Regional Migration in China: A Machine Learning Approach to the Hukou System

Marta Bengoa Calvo, Colin Powell School. City University of New York. University of Johannesburg and CIRANO
Thierry Warin, HEC Montréal and CIRANO

ASSA 2021

Introduction and Literature Review
oooooooo

Research Question and Motivation
oooooo

Model 1
oooo

Model 2
oooooooo

Conclusion
ooo

Section 1

Introduction and Literature Review

Introduction

- China's rapid development have spurred large migration from rural areas to urban areas

- Between 1990 and the end of 2015 the proportion of China's population living in urban areas jumped from 26% to 56%

- Currently estimated by census, there are more than 240 million rural-to-urban migrants and more than 160 million working in cities outside of their hukou. That accounts for approx. 30% of total rural labor force (China National Bureau of Statistics).

- The Hukou household registration system imposes restrictions and limits to where to live, which is determined mainly by birth.

- Hukou card is an internal passport that sets access to education and health services. It started in 1956-58, relaxed during the 60s and enforced again since 1978.

Introduction
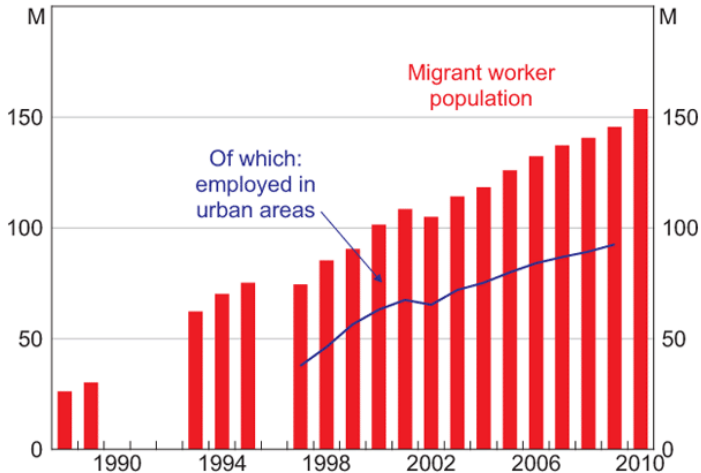
*   The red colums refer to all rural migrants without local 'hukou' where they are living; the blue line by Herd, Koen and Reuterswald (2010) calculates the level of rural–urban migrant employment, RBA estimate for 2009 number following their methodology

Sources: Chan (forthcoming) for 1988, 1989, 1993–1995 data; MOLSS
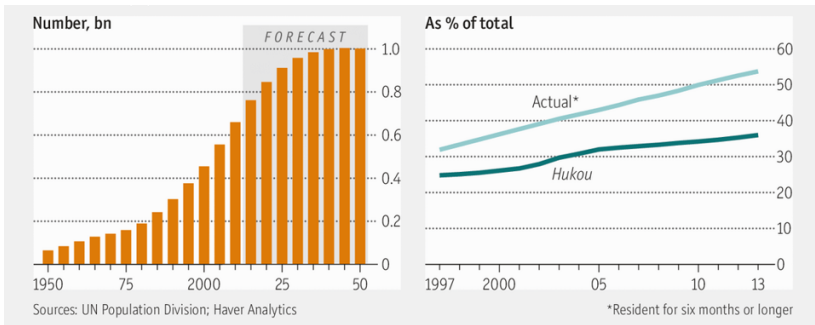
Introduction

Figure 2: China's population

Introduction

- In 2009, China designed a healthcare reform that was intended to address inequalities in health access. Specific goals include:

  1. Provide wide basic health coverage for legal urban workers (mandatory) and to allow voluntary enrolment for urban residents without jobs (including students, elderly or disabled);
  2. Voluntary cooperative medical system for rural citizens; and
  3. Provide wide healthcare coverage to vulnerable and low-income population groups.

- The reform left the rural-to-urban migrant population outside of the provision of public healthcare.

Introduction

- Several major cities have tried to introduce policy reforms to address the health challenges confronted by this phenomenon.

- Since 2006, Beijing is offering primary care health services and free healthcare to the children of migrant workers in community centres.

Introduction

- There are not many studies that have addressed this link between migration with limited access to healthcare and health outcomes in developing economies. The most recent for China published by Sun (2015), uses self-reported outcomes (do I feel well or not, have I been sick..) which suffers from measurement errors.

