



Analysis of inter-provincial migration and regional income inequality in China between 2000 and 2010 with gravity models

Nina Xingna Zhang
School of Geographical Sciences
University of Bristol

- **Defining internal migration in China**
 - people who currently reside in one place but have their household registration in another county (city or district) and have left their registered residence for more than 6 months.
 - 274 million rural migrant workers in 2014, the majority of them have urban destinations. They help to fuel China's skyrocketing growth in the past 30 years.



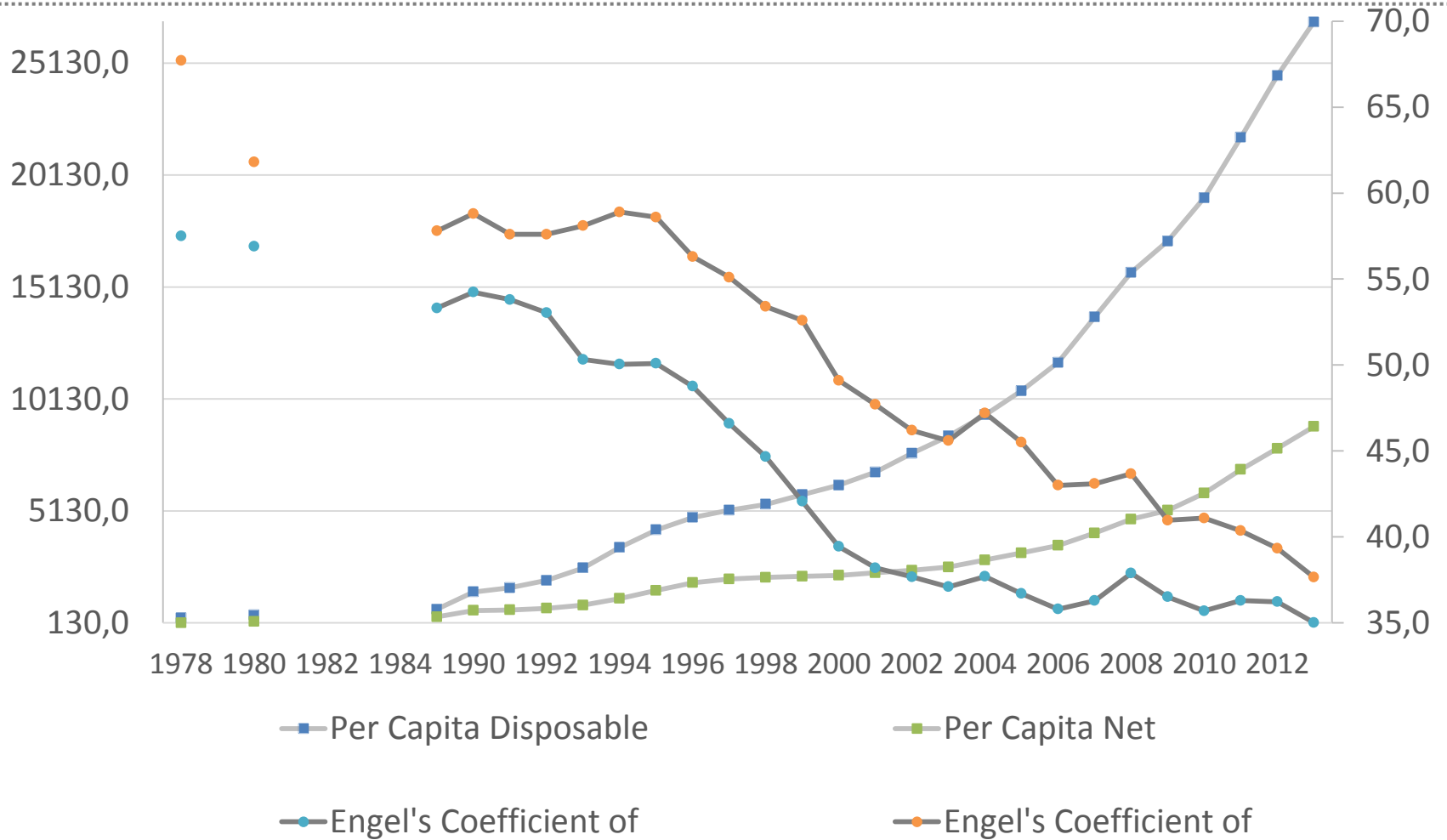
Source: <http://www.globalpost.com/dispatches/globalpost-blogs/rights/china-finds-creative-ways-address-income-inequality>

