students applying to both colleges via equation 6, yielding:

$$b^* = \frac{\kappa - QY_1}{Q^2[Y_{12} + y_{12} - Y_1] + Q(1-Q)Y_2} \tag{11}$$

where Q is set at equilibrium levels.

A.4 Proof of Proposition 1

Parts 1) and 2): In Equation 10, it is clear that equilibrium admissions rates are increasing in F. Thus, a marginal reduction in F leads to a reduction in equilibrium admissions rates. This effect is illustrated in Figure A1 below.

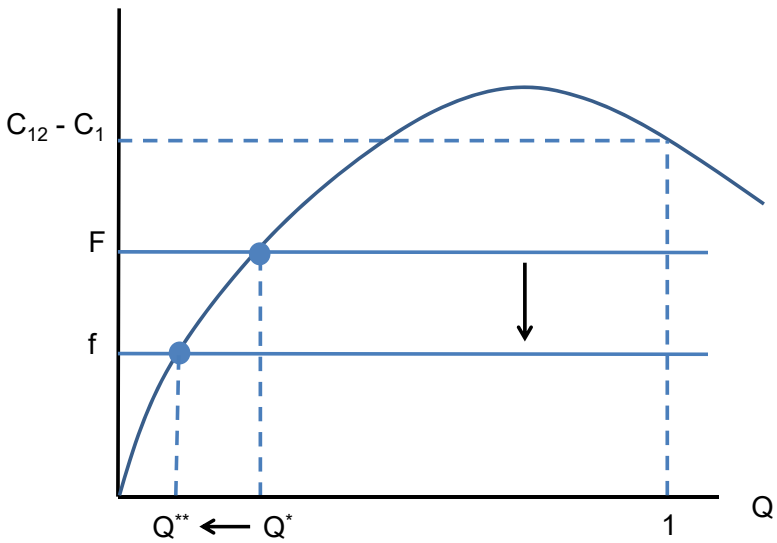


Figure A1: Effects on admissions rates

Given that Q declines under the CA, we must next show that b is decreasing in Q. Taking the derivative of equation 11 with respect to Q, we have:

$$\frac{db}{dQ} = \frac{-Y_1}{D} - \frac{(\kappa - QY_1)[2Q(Y_{12} + y_{12} - Y_1 - Y_2) + Y_2]}{D^2} \tag{12}$$

where the denominator equals $D = Q^2[Y_{12} + y_{12} - Y_1] + Q(1-Q)Y_2$. This denominator is positive since $Y_{12} + y_{12} > Y_1$.

Substituting back in the definition of b, we have that:

$$\frac{db}{dQ} = \frac{-Y_1 - b[2Q(Y_{12} + y_{12} - Y_1 - Y_2) + Y_2]}{D} \tag{13}$$

Re-arranging the numerator, this relationship can be written as follows:

$$\frac{db}{dQ} = \frac{-2bQ(Y_{12} + y_{12} - Y_1) + (bQY_2 - Y_1) - bY_2(1-Q)}{D} \tag{14}$$

Each of these three terms in the numerator are negative. In particular, the first term is negative since $Y_{12} + y_{12} > Y_1$. The second term is negative since $Y_2 < Y_1$, $b < 1$, and Q<1. Finally, the third term is negative since $Q < 1$ in equilibrium. Since the denominator must be positive for b to be

positive, the slope is negative. This change in application rates is illustrated in Figure A2 below.

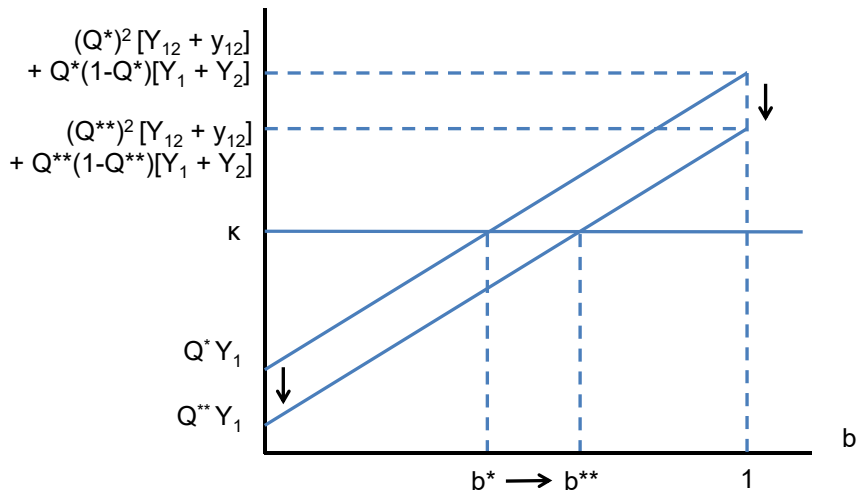


Figure A2: Effects on applications

Part 3): Note that the number of admitted students is equal to $Q(1+b)$, the product of the admissions rate and the number of applications received. Using the closed form solution for b, This can be written as:

$$Q(1+b) = Q + Qb = Q + \frac{\kappa - QY_1}{Q[Y_{12} + y_{12} - Y_1] + (1-Q)Y_2} \tag{15}$$

Taking the derivative, we have that:

$$\frac{dQ(1+b)}{dQ} = 1 - \frac{Y_1}{D} - \frac{\kappa - QY_1}{D^2}[Y_{12} + y_{12} - Y_1 - Y_2] \tag{16}$$

where the denominator equals $D = Q[Y_{12} + y_{12} - Y_1] + (1-Q)Y_2$.

