THREE ESSAYS ON APPLIED MICROECONOMICS

BY

BRENO RAMOS SAMPAIO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Darren H. Lubotsky, Chair
Professor Elizabeth T. Powers
Professor Kristine Brown
Professor Ron Laschever

ABSTRACT

This thesis is divided into three separate chapters. In the first chapter, I analyze the relationship between age and college entrance test score and enrollment. In the second chapter, I focus on how cell phone bans affect driving fatalities rates. Finally, the third chapter studies nationalistic bias in the professional division of surfing. In the next three paragraphs, I summarize these three chapters.

The consequences of single-date school entry systems, which generate a large age difference between children in the same class, are now a widely studied subject. Published research has shown that older children consistently outperform their younger counterparts in several outcomes while in elementary and in the beginning of high school, however, evidence is weak when one considers long-run outcomes such as wages or the probability of being employed. In this chapter I use data from a major university in Brazil to investigate whether age differences significantly affect students' college entrance test scores and their probability of being accepted for higher education. Results show that older students outperform younger students on test scores and, more importantly, this difference significantly affects their likelihood of being accepted in college. These results suggest that age differences might have important long-run effects given its direct link to students' access to higher education.

There has been significant amount of research in the transportation area on the development of strategies that allow good comparisons between states, such that policy analysis are allowed to be carried out and informative policy-oriented questions are allowed to be answered. In this chapter I propose the use of Synthetic Control Methods (SCM) to overcome several identification problems present in previous studies when constructing comparison groups/counterfactuals. I apply the SCM to analyze the effect of New York State's law

prohibiting handheld cell phone use while driving on fatalities rates. Results show that (i) a convex combination of the states of IL, MA and TX provide a good (synthetic) "peer state" for NY when evaluating this specific policy and (ii) that imposing the ban lead to a decrease of about 9% in fatality rates.

The past two decades has seen an increasing interest in detecting and quantifying hidden actions taken by agents when facing decisions that may lead to higher individual payoffs but are not easily observed by all parties involved. One such area that has recently received a lot of attention is on understanding the decision making process of professional referees in sports. In this chapter I estimate nationalistic bias using data from the world's elite division of professional surfing. Different from previous sports analyzed in the literature, surfing competitions are composed mainly by man-on-man heats with surfers having as many as 15 performances scored by the same judging panel in each heat, allowing one to overcome the main difficulty encountered by previous researchers which is to correctly identify whether judges misbehavior is a result of preferences over the way athletes perform in a specific country or is driven by intentionally misreporting scores to benefit a fellow countrymen. Different from what has consistently been reported in the literature, results show that surfing judges neither underscore nor overscore their fellow compatriots. However, they significantly underscore athletes competing against their countrymen. Also, the score given by the judge scoring the surfer competing against his countrymen is statistically smaller than all other judges' scores when the compatriot is losing, but statistically the same when the compatriot is winning, which supports the idea that judges' bias is a result of strategic behavior. Finally, the large score penalties are shown to have a significant effect on final heat positions and, as a consequence, final points and prizes earned in the tournament.

To Cecilia, My Parents, and My Brother and Sister

ACKNOWLEDGEMENTS

This work has benefited from the input of many people and I wish to express my gratitude to all of people who helped me in making this research possible. I owe my deepest gratitude to my advisor, Professor Darren Lubotsky, for his continued mentoring, support, and invaluable guidance throughout graduate school; for the insightful suggestions he provided related to my project; and for teaching many lessons regarding life as a researcher. I am also grateful to the other members of my thesis committee, professors Kristine Brown, Ron Laschever, and Elizabeth Powers, for their time and effort in providing me professional advice related to my research. I am thankful to the Department of Economics and the Lemann Institute for Brazilian Studies at University of Illinois at Urbana-Champaign, which granted me the financial means to complete this project.

I especially wish to thank my Professor and friend Werner Baer, for his guidance throughout graduate school and for his lifetime dedication in supporting and educating Brazilian students in the last few decades. Brazil is now a better place to study and discuss economics because of his contribution.

My colleagues at UIUC, such as Joanna Alexopoulos, Fabricio Almeida, Marco Aurelio, Diloa Athias, Monserrat Bustelo, Igor Cunha, Joao Bernardo Duarte, Luiz Figer, Seyed Karimi, Renato Lima, Leonardo Lucchetti, James Pinkstaff, Rafael da Matta, Euler de Mello, Rafael Ribas, Leandro Rocco, Paulo Vaz, among many others, were always available to discuss methods, data, interpretations of parameters, among many other details that surely shaped and improved my skills as a researcher. I also would like to thank my former professors and academic friends Anisio Brasileiro, Tiago Cavalcanti, Juliana Guimaraes, Luiz Renato Lima and Oswaldo Lima Neto. Their contribution had immense impact on my academic choices.

Finally, I wish thank my parents, Yony and Silvia, for the excellent education and values and for the everyday discussions that sharpened my analytical skills, my brother and sister in law, Gustavo and Andrea, for always being there for me, my sister, Sofia, and my lovely wife, Cecilia, for her support both professionally and emotionally. Without their support none of this would have been possible.

TABLE OF CONTENTS

CHAPTER 1
THE EFFECT OF AGE ON COLLEGE ENTRANCE TEST SCORE AND ENROLLMENT: A REGRESSION DISCONTINUITY APPROACH[1]

1. Introduction

The consequences of being among the older children in a class are now a widely studied subject. Researchers have shown that older children outperform their younger counterparts on standardized test scores in kindergarten (Bedard and Dhuey, 2006; Elder and Lubotsky, 2009; Puhani and Weber, 2006; McEwan and Shapiro, 2008), in the middle and end of primary education (Puhani and Weber, 2005; Bedard and Dhuey, 2006; Elder and Lubotsky, 2009; McEwan and Shapiro, 2008; Smith, 2009) and, recently, in the middle of secondary education (Smith, 2009). With the exemption of McEwan and Shapiro (2008), all published work looking at the evolution of age effects find that it tends to decline over time. Bedard and Dhuey (2006), for example, show that younger children score 4-12 percentiles lower than the oldest children in 4th grade and 2-9 percentiles lower in 8th grade. Elder and Lubotsky (2009) find that being a year older at the beginning of kindergarten leads to an increase of about 0.530 standard deviations (σ) and 0.840σ in reading and math scores, respectively, and that this age advantage tends to decline or even disappear by 3rd grade for the poorest children and be very small by 8th grade for wealthier children. Smith (2009), extending the analysis for high school students, show that older children outscore younger children by about 0.259-0.400σ in 4th grade, and by about 0.104-0.242σ in 10th grade.

With respect to long-term outcomes, researchers have shown that entry age differences do have impacts on teenage pregnancy reduction (Black, Devereux, and Salvanes, 2008), on participating in pre-university program during the final year of high school and on the probability

[1] This work is co-authored with Rafael da Matta, Rafael Perez Ribas and Gustavo Ramos Sampaio.

of taking the SAT and enrolling in college (Bedard and Dhuey, 2006). Less consistent are findings regarding educational attainment and future earnings. Puhani and Weber (2005) find that children who enter school at seven instead of six years of age gain almost half year more of secondary schooling; Fredriksson and Ockert (2005) find a negative impact on wages for the youngest individuals in a cohort in Sweden; Black, Devereux, and Salvanes (2008) find that being among the oldest children at the start of kindergarten has no effect on educational attainment and earnings in Norway. Recently, Dobkin and Ferreira (2010), looking at the tradeoff between being the younger student in the cohort, which typically have poorer academic performance as shown above, and having slightly higher educational attainment due to compulsory schooling laws,[2] find no evidence that the age differences affects job market outcomes such as wages or the probability of employment.

According to Elder and Lubotsky, two different mechanisms with potentially different long-run impacts could lead older children to outperform younger children at younger ages. On one hand, a one year difference in children's entrance age at this stage in life could lead to greater maturity, which could have strong effects on learning capabilities, leading older children to perform better on the beginning and, if the age effect persists,[3] throughout the rest of the schooling process with possible effects on future wages and employment. On the other hand, age differences may arise only because older children spent more time at home and thus start school with a greater accumulation of knowledge. If one believes the latter mechanism is the true one and that all children tend to learn at the same rate, then the age gap should decrease over time

[2] Angrist and Krueger (1991) showed that individuals born in the 1st quarter of the year have lower education attainment when compared to individuals born in the 4th quarter of the previous year. This is due to the fact that children born in the 1st quarter of the year are more likely to start school older than children born in the 4th quarter. This, in turn, makes 1st quarter born children have, on average, less schooling, since school dropout increases significantly after children obtain the minimum schooling age allowed to withdraw.

[3] See Allen and Barnsley (1993) for an interesting description on how age effects can persist to adulthood on hockey league in Canada.

and children from wealthier families should be the ones benefiting the most from entering school at older ages. Elder and Lubotsky`s findings indeed point in the direction of the latter case, i.e., that age differences arise from pre-kindergarten knowledge accumulation and should vanish in later grades. Thus, as they point out, delaying school entry wouldn't generate any benefit in order to compensate the high cost paid in terms of lost future working years, additional child-care costs, and potential reduction on educational attainment.

In this paper we do not seek to analyze which of the two mechanisms are driving age differences. Our main objective is to look until when do age differences matter, whether as a consequence of maturity or pre-kindergarten accumulation of knowledge. More specifically, we extend the age effect analysis for students graduating from high school and estimate their probability of being accepted for higher education using a regression discontinuity (RD) design. Our results show that older students outperform younger students on test scores by .082σ and, more importantly, this difference increases the likelihood of being accepted in college by 20%. Thus, if age differences do arise as a result of pre-kindergarten knowledge accumulation, as pointed out by Elder and Lubotsky, then our results suggest that delaying school entry would indeed generate benefits in the long-run since age effects do persist until the end of secondary education, significantly affecting higher education admissions.

With respect to how different are age effects by gender and parents education, we find that maturity gaps seem to affect boys and girls in similar ways. On the other hand, age effects do differ substantially across different levels of parent education. Being the older in class benefits significantly students coming from families where parents have at most completed primary education. They have a considerable advantage in the college entrance test scores and, as a consequence, are much more likely to be accepted in higher education compared to younger children from similar educational environments. These results suggest that children coming from better educational background (more educated parents and, as a consequence, better schools, for example) are eventually having the gap between older and younger students closed. This, however, is not happening for students coming from worse educational background. This result highlight the need to develop policies, such as directing additional resources to younger students (specially those coming from public schools), that seek to eliminate this huge gap that imposes a barrier on younger students' access to higher education.

The remainder of this paper is organized as follows. Section 2 discusses the empirical strategy. Section 3 describes the data used in the analysis and Section 4 presents the results. Finally, conclusions are presented in Section 5.

2. Empirical Strategy

In this section we present the empirical strategy we use to identify the causal effect of age on college entrance test scores and on students' probability of being accepted for higher education.

Let Y_{ip} be the outcome of interest of student i applying for program p, X_{ip} be vector of personal and family background characteristics such as gender, parents education, race, among others, μ_p be a program fixed effect, and ε_{ip} be unobserved determinants of outcome such as ability. The parameter of interest is given by ρ and represents the marginal effect of age on Y.

(1.1) $$Y_{ip} = \sigma_0 + \rho * AGE_{ip} + X_{ip}\Gamma + \mu_p + \varepsilon_{ip}$$

Omitted variables, such as ability, will bias OLS estimates of the age effect (ρ) as long as it is correlated with the age at which students complete their secondary education, i.e., if $\mathrm{Cov}(\varepsilon_{ip}, AGE_{ip} \mid X_{ip}, \mu_p) \neq 0$. To see why this might be the case, consider first that students' school entry age depend on which time of year they where born, their family background[4] and personal characteristics, and their unobservable variables, such as ability. Parents do have some control over their child's entry age and, given they “observe” their ability, it is likely that the choice to delay entry or not is correlated with it. Also, less able children are more likely to repeat grades while in primary and secondary education,[5] which implies that less able children are more likely to finish high school older. Thus, the existence of a correlation between ability and age implies that OLS will deliver biased estimates of the causal effect of age on the outcomes of interest.

In order to correctly identify ρ we use a Regression Discontinuity (RD)[6] design. This research design uses the laws that regulate the minimum age at which students are eligible to enroll in primary school as a source of exogenous variation in the school entry age. The law in Brazil requires that children enroll in first grade on the year they turn 7, which makes a child born on December 31st to be the youngest one in the class and a child born on January 1st to be the oldest. Thus, the identification strategy relies on comparing the performance of students born right after the cutoff with the performance of students born right before, given the assumption that being born on December 31st or January 1st is a random event which generates a 1 year difference in school entry age.

Since parents have some control over their children school entry age and also less able children are more likely to repeat grades while in primary and secondary education, which generates a correlation between the error terms and the variable of interest as described above,

[4] Machado (2008), using data for Brazil, shows that poorer children enter school, on average, older than the rest of the population.
[5] In Brazil, the problem of grade retention is a very pronounced fact. In 2004, for the age cohort of 11 to 14, which should be enrolled in grades five to eight, 29% were still in grades one to four (Soares, 2006; and Love and Baer, 2009). On the other side, there is little or no grade promotion in Brazilian school, thus more able children are not skipping grades.
[6] See Imbens and Lemieux (2008) and Lee and Lemieux (2009) for a complete description of RDD.

we choose to estimate the model using a *fuzzy* regression discontinuity setup. Thus, we estimate the causal effect of age by the following model via two-stage least squares (2SLS):[7]

$$Y_{ip} = \beta_0 + \beta_1 * AGE_{ip} + f(Bday_{ip}, Cut_{ip}) + X_{ip}\Theta + \mu_p + \varepsilon_{1ip} \quad (1.2)$$

$$AGEip = \delta_0 + \delta_1 * Cut_{ip} + f(Bday_{ip}, Cut_{ip}) + X_{ip}\Omega + \lambda_p + \varepsilon_{2ip} \quad (1.3)$$

where Cut_{ip} is a dummy variable that takes value equal to one for students who were born after January of year X, $f(Bday_{ip}, Cut_{ip})$ is a second order polynomial fully interacted with Cut_{ip},[8] μ_p and λ_p are program fixed effects, and X_{ip} is a vector of personal and family background characteristics. As pointed out by Imbens and Lemieux (2008), the inclusion of additional covariates, X_{ip}, serves to improve precision and, in some sense, to evaluate the "plausibility of the identification strategy" given one should expect estimates of the treatment effect to remain the same as a results of the treatment assignment.

If, however, compliance with the law was perfect and there was no retention of the worst students, i.e., children were not allowed to delay entry neither to repeat grades, a *sharp* regression discontinuity design would correctly identify the effect of age and one would just need to estimate the following model,

$$Y_{ip} = \alpha_0 + \alpha_1 * Cut_{ip} + X_{ip}\Lambda + \phi_p + \nu_{ip} \quad (1.4)$$

which, for comparison purposes, we also provide results for.

An important issue one has to worry about that might compromise identification is the precise choice of birth dates among families with different characteristics, i.e., higher income families might be targeting their children to be born at one side or the other of the cutoff. Thus, for the RD strategy to generate consistent estimates, all families' characteristics should evolve

[7] Following Dobkin and Ferreira, we use a parametric approach since schools entry cutoff is based on the discrete variable age, which is measured in days (see also Lee (2008) and Lee and Card (2008)).

[8] We specify a second order polynomial following McEwan and Shapiro (2008) and Dobkin and Ferreira (2010). We also estimate the models using higher order polynomials, however, we only report estimates using the second order polynomial since results remained statistically the same.

smoothly across the discontinuity. Unfortunately, checking if children born around the cutoff have similar unobservable family characteristics is unfeasible. However, we can test smoothness across observables. This can be done by estimating equation (1.4), but using observable characteristics as dependent variables, and checking whether $\alpha_1 = 0$. In this case, failure to reject the null hypothesis implies that observed covariates vary smoothly around the cutoff (McEwan and Shapiro, 2008).

For comparison purposes, as in Smith (2009), we also estimate age effects using the instrumental variables (IV) approach, as it's done in most papers in the literature described above. In this case, the identification strategy relies on using the age a student should have if he had complied perfectly with the initial assignment given by law and had not repeated any grade as an exogenous variation in his actual age.[9] In this case, the model to be estimated via 2SLS is given by

(1.5) $$Y_{ip} = \gamma_0 + \gamma_1 * AGE_{ip} + X_{ip}\Phi + \mu_{1p} + \eta_{1ip}$$

(1.6) $$AGE_{ip} = \theta_0 + \theta_1 * AGE^9_{ip} + X_{ip}\Psi + \mu_{2p} + \eta_{2ip}$$

where AGE^9_{ip} is students predicted age. Two conditions must be satisfied for consistently estimating the parameter of interest γ_1: predicted age must be correlated with actual age; and predicted age must not be correlated with unobserved determinants of Y_{ip}, i.e., $Cov(\eta_{1ip}, AGE^9_{ip} \mid X_{ip}, \mu_{1p}) = 0$. The latter condition is violated as long as parents that are statistically different in terms of socioeconomic conditions target their children to be born at different times of the year.[10] If this is the case, then predicted age is likely to be correlated with unobservables, for example, compromising the IV estimation.

[9] See Bedard and Dhuey (2006) and Elder and Lubotsky (2009) for a more detailed discussion on the validity of the identification strategy.
[10] See Bound and Jeager (2000) and, more recently, Buckles and Hungerman (2008), for evidences on the correlations between season of birth and family background, education, and earnings.

In the next section we give a brief overview of the data used in the analysis as well as a few concerns regarding its use in identifying the effect of interest.

3. Data

The data used in this study comes from students applying for the *Universidade Federal de Pernambuco* (UFPE), which is the major university in the Northeast of Brazil.[11] UFPE is the main public university in the state of *Pernambuco* and mainly all students graduating from high school who seek higher education apply for it, given there are no tuition fees.

The main entrance requirement in the undergraduate programs is an exam (*vestibular*) that must be taken by all candidates and occurs only once a year. The exam evaluates students in many subjects, such as: Mathematics, Portuguese, a foreign language (either English, French or Spanish), History, Geography, Physics, Chemistry, and Biology. There are two rounds. In the first round, students are evaluated in all subjects and the score in this round is an average of the performance in all subjects. In the second round, the subjects are major specific and the final score is a weighted average of the score in the first and second round. In order to be able to compare all students, we use only the scores of the first round when estimating age effects. Note that all students are required to choose their major before they take the exam and for this reason we include program fixed effects in the equations to be estimated.