- Other studies suggest that migrants are reasonably healthy at the point of migration but more likely to experience adverse effects than non-migrants. As they get injured and can't have access to health some return home while others remain in urban areas.

- Therefore, increases risk of workplace accidents, other contagious diseases (Chen, 2011; Lu and Quin, 2014; Wallace and Kulu, 2014).

Introduction

- Studies mainly focus on the US, on the relationship of health of Mexican immigrants and lack of insurance. Goldman et al. (2014, Demography) find significant and strong correlation between non having insurance and decline in health of Mexican immigrants (compared to never migrants), specially immediately after migration.

- Wassink (2008, Demographic Research) analyses health insurance coverage for returned Mexican migrants to determine that lack of access is mainly driven by unemployment, revealing a negative association between lack of insurance and health

Section 2

Research Question and Motivation

Research Question and Motivation

- Motivation: Using Machine Learning techniques (+ ensemble methods) to study a very complex relationship in Social Sciences: health and migration

- Steps:
  - First, we use traditional econometric techniques for such a topic
  - Second, we use ML techniques to move from **correlations** (econometrics) to **predictive modelling** (ML)

Research Question and Motivation

> *RQ: What we really want to do here is use Machine Learning techniques to question whether hukou status plays a role in the health outcome of migrant workers.*

- We use survey data reported in the Longitudinal Survey on Rural Urban Migration in China from the Institute for the Study of Labor (IZA). The survey collects data for 71,074 individuals (29,556 urban persons; 32,171 rural persons; and 9,347 migrants. Aprox 29% of rural persons) in two waves for the years 2008 and 2009.

- The survey contains data on socioeconomic indicators, such as education, income, ethnicity, and hukou registration, as well as data on many health indicators and outcomes. These include weight (kilograms), height (centimeters), dominant handedness, blood pressure, and grip strength.

Figure 1. Years Since Migration for Migrants

Figure 2. BMI by Residence Status
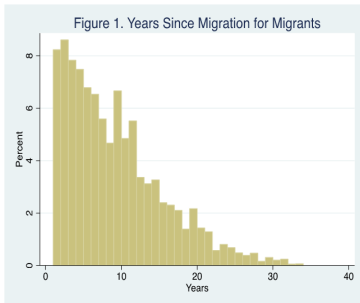
Migrants BMI distribution is skewed to the left, 31% of migrants have a BMI below 18.0, which categorizes them as being underweight and subject to higher health risks
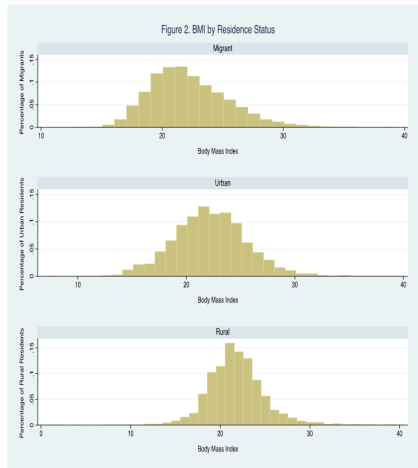
Data

- 816 variables

- 9,475 observations

Data

	Rural Hukou	Urban Hukou
Age (mean)	32.1	39.2
Male %	56.6	48.8
Married %	73.37	80.3
Years of Education (mean)	9.2	10.8
Monthly Income (mean in RMB)	1677.6	1823.7
Wealth (mean in RMB)	4458.1	4667.9
Years Since Migration	8.5	9.8
Excellent or good health rating %	81.1	72.4
Average health rating %	16.8	27.3
Poor or very poor health rating %	2.1	3.1
Self-reported Health Rating (mean from 1-5 scale, 1 best)	2.0	2.1
Height (cm) (mean)	165.5	165.0
Weight (kg) (mean)	61.2	61.6
Smoker %	27.6	20.3
Cigarettes per day (mean)	3.7	2.7
Systolic Blood Pressure (mean)*	119.2	117.4
Diastolic Blood Pressure (mean)*	75.8	75.7
BMI< 18.5 (underweight) %*	6.5	4.5
BMI> 25 (overweight) %*	18.9	17.8
Health Insurance *(per centage)*		
NRCMS (Hukou based, targeted for rural pop, tax-funded and premiums)	58.4	10.8
URBMI (Hukou based, urban pop, tax-funded and premiums)	1.8	6.9
UEBMI (targeted to formal employees, contributions, offered by employer)	0.00	54.9
Commercial insurance	3.7	2.9
Other insurance	1.7	2.2
No insurance	30.5	20.2
Number of observations	7,598	2,706
Actions taken when ill or injured *(per centage)*		
Days missed due to illness or injury (mean)	3.7	6.3
Took No Action	20.1	10.6
Rested, but No Medicines	2.6	3.9
Obtained Medicine by themselves	45.8	38.2