Using the fact that $Qb = [\kappa - QY_1]/D$, the slope can be re-written as:

$$\frac{dQ(1+b)}{dQ} = 1 - \frac{Y_1}{D} - \frac{Qb}{D}[Y_{12} + y_{12} - Y_1 - Y_2] \tag{17}$$

This can be re-written as:

$$\frac{dQ(1+b)}{dQ} = \frac{D - Y_1 - Qb[Y_{12} + y_{12} - Y_1 - Y_2]}{D} \tag{18}$$

Since D is positive, we simply need to show that the numerator is negative. Since the term $Y_{12} + y_{12} - Y_1 - Y_2$ is negative, the numerator is increasing in b. Thus, to show that it is negative for all b between 0 and 1, we simply need to show that it is negative when $b = 1$. In this case, and canceling terms, the numerator can be written as $Y_2 - Y_1$, which is negative.

Part 4): The increase in out-of-state students follows directly from the increase in b resulting from the reduction in F.

A.5 Extension to 3 Colleges

We next consider the case in which two colleges ($c = 1$ and $c = 2$) join the CA but a third college ($c = 3$) does not join. In this case, there are three types of students, corresponding to the ex-ante ranking of the third college. Type 1 students have ex-ante preferences that rank college 3 last (there are two sub-types: either $U_1 > U_2 > U_3$ or $U_2 > U_1 > U_3$). Type 2 students have ex-ante preferences that rank college 3 in the middle (either $U_1 > U_3 > U_2$ or $U_2 > U_3 > U_1$). Type 3 students have ex-ante preferences that rank college 3 first (either $U_3 > U_1 > U_2$ or $U_3 > U_2 > U_1$). Given all of this, we can write the ex-ante preferences of the three different types (six different sub-types) of students as follows:

$$U_1 > U_2 > U_3, 1.1$$

$$U_2 > U_1 > U_3, 1.2$$

$$U_1 > U_3 > U_2, 2.1$$

$$U_2 > U_3 > U_1, 2.2$$

$$U_3 > U_1 > U_2, 3.1$$

$$U_3 > U_2 > U_1, 3.2$$

Let Q_{CA} and Q_N denote admissions rates at the CA colleges and the non-CA college, respectively. Capacities are symmetric and equal κ.

We focus here on the case in which students do not apply to all three colleges.[31] Let $b_1$be the

[31]This could formalized by having a large difference in preferences between the second and third choice. For type 1.1 students, for example, $U_2 - U_3$ would be large.

fraction of type 1 students applying to their first and second choice and likewise for b_2 and b_3. The capacity constraint for college 1 (college 2 is analogous) is now given by:

$$Q_{CA}[\underbrace{(1-b_1)Y_1}_{type\,1.1}+\underbrace{(1-b_2)Y_1}_{type\,2.1}+$$

$$\underbrace{b_1Q_{CA}Y_{12}+b_1(1-Q_{CA})Y_1}_{type\,1.1}+\underbrace{b_2Q_NY_{12}+b_2(1-Q_N)Y_1]}_{type\,2.1}+ \tag{19}$$

$$Q_{CA}[\underbrace{b_1Q_{CA}y_{12}+b_1(1-Q_{CA})Y_2}_{type\,1.2}+\underbrace{b_3Q_Ny_{12}+b_3(1-Q_N)Y_2]}_{type\,3.1}=\kappa \tag{20}$$

And for college 3 it is:

$$2Q_N[\underbrace{(1-b_3)Y_1+b_3Q_{CA}Y_{12}+b_3(1-Q_{CA})Y_1]}_{types\,3}+$$

$$2Q_N[\underbrace{b_2Q_{CA}y_{12}+b_2(1-Q_{CA})Y_2]}_{types\,2}=\kappa \tag{21}$$

Prior to the CA, the three relevant indifference conditions are given, similarly to before, by:

$$Q^2(C_{12}-C_1)+(1-Q)QC_2=F \tag{22}$$

With the introduction of the CA, the conditions become as follows for type 1 students:

$$Q_{CA}^2(C_{12}-C_1)+(1-Q_{CA})Q_{CA}C_2=f \tag{23}$$

However, for types 2 and 3, now they incorporate the two different admission rates. For type 2 students, taking the case of 2.1, we have that:

$$Q_{CA}Q_N(C_{12}-C_1)+(1-Q_{CA})Q_NC_2=F \tag{24}$$

For type 3 students, taking the case of 3.1, we have that:

$$Q_N Q_{CA}(C_{12} - C_1) + (1 - Q_N) Q_{CA} C_1 = F \tag{25}$$

Claim: these three indifference conditions cannot be simultaneously satisfied

Proof: The introduction of the CA causes the right hand side in equation 22 to fall from F to f. This implies that, given the admissions rate at CA schools, more type 1 individuals find it profitable to apply to a second school, increasing b_{CA}. However, under a higher b_{CA}, CA colleges will have excess demand, violating their capacity constraint (equation 20). Thus, Q_{CA} must decrease until type 1 students are indifferent between applying to a second school or not.[32]

Now, note that the fall in Q_{CA} causes the left hand side in equation 24 to increase, implying that more type 2 students want to apply to a second college. However, the opposite happens with type 3 applicants. The fall in Q_{CA} pushes down the left hand side in equation 25, implying that fewer type 3 students want to apply to a second college.