The data includes 54,877 and 50,160 students applying for the university in the years of 2004 and 2005, respectively. We restrict our sample for all students who were graduating from high school.[12] In addition, we exclude all students who were graduating from high school using

[11] The *Universidade Federal de Pernambuco* is, according to the Ministry of Education, the major university in the North and Northeastern regions of Brazil. The university has 62 different undergraduate programs and 108 graduate programs. In 2004, the university had 25,000 students registered (20,500 in undergraduate programs and 4,500 in graduate programs) and 1,647 professors.

[12] There are many students who have graduated from high school before 2004 and are taking the exam again either because they have been accepted for higher education and are trying to switch courses or because they have not yet been accepted before and are still trying to get admitted.

the *supletivo* method. The *supletivo* method is an alternative way to get a high school degree for students with a large age/grade distortion. The system offers short-term courses with a condensed material for different grades, where students are allowed to get, for instance, a secondary school diploma, which in Brazil normally takes 3 years, in a one year course. With these restrictions, we end up with a total of 12,636 and 10,530 students in the years of 2004 and 2005, respectively.

One main concern with the available data is that we only observe students who decided to apply for higher education, i.e., the data contains no information whatsoever on the students who decided to drop out of school before high school graduation or on the students who graduated but decided not to apply for college. This problem will not compromise identification if the probability of dropping out from school and not applying for college (for the ones who graduate) is independent of students' birth date, which is very unlikely to be true. To see this, figure (1.1) shows the histogram of students by birth date, where it can be observed that there is a slight decrease in the number of students born towards the end of the year. This difference is already expected since, according to evidences provided by Elder and Lubotsky (2009), being a year younger at kindergarten entry increases by about 13% the probability of repeating kindergarten, first or second grade.[13] This points in the direction that attrition is more likely to happen for the students with the lowest performance in class, which are, on average, the youngest ones. Thus, among those, the low ability ones will drop out with higher probability leading our estimates of the average test scores for the population of younger students to be overestimated. This, in turn, will cause our RD estimates to be biased downward, since there is a selection of the more able students among the younger ones in class.

4. Results

[13] Similar results are also found in McEwan and Shapiro (2008).

First we present two figures in order to motivate our identification strategy. Figures (1.2) and (1.3) plot the relationship between high school graduation age and birth date and normalized entrance test scores and birth date, respectively. We can observe that there is a discontinuity both in the age and in the (normalized) test score distribution by comparing students born in January of 2005 with students born in December of 2004. Figure (1.2), for example, show that students born in January are approximately .8 years older than students born in December. Note that if compliance with entry laws was perfect and students were not allowed to repeat grades, the age difference between the students born in the beginning of January and the ones born in the end of December would have been exactly 1 year, which clearly is not the case. Thus, the simple comparison of test scores between students born in January and December, as observed in figure (1.3), would not capture correctly the effect of interest, which justifies our choice to use a fuzzy RD set up.

As mentioned in section 2, a main concern about the validity of the identification is if parents are targeting their children birth dates to be on one side or another of the discontinuity. To test if this is the case, we estimate equation (1.4) using observable characteristics as dependent variables to check whether $\alpha_1 = 0$ or not. The test indicates that we fail to reject the null hypothesis, which implies that variables do appear to be evolving smoothly through the cutoff. The estimated coefficients (α_1) for gender (female), family income, and parents' education are, respectively, -.010 (.032), .189 (.125), and -.001 (.277). Thus, this result, besides not excluding that unobservables evolve smoothly across the cutoff, gives us some confidence about the validity of our identification strategy.

4.1. Age and College Entrance Test Scores

In this section we present estimates of the effect of age on college entrance test scores. We start by discussing estimates for the first stage and then proceed to the results of the second stage regressions.

Table (1.1) presents results of the first stage regressions. Columns 1, 2, and 3 presents results for the first stage given by equation 2, the fuzzy RD approach (*F*-RD), and column 4 for the first stage given by equation 5, the IV approach. Column 1 only includes the polynomial while in columns 2 and 3 we add additional covariates and program fixed effects, respectively. The estimated coefficients do not change significantly with the inclusion of the additional covariates or program fixed effects. Thus, students born at the beginning of the year are, on average, .83 years older than the students born at the end of the year. Similar results are obtained when using predicted age as instrument for actual age.

Table (1.2) presents results for the age effect estimated through a naive OLS, a sharp RD, a fuzzy RD, and the traditional IV method. The OLS estimates show that age has a positive effect on entrance test scores, i.e., an additional year on entry age increases test scores by about 0.023σ. However, as described above, OLS estimates are potentially biased downward due to a negative correlation between ability and high school graduation age.

The OLS bias is confirmed by the RD and IV estimates presented in columns 2 to 5. In column 2 we present the sharp RD estimates which imply an age effect of about $.069\sigma$. This estimate, however, should be taken as an underestimate of the true effect given there is a selection resulting from students' retention. In column 3 and 4 we present the fuzzy RD estimates, column 3 controlling for the polynomial function and covariates and column 4, our preferred regression, adding program fixed effects to the specification of column 3. Results from the fuzzy RD approach show that being one year older increases students' performance by .082

standard deviation, which is similar to the estimates given in Smith (2009) for 10th graders (0.104-0.242σ).

The IV estimates, on the other hand, are smaller, however, very significant. The difference between the RD and the IV estimates provides evidences of a negative bias in the IV estimates resulting from a potential correlation between predicted age and unobservable determinants of test scores, as described in section 2. To support this conjecture, we regress students' socioeconomic characteristics on predicted age to test for a significant correlation that might result due to parents targeting their children to be born at certain periods of the year. As expected, family income is positively and significantly correlated with predicted age, a result found by Machado (2008) using data from Brazil. This shows that our RD strategy is indeed preferable to the traditional IV strategy widely used.

As described in section 3, in the first round all candidates are evaluated in Mathematics, Portuguese, a foreign language (either English, French or Spanish), History, Geography, Physics, Chemistry, and Biology. Thus, we are able to identify how the effect of age varies across all different subjects. Table (1.3) presents results for the fuzzy RD estimates. Contrary to results presented by Elder and Lubotsky (2009) for primary school children and Smith (2009) for 10th grade students, age appears not to matter for students performance in math and Portuguese. However, there is a large age effect for Biology, History, Physics, Geography, Chemistry, and foreign language.

4.2. Age and the Probability of Being Accepted for Higher Education

In this section we examine the entry age effect on the probability of being accepted for higher education. We should emphasize that being approved depends on the career chosen by each student, which we cannot control for, given we only observe the career choice for the students who were accepted at the university. However, we still include in the regressions program fixed effects, which captures in part differences in the probability of being accepted since, for example, students applying for the college of health face, on average, greater competition than students applying for the college of social sciences.

The question of whether age differences matter for students when deciding to pursue higher education or not has been analyzed by Berdard and Dhuey (2006). In their analysis, they use two data sets, one from the Canadian province of British Columbia (BC) and the other from the United States. The data from BC contains a five year period record on students who entered grade nine. In order to look at age effects on college pre-university behavior, they define the concept of "university-bound" students, which are students observed in grade 12 who report having graduated by June of the fifth year after they enter grade nine and have taken at least four examinable subjects and earned a 75 percent average or higher. The rationale behind this definition comes in part from the fact that "admission into one of the flagship provincial universities students must generally take at least four examinable subjects and score in excess of 75 percent." Using this definition, they obtain that older students are 9.8-12.8 percent more likely to be university-bound than younger students. The data from US, on the other hand, has information on whether the students took the SAT exam, ACT exam, or both, and whether the students have enrolled in a four-year accredited college/university during the two years following graduation. Results show that older students are 7.7-18.6 percent more likely to have taken the SAT or ACT, and 11.6-27.0 percent more likely to enroll in an accredited four-year college/university. In our analysis we observe all students that actually apply for higher education and we benefit from a system where all students are required to take the entrance exam in order to be accepted at the university. Thus, we can better estimate the probability of being accepted for higher education by age groups.

Figure (1.4) plots the relationship between the proportion of students accepted for higher education and their birth dates. One can observe that among the students born in the beginning of the year, about 9.5% are accepted for college, while among the ones born in the end of the year, only about 8% are accepted. Thus, if this difference is statistically significant, as is confirmed by our estimations below, being among the older children in class have a strong effect on students' access to higher education.

In Table (4.1) we present the results of the age effect on the probability of being accepted for college. Again, we present coefficients estimated by OLS, by a sharp and a fuzzy RD, and by the IV method. The OLS estimate, as expected, is biased downward as is also the sharp RD estimate. The fuzzy RD and the IV estimates are exactly the same value implying that being among the older children in class increases the probability of being accepted by 1.6 percentage points, that is, an increase from 8% to 9.6%. This corresponds to an increase of about 20% in the chances of a students being accepted in college if he had delayed school entry by one year. This result is exactly in the same range as the results obtained by Bedard and Dhuey (2006), thus strengthening the conclusions derived from it.

4.3. Age and its Effects by Gender and Parents Education

In this section we look at how different are estimates of the age effect for males and females and for different levels of parents education. Table (1.5) presents results for gender and parents education of the effect of age on students' college entrance test scores (ETS) and on students' probability of being accepted for higher education.

Age effects appear not to differ substantially between males and females. Besides males having a higher coefficient for college entrance test scores, there is a slightly and not statistically significant difference in the probability of being accepted for higher education. This similarity between gender effects is also obtained by Smith (2009), when considering children at fourth grade. McEwan and Shapiro (2008), however, find that a one-year increase in enrollment age improves fourth grade boys test scores by about one-third more than girls. Looking at students in the middle of high school, Smith (2009) find contrary results to what we find, that is, he obtains that age differences matter significantly for girls but not for boys. Our results point in the direction that maturity gaps do not close faster for boys neither for girls, they seem to be affected equally even for students graduating from high school.

With respect to parents' education, it appears that in terms of college entrance test scores, being the older in class benefits significantly students coming from families where parents have at most completed primary education. Among the students coming from families that have completed secondary education or have acquired a college degree, older students still benefit from delayed entry but not as much as children coming from less educated environments. Elder and Lubotsky (2009) and Smith (2009) look at a similar effect but are interested on how age effects vary by income. Smith (2009), for example, looking at different quartiles of the distribution of average neighborhood household income, find that for 10th grade students age effects are much stronger among students in the lowest income quartile compared to the highest. Our results, besides using a slightly different measure (parents education), show a similar pattern, however, across all levels of parents education age effects appear to matter while, according to Smiths results, the highest income quartile students appear not to benefit at all from being one year older.

When considering the probability of being accepted for higher education, our results show that age effects decrease monotonically across different levels of parents education. Children from families whose parents have at most completed primary education benefit the most while children whose parents have a college degree do not benefit at all. This result suggests that better educated parents appear to be concerned about the performance of their children when being the younger in their class and are trying to compensate this disadvantage with additional support (for example, by hiring additional private tutoring). Another explanation is that children coming from better educational background study, on average, in better schools in Brazil compared to children coming from poorer families. Thus, it might be that better schools are directing resources towards the youngest students in order to compensate their grade disadvantage, while the worst schools (usually public ones) are not doing anything, hence the age effect perpetuates. Thus, an important policy regarding the disadvantage incurred by being the youngest in class is for schools (and parents) to direct additional resources for younger students to try to eliminate this huge gap that imposes a barrier for them to acquire a college degree.

5. Concluding Remarks

Researchers have shown that older children outperform their younger counterparts on standardized test scores on kindergarten, on the middle and end of primary education and, recently, on the middle of secondary education. With respect to long-term outcomes, evidence is still mixed with results varying substantially across countries. In this paper we extend the age effect analysis for students graduating from high school and estimate by how much do older students outperform younger students and if this score advantage significantly affects their likelihood of being accepted for higher education using a regression discontinuity (RD) design.

As in Smith (2009), we also estimate age effects using the traditional IV approach for comparison purposes.

According to Elder and Lubotsky (2009) findings, age differences appear to be arising from pre-kindergarten knowledge accumulation and should vanish in later grades, making school entry laws unimportant in terms of generating long-run effects. However, our results point in the direction that delaying school entry would indeed generate benefits in the long-run since age effects do persist until the end of secondary education significantly affecting higher education admissions. More specifically, we find that age differences appear to matter for students graduating from high school where older students outperform younger students on test scores by .082σ and, more importantly, this difference increases by 20% the likelihood of being accepted in college.

With respect to how different are age effects by gender and parents education, we find that maturity gaps do not close faster for boys or for girls, they seem to be affected in similar ways. On the other hand, age effects do differ substantially across different levels of parents education. Being the older in class benefits significantly students coming from families where parents have at most completed primary education. They have a considerable advantage in the college entrance test scores and, as a consequence, are much more likely to be accepted in higher education compared to younger children from similar educational environments. These results suggest that children coming from better educational background (more educated parents and, as a consequence, better schools) are eventually having the gap between older and younger students closed. This, however, is not happening for students coming from worse educational background. Thus, an important policy regarding the disadvantage incurred by being the youngest in class is for schools (and parents) to direct additional resources for younger students to try to eliminate this huge gap that imposes a barrier on their access to higher education.