Four categories of inter-provincial migration:

According to 2010's census data, **261 million total migrants in 2010, and among them 86 million are inter-provincial migrants, with 930 migration streams** (population movement from an origin to a destination) **formed.**

- Rural-urban migration stream (44.24%)
- Rural-rural migration stream (29.97%)
- Urban-rural migration stream (3.42%)
- Urban-urban migration stream (22.38%)

Rural and urban household income disparity:



National rural and urban household income from 1978 to 2013

- Regional income inequality and migration
 - regional income disparity is the main reason for internal migration (Ye, Wang et al., 2013).
 - internal migration plays a significant role in shaping regional inequality (Jenicek, 2010).
 - it has not been done to analyse the relationship between different migration streams simultaneously.

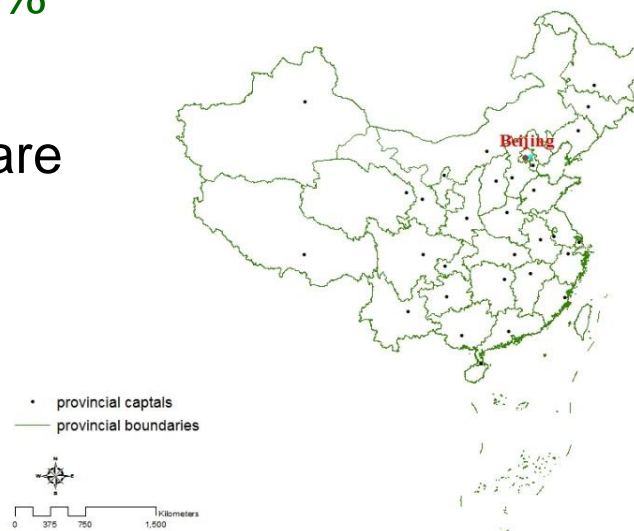
- Gravity model in modelling migration in China
 - gravity models have been the most popular models to predict migration flows (Fan, 2005), and its extended forms have much potential in explaining migration (Christian and Braden, 1966; Claeson, 1969; Johnston, 1970; Ginsberg, 1972).
 - in contrast, studies using it to directly model China's internal migration flows are relatively scarce (Fan, 2005).
 - it has not been done to use segment rather than total populations to analyse specific migration flows.
 - it has not been done to include rural-urban income disparity rather than GDP to analyse internal migration.

Main Research Questions

- How does inter-provincial migration change between 2000 and 2010?
- How origin/destination segment rural and urban populations, affect different inter-provincial migration streams differently?
- How origin/destination rural/urban incomes affect different migration streams differently?

Chinese national censuses in 2000 and 2010

- Short form data and *long form data* (10% sampling of the whole population).
- 930 inter-provincial migration streams are formed in each year's short form data, being used to investigate upon the longitudinal change from 2000 to 2010.
- 897 inter-provincial migration streams, with detailed rural/urban origin and destination information in 2010's long form data, are used to analyze the relationship between different migration streams simultaneously.

