Spatial Analysis of Chinese Fixed Asset Investment

Dr. William Chow

27 July, 2011

Executive Summary

- This short paper evaluates fixed asset investment in Chinese provinces from 1997 to 2009 and relates the development with behavior of other economic variables.

- The quantitative analysis is done using spatial econometric techniques. Specifically, a fixed effects spatial lag model (SAR-FE) is fitted with our panel data. The SAR-FE model essentially indicates that investment in a province i is subject to a multiple order impact of contemporary changes in explanatory variables of neighboring provinces, near and far.

- The estimated model is statistically significant and delivers good fit to the data.

- The signs of coefficients are intuitive and consistent with what one might expect.

- Real fixed asset investment (FAI) increases with real output per worker, the amount of human capital in the region, the size of infrastructural networks, and the freight traffic in the region.

- Inflation and wage pressure, on the other hand, have a deterring effect on FAI.

- The coefficient of the spatial lag term (ρ) is positive indicating that (i) investments in nearby regions are similar rather than dissimilar, and (ii) there are possible spillovers in investments across regions.

- On considering the current status of various provinces in the West/Northwest and the implications of the SAR-FE model, Sichuan, Chongqing and Shaanxi stand in the best position to attract investment. Tibet and Ningxia have much less competitiveness in this aspect.

1. Introduction

This article surveys changes in investment, production, prices, human capital and infrastructural developments in Chinese provinces between 1997 and 2009. Using spatial econometric techniques, we evaluate the determinants of fixed asset investments in China and comment on the direction and scope of investment in less developed provinces. Not surprisingly, we find output, human capital and infrastructure development to be conducive to investment while inflation and wage pressures are hindering factors.

2. The Dataset

All the data used in this exercise are extracted from the Chinese Statistical Yearbook, and the sample runs from 1997 to 2009. We focus on the regional level, and provincial data are the natural choice for this study. The economic variables analyzed are:

(1) Real Fixed Asset Investment (FAI): The FAI figures are quoted in RMB 100 million and in nominal terms. There are price indices for provincial FAI data but they are incomplete. We choose to deflate FAI using the CPI recompiled with a base year of 1994.

(2) Real Regional Output (RGDP) per Worker: Like FAI, the regional outputs are (RMB 100 million) deflated using the CPI index. They are then divided by year-end employment figures (10,000 persons) from the primary, secondary and tertiary industries.

(3) CPI Index: Recompiled using 1994 as the base year.

(4) Real Wages: These are the average wages of staff and workers by status of registration and region. The nominal figures (in RMB) are then deflated by the CPI index.

(5) Number of Graduates from Higher Education Institutions: These are expressed as ratio to population (10,000 persons).

(6) Length of Railways in Operation, and

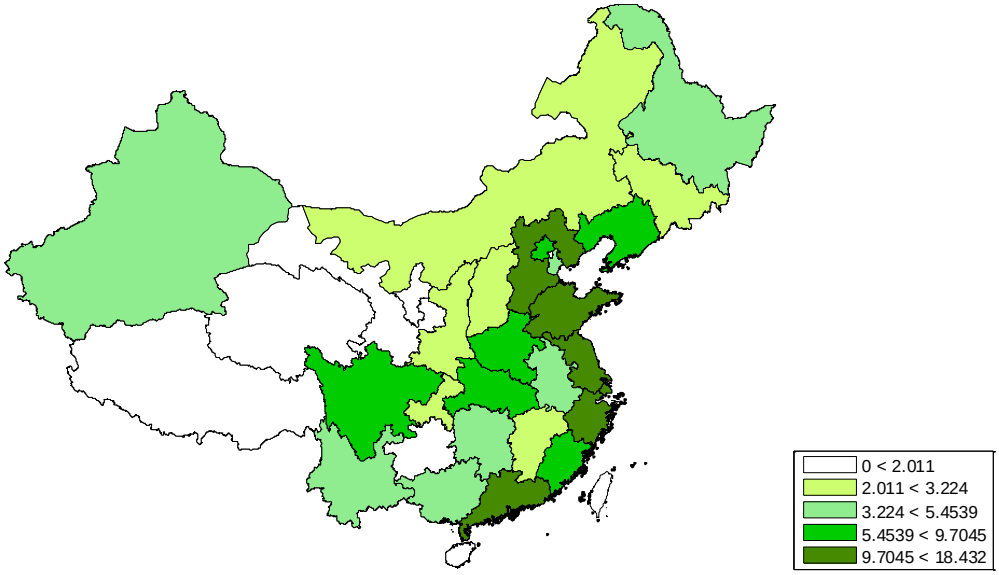
(7) Length of Highways:
Both (6) and (7) are measured in kilometers.