6. Figures and Tables

Figure 1.1: Histogram by Birthday

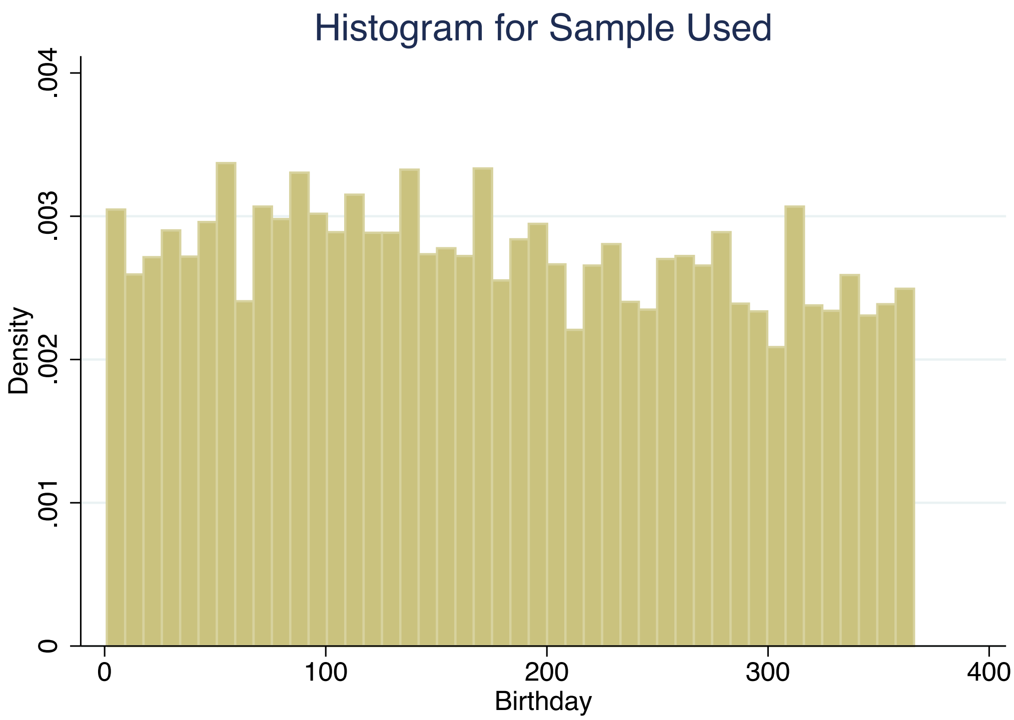


Figure 1.2: High School Graduation Age by Birthday

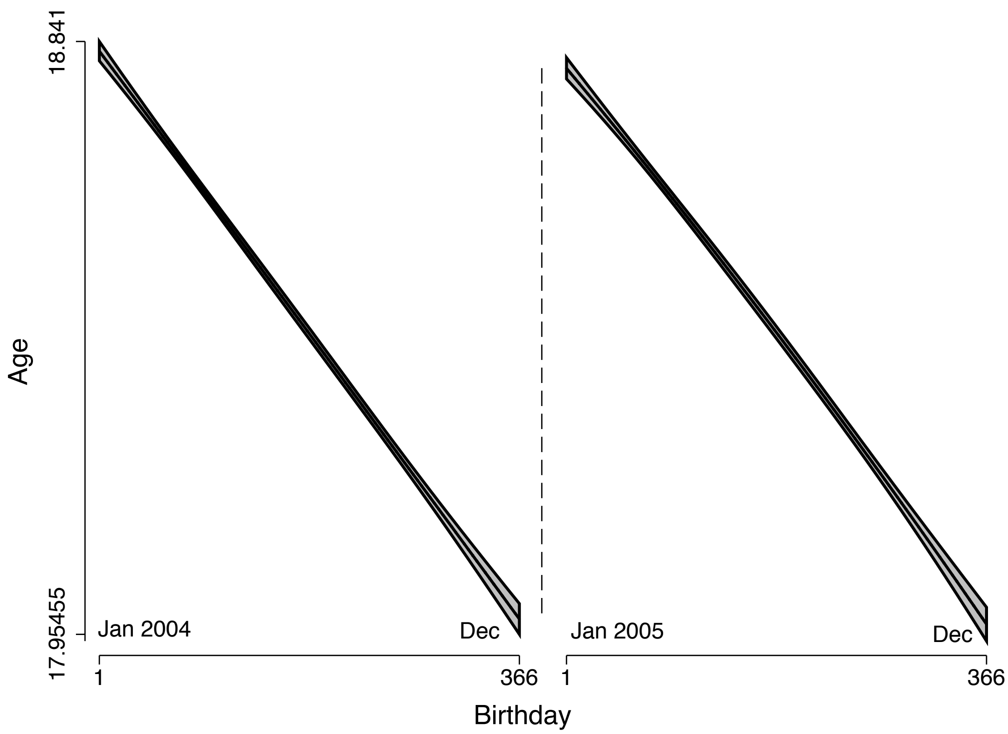


Figure 1.3: Normalized Entrance Test Score (Z-Score) by Birthday

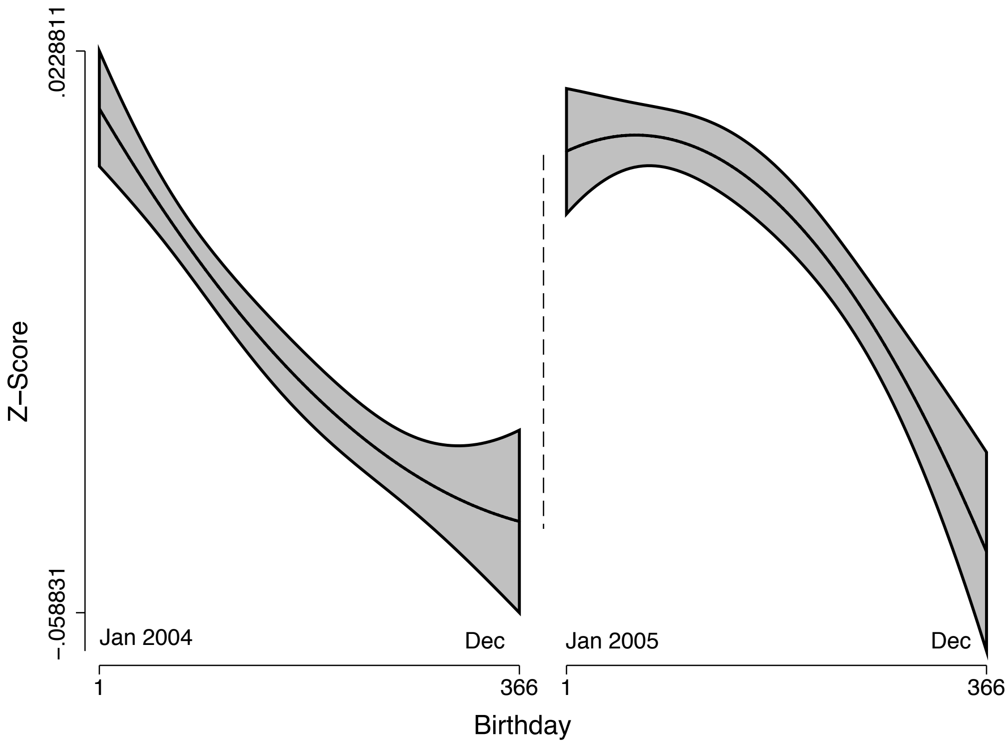


Figure 1.4: Probability of Enrollment by Birthday

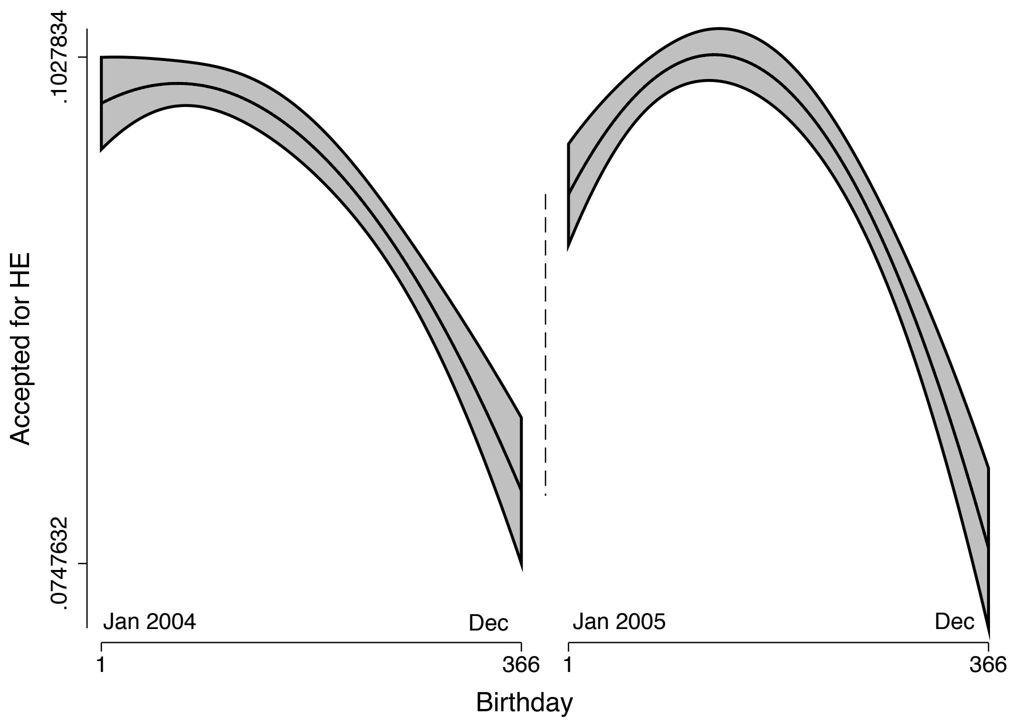


Table 1.1: First Stage Estimates

Variables	F-RD			IV
	(1)	(2)	(3)	(4)
Cut	.862***	.856***	.832***	
	(.015)	(.015)	(.008)	
AGE^9				.818***
				(.012)
Polynomial	YES	YES	YES	NO
Other Covariates	NO	YES	YES	YES
Program FE	NO	NO	YES	YES
N	23,166	23,166	23,166	23,166

Note: The first stage for the fuzzy RD method, represented by equation (1.3), is estimated including only the polynomial function in column 1, adding other covariates and program fixed effects in columns 2 and 3, respectively. Other covariates include gender, parents' education, income, race, and family size. The first stage for the IV approach, represented by equation (1.6), does not contain the polynomial function and the instrument used, the predicted age, is the age a student should have if he had complied perfectly with the initial assignment given by law and had not repeated any grade. Robust clustered standard errors are presented in parentheses. *** represents $p<1\%$.

Table 1.2: OLS, RD and IV Estimates

Variables	OLS	S-RD	F-RD		IV
	(1)	(2)	(3)	(4)	(5)
AGE	.023***		.082***	.082***	073***
	(.004)		(.014)	(.012)	(.006)
Cut		.069***			
		(.006)			
Polynomial	NO	YES	YES	YES	NO
Other Covariates	YES	YES	YES	YES	YES
Program FE	NO	YES	NO	YES	YES
School FE	YES	NO	NO	NO	NO
N	23,166	23,166	23,166	23,166	23,166

Note: The sharp RD (*S*-RD) estimates, estimated using equation (1.4), estimates the age effect directly from the cutoff not considering that parents, for example, have some control over their children entrance age. The fuzzy RD (*F*-RD), on the other hand, considers this effect through the first stage regressions presented in Table (1.1), where the instrument used is given by the cutoff implied by school entry laws. The IV approach uses as instrument for age the predicted age, which is the age a student should have if he had complied perfectly with the initial assignment given by law and had not repeated any grade. All regressions include additional covariates such as gender, parents education, income, race, and family size. The dependent variable in all regressions is the students entrance test score. Robust clustered standard errors are presented in parentheses. *** represents p<1%.

Table 1.3: *F*-RD Estimates for Age by Subjects

Subject	*F*-RD	Subject	*F*-RD
Math	.018 (.015)	Geography	.081*** (.013)
Portuguese	.007 (.014)	Physics	.105*** (.013)
Foreign Language	.049*** (.015)	Chemistry	.113*** (.014)
History	.118*** (.014)	Biology	.123*** (.015)

Note: The fuzzy RD (*F*-RD) instrument used for age is given by the cutoff implied by school entry laws. All regressions include additional covariates such as gender, parents education, income, race, and family size. The dependent variable in all regressions is students entrance test score for all different subjects presented in the table. Robust clustered standard errors are presented in parentheses. *** represents p<1%.

Table 1.4: OLS, RD and IV Estimates

Variables	OLS	S-RD	F-RD		IV
	(1)	(2)	(3)	(4)	(5)
AGE	.007***		.016***	.016***	016***
	(.001)		(.006)	(.005)	(.003)
Cut		.014***			
		(.004)			
Polynomial	NO	YES	YES	YES	NO
Other Covariates	YES	YES	YES	YES	YES
Program FE	NO	YES	NO	YES	YES
School FE	YES	NO	NO	NO	NO
N	23,166	23,166	23,166	23,166	23,166

Note: The sharp RD (*S*-RD) estimates, estimated using equation (1.4), estimates the age effect directly from the cutoff not considering that parents, for example, have some control over their children entrance age. The fuzzy RD (*F*-RD), on the other hand, considers this effect through the first stage regressions presented in Table (1.1), where the instrument used is given by the cutoff implied by school entry laws. The IV approach uses as instrument for age the predicted age, which is the age a student should have if he had complied perfectly with the initial assignment given by law and had not repeated any grade. All regressions include additional covariates such as gender, parents education, income, race, and family size. The dependent variable in all regressions is a dummy variable that identifies if the student was accepted for higher education or not. Robust clustered standard errors are presented in parentheses. *** represents p<1%.

Table 1.5: 2SLS Estimates of Age Effect on ETS by Gender and Parents Education

Variables	2SLS	
	ETS	HE Acceptance
	(1)	(2)
Gender		
- Male	.115***	.014**
	(.019)	(.007)
- Female	.077***	.017***
	(.022)	(.006)
Parents Education		
- Primary	.138***	.025**
	(.024)	(.011)
- Secondary	.076***	.017**
	(.015)	(.007)
- College Degree	.060***	.009
	(.018)	(.008)

Note: All regressions are estimated by a fuzzy RD (*F*-RD) where the instrument used for age is given by the cutoff implied by school entry laws. All regressions, when appropriate, include additional covariates such as gender, parents education, income, race, and family size. The dependent variable in column 1 is students entrance test score while in column 2 is a dummy variable that identifies if the student was accepted for higher education or not. Robust clustered standard errors are presented in parentheses. *** represents $p<1\%$ and ** represents $p<5\%$.

CHAPTER 2
IDENTIFYING PEER STATES FOR TRANSPORTATION POLICY ANALYSIS WITH AN APPLICATION TO NEW YORK'S HANDHELD CELL PHONE BAN

1. Introduction

Many empirical questions in social sciences depend on estimating causal effects of programs or policy interventions. These effects are important since it not only serves as a measure of how effective the program is and the impact it might cause if implemented elsewhere but it also guides new policy designs. For example, understanding how education affects future wages and how laws prohibiting alcohol consumption while driving affects the number of fatal accidents is extremely important for social and political reasons. A critical issue, however, is how to get correct estimates of these causal effects, a question that has been the main challenge faced by statisticians and econometricians in the past decades and is referred to as the "fundamental problem of causal inference" (Holland 1986). This problem exists due to the fact that comparisons of two outcomes of interest for the same unit when exposed, and when not exposed, to a treatment is an infeasible task, given the same unit can either participate or not in a program in the same period (Imbens and Wooldridge 2009). In other words, one never observe a specific state when under and when not under a certain law at the same point in time. To circumvent this problem, the main empirical strategy is to find different units or states (with different treatment status, i.e., some treated - subject to a certain law - and some not) such that after adjusting for differences in other observed characteristics, or pretreatment variables, comparisons are allowed to be made (see Angrist and Pischke (2009) for a detailed explanation on causal inference and on recent developments in the econometrics of program evaluations).

Policy evaluations are of particular importance to the area of transportation, in which knowing policy impacts is a necessary condition to decide whether to adopt a given policy in a different city or state, for example. To overcome the problems described in the preceding paragraph, empirical strategies have been proposed in order to create similar groups (of states, for example) for policy evaluation, in what Hendren and Niemeier (2008) call "peer states." The basic idea is that once states are clustered into peer groups, comparisons are allowed to be made between peer states and informative policy-oriented questions are allowed to be answered. To put in other words, states are selected and clustered into groups based on observed characteristics and, once the groups are formed, comparisons of a certain outcome between states that had a policy change and states that did not have a policy change are taken as causal effects of the change in the law. This approach, however, has an important shortcoming: peer states are selected based on certain characteristics X regardless of the policy one is interested in analyzing. Hence, it could easily be the case that the state of Illinois works as a perfect control (peer state) for the state of Colorado when evaluating the impact of reducing class size on students performance, however it may not be as good peer for the state of Idaho when evaluating the impact of a road expansion on urban growth.

In this paper I contribute in two ways to the current debate on identifying peer states for policy evaluations in transportation research. First, I emphasize that peer groups should be policy-specific. That is, contrary to what is advocated by Hendren and Niemeier, peer groups should not necessarily be the same regardless of the question to be answered (a point previously raised by Hartgen and Neumann, 2002). The second contribution relates to how peer states should be selected. I propose that transportation policy evaluations at the state level should be carried out using the synthetic control method (SCM), a technique developed by Abadie and Gardeazabal (2003) and recently extended by Abadie et al. (2010). This method uses data-driven procedures to construct adequate comparison groups/counterfactuals given that, in practice, it is a difficult task to find a single state unexposed to the policy change of interest that approximates the most, relevant characteristics of the treated state (the state that had a policy change) and that would provide a good control group.

The basic intuition behind the Synthetic Control Method is that a combination of states ó a synthetic control ó offers a better comparison than any single state alone (Belot and Vandenberghe 2011). In other words, when one compare average characteristics of other states in the peer group with the state that had a policy change (which is basically the idea behind the methodology proposed by Hendren and Niemeier), one is implicitly assuming that the “average” states' characteristics are sufficiently similar to the characteristics of the treated state. The synthetic control method, on the other hand, does not assume that states in the constructed peer group have same weights, i.e., the method will provide the researcher with an optimal weight for each state such that the weighted average of the variable one is interested in explaining best approximates the value of this variable for the state that had the policy change. This also implies that several shortcomings of previously proposed methods might be overcome. For example, Hendren and Niemeier find that the states of NY, NC, and WY are each a solo member of their peer group. That is, no other state could be used as a control. This implies that comparisons are unable to be made and impact evaluations of any policy change are impossible to be assessed. Hence, given the advantages of the synthetic control method in creating a (synthetic) peer state, I see it as a promising strategy to overcome some of the shortcomings of previously proposed methods and as a decent instrument to be included in the transportation research toolkit for policy evaluations.

Besides contributing to the literature on identifying peer states for policy evaluation in transportation, this paper contributes also to the current debate about the effects of cell phone bans on driving fatalities. According to data from the National Highway Traffic Safety Administration (NHTSA), more than 10% of the population use a cell phone while driving every day in the United States. This is a very important statistic if one considers that hand-held cell phones are an important source of driver distraction (see, for example, Caird et al. (2008) and Strayer and Drews (2004)), which is thought to be the cause of nearly 80% of automobile accidents and 65% of near-accidents, resulting in approximately 2,600 deaths, 330,000 moderate to critical injuries, and 1.5 million instances of property damage annually in the US (see Cohen and Graham (2003) for more on the consequences of the use of cell phones while driving) (Nikolaev et al., 2010).

This evidence led policymakers across the world to consider whether the use of a cell phone while driving should be regulated or even prohibited. For example, among the developed nations belonging to the Organization for Economic Cooperation and Development (OECD), eight had already enacted in 2001 legislation prohibiting hand-held usage while driving (Cohen and Graham (2003)). By 2009, about 48 countries around the world had imposed such bans. In the United States, New York became the first state to enact legislation banning hand-held cell phones in 2001. It was then followed by the states of California, Connecticut, Delaware, Maryland, New Jersey, Oregon, Washington and, recently, by a few others.

Given the political importance of understanding how the use of hand-held cell phones while driving affect the number of (fatal) accidents, several analysis have been carried out in the past two decades. One of the first attempts was done by Redelmeier and Tibshirani (1997) who analyzed drivers involved in property damage only accidents in the Toronto (Canada) area. They find that the risk of being involved in an accident is four times higher when a cell phone is being used. Similar results were found by McEvoy et al. (2005), using data for Australian drivers, and by Laberge-Nadeau et al. (2002). Looking at the effects of laws banning cell phone use, McCartt and Geary (2004) collected data for the states of New York and Connecticut one month before, shortly after, and 16 months after the imposition of the ban in NY. Their results show that overall cell phone use in NY declined from 2.3% prelaw to 1.1% shortly after, while use rates in Connecticut remained unchanged. However, one year after the ban, cell phone use in NY was 2.1% higher than immediately post-law and not significantly different from pre-law values. Hence, it appears that the initial decrease in cell phone use dissipated during the subsequent year. If this same logic is followed when looking at accidents data, then one might wonder if reductions in the number of accidents right after the imposition of the ban (if they turn out to exist) will perpetuate or vanish in the coming years after the enactment of the law.

In a recent paper, Nikolaev et al. (2010), using data from 1997 to 2007 for the state of New York, compare averages of fatal accidents per 100,000 licensed drivers and personal injury accidents per 1,000 licensed drivers for the period before and after the imposition of the ban. They conclude that the ban reduced both fatal accidents and personal injury accidents. Their analysis, however, suffer from several identification problems as pointed out by Sampaio (2010). The main being that the fatality rate is shown to be decreasing over time, which guarantees that the average of this rate across years after the imposition of the ban is necessarily smaller than the average of this rate in the years before the law. This negative difference, resulting from the natural trend of the data, would lead one to incorrectly reject the hypothesis of no ban effect.

In order to overcome some of the problems that compromise Nikolaev et al. (2010) analysis (such as omitted variables bias), Sampaio (2010) propose a difference-in-difference strategy in which the state of Pennsylvania is used as a counterfactual for the state of New York.

They confirm previously reported results that the ban was effective in reducing fatal accidents per 100,000 licensed drivers. Their analysis, however, do not take into account the fact that the state of Pennsylvania might not be a good control region for the treated state (NY), which might generate bias on the estimated ban effect. This shortcoming is exactly what I aim to overcome in this paper by applying the synthetic control method to obtain a combination of states that best describes pre-treatment variables for the state of New York. More specifically, I study the effect of New York state's law prohibiting handheld cell phone use while driving on fatality rates per 100 million vehicle miles traveled.

Results show that a convex combination of the states of IL, MA and TX provide a decent (synthetic) “peer state” for the state of NY when evaluating this specific policy. Also, I show that the policy caused fatality rates to decrease by almost 9%, on average, and this effect, contrary to what has been previously found, does not dissipate at least in the following 5 years. The analysis presented here improves significantly what has been done in the subject in terms of identifying the interactions between cellular-telephone calls and fatal accidents. These results also shed light on important policy questions, such as the expansion of the ban to other states and to other unregulated countries. These are important questions to be answered given that cell phone subscribers have grown only in the US from 97 million in June 2000 to more than 267 million at the end of the decade.