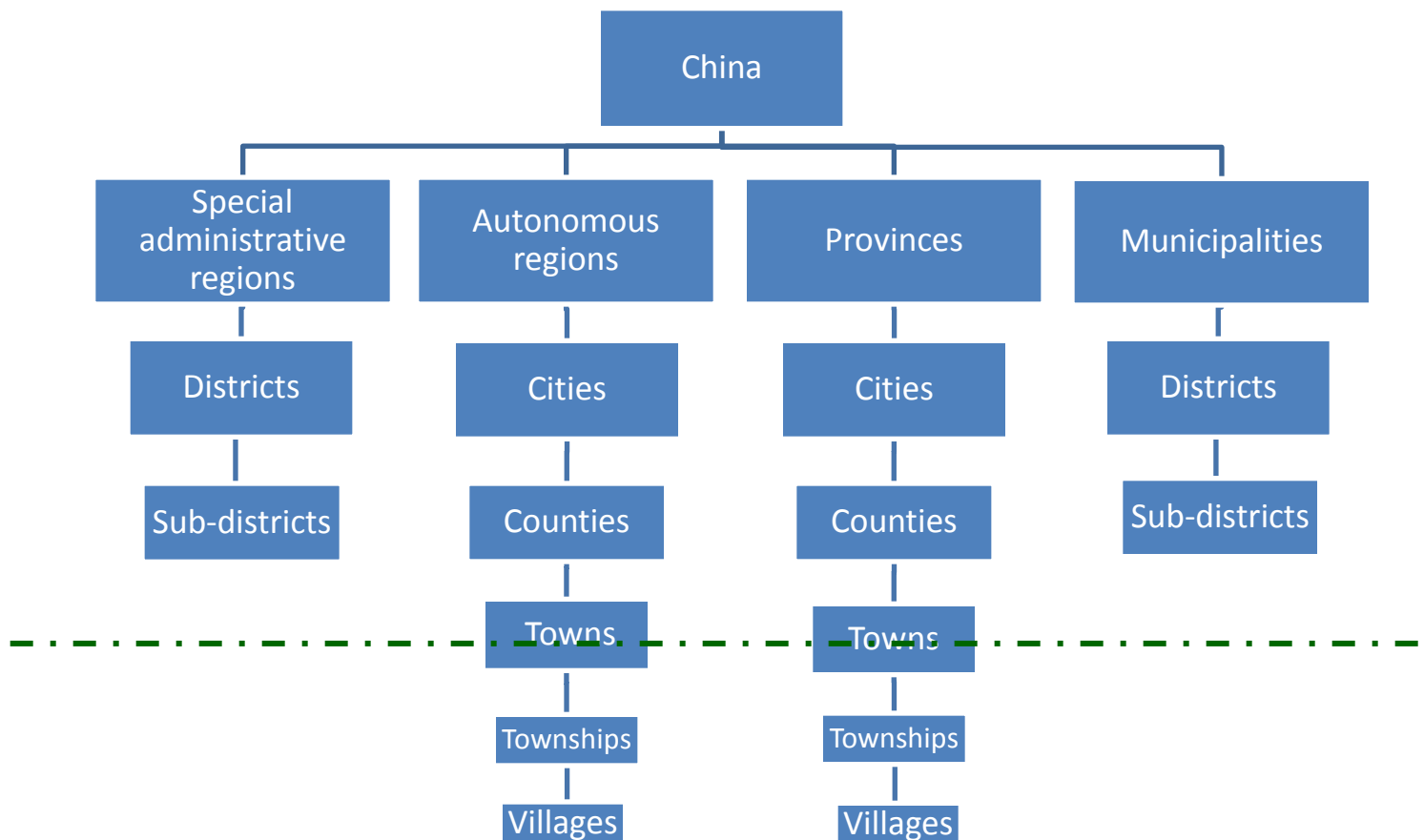


China Statistical Yearbook in 2001 and 2011

- the average per capita income for urban and rural households on the provincial level.
- **regional-specific price deflators** instead of a common deflator for all provinces.

Dependent variables	units	Independent variables	units
interprovincial migration	000s persons	origin-destination distance	000s m
urban-rural migration	00s persons	origin urban population	000,000s persons
rural-urban migration	00s persons	origin rural population	000,000s persons
rural-rural migration	00s persons	origin urban income	000s yuan
urban-urban migration	00s persons	origin rural income	000s yuan
		destination urban population	000,000s persons
		destination rural population	000,000s persons
		destination urban income	000s yuan
		destination rural income	000s yuan

Calculation of Dependent Variables



Hierarchy of China's administrative units

Gravity Model of Migration

- When migration itself is directly modelled, gravity model is quite special as it could incorporate both origin and destination factors into the model.

Linear regression

- It is used to predict the value of one response from a set of predictors, i.e., interprovincial migration in 2000 and 2010.

Multivariate Linear Regression

- Analysis of the relationship between different migration streams simultaneously in 2010.

