Escaping Air Pollution: Immigrants, Students, and Spillover Effects on Property Prices Abroad [*]

Yuk Ying Chang[#]

Sudipto Dasgupta[##]

Abstract. We construct a time-series of news coverage about air pollution in China for the period 1977-2019. Our measure of abnormal news coverage of China's air pollution (ANC) is uncorrelated with growth in economic activity or cyclical components of such activity, but strongly correlated with weather-related and atmospheric conditions known to cause air pollution. ANC is associated with more capital flight from China. Focusing on the U.S. as a destination country, we find that ANC is associated with more Chinese citizens emigrating to U.S. regions with stronger ethnic links to China, and more international students enrolling in U.S. institutions with stronger Chinese student links. U.S. regions with stronger ethnic or educational ties to China experience higher property price growth when ANC is higher. Our study suggests that perception of local environmental risk can have major consequences for the cross-border reallocation of capital and labor.

JEL: Q5, G1, R3, I1, I2, J6

[*] We thank Marcin Kacperczyk (the Editor) and an anonymous referee for comments that greatly improved the paper. We also thank Darwin Choi, Zhenyu Gao, and Wenxi Jiang for helpful comments. Errors are our own responsibility. Chang thanks Massey University's Business Impactful Research Fund for funding.

[#] Massey Business School, Massey University. Email: y.chang@massey.ac.nz.

[##] Chinese University of Hong Kong, ABFER, CEPR, and ECGI. Email: s.dasgupta@cuhk.edu.hk.

Exposure to climate and environmental risk is likely to cause reallocation of labor and capital at unprecedented levels in the coming decades.[1] Very recently, research is emerging documenting such reallocation within specific countries where reliable data on migration and labor skills are available, and quantifying the productivity consequences of such reallocation (Khanna et al., 2021; Albert et al., 2021).[2] In this paper, we examine cross-border migration and capital flight from China, a country regarded as a major sources of toxic and greenhouse gas emissions, and show that such a process has been going on at least for the last two decades. We show that spikes in news about China's air pollution are associated with more cross-border migration and capital flight from China, as well as with significantly more residential property price appreciation in U.S. regions with stronger ethnic links to China. Possibly unique to China, the process has gathered momentum in recent decades due to the willingness of wealthier Chinese parents to send their children abroad in response to worsening air pollution in China, and to invest in foreign residential property in the metropolitan areas where their children study.

Based on keyword search in news articles covered in *Factiva* that mention "air pollution" and "China", we construct an annual time series of the number of such news items (scaled by the number of news items that mention "China"), starting in the late 1970s. This enables us to work with a significantly longer time series than alternatives such a web-based search indices.[3] We use this series – henceforth labelled *APC* (Air Pollution News, China) – to fit an AR(1) model, and the resultant residual (labelled *RAPC*) is regarded as innovation in air pollution news. While all news in principle is "new", some news simply refreshes old news. The

[1] See, for example, "The Great Climate Migration" by Abrahm Lustgarten at https://www.nytimes.com/interactive/2020/07/23/magazine/climate-migration.html.
[2] A major challenge for research on this issue is that data that track individual's locations over time is hard to come by, making it difficult to link the location of origin and destination. In China, the Household Registration or the *Hukou* system provides information on every citizen's registered residence. Recent papers have combined *Hukou* status with the latest Population Census Data, as well as other longitudinal data with history of location changes (Khanna et al., 2021).
[3] Engle et al. (2020) construct a "climate news" index based on coverage of climate change by The Wall Street Journal.

underlying assumption is that innovation or spikes in coverage of air pollution news not only convey new information (e.g., scientific studies about the harmful effects of air pollution attracting news headlines) that increase people's awareness of air pollution, but also reflect severe air pollution episodes that people experience in real time. We validate this measure in a number of ways – for example, both *APC* and *RAPC* correlate very highly with the corresponding measures based on annual averages of daily air quality values (*AQI*) of major Chinese cities available for a much shorter period. While we present results mainly based on news articles in English, we also construct a corresponding measure based on Factiva articles in simplified Chinese, available for the post-1995 period. The two measures are strongly correlated and produce similar results. For several of our tests, we also construct "local" versions of our air pollution news coverage measures based on the joint occurrences of air pollution-related keywords and the name of a particular city or region.

A key challenge in interpreting associations between *RAPC* and individual decisions like cross-border moves is the possibility that air pollution is correlated with economic activity, and thus also with people's income or wealth, which could affect these decisions. We address this issue in several ways. First, our identification strategy in most of our tests relies on the strength of historical ties of a specific U.S. region to China based on early Chinese settlement in the U.S., and we posit that regions with stronger historical ethnic ties would attract more Chinese immigrants and inflow of Chinese capital when there is more concern about air pollution in China. Ethnic links formed a long time ago are less likely to reflect current economic linkages; however, to control for such a possibility, we allow for variables capturing China's economic activity to have different effects on our regional outcome variables of interest depending on the strength of ethnic links to a region. We follow a similar approach vis-à-vis the moves of Chinese international students, although here we have to rely on more recent data to identify regions with stronger education links.

Second, we show that our measure of innovation in air pollution news is uncorrelated with growth in economic variables such as China's gross domestic product (GDP) per capita or trade per capita. When we decompose these series using the Hodrick-Prescott filter (which is also applicable to non-stationary time series) into a "cyclical" component and a "trend" component, we find that *RAPC* is uncorrelated with the cyclical components of China's economic activity variables. Based on data from the National Aeronautics and Space Administration (NASA), we construct a "temperature inversion index" (*TI*) for three major Chinese cities, and a "haze weather index" (*HWI*) (Cai et al., 2017) for Beijing, and show that the local version of *RAPC* is explained by extreme values of *TI* and *HWI*, which reflect atmospheric and weather conditions unrelated to economic activity. Other reasons for variation in *RAPC* could be due to environmental policy being enforced intermittently (Jin et al., 2016), rather than changes in aggregate economic activity related to income or wealth, for which we fail to find much evidence.

Unfortunately, to examine the impact of innovation in air pollution news on capital outflow from China, we cannot sharpen our identification, as discussed above, based on a single time-series. Nonetheless, after controlling for other variables that could be associated with capital outflow, we find that the contemporaneous *RAPC* is significantly and positively associated with outflow of capital from China.

We then consider cross-border migration response to air pollution news. We examine whether Chinese immigration increases more in major Metropolitan Statistical Areas (MSAs) that had a higher proportion of Chinese population in 1870 when innovation in air pollution news is higher. We find this is indeed the case, with larger coefficients for one-and two-period lags of *RAPC* than for the contemporaneous *RAPC*, although all are significant. Since the immigration process takes time, this pattern of coefficients is expected. We further explore whether language ties play a role. We find that emigration to MSAs with relatively more Cantonese speakers is more sensitive to innovation in air pollution news associated with Guangdong province in China, where Cantonese is the primary dialect.

We next explore whether Chinese student movements to the U.S. are similarly affected by innovation in air pollution news. We construct a weighted average *RAPC* for each post-secondary U.S. institution located in a particular MSA, based on (a) local versions of RAPCs of different Chinese cities, and (b) weights reflecting the importance of different Chinese cities as a source of Chinese student to the MSA. We find that the weighted *RAPC* and its lags are all highly significant in explaining institution-level international student enrolment. However, when the weights are assigned randomly, the randomly weighted *RAPC* and its lags are not significant.

We provide additional evidence suggesting that cross-border moves by Chinese students is related to local air pollution news, based on data on the number of students registering for the college entrance examination (*Gaokao*) in a region. In China, due to the tight enforcement of the *Hukou* residency registration system, it is generally difficult for students to take the examination in another region where they do not have *Hukou*. We find that, controlling for the number of senior high school graduates in a region, the number of students registering for the entrance examination is significantly lower when the deviation of local *RAPC* from China's overall *RAPC* is higher. This very likely reflects graduates leaving China to pursue post-secondary education in other countries.

Our results so far suggest that innovation in news about China's air pollution is associated with outflows of both people and capital from China. Capital outflows could precede or follow the movement of people, or simply occur in response to a loss of trust in the government's ability to deal with environmental problems. We hypothesize that a significant part of such capital ends up being invested in residential properties abroad, especially in regions where there are strong ethnic or educational links to China. Accordingly, we examine whether residential property prices increase more in U.S. counties with stronger links to China when the innovation in China-related air pollution news is higher. We find robust evidence showing that this is the case.

Our study makes several contributions. First, climate and environmental risks are correlated – not only do the same activities (such as burning fossil fuel) emit greenhouse gases and pollutants, there is also a feedback loop whereby air pollution and global warming reinforce each other.[4] Thus, beyond the adverse consequences of air pollution, our results also indicate how salient Chinese citizens consider environmental and climate change to be, about which there is conflicting evidence.[5] This is important because of China's position as the second most important emitter of greenhouse gases. Our results suggest that the Chinese citizens are paying attention to climate and environmental risks and are taking costly decisions to avoid or mitigate these risks to their personal lives, or those of their children, over the last two decades.

Second, since it is less costly for labor and capital to move within-country than across borders, our results support recent evidence (Khanna et al., 2021) that regions with significant air pollution problems are likely to lose the more mobile skilled workers, and incur substantial economic costs.

[4] Carbon dioxide and Methane raise the earth's temperature, which worsens smog and increases the production of allergenic air pollutants. Emissions of pollutants into the air can result in changes to the climate. Ozone in the atmosphere warms the climate, while different components of particulate matter (PM) can have either warming or cooling effects on the climate.

[5] Wang and Zhou (2020) argue that available survey research suggests that Chinese citizens surveyed in the last decade have a relatively high awareness of climate change, and the majority understand that the issue is caused by human activity. However, a recent worldwide poll (the Lloyd's Register Foundation World Risk Poll based on interviews with over 150,000 people worldwide conducted by Gallup in 2019) found that 30% of those surveyed in China did not express an opinion on the question of the seriousness of climate change in the next 20 years – far higher than any other country —and only 23% considered it a very serious threat.

Finally, we contribute to recent research documenting that beliefs about long-term environmental or climate risk have significant effects on asset prices. Much of the current financial research on the salience of climate risk has examined whether extreme climate events impact asset prices. For example, Choi et al. (2020) find that when local temperatures are abnormally high, online searches related to climate change increase and local investors sell more carbon-intensive stocks. Engle et al. (2020) document that stocks of firms with lower exposure to regulatory climate risk have higher returns when there is negative news about the future path of climate change. Giglio et al. (2021) construct a measure of attention to climate risk in the housing market, and find that the premiums enjoyed by properties in flood zone areas (likely due to their greater amenity value) get compressed in periods when the attention to climate risk is higher.[6] Like these papers, we also examine extreme episodes as instances when beliefs are likely to change. Such beliefs not only concern the slow-moving processes of environmental and climate change, but also trust in the government's ability to address these issues. Focusing on residential property as an asset class, we show that the impact of such belief revisions in a major economy like China can be significant enough to affect asset prices abroad.[7]

1. Hypothesis, Identification Strategy, and Data

In this section, we introduce our hypotheses and identification strategy, and describe our data.

1.1 Hypothesis and Identification

[6] Ortega and Taspinar (2018), Gibson and Mullins (2020), and Eichholtz et al. (2019) study the effect of hurricanes in the New York area and find that the valuations of exposed properties were adversely affected even when they did not suffer damage, suggesting increasing salience of flood risk following these events.

[7] We focus on residential property prices as housing is a major asset for households, and foreign residential property is anecdotally known to be a popular investment asset for the Chinese. China overtook Canada as the top foreign country investing in U.S. residential real estate in 2014-2015. Anecdotal evidence, reports in the popular press, global investment outlook blogs of real estate companies, and industry reports indicate that one of the important destinations of capital flight from China is foreign housing markets.

Our main hypothesis is that when Chinese citizens' awareness and concern about (the adverse effects of) air pollution increases, they are more likely to consider relocation to other countries with better air quality; wealthier Chinese parents are also more likely to send children abroad to study. Moreover, actual or intended cross-border moves will trigger capital outflow from China. Capital outflow could also be triggered if concern about air pollution is related to scepticism about the government's ability to deal with environmental and related challenges and erodes trust in government. Anecdotal evidence presented in section 2 shows that adverse air pollution episodes or news about the harmful effects of air pollution immediately trigger interest among Chinese citizens in emigration and in foreign real estate, especially in regions with historically strong ties to China. Much of the capital flowing out of China for these reasons is likely to be invested in real estate abroad.[8] Since it may be possible to move capital out of China relatively quickly prior to emigration, we would expect both contemporaneous and lagged effects of innovation in air pollution news in China on housing prices in destination regions.

We exploit two main sources of variation to identify the causal effects of news about air pollution in China on cross-border moves from China to U.S. regions and the impact on U.S. property prices: (a) time series variation in "abnormal" news coverage (ANC) about air pollution in China, and (b) historical links of Chinese immigrants to different U.S. regions. In our empirical tests, we examine whether higher ANC is associated with more emigration to, or higher impact on property prices in, U.S. regions with stronger historical links to China. The identifying assumptions are that (i) ANC is not positively corelated with other variables, such as income or wealth, that make is easier for the Chinese to consider cross-border moves, and

[8] A recent survey conducted jointly by the Shanghai Advanced Institute of Finance of Shanghai Jiao Tong University and Charles Schwab & Co., Inc. finds that real estate is the most preferred asset class for Chinese citizens. The value of their assets in real estate is approximately 4 times that in equities and 16.5 times that in bonds (https://en.cafr.cn/Research/Details.aspx?menuid=16&id=132, retrieved on October 17, 2022).

(ii) ANC is not positively correlated with economic or other linkages between China and U.S. regions with stronger historical ethnic ties to China. In other words, these latter regions are likely destinations of Chinese citizens wishing to escape air pollution in China because they are more familiar to the Chinese rather than because they offer more opportunities precisely when the underlying reasons for more pollution news coverage strengthen.[9]

Below, we discuss our attempts to substantiate these assumptions. To do so, we first describe how we construct the time series of abnormal news coverage and what factors are likely to affect it, followed by a discussion of the cross-sectional variation in ethnic ties to U.S. regions.

1.1.1 Time-Series Variation: Abnormal Air Pollution News Coverage

A reasonable starting point for a study of the effect of air pollution in China on cross-border migration, and the associated impact on foreign housing markets, would be data on air quality in China. However, a sufficiently long time series of an air quality index is not available. Air quality data for China (at the city level) is only available for a relatively short period (from 2000).

Ginsberg et al. (2009) show that flu-related keyword search in Google can estimate influenza epidemics across different geographical regions. The Baidu Index (akin to Google Trend or Google SVI (Search Volume Index)) may capture time-varying broad Chinese interest and attention to a particular topic/issue, including air pollution. However, the index is also only available from 2012. Absent a sufficiently long time-series of the Baidu Index, we use Dow Jones & Company's *Factiva* international news coverage, for which data is available since the

[9] Social and ethnic links are likely to be relevant for Chinese citizen's regional preference to live or invest in residential property in the U.S., obtain information about job opportunities, and in mitigating information asymmetries, e.g., general information about the local property market, or finding realtors and lawyers who speak the language of the property buyers and understand their requirements. Moreover, in the case of investing in residential property, socially connected individuals in the U.S. can also perform an important monitoring role, essentially "looking after" the property or screening and monitoring tenants if the property is rented out. Ethnicity-specialized real estate agents can also facilitate sales of residential units at a later point of time.

late 1970s, to construct a proxy for public attention to the issue of air pollution in China. While search indices such as the Google SVI reflect public interest and the demand for information on an issue, news coverage of issues such as air pollution can be both a source of news and also reflect the salience of issues or events to a particular population.

