**Princeton University**

**Department of Economics**

**Ph.D. Program**

**2023-2024**

**Third Year Paper Comments**

Student: Amy Kim

Paper Title: Who Pays the cost of exclusion? Selection into immigration under the 1885 Chinese head tax

First Reader: Leah Boustan

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Comments:**

1. Your paper has a lot of great material and interesting findings. Most of my comments will be focused on how to rewrite the draft to get the ideas across as clearly as possible.
2. Your first paragraph is too “flowery” for an academic paper and could be condensed. The main point is that, at certain moments in history, US and Canadian immigration policy has been designed to exclude specific groups or to implicitly limit immigration through raising costs.
3. There are two motivations to the paper, as I see it. One can be the main motivation and one can be secondary. First, it is hard to measure migration costs today because the costs take the form of bureaucratic hurdles, lawyers’ fees, etc. Yet we have a sense that heightened migration costs likely affect the number of entrants and the selection of who migrates. You have found a historical setting that allows you to test this relationship. Second, the Chinese are an important and overlooked group in US and Canadian immigration history. They are the first group to face group-specific exclusion and became a precursor to later restrictive immigration policy. You are studying the effect of these exclusionary policies. Both of these motivations are important – make sure they are both clearly stated on the first page.
4. Explain in the introduction why height is a good measure of selectivity/human capital.
5. The other measures (e.g., working as a laborer, owning a house) are also intended to be measures of selection, but they sound like they are effects in the destination (you say “after the tax, Chinese immigrants *became* less likely to…”). Make sure the wording here is clear focusing on why you study these outcomes.
6. Your relationship to the literature on p. 2-3 is still not fully clear. The earlier historical work on migrant selection was undertaken in contexts with low migration costs. Thus, a simple Roy model might fit the facts. The later modern work raises the point about skill-varying migration costs because the implicit migration costs have increased a lot in the modern period (due to immigration policy). In your current writing, this point gets lost. You say “the vast majority of this work is focused on Europe,” which – to me – implies that the value-added is looking at another, non-European group. But the point is bigger than this! It’s not just that you are looking at the Chinese, but that this is a group that faced high migration costs in the past. To address this problem: Maybe start with a few sentences outlining the Roy model & extension to consider costs, and then talk about the empirical literature?
7. Good historical background section
8. There are two issues with the Census data: return migration is one of them, but also the year of immigration variable is often very badly measured. Both of these will attenuate your results. To bound the first problem, you could report on return migration rates and consider what the implied coefficient would be by “re-inflating” the numbers in the Census. To address the second problem (measurement error), you could instrument for the Census data with the register data – if measurement error is a problem, instrumenting with a second measure should increase your coefficient.
9. Table 1: 57% of Chinese migrants are married but 97.5% are male. This must imply that many married men migrated alone and planned to return to China. Make a comment on this.
10. Figure 1: There does not appear to be any data before 1886 in these series. (Maybe they start in 1884 or 1885, I can’t tell?). And yet you talk about the effect of introducing the Head Tax in 1886. Do you only have one or two years of pre-period data? Comment on this.
11. In describing Figure 1, better to start at the beginning (chronologically) and go from there. It’s awkward to start in 1903 and then back track.
12. For equation 1: Put your main RHS variable first, right after the constant. Talk about this variable first and only then talk about the controls in the text below the equation too.
13. The controls that you add are very sensible and help you replicate as close as possible the Hatton/Williamson style work despite not having all of their variables (e.g., GDP per capita). Be sure to cite their papers and other related papers somewhere. You don’t currently have a related literature section. You could add.
14. For POPSTOCK, I would refer to this type of variable as a ‘network’ variable rather than “Canada-specific push” or “pull.” This would coincide with how Hatton/Williamson talk about ethnic networks.
15. Bottom of p. 11: Mention the magnitudes (e.g., -1,300) rather than only saying “negative” or “not significantly negative” or “qualitatively similar.”
16. The content on p. 12 about temporary vs. permanent declines could be cut or substantially condensed or put in appendix. Ditto with the columns in the table that report on these results. I didn’t get much out of these findings and I didn’t follow the motivation for why you looked at them. (I think I remember Ilyana asked about something along these lines…. If it didn’t result in something fruitful, you can cut).
17. I thought that the paragraph on p. 13 about a lack of effect on other immigrant groups was important and could be moved to the main text.
18. Table 2: Indicate clearly at the bottom of the table which controls are included in each column.
19. The content on p. 13-16 summarizing the two models of selection and the contrary predictions of each is very useful and will help elevate the paper. However, right now, the material is not framed clearly enough. You dive in with just “I start with the standard Roy-Borjas model of selection…” For a reader who isn’t already aware of what you are doing, it will not be clear why they should read a page about the Roy model or what is coming next. So, you should start this section with an overview paragraph explaining that there are two models of migrant selection that result in differing predictions about how an increase in migration costs will impact selection. Let’s say that initial selection is negative. In model 1 (Roy-Borjas), an increase in migration costs will accentuate this negative selection. The least-negatively-selected migrant will now choose to stay home, and only the most negatively selected will migrate. *Migration will become more negatively selected.* In model 2 (Chiquiar-Hanson), migration costs vary with skill. Even though the migration tax is skill-neutral, the least skilled may have the hardest time borrowing money to pay for a migration tax. In that case, the most negatively selected migrants will stay home, and *migration will become more positively selected.* Once you explain these ideas upfront, readers will be able to follow the details and will be invested in why they matter.
20. It’s a little strong to say that height “cannot be affected by the decision to migrate.” Many immigrants moved around age 16-18 at that time and they could invest in nutrition in the years leading up to that decision.
21. The material on p. 18 explaining the benchmark heights for the Chinese non-migrants is interesting. It would be useful to place straight lines on Figure 2 emphasizing these reference points.
22. I think you will have to delve further into the odd literacy result with the Japanese comparison. The current explanation is not satisfiying.