

Replication of Import Competition and Internal Migration

Chuxin Liu^a

^aCUNY Graduate Center

Abstract: This paper replicates the main result of "Import Competition and Internal Migration" [1], authored by Andrew Greenland, John Lopresti, and Peter McHenry and published on *Review of Economics and Statistics* in 2019. The replication result is very close to the original work while slight downward estimates are observed. This paper also has an extension which commuting zone size is taken into account. It is shown that for larger cities, the negative effect of import competition on population growth will be smaller.

1. Introduction

This paper replicates the main result of "Import Competition and Internal Migration" [1], authored by Greenland, Andrew, John Lopresti, and Peter McHenry and published on *Review of Economics and Statistics* in 2019. The replication result is very close to the original work while slight downward estimates are observed. This paper also has an extension which commuting zone size is taken into account. It is shown that for larger cities, the negative effect of import competition on population growth will be smaller.

Greenland et al (2019)[1] examine the U.S. internal migration response to increased import competition following the granting of Permanent Normal Trade Relations to China in 2011. Using a variety of data sets and empirical approaches, they find that local labor markets most exposed to the policy change experienced a relative reduction in population growth over the following decade.

This paper is part of the big economic question of winners and losers in trade.

While the gains from trade is spread throughout the economy, costs of trade is highly concentrated. More specifically, the cost is concentrated among low education and low wage labor force. In the literature, there is documented evidence that trade shock leads to lower wages, lower labor force participation, and higher unemployment among low education and low wage population. It is also found that they experience worse physical health, worse mental health, and increased mortality risk. Also, the cost is concentrated geographically. Literature documents that areas with high trade shock also experienced losses in income, lower tax base and less funding for public good. Therefore, there is huge challenge for both trade policy making and place-based policy making.

It is worth noting that when this paper was written, no documented migration due to trade shock. Therefore, this paper is the first in the literature to address how people adjust to changing labor market opportunities by moving. After this paper was written, there are some PhD students working on similar topic, which also tries to address how labor move across space in react to trade shock and technology shock. I have heard two presentations from the Urban Economics Summer School, but I have not found the citation of those two papers.

The major data used in this paper is US Census, NELS:88 and ELS:2002, and IRS Migration Data. While NELS and ELS help address childhood to adult location transition and IRS Migration Data addresses timeline of population change in react to trade shock, the US Census produces the most important results of this paper. It is representative of the whole country, and it is sourced from administrative data, which confirms the reliability of the data used in the research. Ideally, we might want to have more years as US Census is only done in every 10 years.

The research is done with empirical approach without theoretical framework, but it is convincing since the aim is only to document migration, without the need to estimate migration cost, trade cost, trade loss etc. In terms of internal validity, this papers key independent variable of interest is constructed by previous work. The authors fully take advantage of the previous work done by famous economists, including David Autor, Pierce and Schott, which allows minimal discussion on endogeneity. In my opinion, the paper is very convincing.

The remainder of the paper proceeds as follows. Section 2 shows the replication by describing data sets I use and showing the results I have. In section 3, I will discuss the difference between replication and the original paper. I will compare the data I generated and the data from published replication package. In section 4, I will conduct an extension by taking commuting zone size into account and

show that larger commuting zones experience less population loss while holding trade shock constant. Section 5 includes a brief discussion and conclusion.

2. Replication

2.1. Data

To replicate the main result of this paper, the following data sets are used.

- **US Census**

The US Census is representative of the whole country, available every 10 years on IPUMS. It is a well-known data source and provides reliable population measures. From IPUMS, location is coded as county or puma. In order to use them in this paper, county or puma will be converted into commuting zones using crosswalk files.

- **Trade shock measure by Pierce and Schott (2016) [2]**

Pierce and Schott (2016) [2] find that import competition for United State industries changed dramatically when the U.S. granted Permanent Normal Trade Relations (PNTR) to China in 2001. Pierce and Schott (2016) [2] take a difference-in-differences approach to examine the effects of the policy change, exploiting the variation in the potential tariff increase across industries before and after the granting of PNTR in 2001. In addition to increased import competition, they show that the industries most exposed to the policy change experienced a reduction in employment in the years after 2001. They define the difference between the observed NTR tariff rates and the potential non-NTR rates in industry j as the NTR gap facing the industry. As the outcome of interest is inherently geographical, this paper requires a geographically based measure of import competition. To obtain this, the paper create a commuting zone (henceforth CZ) level measure of the NTR gap. CZs, a concept pioneered by Tolbert and Sizer (1996)[3] and employed recently in a trade context by Autor et al. (2016)[4], are approximations of local labor markets, created using journey-to-work data such that individuals living in a given CZ are highly likely to work in the same CZ. There are 722 CZs in this data, spanning the entire continental United States. Hawaii and Alaska are not included. This paper creates a CZ-level measure of the NTR gap by weighting industry NTR gaps by the share of