We construct our main explanatory variable as follows. We use the keywords “air pollution” and “China” to conduct Factiva search and obtain the annual number of news items associated with “air pollution” and “China” (*X1*). Next, we use the keyword “China” to conduct another Factiva search to obtain the annual number of news items associated with “China” (*X2*). The ratio of *X1* to *X2*, which we denote as *APC*, then measures the proportion of news about China that involves air pollution in a given year.

While all news, in principle, is “new”, some news refreshes old news, while other news covers unanticipated events. We focus on Chinese citizens’ response to the latter type of news, i.e., *unpredictable* news coverage about air pollution. We model *APC* as an AR(1) process, and formally confirm that we can reject the null of non-stationarity. The residual from a regression of *APC* on its lagged value, (hereinafter denoted by *RAPC* and referred to as “surprise” or “innovation” in air pollution news), is our key explanatory variable. Since response can take time, our specifications include *RAPC* and its lagged values *L1.RAPC* and *L2.RAPC*. We next verify that *APC* and *RAPC* are valid measures of attention to air quality in various ways.

Correlation with Air Quality and Other Indices

First, to show that these measures reflect air quality of major cities in China, we plot the time series of *APC* and air quality in Figure 1 and their AR(1) residuals in Figure 2, on a standardised scale.[10] To this end, we obtain daily air quality value (*AQI*) of Beijing, Shanghai,

[10] The standardization is done by subtracting from the raw value its mean and then dividing by its standard deviation.

Guangzhou and Shenzhen for the period 2000-2015,[11] take the daily average in a year for each of these four major cities, and then take the average of these cities to obtain the yearly time-series of *AQI*.[12] We also obtain the residual of air quality value (residual *AQI*) from an autoregressive model of order 1. Figure 1 shows that *APC* and *AQI* move similarly. Figure 2 reveals that their residuals move extremely closely together over time. There is no obvious trend before 2012. A spike of both residuals occurs in 2013, which is almost four times the second peak. Based on the Chinese national surveys of the Pew Research Center, we find that 2013 is also the year for which most Chinese respondents (83%) consider air pollution as "a very big problem" or "a moderately big problem".[13]

Second, we construct an *APC* series based on Factiva news articles in simplified Chinese.[14] This is relevant because articles in Chinese are more likely to be in the attention of Chinese citizens. This time series is much shorter. However, the correlation between the one on based on English and the one based on simplified Chinese is 0.85, and the correlation of the *RAPC* corresponding to the two series is 0.88. We prefer to work with the English series because (a) it is a longer time series, (b) the English news media is likely to pick up events related to Chinese air pollution that are more significant, and (iii) for the period over which we can construct a time series of news in simplified Chinese (from around the middle of 1990s), information on news covered in the English news media is particularly likely to come to the attention of Chinese citizens through social media.

[11] The primary source of this data is the Ministry of Environmental Protection Information Centre (now the Ministry of Ecology and Environment) of the People's Republic of China. The data is sourced from https://github.com/mingcheng/AQI, retrieved on 26 April 2021.
[12] The air quality data from China is highly correlated with the PM2.5 data of the corresponding cities from the U.S. Department of State, with correlation coefficients between 0.68 and 0.74.
[13] Our results remain similar when we exclude observations that involve the 2013 *RAPC* as a contemporaneous or lagged explanatory variable. The results are not tabulated, but available from the authors upon request.
[14] These news sources include official news agencies in China as well as news outlets based outside China.

Third, we broaden the set of keywords to construct 4 additional versions of *APC*, based on Factiva search. The broadest set, with the highest *APC*, is "(air pollutants OR air pollution OR air quality OR toxic air OR carbon dioxide emissions OR smog OR soot OR toxic emissions OR particulate OR haze OR clean air) AND China". The second highest *APC* excludes "soot" from the broadest set. The third highest *APC* excludes "haze" from the broadest set. The last *APC* version excludes both "soot" and "haze" from the broadest set. Figure 3 shows fairly parallel time trends of different versions of *APC* although there is a stronger upward trend for the four broader *APC*s.

Fourth, we verify our method beyond the context of China by examining how Google SVI regarding air pollution changes as the Factiva count associated with air pollution changes. To do so, we first calculate the proportion of all Factiva items associated with "air pollution", denoted as *S1*. Second, we use the keyword 'air pollution" to obtain Google SVI, denoted *S2*. To mitigate the concern that there may be a common time trend, we consider the first difference of both *S1* and *S2*. We find that their correlation is 0.44. This suggests that *APC* plausibly triggers and/or tracks attention to air pollution.

Is RAPC Related to Economic Activity?

Air pollution can increase with economic activity, which in turn is associated with more income or wealth, and could affect people's desire and ability to emigrate and invest in foreign assets such as residential property. It is therefore important to address what drives variations in innovations to air pollution news, or *RAPC*. We first check whether *RAPC* is correlated with changes in the level of economic activity, such as growth of GDP per capita, growth of trade per capita, growth of CO_2 emissions per capita, and growth of fossil fuel consumption per capita. We also check whether *RAPC* is correlated with transitory components of economic activity, by decomposing each of the time series of *APC*, GDP per capita, trade per capita, CO_2 emissions per capita and fossil fuel consumption per capita into a cyclical and a trend component using the Hodrick-Prescott filter, which can be applied to both non-stationary and stationary time series. The cyclical components capture temporary deviations from the trend. If economic activity is a main driver of innovation in air pollution news, we would expect *RAPC* and the cyclical component of *APC* to be positively correlated with the growth rates of the economic variables, as well as their cyclical components. Table 1 shows that while *RAPC* and the cyclical component of *APC* are highly correlated, neither is significantly positively correlated with the growth rates of any of the economic related variables or with any of the other cyclical components. Therefore, we do not find any evidence that *RAPC* captures fluctuations in economic activity.

What then drives *RAPC*? A proximate cause of variation in the innovation in air pollution news is clearly variation in unexpected changes in air pollution itself, especially in major cities in China. Air pollution is the contamination of the atmosphere by gases, liquids and solids. The major pollutants are Ozone (O_3), particulate matter, Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Sulphur Dioxide (SO_2), and toxic compounds such as Lead. While emissions of chemicals could be related to economic activity, "changes in emissions do not necessarily translate into changes in pollution concentration at the surface due to other factor such as atmospheric processes and meteorology".[15] In particular, temperature or thermal inversions, wind directions, and ridges of high pressure are recognized to be the most common reasons for abnormal build-up of air pollution, which is also true for the major cities in China.[16]

We empirically test whether weather conditions play a significant role in explaining *RAPC* for three major regions in China (Beijing, Shanghai, and Guangdong). In particular, we

[15] See Gupta and Follette-Cook, September 2021.

[16] Thermal or temperature inversions occur when the normal heat gradient of the atmosphere is reversed. Typically, air is warmer near the ground than in upper levels of the atmosphere. When this gets reversed, cold surface air gets trapped by warmer layers above. See Malek et al. (2006) for a study on the role of thermal inversion in one of the worst air pollution episodes nationally in Logan, Cache Valley, Utah.

collect data at the 6-hour frequency for estimating thermal inversion (*TI*) from the National Aeronautics and Space Administration (NASA) website over the period 1980-2020, and compute a daily average.[17] We create two regional independent variables, *TI_TOP_⅓* and *TI_TOP_¼* to indicate years that are in the top ⅓ and top ¼, respectively, in terms of the proportion of days in the year on which the region's daily *TI* was in the top 20% of the distribution of daily *TI* for that region. We then regress regional *RAPC* on either of these indicator variables, while controlling for regional GDP per capita growth, and trade per capital growth for China, as well as regional fixed effects. Our dependent variable (regional *RAPC*) is constructed in the same way as China's *RAPC,* except that we replace "China" by the name of the region as a search term. We construct both an English language regional *RAPC* as well as a regional *RAPC* using simplified Chinese – the former gives us a longer time series, while the latter may capture more comprehensively local air pollution issues. Results reported in columns (1) and (2) of Table 2 show that both indicator variables for *TI* are highly significant in explaining *RAPC,* but none of the economic variables are.

In addition, for Beijing, we follow Cai et al. (2017) to create a "Haze Weather Index (*HWI*)",[18] which is based on an atmospheric vertical temperature profile, tropospheric meridional flow, and mid-tropospheric zonal flows.[19] We create indicator variables similar to

[17] We follow Chen et al. (2017) to estimate thermal inversions. Specifically, we use M2I6NPANA version 5.12.4 of the MERRA-2, which records the 6-hour air temperature at 42 layers from surface to 36,000 meters by grid (0.5 degree×0.625 degree). For each region, we choose the set of grids that approximately cover the geographical area of the region. Within each 6-hour period, we calculate the temperature difference by subtracting the average temperature of all grids in the first layers (110 meters) from the average temperature of all grids in the second layer (320 meters). If the temperature difference is positive, there is a thermal inversion and the difference measures the degree of the inversion. If the difference is negative, there is no thermal inversion and we replace the negative difference by zero. We then average the inversion degree across four 6-hour lapses daily for each region.

[18] The methodology of Cai et al. (2017) is specific to Beijing.

[19] The temperature profile is the difference in temperature (ΔT) between the near-surface (850 hPa, over the area of 32.5–45° N, 112.5–132.5° E) and the upper atmosphere (250 hPa, over the area of 37.5–45° N, 122.5–137.5° E). The mid-tropospheric flow is measured by a latitudinal difference in 500 hPa zonal winds (*U500*) by subtracting south of Beijing (27.5–37.5° N, 110–137.5° E) from north of Beijing (42.5–52.5° N, 110–137.5° E). The tropospheric meridional flow (*V850*) is over the broader Beijing area (850 hPa, 30°–47.5° N, 115°–130° E). We

those constructed for *TI*. Despite the small number of observations, the indicator variables for years with higher *HWI* are significant in all regressions. The regression R-squares are noticeably higher for Beijing when the *HWI* indicators are the explanatory variables compared to those with the *TI* indicators being the explanatory variables for the panel sample.

The overall low regression R-squares suggest that while innovation in pollution news is associated with local weather conditions, there is still a large unexplained component. Anthropogenic emissions at any given level of economic activity are sensitive to local control strategies, which in turn are related to incentives of local leaders. However, the monitoring of environmental performance in China is weak due to incorrect reporting of emissions data, which was largely at the control of local authorities prior to 2007 (Jin et al., 2016). As a result, sporadic "campaign style" regulation has been a key feature of China's environmental governance (Jin et al., 2016). Such campaigns can be associated with greater news coverage of air pollution and increase the pressure on local leaders.[20]

The lack of significant association between *RAPC* and cyclical components (or growth) of economic activity may still seem surprising. Liu and Cai (2018) study high-resolution CO_2 emissions in 288 Chinese cities and find very weak correlation between city-level per-capita carbon emissions and GDP per-capita. One possible explanation for the lack of association is that implementing emission control measures involves imposing economic costs on citizens (e.g., by closing down or limiting certain activities), and such costs are easier to absorb, and the measures are more likely to be adopted, when the local economy is more robust, offseting the impact of economic activity on emissions.

---

then normalize each of these three time series by its respective standard deviation. The haze weather index is constructed by adding these normalized series. The source of the temperature data is the same as that of *TI*. The 3-hour wind data come from M213NPASM version 5.12.4 of the MERRA-2. We aggregate 3-hour grid-level data into regional daily *HWI* in the same manner as how we aggregate temperature data to compute regional daily *TI*.

[20] The air quality management related to mega events and the official response to the $PM_{2.5}$ crisis in Beijing in 2012-2013 are well-recognized examples of such a process.

1.1.2 Cross-sectional Variation: Strength of Ethnic Ties

We measure the strength of a region's ethnic ties in terms of the share of Chinese population in that region. Ethnic links with China could be correlated with economic links. Reassuringly, as discussed in section 1.1.1, our main variable capturing innovation in air pollution news associated with China, *RAPC*, appears unrelated to economic activity in China.

To address this concern further, we measure the ethnic links of a U.S. region to China in terms of population statistics that go back almost one hundred and fifty years. Relatively comprehensive county-level Chinese population data is available as early as 1870. Counties with higher Chinese population proportion as early as 1870 are less likely to reflect current economic ties with China; however, due to early Chinese settlement in these regions, ethnic ties could endure. Indeed, we find that the correlation between 1870 and 2010 Chinese population by county is 0.34.[21] Based on the 1870 Census, we define counties as having stronger ethnic links to China depending on whether the proportion of Chinese population is above the median among all counties with Chinese residents, and alternatively, whether there is any record of Chinese residents in that county.

Figure 4 provides a map of the 1870 Chinese population distribution, with counties having positive Chinese population indicated in darker shade. When the analysis is at the Metropolitan Statistical Area (MSA) level, we aggregate Chinese and total population from the county level to similarly construct proxies for ethnic ties at the MSA level.

1.1.3 Empirical Design

Based on the two sources of variation discussed in section 1.1.1 and section 1.1.2, our empirical design for our main tests takes the following form:

$$Y_{t,j} = a + b_1 HC_j * RAPC_t + b_2 HC_j * L1.RAPC_t + b_3 HC_j * L2.RAPC_t$$

[21] Many of these counties of early Chinese settlement continue to draw in Chinese emigrants because of the presence of a significant Chinese community and amenities (such as a Chinatown) that are attractive to the Chinese.

$$+\sum_k c_k HC_j * China\ Economic\ Activity_{k,t} + \sum_l d_l Regional\ Controls_{l,j,t} + \sum_m f_m Direct\ Economic\ Links_{m,j,t} + \mu_j + \tau_t + \varepsilon_{j,t}$$

Here, the dependent variable is either the annual inflow of Chinese immigrants or students to U.S. region *j* in year *t*, or residential property price growth in region *j* from year *t-1* to year *t*. HC_j is an indicator of stronger ethnic links of region *j* to China. We include region and year fixed effects, and standard errors are clustered at region and year levels.[22]

To mitigate the concern that economic links (and not air pollution) drive our results, we control for interactions between measures of ethnic links and the degree of economic activity in China, and a variable capturing the extent of a U.S. region's direct economic linkage with China.[23] The regressions also control for regional controls that could influence migration to the region or property prices. Table 3 presents the correlation coefficients between key regional economic and demographic characteristics as of the baseline year and the Chinese proportion of the 1870 regional total population (for the three regression samples where the classification of counties according to the strength of ethnic ties is based on the 1870 Chinese population proportion). The correlation coefficients are generally insignificant.[24]

The success of our identification strategy relies on our measure of innovation in air pollution news, *RAPC*, not being associated with heightened economic activity between China and U.S. regions with stronger ethnic links to China. In support of this premise, we discussed in section 1.1.1 that while *RAPC* is unrelated to several measures of economic activity in China, some of the variation is explained by atmospheric and weather conditions unrelated to

[22] The specification for new international student enrolments to U.S. tertiary educational institutions is similar, except that it is at the post-secondary institution level and an institution-level proxy for the strength of the institution's ties with Chinese students (discussed in section 1.2.2 below) takes the place of the HC dummy.
[23] This is the ratio of the state-level air freight with China to the state-level total freight with all countries.
[24] Only the correlation with population growth is significant. Our results remain very similar if we add the interactions of population growth with *RAPC* and its lags to our main specifications, as shown in Tables A3 & A5 in Supplementary Appendix 2.

economic activity. Moreover, as discussed in section 1.1.2, the ethnic links in our study are based on population distributions in the U.S. as of 1870, so less likely to reflect current economic links which might be correlated with more recent Chinese population distribution in the U.S. Additionally, as noted, in all our specifications based on the strength of ethnic links, we include interactions between variables capturing stronger links and measures of economic activity in China. Such interactions should pick up effects that are attributable to economic linkages between China and regions with stronger ethnic links to China that might strengthen when economic activity in China is more robust and might create more pollution. In addition, we directly control for the strength of connection between U.S. regions and China (estimated using freight information), economic activity in China (often via year fixed effects) and economic and demographic variables of the U.S. regions.