After this introduction, the rest of the article is organized as follows. Section 2 describes in detail the methodology used in this paper and discusses the shortcomings of previous studies. In section 3 I describe the data and in section 4 I present the results. Finally, in section 5 I discuss the main implications of the analysis and present a few concluding remarks.

2. Methodology

In this section I present the empirical strategy I use to identify the causal effect of New York State's law prohibiting handheld cell phone use while driving on fatality rates. While the focus of this section is on how to identify the impact of the law, when presenting the Synthetic Control Method I discuss in the detail how it may help in constructing adequate peer states for policy analysis in transportation.

Let Y_{st} be the outcome of interest (in our case, fatality per 100 million vehicle miles traveled) for state s at time t, LAW_{st} be a dummy variable that assumes value equal to 1 for the years following the enactment of the law prohibiting handheld cell phone use while driving, and ε_{st} be unobserved determinants of the outcome variable. The parameter of interest, β_1, which represents the effect of the enactment of the law on the outcome, may be estimated via the following model

(2.1) $$\mathrm{Yst} = \beta_0 + \beta_1 * LAW_{st} + \varepsilon_{st}$$

One can easily verify that by estimating equation (2.1) using data only for the treated state (NY), which is basically the strategy carried out by Nikolaev et al (2010), the parameter of interest equals the average of the outcome variable after the enactment of the law (when $LAW_{st} = 1$) minus the average of the outcome variable before the enactment of the law (when $LAW_{st} = 0$), i.e., $\beta_1 = E[Y_{st} \mid LAW_{st} = 1] - E[Y_{st} \mid LAW_{st} = 0]$. It is hard to argue, however, that such difference represents the causal effect of the ban, given there are other factors not controlled for that might compromise identification, that is, it might be that $COV(LAW_{st}, \varepsilon_{st}) \neq 0$. This will be the case if, for example, the fatality rate is decreasing over time, then taking the average of this rate across years after the imposition of the ban and comparing it to the same average taken before the imposition, would not reflect any ban effect but only the natural trend of the data. One may overcome such problem by accounting for time effects in equation (2.1), which would eliminate several sources of bias. However, one may still argue that the estimated ban effect does not reflect the policy change but other unobserved shocks correlated with the law variable. For example, a possible explanation for obtaining a negative β_1 would be the September 11. If one argues that after September 11 drivers across the country started driving less carefully or the police changed its behavior and decreased enforcement or, as suggested by Blalock et al. (2009), there was an additional road travel undertaken in response to 9/11, then the parameter estimated via equation (2.1) would be a joint combination of the ban effect and the September 11 effect.

One solution proposed by Sampaio (2010) to overcome the problems described above is to use data on a second state that did not have any ban imposed during the period analyzed, which would then be taken as a counterfactual for the state of NY. This method, widely used in the economics literature, is called difference-in-differences and its main purpose is to remove biases that might result from permanent differences between the state that had and the state that did not have the imposition of the ban, as well as biases from comparisons over time in the state that had the policy change that could be the result of time trends unrelated to the ban (Imbens and Wooldridge, 2009). In this case, the equation to be estimated is given by

(2.2) $$Y_{st} = \alpha_0 + \alpha_1 * LAW_{st} + \alpha_2 * NY_{st} + \alpha_3 * (LAW_{st} * NY_{st}) + \mu_{st}$$

where NY_{st} is a dummy variable that equals 1 for observations in the state of New York and 0 otherwise. The parameter of interest, α_3, equals the average gain over time in the state not exposed to the ban minus the average gain over time in the state exposed to the ban, i.e., $\alpha_3 = \{E[Y_{st} \mid LAW_{st} = 1, NY_{st} = 1] - E[Y_{st} \mid LAW_{st} = 0, NY_{st} = 1]\} - \{E[Y_{st} \mid LAW_{st} = 1, NY_{st} = 0] - E[Y_{st} \mid LAW_{st} = 0, NY_{st} = 0]\}$. One main hypothesis required for the validity of the approach taken by Sampaio (2010) in identifying the ban effect, is that both treated and control states must have exactly the same time trends in the absence of the ban, which is not clearly why should be the case. If, for example, the control state has a different trend compared to the treated state, the researcher will not be able to differentiate between the ban effect and the trend difference. This shortcoming is exactly what I aim to overcome in the present paper by using the synthetic control method to construct a combination of states that best describes pre-treatment variables for the state of New York, i.e., this artificially constructed group is more similar to the treatment group in the pre-treatment periods than any of the control states on their own.

2.1 The Synthetic Control Method (SCM)

In this section I describe the synthetic control method (SCM) developed by Abadie and Gardeazabal (2003) and extended in Abadie et al. (2010). I also discuss its advantages and limitations when compared to other methodologies used in the literature.

Suppose there are $J + 1$ regions and that only the first region is exposed to the policy change (the state of New York), so that there are J remaining regions as potential controls (all other states that had not enacted any law referring to cell phone use in the period considered). Let Y^N_{it} be the outcome that would be observed for region i at time t in the absence of the intervention, for units $i = 1,...,J + 1$, and time periods $t = 1,...,T$. Let Y^I_{it} be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods $T_0 + 1$ to T, where T_0 is the number of pre-intervention periods such that $1 \leq T_0 < T$. It is assumed that the intervention has no effect on the outcome of interest before the implementation period, such that for $t \in 1,...,T_0$ and all $i \in 1,...,N$ I have that $Y^I_{it} = Y^N_{it}$.

Now let $\alpha_{it} = Y^I_{it} - Y^N_{it}$ be the effect of the intervention for unit i at time t, and let D_{it} be an indicator that takes value one if unit i is exposed to the intervention at time t, and zero otherwise. In this case, the observed outcome for unit i at time t is given by $Y_{it} = YN_{it} + \alpha_{it}D_{it}$. For region one, which is the only region exposed to the intervention after period T_0, it follows that $D_{it} = 1$ for $t > T_0$ and zero otherwise.

Our objective is to estimate $(\alpha_{1T0+1},\ldots,\alpha_{1T})$, which is given by $\alpha_{1t} = Y^I_{1t} - Y^N_{1t} = Y_{1t} - Y^N_{1t}$. The problem in estimating α's in this case is that Y^N_{it} is never observed for the treated region once $t > T_0$. Thus, one must estimate its value. To see how a control group might be obtained from the set of untreated regions, suppose as in Abadie et al. (2010) that Y^N_{it} is given by the following model

(2.3) $$Y^N_{it} = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it}$$

where δ_t is an unknown common factor with constant factor loadings across units, Z_i is a vector of observed covariates (not affected by the intervention), θ_t is a vector of unknown parameters, λ_t is a vector of unobserved common factors, μ_t is an vector of unknown factor loadings, and the error terms ε_{it} are unobserved transitory shocks at the region level with zero mean.

Now consider a ($J\ x\ 1$) vector of weights $W = (w_2,\ldots,w_{J+1})'$ such that $w_j \geq 0$ for $j = 2,\ldots,J + 1$ and $w_2 + \ldots + w_{J+1} = 1$. Each value that W might take represents a synthetic control group for region one. For example, if $w_2 = 1$ and $w_j = 0$ for $j = 3,\ldots,J + 1$, then region 2 works as control for region one (the treated one), which is in the lines of the work of Sampaio (2010). If, on the other hand, one sets a subset $J' \subset J$ to have equal weights, such that $w_{j'} = 1 / J'$ for $j' \subset J'$ and 0 otherwise, one should obtain a similar result to what Hendren and Niemeier propose, given the comparison would be between the treated state and the average of all other states that belong to the peer group. As pointed out in the introduction, the limitation of Hendren and Niemeiers' approach is due to the fact that (*i*) the choice of the weights does take into account the policy one is interested in analyzing and (*ii*) considers the average characteristic of the peer states as a control for the treated state, instead of a weighted average with weights chosen to match the characteristics of the treated state before the intervention, which is what I propose by using the SCM.

Using W as weights to construct a weighted average of equation (2.3), one obtains the following expression

(2.4) $$\Sigma_{j=2}^{J+1} w_j Y_{it} = \delta_t + \theta_t \Sigma_{j=2}^{J+1} w_j Z_i + \lambda_t \Sigma_{j=2}^{J+1} w_j \mu_i + \Sigma_{j=2}^{J+1} w_j \varepsilon_{it}$$

If one assumes that exists weights $(w_2^*,\dots,w_{J+1}^*)$ such that the following holds, $\Sigma_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}$ $,\dots,$ $\Sigma_{j=2}^{J+1} w_j^* Y_{jTo} = Y_{11To}$, and $\Sigma_{j=2}^{J+1} w_j^* Z_j = Z_1$, then Abadie et al. (2010) prove that the following equation is true

(2.5) $$Y^N_{1t} - \Sigma_{j=2}^{J+1} w_j^* Y_{jt} = \Sigma_{j=2}^{J+1} w_j \Sigma_{s=1}^{To} \lambda_t (\Sigma_{s=1}^{To} \lambda_n' \lambda_n)^{-1} \lambda_n' (\varepsilon_{js} - \varepsilon_{1s}) - \Sigma_{j=2}^{J+1} w_j^* (\varepsilon_{jt} - \varepsilon_{1t})$$

and that its right hand side will be close to zero if the number of pre-intervention periods is large relative to the scale of the transitory shocks. This implies that $Y^N_{1t} = \Sigma_{j=2}^{J+1} w_j^* Y_{jt}$, which suggests the following estimator for the α vector:

(2.6) $$\hat{\alpha}_{1t} = Y_{1t} - \Sigma_{j=2}^{J+1} w_j^* Y_{jt}$$

To obtain the vector of optimal weights W, let $X_1 = (Z_1', Y_{11},\dots,Y_{1To})'$ be a vector of pre-intervention characteristics for the treated region and X_0 be a matrix that contains the same variables for the untreated regions, such that the jth column of X_0 is $(Z_j', Y_{ij},\dots, Y_{jTo})'$. Then, W^* is chosen to minimize the distance, $||X_1 - X_0W||_V = [(X_1 - X_0W)'V(X_1 - X_0W)]^{½}$, between X_1 and X_0W subject to $w_j \geq 0$ for $j = 2,\dots,J + 1$ and $w2 + \dots + w_{J+1} = 1$, where V is a symmetric and positive semidefinite matrix chosen in a way that the resulting synthetic control region approximates the trajectory of the outcome variable of the affected region in the pre-intervention periods.

The model described above has several advantages when compared to other approaches used in the literature. As pointed out by Nannicini and Ricciuti (2010), the model is transparent, given the weights $(w_2^*,\dots,w_{J+1}^*)$ identify the regions that are used to construct counterfactuals for the treated region, and the model is flexible, as the set of potential control regions can be appropriately restricted to make the comparisons sensible. Also, the model relaxes the assumption that confounding factors are time invariant (fixed effects) or share a common trend (differences-in-differences), given the effect of unobservable confounding factors is allowed to vary with time.

On the other hand, this approach has the limitation that it does not allow one to assess the significance of the results using standard inferential techniques, given the number of untreated regions and the number of periods considered are small. Abadie et al. (2010) suggest that inference should be carried out by implementing placebo experiments. In this case, inference is based on comparisons between the magnitude of the gaps generated by the placebo studies and the magnitude of the gap generated for the treated state. Thus, if the gap estimated for the treated state is large compared to the gap estimated for the placebo experiments, then the analysis would suggest that the treatment had an effect on the outcome of interest and is not driven by chance.

3. Data

We use annual state-level panel data for the period of 1995 to 2006. The law banning cell phone use while driving became effective in 2001 for the state of New York, which gives us several years before and after the imposition of the ban. I discard from the data the states of California, Connecticut, Delaware, Maryland, New Jersey, Oregon and Washington given all imposed a ban on cell phone use during the period I consider.[14] I also discard a few other states given data availability (Colorado, Montana, North Dakota, South Dakota, Rhode Island, Vermont and Wyoming). Hence, our final set of potential control regions is composed of 35 states, which are presented in Table (2.2).

[14] The results are robust to the inclusions of each of these states.

Our main variable of interest is the fatality rate per 100 million vehicle miles traveled obtained from the Fatality Analysis Reporting System (FARS) of the National Highway Traffic Safety Administration (NHTSA). Figure (2.1) shows the evolution of fatality rate for the state of New York (treated unit) and for (the average of) all other states considered in the analysis (untreated units). One can observe that the state of New York has a significantly lower fatality rate when compared to all other states in the sample. Also, the fatality rate is decreasing over time for all states in the sample considered.

With respect to other variables included in the vector of pre-intervention characteristics X_1 and X_0, I opted to include per capita Gross Domestic Product (GDP), population density (number of states residents per square mile), % of high school graduates and per capita annual vehicle-miles traveled (in millions). These variables should predict well the fatality rate in pre-intervention periods. I expect that states with high GDP per capita have a lower fatality rate when compared to states with lower GDP per capita. The same is expected for states with higher population densities and higher % of high school graduates. On the other hand, states with higher per capita annual vehicle-miles traveled are expected to have higher fatality rates. A regression of fatality rates before the intervention on these control variables show that they predict most of the variation in the dependent variable (an R^2 of .76). Hence, I are confident about their inclusion in the analysis.[15]

Summary statistics for all variables used in the analysis are presented in Table (2.1) for the state of New York and for all other states. As expected given Figure (2.1) presented above, the fatality rate before the imposition of the ban is significantly lower in the state of NY compared to that of all other states. Also, the state of New York has a much larger per capita

[15] In order to assess the robustness of the results, additional predictors are included among the variables used to construct the synthetic control, such as age and racial distributions, and unemployment rates. Results are consistent with the ones presented in section 4 of the current paper.

GDP and population density when compared to other states. On the other hand, it has a slightly smaller percentage of high school graduates and a significantly smaller per capita annual vehicle-miles traveled.

4. Results

Figure (2.1), presented in the previous section, suggests that the use of all other states in the sample as a counterfactual for the state of New York might not be a good strategy. This is due to the fact that not only the level is significantly different between NY and other states but also it appears that the state of New York has a slightly higher decrease in fatality rates when compared to other states in the sample. Thus, the use of a synthetic state (a "peer" state) seems a reasonable solution to estimate how would have evolved the fatality rate if there was no ban on cell phone use in the state of New York.

Before looking at the effects of the ban, let us first look at the states that compose the "peer" state (i.e. the synthetic New York) and how their pre-treatment characteristics compare to the pre-treatment characteristics of the real New York and of all other states. Table (2.2) presents the estimated weights for each state in the set of potential control states. The states of Illinois, Massachusetts, and Texas have positive weights while all other states have zero weight.[16] Hence, not only the synthetic control method selects the optimal peer states based on observed characteristics and (differently from what Hendren and Niemeier propose) the variable one in interested in analyzing, but it also assigns weights significantly different from the ones proposed by previous studies (which considers equal weights between peer states).

With respect to how different are pre-treatment characteristics between real and synthetic New York, Table (2.1) shows that the synthetic NY provides a much better control group for the

[16] The number of states receiving positive weights resembles that of Belot and Vandenberghe (2011) and is grater than that obtained by Abadie and Gardeazabal (2003), which had only two states with positive optimal weights.

real NY when compared to the average characteristics of all other states. For example, per capita GPD, population density and % of high school graduates are all well approximated by the synthetic group. Only per capita annual vehicle-miles traveled is not well approximated but still much better estimated when compared to the average of all other states. Hence, the synthetic NY seems to provide a better control group than only comparing real NY with all other states or with a single state, which is what has been before in the literature (if a single state provided the best counterfactual for the state of NY, then it would show up as weight equal to 1 in the weight vector).

Figure (2.2) shows the evolution of the real NY (treated unit) and the synthetic NY (synthetic control unit) for the whole period considered. Note that the synthetic NY follows quite well the level and tendency of the real NY before the treatment (which is represented by the vertical dashed line), which suggests that it might predict well the fatality rate for the state of NY without the existence of the ban. Now looking at the right side of the dashed line, one can observe that the fatality rate for the synthetic control is considerably higher when compared to the real NY fatality rate. This can be seen more clearly in Figure (2.3), which plots the gap between the real and synthetic NY (solid line). It appears that the ban had a significant effect on reducing fatality rate in the state of NY (our calculations imply a decrease of about 9% in the fatality rate per 100 million vehicle-miles traveled). Also, note that contrary to the results presented by McCartt and Geary (2004), the impact of the ban estimated here do not follow the same pattern observed in their data, since clearly the effects obtained here do not dissipated during the subsequent years. These results contribute to the current debate about whether states and other unregulated countries should adopt such a ban on cell phone use, specially if one considers that cell phone subscribers have grown only in the US from 97 million in June 2000 to more than 267 million at the end of the decade.

As pointed out by Abadie et al. (2010) and by Abadie and Gardeazabal (2003) one must “Evaluate the significance” of the estimates using the synthetic control method, given “Results could be driven entirely by chance.” They propose the use of a placebo test, in which the Synthetic Control Method would be applied to all other states that did not impose any ban during the period analyzed, i.e., those 35 states that were initially potential candidates for our NY synthetic control. In this case, inference is basically based on comparisons between the magnitude of the gaps generated by the placebo studies and the magnitude of the gap generated for the state of NY. In other words, if the gap estimated for the state of New York show up to be unusually large compared to the gap estimated for all the other states (the placebo test), then the analysis would suggest that the ban really had an impact on the fatality rate.

Figure (2.3) presents the gaps on fatality rates for the placebo tests. Each grey line represent a placebo state while the solid line represents the state of NY. There are 30 placebo tests given I discarded states that had pre-intervention mean squared prediction error (MSPE) twenty times higher than NY's, as so did Abadie et al. (2010). As it is clear from the figure, the gap estimated for the state of NY after the ban was imposed works almost as a lower bound for all other states (placebo tests). This seems to provide strong evidence that the reduction on fatality rates for the state of NY was due to the ban and not due to some other random event not captured in our analysis.

5. Implications and Concluding Remarks

There has been significant amount of research in the area of transportation on the development of strategies that allow good comparisons between states, in what Hendren and Niemeier call the construction of “peer states”. Such groups of states should be similar enough such that system evaluations and policy analysis are allowed to be carried out and informative predictions with important policy-oriented questions are allowed to be answered.