employment in CZ c accounted for by each industry j .

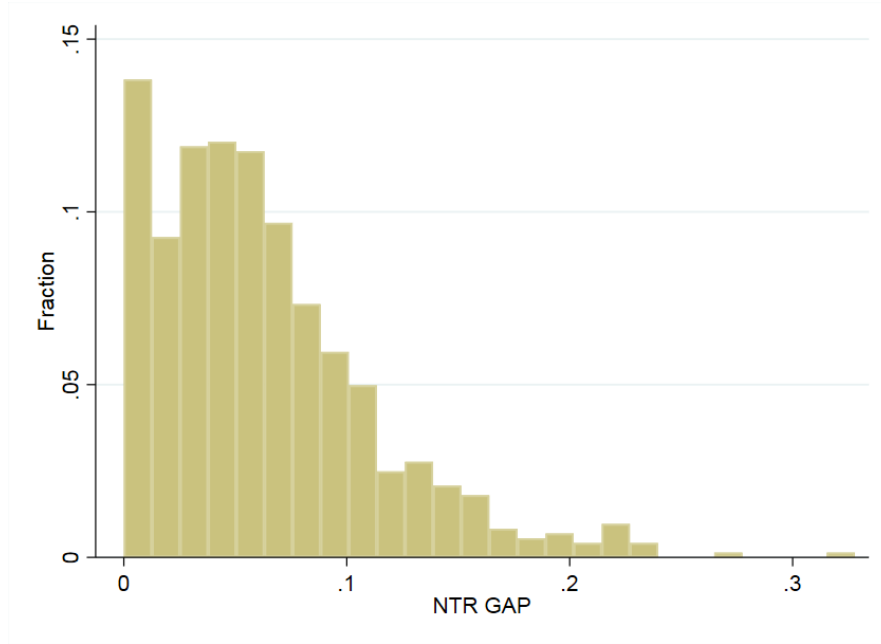
$$NTR\ Gap_j = Non\ NTR\ Tariff_j - NTR\ Tariff_j \quad (1)$$

$$NTR\ Gap_c = \sum_j \frac{L_{cj}}{L_c} * NTR\ Gap_j \quad (2)$$

- **PUMA to CZ crosswalk by Peter McHenry**

While I was exploring ways to convert county or puma into commuting zones, I noticed that one of the authors of the paper, Peter McHenry, posted crosswalk files on his personal websites. These files are directly used in converting location units for my replication.