As a complementary identification strategy, we also exploit cross-sectional variation in news concerning air pollution associated with different Chinese regions, in language ties between Chinese and U.S. regions, and in U.S. regional education links to Chinese cities.

1.2 Data and Variable Construction

We next describe the data, and how other key variables are constructed for our empirical tests.

1.2.1 Main Dependent Variables

We first describe our dependent variables: number of Chinese immigrants, number of international students studying in the U.S., and U.S. county housing ownership and prices.

Our immigration data is sourced from the Yearbook of Immigration Statistics available from the website of the U.S. Department of Homeland Security. In particular, we obtain the yearly number of China-born immigrants (*CIM*), admitted to the U.S. as a lawful permanent resident, by selected Metropolitan Statistical Areas (MSAs) of intended residence, over the period 1996-2020. In general, statistics for the 50 top destination MSAs are provided each year.[25]

Comprehensive data on the number of Chinese international students in U.S. regions is generally not available. However, the Integrated Postsecondary Education Data System (IPEDS) of the National Center for Education Statistics (NCES), the U.S. Department of Education, provides enrolment information by postsecondary educational institution, covering about 6,400 colleges, universities, and technical and vocational institutions in the U.S. In particular, we obtain the number of first-time freshmen whose permanent address is outside the U.S. (*ISTU*). Since China is the major origin of international students studying in the U.S. in the last two decades,[26] and there is no obvious reason why non-Chinese foreign students would respond to the concerns about air pollution in China, we restrict the analysis to the post-2000 period and attribute any association between innovation in air pollution news about China and international student enrolment in U.S. institutions primarily to the influx of Chinese students.[27]

We source the data on housing ownership from IPUMS USA (Ruggles et al., 2021), based on U.S. Census's American Community Survey 1-year Estimates of households. To estimate the housing ownership of new Chinese immigrant households, we calculate the ratio of (a) to (b), for which (a) is the number of respondents who were born in China, entered to the U.S. in the current year, and are living in an owner-occupied housing unit in a county; (b) is one of the following three alternatives: (b1) the county population in the survey, (b2) the total foreign-born population living in the county in the survey, and (b3) and the total China-born

---

[25] The annual average of the number of China-born immigrants to the MSAs covered in the sample for the 2000-2017 period is 57,566. This is comparable to the average number of 69,602 for the same period obtained from the OECD international migration database (https://www.oecd-ilibrary.org/statistics), which provides information on the annual inflow of Chinese to the U.S.

[26] The F1 visa statistics provided by the U.S. Department of State show that Mainland China is the origin country that has the largest number of F1 visa approvals for each year over the period 2008-2019. It is also the overall top source country of U.S. international students after 2000 (27%), followed by India (9.4%) and South Korea (8.9%). South Korea and Japan are larger sources before 2000.

[27] Students (admitted to the U.S. on F1 visa) are not included among immigrants in our immigration data discussed above.

population living in the county in the survey. We compute the ratio for each county every year over the period 2005-2019.

The data used to compute housing price growth at the U.S. county level are the county-level annual House Price Index, obtained from the U.S. Federal Housing Finance Agency (FHFA).

1.2.2 Educational Link Measures

We now introduce our measures of strength of education links for our student samples. We first construct a U.S. post-secondary institution-level proxy for the strength of institution *i*'s ties with Chinese students (CS_i). To do so, we use the ratio of the institution's number ofthe first-time foreign freshmen to its number of the first-time freshmen for 1986 – the first year this data is available to us – to construct a proxy for the importance of an U.S. institution for international students. The source of this data is IPEDS, mentioned above. To capture the importance of the institution from the point of view of Chinese international students, we multiply this ratio with the ratio of the number of F1 visas granted to Chinese students to that to all foreign students for the MSA where the institution is located. This latter data is obtained from the "Global Cities Initiative" study conducted jointly by the Brookings Institute and JPMorgan Chase. The study provides MSA-level information on the number of F1 visas approved during the 2008-2012 period for the top 118 MSAs that had enrolled foreign students, accounting for 85% of the foreign students pursuing a bachelor's degree or above in the U.S. during that period. The primary source of the data has information on the students' addresses in the Form I-20 when they apply for a U.S. F1 visa.

For our analysis of the impact of innovation in news concerning China's air pollution on residential prices in U.S. counties based on the strength of educational links, we define counties as having stronger or weaker educational links to China based on the average of $CS_{i,}$ for all institution *i* in county *j*, for which CS_i is defined as above.

As mentioned, our student samples cover the post-2000 period. As we use information within this sub-period to construct one of the key conditioning variables, we also run our tests that exclude the years 2008-2012.

1.2.3 Main Control Variables

To address the concern that innovation in air pollution in China could be spuriously picking up the effect of the degree of economic activity in U.S. regions and economic links between China and the U.S. regions on our outcome variables of interest, we saturate our regression models with a number of variables capturing such activity or links to the extent possible. To this end, we obtain personal income per capita and population of counties from the data website of the U.S. Bureau of Economic Analysis (https://www.bea.gov/data). For China, we obtain data of gross domestic product (GDP) per capita of Chinese regions and GDP per capita for China from the website of the National Bureau of Statistics of China (http://www.stats.gov.cn/). Chinese population statistics is also retrieved from the website of the National Bureau of Statistics of China. The data on exports and imports of goods and services of China is provided by the World Bank. To construct a variable capturing China's economic links with a U.S. region, we use data on freight with China and all countries in the world covering *all* air carriers (including both U.S. and foreign carriers), for every U.S. state of origin and destination, available from the website of the Bureau of Transportation Statistics of the United States (https://www.transtats.bts.gov/).

We winsorize all growth variables at the 1% and 99% to minimize the influence of outliers or errors in data. For ease of comparison of economic significance, in our regressions, we standardize all continuous variables.

2. Illustrative Evidence

We now provide three illustrative examples of how news about air pollution (or its health effects) triggers local (Chinese region-level) interest in emigration and in real estate in U.S. cities that are major destinations of Chinese citizens. The three cases focus on air pollution episodes in relatively small windows of time in Beijing in February 2014, Chongqing in May 2017, and Chengdu in July 2017. The city description, news information, and Baidu plots of these cases are provided in Supplementary Appendix 1.

For this evidence, we draw on the Baidu Index, supplementing with Factiva and Google search both in simplified Chinese and English. To capture interest in cross-border moves, we use the keywords "America", "study abroad", and "green card". We could not use the keyword "immigration" or "emigration" because Baidu no longer provides results based on that keyword.[28] As for housing interest, we use the keywords "real estate" and the name of the city ("Los Angeles", "New York", or "Seattle").[29]

Case I: Beijing, February 2014. Between February 13, 2014 and February 25, 2014, several news articles were published about smog levels in Beijing, with most occurring on February 25. Case 1 in Supplementary Appendix 1 gives a list of some of the important news sources and headlines. There was a sharp increase in Baidu search volume for the keyword "aqi" on February 26 and, more importantly for our purposes, an increase in Baidu search using keywords "real estate" and "Los Angeles" (alternatively, NY and Seattle). Figures A1a and A1b in Supplementary Appendix 1 provide visual evidence.

[28] Qin and Zhu (2018) were able to conduct Baidu search for 2014 based on "emigration" and find that city-level search of this term is positively related to local AQI. However, they do not examine search related to U.S. real estate. The Chinese term most likely to be used to conduct Baidu search to get information concerning emigration from China to another country is *yín mín* (the Chinese character 移民). The meanings of this Chinese term include emigrant, emigrate, emigration, immigrant, immigrate, immigration, migrant, and resettlement.

[29] Barcelona et al. (2021) report that for the U.S. housing markets, properties in the Core Business Statistical Area "Los Angeles – Long Beach – Anaheim, CA" have been most viewed (18.9%) by Chinese through Juwai website between November 2016 and January 2017, while those in CBSAs in New York and Seattle, "New York-Newark-Jersey City CBSA" (12.3%) and "Seattle-Tacoma-Bellevue CBSA" (5.5%) are among the next most viewed.

Caste II: Chongqing, May 2017. The likely precursor to the Chongqing case were two scientific articles published in 2017 on air pollution and health of children in Chongqing. One was published in January 2017 and was covered by China Weekly News on 24 January 2017. The other was published on 25 April 2017. On 16 May 2017, China Weekly News revealed that Chongqing is the top contributor (among 190 cities in China) to pollution-related mortality, accounting for 3.10% of the total deaths caused by PM10 pollution among 190 cities in China. Chongqing also had the highest premature mortality of 25,162 per year. Chongqing's Baidu search index (using "America", "study abroad", & "green card" as keywords) started to increase in May and peaked around May 12. Baidu search using "real estate" and "Los Angeles" (alternatively, NY and Seattle) as keywords also peaked sharply around this time.[30] Figures A2a and A2b in Supplementary Appendix 1 provide visual evidence.

Case III: Chengdu, July 2017. On 12 July 2017, the official website of China Environmental News (www.cenews.com.cn) reported that air quality in Chengdu had deteriorated in the first half of 2017. Official representatives of three districts in Chengdu were called upon for discussion with the provincial officer. This might have suggested that policies for tackling the air pollution problems in Chengdu had been ineffective. Both Chengdu's Baidu search index using "America", "study abroad", & "green card" as keywords, as well as the search index using "real estate" and "Los Angeles" (alternatively, NY and Seattle) surged around July 11, 2017. Figures A3a and A3b in Supplementary Appendix 1 provide visual evidence.

3. Air Pollution's Effects on Capital Outflow, Emigrants and Students

[30] The publication of scientific studies about the (adverse) effects of air pollution picked up by the media is a possible channel through which *RAPC* is affected. We formally test whether innovation in a time series of academic publications on air pollution in China (based on the number of hits per year from Google Scholar search on keywords "air pollution" and "China", scaled by the numbers of hits on all academic publications with keyword "China") is associated with *RAPC*. We find a positive coefficient significant at the 10% level when *RAPC* is regressed on the AR(1) residual from this Google Scholar series.

In this section, we first show that innovation in news about China's air pollution is associated with outflow of capital from China. We then provide evidence suggesting that such innovation is associated with outflows of Chinese emigrants and students – especially to U.S. regions with stronger ethnic and educational ties with China, and language ties with Chinese regions associated with air pollution news. In section 4, we associate innovation in news about China's air pollution with faster appreciation in residential property prices in U.S. regions with stronger ethnic and educational links to China.

3.1 Capital Outflow from China

Concern about air pollution in China may motivate Chinese citizens to invest in residential property in a country such as the U.S. with better air quality. This could happen when the decision to emigrate or to send children abroad for education is made, or when the cross-border move materializes, which could occur with some lag. Moreover, as discussed, spikes in air pollution can erode trust in government, and lead to capital flight.[31]

To begin with, we present preliminary evidence that *RAPC* is associated with outflow of capital from China (*CKOUT*). Chang and Dasgupta (2022) show that various aggregate measures of capital flight from China generally move in tandem. For *CKOUT*, we use Cudddington's (1986) estimate, which contains the main component of different capital flight measures, namely, the balancing entry in the balance of payment, i.e., net errors and omissions (Claessens et al, 1993).[32] We estimate the following model:

$$CKOUT_t = a + c_0 RAPC_t + c_1 L1.RAPC_t + c_2 L2.RAPC_t + Controls + \epsilon_t \qquad (1)$$

[31] Badarinza and Ramadorai (2018) find that following a shock to political risk in a particular country, capital flight from that country drives up housing prices in London city wards with high concentration of ethnic groups representative of the country as early as in the next quarter. Chang and Dasgupta (2022) document that Chinese capital flight follows high political risk and pushes up foreign property prices in regions with stronger Chinese ties.

[32] We obtain the data from Gunter (2017).

where among controls, we include the mean of the annual growth of China's GDP per capita (*MGDPG3*) and the mean of the annual growth of China's trade per capita (*MTRG3*) for years t, t-1 and t-2. Since we do not exploit cross-sectional variations in the strength of ethnic ties for this result, the evidence presented here has to be interpreted with caution. Even though we control for measures of Chinese economic activity, failure to properly control for relevant dimensions of economic activity could bias our estimates of the sensitivity of the capital outflows to air pollution news. In other words, our interpretation of these results relies more heavily on *RAPC* not reflecting economic activity in China discussed in section 1.1.1 above.

In Supplementary Appendix 2, Table A1 reports the regression results, based on robust standard errors. The contemporaneous *RAPC* has a positive coefficient significant at the 1% level. While the coefficient of one-period lagged *RAPC* is also positive and has an economically significant magnitude, it is not statistically significant. The magnitude of the *RAPC* coefficient is large — a one-standard deviation increase in *RAPC* leads to an increase of 26.7 percent of one-standard deviation of the dependent variable, i.e., approximately $16 billion (using statistics from Table A2 in Supplementary Appendix 2) increase in Chinese capital flight. Hence, China's capital flight reacts strongly and quickly to "surprise" in air pollution news.

Given the limitations of causal identification as in Eqn. (1) based on a single time series, the remainder of our results exploits? various types of cross-sectional variation (e.g., in air pollution news associated specifically with a particular Chinese region, the strength of ethnic ties of U.S. regions to China, language ties of U.S. regions to specific regions in China, and educational links of U.S. regions to Chinese cities) to identify how innovation in air pollution news affects the movement of Chinese citizens and housing prices in the U.S.

3.2 Chinese Emigration to the U.S.

We begin by documenting how innovation in news about China's air pollution is associated with emigration from China to U.S. regions with stronger (historical) ethnic and language ties to China. This is followed by an analysis of cross-border moves by students to U.S. regions that are likely to be major destinations of Chinese students. We close this section by providing evidence that variation in air-pollution news concerning different regions in China is associated with cross-sectional variation in the number of local students who leave China from that region to study abroad.

3.2.1 Ethnic Ties and Chinese Emigration to the U.S.

As mentioned above, the Yearbook of Immigration Statistics on the website of the U.S. Department of Homeland Security provides data on the number of new immigrants by country in which the immigrant was born for selected Metropolitan Statistical Areas (MSAs).

In general, statistics for the top-50 destination MSAs are provided each year. We hypothesize that innovation in news about Chinese air pollution would be associated with more emigration from China to regions with stronger ethnic ties to China. Aggregating from the county populations using 1870 Census data[33], we classify the MSAs as having stronger Chinese ties (with *HC*=1, where *HC* stands for "High Chinese Ties") if the proportion of the Chinese population in the MSA is above median among the MSAs in the sample. *HC* takes the value of 0 for the remaining MSAs. Alternatively, we assign the value of 1 to *HC* to an MSA if the 1870 Chinese proportion is positive, and 0 otherwise. We then estimate the following model:

$$CIM_{t,j} = a + b_1 HC_j * RAPC_t + b_2 HC_j * L1.RAPC_t + b_3 HC_j * L2.RAPC_t + c_1 HC_j * MGDPG3_t + c_2 HC_j * MTRG3_t + c_3 CFRSH3_t + \mu_j + \tau_t + \varepsilon_{j,t} \qquad (2)$$

[33] Chinese population data with comprehensive coverage at the MSA level are only available from the 1990 U.S. Census.

The dependent variable $CIM_{t,j}$ is the number of new Chinese immigrants settled in MSA j in year t. As defined before, *MGDPG3* and *MTRG3* are the 3-year (including the current year) averages of annual growth in China's GDP per capita and trade (exports plus imports) per capita, respectively. These variables are interacted with *HC* to absorb possible economic links between China and *HC* regions that might strengthen when economic activity in China is higher (possibly together with air pollution), and explain emigration to a particular region. Inclusion of these interactions should enable the interactions of *HC* and *RAPC* to pick up other non-economic reasons (such as atmospheric and weather conditions, or news about the effects of pollution on health) that drive the effect of *RAPC* on *CIM*. Accordingly, the interaction terms *HC**MGDPG3* and *HC**MTRG3* are included as controls in regressions whenever possible. *CFRSH3* is the 3-year mean of the change in the ratio of the state-level freight with China to the state-level total freight with all countries from the previous year to years t, t-1, and t-2, for the state where the MSA is located. This variable is included to directly control for change in economic connection between China and the region where an MSA is located.