(8) Total Freight Traffic: Include all traffic by railways, highways, waterways and aviation. This is measured in 10,000 tons.

The following figures illustrate the status of each of these economic variables in 1997 and 2009. The colors go from light to dark and each tone signifies a 20th percentile with the lightest being the lowest 20%.

Figure 1: Fixed Asset Investment (Real and rebased, RMB 100 million)

1997



2009

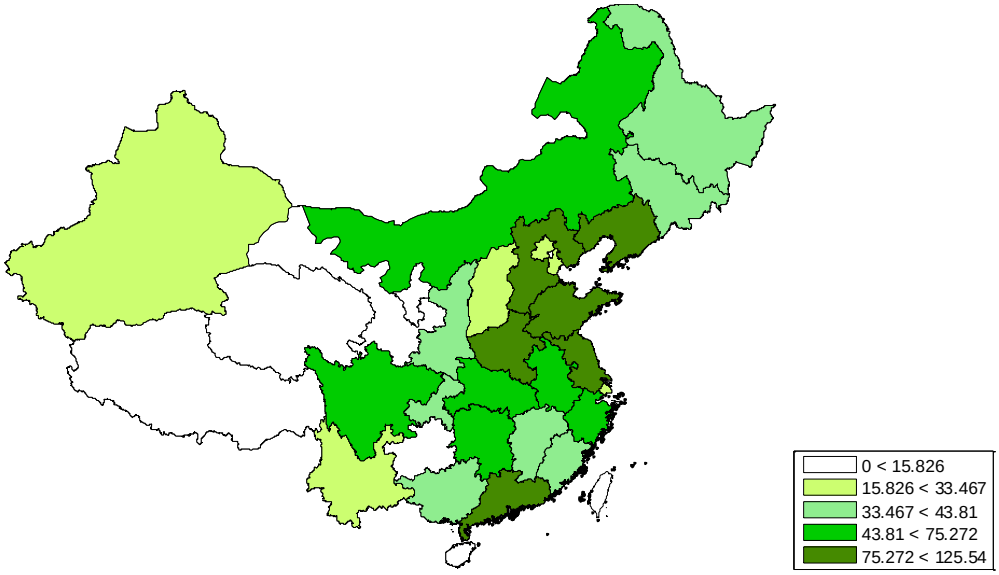
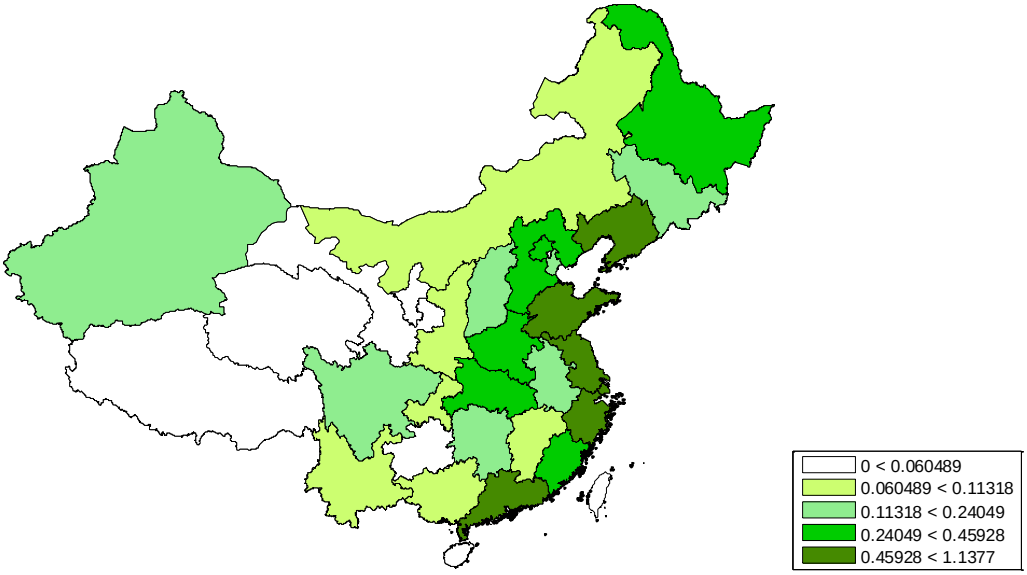