The Extended and Enhanced Gravity Model of Migration

- The extended and enhanced form of the gravity model,

$$m_{ij} = k \frac{p_{ir}^{a_1} p_{iu}^{a_2} p_{jr}^{b_1} p_{ju}^{b_2}}{d_{ij}^c} I_{ir}^{f_1} I_{iu}^{f_2} I_{jr}^{g_1} I_{ju}^{g_2}$$

Where,

- $k, a_1, a_2, b_1, b_2, c, f_1, f_2, g_1$ and g_2 , are parameters;
- m_{ij} is migration flow between place i and place j ;
- p_{ir} and p_{iu} are the rural and urban population of origin i ;
- p_{jr} and p_{ju} are the rural and urban population of destination j ;
- I_{ir} and I_{iu} are the rural and urban household income per capita of origin i ;
- I_{jr} and I_{ju} are the rural and urban household income per capita of destination j ;
- d_{ij} is distance between place i and place j .

-China's internal migration is highly directional and economically driven (Fan, 2005).

Taking logs on both sides – *model used in this study*

$$\log m_{ij} = \log k + a_1 \log p_{ir} + a_2 \log p_{iu} + b_1 \log p_{jr} + b_2 \log p_{ju} - c \log d_{ij} + f_1 \log I_{ir} + f_2 \log I_{iu} + g_1 \log I_{jr} + g_2 \log I_{ju}$$

-Adding economic inequality in the equation - China's internal migration is largely economic-induced;

-Decomposing total population into segmented urban/rural populations in the model - more relevant in migration studies in developing countries;

-Taking log on both sides - the mathematical assumption that migration flows have a Poisson distribution.

Model comparison – linear regression models without taking logs

Model	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
2000 interprovincial migration	origin urban population	-1153.473	1871.119	0.620	0.538	-4825.629	2518.683
	origin rural population	1620.645	458.608	3.530	<0.001	720.606	2520.683
	destination urban population	8543.815	1871.119	4.570	<0.001	4871.659	12215.970
	destination rural population	-351.475	458.608	0.770	0.444	-1251.513	548.564
	distance	-29.193	8.972	3.250	0.001	-46.802	-11.585
	origin urban income	-7158.314	7806.751	0.920	0.359	-22479.420	8162.794
	origin rural income	5906.455	18647.030	0.320	0.752	-30689.190	42502.100
	destination urban income	51622.010	7806.751	6.610	<0.001	36300.900	66943.120
	destination rural income	-66902.260	18647.030	3.590	<0.001	103497.900	-30306.610
constant	189019.100	45326.220	4.170	<0.001	277973.900	100064.300	
2010 interprovincial migration	origin urban population	-2990.402	2612.489	1.140	0.253	-8117.532	2136.728
	origin rural population	3439.309	841.180	4.090	<0.001	1788.455	5090.163
	destination urban population	17445.650	17895.620	0.970	0.330	-17675.320	52566.620
	destination rural population	857.848	705.603	1.220	0.224	-526.930	2242.626
	distance	-58.578	13.768	4.250	<0.001	-85.598	-31.559
	origin urban income	-24164.920	7371.542	3.280	0.001	-38631.910	-9697.929
	origin rural income	35979.490	14141.540	2.540	0.011	8226.069	63732.910
	destination urban income	38747.600	7253.301	5.340	<0.001	24512.670	52982.540
	destination rural income	-33959.620	13713.500	2.480	0.013	-60872.990	-7046.246
constant	355741.600	175037.000	2.030	0.042	699259.700	-12223.490	

- Adjusted R² for 2000 interprovincial migration is 0.143;
- Adjusted R² for 2010 interprovincial migration is 0.192;
- The result shows that the linear regression model without taking logs on variables could not explain interprovincial migration very well.