Figure 1: Histogram of Trade Shock



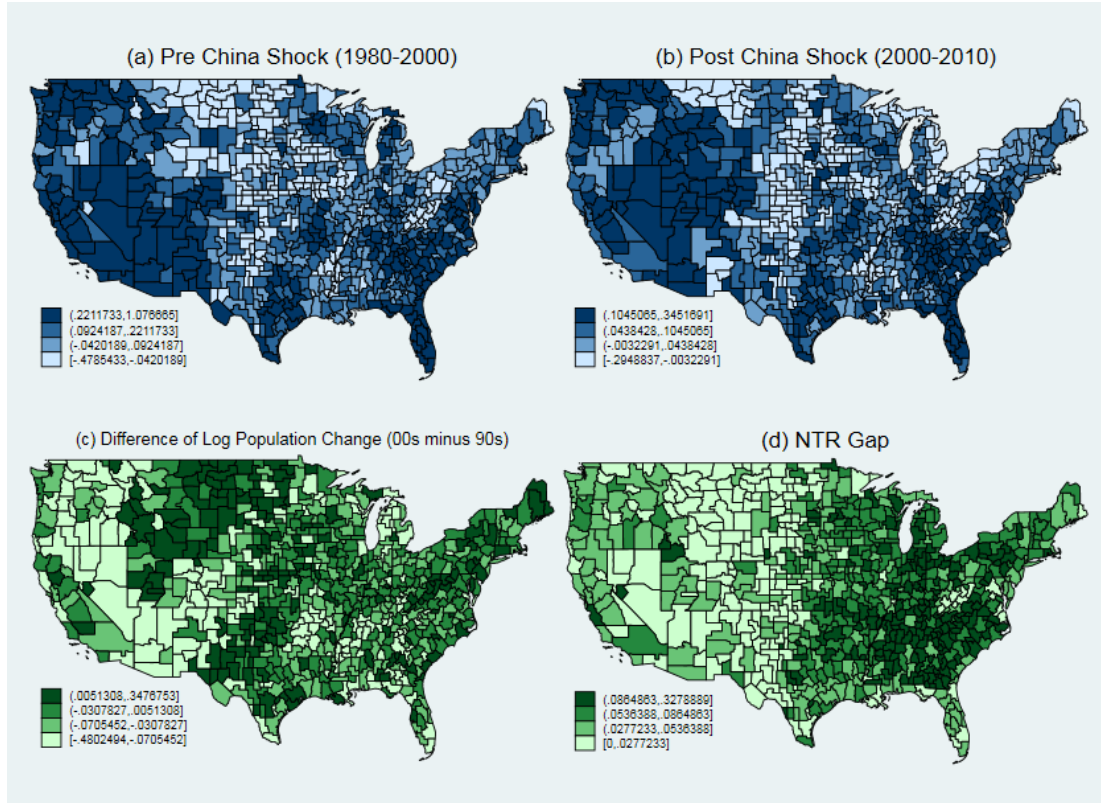
2.2. Result

Firstly, I replicate the main figures of the paper. Figure 1 is the histogram of NTR Gap, which is the key independent variable of this paper: the measure of

trade shock. Figure 1 is successfully replicated.

Figure 2 panel d is the map of NTR Gap, which is also successfully replicated. I extend the map by adding population change pre shock (panel a), population change post shock (panel b), and the difference of post and pre shock (panel c) to show spatial correlation. As it is shown, that places that are darker in panel (d), where places experienced larger import competition, are also the places that are lighter in panel (c), where population growth is lower.

Figure 2: Geography of Population Change and Trade Shock



I go on to replicate the important tables in the paper. Firstly, Table 1 and Table 2 replicate that first table in the original paper, which regress ten-year changes in Log CZ Population on import competition. Different specifications across columns are due to different levels of controls and 2 age groups: 15-64 and 15-34. I am able to obtain very similar estimate and significance, while my replication estimates are slightly smaller and slightly less significant. The R^2 in my

replication is around 0.03 less than the original paper too. Overall, the replication is reasonable.

I also replicate a different NTR Gap measure created by Kovak (2013) [5] in Table 3 and Table 4. Comparing them to the tables in the appendix of the original paper, the pattern is similar, that the replication shows slight smaller estimate and slightly less significance. This shows that the population I generated might have some difference from the population generated by the authors.

I go on to replicate the main result by dividing the population into heterogeneous groups: gender, race, education, and 12 groups by education and gender. These are my Table 5, Table 6 and Table 7. Again, slightly smaller estimates and slight lower significance are observed. However, it is worth noting that the same conclusion can be come to:

- Male population reacts to NTR Gap more than female population.
- Only asian population reacts to NTR Gap significantly. And the scale is considerably large, which is around 6 times the estimate for working age population.
- Black population actually sort themselves into high NTR Gap areas, although this estimate is not significant.
- Lower the education, higher the reaction to NTR Gap. College or more population actually has non significant response to NTR Gap, which makes sense since import competition is mainly in the low education, low wage sector.
- Younger the population, higher the reaction to NTR Gap. This is also consistent with most of the literature.