In this paper I contribute to the current debate on identifying peer states for policy evaluations in transportation research and also contribute to the debate on how effective are hand-held cell phone bans on reducing fatality rates. With respect to the problem of identifying peer states, I propose the use of a recently developed method by Abadie et al. (2010) called Synthetic Control Method. Contrary to what Hendren and Niemeier propose, which is basically clustering states based on observable characteristics without actually knowing what policy will be evaluated, the Synthetic Control Method uses data-driven procedures to construct an adequate comparison group/counterfactual, which might be combination of states, that offers a better comparison than any single state alone. The selection of the states that should belong to the synthetic control (the “peer” state) is done based, in part, on the variable one is interested in analyzing. Hence, a main difference between the approach proposed here and the approach taken by Hendren and Niemeier is that the former control group is policy-specific while the latter is not. For example, in their paper the states of NY, NC, and WY are each a solo member of their peer group, that is, no other state has similar characteristics according to their clustering process. In the Synthetic Control Method, a good (synthetic) “peer state” could be composed of a convex combination of other states such that in the periods before any intervention the behavior of the synthetic control is as similar as possible to the behavior of the treated state.

Finally, I contribute to the debate on how effective are cell phone bans by applying the synthetic control method to study the effects of New York state's law on fatality rates per 100 million vehicle miles traveled. I obtain that a convex combination of the states of IL, MA and

TX provide a (synthetic) “peer state” for the state of NY when evaluating this specific policy. With respect to our results, I show that, after the ban was imposed in the state of NY, the fatality rate for the synthetic NY is considerably higher when compared to the real NY fatality rate, which implies that the ban did have a negative effect of about 9% on the fatality rate.

6. Figures and Tables

Figure 2.1: Trends in Fatality per 100 Million Vehicle Miles Traveled:
NY vs. Other Untreated States.

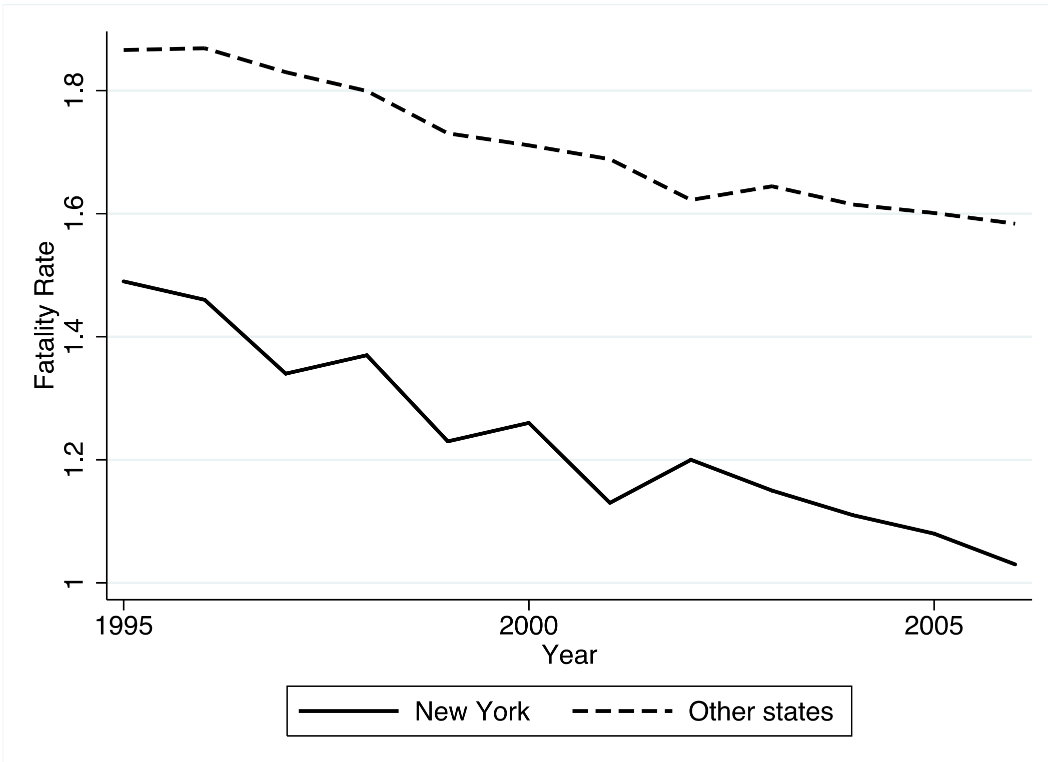


Figure 2.2: Trends in Fatality per 100 Million Vehicle Miles Traveled:
NY vs. Synthetic NY.

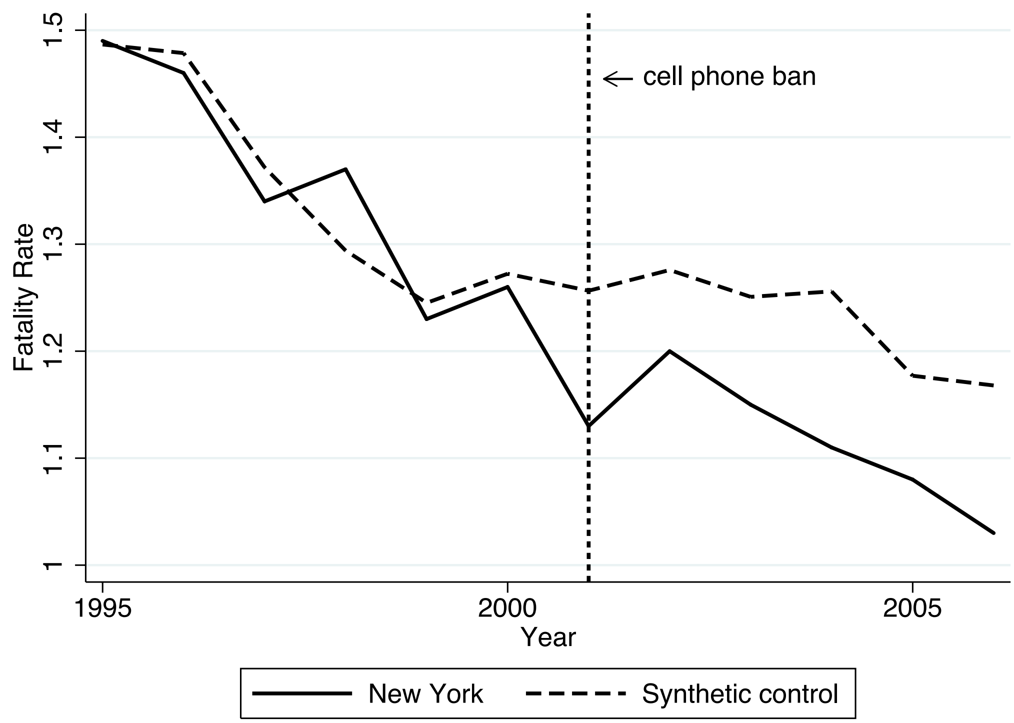
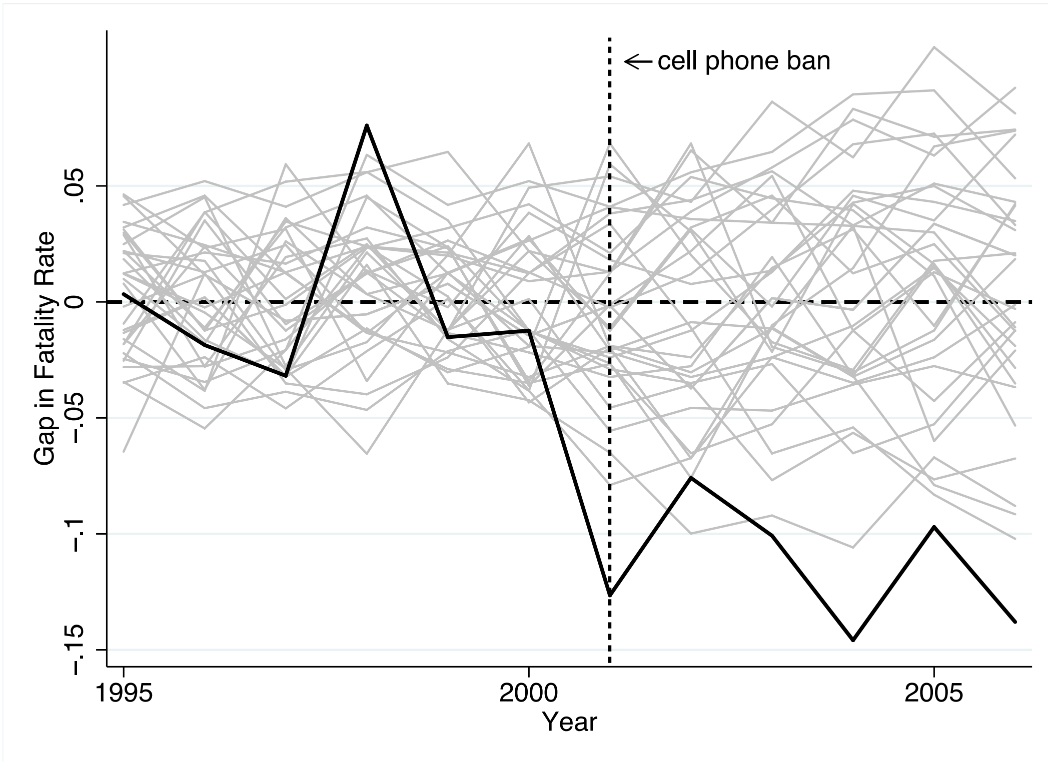


Figure 2.3: Fatality per 100 Million Vehicle Miles Traveled:
Gap between NY and Synthetic NY and Placebo Gaps for 30 Control States.



Note: Discards States with Pre-Treatment Mean Square Prediction Error (MSPE) Twenty Times Higher than NY's.

Table 2.1: Summary Statistics for NY and other States Included in the Analysis

Variables		New York		Average of all other states
Fatality Rate	Before Ban	1.358		1.801
	After Ban	1.134		1.634
		Real	Synthetic	
Per Capita GDP		31,205.00	29,823.68	24,416.04
Population Density		401.90	398.54	126.58
% of High School Graduates		79.1	82.2	81.21
Per Capita Annual Vehicle-Miles Traveled (in millions)		6,801.00	8,466.65	10,500.46

Note: Fatality rate is the number of fatal automobile accidents per 100 million vehicle miles traveled obtained from the Fatality Analysis Reporting System (FARS) of the National Highway Traffic Safety Administration (NHTSA). The law banning cell phone use while driving became effective for the state of New York in 2001.

Table 2.2: Synthetic NY: State Weights

State	Weight	State	Weight
AK	0	MO	0
AZ	0	NE	0
AR	0	NV	0
FL	0	NH	0
GA	0	NM	0
HI	0	AL	0
ID	0	NC	0
IL	.628	OH	0
IN	0	OK	0
IA	0	PA	0
KS	0	SC	0
KY	0	TN	0
LA	0	TX	.06
ME	0	UT	0
MA	.313	VA	0
MI	0	WV	0
MN	0	WI	0
MS	0		0

CHAPTER 3

FOR LOVE OF COUNTRY: NATIONALISTIC BIAS IN PROFESSIONAL SURFING

1. Introduction

The past two decades has seen an increasing interest in detecting and quantifying hidden actions taken by agents when facing decisions that may lead to higher individual payoffs but are not easily observed by all parties involved. For example, researches have shown that scholars engage in a "paper inflation" behavior by downloading their papers multiple times in order to increase download statistics and the visibility of the paper (Edelman and Larkin, 2009) and that real estate agents, who are often better informed than their clients, tend to sell their clients houses faster and for lower prices when compared to their own houses (Levitt and Syverson, 2008). According to Zitzewitz (2011), this new field of research is called "forensic economics" and it has provided evidence on hidden behavior in a variety of domains, such as finance, health, education, among many others.

One such area that has recently received a lot of attention is on understanding the decision making process of professional referees in sports. Whilst one may expect referees to make mistakes when judging, it is important to detect if such "mistakes" follow any systematic pattern that might benefit a specific set of athletes or teams (Page and Page, 2010). This has been the main focus of many empirical studies in the past decade and results strongly support the idea that judges scores are influenced by the nationality of the athlete they are evaluating. For example, Setzer and Glass (1991), Basset and Persky (1995), Campbell and Galbraith (1996), and Zitzewitz (2006) provide evidence that in figure skating scores given by judges that share the same nationality as the athlete being evaluated are higher when compared to the score of all other remaining judges in the judging panel. Similar results were found when looking at Gymnastics

(Ste-Marie, 1996), Ski jumping (Zitzewitz, 2006), Diving (Emerson et al., 2009, and Emerson and Meredith, 2011), Muaythai (Myers et al., 2006), Soccer (Garicano et al., 2005), and Cricket (Crowe and Middeldorp, 1996), to name a few.

All these evidences support the conclusion that judges give higher scores to their compatriots. However, as pointed out by Emerson et al. (2009), Emerson and Meredith (2011) and Zitzewitz (2006), these evidences not necessarily imply that judges are behaving strategically, that is, the score difference might be totally driven by preferences over the way athletes perform in a specific country. Hence, the main challenge faced by researchers in this literature is to identify whether judges scores are lower due to favoritism (nationalistic bias) or due to preferences over a country-specific surfing style (taste).

In this paper, I look at the world's elite division of professional surfing, which, different from other sports previously analyzed, is composed of one-on-one heats with surfers having as many as 15 performances scored by the same judges in each heat. Thus, I am able to observe scores for many waves given by the same judging panel for two athletes competing against each other in a heat in which one of them will be taken out of the competition (and hence, will acquired less points in the tour, which are necessary to guarantee a spot in next year's competition, and receive a lower monetary payoff). This is different from figure skating, for example, in which an athlete competes against all other competitors and the number of performances is very limited in each competition. My setup, thus, estimates the bias based on many observations for the same pair of surfers and holding fixed the same judging panel. In addition to that, I also use an important information in the data that reports the exact time in each heat that each wave was caught by the surfers, allowing me to calculate for each point in time which surfer was winning the heat and by how many points was the winning advantage. Thus, I can identify whether or not judges that share the same nationality as one of the surfers in the heat behave differently when they are winning or when they are losing. If results turn out to show that judges behave differently depending on the position their countrymen is in, then one may easily argue that judges are behaving strategically and that the bias is not a result of taste/preferences.

Different from what has been obtained in previous research, i.e., that judges overscore their countrymen, surfing judges do not overscore/underscore their compatriot, but significantly underscore athletes competing against their home athlete. Hence, it is less clear why one should believe it to be completely driven by preferences. For that to be the case, one would have to argue that all judges selected to participate in a heat in which there is a surfer from their same nationality are extremely pessimistic in a way that their scores are ex-ante statistically smaller when compared to the rest of the judging panel. In this case, one may claim that the net benefit for the surfer who's nationality matches one of the judges might be a result of preferences over surfing style. However, this selection is not very likely to be true, and the results provide compelling evidence on favoritism in professional surfing.

This is confirmed by the analysis that considers the time in which each wave was scored in the heat. In this case, comparing waves of the same surfer and holding the same judging panel constant, which excludes any judging-panel selection effect, it is observed that the score given by the judge scoring his countrymen is statistically the same as all other judges' scores whether the athlete is winning or losing the heat. On the other hand, when looking at the athlete competing against the surfer backed up by a judge, results show a completely different picture. When the surfer born in the same country as one of the judges is losing, the judge whose countrymen is in the heat underscore his opponents waves by about .133 (approximately .25σ). This negative effect vanishes if the surfer backed up by a judge is winning. Also, when looking specifically at waves that changes current total heat score for a surfer, it is observed that the judge whose countrymen is in the heat may underscore his opponents waves by more than .3, which represents a score disadvantage of approximately .56σ. This number is significantly higher than the results provided by previous research, which shows that judges overscore their countrymen by .1 to .2σ. Finally, these robust and large score penalties imposed on surfers competing against athletes backed up by a judge are shown to have a significant effect on final heat positions and, as a consequence, final points and prizes earned in the tournament.

The paper contributes more generally to two strands in the literature. The first is the literature on corruption and the second is the literature on the process of decision making in organizations. While acknowledging that the specific behavior of surfing judges is by itself not of great interest to the economics profession, the paper provides clear evidence on how agents behave under a framework in which there is huge incentives to report biased evaluations to promote certain individuals. Hence, in a broader sense, these evidences shed light on the role that corruption plays in real world economies, an issue that has been for many years of great interest to the profession but has only recently received significant empirical support, as discussed in Zitzewitz (2011), due to its illicit nature.

The second contribution of the paper relates to how organizations should aggregate opinions given by better informed individuals/managers regarding the performance of a specific employee, for example. The individuals who best understand or have better information about specific characteristics of a project or employee are exactly the ones more likely to report biased evaluations, if there is any specific payoff from such deviation. In fact, this has been the focus of many theoretical analysis, such as, for example, Prendergast and Topel (1996), who tackle the problem of relying on the opinion of a "potentially" biased supervisor about employee performance, and Aghion and Tirole (1997) and Athey and Roberts (2001), who analyze the problem of taking the opinion of a employee about projects quality. Thus, the analysis carried out in this paper attempts to be informative and to provide empirical support about deviations from truth-telling in organizational designs which involve more than one person in the decision making process.

The remainder of this paper is organized as follows. Section 2 presents the data used in the analysis. Section 3 describes the methodology as well as estimates of the nationalistic bias and its variants. Finally, implications of the results and some concluding remarks are presented in Section 4.

2. Data

The data used in this paper was hand-collected from the website of the Association of Surfing Professional (ASP) World Tour, which is the elite division of professional surfing. The tour is composed of approximately 40 surfers who compete for the world title in 11 events or stops throughout the year.[17] In each event, surfers gain points (depending on their final position, i.e., if the surfer finished in 1st place, they gain more points when compared to another surfer that finishes in 33rd), which are added up at the end of the season to decide whose the world surfing champion. These points also determine which surfers must leave the tour to be substituted for the best placed surfers in the World Qualifying Series (WQS), the second division of professional surfing.

