

# *Effect of Health Care on Population Flow in China: A Modern Panel Data Approach*

Heng Yu<sup>1, a, \*</sup>

<sup>1</sup>*School of Management, Xi'an Jiaotong University, 28 West Xianning Road, Xi'an, China*

*a. 942136943@qq.com*

*\*corresponding author*

**Keywords:** migration, health care, crowd behaviour, social media

**Abstract:** The underlying principle behind rural-to-migration in China can be summarized by the inequality of social welfare distribution like health care and education resources amongst different cities and provinces, and China's huge population base makes the population flow and crowd behavior even more convoluted. This paper provides theoretical and empirical evidence on the effect health care can have on population net flow for a given city (positive net flow denotes population net inflow and vice versa). The panel data we use was collected from the annual statistic yearbook of 19 different Chinese cities on a timespan of 13 years from 2005 to 2017. Cities of different development levels are selected to make an unbiased sample for net population flow. Our theory suggests that the effect of social economic development (in this case we take health care development as proxy variable) varies between different development phases and hence a three-point economic development theory on population flow has been proposed through induction. Through pooled panel OLS and Fixed Effect regression analysis, we prove that our model on the relationship between city development and population net flow is correct. For cities with either early or late development stages, our analysis suggests that social effect through internet plays an important role upon individual migration decisions in China.

## 1. Introduction

Due to its huge population base, internal migration in China has remained the largest population flow across the world since the late stages of reform and opening up in the 1980s [1]. The advent of modern technology, especially the instant-communication technology brought by the development of internet and mobile network makes study on the nature of population flow even harder because of the well-developed social media network which has already been woven into our lives. While it is easy to see that social media can have positive effects on migration decision through network effect and acting as a palliative against short-term depression, the negative effects like community exclusion and the adverse emotions channelled through social media still remains untapped for researchers [2][3]. In developed countries like Holland, study has found that some highly educated workers accept jobs below their education level and skill, which leads to the crowding-out effect for less educated workers [4].

Various researches have been carried out on rural-to-urban migration concerning the social economic effect and the health consequences on migrants themselves. Yang Song and Wenkai Sun found that effect of migration on health status of migrants is strongly positive in the short run (1 year) probably due to the increase in income level. But no significant health gain has been observed over longer terms (>3 years) probably due to mental health issues caused by metropolis working pressure and community [5]. For mental health issues of migrants, Yao Lu found that mental health issues are prevalent amongst Indonesian migrant labour and the effect differs between men and women: men tend to externalize stress through smoking while women tend to internalize and display depressive symptoms [6]. Ghuncha Firdaus suggested that poor mental health risk is significantly high amongst unskilled, single, divorced or separated migrants, while longer year of migration, together with younger age and higher income appeared to be positively related with mental health [7]. Studying the effect of ethnic minority groups in Canada, researchers have found that stressors like neurocognitive and physical impairments, HIV stigma and discrimination, fear of deportation and a lack of mental health palliative services all leads to social isolation of minority migrants which worsens the mental health care conditions for migrants [8][9]. The ramifications of rural-to-urban migration will extend to the close relatives in sending areas regarding to the health and mental status of left behind children (which amounts to 7 million in 2018) and health conditions of left behind parents. Xiang Ao et al. found that having one more migrating child will increase the risk of poor health condition by 8%, or higher for those with single child migrating [10]. While the original migrating purpose is to increase income for family and provide better living conditions for children, Chenyue Zhao et al. found that due to parent-children separation, parental migration resulted in worsened mental health conditions for children, and this parental absence effect could hardly be mitigated by the presence of grandparents as caregivers [11]. For younger children migrating with their parents, the depressive psychological effect caused by social discrimination and other means of victimization stress may be mitigated by local and trans-local ties with communities, and thus adolescents do not necessarily suffer from worse psychological health than their urban native counterparts [12].

This paper will discuss the effect health care have on migration decision, here as individual data is not available, we will study the effect on net population flow which will be positive for net inflow and vice versa. We collected data from 19 Chinese cities on a timespan of 13 years (2005 to 2017) based on geographical location, sending cities and receiving cities are selected to compile a representative sample as contemporary migration to bigger cities, as research shows, will increase income inequality amongst different cities [13]. Both economic and social factors are taken into account in our model below as migration group shares a huge gap in education level and social economic status which leads to consumption gap of up to 46% between the two groups, whereby up to 20% of the inequality can be explained by difference in education level [14]. As difference in education level and wage gap usually means the gap between social economic statuses (SES) in the social hierarchy, this difference becomes one of the main social stressors and as result migrants with lower social economic status are more likely to catch cardiovascular diseases and mental disorder for both the migrants themselves and their children, as shown by Italian researchers [15][16]. Marriage separation, which is more often seen amongst rural-to-urban migrants in China than divorce, makes things worse as long-term physical separation increases the sense of insecurity and the risk of suicide as a result of distress, which are both detrimental to family relationship and personal health condition [17]. Feedback through social media, as part of the social factors, was discussed in our model and the regression results, but due to lack of individual tracking data, we cannot analyse the effect of feedback on individual migration decision in this paper, but the effect itself is worth mentioning as the role social media plays in our life increases.

The rest of the paper is organized as follows. Section two introduces our model of health care

effect on population flow. Section three reviews the empirical evidence we gathered through panel data and regression models. Section four presents a detailed discussion on the empirical results and our conclusion on this research, together with suggestions on further research of this topic.

## 2. A New Model for Population Flow

The fluent nature of population flow makes analysis on the crowd behaviour for certain cities more convoluted than consumable flow in business world, one has to take into account both economic factors and sociology effect woven into the complicated human society. In a world where social media permeates into every corner of our life, one cannot underestimate the role internet and modern technology plays upon our life in both real world and the virtual society where people make friends, contact, and share their feelings. So both economic factors and sociological effects will be taken into consideration in the methodology on population flow.

