

A Responses to Editors and Referee

I would like to thank the editor and the anonymous referees for their insightful comments, suggestions, and effort and time in reviewing this paper. I have addressed all the comments and suggestions in the revised manuscript. Below, I provide a summary of the changes made to the manuscript in response to the comments and suggestions.

B Responses to Referee One

I would like to thank referee one for the insightful comments and suggestions. Below is a detailed response to the comments and suggestions.

R1: 1. My first concern relates to the threats to identification. It is feasible that children born to HW couples may differ systematically to children born to WH couples, particularly with respect to unobservable characteristics important for labor market outcomes. For example, HW pairs may exhibit differing parental characteristics related to parenting styles, parental preferences for education, gender norms and beliefs. The paper would be improved by including an expanded discussion around the threat to identification. In light of these issues, the paper would be improved by providing a more thorough discussion of the relative advantages of the current approach, compared to the conventional Oaxaca-Blinder decomposition.

Thank you for raising these concerns. I have significantly expanded the discussion of identification threats and the advantages of my approach compared to traditional methods. In Section 4 (“From the Data: The Differences Between HW and WH Couples”), I now explicitly acknowledge that “Unlike traditional decomposition methods, this comparison directly isolates the role of surname signaling, reducing the confounding effect of background disparities. By focusing specifically on children from interethnic marriages with similar parental characteristics but different surname ethnicities, I use a more targeted comparison group that better isolates the effect of perceived ethnicity from family background. Traditional decomposition approaches would struggle to separate discrimination effects from the broader socioeconomic gaps between Hispanic and White families overall. The smaller educational and income disparities between HW and WH families compared to HH and WW families, allowing for a clearer attribution of outcome differences to discrimination rather than unobserved heterogeneity. Nevertheless,

I acknowledge that residual selection bias remains a concern, particularly regarding cultural and attitudinal factors that may systematically differ between HW and WH households despite similar educational and economic outcomes.”

However, I argue that this approach offers distinct advantages over traditional Oaxaca-Blinder-Kitagawa decomposition methods. First, by focusing specifically on children from interethnic marriages with similar parental characteristics but different surname ethnicities, I use a more targeted comparison group that better isolates the effect of perceived ethnicity from family background. Traditional decomposition approaches would struggle to separate discrimination effects from the broader socioeconomic gaps between Hispanic and White families overall. Second, the smaller educational and income disparities between HW and WH families compared to HH and WW families, allowing for a clearer attribution of outcome differences to discrimination rather than unobserved heterogeneity.

The empirical evidence supports this approach: Table 4 shows that interethnic couples are much more similar to each other than endogamous couples are to each other, which is exactly why comparing children of interethnic couples provides a better comparison than comparing children of endogamous marriages. Nevertheless, I acknowledge that residual selection bias remains a concern, particularly regarding cultural and attitudinal factors that may systematically differ between HW and WH households despite similar educational and economic outcomes. Moreover, it is likely that the parenting and cultural backgrounds of interethnic children are more similar to each other than those of children of endogamous marriages.

R1: 2. The analysis sample is restricted to individuals who self-

identify as 'White'. It is likely that this self-identification of ethnicity is endogenous to labor market outcomes. Importantly, the unobservable determinants of racial identification are potentially correlated with labor market outcomes. The paper would be improved by including a more thorough discussion of this (sample) selection issue. Related to this is the depiction of parents born in the United States as 'White'. This seems particularly over-simplifying and ignores the long history of immigration. A more accurate description would be to refer to this group as US born.

Thank you for your comments and suggestions. I have addressed both concerns raised by the reviewer. First, regarding the endogeneity of self-identification, I now explicitly discuss this issue in the data section. I explain that "The sample is restricted to Hispanic and non-Hispanic White individuals to avoid confounding racial factors." I argue that including those who do not identify as White could contaminate the estimate of bias against Hispanics due to racial signals. Moreover, as shown by Hadah (2024), there exists strong correlation between bias against Hispanics and self-reported Hispanic identity, and excluding Hispanics who experience ethnic attrition could lead to overestimated bias in the most discriminatory states. To address the comment on the long history of immigration in the US, I updated the manuscript to make clear that the interethnic children I am studying are US born. I also added the following: "This study's focus on US-born children with foreign-born parents means the findings may not generalize to Hispanic children with US-born parents, who likely face different socioeconomic and cultural circumstances. The analysis does not account for heterogeneity in immigrant characteristics in different Spanish-speaking countries of origin, such as

variations in educational attainment, socioeconomic status, and gender-specific migration patterns, which could influence both parental selection in migration and subsequent child outcomes.”