Section 3
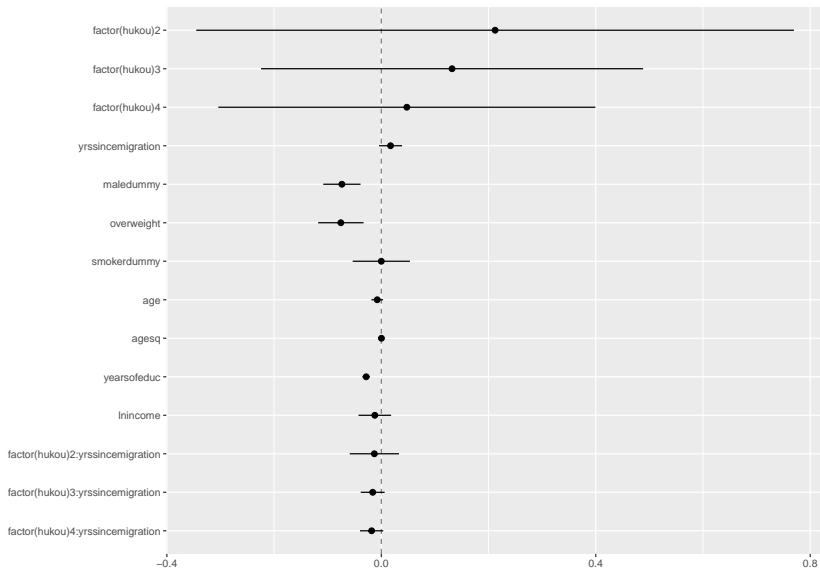
Model 1

Regular regression analysis

- overall sample of 7,917 obs

- no cross-validation

- no out-of-sample predictions

Regular regression analysis

Regression results

Model 1: Regression with no regularization

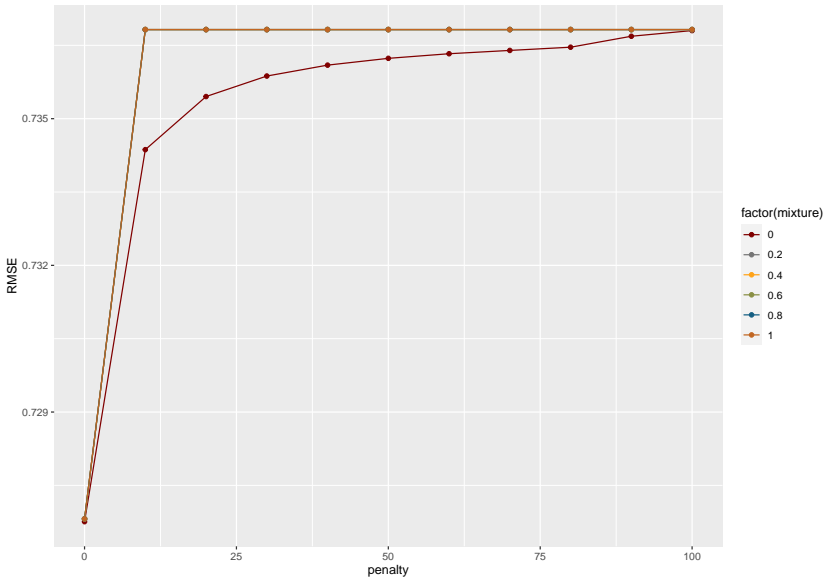| term	estimate	std.error	statistic	p.value
(Intercept)	2.178428e+00	2.263927e-01	9.6223420304	8.471052e-22
factor(hukou)2	2.122157e-01	2.844094e-01	0.7461626237	4.555914e-01
factor(hukou)3	1.319305e-01	1.818107e-01	0.7256474935	4.680764e-01
factor(hukou)4	4.762394e-02	1.794176e-01	0.2654363146	7.906803e-01
yrssincemigration	1.722914e-02	1.097948e-02	1.5692136993	1.166382e-01
maledummy	-7.356271e-02	1.775992e-02	-4.1420623314	3.477819e-05
overweight	-7.550891e-02	2.161436e-02	-3.4934598940	4.794152e-04
smokerdummy	-1.953234e-05	2.723611e-02	-0.0007171486	9.994278e-01
age	-7.617401e-03	5.543715e-03	-1.3740607162	1.694618e-01
agesq	2.199417e-04	7.666419e-05	2.8688967394	4.129980e-03
yearsofeduc	-2.825388e-02	3.556032e-03	-7.9453383037	2.204072e-15
lnincome	-1.214460e-02	1.551170e-02	-0.7829315538	4.336908e-01
factor(hukou)2:yrssincemigration	-1.299365e-02	2.342344e-02	-0.5547284484	5.790961e-01
factor(hukou)3:yrssincemigration	-1.605196e-02	1.138429e-02	-1.4100097241	1.585762e-01
factor(hukou)4:yrssincemigration	-1.811218e-02	1.108752e-02	-1.6335644521	1.023902e-01