The basic assumption for our theory is the existence of economic disparity between different cities. Without this assumption, if doppelganger cities prevail, there will only be limited natural population flow between cities which will contradict the following theories. To avoid unnecessary discussion, the population flow of the cities we will be discussing will be in the form of net flow, which can be positive if there are population inflows into the city and negative if population outflows exist. We assume that certain correlation exists between the population and development of one city. City capacity to sustain a certain number of people depends on the development level which will be the reflected in economic factors like GDP and basic infrastructure conditions. Similarly, the population of a city will effect city development in a similar way: more population provide labour for construction and industries, contribute more to local tax authorities which becomes government funds, and the money accumulated in the government coffer will again be distributed amongst social economy, all of which forms a cycle of social development.

Assume that city capacity is denoted by  $C_{it}$ , which denotes the scale of population city I can hold in time period  $t$ . Two deciding factors are denoted by  $E_{it}$  and  $S_{it}$  with the same index as city capacity to represent the economic development and social effect. Now we can express city capacity as:

$$C_{it} = G(E_{it}, S_{it}), \text{ with } \frac{\partial C}{\partial E}, \frac{\partial C}{\partial S} > 0 \quad (1)$$

$G$  is a function of  $E$  and  $S$ . The two partial derivatives are strictly positive because social and economic development will build up city capacity as disused above. Here the larger  $E$  and  $S$  values are, the stronger their effects are. However, on reaching the saturation level the second derivative of two variables will be close to zero. Saturation here means the highest level of capacity city might reach under temporal science and technology development.

The underlying principle behind population outflow may be explained by the following: current city development or capacity cannot meet the utility demand of the local residents. Similarly, massive population inflow can be explained by opportunity abundance which is the result of population level below city capacity.

Population net flow, as described above, can be represented in a way similar to city capacity, where  $F$  is a projection function transforming economic and social factors to population net flow:

$$I_{it} = F(E_{it}, S_{it}) \quad (2)$$

Take a deep look into this Economic factor, and we can split it into the main effect ( $M$ ) which

represents the effect of city economy on its own and gravity effect (G) which takes into account its surrounding areas. Gravity here means the development gap between city and its vicinity, for which positive or negative gap will result in pulling-in or pushing-out forces similar to that of gravity centre.

$$E_{it} = E(M_{it}, G_{it}) \quad (3)$$

And,

$$\frac{\partial E}{\partial M} = \varphi(M, G), \frac{\partial E}{\partial G} = \phi(M, G) \quad (4)$$

This is because main effect and gravity effect are strongly correlated as gravity effect will be amplified as the city develops and the stronger gravity effect is, the more opportunities the gravity center will get through the inflow of peripheral population. In those undeveloped remote areas, where both economic factors E and social effects S remain low, increase in either economy or social development will not yield significant changes to the economic status, and hence:

$$\frac{\partial E}{\partial M} \in (-\Gamma, \Gamma), \text{ when } m \text{ remains low} \quad (5)$$

Here we call  $(-\Gamma, \Gamma)$  the inactivity range. Since the main effect will be channeled through E function and F function, this results in inactivity of population flow for the economic backwaters, as social effect will not have much influence on an immature society. The impact on net population flow will be presented as:

$$\frac{\partial I}{\partial M} = \frac{\partial I}{\partial E} \frac{\partial E}{\partial M} + \frac{\partial I}{\partial S} \frac{\partial S}{\partial M} \quad (6)$$

So, when M remains low:

$$\frac{\partial I}{\partial M} \in (-\tau, \tau) \quad (7)$$

Further look into the main effect, we can break it down into three factors regarding the economy, health care level and education development.

$$M_{it} = M(\text{ECO}_{it}, \text{EDU}_{it}, \text{MED}_{it}) \quad (8)$$

These three are cornerstones on which the society is built. However, serious correlation exists as increase in any one of these will case a ripple effect. Take economy as an example. Increase in local economy (which is often represented as the local GDP or GDP per capita) will increase government expectation for the next year. This will result in a rise for government budget and funds available for disposal as tax income increases. Suppose distribution of funds amongst investments remains constant, increase in funds will also provide more budgets for health care and education development. Just as what ripple effect means, the increase of education and health care budget will also have a second effect on the whole economic system as health care development reduces mortality rate, education development betters resource allocation by reducing the number of students per teacher and providing more funding for research and development. Reduction of mortality rate and better education resources will provide solid ground for migrants with high education given that education

of children and health care facilities has deciding impact on migrant decisions. Inflow of migrants with high education will further contribute to local economy and hence the cycle forms.

On condition that budget allocation will not change seriously in the short term (less than five years), we can create the binding condition for the three factors to simplify the model.