2000 and 2010 comparison - short form data & enhanced and extended gravity model

Model	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
2000 interprovincial migration	origin urban population	0.166	0.113	1.460	0.144	-0.057	0.388
	origin rural population	0.943	0.083	11.370	<0.001	0.780	1.105
	destination urban population	0.636	0.113	5.610	<0.001	0.413	0.858
	destination rural population	0.017	0.083	0.200	0.839	-0.146	0.180
	distance	-	0.067	-	<0.001	-1.228	-0.965
	origin urban income	1.096	0.322	-3.480	0.001	-1.749	-0.487
	origin rural income	1.118	0.272	1.810	0.071	-0.042	1.025
	destination urban income	0.492	0.322	6.230	<0.001	1.371	2.633
	destination rural income	2.002	0.272	0.570	0.569	-0.379	0.689
	constant	0.155	0.912	3.350	0.001	1.261	4.839
	2010 interprovincial migration	origin urban population	3.050	0.091	5.260	<0.001	0.301
origin rural population		0.480	0.077	8.720	<0.001	0.519	0.820
destination urban population		0.669	0.091	10.330	<0.001	0.763	1.122
destination rural population		0.942	0.077	-4.300	<0.001	-0.480	-0.179
distance		-	0.053	-	<0.001	-1.106	-0.899
origin urban income		1.002	0.377	-4.550	<0.001	-2.455	-0.974
origin rural income		1.715	0.289	0.360	0.720	-0.463	0.671
destination urban income		0.104	0.377	11.470	<0.001	3.586	5.067
destination rural income		4.327	0.289	-4.970	<0.001	-2.004	-0.870
constant		-	1.052	0.520	0.604	-1.519	2.609

- *2000 interprovincial migration model (Model 1)*, adjusted R² is 0.653;
- *2010 interprovincial migration model (Model 2)*, adjusted R² is 0.737;
- both of the linear regression models predict the dependent variables significantly well, indicating the models are statistically precise and could explain interprovincial migration in 2000 and 2010.

2000 and 2010 comparison – short form data & enhanced and extended gravity model

$$M_{2000} = 21.117 \times p_{or}^{0.943} \times p_{ou}^{0.166} \times p_{dr}^{0.017} \times p_{du}^{0.636} \times I_{or}^{0.492} \times I_{ou}^{-1.118} \times I_{dr}^{0.155} \times I_{du}^{2.002} \times d^{-1.096}$$

$$M_{2010} = 1.725 \times p_{or}^{0.669} \times p_{ou}^{0.480} \times p_{dr}^{-0.330} \times p_{du}^{0.942} \times I_{or}^{0.104} \times I_{ou}^{-1.715} \times I_{dr}^{-1.437} \times I_{du}^{4.327} \times d^{-1.002}$$

Where,

- M_{2000} is the number of migrants from origin province's rural to destination province's urban areas in 2000;
- M_{2010} is the number of migrants from origin province's rural to destination province's urban areas in 2010.

2000 and 2010 comparison – Changes in interprovincial migration for every time's increase in dependent variables in 2000 and 2010

	Parameter		Changes in dependent variable	
	2000	2010	2000	2010
origin urban population	0.166	0.480***	1.122	1.395
origin rural population	0.943***	0.669***	1.923	1.590
destination urban population	0.636***	0.942***	1.554	1.921
destination rural population	0.017	-0.330***	1.012	0.796
distance	-1.096***	-1.002***	0.468	0.499
origin urban income	-1.118**	-1.715***	0.461	0.305
origin rural income	0.492	0.104	1.406	1.075
destination urban income	2.002***	4.327***	4.006	20.070
destination rural income	0.155	-1.437***	1.113	0.369

Note: *** denotes $p < 0.001$, ** represents $p < 0.01$, and * symbolises $p < 0.05$.

Migration stream comparison – 2010 long form data & multivariate linear regression

Model	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
urban-urban migration	origin urban population	0.635	0.079	8.02	<0.001	0.479	0.79
	origin rural population	0.16	0.064	2.49	0.013	0.034	0.286
	destination urban population	1.037	0.079	13.18	<0.001	0.883	1.192
	destination rural population	-0.438	0.064	-6.85	<0.001	-0.563	-0.312
	distance	-0.854	0.043	-19.78	<0.001	-0.939	-0.769
	origin urban income	-2.258	0.313	-7.21	<0.001	-2.872	-1.643
	origin rural income	1.119	0.24	4.67	<0.001	0.649	1.59
	destination urban income	3.738	0.312	12	<0.001	3.127	4.349
	destination rural income	-1.42	0.24	-5.92	<0.001	-1.891	-0.95
	constant	5.586	0.869	6.43	<0.001	3.88	7.292
urban-rural migration	origin urban population	0.804	0.104	7.7	<0.001	0.599	1.009
	origin rural population	0.12	0.085	1.42	0.156	-0.046	0.286
	destination urban population	0.249	0.104	2.4	0.016	0.046	0.453
	destination rural population	0.213	0.084	2.52	0.012	0.047	0.378
	distance	-1.045	0.057	-18.37	<0.001	-1.157	-0.933
	origin urban income	-1.491	0.413	-3.61	<0.001	-2.301	-0.682
	origin rural income	-0.067	0.316	-0.21	0.832	-0.687	0.553
	destination urban income	2.417	0.41	5.89	<0.001	1.612	3.223
	destination rural income	-0.89	0.316	-2.82	0.005	-1.51	-0.269
	constant	6.749	1.145	5.89	<0.001	4.501	8.997