Table 1: Import Competition and 10-year Changes in Log CZ Population (aged 15-34), Census

| | (1) | (2) | (3) | (4) |
|--------------------------|---------------------------|-----------|-----------|-----------|
| | $\Delta \ln(pop_{15-64})$ | | | |
| NTR Gap | -0.186* | -0.251*** | -0.190 | -0.186 |
| | (0.098) | (0.081) | (0.143) | (0.135) |
| Lag 15-64 Change | 0.552*** | 0.545*** | 0.524*** | 0.519*** |
| | (0.037) | (0.034) | (0.035) | (0.037) |
| Hispanic Share | | 0.083* | 0.067 | 0.062 |
| | | (0.046) | (0.047) | (0.051) |
| Under 25 Share | | 0.336*** | 0.412*** | 0.422*** |
| | | (0.114) | (0.118) | (0.125) |
| Black Share | | -0.117* | -0.150*** | -0.156*** |
| | | (0.067) | (0.054) | (0.050) |
| American Indian Share | | -0.206*** | -0.108* | -0.113* |
| | | (0.056) | (0.064) | (0.064) |
| Asian Share | | 0.058 | -0.484*** | -0.479*** |
| | | (0.237) | (0.174) | (0.169) |
| Neighbor CZ Effect | | | 0.083 | 0.254 |
| | | | (0.305) | (0.288) |
| Routineness | | | 0.004** | 0.003* |
| | | | (0.002) | (0.002) |
| Offshorability | | | 0.015 | 0.017 |
| | | | (0.013) | (0.014) |
| Capital-Labor Ratio | | | 0.000 | 0.001 |
| | | | (0.013) | (0.014) |
| Skill Intensity | | | 0.096 | 0.104* |
| | | | (0.059) | (0.060) |
| Female Labor Force Share | | | -0.374 | -0.392 |
| | | | (0.293) | (0.339) |
| College Educated Share | | | 0.001 | 0.001 |
| | | | (0.001) | (0.001) |
| Debt to Income | | | | 0.018 |
| | | | | (0.031) |
| HPI Break | | | | 0.144** |
| | | | | (0.062) |
| Observations | 1444 | 1444 | 1444 | 1322 |
| R^2 | 0.678 | 0.693 | 0.708 | 0.711 |
| Region-Year FE | Y | Y | Y | Y |

Standard errors in parentheses

* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Import Competition and 10-year Changes in Log CZ Population (aged 15-34), Census

| | (1) | (2) | (3) | (4) |
|--------------------------|---------------------------|----------------------|----------------------|----------------------|
| | $\Delta \ln(pop_{15-34})$ | | | |
| NTR Gap | -0.580*** (0.126) | -0.682*** (0.107) | -0.421** (0.193) | -0.399** (0.185) |
| Lag 15-34 Change | 0.353*** (0.046) | 0.344*** (0.044) | 0.331*** (0.047) | 0.323*** (0.047) |
| Hispanic Share | | 0.087 (0.056) | 0.081 (0.062) | 0.068 (0.064) |
| Under 25 Share | | 0.261* (0.155) | 0.397** (0.172) | 0.425** (0.172) |
| Black Share | | -0.157** (0.080) | -0.171*** (0.065) | -0.181*** (0.062) |
| American Indian Share | | -0.189** (0.086) | -0.089 (0.095) | -0.084 (0.091) |
| Asian Share | | -0.228 (0.307) | -0.827*** (0.243) | -0.782*** (0.238) |
| Neighbor CZ Effect | | | -0.319 (0.422) | -0.087 (0.388) |
| Routineness | | | 0.003 (0.002) | 0.002 (0.002) |
| Offshorability | | | 0.001 (0.019) | 0.006 (0.020) |
| Capital-Labor Ratio | | | -0.007 (0.017) | -0.007 (0.018) |
| Skill Intensity | | | 0.169** (0.086) | 0.178** (0.089) |
| Female Labor Force Share | | | -0.099 (0.399) | -0.117 (0.444) |
| College Educated Share | | | 0.002 (0.001) | 0.002 (0.001) |
| Debt to Income | | | | 0.052 (0.035) |
| HPI Break | | | | 0.224** (0.088) |
| Observations | 1444 | 1444 | 1444 | 1322 |
| R^2 | 0.607 | 0.619 | 0.634 | 0.641 |
| Region-Year FE | Y | Y | Y | Y |