Due to these opposing effects of a decrease in Q_{CA} for type 2 and type 3 students, both conditions cannot be simultaneously satisfied, meaning that either b_2 or b_3 must be at a corner solution. More formally, comparing the conditions for type 2 and type 3, we have that:

$$Q_{CA} Q_N (C_{1,2} - C_1) + (1 - Q_{CA}) Q_N C_2 = Q_N Q_{CA}(C_{1,2} - C_1) + (1 - Q_N) Q_{CA} C_2$$

This is only satisfied when $Q_{CA} = Q_N$. However, under this condition, the left hand side of the three conditions are equal. But this is a contradiction with the fact that the first equation equals f, the second and third equations equal F, with $f < F$.

Claim: There is no equilibrium with $b_2 = 1$ and b_3 interior.

Proof: Assuming that b_1 is interior, and imposing symmetry, this would require the following:

$$Q_{CA}^2 (C_{1,2} - C_1) + (1 - Q_{CA}) Q_{CA} C_2 = f$$

$$Q_{CA} Q_N (C_{1,2} - C_1) + (1 - Q_{CA}) Q_N C_2 > F$$

[32] Note that an increase in b_{CA} must accompany the fall in Q_{CA}. Else, if only Q_{CA} were to fall, colleges would not meet the capacity constraint, as they would have open vacancies given the smaller admission rate.

$$Q_N Q_{CA}(C_{1,2} - C_1) + (1 - Q_N) Q_{CA} C_2 = F$$

Comparing the conditions for type 1 and type 3 and using the fact that $f < F$, we have that:

$$Q_{CA}^2(C_{1,2} - C_1) + (1 - Q_{CA}) Q_{CA} C_2 < Q_N Q_{CA}(C_{1,2} - C_1) + (1 - Q_N) Q_{CA} C_2$$

Re-arranging, this can be written as:

$$Q_{CA}(Q_{CA} - Q_N)(C_{1,2} - C_1 - C_2) < 0$$

Since $C_{1,2} - C_1 - C_2 < 0$, this requires $Q_{CA} > Q_N$.

Comparing types 2 and 3, we have that:

$$Q_{CA} Q_N (C_{1,2} - C_1) + (1 - Q_{CA}) Q_N C_2 > Q_N Q_{CA}(C_{1,2} - C_1) + (1 - Q_N) Q_{CA} C_2$$

Re-arranging, this can be written as:

$$(1 - Q_{CA}) Q_N > (1 - Q_N) Q_{CA} \tag{26}$$

This requires that $Q_N > Q_{CA}$. This contradicts that earlier requirement that $Q_{CA} > Q_N$.

Summary: The introduction of the CA leads to an increase in b_1, the fraction applying to both CA schools and a reduction in Q_{CA}. Given this, it must be case that b_2 increases to 1 or that b_3 decreases to zero since both cannot be interior. However, we have shown that b_2 cannot equal 1, meaning that $b_3 = 0$. Thus, there is a reduction, all the way to zero, in the fraction applying to a school outside of the CA and a school inside the CA. Given this, there are network effects with more type 1.1 students attending college 2 and more type 1.2 students attending college 1. Likewise, there are fewer type 3.1 students attending college 1 and fewer type 3.2 students attending college 2.

Quantitative analysis: To provide further evidence on the three college case, we choose parameters that guarantee interior solutions before the policy change. In particular, we set $U_1 = 2$, $U_2 = 1$, $\kappa = 0.55$, and $F = 0.3$. Prior to the CA, colleges and students are symmetric, with $Q = 0.2985$ and $b = 0.0776$. Introduction of the CA lowers F to $f = 0.29$. Under the CA, imposing

that $b_3 = 0$ in the new equilibrium, the admission rate of the CA schools falls from $Q = 0.2985$ to $Q'_{CA} = 0.2844$, while the admission rate of the non-CA school also falls, but by a smaller degree, to $Q'_N = 0.2942$. These changes reflect the direct and indirect effects of the decrease in F on the different types of students. Type 1 students are the only ones that benefit directly by the policy and strongly increase their applications to a second school from $b = 0.0776$ to $b'_1 = 0.3346$. On the other hand, type 2 students see their application rate grow only marginally, to $b'_2 = 0.0908$. The reason is that type 2 students have as first choice a CA college followed by the non-CA college, so they do not enjoy the lower application fee but do face the lower admission rate from the CA school, providing them with incentives to apply to their second choice.

A.6 Extension to Test Scores

We consider three colleges (c) and two test score types: low and high. We assume that colleges want to attract as many high test score students as possible and thus admit them with probability one. Low test score students are then admitted at a lower rate in order to fill any remaining capacity. Given our interest in stratification, we can then simply study the behavior of high test score students. Given that high test score students are admitted with certainty, the model plays out differently in this case. In particular, students will not be indifferent when choosing their application sets, and corner solutions are now relevant for these high test score students. Given these corner solutions, we focus on non-marginal changes in application costs.

We focus here on two cases. In the first case, application costs are sufficiently high that high test score students only apply to their first choice in the absence of the CA. For students with the preference order $U_1 > U_2 > U_3$, this requires:

$$C_{12} - C_1 < F \tag{27}$$

Suppose now that colleges 1 and 2, but not college 3, join the CA. Further, suppose that application costs fall sufficiently such that $C_{12} - C_1 > f$, and likewise for students with preference ordering $U_2 > U_1 > U_3$. Then, these two sets of students will apply to both colleges, and students with other preference orderings are unaffected. Since these two sets of students are now more likely to attend college (recall that $Y_{12} + y_{12} > Y_1$), the fraction of high test score students at CA colleges increases.