Figure 2: Real Output Per Worker– All Industries (RMB 10,000)

1997



2009

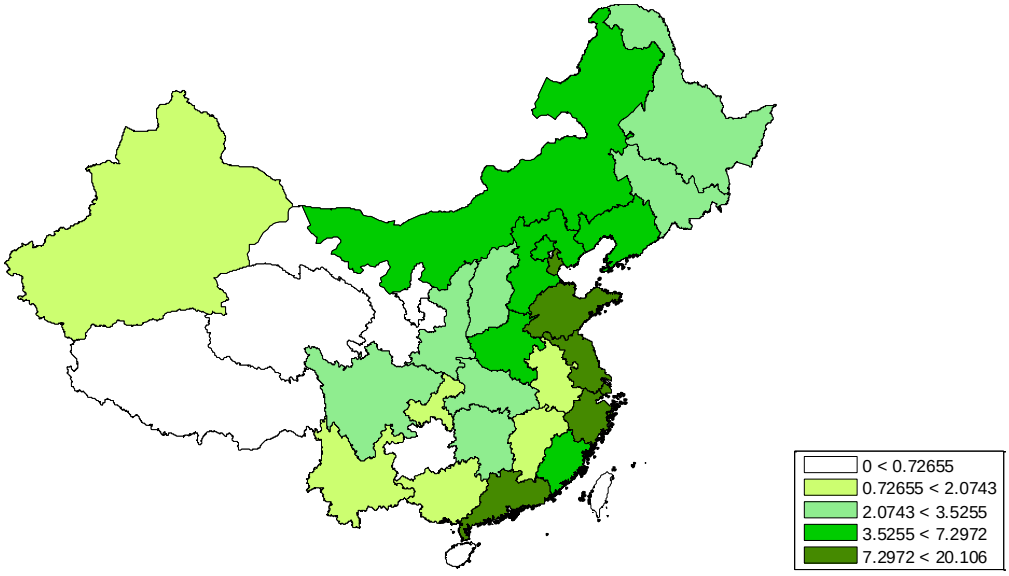
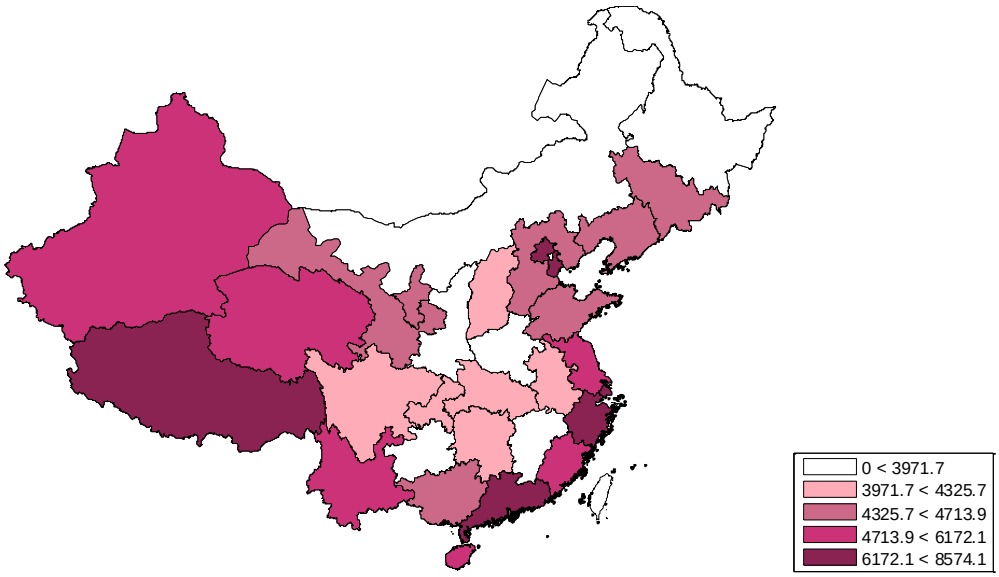