The data used in the estimations performed below are from the first 5 events of the 2010 season. In each event, composed of exactly 48 surfers,[18] there are 7 rounds. The 5th, 6th and 7th

---

[17] The number of world tour events may vary between years, with a maximum of 13, as stipulated by the ASP rule book.
[18] In each event, additional surfers are selected to participated by the event sponsor (usually a surfing brand) or by ASP.

round are, respectively, the quarter-finals, semi-finals, and finals, as explained below. The number of heats in each round varies according to the following format

- Round 1: 16 heats of 3 surfers with 1^{st} advancing to Round 3 while 2^{nd} and 3^{rd} advance to a one-on-one Round 2
- Round 2: 16 heats (one-on-one) of 2 surfers with 1^{st} advancing to Round 3 while 2^{nd} is eliminated
- Round 3: 16 heats (one-on-one) of 2 surfers with 1^{st} advancing to Round 4 while 2^{nd} is eliminated
- Round 4: 8 heats (one-on-one) of 2 surfers with 1^{st} advancing to quarter-finals while 2^{nd} is eliminated
- Round 5 (quarter-finals): 4 heats (one-on-one) of 2 surfers with 1^{st} advancing to semi-finals while 2^{nd} is eliminated
- Round 6 (semi-finals): 2 heats (one-on-one) of 2 surfers with 1^{st} advancing to finals while 2^{nd} is eliminated
- Round 7 (finals): a one-on-one heat in which 1^{st} place is crowned champion

Points and prizes earned in each event depend on which round the surfer was eliminated in. Elimination in the 2^{nd} round guarantees the surfer a 33^{rd} place, while elimination in the 7^{th} round guarantees a 2^{nd} place. Table (3.1) presents final place, prizes and points earned in each elimination round for each event.[19] As one may observe, advancing a round in the tournament increases significantly the amount of points earned (besides also raising the money prize), which are extremely valuable at the end of the season due to the elimination cutoff. Total prizes usually

[19] Note that these numbers are valid only for the first 6 events of the 2010 season. From the 7^{th} event on, the best placed 32 surfers were selected to continue on the competition and the heat format changed slightly: rounds 1 through 3 are the same, round 4 is now another three-man heat in which the 1^{st} advances to Round 6 (now the quarter-finals) while the 2^{nd} and 3^{rd} advance to a one-on-one heat on round 5 to compete for a spot in the quarter-finals. Finally, rounds 7 and 8 are, respectively, the semi-finals and finals.

add up to $400,000 (in the 2010 season), however they may vary between years and events. In the 2011 season, for example, the minimum total event prize was raised to $425,000, with one event paying a total of $500,000 and another paying $1,000,000.

In each heat, athletes are allowed to surf a maximum number of 15 waves, which are each scored by five judges that compose the judging panel. The high and low judges scores are eliminated for each wave and the remaining three judges scores are averaged out to deliver a final wave score between 0.1-10. At the end of each heat, the best 2 waves caught by surfer *i* are summed up to give his final heat score which lies between 0.2-20 and the surfer with the highest total score is the winner and advances to the next round according to the format described above.

Table (3.2) summarizes the total data and sample used. The total data includes 5 events, each consisting of 7 rounds, which add up to 35 rounds. Given the number of heats in each round (presented above), the data includes 315 heats in which 4,729 waves are surfed and a total of 23,645 scores are recorded (since each wave is scored by exactly five judges). There are 66 surfers who participate in at least one event (most participate in all events, however a few are wildcards and only observed in one event) and 21 judges.

Given the main objective of the paper is to estimate nationalistic bias and given that the same judging panel score all waves of the surfers competing against each other in a heat and that the scoring potential may change significantly between heats (depending on wave quality), I restrict the sample to include only heats in which one of the surfers matches the nationality of one of the judges. Hence, comparisons are all performed within heats and not between heats. I also discard all three-man heats, given they compose the first round of each tournament which is not a single-elimination competition and bias is less likely to be captured. The final sample consists of 4,660 scores given to 932 waves surfed by 52 athletes. Note that the number of rounds in the sample is 30, since the first round of each event was discarded.

Table (3.3) presents summary statistics for judges' scores. There are no zeros given to any wave caught by a surfer (the minimum score allowed is .1) while there are 6 judges who gave a 10 for a performance. The average score is 3.9. However there is huge variation as can be observed in the standard deviation. On the other hand, score variation within performance (i.e., within the same wave) is significantly lower when compared to overall performances.

3. Estimating Nationalistic Bias

3.1 Main Challenges

The main challenge an econometrician faces when estimating nationalistic biases comes from the fact that an objective measure of performance is never observed. In surfing, for example, there are no objective measures of the quality of a wave surfed by a specific surfer in a heat. Additionally, scores are known to vary significantly with the quality of the waves, which are directly influenced by, for example, the direction of the wind and the size and direction of the swell.[20] Hence, it could easily be the case that a surfer may repeat exactly the same performance and maneuvers in two different waves, but receive very different scores depending on the current wave quality. This suggests that judges adjust their scoring behavior in each heat depending on the scoring potential given by the quality of the waves.

To estimate nationalistic bias one needs an objective measure of performance which, as argued above, is unobtainable for the case of surfing. To circumvent this problem, I follow Zitzewitz (2006) and assume that a "good" measure of the quality of a performance is given by the average score of all judges once nationalistic biases are accounted for. Thus, nationalistic

[20] According to the definition given by Wikipedia, "a swell, in the context of an ocean, sea or lake, is a formation of long-wavelength surface waves. Swells are far more stable in their direction and frequency than normal wind waves, having often traveled long distances since their formation by tropical storms or other wind systems."

bias is assumed to exist only when a judge j is present in the judging panel of a surfer i, and both share the same nationality. This requires that the average scores provided by judges whose nationality does not coincide with the nationality of the surfer being scored be equal to the "true" quality of the performance. This implicitly assumes away the existence of a judge from country c overscoring an specific surfer from country c' ($c \neq c'$) due to preferences over his surfing style.

On the other hand, nationalistic bias estimated below include also potential unconscious same country-style bias from judges scoring surfers from their same country, i.e., judges might unconsciously be overscoring their countrymen due to a preference over their country's surfing style. However, as I will argue later in the paper, the results obtained strongly support the claim that the bias reflect judges' behaving strategically and not an unconscious country-taste bias.

3.2 Measurement

To estimate nationalistic bias, define $Diff_{iwh}$ to be equal to the difference between the score given by the judge who's nationality matches that of one of the surfers and the average of the scores given by all other four judges for each wave in each heat.[21] This calculation is illustrated in Table (3.4), in which (only) one of the surfers in the heat (Surfer 1) matches the same nationality as (only) one of the judges scoring the heat (Judge 5).

For this example,

(3.1)
$$Diff_{iwh} = \begin{cases} a_5 - \frac{1}{4}\Sigma_{i=1}^{4} a_i, & \text{for Surfer 1,} \\ b_5 - \frac{1}{4}\Sigma_{i=1}^{4} b_i, & \text{for Surfer 2.} \end{cases}$$

This number is calculated for each wave of each surfer in each heat. Thus, the maximum number of score differences in each heat is 30, given each surfer may surf a maximum of 15 waves per

[21] In the appendix of the paper I provide a slightly modified strategy to estimate nationalistic bias. The results are roughly the same, however, instead of using the score difference, I use the raw score and include wave fixed effects.

heat. Now define $Compatriot_{iwh}$ to be a dummy variable that takes value equal to 1 for the surfer whose nationality coincide with a judge in the judging panel and 0 otherwise, such that

$$Diff_{iwh} = \beta_0 + \beta_1 * Compatriot_{iwh} + \varepsilon_{iwh} \tag{3.2}$$

The parameter, $\beta_0 = E[Diff \mid Compatriot = 0]$, represents the cost (or, perhaps, the benefit) of having a judge born in the same country as the opponent surfer in a heat. One may find, for example, that $\beta_0 < 0$, which suggests that a judge would bias against athletes from other nationalities, consequently benefiting their own country athletes. Another possibility is to find $\beta_0 > 0$, which would imply that judges overscore athletes from other countries when competing against surfers from their own nationality. In this case, judges might be overscoring surfers in a attempt to decrease as much as possible their chances of giving the lowest score for a given wave and the possibility of being viewed as having inappropriate behavior. It is straightforward to show that the sum $\beta_0 + \beta_1 = E[Diff \mid Compatriot = 1]$ equals the benefit (or the cost) of having a judge from the same nationality in the judging panel. Thus, β_1 represents the net benefit (or cost) of having a countrymen in the judging panel.

Before proceeding to the estimation results, Figure (3.1) shows the empirical density estimate of the score difference ($Diff_{iwh}$) for surfers that share the same nationality as one judge in the judging panel (solid black line) and their opponents (dashed grey line). This figure shows that the score difference distribution of the opponent surfer (competing against the surfer backed up by a judge) is located to the left of the distribution of score differences for the athlete supported by a judge. This implies that there is a large mass of negative scores (compared to the average) given to opponents when compared to athletes that share the same nationality as one of the judges. The p-value for the two-sample Kolmogorov-Smirnov test equals .027, which implies that the distributions do not come from populations with same distribution.

Table (3.5) presents estimates of the bias calculated via equation (3.2). As one may observe, $\beta_0 + \beta_1 = 0$, which suggests that there are no differences between the score given by a judge scoring a surfer from his same nationality and the scores of all other judges in the judging panel. Thus, having a compatriot in the judging panel leads to no benefit (in terms of the scores given for his waves) for the surfer. On the other hand, there is a significant score penalty imposed on a athlete when competing against a surfer whose nationality coincide with that of a judge ($\beta_0 < 0$). Hence, given $\beta_1 > 0$, having a judge from the same country in the judging panel does lead to benefits for the surfer, however, not because judges scores are higher for surfers from their same country but because judges are underscoring surfers competing against their countrymen by approximately .08.

This result is completely different from what has been previously reported in the literature. For example, Seltzer and Glass (1991), Basset and Persky (1995), Campbell and Galbraith (1996), and Zitzewitz (2006) provide evidence that in figure skating scores given by judges that share the same nationality as the athlete being evaluated are higher when compared to the score of all other remaining judges in the judging panel. Similar results were found when looking at many other sports, such as Gymnastics (Ste-Marie, 1996), Ski jumping (Zitzewitz, 2006), Diving (Emerson et al., 2009, and Emerson and Meredith, 2011), Muaythai (Myers et al., 2006), Soccer (Garicano et al., 2005), and Cricket (Crowe and Middeldorp, 1996), to name a few.

Obtaining that judges give higher scores to their compatriots, as reported in all previous research, does not necessarily imply that judges are behaving strategically, that is, the score difference might be totally driven by preferences over the way athletes perform in a specific country. Hence, the researcher is unable to identify whether judges scores are lower due to favoritism (nationalistic bias) or due to preferences over a country-specific surfing style (taste). For the specific case of surfing, in which judges underscore competitors surfing against their countrymen, it is less clear why one should believe it to be completely driven by preferences. For that to be the case, one would have to argue that all judges selected to participate in a heat in which there is a surfer from their same nationality are extremely pessimistic, in a way that their scores are ex-ante statistically smaller when compared to the rest of the judging panel. In this case, one may claim that the net benefit for the surfer who's nationality matches one of the judges might be a result of preferences over surfing style. However, given this selection is not very likely to be true, I believe results do provide evidence on favoritism in professional surfing.

I propose, however, a second approach that is arguably more convincing than the one given above to insure that the results do follow from judges' favoritism and are not driven by preferences. For that, I use an additional feature in the data that contains the exact time in each heat that each wave was caught by the surfers. Hence, I am able to calculate for each point in time which surfer was winning the heat and by how many points was the winning advantage. With that information in hand, I can identify whether or not judges that share the same nationality as one of the surfers in the heat behave differently when they are winning or when they are losing. Note that this strategy requires all judges to have full information about scores in all points in time, which is the case in the elite surfing division studied here given judges are constantly informed about surfing scores and the score advantage of the surfer winning the heat.

To measure such effects, let $Compatriot\_Winning_{iwh}$ be a dummy variable that equals 1 when the surfer whose nationality is the same as one of the judges' is winning and 0 otherwise, and ε_{iwh} be an error term, such that

(3.3) $$Diff_{iwh} = \theta_0 + \theta_1 * Compatriot\_Winning_{iwh} + \varepsilon_{iwh}$$

The parameters of equation (3.3), θ_0 and θ_1, are estimated in two different ways to capture differences in judges behavior when scoring the surfers: (a) firstly using the sample of waves for the surfer whose nationality coincide with a judge in the judging panel; and secondly (b) using the sample of waves for the surfer competing against the athlete that is backed up by a judge. Using the former sample, the parameter of interest $\theta_0 = E[Diff_{iwh} \mid Compatriot\_Winning_{iwh} = 0]$ would capture any behavioral difference in judges scores when his countrymen is losing the heat, while the sum of the parameters $\theta_0 + \theta_1 = E[Diff_{iwh} \mid Compatriot\_Winning_{iwh} = 1]$ would capture judges behavior when the surfer is winning the heat. Similarly, using the latter sample, θ_0 and $\theta_0 + \theta_1$ would, respectively, represent judges behavior with respect to the surfer competing against his countrymen when his countrymen is losing or winning the heat.

Estimating nationalistic bias via equation (3.3) has several advantages when compared to estimates obtained by equation (3.2). Given the comparison is made within the waves of the same surfer and holding the same judging panel constant, which excludes any judging-panel selection effect, difference in behavior is likely due to favoritism and not country-style preference, specially if results turn out to show that judges bias more when their preferred surfer is losing. Differently from what is done in equation (3.2), this approach uses data at the surfer level instead of using data at the heat level. Hence, it is the most data-intensive approach of the two.

Table (3.6) presents estimates of the nationalistic bias using data at the surfer level. Columns 1 and 2 use the sample of waves for the surfer whose nationality coincide with a judge in the judging panel, while columns 3 and 4 use the sample of waves for the surfer competing against the athlete that is backed up by a judge. In column 1 and 2, all coefficients are not statistically different from zero, i.e., the score given by the judge scoring his countrymen is statistically the same as all other judges' scores and does not vary if the athlete is winning or losing the heat. On the other hand, when looking at the athlete competing against the surfer backed up by a judge, results show a completely different picture. When the surfer born in the same country as one of the judges is losing, the judge whose countrymen is in the heat underscore his opponents waves by about .120-.133 (about .23-.25σ). This negative effect almost vanishes if the surfer backed up by a judge is winning, i.e., the effect is around -.04 but not statistically different from zero.

Until now I have investigated if judges behave differently when scoring a surfer whose nationality matches the nationality of a judge when he is losing or winning the heat. However, given I have detailed data on judges scores and the exact time that each wave was caught, I can calculate how the score difference ($Diff_{wh}$) vary when the average score of all other judges is above or below the necessary to change the sum of the best two waves of a surfer (that is, to make the wave count for the total score). In other words, if judges do behave strategically and given the results presented so far, one must expect to find that judges bias significantly more when scoring a wave that might change total scores in the heat when compared to those waves that surely will not change anything.

Before proceeding to the parametric estimations, Figure (3.2) presents the relationship between the score difference ($Diff_{wh}$) for the biased judge and the average score of the other four judges in the judging panel (whose nationality does not match any of the surfers) for the surfer competing against the athlete backed up by a judge. Results are quite impressive and strongly suggestive of strategic behavior by judges in favor of their countrymen: when the athlete whose nationality matches one of the judges is losing (panel A), the negative effect incurred by the opponent surfer increases monotonically with the average score, i.e., for bad waves, the biased judge behaves similarly to all other judges, however, for waves with higher scoring potential that might increase the lead of the opponent surfer, the biased judge underscores significantly compared to the average of all other judges. On the other hand, when the athlete whose nationality matches one of the judges is winning (panel B), the biased judges behaves exactly the same as all other judges up to a point when the wave score might actually affect the final outcome of the heat. In this case, judges underscore the opponent surfer significantly.

There are, however, two factors not accounted for in the analysis of Figure (3.2) that might change the conclusions derived from it. First, there are no heat fixed effects such that differences between heats can be accounted for. Second, and most importantly, the graphs partially reflect judges behavior when scoring "important" waves (the ones that might affect final heat positions) and when scoring waves that will surely not affect final scores. Hence, besides informative, Figure 2 does not correctly inform how judges behave when the probability of a wave affecting final scores changes from zero to a positive value.

To investigate such effects, I estimate the following model

$$(3.4) \qquad Diff_{wh} = \delta_0 + \delta_1 * ReqScore_{wh} + \delta_2 * Compatriot\_Winning_{wh} + \delta_3 * ReqScore_{wh} * Compatriot\_Winning_{wh} + \xi_{wh}$$

where $Diff_{wh}$ and $Compatriot\_Winning_{wh}$ are defined above, and ξ_{wh} is the error term. $ReqScore_{wh}$ is defined as the difference between the average of scores given by all four judges (excluding the one that share the same nationality as one of the surfers) and the minimum score composing the sum of the best two waves so far in the heat. Hence, $ReqScore_{wh}$ is positive when the score given to his current wave is larger than the second best score obtained before the present wave (which means that the total heat score increased) and negative when the score given to his current wave does not affect current total scores. As a robustness check I consider two alternative specifications for $ReqScore_{wh}$. The first considers a dummy variable that equals 1 for waves that change the sum of the best two waves in the heat and 0 otherwise, i.e., $\boldsymbol{1}(ReqScore_{wh} \geq 0)$, and the second allows the effect to differ non-linearly when that size of $ReqScore_{wh}$ increases.

Results for the three specifications are presented in Table (3.7). For expositional purposes, results are better seen in Figure (3.3). The figure on the top represents the results of specification 1 in Table (3.7), while the figures on the middle and on the bottom represents specifications 2 and 3, respectively. As can be observed if Figure (3.1), when the athlete whose nationality matches one of the judges is losing (panel A), the negative effect incurred by the opponent surfer increases monotonically with the required score, i.e., as the wave score increases, the score difference between the biased judge and all other judges increases. As the wave score reaches the required score (i.e., the wave has a positive probability of increasing total heat score or $ReqScore_{wh} > 0$), the score difference is negative and statistically significant. On the other hand, when the athlete whose nationality matches one of the judges is winning (panel B), the biased judges behaves exactly the same as all other judges. The figures on the middle and on the bottom provide similar conclusions, however, it presents in a clearer way that scores that are above the required to change heat scores are significantly underscored by the biased judge when his countrymen is losing the heat.