$$\frac{\overline{MED}_{iT}}{\overline{ECO}_{iT}} = C_{1iT} \quad (9)$$

And

$$\frac{\overline{EDU}_{iT}}{\overline{ECO}_{iT}} = C_{2iT} \quad (10)$$

Here we average the three variables to eliminate periodic fluctuations,  $C_{1iT}$  and  $C_{2iT}$  represents the proportion of fund the two factors have taken during period T which remains constant periodically. We assume that no significant change in any development rate within the period, such as GDP growth, hence from this assumption we can obtain time t stats on accrued basis. Thus we can simplify the model through elimination of ECO and EDU and write main effect as:

$$M_{it} = M(MED_{it})_T \quad (11)$$

As gravity effect can be expressed through difference value of economic factors (in this case denoted by MED), as geometric location is taken into social economic development factors, it can also be represented as function of MED. Hence we can further simplify the model by writing economy effect E as:

$$E_{it} = \varphi(MED_{it})_T \quad (12)$$

Social effect is rather complicated given the unpredictable nature of human crowd effect. For migration model, we have to take into consideration factors like population structure, composition of local communities, economic development level, social economic status, transport and even the use of social media. What is even more baffling is that the stronger local economy and city capacity is, the stronger social effect will be, probably amplified through the channel of population. The above factors, by nature, can be categorized into three groups: psychological factors, network effect and social economic effect. As mental disorder of migrants in larger cities are often caused by social economical effect like stress from long working hours and relevantly low wage and social status, psychological factors can be represented through projection of economic factors like MED, so simplification of model can be presented as:

$$S_{it} = S(P_{it}, N_{it}, E_{it}) = S(P_{it}, N_{it}, \varphi(MED_{it})_T) = \phi(N_{it}, MED_{it}) \quad (13)$$

So the overriding theory can be presented as:

$$I_{it} = F(E_{it}, S_{it}) = \Theta(MED_{it}, N_{it}) \quad (14)$$

$$C_{it} = G(E_{it}, S_{it}) = \Gamma(MED_{it}, N_{it}) \quad (15)$$

First partial derivative as:

$$\frac{\partial I_{it}(E_{it}, S_{it})}{\partial E_{it}} = \frac{\partial I_{it}(\varphi(MED_{it})_T, \phi(N_{it}, MED_{it}))}{\partial \varphi(MED_{it})_T} = \frac{\partial \Theta(MED_{it}, N_{it})}{\partial \varphi(MED_{it})_T} = \Theta_1(\varphi(MED_{it})_T, N_{it}) \quad (16)$$

$$\frac{\partial G_{it}(E_{it}, S_{it})}{\partial E_{it}} = \frac{\partial G_{it}(\varphi(MED_{it})_T, \phi(N_{it}, MED_{it}))}{\partial \varphi(MED_{it})_T} = \frac{\partial \Gamma(MED_{it}, N_{it})}{\partial \varphi(MED_{it})_T} = \Gamma_1(\varphi(MED_{it})_T, N_{it}) \quad (17)$$

The simplified forms will be used for discussion over partial effects below.

For G function of capacity variable, as city capacity by nature reflects the economic development from an objective perspective, the effect of health care development (as stated before to be proxy for economic status) will remain strictly positive across time and development stages, given that health care development increases resident welfare and happiness by lowering mortality rate, increasing disposable wages and providing promising chances for education. Thus we will have:

$$\frac{\partial G_{it}(E_{it}, S_{it})}{\partial MED_{it}} = \frac{\partial \Gamma(MED_{it}, N_{it})}{\partial MED_{it}} > 0 \quad (18)$$

For population net inflow we will switch to the second expression as it breaks down the effect of economy into environmental effect and sociological effect:

$$\frac{\partial I_{it}(E_{it}, S_{it})}{\partial MED_{it}} = \frac{\partial I_{it}(\varphi(MED_{it})_T, \phi(N_{it}, MED_{it}))}{\partial MED_{it}} = \varphi'(MED_{it})_T + \frac{\partial \phi(N_{it}, MED_{it})}{\partial MED_{it}} \quad (19)$$

While as explained before,

$$\varphi'(MED_{it})_T > 0, \forall MED_{it} \notin \text{inactivity range} \quad (20)$$

The effect of health care on social effect will depend on the development stage of the society. As discussed in social effect, upon reaching the higher stages of development, stress and other psychological effects will take the helm as pressure mounting. These negative effects are channeled through social media or community congregation back home, and this is called the feedback mechanism. Adverse feedback from migrants will thwart their effort of migrating because of the dampened future expectations from the adverse feedback from those already migrated, also, as a result of long-term trauma and torment, migrants with lower education level (which is about 90% without high education in China, and who often end up with lower social economic status) are likely to catch mental diseases or stagger from the metropolitan cage, this results in what we called urban retreat. When MED is high, the effect can be summarized as:

$$\frac{\partial \phi(N_{it}, MED_{it})}{\partial MED_{it}} < 0 \quad (21)$$

In the early stages of development when social effect is restricted compared to economic effects, we will have, like the inactivity range,

$$\frac{\partial I}{\partial MED} \in (-\epsilon, \epsilon) \quad (22)$$

With,

$$\frac{\partial\phi(N_{it},MED_{it})}{\partial N_{it}} < -\epsilon \quad (23)$$

So, without detailed discussion into social economic development, we cannot derive the ultimate effect health care will have upon population net flow, with both positive effects in economy and negative effects in late stages of fully-fledged society.

Ignoring the geographical location, now we start from a small undeveloped town Alpha. The following analysis will be on an induction basis to form a city development model. As our town is in the primitive stages of development, the demand of people will stay at subsistence level so that demand here in Alpha will not exceed the city capacity. With the development of local economy, the advent of internet and mobile network provides locals with information of other parts of the world. At this stage people's utility demand will increase exponentially given the network effect provided by social media, utility demand of locals will finally reach the first milestone where development of local economy and infrastructure cannot provide local people with their marginal demand of utility, hence population outflow will happen at this time spot, called E1. And this can be explained by the low network social effect in the early stages of development where social economic conditions take hold of decision making. But with the introduction of social media and instant messaging technology, the adverse effect of network takes place because local economy was not enticing enough to provide people with enough utility level:

$$\frac{\partial\phi(N_{it},MED_{it})}{\partial N_{it}} < -\epsilon \text{ and } \frac{\partial I}{\partial MED} \in (-\epsilon, \epsilon) \quad (24)$$