The fraction of high test score students at colleges outside of the CA is unchanged.

In the second case, suppose that application costs are sufficiently low that high test score students apply to their top two choices, but not their third choice, in the absence of the CA. For students with the preference order U_1>U_3>U_2, not applying to the third college requires:

$$C_{123} - C_{13} < F \tag{28}$$

where C_{123} represents the value from having a full choice set of all three colleges. Suppose now that colleges 1 and 2, but not college 3, join the CA, and application costs fall sufficiently such that $C_{123} - C_{13} > f$, and likewise for all students that have college 1 or 2 as their third choice. Then, all students except those with preference orderings U_1>U_2>U_3 and U_2>U_1>U_3 will apply to all three colleges. Thus, there is an increase in applications for colleges 1 and 2 and no increase in applications for college 3. Given that the yield on students accepted to college 3 now falls (resulting from more college 3 applicants also applying to colleges 1 and 2), this implies that colleges 1 and 2 will now draw some high test score students who would have attended college 3 in the absence of the CA. Thus, as in the first case, the fraction of high test score students at CA colleges increases. The new effect here is that the fraction of high test score students falls at schools outside of the CA.

A.7 Extension to Student Income

We consider three colleges (c) and two income types: low-income and high-income. There are two types of application costs. As before, the time cost of applying to a first CA college equals F and the time cost of applying to a second CA college equals $f \leq F$. The financial costs of applying to a first CA colleges equals ϕ and the cost of applying to a second CA college also equals ϕ. We simplify the model by assuming that low-income students can only apply to one college, perhaps due to credit constraints. We also assume that colleges do not distinguish between low-income and high-income students in terms of admissions probabilities. Given all of this, only high-income students decide whether or not to apply to a second college. Thus, all of the results from the first extension apply to high-income students but not low-income students. In particular, colleges 1 and 2, which are members of the CA, experience an increase in applications from high-income

students, relative to college 3, which is not a member of the CA. Given this, colleges 1 and 2 ultimately enroll more high-income students, relative to college 3, which ends up attracting more low-income students.

A.8 Summary Statistics for Main Variables

Table A1: Summary Statistics

	mean	sd	min	max	count
Applications	4462.01	6469.93	29.00	97121.00	34519
Yield	0.41	0.17	0.03	1.00	34361
Admits	2672.44	3380.56	25.00	36088.00	34556
Freshmen Enrollment	931.38	1130.04	25.00	11807.00	38359
Selectivity	0.70	0.18	0.04	1.00	34468
Fraction Out of State	0.31	0.25	0.00	1.00	37628
SAT 25th pctile	967.22	136.67	410.00	1510.00	28534
SAT 75th pctile	1184.60	130.36	720.00	1600.00	28540
Non-White Enroll %	0.32	0.23	0.00	1.00	27517
PhD Faculty	222.80	348.97	0.00	3792.00	33893
Application Fee	29.82	16.84	0.00	150.00	40881
Ugrad Enroll % Pell	0.43	0.22	0.00	1.00	26821

All variables from College Board data except Ugrad Enroll % Pell;
this variable uses separate IPEDS and Dept. of Ed. data.

A.9 Event Study Plots for Additional Outcome Variables

A.10 Trends in Geographic Integration

Hoxby (2000) documents that the percentage of students attending in-state institutions fell consistently from 1949 to 1994 and that the role of distance in explaining college choice decreased as well. We extend this study of geographic integration into our sample period by measuring trends in distance traveled from a student's home state to the state of their university using IPEDS data over the period from 1986 to 2014.[33] In Figure A6 we plot the mean distance traveled in each

[33] In particular, we measure the great circle distance in kilometers between state centroids, defining the distance for all in-state students as zero

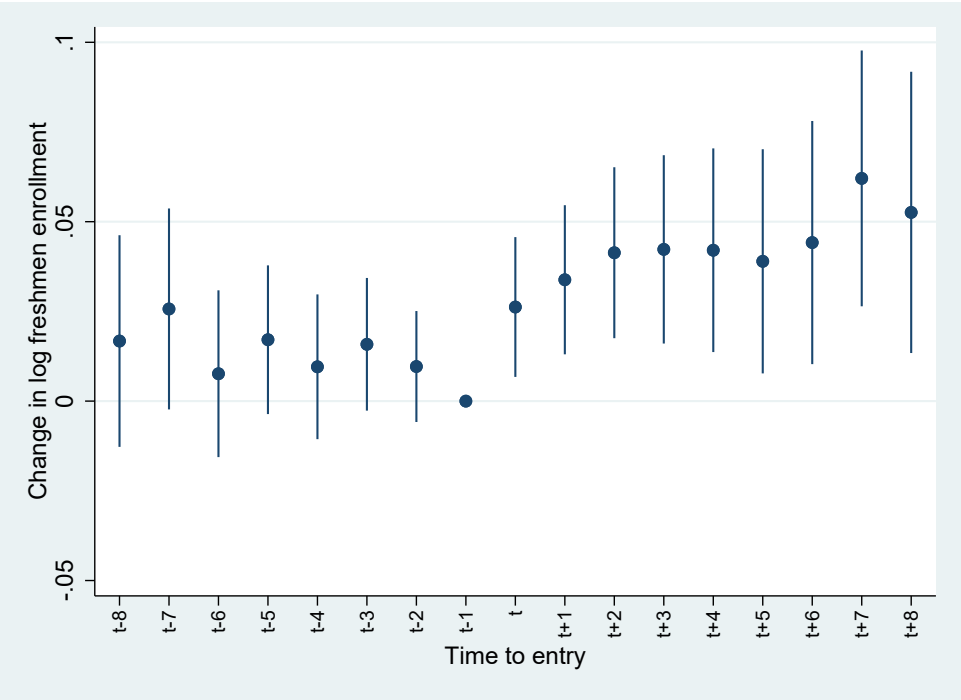


Figure A3: CA Entry and Enrollment

Notes: Coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4) using the full sample.