Our regressions generally include fixed effects at the level of the unit of observations (e.g., educational institution, region) and year fixed effects. Robust standard errors are clustered at the unit and year levels. All variables used in the regression are standardized.

Column (1) of Table 4 reports the baseline results when *HC* indicates those MSAs with a Chinese proportion of the total population, as of 1870, above the median. Column (2) reports the baseline results when *HC* indicates those MSAs with positive Chinese population as of 1870. All interactions of *RAPC* and its lagged values with *HC* are positive and significant. The economic magnitude is large: a one-standard-deviation increase in *RAPC* is associated with 6% to 10% of one standard deviation higher emigration to *HC* MSAs at different lags. Over a 3-year period, the magnitude in column (1) is 24.7% of one standard deviation, i.e., about 21,000 more immigrants move to all *HC* MSAs (based on the statistics in Table A2).[34] The interactions with lagged values of *RAPC* are of larger magnitude, which is consistent with the possibility that certain categories of emigration might take longer than one year.

Several counties have relatively high concentration of Chinese population. For example, in San Francisco County in California, the Chinese accounted for 19.6% of the population as of 2000. To examine whether our baseline immigration results in Columns (1)– (2) of Table 4 are driven by these counties, we exclude the MSAs where the top five counties in terms of the Chinese proportion of the total county population, based on the 2000 Census data, are located. These counties together represent approximately 22% of total Chinese population of the U.S., as of 2000. We then re-estimate the regression model in Eqn. (2). Columns (3)– (4) of Table 4 show the results. The estimated coefficients of the interactions of *RAPC* and its lagged values with *HC* become slightly smaller while their statistical significance remains similar.

3.2.2 Language Ties and Chinese Emigration

Next, we exploit cross-sectional variation in language ties. Specifically, we examine whether innovation in air pollution news associated with certain Chinese regions is related to more Chinese immigration to U.S. regions where that Chinese region's language is more predominantly spoked by Chinese households. The 2000 U.S. Census records population by dialect for seven distinct Chinese dialects[35] that are spoken at home in the U.S. Of these, Mandarin (33%) and Cantonese (50%) are the major ones. The Cantonese-speaking population in China is mostly concentrated in the Guangdong province, while the Mandarin-speaking population is spread out over a wide area, including the capital, Beijing. We now examine

[34] It is estimated as the product of 3 terms: (i) the sum of the estimated coefficients of the 3 *HC* interaction terms, (ii) the standard deviation of *CIM*, and (iii) 25 (half of the number of MSAs corresponding to the above-median *HC* value in the sample in a typical year).
[35] The earliest comprehensive data on the U.S. regional distribution of foreign dialect speaking population is 2000.

whether innovation in *local* air pollution news (more correctly, air pollution news that mentions a local region) in China is associated with emigration to U.S. counties with stronger ties of Chinese language to that region. Specifically, we capture innovation in local air pollution news for Guangdong by first constructing the *APC* measure based on a Factiva search in simplified Chinese of "air pollution" together with "Guangdong" (in the same way as we construct China's *APC*), and fit an AR(1) model to the *APC* time series and extract the residual *RAPC*. Based on data available from the 2000 U.S. Census on the Chinese language spoken at home (available at the county level, but not at the MSA level), we obtain the proportion of Cantonese speakers to total population and alternatively to total Chinese population by aggregating data from the county level to the MSA level. We expect that Chinese emigration to MSAs with stronger Cantonese ties (as indicated by higher proportion of Cantonese speakers in the MSA) will be more sensitive to innovation in air pollution news associated with Guangdong, but not to innovation in air pollution news associated with other (non-Cantonese-speaking) Chinese regions. To examine whether this is the case, we estimate the following regression:

$$CIM_{t,j} = a + \sum_{k=0}^{2} b_k HC_j * RAPC_{t-k} + \sum_{k=0}^{2} c_k CAN_j * GRAPC_{t-k} + d_1 HC_j * MGDPG3_t + d_2 CAN_j * MGGDPG3_t + d_3 HC_j * MTRG3_t + d_4 CFRSH3_{j,t} + \mu_j + \tau_t + \varepsilon_{j,t} \quad (3)$$

As defined in Section 3.2.1, HC_j is an indicator variable that takes the value of 1 if MSA *j*'s proportion of Chinese in the total MSA population is in the top half of the sample. Likewise, CAN_j is an indicator variable taking a value of 1 if MSA *j*'s Cantonese speaking population as a fraction of the MSA's total population (alternatively, as a fraction of the MSA's Chinese population) is in the top half of the sample. *GRAPC* is the AR(1) residual for Guangdong's *APC* time series based on simplified Chinese. *RAPC* is defined as before. We hypothesize that in response to innovation in air pollution news associated with Guangdong, Chinese citizens in Guangdong province will be more likely to emigrate to MSAs with higher proportion of

Cantonese-speakers, that is, the coefficients c_k, k=0,1,2, will be positive. As a placebo, we alternatively interact CAN_j with the *RAPC* of Beijing and Shanghai. The major dialects spoken in these two cities are Mandarin and Shanghainese, respectively, and thus if language ties are important, there is no reason to believe that the coefficients of the corresponding interaction terms would be positive.[36] *MGDPG3*, *MTRG3*, and *CFRSH3* are defined as above. *MGGDPG3* is the 3-year (including the current year and preceding two years) mean of annual growth in Guangdong's GDP per capita.

Table 5 presents the results. In the odd-numbered columns, we present results for CAN_j defined as the proportion of Cantonese speakers to all Chinese in MSA *j*, while in the even-numbered columns, CAN_j is the proportion of Cantonese speakers in the total MSA population. In columns (1) and (2), the coefficients of the interaction terms of CAN_j with Guangdong's *RAPC* (*GRAPC*) and its lags are all positive. The interactions with the contemporaneous and two-period lagged *GRAPC* are significant at the 5% level. None of the coefficients for the interactions of CAN_j with Beijing's *RAPC* and its lags (columns (3) and (4)) or with Shanghai's *RAPC* and its lags (columns (5)– (6)) are significant. In all regressions, the interaction coefficients of *HC* with *RAPC* for China and its lags remain positive; the lag interactions are significant at the 10% level. Overall, these results are consistent with our prediction.

3.3 Educational Links and Student Moves

Next we study Chinese students' cross-border moves. While cross-border migration decisions are major decisions and are likely to be costly for the individuals concerned, requiring a higher threshold for a "push" factor, the decision to send children abroad for studies is likely

[36] We cannot design a similar test based on the Mandarin speaking population in the U.S. MSAs and *RAPC* for Beijing because Mandarin is spoken in most of China and the *RAPC* for China overall captures the exposure of Mandarin speakers to pollution news. Consequently, there may not be much explanatory power for Beijing's *RAPC* beyond the overall *RAPC*. Shanghainese is not one of the Chinese languages separately covered in the U.S. census (although its family Wu is included).

to require a lower threshold, especially for wealthier households. Such moves not only stop children's exposure to worsening air pollution, but also provide them with opportunities to migrate in the future if air quality does not improve or continues to worsen.

We face two types of data limitation to study associations between abnormal air pollution news coverage in China and the influx of Chinese students to the U.S. First, even though we have post-secondary institution level data on international freshman enrolments, breakdown based on country of origin is not available. However, since Chinese students constitute the majority of international students at least since the mid-2000, we restrict our analysis to the post-2000 period and argue that any association between *RAPC* and international student enrolment is most likely driven by Chinese international students.

A second data limitation is that the only available information we have on Chinese students' links to specific U.S. MSAs is a single cross-section based on a Brookings Institute and JPMorgan Chase study of student visas granted to international students for the period 2008-2012 (Ruiz, 2014). Using this data to construct a measure of an institution's ties to Chinese students is not ideal for our purposes, since it raises an endogeneity concern, discussed below. Fortunately, this data also provides information on the student's city of origin. We exploit this information to construct a weighted average *RAPC* based on the city's *RAPC* with the weights reflecting the importance of the city's *RAPC* for the U.S. MSA, and examine whether international student enrolment in an institution is related to the corresponding weighted *RAPC* for its MSA.[37]

3.3.1 Educational Links and New International Students in the U.S.

[37] Arguably, such an approach is less subject to endogeneity, since the weighted *RAPC* exploits the time varying cross-sectional city-level *RAPC* although the weights are static, in comparison to China's overall *RAPC*.

We briefly discuss how the number of freshmen international student (*ISTU*) enrolments in U.S. institutions is related to *RAPC* depending on our estimated strength of the institution's ties to China. We estimate the following model, which is very similar to the immigration model in Eqn. (2):

$$ISTU_{j,t} = a + \sum_{k=0}^{2} b_k HCS_j * RAPC_{t-k} + c_1 ALLSTU_{j,t} + c_2 HCS_j * MGDPG3_t + c_3 HCS_j * MTRG3_t + c_4 CFRSH3_{j,t} + \mu_j + \tau_t + e_{j,t} \tag{4}$$

Subscripts *j* and *t* index educational institutions and years. Section 1.2.2 discusses how we construct our measure of the strength of an institution's ties to Chinese students (CS_j). The indicator variable HCS_j for institution *j* is defined in two alternative ways: (i) it takes the value of 1 if CS_j is among the top half in the sample, and 0 otherwise, (ii) it takes the value of 1 if CS_j is positive, and 0 otherwise.

Since the construction of the HCS_j indicators for educational institution *j* involves information on Chinese student visas granted during the 2008-2012 period, albeit at the MSA level, there could be a concern that HCS_j is correlated with the error term in Eqn. (4). However, time-invariant institution-level factors that influence Chinese student enrolment are absorbed by the institutional fixed effects, so this is less likely to be an issue for years other than 2008-2012. As shown below, excluding these years from our regression sample does not change our results materially.

Table A4 in Supplementary Appendix 2 presents our results for the entire post-2000 period as well as for the sample that excludes the period 2008-2012. The results for the entire post-2000 sample indicate that the estimated coefficients of the interaction terms of HCS_j with *RAPC* and its lags are all positive. Based on columns (1)–(2), over three-year period, a one-standard deviation increase in *RAPC* is associated with about 8%–9% of one standard deviation more in the increase in international freshmen enrolment at U.S. post-secondary educational institutions that were more preferred destinations of Chinese students. As can be seen from columns (3)–(4), the results are similar when we exclude the period 2008-2012.[38] We also examine whether the effect of innovation in China's air pollution news on influx of students to the U.S. is driven by a few major student destinations, and find that our results remain even when we exclude the institutions located in the top-five MSAs, in terms of the MSA-level Chinese share of foreign students. The results are reported in columns (5)–(8).

3.3.2 Cross-Sectional Variation in Educational Ties to Chinese Cities

Next, we exploit cross-sectional variation in the exposure of U.S. MSAs to local air pollution news in Chinese cities based on the strength of the MSA's ties with these cities. To do so, we replace the interaction term between HCS_j and *RAPC* in Eqn. (4) by an MSA-specific weighted *RAPC*. The weighted $RAPC_j$ for an institution in MSA j is a weighted average of *RAPC* of Chinese cities, where the city weight is the proportion of Chinese international students in MSA j who come from that Chinese city, for which our data source is also the Brookings Institute. We replace the lag interaction terms in the same manner. We then re-run the institution-level regression. The results are shown in Table 6. In column (1), we find that the estimated coefficients of weighted *RAPC* and its lags are all positive and significant at the 1% level. The coefficients range between 2.3% and 4.8%. In column (2), when we exclude the period 2008-2012, the estimated coefficients have similarly high levels of statistical significance, and are slightly larger in magnitude. To address the concern that the local *RAPC*s move together so that the weighted *RAPC* essentially reflects China's overall *RAPC*, we conduct placebo tests by randomly assigning the weights for a particular MSA to the cities, and

[38] It is worth emphasizing that these results probably underestimate the impact of innovation in Chinese air pollution news on the cross-border moves of Chinese students because of the possibility that higher applications from Chinese students could crowd out other international student admissions to U.S. post-secondary institutions.

repeat the regression. We bootstrap the process 1000 times. The bootstrapped results reported in the last six columns show that the means of the estimated coefficients of "weighted pseudo-$RAPC_j$" and its lags are generally insignificant, with magnitude close to 0.[39]

3.3.3 College Entrance Examination Sitters in Chinese Regions

As an alternative strategy for causal identification, we exploit cross-sectional variation in innovation to air pollution news associated with a particular region in China (*RLAPC*, the local version of *RAPC*) and the response of local students. In particular, we examine whether larger *RLAPC* for a region is associated with an increase in the number of local senior high school graduates who leave China to study abroad, whereby the number of students who registered for the National College Entrance Examination (*Gaokao)* in that region decreases.[40] Our sample covers all 31 provinces, municipalities and autonomous regions (hereafter "regions") under direct jurisdiction in China over the period 2006-2020. China has a system of registration of residency status of individuals, known as *Hukou*, which limits internal migration of Chinese citizens.[41] Therefore, it is generally difficult for students to take the examination in another region where they do not have *Hukou* (Sun and Chen, 2015). Our regression model is as follows:

$$EXAMSTU_{j,t} = a + b_1 RLAPC_{j,t} + b_2 L1.RLAPC_{j,t} + b_3 L2.RLAPC_{j,t} + c\ GRADSTU_{j,t} + d_1 MLGDPG3_{j,t} + d_2 MGDPG3_t + d_3 MTRG3_t + \mu_j + e_{t,j} \quad (5)$$

[39] One caveat is that there are only 15 major Chinese source cities for which we have estimates of the MSA-level number of Chinese international students who had studied in the U.S. over the period 2008-2012. The 15 Chinese cities are Beijing, Chengdu, Dailian, Fuzhou, Guangzhou, Hangzhou, Nanjing, Ningbo, Qingdao, Shanghai, Shenyang, Shenzhen, Suzhou, Wuhan, and Xian.

[40] The examination is required for entrance into almost all higher education institutions in China and is usually taken by students in their last year of senior high school.

[41] In China, the central government generally tightly controlled internal migration although the migration restrictions have been relaxed in recent few decades (Chan, 2015).

Subscripts j and t index regions and years. $EXAMSTU_{j,t}$ is the number of students who registered for the National College Entrance Examination in region j in year t. $LAPC_{j,t}$ is the ratio of the number of articles based on Factiva Simplified Chinese search using the keywords "air pollution" and the name of region j to that only using the keyword the name of region j for year t. *RLAPC* is residual *LAPC* of an AR(1) model. *L1.RLAPC* is *RLAPC* lagged once. *L2.RLAPC* is *RLAPC* lagged twice. The regression controls for $GRADSTU_{j,t}$, which is the number of senior high school graduates in region j in year t.[42] *MLGDPG3* is the mean of the annual growth of region j's GDP per capital in year t and the preceding two years. The other variables are defined as before.

One concern is that the coefficients of *RLAPC* and its lags (b_1, b_2, and b_3) may be driven by variation of China's overall *RAPC* over time. Therefore, we construct *RLAPC* in relation to *RAPC*, in the following two alternative ways. First, we replace $RLAPC_{j,t,}$ by the difference between $RLAPC_{j,t,}$ and *RAPC*, and conduct similar replacement for its lags. Second, we replace $LAPC_{j,t}$ by the ratio of $LAPC_{j,t,}$ to $APC_{t,}$, and then redefine the residual ratio, from an AR(1) model, as $RLAPC_{j,t,}$.