Figure 3: Real Wage – All Industries (Average, RMB)

1997



2009

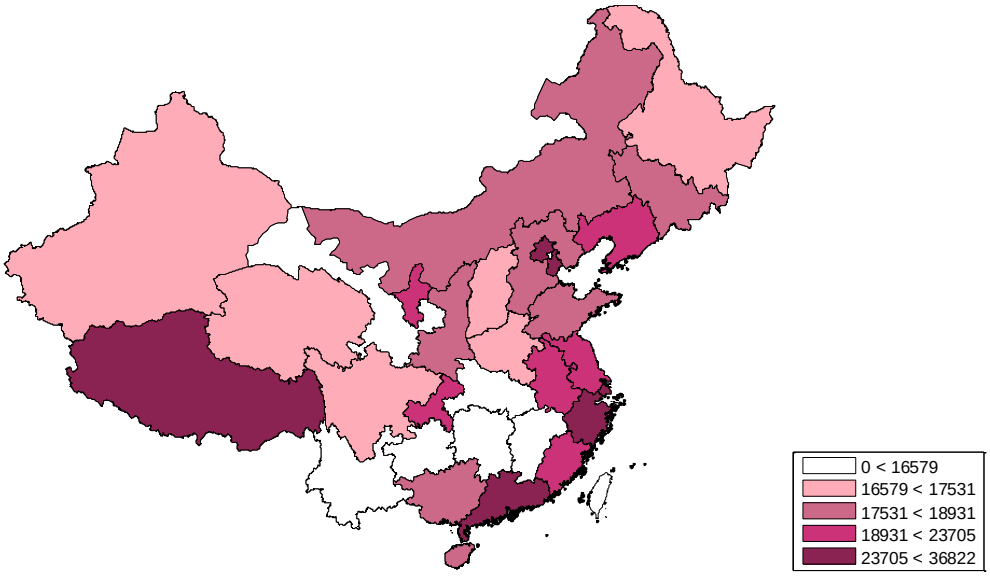
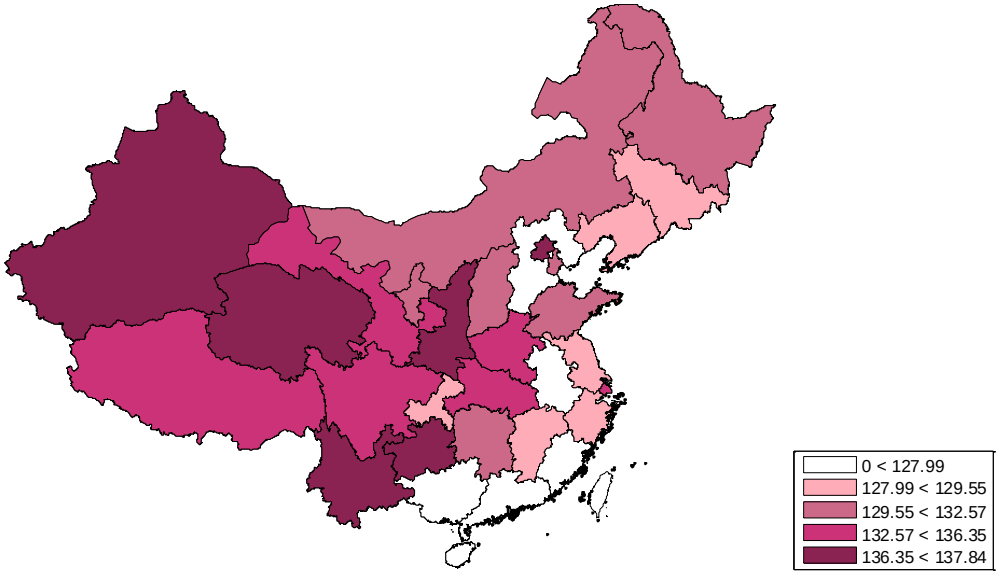