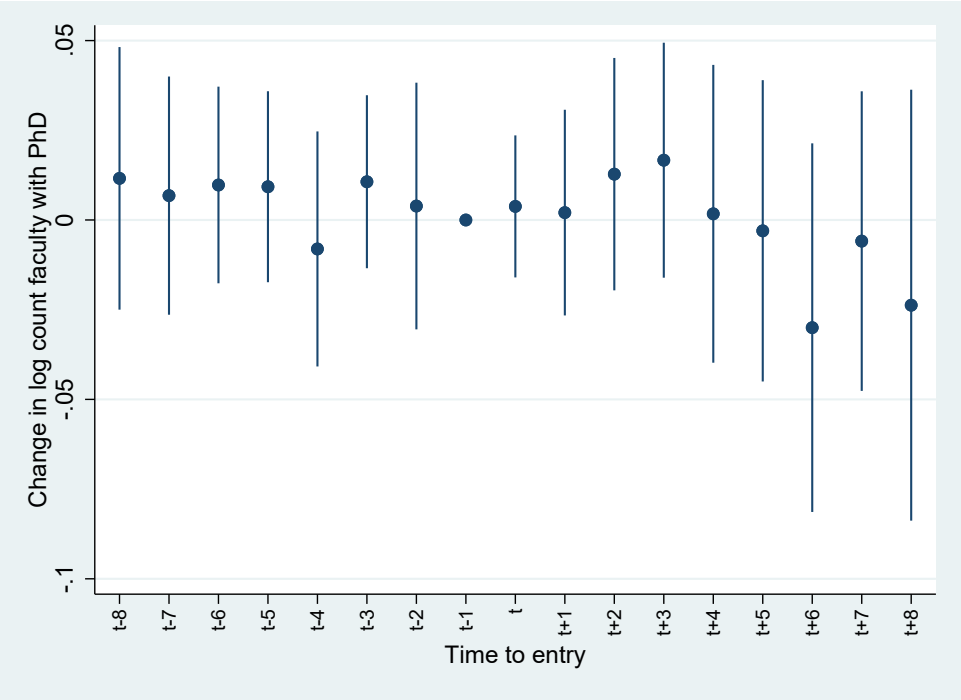


Figure A4: CA Entry and PhD Faculty Count

Notes: Coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4) using the full sample.