Second, regarding the characterization of US-born parents, I have revised the language throughout the manuscript to make it clear that the analysis focuses on US-born individuals. I now explicitly address this concern: “While the United States has a long history of immigration, the probability that a US-born parent in this sample is a second-generation or later immigrant from a Spanish-speaking country is very low, as shown in Table 5. The majority of Hispanics in the US during this period were first- and second-generation immigrants. Only 3% of native-born Americans identified as Hispanic during this period, making it statistically unlikely that an interethnic child with a native-born parent is also a second-generation Hispanic immigrant.” I have also changed terminology throughout to use “US-born” rather than “White” when referring to native-born parents, and I clarify that this approach “provides a cleaner identification of surname effects, which is the primary focus of this study.”

R1: 3. The discussion on the construction of the ‘synthetic parents’ is somewhat brief. The paper suggests that the potential parents are matched using the birth year of the child and the parent’s place of birth to the children’s information collected in the CPS sample. While not explicitly stated in the paper, I assume that this sample of synthetic parents is used to construct mean parental education and family income at the time of the birth of the child. The process for matching the potential parents to the children is somewhat aggregated and

there is likely considerable heterogeneity in the mean educational attainment of the potential parents within the set of children with the same birth year with the same parent's country of birth. The paper could be improved by exploring the possibility of improving the quality of the match of potential parents by including further characteristics to match the potential parents to their children.

Thank you for raising this point. I have significantly expanded the discussion of synthetic parents construction and provided a detailed concrete example. In the data section, I now include: "To illustrate the construction of synthetic parents more concretely, consider someone who was 35 years old in 1999, meaning they were born in 1964. If this person's mother was born in Mexico and their father was born in the United States, their 'synthetic parents' would be identified using the 1970 Census data, when the person was 6 years old. The 'synthetic mother' would have the average characteristics (education, income, etc.) of Mexican-born women who were married to US-born men and had children around 1964, when they were between 20 and 35 years old (meaning they were born between 1929 and 1944). Similarly, the 'synthetic father' would have the average characteristics of US-born men who were married to Mexican-born women and had children in that same year."

I acknowledge the reviewer's suggestion about including additional characteristics to improve the matching quality. While incorporating more matching variables (such as parent's age at birth) could potentially reduce heterogeneity within the synthetic parent groups, the current approach balances precision with sample size considerations. Adding more matching criteria would create increasingly small cells in the Census data, potentially leading to unstable estimates or

empty cells, particularly for less common parent birthplace combinations. Moreover, since the synthetic parent characteristics are used exclusively for descriptive summary statistics and not as control variables in the main regression analyses, the current level of aggregation is appropriate for the paper's empirical strategy.

R1: 4. The description of the estimated model is somewhat brief. However, there are two main issues associated with statistical inference on the estimated parameters in model (1). First, there is a 'generated regressor' issue associated with using estimated group-level parental education. A failure to account for this sampling variation will lead to misleadingly small standard errors. Second, given this group structure for parental education, it seems reasonable to assume that the model errors are uncorrelated across clusters but correlated within (potential parents) clusters. It is well understood that failing to account for this 'clustering' problem can lead to misleadingly small standard errors, narrow confidence intervals, and low p-values.

I am grateful to the referee for raising this important concern. I have expanded the discussion of these methodological issues in the empirical approach section. Regarding the generated regressor issue, I updated the regression results in the manuscript with results that do not include the specification where 'synthetic parents' are included in the controls. This alleviates the concern regarding generated, as they are no longer in the regression. I also updated the manuscript to make sure that it is clear that 'synthetic parents' parents are only used for descriptive summary statistics to enhance comparability between interethnic families and are not included as control variables in any of the main regressions.