As mentioned above. Thus from E1 we can see net population outflow in Alpha other than the natural increase of population through birth and the rate is going to remain high until the time come when local economy is strong enough to take external investment and social effect becomes steady. From then with the investment of external market, local economy provides better chances for local people and hence the population outflow goes down. In addition, inflow of external population due to investment activities and speculation activities will finally lead to the establishment of migrant-based local community, and then, what we discussed before as network effect will result in population inflow which further counters local population outflow. This can be expressed as:

$$\frac{\partial\phi(N_{it},MED_{it})}{\partial MED_{it}} > 0 \quad (25)$$

And

$$\frac{\partial\phi(N_{it},MED_{it})}{\partial N_{it}} > 0 \quad (26)$$

The time spot when population outflow is fully offset is then called E2. From E2 Alpha will be on similar track as the modern development of metropolis. Without catastrophic events like natural apocalypse or war city will be on a steady grow for both economy and scale, which means population flow will remain positive for a foreseeable future, and this can expressed as:

$$\varphi'(MED_{it})_T > 0, \frac{\partial\phi(N_{it},MED_{it})}{\partial MED_{it}} > 0, \frac{\partial\phi(N_{it},MED_{it})}{\partial N_{it}} > 0 \quad (27)$$

However, with the development of economy, separate local communities of migrant forms, and social effect emerges to become one of the dominating factors for migrant decision. Complication of social structure, together with hiking working stress will in the long run have negative effect on the mental health of migrants. As discussed before, the education structure of migrants will make things worse as migrants with low SES (which is the result of low education status) are more likely to catch psychological diseases. This can be explained by the wage gap of different social stratum, and those wearing high hats on average receive more years of education than those of lower caste. In addition, with mounting consumption price and real estate price, life becomes harder for those of lower wage in larger cities. Living conditions, together with mental health conditions will together have adverse effect on personal fitness hence further increase living expenses. The cycle forms again, and the fact that most migrants (about 90% in China) have not received high education will lead to further stratification of this society. Thus under mounting pressure no wonder why so many people end up on their way back home, and this is what we called urban retreat. As subject social mobility, which is one's expectation to move upwards in social hierarchies, has positive effect on one's sense of social well-being, pressure and stress will reverse this effect by dampening one's ambitions [18]. So after a certain time spot E3 the population net inflow which has remain steady will start to be offset by the outflow of urban retreat. But as capacity of the city has been formed, it is not likely that through this urban retreat process massive outflow will be observed without the force of policymaking. So net population flow will go down, even become negative, but job vacancy and opportunity left by those gone will attract new migrants hence population flow here is also a cycle, oscillating between high and low left by population inflow and urban retreat. The partial effects regarding stage E2 to E3 can be expressed as follows:

$$\varphi'(MED_{it})_T > 0, \quad \frac{\partial \phi(N_{it}, MED_{it})}{\partial MED_{it}} < 0, \quad \left| \frac{\partial \phi(N_{it}, MED_{it})}{\partial MED_{it}} \right| > \varphi'(MED_{it})_T \quad (28)$$

The above discussion, in macro scale, can be summarized as follows:

For E1,

$$P_0(1+i)^{t_1} + \int_0^{t_1} F(E_{it}, S_{it}) dt = C_{it_1} = G(E_1, S_{it_1}) \quad (29)$$

And,

$$\frac{\partial I_{it_1}(E_1, S_{it_1})}{\partial E_1} > \frac{\partial G(E_1, S_{it_1})}{\partial E_1} \quad (30)$$

In this case P0 denote the base period population at the start of our observation.

Utility function U represents the demanding side of people, where u denotes the maximum utility temporal social economic conditions can provide.

$$U(E_1, S_{it_1}) = u(E_1, S_{it_1}) \quad (31)$$

$$\frac{\partial U(E_1, S_{it_1})}{\partial E_1} > \frac{\partial u(E_1, S_{it_1})}{\partial E_1} > 0 \quad (32)$$

For E2 we have:



$$P_0(1+i)^{t_2} + \int_0^{t_2} F(E_{it}, S_{it}) dt = C_{it_2} = G(E_2, S_{it_2}) \quad (33)$$

$$\frac{\partial I_{it_2}(E_2, S_{it_2})}{\partial E_2} < \frac{\partial G(E_2, S_{it_2})}{\partial E_2} \quad (34)$$

With utility as follows:

$$U(E_2, S_{it_2}) = u(E_2, S_{it_2}) \quad (35)$$

$$0 < \frac{\partial U(E_2, S_{it_2})}{\partial E_2} < \frac{\partial u(E_2, S_{it_2})}{\partial E_2} \quad (36)$$

For E3 the average effect of oscillation can be taken as follows, given that maximum level of population capacity has been reached.

$$\frac{\partial I_{it_3}(E_3, S_{it_3})}{\partial E_3} = \frac{\partial I_{it_i}(E_i, S_{it_i})}{\partial E_i} = 0 \quad (i > 3) \quad (37)$$

Using the simplified form of health care, the above time spots can be written as:

$$\Theta_1 \left( \varphi(\text{MED}_{it_1})_T, N_{it_1} \right) > \Gamma_1 \left( \varphi(\text{MED}_{it_1})_T, N_{it_1} \right) \quad (38)$$

$$\Theta_1 \left( \varphi(\text{MED}_{it_2})_T, N_{it_2} \right) < \Gamma_1 \left( \varphi(\text{MED}_{it_2})_T, N_{it_2} \right) \quad (39)$$

$$\Theta_1 \left( \varphi(\text{MED}_{it_3})_T, N_{it_3} \right) = \Theta_1 \left( \varphi(\text{MED}_{it_i})_T, N_{it_i} \right) = 0 \quad (i > 3) \quad (40)$$