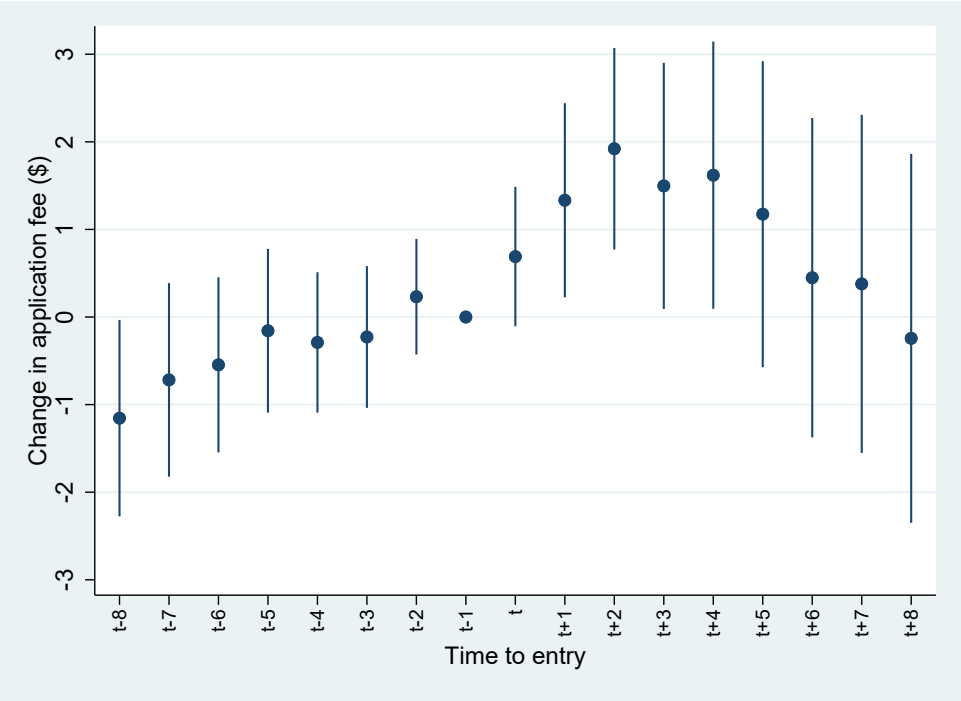


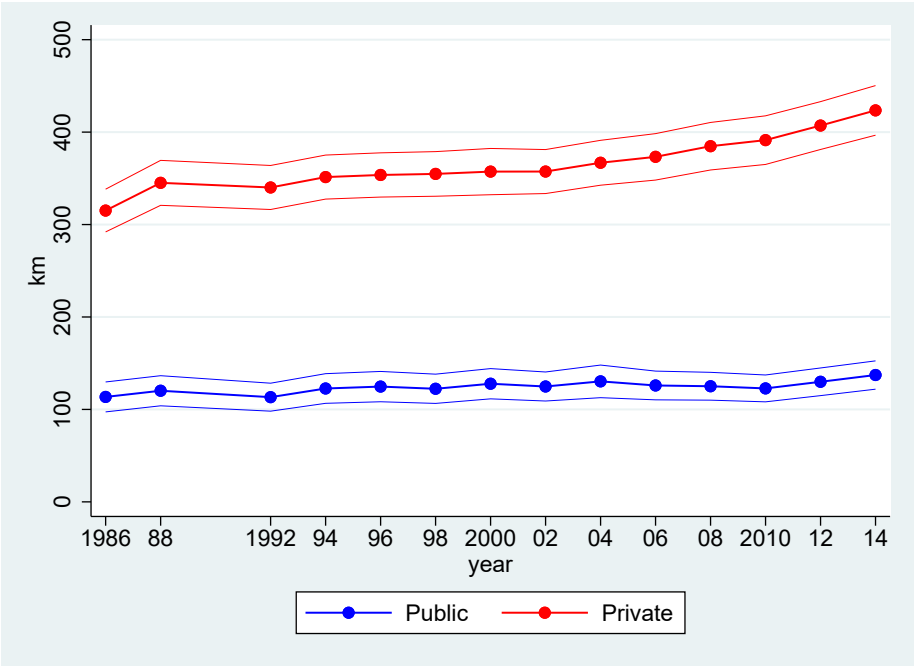
Figure A5: CA Entry and Application Fee

Notes: Coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4) using the full sample.

year, along with 95 percent confidence intervals.[34] As shown, there is a clear increase in distance traveled over this time period, with the average distance traveled increasing by over 100 kilometers for private universities and roughly 40 kilometers for public universities. Thus, the trends towards greater geographic integration documented by Hoxby between 1949 and 1994 also appear over our sample period 1986-2014.

[34]Letting the subscript s denote a student's state, we define the mean distance traveled by all students from US states at college c in year t as $avdist_{c,t} = (1/nat\_enroll_{c,t})\sum_{s \in S} enroll_{c,s,t} * dist_{c,s}$. The variable nat_enroll is total enrollment from the 50 US states and D.C. The home location of foreign students and students from US territories is usually not available, and therefore we excluded these groups from the total. However, students from these groups are counted in total enrollment when calculating percentage of students attending in-state.

Figure A6: Distance Traveled per Student



Notes: Plot shows mean distance traveled and 95% confidence intervals. Distance calculated between university-state centroid and student-state cenroid. Distance for in-state students set to zero. Data covers every two years from 1986 to 2014, except 1990 (missing).

In Table A2 we calculate the average increase in geographic integration over time for public and private institutions, using the specification $y_{ct} = \beta_1 years_t + \beta_2 years_t \times public_c + \mu_c + \varepsilon_{ct}$. In column 1 we find that distance traveled increases by about 3 kilometers per year for private institutions and 1 kilometer per for public institutions, while column 2 specifies average distance in logs and shows that both types of institutions have roughly the same percentage increase over time of 1.4 percent. In columns 3 and 4 we examine the average distance traveled by out-of-state students only, which allows us to distinguish the effect of a change in the percentage of out-of-state students from a change in the geographic composition of the out-of-state students. Interestingly, the results show that while out-of-state students at private institutions are traveling further each year, there is essentially no increase in distance for out-of-state students at public institutions (the interaction effect is the same magnitude as the main effect). This implies that the increasing distance traveled by public university students comes entirely from an increase in the out-of-state percentage, which increases at about 0.14 percentage points each year for both types of institutions.

Table A2: Geographic Integration by Institution Type

	(1) Distance	(2) Log Distance	(3) Distance	(4) Log Distance	(5) Out-of-state %
years	3.1867***	0.0136***	5.1360***	0.0070***	0.0014***
	(0.2797)	(0.0010)	(0.3278)	(0.0004)	(0.0002)
years X public	-1.9582***	0.0001	-5.5110***	-0.0056***	-0.0001
	(0.3571)	(0.0018)	(0.6052)	(0.0007)	(0.0003)
Constant	236.9366***	4.6789***	876.4421***	6.5852***	0.2941***
	(2.9090)	(0.0127)	(4.2712)	(0.0048)	(0.0022)
Observations	20685	20236	20236	20236	20685
Clusters	1708	1688	1688	1688	1708

Notes: Results from estimating $y_{ct} = \beta_1 years_t + \beta_2 years_t \times public_c + \mu_c + \varepsilon_{ct}$, where μ_c is an institution fixed effect. The dependent variable in columns 1 and 2 is distance per student, in columns 3 and 4 it is distance per out-of-state student, while column 5 shows percentage of out-of-state students at the institution. Distance is measured in kilometers. Standard errors clustered by institution in parentheses.

A.11 CA and Geographic Integration (IPEDS data)

As shown in Table A3, universities, after joining the CA, experience a significant increase in the average distance students travel to attend, with column 1 documenting an increase of 30 kilometers, a roughly 10 percent increase (column 2). Restricting to only out-of-state students, the distance increases by 55 kilometers (column 3), an increase of 7 percent for this population (column 4). In addition to out-of-state students traveling further, joining the CA also decreases the fraction of in-state students by about 2.3 percentage points (column 5). Generally, the magnitudes of the effects in Table A3 are large. Comparing each coefficient in Table A3 to its counterpart in Appendix Table A2, the effect of joining is about 10 times larger than the yearly trend for distance measures and about 15 times larger for in-state percentage.