- Model 3, *urban-urban migration stream*, adjusted $R^2 = 0.718$;
- Model 4, *urban-rural migration stream*, adjusted $R^2 = 0.566$;
- both multivariate linear regression models predict the dependent variables significantly well ($p=0.0000$).

Migration stream comparison – 2010 long form data & multivariate linear regression

Model	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
rural-urban migration	origin urban population	-0.071	0.116	-0.61	0.54	-0.298	0.156
	origin rural population	1.361	0.094	14.52	<0.001	1.177	1.545
	destination urban population	0.687	0.115	5.98	<0.001	0.461	0.912
	destination rural population	-0.219	0.093	-2.35	0.019	-0.402	-0.036
	distance	-0.991	0.063	-15.74	<0.001	-1.115	-0.868
	origin urban income	-3.535	0.457	-7.74	<0.001	-4.431	-2.638
	origin rural income	1.012	0.35	2.89	0.004	0.325	1.699
	destination urban income	4.455	0.455	9.8	<0.001	3.563	5.347
	destination rural income	-1.084	0.35	-3.1	0.002	-1.77	-0.397
	constant	6.049	1.268	4.77	<0.001	3.56	8.538
rural-rural migration	origin urban population	0.316	0.134	2.36	0.018	0.053	0.578
	origin rural population	0.954	0.108	8.81	<0.001	0.741	1.167
	destination urban population	0.586	0.133	4.41	<0.001	0.326	0.847
	destination rural population	0.036	0.108	0.33	0.741	-0.176	0.247
	distance	-1.302	0.073	-17.89	<0.001	-1.445	-1.16
	origin urban income	-0.951	0.528	-1.8	0.072	-1.987	0.086
	origin rural income	-1.046	0.404	-2.59	0.01	-1.84	-0.252
	destination urban income	3.457	0.525	6.58	<0.001	2.426	4.489
	destination rural income	-0.938	0.405	-2.32	0.021	-1.732	-0.144
	constant	5.395	1.466	3.68	<0.001	2.517	8.272

- Model 5, *rural-urban migration stream*, adjusted R² = 0.711;
- Model 6, *rural-rural migration stream*, adjusted R² = 0.634;
- both multivariate linear regression models predict the dependent variables significantly well (p=0.0000).

Migration stream comparison

$$M_{uu} = 2.666 \times p_{or}^{0.160} \times p_{ou}^{0.635} \times p_{dr}^{-0.438} \times p_{du}^{1.037} \times I_{ou}^{-2.258} \times I_{or}^{1.119} \times I_{du}^{3.738} \times I_{dr}^{-1.420} \times d^{-0.854}$$

$$M_{ur} = 8.534 \times p_{or}^{0.120} \times p_{ou}^{0.804} \times p_{dr}^{0.213} \times p_{du}^{0.249} \times I_{ou}^{-1.491} \times I_{or}^{-0.067} \times I_{du}^{2.417} \times I_{dr}^{-0.890} \times d^{-1.045}$$

$$M_{ru} = 4.236 \times p_{or}^{1.361} \times p_{ou}^{-0.071} \times p_{dr}^{-0.219} \times p_{du}^{0.687} \times I_{ou}^{-3.535} \times I_{or}^{1.012} \times I_{du}^{4.455} \times I_{dr}^{-1.084} \times d^{-0.991}$$

$$M_{rr} = 2.202 \times p_{or}^{0.954} \times p_{ou}^{0.316} \times p_{dr}^{0.036} \times p_{du}^{0.586} \times I_{ou}^{-0.951} \times I_{or}^{-1.046} \times I_{du}^{3.457} \times I_{dr}^{-0.938} \times d^{-1.302}$$

Where,

- M_{uu} is the number of migrants from origin province's urban to destination province's urban areas;
- M_{ur} is the number of migrants from origin province's urban to destination province's rural areas;
- M_{ru} is the number of migrants from origin province's rural to destination province's urban areas;
- M_{rr} is the number of migrants from origin province's rural to destination province's rural areas.