Until now we have provided compelling evidence that judges that share the same nationality as one of the surfers in the heat bias in favor of this surfer by underscoring the waves of the opponent athlete. According to Figure 3, the judge whose countrymen is in the heat may underscore his opponents waves by more than .3, which represents a score disadvantage of $.56\sigma$. This number is significantly higher than the results provided by previous research, which shows that judges overscore their countrymen by .1 to .2σ. It remains, however, to show if this score difference affects final heat positions, which is what is discussed in the next section.

3.3 Do the Score Difference Matter?

In this section I analyze if having a supporting judge in the judging panel affects final heat positions. Although the score disadvantage is high, as discussed in the previous section, it is important to assess if such difference do change the relative position of athletes in each heat. To analyze such effect, I estimate the following linear probability model

(3.5) $$Rank_{ih} = \rho_0 + \rho_1 * Compatriot_{ih} + \lambda_i + \varepsilon_{ih}$$

where $Rank_{ih}$ is a dummy variable that equals 1 when the surfer wins the heat and 0 otherwise, $Compatriot_{ih}$ is a dummy variable that takes value equal to 1 for the surfer whose nationality coincide with a judge in the judging panel and 0 otherwise, and λ_i is a surfer fixed effect. The equation is estimated using a sample that includes all heats in which exactly one of the surfers matches the nationality of one of the judges. The parameter of interest is given by ρ_1, which represents the increment in the probability of winning a heat resulting from having a countrymen in the judging panel. Thus, if the bias estimated in the previous section was not sufficiently large to alter the probability of winning the heat, then one should observe a value of ρ_1 that is not statistically different from zero. If this is the case, then one must observe also that ρ_0 is not statistically different from .5, which would be the probability of winning a heat once surfer differences are accounted for.

Results are presented in Table (3.8). As expected given previous estimates, the parameter ρ_1 equals .222 and is statistically different from zero at the 5% significance level, implying that there is a significant increase in the probability of being first place in a heat that contains a friendly judge when compared to the situation in which the opponent surfer has a biased judge backing him up. Thus, although scores are aggregated in a way that the highest and lowest judges' scores are dropped, it seems that having a judge backing a surfer up does alter with positive probability the final positions in a heat.

4. Concluding Remarks

The past two decades has seen an increasing interest in detecting and quantifying hidden actions taken by agents when facing decisions that may lead to higher individual payoffs but are not easily observed by all parties involved. One branch of this literature that has recently received a lot of attention is on understanding the decision making process of professional referees in sports, i.e., if such referees make "non-random" mistakes that ultimately benefit a specific athlete or team. The evidences provided so far support the conclusion that judges give higher scores to their compatriots. However, as largely documented in the literature, these evidences not necessarily imply that judges are behaving strategically, i.e., judges behavior might be totally driven by preferences over the way athletes perform in a specific country and not by strategically misreporting scores to benefit a fellow countrymen. Thus, the evidences are weak in terms of correctly identifying if lower scores are due to favoritism (nationalistic bias) or due to preferences over a country-specific performing style (taste).

In this paper, I look at the world's elite division of professional surfing, which, different from other sports previously analyzed, is composed of one-on-one heats with surfers having as many as 15 performances scored by the same judging panel in each heat. Thus, I am able to observe scores for many waves given by the same set of judges for two athletes competing against each other in a heat in which one of them will be taken out of the competition (and hence, will acquired less points in the tour and receive a lower monetary payoff). This is different from sports previously analyzed, given an individual competes against all other athletes and the number of performances is very limited in each competition. My setup, thus, estimates the bias based on many observations for the same pair of surfers and holding fixed the same judging panel. In addition to that, I also use an important information in the data that reports the exact time in each heat that each wave was caught by the surfers, allowing me to calculate for each point in time which surfer was winning the heat and by how many points was the winning advantage. Thus, I can identify whether or not judges that share the same nationality as one of the surfers in the heat behave differently when they are winning or when they are losing.

Different from what has been obtained in previous research, my results show that surfing judges neither underscore nor overscore their fellow compatriots, but significantly underscore athletes competing against their home athlete. Hence, it is less clear why one should believe it to be completely driven by preferences. This is also confirmed by the analysis that considers the time in which each wave was scored in the heat. In this case, comparing waves of the same surfer and holding the same judging panel constant, which excludes any judging-panel selection effect, it is observed that the score given by the judge scoring his countrymen is statistically the same as all other judges' scores whether the athlete is winning or losing the heat. However, when looking at the athlete competing against the surfer backed up by a judge, results show a completely different picture, i.e., the judge whose countrymen is in the heat underscores the opponents' surfer waves by about .133 (approximately .25σ). This negative effect vanishes if the surfer backed up by a judge is winning. Also, looking specifically at waves that changes current total heat score for a surfer, results show that the judge whose countrymen is in the heat underscores his opponents waves by more than .3, which represents a score disadvantage of approximately .56σ. This number is significantly higher than the results provided by previous research, which shows that judges overscore their countrymen by .1 to .2σ. Finally, these robust and large score penalties imposed on surfers competing against athletes backed up by a judge are shown to have a significant effect on final heat positions and, as a consequence, final points and prizes earned in the tournament.

5. Figures and Tables

Figure 3.1: Empirical Kernel Density Estimates for the Bias: Comparing Surfers from Same Country as Judges and Their Opponents. Distributions do not come from populations with same distributions (p-value for two-sample Kolmogorov-Smirnov test equals .027).

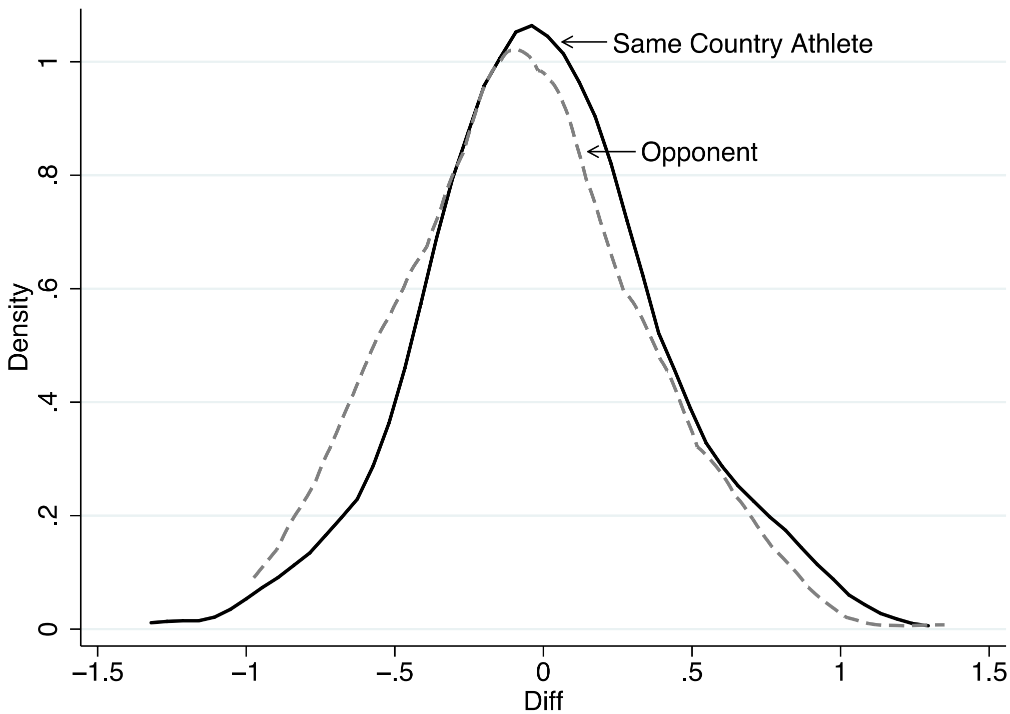


Figure 3.2: Judges Bias and Average Wave Score for the Surfer Competing Against the Athlete Backed Up by a Judge.

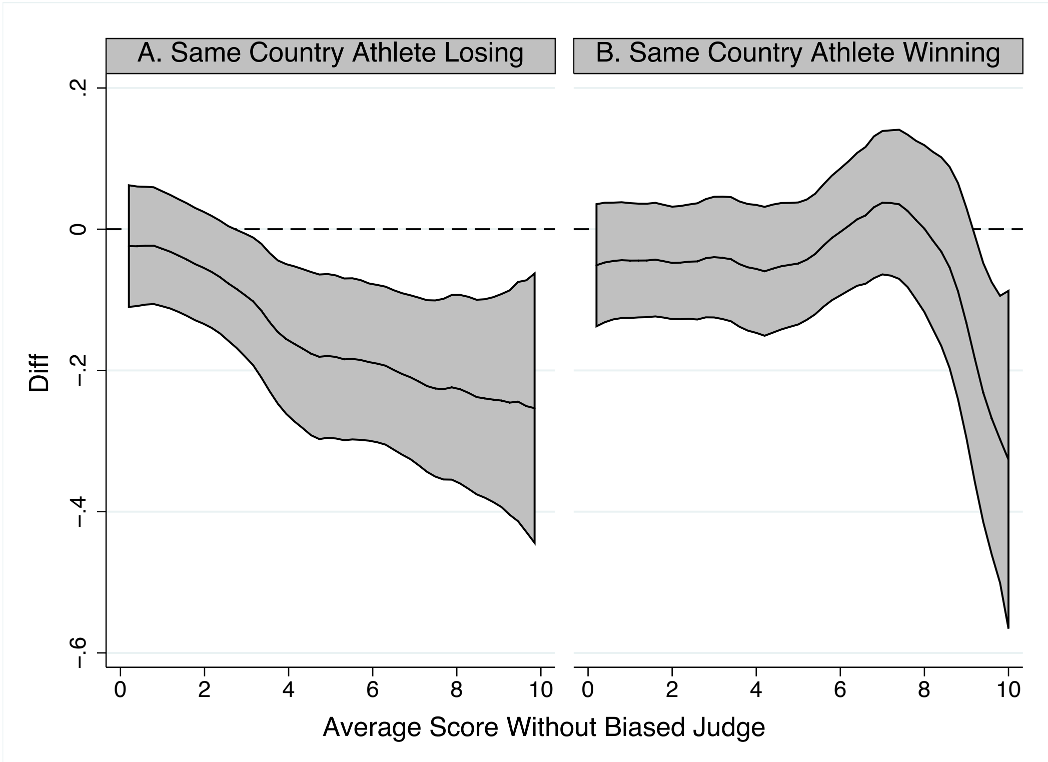


Figure 3.3: Judges Bias and Required Score for the Surfer Competing Against the Athlete Backed Up by a Judge.

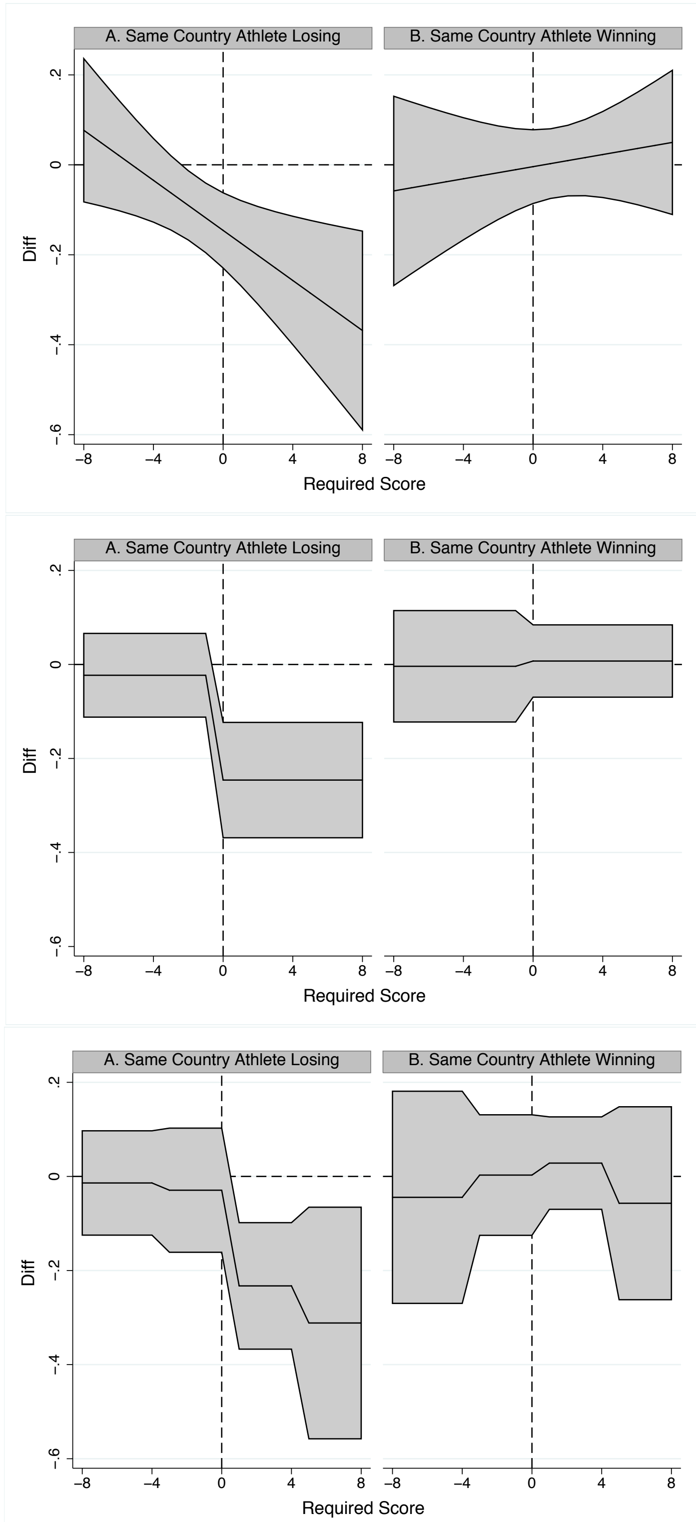


Table 3.1: Men's World Tour: Round of Elimination, Final Place, Prize and Points Earned

Round Eliminated	Place	Prize	Points
	1^{st}	\$50,000	10,000
7^{th}	2^{nd}	\$25,000	8,000
6^{th}	3^{rd}	\$14,500	6,500
5^{th}	5^{th}	\$10,000	5,250
4^{th}	9^{th}	\$8,000	3,750
3^{rd}	17^{th}	\$6,500	1,750
2^{nd}	33^{rd}	\$5,500	500

Table 3.2: Total Data and Sample Size

Variables	Total Data	Sample
Events	5	5
Rounds	35	30
Heats	315	93
Waves	4,729	932
Scores	23,645	4,660
Surfers	66	52
Judges	21	18

Table 3.3: Summary Statistics

Average Score	3.931
Minimum Score	.1
Maximum Score	10
Standard Deviation	
- Overall	2.614
- Within Waves	.331

Table 3.4: Scores by Judge in each Heat for one of the Waves of each Surfer

	Surfer 1 (Nationality = X)	Surfer 2 (Nationality = Y)
Judge 1 (Nationality = T)	a_1	b_1
Judge 2 (Nationality = U)	a_2	b_2
Judge 3 (Nationality = V)	a_3	b_3
Judge 4 (Nationality = Z)	a_4	b_4
Judge 5 (Nationality = X)	a_5	b_5

Table 3.5: Nationalistic Bias

Variables	(1)	(2)
$Compatriot_{iwh}$	.078***	.086***
	(.029)	(.029)
Intercept	-.073***	-.082***
	(.021)	(.027)
Heat Fixed Effects	NO	YES
N	743	743

Note: Estimation is performed via equation (3.2). The bias is estimated by regressing the score difference between the "possible" biased judge (who share the same nationality as one of the surfers) and all other judges on a dummy variable that takes value equal to 1 when the nationality of the surfer matches the nationality of the judge and 0 otherwise. Heat fixed effects are included in the estimation to capture any difference between heats. Robust standard errors are presented in parentheses. *** represents p<1%.

Table 3.6: Nationalistic Bias: Do Judges Behave Differ when their Countrymen is Winning or Losing?

Variables	$Compatriot_{wh} = 1$		$Compatriot_{wh} = 0$	
$Compatiot\_Winning_{wh}$	.027	.028	.083**	.091**
	(.040)	(.042)	(.041)	(.043)
Intercept	-.007	-.011	-.120***	-.133***
	(.028)	(.032)	(.031)	(.035)
Heat Fixed Effects	NO	YES	NO	YES
N	374	374	369	369

Note: Coefficients presented in columns 1 and 2 are estimated via equation (3.3) using the sample of waves for the surfer whose nationality coincide with a judge in the judging panel and columns 3 and 4 are estimated using the sample of waves for the surfer competing against the surfer whose nationality coincide with a judge in the judging panel. Heat fixed effects are included in order to compare waves within heats. Robust standard errors are presented in parentheses. *** represents p<1%, ** represents p<5%.

Table 3.7: Nationalistic Bias: Do Judges Behave Differ when Scores are Above or Below the Necessary to Change the Sum of the Best Two Waves?

Variables	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
$ReqScore_{wh}$	-.025***	-.028***				
	(.010)	(.011)				
$1(ReqScore_{wh} \geq 0)$			-.212***	-.223***		
			(.072)	(.075)		
$Compatriot\_Winning_{wh}$	.140***	.142***	.040	.019	.036	-.030
	(.052)	(.050)	(.074)	(.068)	(.131)	(.129)
$ReqScore_{wh}*Compatriot\_Winning_{wh}$	.024	.035**				
	(.016)	(.015)				
$1(ReqScore_{wh} \geq 0)*Compatriot\_Winning_{wh}$			.189**	.234***		
			(.103)	(.093)		
$1(ReqScore_{wh} \in (-4,0])$					-.007	-.015
					(.088)	(.088)
$1(ReqScore_{wh} \in (0,4])$					-.206**	-.219***
					(.090)	(.085)
$1(ReqScore_{wh} \geq 4)$					-.265***	-.298***
					(.140)	(.140)
$1(ReqScore_{wh} \in (-4,0])*Compatriot\_Winning_{wh}$					.005	.063
					(.157)	(.160)
$1(ReqScore_{wh} \in (0,4])*Compatriot\_Winning_{wh}$					.211***	.292***
					(.155)	(.143)
$1(ReqScore_{wh} \geq 4)*Compatriot\_Winning_{wh}$					.137	.285
					(.206)	(.219)
Intercept	-.135***	-.146***	-.019	-.023	-.015	-.014
	(.038)	(.043)	(.044)	(.045)	(.063)	(.057)
Heat Fixed Effect	NO	YES	NO	YES	NO	YES
N	252	252	252	252	252	252

Note: In column 1 I estimate the model presented in equation (3.4) where $ReqScore_{wh}$ is defined as the difference between the average of scores given by all four judges and the minimum score composing the sum of the best two waves so far in the heat. Column 2 considers $ReqScore_{wh}$ defined as a dummy variable that equals 1 for waves that change the sum of the best two waves in the heat and 0 otherwise. Column 3 considers a slightly modified version of column 2, but adding additional dummy variables for intervals of $ReqScore_{wh}$. The sample used in the estimations are all waves for the surfer competing against the surfer whose nationality coincide with a judge in the judging panel. Note that the number of observations differs slightly from Column 2 of Table (3.7), given the first two waves of every surfer in every heat are discarded to correctly generate the variable $ReqScore_{wh}$. All estimations include heat fixed effects. *** represents p<1%, ** represents p<5%.