Section 4

Model 2
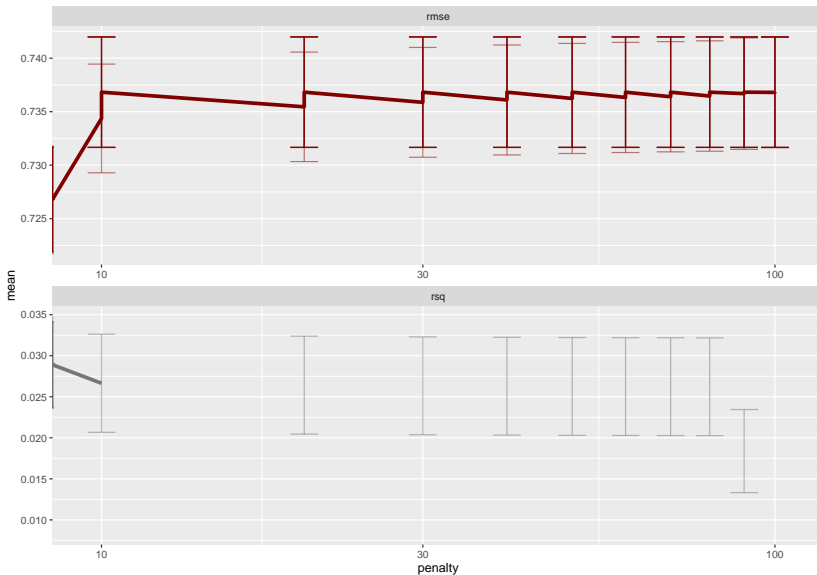
ML model with elastic net

$$RSS + \lambda \left((1 - \alpha) \sum_{i=1}^{p} + \alpha \sum_{i=1}^{p} |\beta_j| \right)$$

- train, test, validate datasets
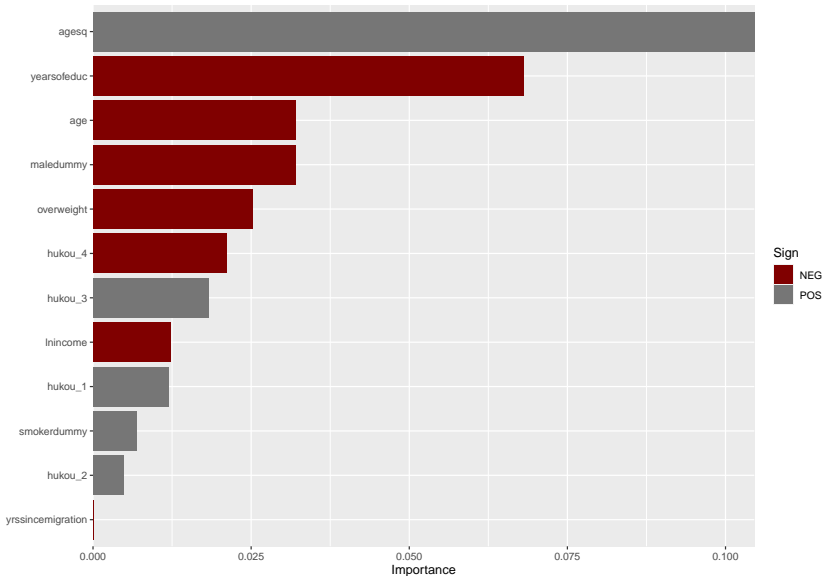- 5 fold cross-validation