Standard errors in parentheses

* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Import Competition Following Kovak (2013) (15-64)

| | (1) | (2) | (3) | (4) |
|---|---------------------------|---------------------|---------------------|---------------------|
| | $\Delta \ln(pop_{15-64})$ | | | |
| (mean) ntr_kovak | -0.083 (0.066) | -0.123** (0.056) | -0.055 (0.092) | -0.056 (0.089) |
| Lag 15-64 Change | 0.553*** (0.037) | 0.545*** (0.034) | 0.526*** (0.035) | 0.520*** (0.038) |
| Observations | 1444 | 1444 | 1444 | 1322 |
| R^2 | 0.677 | 0.692 | 0.707 | 0.710 |
| Region-Year FE | Y | Y | Y | Y |
| Standard errors in parentheses | | | | |
| * $p < .1$, ** $p < 0.05$, *** $p < 0.01$ | | | | |

Table 4: Import Competition Following Kovak (2013) (15-34)

| | (1) | (2) | (3) | (4) |
|--|---------------------------|----------------------|---------------------|---------------------|
| | $\Delta \ln(pop_{15-34})$ | | | |
| (mean) ntr_kovak | -0.297*** (0.082) | -0.367*** (0.070) | -0.188 (0.118) | -0.183 (0.115) |
| Lag 15-34 Change | 0.353*** (0.046) | 0.344*** (0.044) | 0.333*** (0.048) | 0.325*** (0.048) |
| Observations | 1444 | 1444 | 1444 | 1322 |
| R^2 | 0.604 | 0.618 | 0.633 | 0.640 |
| Region-Year FE | Y | Y | Y | Y |
| Standard errors in parentheses | | | | |
| * $p \leq .1$, ** $p \leq 0.05$, *** $p \leq 0.01$ | | | | |

Table 5: Heterogeneous Groups (Gender and Race)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|----------------|----------|----------|----------|----------|----------|-----------|
| | Male | Female | Hispanic | Black | White | Asian |
| ntr_gap_post | -0.220* | -0.103 | -0.262 | 0.276 | -0.093 | -1.356*** |
| | (0.132) | (0.129) | (0.293) | (0.504) | (0.132) | (0.467) |
| L.lcmale | 0.549*** | | | | | |
| | (0.033) | | | | | |
| L.lcfemale | | 0.657*** | | | | |
| | | (0.031) | | | | |
| L.lchispanic | | | 0.417*** | | | |
| | | | (0.056) | | | |
| L.lcblack | | | | 0.349*** | | |
| | | | | (0.066) | | |
| L.lcwhite | | | | | 0.538*** | |
| | | | | | (0.030) | |
| L.lcasian | | | | | | 0.097* |
| | | | | | | (0.055) |
| Observations | 1444 | 1444 | 1444 | 1428 | 1444 | 1444 |
| R ² | 0.720 | 0.772 | 0.495 | 0.403 | 0.733 | 0.360 |

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6: Heterogeneous Groups (Education)

| | (1) | (2) | (3) |
|----------------|--------------|-------------------|--------------|
| | Less than HS | HS - Some College | College More |
| ntr_gap_post | -0.308*** | -0.179** | -0.175 |
| | (0.117) | (0.091) | (0.174) |
| L.lcunderhs | 0.620*** | | |
| | (0.045) | | |
| L.lchs_scol | | 0.240*** | |
| | | (0.024) | |
| L.lccol_mor | | | 0.281*** |
| | | | (0.029) |
| Observations | 1444 | 1444 | 1444 |
| R ² | 0.740 | 0.665 | 0.504 |

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 7: 12 Education * Age sub-groups

| | (1) | (2) | (3) |
|-------------------|----------------------|----------------------|----------------------|
| | UnderHS | HS - Some College | College More |
| ntr_gap_post_2534 | -1.035*** (0.278) | -0.693*** (0.157) | -1.032*** (0.343) |
| R^2 | 0.4356 | 0.7259 | 0.3256 |
| ntr_gap_post_3544 | -0.970*** (0.285) | -0.108 (0.158) | -0.186 (0.183) |
| R^2 | 0.8207 | 0.8846 | 0.2840 |
| ntr_gap_post_4554 | -0.381* (0.209) | -0.053 (0.151) | 0.233 (0.215) |
| R^2 | 0.5957 | 0.7271 | 0.8823 |
| ntr_gap_post_5564 | 0.032 (0.225) | -0.279** (0.111) | -0.435** (0.176) |
| R^2 | 0.6485 | 0.7744 | 0.6065 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