APPENDIX

1. Introduction

In this section I present a slightly different strategy to estimate the nationalistic bias described in the third chapter of this thesis. For that, let S_{ijwh} be the score obtained by surfer i, from judge j, in wave w, and in heat h, $Judge\_Compatriot_{ijwh}$ be a dummy variable that takes value equal to 1 when surfer i and judge j share the same nationality and 0 otherwise, $Judge\_Competitor_{ijwh}$ be a dummy variable that takes value equal to 1 when judge j share the same nationality as the surfer competing against surfer i and 0 otherwise, and η_{ijw} be a wave fixed effect, such that

(A.1) $$S_{ijwh} = \gamma_0 + \gamma_1 Judge\_Compatriot_{ijwh} + \gamma_2 Judge\_Competitor_{ijwh} + \eta_{ijw} + \varepsilon_{ijwh}$$

The parameters of interest are γ_1 and γ_2. Parameter γ_1 represents the difference between the score given by the judge whose nationality matches that of one of the surfers and the score average of all other judges scoring wave w, while γ_2 represents the difference between the score given by the judge whose nationality matches the nationality of the surfer competing against surfer i and the score average of all other judges scoring wave w. The wave fixed effect, η_{ijw}, induces comparisons to be made within each wave, which implies that not only the same pair of surfers competing against each other is held fixed, but the judging panel is also taken to be constant. As a robustness check, I estimate the model using three variations of the specification above: considering heat fixed effects, considering wave fixed effects but clustering the standard errors at the heat level, and considering judge-surfer fixed effects.

Given the results presented above, one should expect to find $\gamma_1 = 0$, which would imply that there exists no differences between the score given by a judge scoring a surfer from his same nationality and the scores of all other judges in the judging panel. Similarly, γ_2 should be smaller than zero, given judges underscore athletes when competing against their countrymen. That is exactly what is observed in table (A.1). The coefficient on $Judge\_Competidor_{ijwh}$, γ_2, is negative and statistically significant, implying that there is a significant score penalty imposed on a athlete when competing against a surfer whose nationality coincide with that of a judge. The sign and significance of the coefficient is maintained when including wave fixed effects, heat fixed effects, and judge-surfer fixed effects. As expected, the coefficient on $Judge\_Compatriot_{ijwh}$ is small in magnitude and not statistically different from zero in all specifications.

Using this framework, it is also useful to look at judges' behavior when surfers are winning or losing the heat. For that, define $Losing_{ijwh}$ to be a dummy variable that equals 1 when surfer i is losing the heat and 0 otherwise, and consider the following model,

$$\text{(A.2)} \qquad S_{ijwh} = \gamma_0 + \gamma_1 Judge\_Compatriot_{ijwh} + \gamma_2 Judge\_Compatriot_{ijwh} * Losing_{ijwh} + \gamma_3 Judge\_Competitor_{ijwh} + \gamma_4 Judge\_Competitor_{ijwh} * Losing_{ijwh} + \eta_{ijw} + \varepsilon_{ijwh}$$

The parameters of interest are given by γ_2 and γ_4 and represent, respectfully, the difference in same-country judges behavior when surfer i is winning or losing the heat and the difference in behavior of judges that share same nationality as an opponent surfer when surfer i is winning or losing the heat. If one is to identify if the bias observed in table (A.1) is due to strategic bias and is not a result of country specific tastes, then the parameters of interest, specially γ_4, should be statistically different from zero, given that judges in this case would be behaving differently when surfers positions are exchange.

Results show that judges do not underscore/overscore their compatriots when they are losing or winning the heat, which was expected given the results presented in chapter 3. On the other hand, when looking at the athlete competing against the surfer backed up by a judge, results show a completely different picture. When the surfer born in the same country as one of the judges is losing, the judge whose countrymen is in the heat underscore his opponents waves by .377-.068. This negative effect almost vanishes if the surfer is winning. Only for specification (2) we observe a statistically insignificant coefficient. This, however, is expected, given wave quality changes significantly within heats and not controlling for wave fixed effects introduces problems in the interpretation of the coefficients.

2. Figures and Tables

Table A.1: Nationalistic Bias

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Judge_Compatriot$_{ijwh}$	-.003	.018	-.007	-.007	-.002	-.026
	(.060)	(.036)	(.010)	(.009)	(.010)	(.073)
Judge_Competidor$_{ijwh}$	-.166***	-.068**	-.046***	-.046***	-.042***	-.171***
	(.055)	(.031)	(.010)	(.008)	(.010)	(.064)
Wave Fixed Effect	NO	NO	YES	YES	YES	NO
Heat Fixed Effect	NO	YES	NO	NO	NO	NO
Judge-Surfer Fixed Effect	NO	NO	NO	NO	NO	YES
Intercept	3.746***	3.725***	3.725***	3.725***	3.856***	3.716***
	(.026)	(.003)	(.003)	(.002)	(.002)	(.031)
N	14,960	14,960	14,960	14,960	11,870	14,960

Note: Estimation is performed via equation (A.1). The bias is estimated by regressing the scores given by the judges for surfer *i* on a dummy variable that takes value equal to 1 when the nationality of the surfer matches the nationality of the judge and 0 otherwise, and a dummy that takes value equal to 1 when the nationality of the judge matches the nationality of the surfer competing against surfer *i*. Wave fixed effects are included in the estimation so that comparisons are made within the same wave score, hence holding the judging panel as well as the surfer pair of competitors fixed. In column 3, I include wave fixed effects, however, standard errors are clustered in the heat level. Column 5 does not consider waves surfer in the first round, i.e., the heats that are composed of three surfers. Column 6 I include judge-surfer fixed effects.

Table A.2: Nationalistic Bias

Variables	(1)	(2)	(3)	(4)	(5)
Judge_Compatriot$_{ijwh}$	-.052	-.055	-.005	-.005	-.008
	(.087)	(.075)	(.015)	(.014)	(.016)
Judge_Compatriot$_{ijwh}$	.101	-.009	.001	.001	.016
**Losing*$_{ijwh}$	(.137)	(.152)	(.023)	(.022)	(.024)
Judge_Competidor$_{ijwh}$	-.377***	-.121**	-.070***	-.070***	-.068***
	(.079)	(.067)	(.015)	(.013)	(.016)
Judge_Competidor$_{ijwh}$	.282**	.053	.042**	.042**	.046**
**Losing*$_{ijwh}$	(.117)	(.163)	(.021)	(.020)	(.023)
Wave Fixed Effect	NO	NO	YES	YES	YES
Heat Fixed Effect	NO	YES	NO	NO	NO
Intercept	3.698***	3.660***	3.725***	3.725***	3.856***
	(.030)	(.003)	(.003)	(.002)	(.002)
N	10,770	10,770	10,770	10,770	8,525

Note: Estimation is performed via equation (A.2). The variation in bias given by judges when the surfer is winning or losing the heat is captured by interacting a dummy variable that equals 1 when the surfer is losing and 0 otherwise with the dummies that identify when the nationality of the surfer matches the nationality of the judge and when the nationality of the judge matches the nationality of the surfer competing against surfer *i*. Wave fixed effects are included in the estimation so that comparisons are made within the same wave score, hence holding the judging panel as well as the surfer pair of competitors fixed. In column 3, I include wave fixed effects, however, standard errors are clustered in the heat level. Column 5 does not consider waves surfer in the first round, i.e., the heats that are composed of three surfers. Clustered standard errors at the wave level are presented in parentheses when using wave fixed effects. Note that the sample size is table (A.2) is considerably smaller then the one used in table (A.1), given we only observe winning/losing positions for 3 out of the 5 tournaments considered. Results presented in table (A.1), however, are exactly the same if we censor the sample to be the same as in table (A.2).

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 104: 493-505.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *The American Economic Review*, 93: 113-132.

Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105(1): 1-29.

Allen, Jeremiah, and Roger Barnsley. 1993. "Streams and Tiers: The Interaction of Ability, Maturity, and Training in Systems with Age-Dependent Recursive Selection." *The Journal of Human Resources*, 28(3): 649-659.

Angrist, Joshua, and Alan Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings." *The Quarterly Journal of Economics*, 106(4): 979-1014.

Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics:* an Empiricist's Companion. Princeton University Press.

Athey, Susan, and John Roberts. 2001. "Organizational Design: Decision Rights and Incentive Contracts." *American Economic Review, Papers and Proceedings*, 91(2): 200-205.

Bassett, Gilbert, and Joseph Persky. 1994. "Rating Skating." *Journal of the American Statistical Association*, 89(427): 1075-1079.

Bedard, Kelly, and Elizabeth Dhuey. 2006. "The Persistence of Early Children Maturity: International Evidence of Long-Run Age Effects." *The Quarterly Journal of Economics*, 121(4): 1437-1472.

Belot, Michele, and Vincent Vandenberghe. 2011. "Grade retention and educational attainment." *Education Economics*, forthcoming.

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2008. "Too young to leave the nest: The effects of school starting age." NBER working papers No. 13969.

Blalock, Garrick, Vrinda Kadiyali, and Daniel Simon. 2009. "Driving Fatalities After 9/11: A Hidden Cost of Terrorism." *Applied Economics*, 41: 1717-1729.

Bound, John, and David Jaeger. 2000. "Do Compulsory School Attendance Laws Alone Explain the Association Between Quarter of Birth and Earnings?" *Research in Labor Economics*, 19: 83-108.

Buckles, Kasey, and Daniel Hungerman. 2008. "Season of Birth and Later Outcomes: Old Questions, New Answers." NBER Working Paper No. 14573.

Caird, Jeff K., C. R. Willness, P. Steel, and C. Scialfa. 2008. A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis and Prevention*, 40: 1282-1293.

Campbell, Bryan, and John W. Galbraith. 1995. "Nonparametric Tests of the Unbiasedness of Olympic Figure-Skating Judgments." *Journal of the Royal Statistical Society. Series D (The Statistician)*, 45(4): 521-526.

Cohen, Joshua, and John D. Graham. 2003. "A revised economic analysis of restrictions on the use of cell phones while driving." *Risk Analysis*, 23: 5-17.

Crowe, S. M., and Jennifer Middeldorp. 1996. "A comparison of leg before wicket rates between Australians and their visiting teams for test cricket series played in Australia, 1977-94." *Journal of the Royal Statistical Society. Series D (The Statistician)*, 45(2): 255-262.

Dobkin, Carlos, and Fernando Ferreira. 2010. "Do school entry laws affect educational attainment and labor market outcomes?" *Economics of Education Review*, 29(1): 40-54.

Duggan, Mark, and Steven Levitt. 2002. "Winning Isn't Everything: Corruption in Sumo Wrestling." *American Economic Review*, 92(5): 1594-1605.

Edelman, Benjamin G., and Ian Larkin. 2009. "Demographics, Career Concerns or Social Comparison: Who Games SSRN Download Counts?" Harvard Business School NOM Unit Working Paper No. 09-096.

Elder, Todd, and Darren Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers." *Journal of Human Resources*, 44(3): 641-683.

Emerson, John W., and Silas Meredith. 2011. "Nationalistic Judging Bias in the 2000 Olympic Diving Competition." *Math Horizons*, forthcoming.

Emerson, John W., Miki Seltzer, and David Lin. 2009. "Assessing Judging Bias: an Example from the 2000 Olympic Games." *The American Statistician*, 63(2): 124-131.

Fredriksson, Peter, and Björn Öckert. 2005. "Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labour Market Performance." IZA Discussion Paper No. 1659.

Garicano, Luiz, Ignacio Palacios-Huerta, and Canice Prendergast. 2005. "Favoritism Under Social Pressure." *The Review of Economics and Statistics*, 87(2): 208-216.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201-

209.

Hartgen, David, and Lance Neumann. 2002. “Performance: a TQ point/counterpoint exchange with David T. Hartgen and Lance A. Neumann.” *Transportation*, 56: 5-19.

Hendren, Patricia, and Debbie A. Niemeier. 2008. “Identifying peer states for transportation system evaluation & policy analyses.” *Transportation*, 35: 445-465.

Holland, Paul. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association*, 81: 945-970.

Imbens, Guido, and Thomas Lemieux. 2008. “Regression discontinuity designs: A guide to practice.” *Journal of Econometrics*, 142(2): 615-635.

Imbens, Guido, and Jeffrey Wooldridge. 2009. “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature*, 47: 5-86.

Laberge-Nadeau, Claire, Urs Maag, François Bellavance, Sophie D. Lapiere, Denise Desjardins, Stéphane Messier, and Abdelnasser Saïdi. 2003. “Wireless Telephones and the Risk of Road Crashes.” *Accident Analysis and Prevention*, 35: 649-660.

Lee, David S. (2008). “Randomized experiments from non-random selection in U.S. house elections,” *Journal of Econometrics*, 142(2): 675-697.

Lee, David S., and David Card. 2008. “Regression discontinuity inference with specification error.” *Journal of Econometrics*, 142(2): 655-674.

Lee, David S., and Thomas Lemieux. 2009. “Regression discontinuity designs in economics.” NBER Working Paper No. 14723.

Levitt, Steven D., and Chad Syverson. 2008. “Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions.” *The Review of Economics and Statistics*, 90(4): 599-611.

Love, Joseph, and Werner Baer. 2009. *Brazil Under Lula: Economy, Politics, And Society Under The Worker-president*. Palgrave Macmillan.

Ludwig, Jens, and Douglas L. Miller. 2007. “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design.” *The Quarterly Journal of Economics*, 122(1): 159-208.

Machado, Danielle C. 2008. “Efeitos da Saude na Idade de Entrada a Escola.” *Pesquisa e Planejamento Economico,* 38(1): 67-96.

McCartt, Anne T., and L. L. Geary. 2004. “Longer term effects of New York State’s law on drivers’ handheld cell phone use.” *Injury Prevention*, 10: 11-15.

McEwan, Patrick J., and Joseph S. Shapiro. 2008. "The Benefits of Delayed Primary School Enrollment." *Journal of Human Resources*, 41(1): 1-29.

McEvoy, Suzanne P., Mark R. Stevenson, Anne T. McCartt, Mark Woodward, Claire Haworth, Peter Palamara, and Rina Cercarelli. 2005. "Role of Mobile Phones in Motor Vehicle Crashes Resulting in Hospital Attendance: A Case-Crossover Study." *British Medical Journal*, doi:10.1136/bmj.38537.397512.55.

Myers, Tony D., Nigel J. Balmer, Alan M. Nevill and Yahya Al-Nakeeb. 2006. "Evidence of Nationalistic Bias in Muaythay." *Journal of Sports Science and Medicine*, 5: 21-27.

Nannicini, Tommaso, and Roberto Ricciuti. 2010. "Autocratic Transitions to Growth." CESifo Working Paper No. 2967.

Nikolaev, Alexander G., Matthew J. Robbins, and Sheldon H. Jacobson. 2010. "Evaluating the impact of legislation prohibiting hand-held cell phone use while driving." *Transportation Research Part A: Policy and Practice*, 44: 182-193.

Page, Lionel, and Katie Page. 2010. "Evidence of referees' national favouritism in rugby." Discussion Paper, University of Westminster.

Porter, Jack. 2003. "Estimation in the Regression Discontinuity Model." Working Paper, Harvard University, Department of Economics.

Prendergast, Candice, and Robert H. Topel. 1996. "Favoritism in Organizations." *Journal of Political Economy*, 104(5): 958-978.

Price, Joe, and Justin Wolfers. 2010. "Racial Discrimination Among NBA Referees." *Quarterly Journal of Economics,* 125(4): 1859-1887.

Puhani, Patrick A., and Andrea M. Weber. 2005. "Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Educational Effects of Age of School Entry in Germany." *Empirical Economics*, 32: 359-386.

Redelmeier, Donald A., and Robert J. Tibshirani. 1997. "Association Between Cellular-Telephone Calls and Motor Vehicle Collisions." *New England Journal of Medicine*, 336: 453-458.

Sampaio, Breno. 2010. "On the identification of the effect of prohibiting hand- held cell phone use while driving: Comment." *Transportation Research Part A: Policy and Practice*, 44: 766-770.

Seltzer, Richard, and Wayne Glass. 1991. "International politics and judging in Olympic skating events: 1968-1988." *Journal of Sport Behavior*, 14(3): 189-200.

Smith, Justin. 2009. “Can Regression Discontinuity Help Answer an Age- Old Question in Education? The Effect of Age on Elementary and Secondary School Achievement.” *The B.E. Journal of Economic Analysis & Policy*, 9(1) Article 48. Available at: http://www.bepress.com/bejeap/vol9/iss1/art48.

Soares, Sergei. 2006. “Aprendizado e seleção: uma análise de evolução educacional brasileira de acordo com uma perspectiva de ciclo de vida.” Texto para Discussão No. 1185. Brasília, Brasil: Instituto de Pesquisa Econômica Aplicada.

Ste-Marie, Diane. 1996. “International Bias in Gymnastic Judging: Conscious or Unconscious Influences?” *Perceptual & Motor Skills*, 83(3): 963-975.

Strayer, David L., and Frank A. Drews. 2004. “Profiles in driver distraction: effects of cell phone conversations on younger and old drivers.” *Human Factors*, 46: 640-649.

Zitzewitz, Eric. 2006. “Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making.” *Journal of Economics & Management Strategy*, 15(1): 67-99.

Zitzewitz, Eric. 2011. “Forensic Economics.” *Journal of Economic Literature*, forthcoming.