ML model with elastic net

ML model with elastic net

ML model with elastic net

Introduction and Literature Review
○○○○○○○○

Research Question and Motivation
○○○○○○

Model 1
○○○○

Model 2
○○○○○○●○○

Conclusion
○○○

ML model with elastic net

Regression results

Model 2: Regression with lambda = 0

term	estimate	std.error	statistic	p.value
(Intercept)	2.178428e+00	2.263927e-01	9.6223420304	8.471052e-22
factor(hukou)2	2.122157e-01	2.844094e-01	0.7461626237	4.555914e-01
factor(hukou)3	1.319305e-01	1.818107e-01	0.7256474935	4.680764e-01
factor(hukou)4	4.762394e-02	1.794176e-01	0.2654363146	7.906803e-01
yrssincemigration	1.722914e-02	1.097948e-02	1.5692136993	1.166382e-01
maledummy	-7.356271e-02	1.775992e-02	-4.1420623314	3.477819e-05
overweight	-7.550891e-02	2.161436e-02	-3.4934598940	4.794152e-04
smokerdummy	-1.953234e-05	2.723611e-02	-0.0007171486	9.994278e-01
age	-7.617401e-03	5.543715e-03	-1.3740607162	1.694618e-01
agesq	2.199417e-04	7.666419e-05	2.8688967394	4.129980e-03
yearsofeduc	-2.825388e-02	3.556032e-03	-7.9453383037	2.204072e-15
lnincome	-1.214460e-02	1.551170e-02	-0.7829315538	4.336908e-01
factor(hukou)2:yrssincemigration	-1.299365e-02	2.342344e-02	-0.5547284484	5.790961e-01
factor(hukou)3:yrssincemigration	-1.605196e-02	1.138429e-02	-1.4100097241	1.585762e-01
factor(hukou)4:yrssincemigration	-1.811218e-02	1.108752e-02	-1.6335644521	1.023902e-01

ML model with elastic net on the test dataset

.metric	.estimator	.estimate	.config
rmse	standard	0.7296538	Preprocessor1_Model1
rsq	standard	0.0381309	Preprocessor1_Model1

Using the results from the regularization assesment to predict

Introduction and Literature Review
00000000

Research Question and Motivation
000000

Model 1
0000

Model 2
00000000
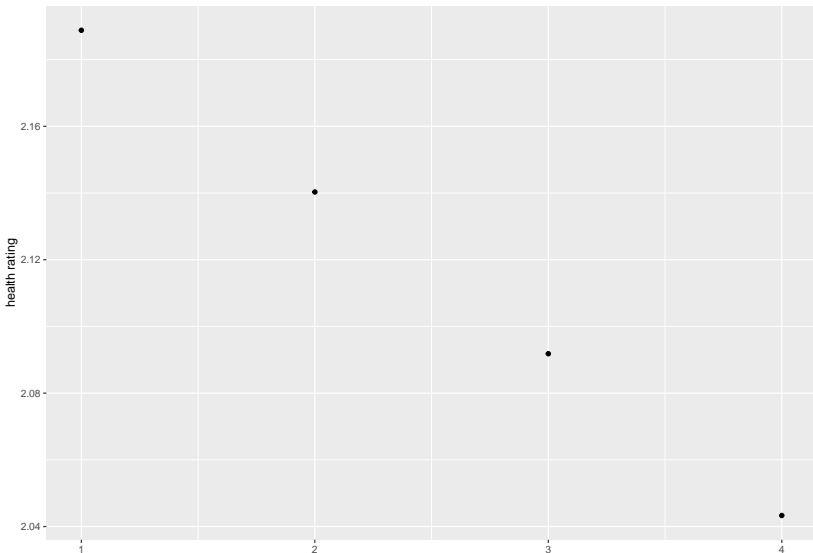
Conclusion
●00

Section 5

Conclusion

Conclusion

- Migration will require adjustments in health provisions to accommodate the changing spatial demographics.

- Restricting migrants access to healthcare will clearly have an effect in the long run, including on migrant's health, productivity, and potential economic growth.

Conclusion

- Further research with classification methods with more features in the ML models
- Public policies design based on these ML estimations