Table A3: CA Entry and Geographic Integration

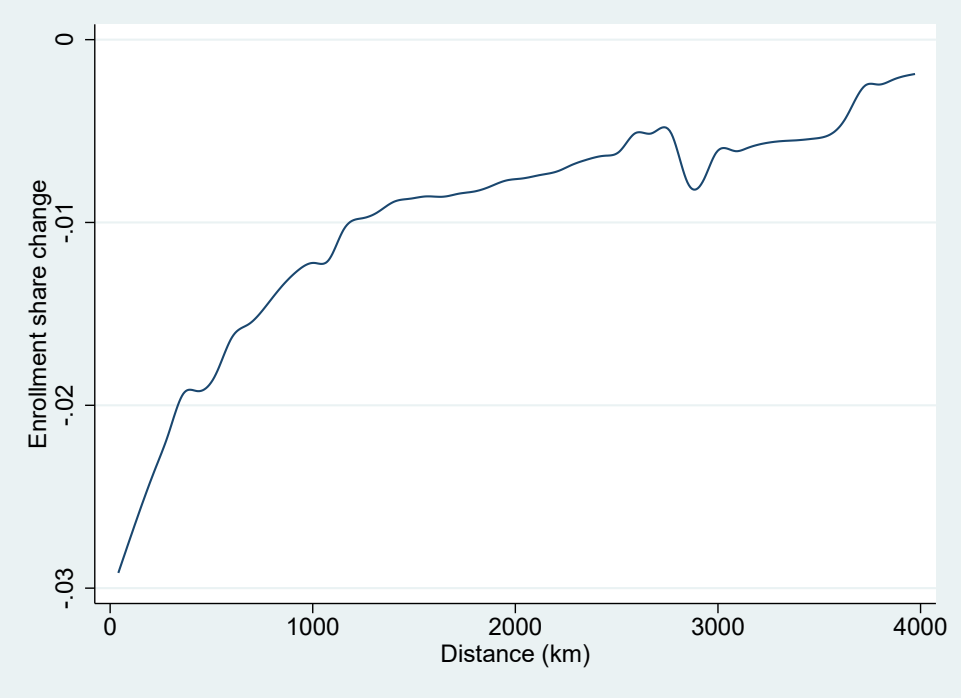
	(1) Distance	(2) Log Distance	(3) Distance	(4) Log Distance	(5) Out-of-state %
CA member	30.1654***	0.1044***	55.1233***	0.0700***	0.0233***
	(6.1947)	(0.0255)	(8.5535)	(0.0105)	(0.0048)
Observations	20629	20176	20176	20176	20629
Clusters	1652	1628	1628	1628	1652

Notes: Results from constant coefficient specification (Eq. 3) on full sample, using IPEDS data. The dependent variable in columns 1 and 2 is distance per student, in columns 3 and 4 it is distance per out-of-state student, while column 5 shows percentage of out-of-state students at the institution. Distance is measured in kilometers. All specifications include institution and year fixed effects, standard errors clustered by institution in parentheses.

As a further analysis of the effect of joining CA on distance traveled, we now consider the change in the entire distribution of a college's enrollees over distance. To do so, we restrict our sample to only those institutions which joined the CA and which have migration data both three or fours years before and three or four years after joining, depending on whether the institution joined in an odd or even year. We then sum the enrollees across all universities in each period (pre, post) and calculate the percentage coming from each state-to-state distance. This allows us to calculate two cumulative distribution functions (CDF), where each bin of the CDF represents a given state-to-state distance and we are calculating the percentage of students traveling that distance to all universities in a period. We then plot the difference between the before and after CDF in Figure A7.[35] The largest difference occurs at zero, indicating that most of the effect comes from a 3 percentage point decrease in the in-state percentage. The slope of this differenced CDF increases sharply and approaches zero, so that a distance of 1200 kilometers the change is less than 1 percent, and then flattens. This shape suggests that CA entry increases the distance traveled by enrollees by mostly increasing the number of enrollees from nearby states.

[35] We truncate the graph at 4000km since the share of students coming from a greater distance is very small in both periods.

Figure A7: Effect of CA Entry on Enrollment Share Change by Distance



Notes: Enrollment share change defined as enrollment share before joining the CA minus the enrollment share after joining the CA. The before period is 3 to 4 years before the join year; the after period is 3 to 4 years after the join year. Distance is calculated between state centroids; in-state distance defined as zero. The graph is smoothed with the median spline method using 50 bands. Sample has 265 unique institutions.

A.12 Future Joiners Comparison: Methodology and Data Preparation

As described earlier, a key threat to identification in our analysis is that joiners might have different pre-trends relative to the comparison group, which includes both schools that never join during our sample period and schools that will join in the future but before the end of our sample period. We address this concern with an additional specification in which we compare outcomes for joiners to outcomes for colleges that will join the CA in the near future. Specifically, we compare schools that joined in year j to schools that will join in years $[j+3, j+5]$, over a pre-join period $[j-5, j-1]$ and a post-join period $[j, j+2]$. This duration is short enough to make joiners and future joiners quite comparable, but still long enough to estimate a post-join effect. For example, for a school joining in 2000, the comparison group includes colleges that join in 2003, 2004, and 2005, and we analyze outcomes over the 1995-2002 period.