Figure 4: Consumer Price Index (Rebased)

1997



2009

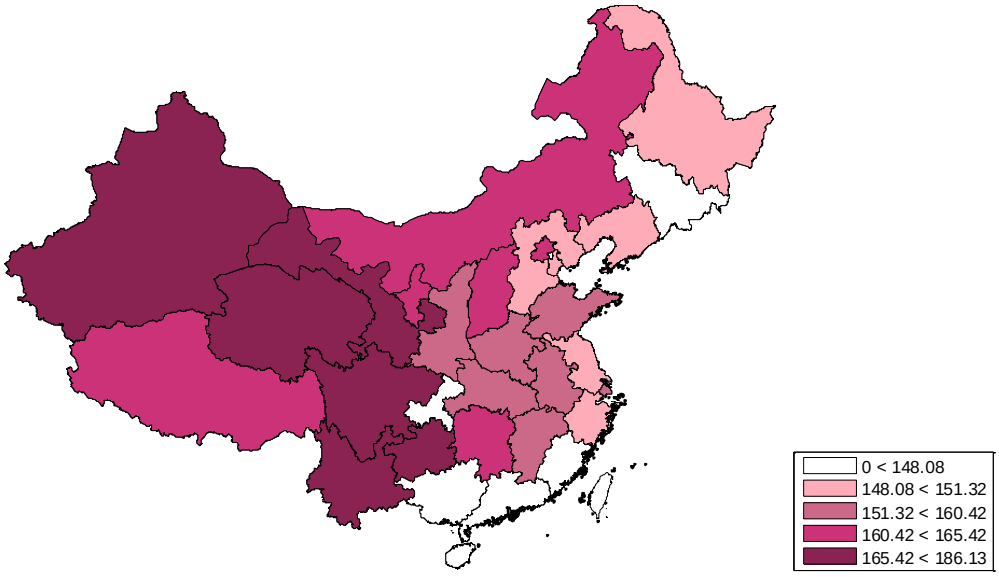
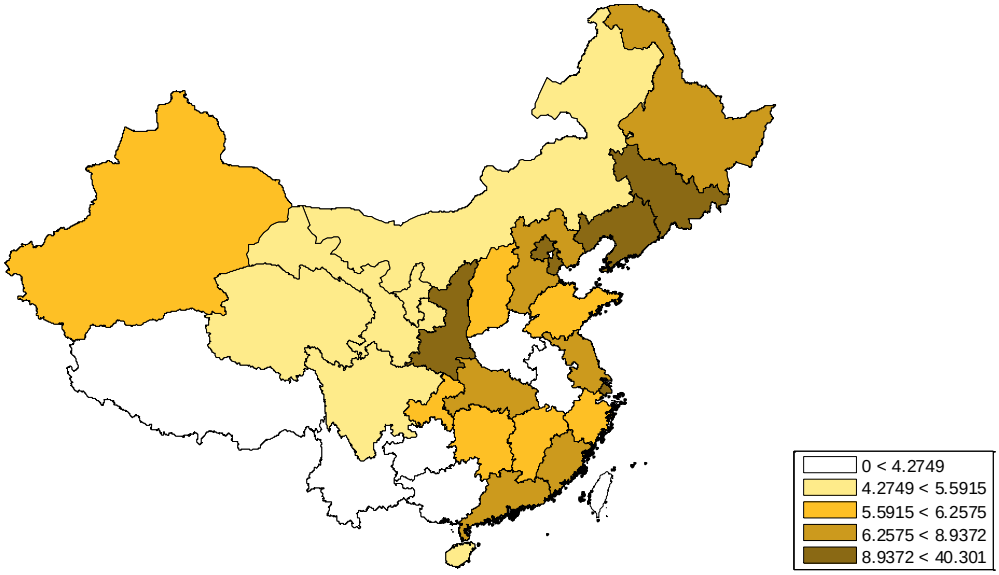