Table 7 reports the results, based the two alternative ways to construct local news innovation. Consistent with the expectation, all estimated coefficients of *RLAPC* and its lags are negative and generally statistically significant. Among these coefficients, the contemporaneous one has the largest magnitude (0.029 in column (2)), and translates to a decrease of the number of registered students by approximately 6,150 per region per year

[42] The difference between $GRADSTU_{j,t}$ and the first-time exam-sitters (estimated as $0.8*EXAMSTU_{j,t}$ to reflect the fact that students in China can sit for the *Gaokao* more than once) is an estimate of the number of Chinese students who enrol as freshmen in other countries. Using UNESCO and OECD data on overall numbers of students from China who are enrolled in tertiary institutions in different countries for 2015-2018, we estimate that the fraction of outbound Chinese students going to the U.S. is about 39%. Multiplying the difference with this fraction gives us an annual average of 158,102 as an estimate of the annual number of Chinese freshmen enrolling in U.S. post-secondary institutions for this period. The number is almost the same as the annual average number of F1 visas granted by the U.S. to Chinese students (158,549) over the same period.

(based on the statistics in Table A2 in Supplementary Appendix 2). The one-period lagged coefficient of *RLAPC* has comparable, but smaller magnitude. Over the two-year period, an increase in *RLAPC* by one standard deviation is associated with a decrease in the number of students who registered for the National College Entrance Examination by approximately 5 percent of one standard deviation. These results are consistent with the notion that an increase in local air pollution news innovation is likely associated with an increase in the number of local students who leave China to study abroad.

4. China's Air Pollution News and Residential Property Prices in U.S. Counties

As Badarinza and Ramadorai (2018) and Chang and Dasgupta (2022) show, capital flight triggered by political risk in "source" countries results in property price appreciation in regions where there are stronger ethnic and education ties with the source country. Innovation in news about China's air pollution can lead to the relocation of some sections of Chinese citizens to countries which are believed to have clearer air and can also cause capital flight. Our results so far establish that these plausible channels are indeed at work.

While some of the capital inflow to the U.S. presumably is contemporaneous with the movements of emigrants or students, it is also conceivable that it precedes or follows such movements. In any case, the main channel of influence of innovation in Chinese air pollution news on U.S. property prices is likely to be capital inflows into the residential property market. We therefore expect that innovation in news about air pollution in China will be positively associated with housing price growth in U.S. regions with stronger ethnic or educational links to China.

We present results on the effect of innovation in China's air pollution news on housing prices in U.S. counties for three samples. Two of these correspond to our analysis of emigration in section 3.2.1 (the "immigration sample") and student movements in section 3.3.1 (the "student sample"). The third is the sample of all counties for which housing price and other relevant data are available ("full sample"). For this latter sample, we again draw on early Chinese settlement to mitigate the concern that recent Chinese population concentration in a particular region could reflect the strength of the region's economic ties with China. We use the county-level Chinese population as of 1870 to determine whether a county has strong or weak Chinese ties in two alternative ways. First, when the 1870 Chinese population, as a proportion of the total county population, is above the median of all counties with recorded Chinese population, we consider the county as having strong Chinese ties and assign the value of 1 for a high Chinese dummy (*HC*). We consider the remaining counties as having weak Chinese ties and assign them the value of 0 for their *HC*. Alternatively, if the 1870 Chinese proportion is positive, we consider the county as having stronger Chines ties, and assign the value of 1 to be the *HC* value. We consider the remaining counties without a positive Chinese share as having weaker Chinese ties, and assign the value of 0 to be their *HC* value.[43]

For the "immigration sample", since our analysis is now at the county level, we no longer need to aggregate 1870 county population numbers to the MSA level to determine a region's strength of ethnic links to China. Accordingly, a county's HC value is determined in exactly the same way as for the full sample. For the "student sample", CS_{county} for each county is the average value of CS_i of all educational institutions in that county, and its *HC* is either (i) equal to 1 if CS_{county} is above median, and zero otherwise, or alternatively, (ii) equal to 1 if that CS_{county} is positive, and zero otherwise.

With HC defined as above for the counties in the respective samples, we interact HC with the key variables of interest and estimate the following model:

---

[43] Alternatively, we also classify counties based on the recorded number of Chinese in the county as of 1870, and control for the county-level total population in our regressions (to address possible "size effects"). Our results remain. The results are reported in Table A16 at the end of Supplementary Appendix 2.

$$HPG_{j,t} = a + \sum_{k=0}^{2} b_k HC_j * RAPC_{t-k} + c_1 HC_j * MGDPG3_t + c_2 HC_j * MTRG3_t \; + c_3 CFRSH3_{j,t} + Local\ Controls + \mu_j + \tau_t + \varepsilon_{j,t} \tag{6}$$

$HPG_{j,t}$ is the annual nominal housing price growth of county *j* for year *t*. As above, control variables include interactions of *HC* dummy with (i) the 3-year average of annual growth of China's GDP per capita and (ii) the 3-year average of annual growth of China's trade per capita, and 3-year mean of the change in the state-level ratio of the freight with China to the total freight with all countries (for which the state is where the county is located), and additionally, county-level control variables including annual growth of personal income and annual growth of population, for the contemporaneous year and for the next five years, and population of the previous year.

In Table 8, columns (1)–(2) report the results based on counties in the immigration sample; columns (3)–(4) report the results based on counties in the student sample, and columns (5)–(6) report the results based on the full county sample. The odd-numbered columns show the results when *HC* is the above-median dummy reflecting ethnic or education ties. The even-numbered columns show the results when *HC* is the positive dummy for the ties.

The estimated coefficients of *HC* interaction terms with *RAPC* and one-period lagged *RAPC* are all positive and generally statistically significant. The magnitudes of *HC* interactions are larger when *HC* is classified based on a higher Chinese share cut-off. Overall, the results suggest that the strength of ethnic ties is at work — when there is larger innovation in news about air pollution in China, U.S. counties with stronger Chinese ties experience more housing price appreciation, compared with those counties with weaker Chinese ties.

The magnitude of the incremental effects for *HC* counties is similar for the immigration sample and the full sample, and larger compared to the student sample. Based on the full sample, a one-standard deviation increase in *RAPC* increases housing price growth by 17.4 percent of one-standard deviation more (about 0.8 percent per year based on the statistics in Table A2 in

Supplementary Appendix 2) in counties with Chinese ties than in counties without clear Chinese ties. When we use the higher (above-median) cut-off to classify *HC*, housing price growth increases by an additional of 27.7 percent of one-standard deviation (about 1.3 percent per year based on the statistics in Table A2) in counties with stronger Chinese ties than in counties with weaker Chinese ties. The magnitude of the one-period lagged *RAPC* interactions remains economically significant (70% - 75% that of the contemporaneous interactions). Since the sample average housing price growth is about 3 percent per year, these estimates of the incremental effect of a one-standard deviation increase in *RAPC* on counties with stronger ethnic ties to China represent economically significant effects.

Compared to that for the full sample, for the emigration sample, the corresponding contemporaneous magnitude is slightly lower (about 0.7 percent per year higher housing price growth in counties with stronger Chinese ties for a one-standard deviation increase is *RAPC*), while for the student sample, the corresponding magnitude is 0.21 percent per year. The results suggest that the capital inflow into residential real estate associated with immigration (when there is more abnormal air pollution news coverage) is more important than that associated with student inflows.[44] It is also possible that these magnitudes reflect the higher *RAPC*-sensitivity of Chinese immigration to regions with stronger ethnic ties than that for student flows to institutions with stronger educational links to Chinese students. While immigrants can choose where they want to reside or invest, students are likely to face more constraints in being admitted to their preferred institutions/regions.[45]

[44] According to the National Association of Realtors, 8 percent of U.S. home purchases by Chinese buyers in 2017 were intended for use by a student, compared to 39 percent as primary residence, and 39 percent as residential investment and vacation home, respectively (2017 Profile of International Activity in U.S. Residential Real Estate, page 25, available at https://www.nar.realtor/sites/default/files/documents/2017-Profile-of-International-Activity-in-US-Residential-Real-Estate.pdf).

[45] The cumulative 3-year *RAPC*-sensitivity of Chinese immigration (relative to mean) to regions with stronger ethnic links is about four times that of international student inflows to institutions with stronger Chinese links, based on (i) the coefficients of *RAPC* and its lags in column (2) of Table 4 and column (2) in Table A4 in Supplementary Appendix 2, and (ii) the sample means and standard deviations of immigration and international

Overall, we find that housing prices of U.S. counties respond relatively quickly to innovation in news about China's air pollution, as reflected by the larger magnitude of the estimated coefficients of the interactions of the contemporaneous *RAPC* compared with the interactions of lagged *RAPC*s. This contrasts with the results in section 3.2.1 for cross-border moves of Chinese emigrants, where the lagged *RAPC* interactions are of larger magnitude than the contemporaneous *RAPC*, suggesting that capital outflows react more quickly to innovation in air pollution news than the movement of people does.

5. Robustness

We now discuss additional tests that rule out alternative explanations and further support our conclusions. Because the immigration sample covers a relatively small number of MSAs and the student sample is restricted to the post-2000 period, we address these robustness issues mainly with respect to the effect of innovation in air pollution news on housing price growth for the full sample.

5.1 Matched County Sample

While we control for local economic and demographic variables in our regressions, an alternative way to address the concern that our results are driven by systematic differences between counties with strong ethnic ties and counties with weak ethnic ties is to base our tests on a propensity score matched sample. To this end, using the full sample, we first match each county with an 1870 Chinese share above the median (*HC*=1) with the "nearest" county among all *HC*=0 counties, in terms of the following county-level key economic and demographic characteristics as of year 2000: personal income per capita and its growth; unemployment rate and its change; and population and its growth. We then re-run the housing regression based on this matched sample for the post-2000 period. We repeat the above procedure for the alternative

student inflows (at the level of the corresponding units of analysis), respectively, reported in Table A2 in Supplementary Appendix 2.

HC classification (i.e., *HC* counties are those counties with a positive 1870 Chinese share). We find that our conclusions remain. The results are reported in Table A6 in Supplementary Appendix 2.

5.2 Regions with High Chinese Concentration

As in our analysis of cross-border moves, we also examine whether the above housing results are mainly driven by counties with high Chinese concentration. We thus exclude top 5 Chinese share counties based 2000 Census county-level data, and re-run all the regressions in Table 8. The results are reported in Table A7 in Supplementary Appendix 2, which shows that the estimated coefficients of *HC* interactions with *RAPC* and one-period lagged *RAPC* remain similar, in terms of both economic and statistical significance.[46]

The state of California has been historically important for Chinese migration, and anecdotal evidence suggests that Chinese capital inflows have boosted property prices in that state. We thus examine whether our results hold if we exclude California from our sample. Excluding California, we find that the new estimated coefficients of *HC***RAPC* corresponding to the full-sample baseline regressions in columns (5)– (6) in Table 8 remain positive and significant. The magnitude of the estimated coefficients of the contemporaneous and one-period lagged interactions is 64%–72% that in the original sample with California. The estimated coefficients of the two-period lagged interaction are larger and statistically more significant (at the 5% level) than the corresponding baseline coefficients. These results are reported in columns (5)– (6) of Table A8 in Supplementary Appendix 2.

5.3 Subperiods

Next, we examine whether the results are stronger in recent years. Specifically, in columns (1)– (4) of Table A8 in Supplementary Appendix 2, we run the full-sample regressions

[46] We find similar results if we exclude top 10 Chinese share counties. These results are not tabulated, but available from the authors upon request.

separately for two sub-periods —before and after 2000-end, respectively. Since air quality in China has worsened steadily over time, we expect that the sensitivity of individual decisions to innovation in air pollution to be more significant in the latter period, when worsening air pollution reaches a more alarming level. The results are consistent with our expectation. The coefficients for the earlier sub-period are insignificant.

5.4 Alternative News Measures

We consider whether our results are sensitive to alternative ways of constructing the news measures. We find that our housing baseline results, in Columns (5)– (6) in Table 8, are robust with respect to the following alternative measures of innovation to China's air pollution news. First, we replace *RAPC* and its lags by those based on news in simplified Chinese, and then re-run the regressions. In Supplementary Appendix 2, Table A9 shows the results. We find economically and statistically stronger results for all key variables of interest. The interaction between two-period lagged *RAPC* and *HC* has larger estimated coefficients and higher level of statistical significance, compared with that of the baseline results in Table 8. Second, we replace *APC* by the alternative *APC* based on the broadest set of "air pollution" search terms, discussed in section 1.1.1, and then re-run the regressions. We find that the results remain, as shown in Table A10 in Supplementary Appendix 2. Third, as discussed in section 1.1.1., *RAPC* has particularly large values in certain years, for example, in 2013. To examine whether our results are significantly affected by these large values, we replace *RAPC* and its lags by their ranks, and then re-run the regressions. In Supplementary Appendix 2, Table A11 shows that our results are qualitatively the same. Finally, we replace *RAPC* and its lags by the first difference in *APC* and the lags of the difference, and then re-estimate the regressions. We find that our results, shown in Table A12 in Supplementary Appendix 2, are similar as before.

5.5 Supply Elasticities

Property prices may respond more to *RAPC* and its lags if supply elasticities are lower. It is possible that *HC*=1 counties are associated with tighter regulation of residential construction relative to *HC*=0 counties. To determine whether this could be driving our results, we re-run our regressions by additionally controlling for interaction terms of a composite regulatory index (*LURI*) with *RAPC* and its lags. *LURI* is a standardized measure of residential land use regulatory restrictiveness, based on a 2018 survey of communities across nationwide metropolitan areas in the U.S.[47] The index is the first factor of a factor analysis of a dozen subindexes that capture the different components of the local regulatory environment (Gyourko et al., 2021). We report the results in Table A13 in Supplementary Appendix 2. We find that the estimated coefficients of the interaction terms of *HC* with *RAPC* and its lags remain positive and highly significant, with comparable magnitudes as before.

5.6 Political Risk

Following Badariza and Ramadorai (2018) and Chang and Dasgupta (2022), we re-run our regressions by incorporating an interaction term between an indicator of the strength of ethnic/educational links to China and China's political risk in relation to the U.S.'s political risk as an explanatory variable in the regressions. Inclusion of this variable leads to a smaller sample size as the political risk data does not cover our entire sample period; however, our results remain (see Table A14 in Supplementary Appendix 2).

5.7 A Placebo Test

One possibility that may not be adequately addressed by the above robustness tests is that extreme air pollution years are becoming more common over time, and what we could be capturing is not sensitivity to extreme air pollution but rather a concurrent time period effect

[47] Gyourko et al. (2021) report that highly (lightly) regulated housing markets in the 2005 survey do not generally change to be less (more) strictly regulated in the 2018 survey.

that reflects the per capita GDP of China crossing some threshold beyond which the upper middle class considers living in foreign countries attractive and affordable – but not necessarily because they have become particularly sensitive to air pollution. Controlling GDP per capita would not necessarily solve the problem, as this effect could be nonlinear in GDP per capita. To address this issue, we conduct a comprehensive re-examination of our main results by lagging all air pollution news variables by five periods (we chose five periods because we have already established that the lags of *RAPC* have significant effects in some of our specifications). As shown in Table A15 in Supplementary Appendix 2, we find no significant results, suggesting that we are not simply capturing a time-period effect coinciding with rising wealth in China. Taken together, we find a robust positive association between innovation in news about China's air pollution and appreciation of housing prices in U.S. counties with stronger ethnic and educational ties to China.

6. Household-Level Evidence

In this section, we utilize survey data from IPUMS USA (Ruggles et al., 2021), based on U.S. Census Bureau's American Community Survey (ACS) 1-year Estimates of households.[48] This data enables us to relate innovation in news about China's air pollution to household-level evidence on home ownership, housing values, and rents. The unit of analysis in what follows remains at the county level; however, the data is aggregated from surveyed households to construct dependent variables at the county level.