R1: 5. The paper would be improved through an expanded discussion of the impact of measurement error on the reported estimates. While there is a well understood result in the measurement error literature that measures of group mean parental education will provide estimates that are more robust to the presence of measurement error in individual level measures of parental education, there is still an issue with the non-random attrition of potential parents. Specifically, not all potential parents have children in the CPS sample, nor do all children in the CPS sample have parents in the sample of potential parents. Moreover, the non-random attrition of potential parents implies that mean parental education may be systematically higher or lower than the actual parental education of the children in the CPS sample. The paper would be improved by providing an expanded discussion of the likely impacts of measurement error on the reported estimates.

Thank you for your comments. I have clarified this methodological point and expanded the discussion of measurement error concerns. First, I want to clarify that the concern about attrition does not apply to my analysis: Synthetic parents are constructed using the birth year of the child and are not based on actual parents in the CPS sample at the time of the survey. The place of birth of parents—mothers and fathers—are questions asked of all participants in the Current Population Survey (CPS) starting 1994. Therefore, I do not use the information of parents who are in the CPS sample at the time of the survey, and I construct synthetic parents from Census data. Because this approach uses Census data rather than contemporaneous parent-child pairs, concerns about non-random attrition of parents from the CPS sample do not apply to my analysis.

Regarding measurement error more broadly, I have expanded the discussion in the empirical approach section to include the following: “A potential concern is measurement error in using CPS data to infer parental ethnicity, since it relies on place of birth rather than self-identified Hispanic origin. However, during the period studied, second-generation Hispanics were rare: only about 3% of native-born Americans identified as Hispanic (see Table 5, which presents Hispanic identification rates by nativity). This makes it unlikely that US-born parents in interethnic unions are second-generation+ Hispanic immigrants.”

C Responses to Referee Two

I would like to thank referee two for the insightful and constructive comments and suggestions. Below is a detailed response to the comments and suggestions.

R2: 1. The paper needs to be more clearly motivated in the introduction. There are a lot of reasons why studying discrimination by race and ethnicity is important and the paper would benefit from clearly articulating this, including discussing the implications of this discrimination. You mention economic mobility and I think there is more you can say about this. I also think clearly stating the contributions of the paper earlier on would be important.

Thank you for the important suggestion. I have expanded the introduction to better motivate the study. I have included a discussion of the implications of discrimination on economic mobility and the contributions of the paper. I have also included a discussion of the importance of studying discrimination

R2: 2. I would also like to see a clearer theory and review of the literature on the topic. For example, why should we expect that discrimination will affect educational attainment? Hispanics are one of the fastest growing groups entering college, but the extent to which they complete an associate or BA degree may reflect discrimination that happens when students are in school that ultimately translates into differences in attainment. Why might we expect differences in employment and years of education? Why might we expect differences by gender?

I appreciate this suggestion. Here are my replies to the different comments.

First, regarding including more on why discrimination could affect educational attainment. I have expanded the part of the introduction that discusses

differences in education and the channels in which discrimination can affect educational attainment. The literature on discrimination and education shows that discrimination can affect educational attainment of minorities through various channels, including differences in school quality and bias from teachers, administrators, counselors, etc. These biases could lead to differences in educational attainment by preventing students access to some schools, recommendation letters, or counselors restricting access to more advanced courses, etc. Consequently, these biases could lead to differences in educational attainment. For example, having access to advanced courses in high school or better recommendation letters could affect the likelihood of attending college.

Second, if minorities face discrimination in access to education and the labor market, then we would expect differences in employment and years of education. The literature shows that discrimination can affect labor market outcomes through various channels, including differences in access to jobs and wage differences.

Finally, to address the comment here, and in other places, on why we might expect differences between genders. I added a discussion of why we might expect differences between men and women in the results section. I believe that showing that there might be heterogeneity in gaps between men and women in educational outcomes to be an important contribution of the paper since the literature studying gaps in earnings mainly focuses on the average gap between men. Moreover, showing that couples with a Hispanic husband and a White wife do not invest differently in their children than couples with a White husband and a Hispanic wife could be a way to test for cultural differences between the two groups.

R2: 3. Occupational segregation can also reflect discrimination in the labor market with important implications for economic mobil-

ity and other outcomes, and it would be important to mention this in the literature review. In general, the review of the literature needs to provide more details about the different studies and how your paper contributes to that literature above and beyond using a cleaner comparison.

Thank you for this suggestion. I have included a discussion of occupational segregation in the literature review and added how my paper contributes to the different strands of the literature.