The graph of  $\Theta$  function, if drawn, will feature a reverse U over the x axis and a U beneath x axis, with tails repeating this figure. Of course, the head of this function will be heavier than the other parts given it takes long to develop into metropolis

### 3. Empirical Evidence

#### 3.1 Data Summary

The data for this research is gathered from statistic yearbooks of 19 Chinese cities from 2005 to 2017. The cities are selected from certain areas of China mainland, and five of these cities are selected as the economic centres for each metropolitan area, namely Shanghai, Guangzhou, Chengdu, Zhengzhou and Beijing. For some cities in Sichuan Province and Henan Province, detailed records on high school enrolment are not available, but since the number of missing records is small compared to the sample size and the inclusion of more representative variable high school student/teacher ratio, the omission of missing high school data will not affect the robustness of our model vide infra.

The cities are selected according to geographical locations and are categorised into nearby metropolitan areas. In each metropolitan area, one centre city and three other less developed cities in its vicinity are selected to create a gravity chain of population, whereby the peripheral cities are taken as supplier of migrants and the centre cities are taken as the gravity centre absorbing those population inflows. One exception in this assumption goes to the “Yangtze River Delta” region containing three provinces and 15 cities, in which there are three centre cities—Nanjing, Shanghai and Hangzhou. All three vertex cities are selected into the model, and two of the supplier cities inside the triangle are selected based on the assumption that massive population outflow can occur in all three directions.

The summary statistics below includes mean, median and standard error of each of gathered data. Note that there are two GDP variables, the first one is just GDP gathered right from the workbook, and GDP2 is GDP1 divided by accumulated CPI (taking 2005 as 1) and the tuned GDP factor will take into account of the change in price level since in the social effect model the rise in the price level will form part of the social effect as a sort of feedback given that fast growth of consumption price will dissatisfy the migrants and so their mental health and feedback on the living conditions in city will likely to dent. Increase in wage levels at the same rate of CPI increase will not have as much effect as price level on the happiness and welfare of migrants, since on average at least 90% of migrants have not taken higher education and then most of their wages will cover living expenses like food and rent and some of the wages will be saved in bank accounts or sent as remittance to family back home. This assumption means that the disposable income for migrants on average will not be high enough to create financial need like investment for them, and hence what matters most for most of the migrants will be the living cost level which will be reflected in CPI change.

Other than the GDP variables, there are two Delta variables too in our data set. The Delta variable is simply the difference of resident population between two continuous years (like Delta for 2017 is RP2017 minus RP2016), so we expect to see some negative Deltas in our data panel given that in less developed cities people tend to leave for larger cities in search of better opportunities and better wages so we expect net population outflow in those economic backwaters. Delta2 measures the difference between resident population and the registered population according to the census compilation. We use this variable as a measure of the scale of external population. Direct difference between the two numbers may not pinpoint the number of external populations directly, but it can act as a qualitative measure of the size of migrants. For those with low or negative Delta2 this may be a sign of serious resident outflow. In addition, a portion variable is calculated through dividing Delta2 by resident population. This is to measure the effect of local migrant community which can be seen from two mirrored perspectives. First one is the aggregation effect or the network effect which will have positive effect on the new migrants given the existing size of migrant societies. This effect will be channelled through communication tunnels like social media and the use of network and mobile phone. But large crowds will also have negative impact on crowd behaviour. In this case we call it “elbow effect”, or exclusion effect. This is because the selfishness deep down in our heart will take the helm when one is in his full swing or gaining advantage from a certain condition, and it is just human nature to exclude any outsider or ones without intimacy from the benefits and boon.

The high school portion here is calculated through dividing the population with high education by overall resident population as a measure of labour education level for the city. As I mentioned before, most of the migrants do not have high education so those with lower education, when entering a city with relative high education level, will likely to rank low amongst the city social economic status levels. Low social economic status will in the long run affect both one’s living conditions and one’s mental health status because of the accumulated distress and torment from the “underdog” lifestyle in a large city. In addition, talking of one’s motives of migrating to larger cities, the sole purpose for migrating is just to get better welfare, which is reflected in wage level, education for younger generation and better health care. The pursuit for better welfare will in some way aggravate one’s psychological conditions because of the traumatizing conflicts between better live and dire living conditions. Hence social economic status will affect migrants in the long run.

Table 1: Summary Statistics.

	Mean	Median	Standard Error
Resident population	852.8	749.4	535.515
GDP (per capita)	55329	47582	34367.370
Delta	14.14	3.75	34.043
College (student/teacher ratio)	17.45	17.49	3.758
Highschool (student/teacher ratio)	14.26	13.61	3.523
Doctor (per 10,000 population)	22.04	21.55	7.886
Bed (per 10,000 population)	44.46	45.39	14.926
Registered population (per census)	724.4	683.5	314.329
Delta 2	128.4	24.2	272.529
Portion	0.05325	0.03428	0.209
Highschool portion	0.12973	0.08947	0.0919
GDP2 (CPI adjusted)	45143	40710	26003.45

### 3.2 Estimation Method

Fixed effects model and Pooled OLS model are used to estimate the partial effect of variables on population net flow. Since there are no time-constant variables, both models can be expressed in the same form, the main difference in the two regression models is that fixed effect model use the form of within estimator, which means that the variables in regression are demeaned to eliminate the time-constant effects, in this case the use of different estimation methods may produce different results for the two models.