6.1 Emigration and Housing Ownership in U.S. Counties

[48] About 1 in 38 U.S. households per year receives the questionnaire for participating in the ACS. An ACS Information Guide can be downloaded from https://www.census.gov/programs-surveys/acs/about.html. For more details on ACS single year and multiyear estimates, please read https://www.census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch03.pdf

Badarinza and Ramadorai (2018) point out that it is difficult to know the country of origin of the actual property buyers because of data limitations such as routing transactions via offshore vehicles. However, IPUMS USA provides three pieces of information that enable us to identify the housing ownership status of Chinese who newly entered the U.S. The three pieces of information are: i) the country in which the respondent was born, ii) the year in which the respondent entered the U.S., and iii) whether the respondent is living in an owner-occupied housing unit. From i), we obtain the number of individuals born in China in the sample (*CPOP*, denominator). From i) to iii), we obtain the number of individuals born in China who are living in an owner-occupied housing unit and who entered the U.S. in the survey year (*CHO*, numerator). The ratio of these two numbers (*CHO/CPOP*) can be computed annually at the county level. Alternatively, we use the foreign-born population (*FPOP*) or the total population (*TPOP*) in the sample county as the denominator of the ratio (*CHO/FPOP* or *CHO/TPOP*). Since it is possible for a new immigrant to (temporarily) live with another person living in an owner-occupied house, the numerator of the ratio does not necessarily reflect property purchases by new Chinese immigrants; however, it should be positively associated with both the number of new Chinese immigrants as well as the number of such immigrants who have purchased a property and are living in that property.

We estimate the following regression model:

$$CHO/DENOM_{j,t} = a + b_1 HC_j * RAPC_t + b_2 HC_j * L1.RAPC_t + b_3 HC_j * L2.RAPC_t$$
$$+ c_1 HC_j * MGDPG3_t + c_2 HC_j * MTRG3_t + c_3 CFRSH3_{j.t} +$$
$$Local\ Controls + \mu_j + \tau_t + \varepsilon_{j,t} \quad (4)$$

DENOM = {*CPOP*, *FPOP*, *TPOP*}. Subscripts *j* and *t* index counties and years. Based on 1870 Census data, we classify the counties as having stronger Chinese ties (*HC*=1) if the Chinese proportion of the total population in the county is positive, and the remaining counties as having weaker Chinese ties (*HC*=0).[49] The other variables are defined as before.

Column (1a) of Table 9 show the results when *DENOM* is *CPOP*. The estimated coefficients of the interactions of *HC* with contemporaneous *RAPC* and one-period lagged *RAPC* are positive and statistically significant. Over the two-year period, an increase in one standard deviation in *RAPC* is associated with 8% of one standard deviation higher proportion of Chinese entering the U.S. in a particular year and living in an owner-occupied housing unit in *HC* counties, compared with the other counties. Columns (1b)–(1c) show similar results when we replace *CPOP* by *FPOP* and by *TPOP*. The results suggest that innovation in news concerning air pollution in China is positively associated with housing ownership of Chinese emigrants in U.S. regions with stronger ethnic ties to China.

6.2 Rent and House Value Growth

Next, we examine the effect of innovation in news about China's air pollution on rents and house values as reported by surveyed households. To reduce heterogeneity, both variables are scaled by household income. The dependent variables are, respectively, the percentage change in the median ratio of rent over household income and in the median ratio of house value over household income, among all households in the county in the survey year. The results are reported in columns (2a) and (2b) of Table 9. The contemporaneous *RAPC* interacted with *HC* has a significant positive coefficient when the dependent variable is the rent-to-income ratio. When the dependent variable is the house value-to-income ratio, the interactions of *HC* with both the contemporaneous *RAPC* and one-period lagged *RAPC* are positive and significant, while that with two-period lagged *RAPC* is negative and significant (with a smaller

[49] We only work with the "positive Chinese proportion in 1870" version of the *HC* indicator variable because the "above median" criterion leaves us with only 2.6% of the county-years that are *HC*=1 and have home ownership by a new Chinese emigrant (i.e., positive *CHO*). Overall, 15.6% of the sample county-years have a positive *CHO*.

magnitude and a lower level of significance). The latter result suggests either some spillover to counties with weaker Chinese ties, or some reversal to an initial over-reaction in the counties with stronger Chinese ties.

Overall, the survey data provides micro-evidence suggesting a positive effect of innovation in news concerning China's air pollution on housing ownership of Chinese emigrants, rents and house values in counties with stronger ethnic ties to China.

7. Conclusion

How quickly individuals revise their beliefs about the salience of environmental risks is likely to be a key determinant of the success of efforts towards arresting climate change. Such belief revision is easier to detect in countries that are major contributors to emissions, and where individual economic decisions are likely to be more responsive to exposure to these risks. China has been one of the major contributors to emissions and Chinese citizens have reached a standard of living where, for a significant percentage of the population, environmental and climate issues are important considerations for the quality of living. How they respond to indicators of worsening climate and environmental conditions, and what the likely economic implications of such responses are, therefore is an important issue to study. In this paper, we construct a time series of news coverage about air pollution in China. We find that when news coverage is unexpectedly high, more Chinese citizens leave the country, and capital flight from China picks up. Focusing on the U.S., we find that emigration to regions with stronger ethnic (language) ties to China (Chinese regions) increases when innovation in air pollution news about China (or a Chinese region) is higher. Similarly, international student enrolments in U.S. universities with stronger ties to Chinese students also increases. We find that regions in U.S. with stronger ethnic and education links to China experience faster growth in residential property prices in response to higher innovation in China's air pollution news.

Overall, the results suggest that environmental issues are increasingly shaping individual decisions in China and having cross-border repercussions in terms of relocation of capital and labor.

Data Availability Statement

The data underlying this article are available from sources in the public domain: the U.S. Department of Homeland Security (*https://www.dhs.gov/immigration-statistics/yearbook*), the U.S. Department of Education (*https://nces.ed.gov/ipeds/use-the-data*), and the Federal Housing Finance Agency (*https://www.fhfa.gov/DataTools*). The news data for constructing the key variable of interest can be downloaded from a Dow Jones Factiva platform under licence.

References

Albert, C., Bustos, P., and Ponticelli, J. (2021) *The effects of climate change on labor and capital reallocation* (No. w28995), National Bureau of Economic Research.

Badarinza, C., and Ramadorai, T. (2018) Home away from home? Foreign demand and London house prices, *Journal of Financial Economics* 130(3), 532–555.

Barcelona, W., Converse, N., and Wong, A. (2021) US housing as a global safe asset: Evidence from China shocks. *Available at SSRN 3965183*.

Cai, W., Li, K., Liao, H., Wang, H., and Wu, L. (2017) Weather conditions conducive to Beijing severe haze more frequent under climate change, *Nature Climate Change* 7(4), 257–262.

Chan, K. W. (2015) Five decades of the Chinese Hukou System. In *Handbook of Chinese Migration: Identity and Wellbeing*, 23–47. Northampton, MA: Edward Elgar Publishing, Inc.

Chang, Y. and Dasgupta, S. (2022) Capital inflows and property prices: Ethnicity, education, and spillovers, *Journal of Financial and Quantitative Analysis*, 1–33. doi:10.1017/S0022109022000072

Chen, S., Oliva, P., and Zhang, P. (2017) *The effect of air pollution on migration: Evidence from China* (No. w24036), National Bureau of Economic Research.

Choi, D., Gao, Z., and Jiang, W. (2020) Attention to global warming, *The Review of Financial Studies* 33(3), 1112–1145.

Claessens, S., Naude, D., and Mundial, B. (1993) *Recent estimates of capital flight*. Washington, DC: World Bank.

Cuddington, J. T. (1986) "Capital Flight: Estimates, Issues and Explanations." In *Princeton Studies in International Finance*, 58.

Eichholtz, P., Steiner, E., and Yönder, E. (2019) Where, when, and how do sophisticated investors respond to flood risk? *Available at SSRN 3206257*.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020) Hedging climate change news, *The Review of Financial Studies* 33(3), 1184–1216.

Favilukis, J. and Van Nieuwerburgh, S. (2021) Out-of-town home buyers and city welfare, *The Journal of Finance* 76(5), 2577–2638.

Giglio, S., Maggiori, M., Rao, K., Stroebel, J., and Weber, A. (2021) Climate change and long-run discount rates: Evidence from real estate, *The Review of Financial Studies* 34(8), 3527–3571.

Gibson, M. and Mullins, J. T. (2020) Climate risk and beliefs in New York floodplains, *Journal of the Association of Environmental and Resource Economists* 7(6), 1069–1111.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009) Detecting influenza epidemics using search engine query data, *Nature* 457, 1012–1014.

Gunter, F. R. (2017) Corruption, costs, and family: Chinese capital flight, 1984–2014, *China Economic Review* 43, 105–117.

Gupta, P. and Follette-Cook, M. September 23, 2021. Introduction to Air Quality Forecasting. National Aeronautics and Space Administration. https://appliedsciences.nasa.gov/sites/default/files/2021-09/AQF_Part1_Edited_JO.pdf

Gyourko, J., Hartley, J. S., and Krimmel, J. (2021) The local residential land use regulatory environment across US housing markets: Evidence from a new Wharton index, *Journal of Urban Economics* 124, 103337.

Jin, Y., Andersson, H., and Zhang, S. (2016) Air pollution control policies in China: a retrospective and prospects. *International Journal of Environmental Research and Public Health* 13(12), 1219.

Juwai. "6 Reasons Why Education Underpins Chinese Overseas Property Investment." (July 7, 2016). https://list.juwai.com/news/2016/07/6-reasons-why-education-underpins-chinese-overseas-property-investment

Khanna, G., Liang, W., Mobarak, A. M., and Song, R. (2021) *The productivity consequences of pollution-induced migration in China* (No. w28401), National Bureau of Economic Research.

Liu, Z. and Cai, B. (2018) High-resolution carbon emissions data for Chinese cities, *Environment and Natural Resources Program Belfer Centre for Science and International Affairs Harvard Kennedy School*.

Lloyds Register Foundation. 2020. The Lloyd's Register Foundation World Risk Poll [Online] Available at: https://wrp.lrfoundation.org.uk

Malek, E., Davis, T., Martin, R. S., and Silva, P. J. (2006) Meteorological and environmental aspects of one of the worst national air pollution episodes (January, 2004) in Logan, Cache Valley, Utah, USA, *Atmospheric Research* 79(2), 108–122.

The New York Times and Lustgarten, A. (2020) The great climate migration. New York Times Magazine.

Ortega, F. and Taṣpınar, S. (2018) Rising sea levels and sinking property values: Hurricane Sandy and New York's housing market, *Journal of Urban Economics* 106, 81–100.

Pew Research Center, September 2015. Corruption, pollution, inequality are top concerns in China. Qin, Y. and Zhu, H. (2018) Run away? Air pollution and emigration interests in China, *Journal of Population Economics* 31(1), 235–266.

Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., and Sobek, M. (2021) *IPUMS USA: Version 11.0* [dataset]. Minneapolis, MN. https://doi.org/10.18128/D010.V11.0

Ruiz, N. G. (2014) A Joint Project of Brookings and JPMorgan Chase: The Geography of Foreign Students in U.S. Higher Education: Origins and Destinations.
https://www.brookings.edu/wp-content/uploads/2014/08/foreign_students_final.pdf (https://brookings.edu/interactives/the-geography-of-foreign-students-in-u-s-higher-education-origins-and-destinations/)

Shanghai Advanced Institute of Finance of Shanghai Jiao Tong University and Charles Schwab & Co., Inc. (2019) 2019 China Rising Affluent Financial Well-Being Index.
https://en.cafr.cn/Research/Details.aspx?menuid=16&id=132

Sun X. and Chen M. (2015) Inequality in educational opportunities of migrant children in China. In: Chan EKL, editor. Chinese migration and families-at-risk. Newcastle upon Tyne: Cambridge Scholars Publishing Ltd., 52–77.

Wang, B. and Zhou, Q. (2020) Climate change in the Chinese mind: An overview of public perceptions at macro and micro levels. *Wiley Interdisciplinary Reviews: Climate Change* 11(3), e639.

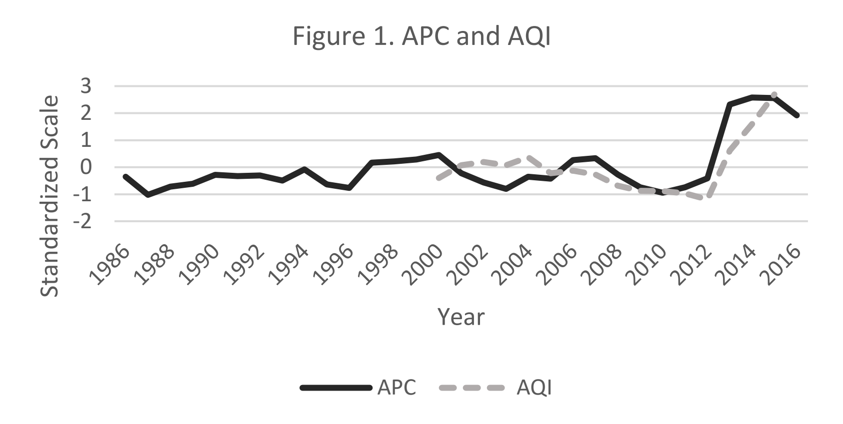

Figure 1. APC is the ratio of the number of articles based on Factiva search using the keywords "air pollution" and "China" to that only using the keyword "China". AQI is the average of the average daily air quality indices of Beijing, Shanghai, Guangzhou and Shenzhen.

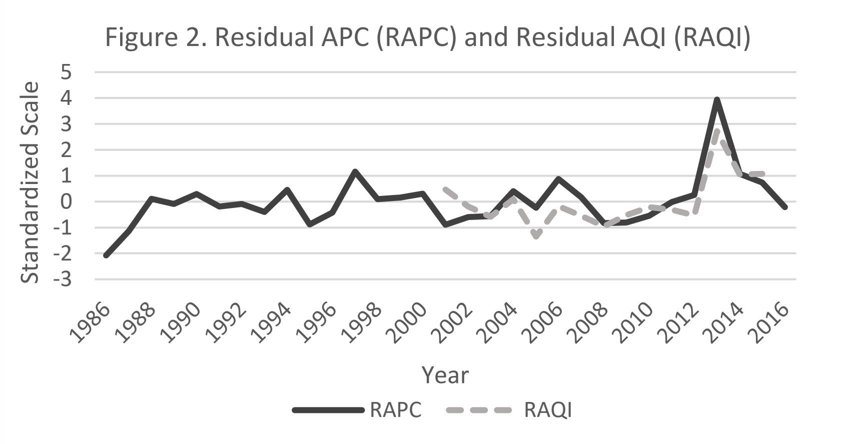

Figure 2. Both RAPC and RAQI are residuals of AR(1) models. APC is the ratio of the number of articles based on Factiva search using the keywords "air pollution" and "China" to that only using the keyword "China". AQI is the average of the average daily air quality indices of Beijing, Shanghai, Guangzhou and Shenzhen.