Figure 5: Higher Education Graduates per 10,000 Persons

1997



2009

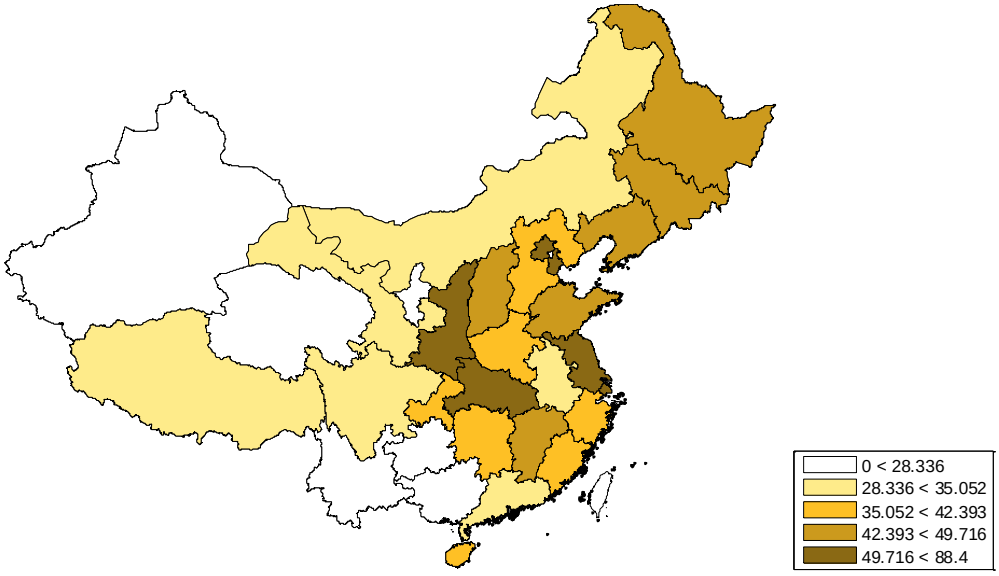
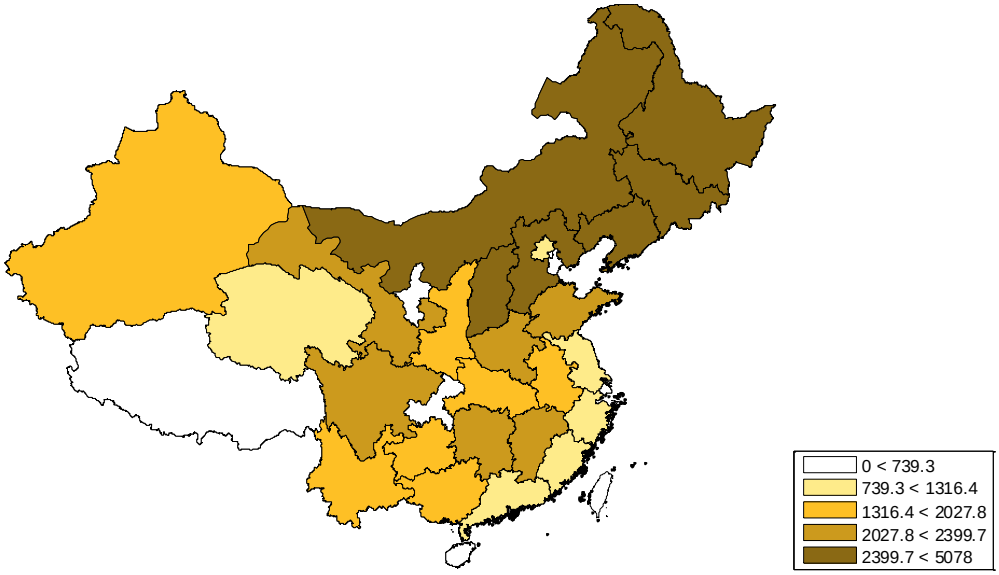