This empirical strategy is similar, although not identical, to that used by Deshpande and Li (2019) and we constructed our dataset in a similar way. We first assembled separate datasets of

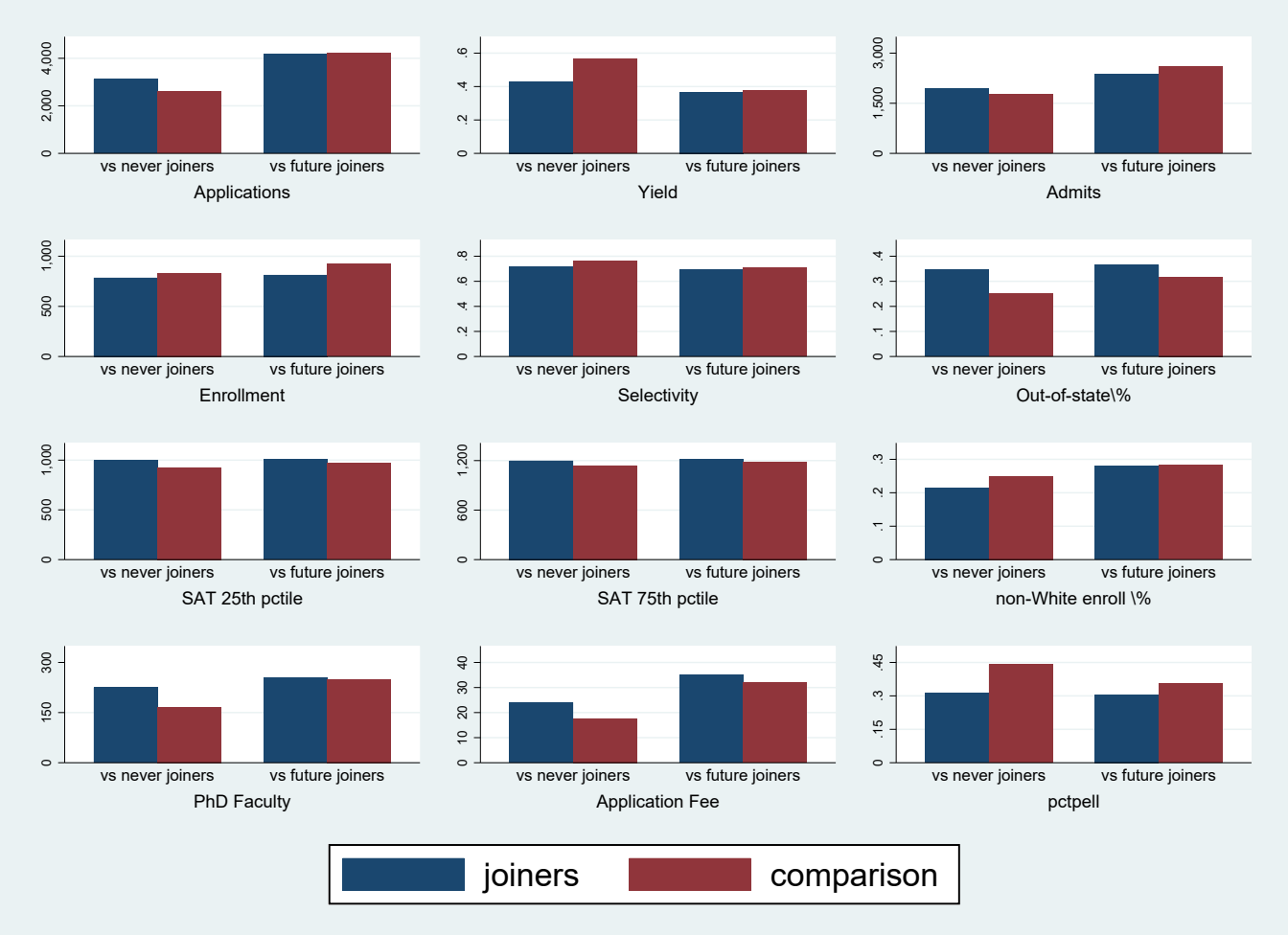
joiners and future joiners for every join year j, and then appended each join year's data into a single dataset. The resulting dataset has some duplicate observations since the same school-year may serve as a comparison observation for multiple join-years. For example, a school joining in 2004 is a comparison observation for schools joining from 1999-2001. In all specifications we cluster standard errors at the school level and therefore these duplicates do not affect inference. Additionally, since we compare joiners to future joiners, most schools serve as both treated and comparison observations, over different join years j. Therefore, following Deshpande and Li (2019), we also include an additional indicator for whether a school is a joiner for a specific join year j as a comparison. The regression model using this strategy is:

$$ln(y_{jct}) = \beta * (CA_{cjt}) + \mu_c + \mu_t + \alpha * 1(J_c = j) + \varepsilon_{jct} \tag{29}$$

The corresponding event-study specification is:

$$ln(y_{jct}) = \sum_{k=-w}^{w-1} \beta_k [1(J_c = j) \times 1(t - J_c = k)] + \mu_c + \mu_t + \alpha * 1(J_c = j) + \varepsilon_{jct} \tag{30}$$

Figure A8: Pre-join Differences for each Identification Strategy



Notes: First set of bars compares joiners to never joiners using 1990 values (1999 for Pell%). Second set of bars compares joiners to future joiners, using values from five periods before joining.