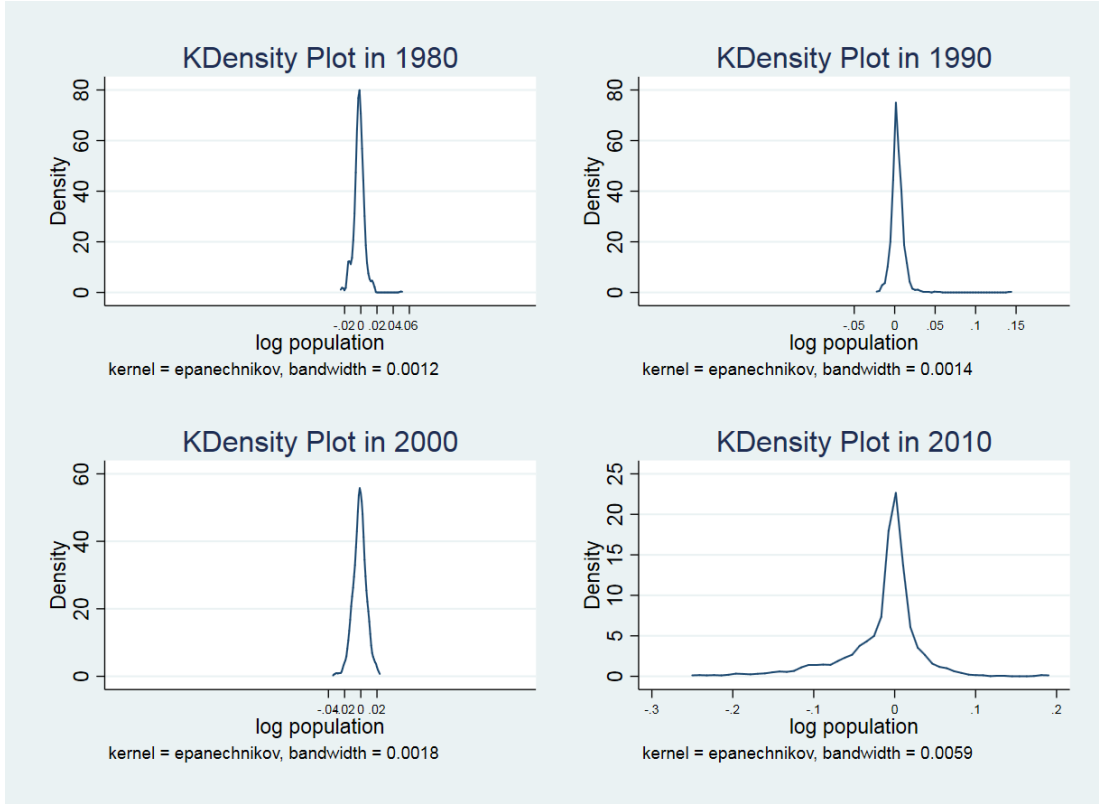
3. Discussion of Difference between Replication and Original Paper

This section discuss potential reasons why replication estimate differs from the original paper.

What is the reason that my replication always gets smaller estimate and less significance? I obtain the replication package published online, which can be found on the website of the journal. I take log difference of the population data used in the paper, and draws kernel density graph for each year, shown in Figure 3.

It is clear that while two data sets are almost the same, some difference is observed and the difference in 2010 is considerable. In my opinion, the difference in 2010 should be producing the main difference between replication and original paper. While I have followed most of the steps described in the paper, the author did not explain why they use ACS 3-year survey in 2011 to create population measure in 2010, instead of using 2010 census directly. Choosing between data sets should be done in cautious, while the amount of difference should be addressed in a better way.

Figure 3: Different Population Distribution (Replication and Original)



4. Extension

This section provides an extension of the paper by taking city size into account. Large cities, with its thick labor market, act as a role of "sponge" to mitigate the import competition. Displaced workers can find jobs in larger cities without the need to migrate. This provides the motivation for my extension exercise.

To conduct this extension, I start with dividing my sample into two groups: large commuting zones and small commuting zones. The size of commuting zone I use is the commuting zone size in 1990, which is before the China shock kicked in. I include all the control variables, same as the preferred specification from the paper.

Table 8 shows the result of whole sample, small commuting zones subsample, and large commuting zone subsample. Two age groups are examined: 15-64 working population and 15-34 young labor force. There is little evidence that NTR Gap has different effect on large and small commuting zones.