R2: 4. There are places throughout the paper where additional references are needed. For example, when you state in the introduction on page 3 that discrimination can lead to lower wages, reduced opportunities, and hinder assimilation this statement needs references.

Thank you for pointing this out. I have added references to such statements.

R2: 5. Can you explain why you discuss assimilation and in what ways this is connected to your theory given that you are focusing on U.S. born children? If this matters because you are focusing on U.S. born Hispanic children with one foreign-born parent, then you need to clearly state this.

Thank you for the comment. I added a footnote to the introduction to address it.

R2: 6. You also need to state clearly and early in the paper that your study focuses on children with a foreign-born parent. Children with U.S. born Hispanic parents may be different than those with foreign-born parents and you could more explicitly discuss this in the paper.

I appreciate this comment. I made it more clear throughout the paper that I focus on children with a foreign-born parent.

R2: 7. Also given the differences in the characteristics of immigrants from different Spanish-speaking countries living in the U.S. (differences in socioeconomic status, education, etc.), you probably want to mention that your study is not capturing this. Further, to the extent that men and women migrate from different countries and have different pre- and post-migration characteristics then this might affect your results.

One possible way to address this is to conduct sensitivity analyses limited to Children whose parents were likely born in Mexico or who respond themselves that they are Mexican. Given that immigrants from Mexico are the largest immigrant group from a Spanish speaking country in the U.S., this may offer an even cleaner comparison. There could still be differences in the characteristics of Mexican mothers and fathers, but you could potentially check this in the data.

Thank you for this suggestion. I made it clear in the paper that I do not capture differences in the characteristics of immigrants from different Spanish-speaking countries. I also added some results and discussion as a sensitivity analysis breaking down the results for Hispanic children with Mexican parents versus Hispanic children with non-Mexican parents.

R2: 8. There are several sections that are repeated in the paper and I would suggest you streamline the text. For example, you review results twice, but this isn't necessary.

Similarly, I think you can more systematically organize the section describing your empirical strategy and clearly explaining your identification strategy, the concerns that it helps you overcome, your assumptions, and how you are testing whether these assumptions likely

hold. Currently this is explained in multiple sections throughout the paper.

I appreciate this comment. I have streamlined the text and reorganized the paper to address these concerns.

R2: 9. Can you provide an example that illustrates how you link individuals in the CPS to the synthetic parents?

I added an example to the data section that illustrates how I link individuals in the CPS to the synthetic parents.

R2: 10. I would like more details about the sample and the decisions you make. For example, you identify your sample as U.S. born children who identify as White in the CPS but some of these may also identify as Hispanic no? Do you restrict your sample to respondents who identify as White Hispanic or not? How many individuals in your sample identify as White non-Hispanic even if they have a parent born in a Spanish speaking country and how many do not? If someone has a Hispanic parent but do not identify as Hispanic those may be different than people with a Hispanic parent who identify as Hispanic.

If there are people who identify as White non Hispanic in your sample even though they have a parent born in a Spanish speaking country, can you do a sensitivity analysis removing them?

For example, on page 18 you say “I also find a significant earnings gap between those that identify as Hispanic.” Can you explain this statement? Is this related to the point I made above?

Thank you for the comment. I added more description of the sample to the data section and the reasoning of why I chose those who self-identify as White.

R2: 11. If there are people who identify as White non Hispanic in your sample even though they have a parent born in a Spanish speaking country, can you do a sensitivity analysis removing them?

Thank you for the comment. I added more description of the sample to the data section and the reasoning of why I chose those who self-identify as White.

R2: 12. For example, on page 18 you say “I also find a significant earnings gap between those that identify as Hispanic.” Can you explain this statement? Is this related to the point I made above?

Thank you for the comment. I added more description of the sample to the data section and the reasoning of why I chose those who self-identify as White.

R2: 13. You report results for men and women and, I am sorry if I missed it, but I would like to see this motivated in the paper as there are many reasons for doing the analyses separately. See my earlier point on the lit review/theory.

Please see my response to the earlier comment on why we might expect differences between men and women.

R2: 14. It would also be important to discuss effect sizes. Some of the findings seem to be small and you should discuss whether they are economically meaningful. Further, there should be more discussion about what your findings mean to understand discrimination for Hispanics and how it plays out, including by gender and in what ways they fall short in answering this question.