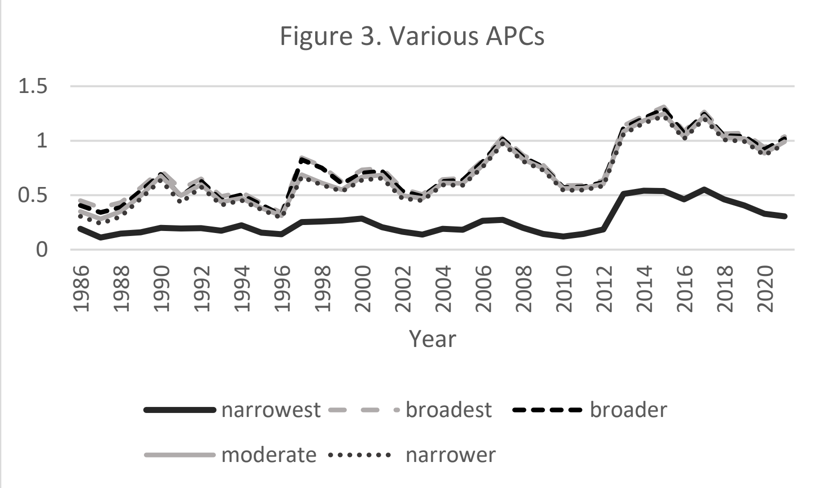

Figure 3. Broadest APC is the number of articles based on Factiva search using the broadest set of keywords "(air pollutants OR air pollution OR air quality OR toxic air OR carbon dioxide emissions OR smog OR soot OR toxic emissions OR particulate OR haze OR clean air) AND China". Broader APC excludes "soot" from the broadest set. Moderate APC excludes "haze" from the broadest set. Narrower APC excludes both "soot" and "haze" from the broadest set. Narrowest APC is the number of articles based on Factiva search using the keywords "air pollution AND China". All APCs are measured as a proportion of the number of articles based on Factiva search only using the keyword "China".

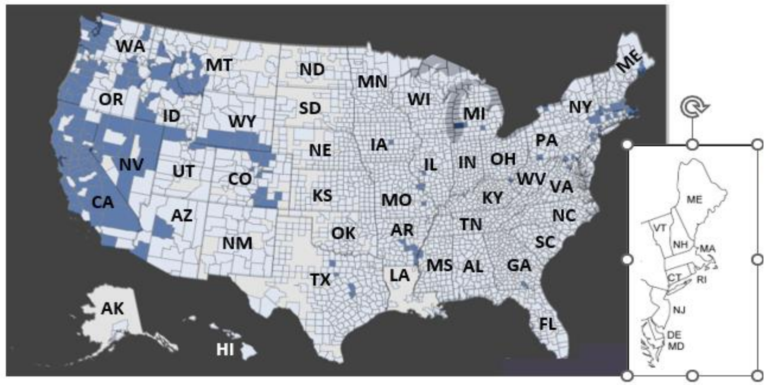

Figure 4. U.S. Counties with 1870 Chinese Population (in Dark Blue)

Table 1. Correlation of air pollution news and economic (related) variables

This table reports correlation coefficients of annual variables. The sample period is between 1960 and 2022. APC is the ratio of the number of articles based on Factiva search using the keywords “air pollution” and “China” to that only using the keyword “China”. RAPC is residual APC of an AR(1) model. GDP is China’s gross domestic product per capita. Trade is the sum of China’s exports and imports of goods and services per capita. CO2 is China’s carbon dioxide emission per capita. FOSSIL is China’s fossil fuel consumption per capita. The cyclical variable of X, denoted as “C-X”, represents the cyclical component of the original series (X), based on Hodrick-Prescott filter. The growth variable of X, denoted as “G-X”, is the natural logarithm of the ratio X(t) to X(t-1). ***, **, and * indicate 1%, 5%, and 10% levels of significance, respectively.

	RAPC	C-APC	C-GDP	G-GDP	C-TRADE	G-TRADE	C-CO2	G-CO2	C-FOSSIL
C-APC	0.682***								
C-GDP	0.097	0.149							
G-GDP	–0.002	–0.077	–0.143						
C-TRADE	0.266*	0.027	0.310**	0.112					
G-TRADE	–0.049	–0.119	–0.107	0.538***	0.008				
C-CO2	0.193	–0.127	0.083	0.188	0.805***	–0.011			
G-CO2	–0.163	–0.419***	–0.354**	0.375***	0.029	0.429***	0.032		
C-FOSSIL	0.181	–0.115	–0.055	0.190	0.815***	–0.004	0.951***	0.112	
G-FOSSIL	–0.082	–0.309*	–0.405***	0.245*	–0.062	0.174	0.000	0.932***	0.059

Table 2. Air pollution news, haze weather index, and thermal inversion

This table shows that RAPC is associated with weather conditions, but not with key economic variables. In Columns (1) – (4), the sample consists of Beijing (BJ), Shanghai (SH), and Guangdong (GD). In Columns (5) – (8), the sample is Beijing (BJ). APC is the ratio of the number of articles based on Factiva search using the keywords "air pollution" and "China" to that only using the keyword "China". RAPC is residual APC of an AR(1) model. TI is regional thermal inversion, defined in section 1.1.1. HWI is regional haze weather index, defined in section 1.1.1. X_TOP_⅓ (¼) an indicator that takes a value of 1 for a year for a region if the fraction of days for which X (representing TI or HWI) is in the top 20% of the daily distribution in that region for that year is in the top ⅓ (¼) of all years for that region. REGIONAL GDP PER CAPITA GROWTH is the contemporaneous growth of gross domestic product per capita of the region. CHINA'S TRADE PER CAPITA GROWTH is the contemporaneous growth of sum of China's exports and imports of goods and services per capita. Except the indicators, all other variables are standardised. In Columns (1) – (4), regional fixed effects are included, and the standard errors are based on clustering at the regional level. The estimated coefficients and the corresponding p-values (in parentheses) are reported. ***, **, and * indicate 1%, 5%, and 10% levels of significance, respectively.

Region(s):	BJ, SH, GD	BJ, SH, GD	BJ, SH, GD	BJ, SH, GD	BJ	BJ	BJ	BJ
RAPC version:	English	English	Chinese	Chinese	English	English	Chinese	Chinese
X:	TI	TI	TI	TI	HWI	HWI	HWI	HWI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	RAPC	RAPC	RAPC	RAPC	RAPC	RAPC	RAPC	RAPC
X_TOP_⅓	0.237**		0.401***		0.766**		1.022**	
	(0.047)		(0.001)		(0.023)		(0.044)	
X_TOP_¼		0.268**		0.606**		0.772**		0.970*
		(0.012)		(0.035)		(0.040)		(0.055)
REGIONAL GDP PER CAPITA GROWTH	–0.111	–0.095	0.008	0.063	0.036	–0.003	–0.220	–0.083
	(0.476)	(0.558)	(0.678)	(0.141)	(0.845)	(0.988)	(0.443)	(0.758)
CHINA'S TRADE PER CAPITA GROWTH	–0.064	–0.075	–0.042	–0.089	–0.048	–0.038	0.171	0.001
	(0.374)	(0.342)	(0.461)	(0.331)	(0.798)	(0.842)	(0.566)	(0.997)
CONSTANT	–0.079**	–0.067**	–0.134***	–0.152**	–0.264	–0.191	–0.333	–0.233
	(0.047)	(0.012)	(0.001)	(0.035)	(0.169)	(0.292)	(0.209)	(0.334)
Observations	123	123	72	72	41	41	24	24
R^2	0.033	0.034	0.026	0.054	0.134	0.111	0.195	0.179

Table 3. Correlation between initial regional characteristics and Chinese share of 1870 regional population

This table reports correlations between the Chinese proportion of the total population in U.S. regions as of 1870 and key regional economic and demographic variables at the beginning of our three main regression samples. Column (1) shows the correlation coefficients at the MSA level for the sample used in Table 4 for year 1996, the beginning of the sample period. Column (2) shows the correlation coefficients at the county level for the regression in Table 8 for the full county sample (beginning year is 1991), and column (3) for the immigration sample in the same table (beginning year is 1996). CIM refers to Chinese Immigration, and HPG to housing price growth. "IM" refers to the regression sample for which we have data on immigration, and "FULL" refers to the regression sample that includes all U.S. counties. ***, **, and * indicate the 1%, 5%, and 10% levels of significance, respectively.

	(1)	(2)	(3)
Outcome variables of interest in the regression	CIM	HPG	HPG
Year	1996	1991	1996
Regional level	MSA	County	County
Sample	IM	FULL	IM
Panel A: Key Economic and Demographic Variables Used as Explanatory Variables in Regressions			
Personal income per capita	0.034	–0.160*	–0.154
Growth of personal income per capita	0.113	–0.062	0.122
Population	0.014	–0.114	–0.118
Growth of population	–0.008	0.240***	0.342**
Panel B: Other Key Economic Variables Not Included in Regressions			
Labor-to-population ratio	–0.089	–0.151*	–0.168
Growth of labor-to-population ratio	–0.086	–0.004	–0.111
Employment	0.008	–0.116	–0.122
Growth of employment	0.002	0.110	0.111

Table 4. Ethnic ties, innovation in China's air pollution news and Chinese immigration to the U.S.

The table shows the effect of innovation in China's air pollution news on the inflow of Chinese immigrants to U.S. Metropolitan Statistical Areas (MSAs). The sample period is 1996–2020. In Columns (1)– (2), the results are based on the full sample. In Columns (3)– (4), we exclude those MSAs consisting of a county in the top five in terms of Chinese proportion of the 2000 total county population. $CIM_{j,t}$ is the number of Chinese who were admitted as an immigrant to MSA *j* in year *t*. Chinese (total) population of MSA *j* as of 1870 is estimated as the sum of 1870 Chinese (total) population of all counties in MSA *j*. In Columns (1) and (3), HC is an indicator for the top half of MSAs, in terms of the Chinese proportion of the 1870 total MSA population, among all MSAs in the sample. In Columns (2) and (4), HC is an indicator for a positive Chinese proportion of the total population in the MSA. APC is the ratio of the number of articles based on Factiva search using the keywords "air pollution" and "China" to that only using the keyword "China" for year t. RAPC is residual APC of an AR(1) model for year t. L1.RAPC is RAPC lagged once. L2.RAPC is RAPC lagged twice. MGDPG3 is the 3-year mean of the annual growth of the China's GDP per capita in years *t*, *t*-1, and *t*-2. MTRG3 is the 3-year mean of the growth of the sum of China's imports and exports per capita in years *t*, *t*-1 and *t*-2. CFRSH3 is the 3-year mean of the first difference of the ratio of the state-level freight with China to the state-level total freight with all countries in years *t*, *t*-1, and *t*-2, for which the state is where MSA *j* is located. L1.POP is the MSA total population in year *t*-1. PIG0 (POPG0) is the contemporaneous MSA-level annual nominal personal income growth (population growth). FUTURE_PIG5 (FUTURES_POPG5) is the MSA-level average annual nominal personal income growth (population growth) of the next five years or remaining years for which data are available. All variables are standardized. MSA and year fixed effects are included; robust standard errors are based on clustering at the MSA and year levels. The estimated coefficients and the robust p-values (in parentheses) are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of significance, respectively.

Sample:	Full	Full	x Top 5	x Top 5
1870 Chinese Share:	> Median	> 0	> Median	> 0
	(1)	(2)	(3)	(4)
	CIM	CIM	CIM	CIM
HC*RAPC	0.036*	0.041*	0.032*	0.029*
	(0.089)	(0.067)	(0.057)	(0.071)
HC*L1.RAPC	0.079**	0.080**	0.060***	0.060***
	(0.024)	(0.033)	(0.008)	(0.010)
HC*L2.RAPC	0.066**	0.082**	0.055*	0.056*
	(0.050)	(0.048)	(0.080)	(0.083)
HC*MGDPG3	0.052	0.067	0.041	0.039
	(0.145)	(0.246)	(0.211)	(0.211)
HC*MTRG3	–0.031	–0.021	–0.044*	–0.039
	(0.450)	(0.579)	(0.088)	(0.145)
CFRSH3	0.006	0.007	0.010	0.010
	(0.468)	(0.460)	(0.254)	(0.261)
L1.POP	1.219*	1.219*	1.111***	1.112***
	(0.080)	(0.079)	(0.000)	(0.000)
PIG0	0.026	0.027	0.003	0.004
	(0.363)	(0.344)	(0.893)	(0.876)
POPG0	–0.104*	–0.099*	–0.073**	–0.072**
	(0.081)	(0.080)	(0.029)	(0.027)
FUTURE_PIG5	0.004	0.006	0.012	0.012
	(0.809)	(0.721)	(0.527)	(0.504)
FUTURE_POPG5	0.163*	0.158**	0.089**	0.089**
	(0.053)	(0.049)	(0.012)	(0.012)
CONSTANT	–0.044	–0.045	–0.049**	–0.048**
	(0.550)	(0.551)	(0.014)	(0.016)
MSA Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
S.E. Cluster by MSA	Yes	Yes	Yes	Yes
S.E. Cluster by Year	Yes	Yes	Yes	Yes
Observations	817	817	749	749
R^2	0.914	0.915	0.920	0.920

Table 5. Language ties, innovation in air pollution news associated with Chinese regions, and Chinese immigration to the U.S.

The table shows the effect of innovation in air pollution news associated with Chinese regions on Chinese immigration of to U.S. Metropolitan Statistical Areas (MSAs). The sample period is 1996–2020. $CIM_{j,t}$ is the number of Chinese who admitted as an immigrant to MSA *j* in year *t*. Chinese (total) population of MSA *j* as of 1870 is estimated as the sum of 1870 Chinese (total) population of all counties in MSA *j*. HC is an indicator for the top half of MSAs, in terms of the Chinese proportion of the total population, among all MSAs in the sample. CAN is an indicator for the top half of MSAs, in terms of Cantonese speaking proportion of 2000 Chinese population (CPOP) in the odd-numbered columns or total population (TPOP) in the even-numbered columns, among all MSAs in the sample. R**X** is Region **X**'s version of RAPC, where **X** ={Guangdong, Beijing, Shanghai}. M**X**GDPG3 is Region **X**'s version of MGDPG3. The other variables are defined in Table 4. All variables are standardized. MSA and year fixed effects are included; robust standard errors are based on clustering at the MSA and year levels. The estimated coefficients and the robust p-values (in parentheses) are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of significance, respectively.

X:	Guangdong	Guangdong	Beijing	Beijing	Shanghai	Shanghai
CAN in relation to:	CPOP	TPOP	CPOP	TPOP	CPOP	TPOP
	(1)	(2)	(3)	(4)	(5)	(6)
	CIM	CIM	CIM	CIM	CIM	CIM
HC*RAPC	0.019	0.013	0.033	0.029	0.026	0.022
	(0.169)	(0.278)	(0.177)	(0.224)	(0.261)	(0.325)
HC*L1.RAPC	0.055*	0.055*	0.078*	0.077*	0.074*	0.075*
	(0.074)	(0.078)	(0.068)	(0.074)	(0.075)	(0.086)
HC*L2.RAPC	0.038*	0.038*	0.067*	0.069*	0.057	0.060
	(0.087)	(0.095)	(0.086)	(0.076)	(0.121)	(0.119)
CAN*R**X**	0.091**	0.091**	–0.011	–0.002	0.034	0.035
	(0.021)	(0.014)	(0.251)	(0.845)	(0.454)	(0.446)
CAN*L1.R**X**	0.045	0.042*	–0.022	–0.016	0.012	0.009
	(0.106)	(0.094)	(0.227)	(0.380)	(0.780)	(0.837)
CAN*L2.R**X**	0.090**	0.079**	–0.019	–0.016	0.036	0.028
	(0.022)	(0.013)	(0.267)	(0.328)	(0.430)	(0.539)
HC*MGDPG3	0.065	0.069	0.098**	0.101**	0.091*	0.092*
	(0.108)	(0.113)	(0.040)	(0.037)	(0.051)	(0.051)
CAN*M**X**GDP3	0.011	–0.005	–0.093	–0.090*	0.098	0.065
	(0.828)	(0.921)	(0.107)	(0.071)	(0.490)	(0.646)
HC*MTRG3	–0.059	–0.059	–0.023	–0.018	–0.057	–0.053
	(0.115)	(0.122)	(0.347)	(0.466)	(0.177)	(0.237)
CFRSH3	0.004	0.005	0.002	0.003	0.005	0.005
	(0.442)	(0.429)	(0.747)	(0.730)	(0.492)	(0.460)
L1.POP	1.024	1.032	0.940	0.910	1.013	1.000
	(0.110)	(0.106)	(0.119)	(0.125)	(0.102)	(0.104)
PIG0	0.015	0.015	0.021	0.025	0.022	0.022
	(0.403)	(0.405)	(0.304)	(0.258)	(0.283)	(0.277)
POPG0	–0.092	–0.094	–0.094	–0.097	–0.094	–0.095
	(0.111)	(0.105)	(0.117)	(0.114)	(0.117)	(0.120)
FUTURE_PIG5	0.003	0.002	–0.006	–0.002	0.003	–0.000
	(0.853)	(0.892)	(0.751)	(0.925)	(0.889)	(0.996)
FUTURE_POPG5	0.177*	0.171*	0.165*	0.169*	0.163*	0.165*
	(0.061)	(0.064)	(0.074)	(0.071)	(0.069)	(0.069)
CONSTANT	–0.012	–0.012	–0.000	0.004	0.005	0.005
	(0.860)	(0.855)	(0.996)	(0.954)	(0.935)	(0.940)
MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
S.E. Cluster by MSA	Yes	Yes	Yes	Yes	Yes	Yes
S.E. Cluster by Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	717	717	717	717	717	717
R^2	0.927	0.927	0.926	0.926	0.926	0.926

Table 6. Education links, innovation in air pollution news associated with Chinese cities, and influx of students to the U.S.