To estimate this idea more precisely, I introduce an interaction term to estimate how the effect of NTR Gap changes with size of commuting zone.

$$\begin{aligned} \ln(population_{ct}) = & \beta_0 + \beta_1 NTRGap_c * Post2001_t + \beta_2 X_c * Post2001_t \\ & + \beta_3 \Delta \ln(population_{c,t-10}) + \beta_4 CZsize_{1990} \\ & + \beta_5 NTRGap_c * Post2001_t * CZsize_{1990} + \delta_{rt} + \epsilon_{ct} \end{aligned} \quad (3)$$

Table 9 shows the result for this interaction model. For population aged 15-64, bigger cities react less to NTR Gap, as the interaction term is positive and significant at 1% level. For population aged 15-34, the interaction term estimate becomes smaller and not significant, while remains positive. I argue that commuting zone size might be an important factor to consider because displaced workers might be less likely to move if the market is thick.

Table 8: Log CZ Population Changes (Whole Sample, Large, Small CZs)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---------|---------------------------|---------------------|---------------------|---------------------------|----------------------|----------------------|
| | $\Delta \ln(pop_{15-64})$ | Large CZs | Small CZs | $\Delta \ln(pop_{15-34})$ | Large CZs | Small CZs |
| NTR Gap | -0.264** (0.114) | -0.263** (0.115) | -0.248** (0.112) | -0.503*** (0.170) | -0.505*** (0.168) | -0.494*** (0.159) |
| Lag | 0.529*** (0.034) | 0.530*** (0.037) | 0.569*** (0.044) | | | |
| Lag | | | | 0.338*** (0.047) | 0.324*** (0.052) | 0.344*** (0.050) |
| Obs | 1322 | 961 | 961 | 1322 | 961 | 961 |
| R^2 | 0.743 | 0.758 | 0.768 | 0.653 | 0.669 | 0.637 |
| RY FE | Y | Y | Y | Y | Y | Y |

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

5. Conclusion

This paper replicates the main result of "Import Competition and Internal Migration" [1], authored by Greenland, Andrew, John Lopresti, and Peter McHenry and published on *Review of Economics and Statistics* in 2019. The replication result is very close to the original work while slight downward estimates are observed.

Table 9: Interaction Model Estimates

| | (1) | (2) | (3) | (4) |
|------------------|---------------------------|----------------------|---------------------------|---------------------|
| | $\Delta \ln(pop_{15-64})$ | | $\Delta \ln(pop_{15-34})$ | |
| NTR Gap | -3.343*** (0.712) | -2.147*** (0.568) | -2.557*** (0.969) | -1.398* (0.750) |
| Lag 15-64 Change | 0.591*** (0.029) | 0.563*** (0.032) | | |
| lpop90 | -0.007** (0.003) | -0.007* (0.003) | -0.004 (0.004) | -0.003 (0.005) |
| NTR * lpop90 | 0.244*** (0.056) | 0.146*** (0.047) | 0.151** (0.075) | 0.070 (0.060) |
| Lag 15-34 Change | | | 0.385*** (0.044) | 0.352*** (0.045) |
| Controls | No | Yes | No | Yes |
| Observations | 1444 | 1322 | 1444 | 1322 |
| R^2 | 0.730 | 0.750 | 0.625 | 0.654 |

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

This paper also has an extension which commuting zone size is taken into account. It is shown that for larger cities, the negative effect of import competition on population growth will be smaller.

References

- [1] A. Greenland, J. Lopresti, P. McHenry, *Import competition and internal migration*, *Review of Economics and Statistics* 101 (1) (2019) 44–59.
- [2] J. R. Pierce, P. K. Schott, *The surprisingly swift decline of us manufacturing employment*, *American Economic Review* 106 (7) (2016) 1632–62.
- [3] C. M. Tolbert, M. Sizer, *Us commuting zones and labor market areas: A 1990 update*, *Tech. rep.* (1996).
- [4] D. H. Autor, D. Dorn, G. H. Hanson, *The china shock: Learning from labor-market adjustment to large changes in trade*, *Annual Review of Economics* 8 (2016) 205–240.
- [5] B. K. Kovak, *Regional effects of trade reform: What is the correct measure of liberalization?*, *American Economic Review* 103 (5) (2013) 1960–76.