Migration stream comparison – Comparison of parameters and changes in the dependent variable for every time's increase in all the independent variables in different migration stream models

	Parameter					Changes in dependent variable				
	Model 3	Model 4	Model 5	Model 6	Model 2	Model 3	Model 4	Model 5	Model 6	Model 2
	urban-urban stream	urban-rural stream	rural-urban stream	rural-rural stream	interprovinci al migration	urban-urban stream	urban-rural stream	rural-urban stream	rural-rural stream	interprovinci al migration
origin urban population	0.635***	0.804***	-0.071	0.316*	0.480***	1.553	1.746	0.952	1.245	1.395
origin rural population	0.160*	0.120	1.361***	0.954***	0.669***	1.117	1.087	2.569	1.937	1.590
destination urban population	1.037***	0.249*	0.687***	0.586***	0.942***	2.052	1.188	1.610	1.501	1.921
destination rural population	-0.438***	0.213*	-0.219*	0.036	-0.330***	0.738	1.159	0.859	1.025	0.796
distance	-0.854***	-1.045***	-0.991***	-1.302***	-1.002***	0.553	0.485	0.503	0.406	0.499
origin urban income	-2.258***	-1.491***	-3.535***	-0.951	-1.715***	0.209	0.356	0.086	0.517	0.305
origin rural income	1.119***	-0.067	1.012**	-1.046**	0.104	2.172	0.955	2.017	0.484	1.075
destination urban income	3.738***	2.417***	4.455***	3.457***	4.327***	13.343	5.341	21.933	10.981	20.070
destination rural income	-1.420***	-0.890**	-1.084**	-0.938*	-1.437***	0.374	0.540	0.472	0.522	0.369

- China's internal migration is so highly directional and economic-induced that **rural/urban segment populations** are very responsive to China's inter-provincial migration.
- The (average) effect of rural/urban segment populations is markedly big, and **regional income disparity** in development is an increasingly significant factor, while **distance** gradually falls in prominence from 2000 to 2010.
- In analysing 2010's migration streams, **urban income** is playing a predominant role among the four categories of regional income; urban-rural migration stream has the steepest distance decay, while rural-urban migration stream has the least steep one.

- Rodrigue, J.-P., Comtois, C., Slack, B. (2009). *The Geography of Transport Systems*. London, New York: Routledge. ISBN 978-0-415-48324-7.
- Wang, Y., et al. (2013). "Evolution Characteristics of Population and Economic Gravity Centers in Tarim River Basin, Uygur Autonomous Region of Xinjiang, China." *Chinese Geographical Science* **23**(6): 765-772.
- Jenicek, V. (2010). "World population - development, transition." *Agricultural Economics-Zemедельска Ekonomika* **56**(1): 1-15.
- Fan, C. C. (2005). "Modeling interprovincial migration in China, 1985-2000." *Eurasian Geography and Economics* **46**(3): 165-184.
- GEDIK, A. ". "Mobility Transition (Test Of Zelinsky's Theory) and Economic and Demographic Factors: Japan and Turkey."
- Christian, W. and W. Braden (1966). "Rural Migration and Gravity Model." *Rural Sociology* **31**(1): 73-80.
- Claeson, C. F. (1969). "2-Stage Model of in-Migration to Urban Centres - Deductive Development of a Variant of Gravity Formulation." *Geografiska Annaler Series B-Human Geography* **51**(2): 127-138.
- Johnston, R. J. (1970). "Latent Migration Potential and Gravity Model - New-Zealand Study." *Geographical Analysis* **2**(4): 387-397.
- Ginsberg, R. B. (1972). "Incorporating Causal Structure and Exogenous Information with Probabilistic Models - with Special Reference to Choice, Gravity, Migration, and Markov Chains." *Journal of Mathematical Sociology* **2**(1): 83-103.
- Flores, M., et al. (2013). "Economic Liberalization and Contemporary Determinants of Mexico's Internal Migration: An Application of Spatial Gravity Models." *Spatial Economic Analysis* **8**(2): 195-214.
- Che, Y., et al. (2015). "The development of farm labor market in rural China." *China Agricultural Economic Review* **7**(2): 280-302.
- Fan, C. C. and M. Sun (2008). "Regional inequality in China, 1978-2006." *Eurasian Geography and Economics* **49**(1): 1-18.
- Fan, S., et al. (2009). *Regional inequality in China: trends, explanations and policy responses*, Routledge.