This table shows the effect of innovation in air pollution news associated with Chinese cities on influx of international students to the U.S. In Columns (1) and (3), the sample period is 2001-2020. In Columns (2) and (4), the period 2008–2012 is excluded. The dependent variable (ISTU) is the number of first-time freshmen whose permanent address is outside the U.S., for institution j in year t. WRAPC is a weighted average of RLAPC (local version of RAPC for the whole China). LAPC is the ratio of the number of articles based on Factiva search using the keywords "air pollution" and the name of the Chinese city to the number of articles that contain the name of the Chinese city for year t. RLAPC is residual LAPC of an AR(1) model for year t. L1.RLAPC is RLAPC lagged once. L2.RLAPC is RLAPC lagged twice. In Columns (1)– (2), the weight of RLAPC of Chinese city k (***weight***) for MSA m where institution j is located is the fraction of all Chinese international studentsfor MSA m who come from Chinese city k. In Columns (3)– (4), WRAPC is constructed by randomly pairing up city k's RLAPC and any other city's ***weight*** (among the other 14 cities) for the same MSA (without replacement). CFRSH3 is the 3-year mean of the first difference of the ratio of the state-level freight with China to the state-level total freight with all countries in years *t*, *t*-1, and *t*-2, for which the state is where institution *j* is located. ALLSTU is the number of first-time freshmen for institution j in year t. The other variables are defined in Table 4. All variables are standardized. Institution and year fixed effects are included; robust standard errors are based on clustering at the institution and year levels. In Columns (1)–(2), based on the actual sample, the estimated coefficients and the robust p-values (in parentheses) are reported. In Columns (3)–(4), mean, the 0.5st percentile, and the 99.5th percentile of the 1000 coefficients of the bootstrapped results are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of significance, respectively.

weight	actual	actual	random	random	random	random	random	random
Period	Full	x 08-12	Full	Full	Full	x 08-12	x 08-12	x 08-12
	(1)	(2)	(3): mean	(3): p0.5	(3): p99.5	(4): mean	(4): p0.5	(4): p99.5
	ISTU	ISTU	ISTU	ISTU	ISTU	ISTU	ISTU	ISTU
WRAPC	0.032**	0.035**	0.0034	–0.0021	0.0104	0.0034	–0.0029	0.0122
	(0.011)	(0.027)						
L1.WRAPC	0.023**	0.027**	0.0053	–0.0001	0.0092	0.0060	–0.0005	0.0123
	(0.020)	(0.033)						
L2.WRAPC	0.048***	0.053***	0.0019	–0.0051	0.0078	0.0029	–0.0049	0.0132
	(0.003)	(0.007)						
ALLSTU	0.277**	0.363***	0.2780	0.2777	0.2782	0.3656	0.3648	0.3659
	(0.012)	(0.008)						
CFRSH3	0.007	0.021*	0.0066	0.0066	0.0067	0.0199	0.0198	0.0200
	(0.346)	(0.064)						
L1.POP	–0.135	–0.181	–0.1077	–0.1082	–0.1070	–0.1425	–0.1434	–0.1417
	(0.189)	(0.139)						
PIG0	0.026	0.043	0.0269	0.0267	0.0272	0.0429	0.0426	0.0431
	(0.132)	(0.136)						
POPG0	0.032	0.042	0.0376	0.0372	0.0378	0.0519	0.0512	0.0524
	(0.166)	(0.178)						
FUTURE_PIG5	0.045	0.057	0.0457	0.0455	0.0459	0.0581	0.0578	0.0583
	(0.120)	(0.121)						
FUTURE_POPG5	–0.024	–0.001	–0.0241	–0.0243	–0.0235	0.0049	0.0045	0.0057
	(0.418)	(0.984)						
CONSTANT	–0.011	–0.029	–0.0194	–0.0196	–0.0188	–0.0281	–0.0284	–0.0276
	(0.411)	(0.154)						
Fixed Effects:								
(a) Institution	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(b) Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S.E. Cluster by:								
(a) Institution	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(b) Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,910	24,454	—	—	—	—	—	—
R^2	0.725	0.723	—	—	—	—	—	—

Table 7. innovation in air pollution news associated with Chinese regions and *Gaokao* student number

This table shows the effect of innovation in air pollution news associated with Chinese regions on the number of local students registered for the National College Entrance Examination (*Gaokao*). The sample consists of all 31 provinces, municipalities and autonomous regions (hereafter "regions") under direct jurisdiction in China over the period 2006–2020. $EXAMSTU_{j,t}$ is the number of students who registered for *Gaokao* in region *j* in year *t*. In Columns (1)– (2), Z is the AR(1) residual of the ratio of regional APC (LAPC) to APC. $LAPC_{j,t}$ is regional APC and constructed in the same way as APC, except that we replace the keyword "China" by the name of region *j* and conduct the Factiva search in Simplified Chinese. RLAPC and its lags are constructed in the same way as RAPC and its lags. In Columns (3)– (4), Z is regional RAPC minus the overall RAPC for China. $GRADSTU_{j,t}$ is the number of students who graduated from a high school in region *j* in year *t*. MLGDPG3 is region j's version of MGDPG3. The other variables are defined in Table 4. All variables are standardized. Region fixed effects are included; robust standard errors are based on clustering at the region and year levels. The estimated coefficients and the robust p-values (in parentheses) are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of significance, respectively.

Z:	Residual of LAPC/APC		Residual LAPC – Residual APC	
	(1)	(2)	(3)	(4)
	EXAMSTU	EXAMSTU	EXAMSTU	EXAMSTU
Z	–0.031**	–0.029**	–0.024**	–0.022*
	(0.010)	(0.012)	(0.049)	(0.065)
L1.Z	–0.022***	–0.021***	–0.017**	–0.015*
	(0.001)	(0.010)	(0.037)	(0.069)
L2.Z	–0.016**	–0.015*	–0.016*	–0.015
	(0.020)	(0.066)	(0.056)	(0.112)
MLGDPG3		–0.015		–0.018
		(0.531)		(0.472)
MGDPG3		0.007		0.008
		(0.774)		(0.771)
MTRG3		0.016		0.020
		(0.388)		(0.354)
GRADSTU	0.896***	0.908***	0.913***	0.925***
	(0.000)	(0.000)	(0.000)	(0.000)
CONSTANT	0.000	0.000	0.000	0.000
	(1.000)	(1.000)	(1.000)	(1.000)
Observations	464	464	464	464
R^2	0.982	0.982	0.981	0.981

Table 8. Ethnic and education ties, innovation in China's air pollution news and housing price growth in U.S. counties

This table shows the effect of innovation in air pollution news concerning China on residential housing price growth in U.S. counties. In Columns (1)– (2), the sample consists of the counties corresponding to those in the full immigration sample in Table 4. In Columns (3)– (4), the sample consists of the counties corresponding to those in the full student sample in Table 6. In Columns (5)– (6), the sample consists of all counties that have 1870 population information; the sample period is 1991–2019. The dependent variable is the county-level annual nominal housing price growth (HPG) for year t. In Columns (2) and (6) [(1) and (5)], HC is a dummy variable that has a value of 1 if the Chinese proportion of the 1870 county-level total population is positive [above the median], and 0 otherwise. In Column (4) [(3)], HC is a dummy variable that has a value of 1 if the average of a measure of Chinese students' ties of all institutions in the county is positive [above the median), and 0 otherwise. The measure of Chinese students' ties CS_j is a product of two ratios: (i) the ratio of institution j's number of new international students to the number of freshmen as of the year 1986, and (ii) the fraction of Chinese international students to all international students for the MSA where the institution is located. PIG0 (POPG0) is the contemporaneous county-level annual nominal personal income (population) growth. FUTURE_PIG5 (FUTURE_POPG5) is the county-level average annual nominal personal income (population) growth of the next five years or remaining years for which data are available. L1.POP is the county total population in year *t*-1. CFRSH3 is the 3-year mean of the first difference of the ratio of the state-level freight with China to the state-level total freight with all countries in years *t*, *t*-1, and *t*-2, for which the state is where the county is located. The other variables are defined in Table 4. All variables are standardized. County and year fixed effects are included; robust standard errors are based on clustering at the county and year levels. The estimated coefficients and the robust p-values (in parentheses) are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of significance, respectively.

Sample:	Immigration	Immigration	Student	Student	Full	Full
1870 Chinese Share:	> Median	> 0	> Median	> 0	> Median	> 0
	(1)	(2)	(3)	(4)	(5)	(6)
	HPG	HPG	HPG	HPG	HPG	HPG
HC*RAPC	0.302***	0.149***	0.054*	0.039**	0.277***	0.174***
	(0.000)	(0.000)	(0.087)	(0.027)	(0.000)	(0.000)
HC*L1.RAPC	0.224***	0.104*	0.068***	0.018	0.194***	0.133***
	(0.009)	(0.061)	(0.000)	(0.133)	(0.001)	(0.001)
HC*L2.RAPC	–0.039	–0.037	–0.049*	–0.026	–0.013	0.008
	(0.574)	(0.426)	(0.051)	(0.211)	(0.836)	(0.836)
HC*MGDPG3	–0.658***	–0.492***	–0.072*	–0.185***	–0.694***	–0.517***
	(0.000)	(0.000)	(0.079)	(0.000)	(0.000)	(0.000)
HC*MTRG3	0.722***	0.496***	–0.133***	0.097**	0.673***	0.521***
	(0.000)	(0.000)	(0.005)	(0.024)	(0.000)	(0.000)
PIG0	0.116***	0.117***	0.113***	0.118***	0.084***	0.083***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
POPG0	0.292***	0.297***	0.299***	0.289***	0.257***	0.256***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
FUTURE_PIG5	0.002	–0.002	0.010	0.024	–0.020	–0.020
	(0.943)	(0.924)	(0.717)	(0.371)	(0.250)	(0.250)
FUTURE_POPG5	–0.088**	–0.075**	–0.103**	–0.093**	0.038	0.044*
	(0.018)	(0.039)	(0.028)	(0.046)	(0.120)	(0.091)
L1.POP	0.606***	0.626***	–0.008	0.154	0.620***	0.624***
	(0.007)	(0.008)	(0.982)	(0.667)	(0.000)	(0.000)
CFRSH3	–0.010*	–0.011*	–0.011	–0.009	–0.011*	–0.011
	(0.097)	(0.091)	(0.304)	(0.372)	(0.072)	(0.103)
CONSTANT	–0.064***	–0.060***	0.000	0.000	–0.020***	–0.017**
	(0.000)	(0.000)	(0.981)	(0.843)	(0.005)	(0.011)
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
S.E. Cluster by County	Yes	Yes	Yes	Yes	Yes	Yes
S.E. Cluster by Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,641	15,641	12,070	12,070	55,329	55,329
R^2	0.612	0.607	0.618	0.614	0.416	0.416

Table 9. Ethnic ties, innovation in China's air pollution news and (1) Chinese emigration and US housing ownership, and (2) growth of rent-to-income ratio and growth of house-value-to-income ratio (using household data)

This table shows the effect of China's air pollution news on (1) housing ownership of new Chinese immigrants in U.S. counties, and (2) growth of rent-to-income ratio and growth of house-value-to-income ratio in U.S. counties. The sample period is 2005–2019, based on the U.S. Census's American Community Survey 1-year estimates. $CHO_{j,t}$ is the number of respondents who were born in China, entered the U.S. in year t, and were living in an owner-occupied housing unit in county *j* in year *t* when the survey was conducted. $CPOP_{j,t}$ ($FPOP_{j,t}$) [$TPOP_{j,t}$] is the China-born population (foreign-born population) [total population] in county *j* in year *t.* RENT/INCOME MEDG is the growth of median rent-to-income ratio in county *j* from year *t*-1 to year *t*. HOUSE-VALUE/INCOME MEDG is the growth of median house-value-to-income ratio in county *j* from year *t*-1 to year *t*. HC is an indicator for those counties with a positive Chinese proportion of the 1870 county-level total population. The local control variables are defined in Table 8. The other variables are defined in Table 4. We perform logistic transformation of the proportion dependent variable for normality. All variables are standardized. County and year fixed effects are included; robust standard errors are based on clustering at the year level. The estimated coefficients and the robust p-values (in parentheses) are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of significance, respectively.

	(1a)	(1b)	(1c)	(2a)	(2b)
	$\frac{CHO}{CPOP}$	$\frac{CHO}{FPOP}$	$\frac{CHO}{TPOP}$	$\frac{RENT}{INCOME}\ MEDG$	$\frac{HOUSE\ VALUE}{INCOME}\ MEDG$
HC*RAPC	0.039**	0.039**	0.040**	0.039**	0.129***
	(0.044)	(0.034)	(0.028)	(0.012)	(0.002)
HC*L1.RAPC	0.040*	0.035	0.036	0.030	0.077**
	(0.094)	(0.152)	(0.142)	(0.521)	(0.010)
HC*L2.RAPC	–0.041	–0.040	–0.041	0.028	–0.060*
	(0.248)	(0.276)	(0.275)	(0.371)	(0.083)
HC*MGDPG3	–0.008	–0.011	–0.011	0.124**	–0.297***
	(0.905)	(0.880)	(0.882)	(0.038)	(0.004)
HC*MTRG3	–0.015	–0.018	–0.018	–0.107	0.244**
	(0.582)	(0.499)	(0.498)	(0.250)	(0.022)
PIG0	0.015	0.017	0.016	–0.013	0.073**
	(0.278)	(0.217)	(0.226)	(0.529)	(0.021)
POPG0	0.014	0.019	0.019	0.054	0.252***
	(0.434)	(0.262)	(0.261)	(0.148)	(0.000)
FUTURE_PIG5	–0.020	–0.016	–0.015	–0.013	0.050
	(0.393)	(0.450)	(0.468)	(0.517)	(0.115)
FUTURE_POPG5	–0.057	–0.063*	–0.064*	–0.110***	0.016
	(0.108)	(0.083)	(0.079)	(0.002)	(0.851)
L1.POP	1.153***	1.167***	1.200***	0.274	1.527***
	(0.000)	(0.000)	(0.000)	(0.240)	(0.004)
CFRSH3	–0.006	–0.007	–0.007	0.021*	0.016
	(0.659)	(0.615)	(0.615)	(0.090)	(0.535)
CONSTANT	–0.001**	0.000	0.000	–0.008	–0.006
	(0.023)	(0.722)	(0.716)	(0.114)	(0.198)
County Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
S.E. Cluster by County	Yes	Yes	Yes	Yes	Yes
S.E. Cluster by Year	Yes	Yes	Yes	Yes	Yes
Observations	5,980	5,994	5,994	5,496	5,500
R^2	0.373	0.367	0.380	0.054	0.296