Without taking into consideration the within estimator, the model can be expressed as follows:

$$y_{ig} = x'_{ig}\beta + c_i + \varepsilon_{ig}, i = 1 \dots N, g = 1 \dots G \quad (41)$$

In this case we use the change in resident population Delta as dependent variable  $y$ , indexed with city number  $I$  and time period  $g$ . The variables are selected and include in vector  $x$ , indexed in the same way as the dependent variable.  $C$  indicates unobserved time effects, also called individual effect, unique for each of the cities observed which represents the average effect in the time unit for the panel data, in this case one year. For each city the unobserved effect is assumed to be uncorrelated with  $x$  variables across time for pooled OLS model and arbitrary correlation is allowed for fixed effect analysis. There are series correlations in some of the independent variables, as mentioned in the above theorems. For instance, series correlation of 0.7315 has been found between CPI adjusted per capita GDP (GDP2) and the number of doctors per ten thousand people (DOC), for GDP per capita without divided by CPI, this correlation is even stronger, reaching 0.7377. Similar results have been found between GDP2 and the number of high school student per teacher (-0.7009231), between GDP2 and the portion of high education amongst resident population (0.7896) and between the number of beds per ten thousand people and the number of licenced doctors per ten thousand people (0.7738). However, due to some reason (which will be explained in section 4.2 together with analysis of model coefficients) in the regression models we will still include seriously correlated variables, this can be explained by the nature of panel data we gathered, see section 4.2 for more discussion.

### 3.3 Fixed Results

Table 2: Summary Statistics.

	Pooled OLS	Fixed Effect
Delta2	0.07395***	0.06019
GDP2	-0.00010472	-0.00024511
Doctor	0.13118	-0.18559
Bed	1.4739***	2.7524***
Portion	-39.584	-7.625
High School Portion	66.834	130.997
Doctor*Bed	-0.019041	-0.031478
R squared (Adj R-sq)	0.37964	0.20763

\*\*\*P <0.01

Pooled OLS and fixed effect model are used for the panel data model, and the results are shown in the above table. The variables we use in the two models are the same, because comparison is needed to figure out the serial correlation across time. Note that we did not include GDP and high school student per teacher in the two models. The reason for the exclusion of GDP per capita is mentioned above in section 4.1, as we include in the model the CPI adjusted variable GDP2 which takes account of the rising price level effect on the living conditions of migrants. The exclusion of high school student per teacher is partly due to lack of information from remote areas in Sichuan province, but more importantly, high education will not be a good proxy to represent education development in less-developed areas due to lack of strong economy and resources to support the higher education facilities like R&D. So we resorted to the high school student per teacher as a proxy variable for education against the backdrop of Chinese National College Entrance Exam system which provides students from across the country with a chance to get high education through fair examination. Exam results for each city will form part of the government evaluation indicator and reported in the annual report compiled by the local authorities. So student per teacher for high school may be a decent indicator for the measurement of education development as those with less investment tend to have larger classes than those with well-developed education system. Section 4.1 mentioned strongly correlated variables Doctor and bed, and both doctor and bed features in the two models together with interaction factor Doctor\*Bed. The reason why these are included goes beyond the model to the panel data itself. The main argument here is that in some less developed cities like Meishan, Zigong and San Menxia, the number of beds per ten thousand people is on the same level with or higher than some well-developed cities like Nanjing and Hangzhou. This anomaly is probably due to the chain reaction of less-developed economy. As mentioned in part 2, low economic development means that local government will not have enough funds to upgrade education and health care facilities, and this means that more educated locals will migrate to larger cities, left at home those with higher age and less educated. Hence the remaining population are more prone to catch diseases due to lack of knowledge in health care and relative hypoimmunity due to age and exposure to hygienic conditions that do not meet the sanitary standards to prevent contagion. Thus the expectation of catching diseases for susceptible population in the economic backwaters may be higher than those in well-developed areas with higher education level, and this explains the anomaly of higher bed numbers in remote

areas of less-developed provinces like Sichuan. The two proxy beds and doctors on their own may not decently represent the health care development in certain areas, but the combination of these two, together with the interaction term are proven to be cogent indicators based on the regression results shown above, which we will discuss later.

The above model summary shows great different in the significance of pooled OLS model and FE model, as indicated by large standard errors and adjusted R-squared for FE model. The difference between the significance levels for two models may due to the lack of correlation between independent variables and the unobserved individual effects.

Since economic, educational, medical, social issues are all covered in our model, invoking the proposed theory in section 2, the individual effect in this case may be attributed to the geographical areas in which cities are based, rituals for the local societies and policy factors which are hard to measure without real life experience and measurement benchmarks, all of which are quite unique for all the cities with such cultural diversity in China. These unique estimators are of course not correlated with existing variables which mainly describe the environmental issues and base stats, so no wonder why the fixed effect model shows low adjusted R-squared and model significance.

## 4. Discussion and Conclusion

### 4.1 Discussion

The following analysis will be mainly based on Pooled OLS model, given that the significance level for this model is significantly higher than the FE model. FE model, when needed, will be mentioned as reference to certain effects.