Figure 6: Infrastructure – Railroads in Operation (kilometers)

1997



2009

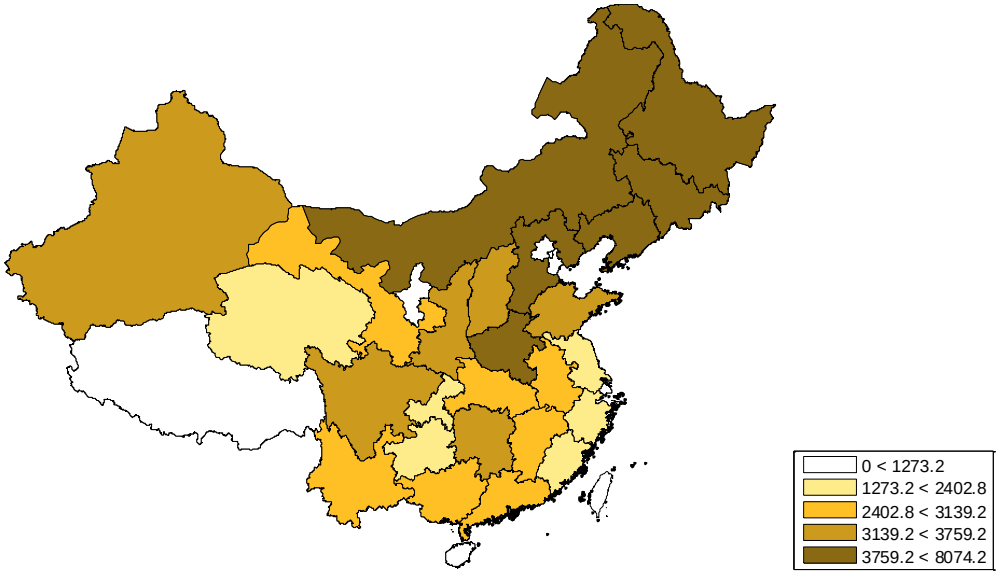
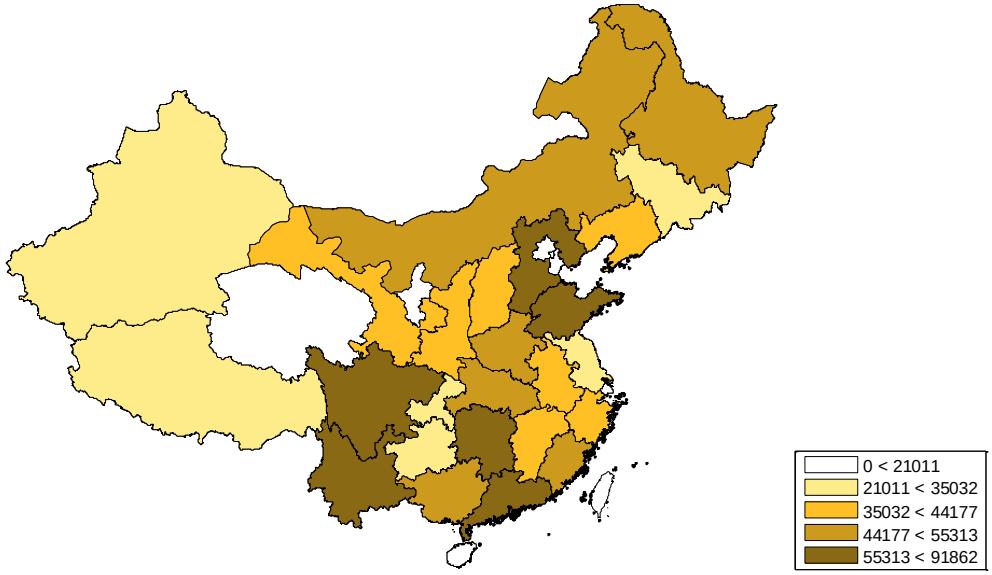