- Jian, T., et al. (1996). "Trends in regional inequality in China." China Economic Review **7**(1): 1-21.
- Kleinwechter, U. and H. Grethe (2015). "National Wage Trends and Migration in a Chinese Village Economy: A Micro Level Modeling Approach Based on a Composite Utility Function." American Journal of Agricultural Economics **97**(3): 701-726.
- Li, Y., et al. (2014). "Analysis of Determinants on China's Interprovincial Migration during 1985-2005." Sustainable Development of Industry and Economy, Pts 1 and 2 **869-870**: 1096-1105.
- Liu, S. Q., et al. (2014). "Influences on rural migrant workers' selection of employment location in the mountainous and upland areas of Sichuan, China." Journal of Rural Studies **33**: 71-81.
- Liu, Y., et al. (2014). "Interprovincial Migration, Regional Development and State Policy in China, 1985-2010." Applied Spatial Analysis and Policy **7**(1): 47-70.
- Mai, Y. H., et al. (2014). "The economic effects of facilitating the flow of rural workers to urban employment in China." Papers in Regional Science **93**(3): 619-642.
- Song, W. and M. L. Liu (2014). "Assessment of decoupling between rural settlement area and rural population in China." Land Use Policy **39**: 331-341.
- Su, W. L., et al. (2015). "Household-level linkages between off-farm employment and agricultural fixed assets in rural China." China Agricultural Economic Review **7**(2): 185-196.
- Sun, W. K., et al. (2014). "Income inequality and mobility of rural households in China from 2003 to 2006." China Agricultural Economic Review **6**(1): 73-91.
- Zhu, J. M. (1998). "Rural out-migration in China: A multilevel model." Migration, Urbanization, and Development : New Directions and Issues: 157-186.



Thank you!

Any Questions?

The Original Gravity Model of Migration

The original form of the gravity model,

$$m_{ij} = k \frac{p_i p_j}{d_{ij}^2}$$

Where,

- k is the parameter;
- m_{ij} is migration flow between place i and place j ;
- p_i is the population in origin place i ;
- p_j is the population in destination place j ;
- d_{ij} is distance between place i and place j , or represents friction of distance.

-The original gravity model takes little account of the directional movement of migration:

for instance,

$$m_{ij} = k \frac{p_i p_j}{d_{ij}^2} = m_{ji}$$

The General Gravity Model of Migration

- The general form of the gravity model,

Where,

$$m_{ij} = k \frac{p_i^a p_j^b}{d_{ij}^c}$$

- a , b , c and k are parameters;
- m_{ij} is migration flow between place i and place j ;
- p_i is the population in origin place i ;
- p_j is the population in destination place j ;
- d_{ij} is distance between place i and place j .

-Distance does not necessarily follow the inverse-square law;

-Populations in two different locations do not necessarily have the same impact on the magnitude of migration, namely a does not necessarily equal to b or 1 .

-The general gravity model takes little account of other socio-economic factors other than total population, such as income.

The Extended Gravity Model of Migration

- The extended form of the gravity model,

$$m_{ij} = k \frac{p_i^a p_j^b}{d_{ij}^c} I_i^f I_j^g$$

Where,

- a, b, c, k, f and g are parameters;
- m_{ij} is the number of migrants from origin i to destination j ;
- p_i and p_j are the population in origin i and destination j ;
- I_i and I_j are the income in origin i and destination j ;
- d_{ij} is distance between place i and place j .

-Inclusion of the economic factor into the model, i.e. origin and destination income.

-This extended gravity model takes little account of rural-urban divide.