What is worth noting in this model is that the positive single effect of Bed and Doctor, if we ignore the huge insignificance of Doctor, produces negative interaction estimator, and the significance level for the interaction term is not unacceptable at around 0.2. In section 4.1 we explained the reasons for inclusion of two strongly correlated variables Doctor and Bed, and their interaction. Here, based on the regression results and our data panel, there may be a cogent explanation for it. As Bed on its own cannot represent the health care development of one particular city, we found that Doctor, when combined with Bed, produces interaction term which shows completely new results. The partial effects can be explained by taking differentiated form. Note that in our data set doctor and bed on its own cannot explain the whole effect due to the anomalies in remote areas of China mainland, we now analyse the effect of Doctor and Bed with both terms. The strongly significant positive effect of Bed may due to the increase in hospital capacity in taking patients with hospitalization needs, but increase of beds without taking into account of existing number of registered doctors will absolutely have negative effect on resident welfare as the number of patient one doctor may be responsible to will increase due to the rising number of beds which means the average expected number of hospitalized patients will increase. The increase in expectation of hospitalized patients is due to the palliative effect brought about by the increase in bed number, and also the same goes to the number of doctor per ten thousand people. This will lower the expected mortality rate when catching serious disease for local residents given that from their perspective what matters is just sheer number of health care resource increases. On aggregate this will lower the survival rate since infection risk increases, and this problem will be worsened if only one of the two indicators increases given that doctors and beds are complements for each other, if a city cannot strike a balance between the number of doctors and the number of beds then one of the health care resources will be over slanted hence evaluation of health care system as a whole will be lowered as mortality and infection rate increases. Here the interaction term can be explained as an offset balance term for single effect of doctors and beds as

these two when singled out will have skewed representation of health care, and the complementary effect plays a important role in the correlation between Bed and Doctor. So inclusion of two highly correlated variables and their interaction term seems reasonable in this case as we don't often see perfect complement included in one single regression model.

Variables Delta2, Portion and High school portion are supposed to reflect the social effect large crowd has on new-comers. As we have seen in part 3 data description, the Delta2 variable is derived from the difference of resident population and registered population. Although this may not precisely be direct measurement of migrants, the sheer size of value calculated will be a decent indicator of the scale of external population. Portion is calculated to represent the scale of local community, or more precisely, the migrant societies formed through time and the high school portion variable is calculated to represent the education level which according to section 4.1 will have a huge effect on the social economic status for those new-comers, further consideration on mental health and psychological status is not taken into the model as we need tracking information to compile a persuading dataset. As seen from the Pooled OLS model, Delta2 and high school portion have positive partial effect on population net flow (which in this case means net inflow) and Portion has negative partial effect on population net flow (which in this case means increasing the migrants population will result in population outflow in a city of selected size). Explanation for these three partial effects is quite straight forward compared with the partial effect on Bed and Doctor. To explain the significantly positive effect of Delta2, we will invoke the network effect which is often quoted in sociology or telecommunication studies. Consider network system like social media, five, fifty or five hundred users will not constitute an online community, and hence the network itself will not be as enticing and appealing as those social media with thousands of users. For social media the accumulation of user base takes time, and the utility one may get on entering a social network remains low until user base reaches a certain threshold. For migrant population this rule applies too. As we gather data from Chinese cities on county level with at least two million residents, the networks are assumed to be well-built so the "entree problem" described above as a starting conundrum will not occur in this case. So it is not hard to conclude that the existing migrant scale will have positive network effect on potential migrants as the message in larger cities can be channelled through social media like WeChat and Tik Tok. But the negative effect brought by portion variable which stands for community scale sheds a different light on this issue. From a statistic point of view, the significance level of about 0.2 is not unacceptable so we can assume that the negative effect itself is of statistical significance. Then, from sociological perspective, the negative effect of crowd scale indicates there is crowding-out effect in it, that is, out of the human nature of selfishness, the early entrants will try to keep their pole position by excluding the others from sharing similar benefits, in this case the big cake one may share on migrating into a larger city, and this can explain the negative effect of portion on population flow. That is to say, the existing community reaches a maximum level of members to share the "cake", as new members coming in, each person's share will dwindle, though not uniformly, due to the nepotistic nature of Chinese society. Then the stimulus to keep one's social position will form a formidable force excluding the new entrants from the well-built local society, so the partial effect of local community scale on population flow, under the aforementioned assumption, is negative in our model.

As we mentioned before high education proportion in the long term will have negative effect on migrants' health status both physically and mentally due to the channelling mechanism through social economic status. However, in our model here the effect of high school portion is positive with significant level of about 0.2, from which we can ascertain that the overall effect is not negative. This deviation can be explained from the perspectives of policy making. From 2008 Chinese government has issued a series of high level talented recruitment plans to support local science development.

Since most of the universities and research institutes (other than those for military purpose) are based in larger cities like province capitals because of the funds and policy advantage institutes can get from local governments. Upon graduation, students are encouraged to seek out working or research opportunities in larger cities given the wage level and opportunity abundance in these well-developed areas. Hence the larger cities in China tends to hold larger share of highly educated population, and this share will grow even larger as Chinese universities increase enrolment size according to central policymaker. Moreover, on graduation, students with higher education will seek out working opportunities in larger cities because of wage level and industrial diversification, thus aggregation is likely to occur in larger cities in China, and in short term this can explain the positive partial effect of higher education proportion variable.

## 4.2 Conclusions

Empirical regression results seem to agree with our theoretical model concerning health care's effect on population net flow and the role sociological factors plays in it. For cities with developed health care conditions, the effect of changes in beds and doctors per ten thousand people, on calculation from the first derivative, is negative on net population flow which in this case means population out flow, this is in concert with our three-point theory in part two which mentioned that metropolis are likely to enter E3 stage with population out flow caused by the social effect like pressure brought over by the economic development. The positive early stage effect of health care is proven from the same perspective. Talking about the social effect on migration decisions, from state-gathered we cannot track individual effect hence exact impact of social media and instant communication on potential and incumbent migrant cannot be measured without individual tracking studies. From the state panel data and regression results we can verify the positive network effect existing migrant communities have on potential migrants and the crowding-out exclusion effect which has negative effect on migration decisions.