Thank you for the suggestion. I have included a discussion of the effect sizes and the economic meaning of the results. I have also included a discussion of

what the findings mean for understanding discrimination against Hispanics and how it plays out.

R2: 15. How are you measuring the different outcomes? I don't believe you discuss in the paper. This may be obvious, but I think it is still important to mention for clarity.

Thank you for the comment. I have included a discussion of how I measure the different outcomes in the data section.

R2: 16. When you control for education in models that examine earnings, the effect becomes statistically insignificant, what happens if you include industry or occupation fixed effects? Can you do that? Then you'd be comparing people within the same industry or occupation.

I appreciate this comment. I have added a discussion of the results when I include occupation fixed effects to the sensitivity analysis section.

R2: 17. The conclusion repeats much of what was said in the body of the paper, and I would like to see more discussion about the implications of your findings to understand disparities in education and labor outcomes between people with different ethnic background, and to what extent you can conclude these differences can be attributed to discrimination. It would also be important to compare your results with other literature on this topic.

Thank you so much for this comment. I have expanded the conclusion to include more discussion about the implications of the findings and how they can be

attributed to discrimination. I have also included a discussion of how my results compare to the literature on this topic.

R2: 18. On page 14, this sentence seems to be incomplete: “Consequently, comparing WH and HW children to each other to analyze discrimination against Hispanics in the labor market.”

Thank you for pointing this out. I have corrected this sentence.

R2: 19. Throughout the paper you refer to people with a parent born in a Spanish speaking country as “children who have a Spanish-sounding last name” I would encourage you to soften this language and say “who likely have” a Spanish sounding last name. First, you do not know if a person actually does have the last name of their father. Second, while it is true that Spanish sounding last names are very common in Spanish-speaking Latin American countries and among Latinos in the U.S., this may not be true across the board and many people born in Spanish speaking countries, who are Spanish speakers and Hispanic, may not have a traditionally sounding Spanish last name.

Thank you for your comment. I changed the mention of "children who have a Spanish-sounding last name" to "children who likely have a Spanish-sounding last name" throughout the paper.

D Responses to Referee three

I would like to thank referee three the comments. Below is a detailed response to the comments and suggestions.

R3: 1. As the author states on page 4, the key identifying assumption is that people born to HW parents are similar to their WH peers. Only when this assumption is fulfilled can the author attribute the estimated coefficients to evidence of discrimination. However, Table 4 clearly shows that the differences between HW and WH synthetic parents' characteristics, such as the father's/mother's education and total family income, are all significantly different. More specifically, HW families exhibit lower levels of education and income compared to their WH counterparts. Based on this, it is not clear whether the estimated coefficients are indeed evidence of discrimination or they merely reflect the fact that HW children grew up in an environment with less resources/educated parents.

The manuscript heavily adopts the methodology of Rubinstein and Brenner (2014), which examines the impacts of having a Sephardic-sounding surname on wages by comparing the Israeli-Jewish men born to Sephardic fathers and Ashkenazi mothers (SA) with those born to Ashkenazi fathers and Sephardic mothers (AS). They too find that the AS and SA families have statistically different educational and labor market outcomes. However, in their case, despite the fact that Sephardic Israelis face tougher labor market conditions, SA parents exhibit better education and labor market outcomes than AS parents. Therefore, any evidence indicating that SA offspring have worse labor market outcomes provides convincing (and potentially a lower bound) evidence of discrimination.

Thank you for the comment. Even though the synthetic parents have different characteristics, the estimation strategy rests on the fact that selection in the marriage market decreases the differences between HW and WH couples. This is the reason why comparing children of intermarried couples provides a better com-

parison than comparing children of endogamous marriages, i.e. marriages where both parents are either non-Hispanic White or Hispanic. I also included a discussion of the relative advantages of the current approach compared to the conventional Oaxaca-Blinder-Kitagawa decomposition. I argue that the current approach provides a more accurate estimate of gaps that are due to discrimination. In fact, the concern the reviewer raises is one of the reasons why my approach is preferred over the Oaxaca-Blinder-Kitagawa since children of endogamous marriages are more likely to have different characteristics than children of intermarried couples, including those that are unobservable. I argue, and show from the data, that children of intermarried couples are more likely to have similar characteristics than children of endogamous couples.