We conclude here with external validity question on our research, whether or not our research on Chinese cities can be applied to other countries as not many of them shares similar population densities. The significant effect of social factors here in our model may attribute to widely used Chinese social media like WeChat with instant communication and social network system combined in one single app [19]. In addition, fast development and popularization of short video apps like Tik Tok increases the screen time of Chinese population even further. With several tweaks of our theory I believe our model can be applied to other parts of the world. Study by Mingli Liu et al. has proven that adolescent psychiatric symptoms and self-harm behaviour are strongly correlate with the time exposure to television or recreational computer use like video games, hence increased time spent on mobile devices will likely to increase the risk to psychiatric problems hence this partly explains the need to include social media feedback in our model [20]. While not all countries has the same population density as China, we may need to add geographical location indicator and population density in order to take into account of transport and radiation of city's impact, as in developed countries like USA population takes a rather scattered distribution. Also, the effect social economic status on migrants will have to be tuned since labour in different countries has different educational backgrounds. The effect of social media for other countries may not be as much as in China as other countries do not have similar population structure in China as rural population amounts to about 900 million in China, which makes the use of social media a stronger determinant in migration choice.

Due to time and funding constraints, tracking study has not been done for this research, but tracking is necessary if researchers want to investigate the effect of social and health care on individual migration decisions in order to fully understand crowd behaviour under migration



situations.

## References

- [1] Bingwen Z. 30 Years of Reform and Openness in China: Development and Challenges of Social Security System for Floating Population[J]. *Chinese Journal of Population Science*, 2008, 000(005):2-17.
- [2] Brueckner J K , Smirnov O . *Social Networks and the Convergence of Population Attributes: A Generalization*[J]. *Working Papers*, 2006, 48(2):359-365.
- [3] Primack B A , Shensa A , Escobar-Viera, César G, et al. Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults[J]. *Computers in Human Behavior*, 2017, 69:1-9.
- [4] Ponds R , Marlet G , Van Woerkens C , et al. Taxi drivers with a PhD: trickle down or crowding-out for lower educated workers in Dutch cities?[J]. *Cambridge journal of regions, economy and society*, 2016, 9(2):405-422.
- [5] Song Y , Sun W . *Health Consequences of Rural-to-Urban Migration: Evidence from Panel Data in China*[J]. *Health Economics*, 2015, 25(10):1252-1267.
- [6] Lu, Yao. *Mental health and risk behaviours of rural–urban migrants: Longitudinal evidence from Indonesia*[J]. *Population Studies*, 2010, 64(2):147-163.
- [7] Firdaus G . *Mental well-being of migrants in urban center of India: Analyzing the role of social environment*[J]. *Indian Journal of Psychiatry*, 2017, 59(2):164.
- [8] Reitmanova S , Gustafson D L . *Mental Health Needs of Visible Minority Immigrants in a Small Urban Center: Recommendations for Policy Makers and Service Providers*[J]. *Journal of Immigrant and Minority Health*, 2009, 11(1):46-56.
- [9] Wong P H , Li T W , Poon K L , et al. *An exploratory study on the mental health of immigrants, refugees and non-status people living with HIV in Toronto*[J]. *International Journal of Migration, Health and Social Care*, 2013, 9(3):122-134.
- [10] Ao X , Jiang D , Zhao Z . *The impact of rural–urban migration on the health of the left-behind parents*[J]. *China Economic Review*, 2016, 37.
- [11] Chenyue Z , Feng W , Xudong Z , et al. *Impact of parental migration on psychosocial well-being of children left behind: a qualitative study in rural China*[J]. *International Journal for Equity in Health*, 2018, 17(1):80-.
- [12] Cheung, Nicole W T . *Rural-to-urban migrant adolescents in Guangzhou, China: Psychological health, victimization, and local and trans-local ties*[J]. *Social Science & Medicine*, 2013, 93:121-129.
- [13] Ha W , Yi J , Yuan Y , et al. *The dynamic effect of rural-to-urban migration on inequality in source villages: System GMM estimates from rural China*[J]. *China Economic Review*, 2016, 37:27-39.
- [14] Fang Z , Sakellariou C . *LIVING STANDARDS INEQUALITY BETWEEN MIGRANTS AND LOCAL RESIDENTS IN URBAN CHINA-A QUANTILE DECOMPOSITION*[J]. *Contemporary Economic Policy*, 2016, 34(2):369-386.
- [15] Fulvio R , Carlotta S , Teresa G M , et al. *The Association between Educational Level and Cardiovascular and Cerebrovascular Diseases within the EPICOR Study: New Evidence for an Old Inequality Problem*[J]. *PLOS ONE*, 2016, 11(10):e0164130-.
- [16] Cardano M , Scarinzi C , Costa G , et al. *Internal migration and mental health of the second generation. The case of Turin in the age of the Italian economic miracle*[J]. *Social Science & Medicine*, 2018:S0277953618302284.
- [17] Chen F , Liu H , Vikram K , et al. *For Better or Worse: The Health Implications of Marriage Separation Due to Migration in Rural China*[J]. *Demography*, 2015, 52(4):1-23.
- [18] Silin H , Jiawei H , Ling S , et al. *The Effects of Objective and Subjective Socioeconomic Status on Subjective Well-Being among Rural-to-Urban Migrants in China: The Moderating Role of Subjective Social Mobility*[J]. *Frontiers in Psychology*, 2017, 8:819-.
- [19] Qiu J , Li Y , Tang J , et al. *The Lifecycle and Cascade of WeChat Social Messaging Groups*[J]. 2015.
- [20] Mingli L , Qingsen M , Jinyao Y , et al. *Screen Time on School Days and Risks for Psychiatric Symptoms and Self-Harm in Mainland Chinese Adolescents*[J]. *Frontiers in Psychology*, 2016, 7.