Figure 7: Infrastructure – Highways in Operation (kilometers)

1997



2009

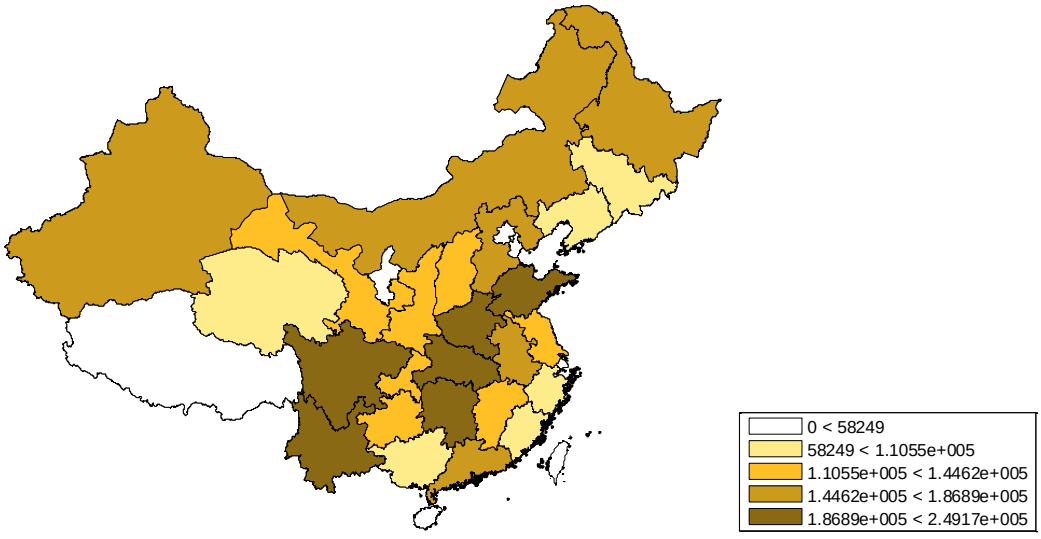
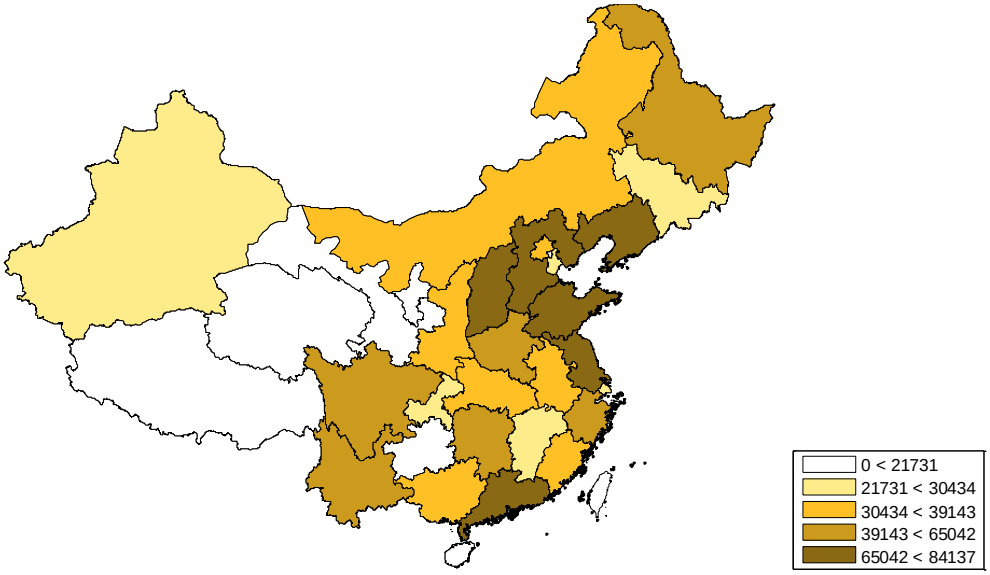
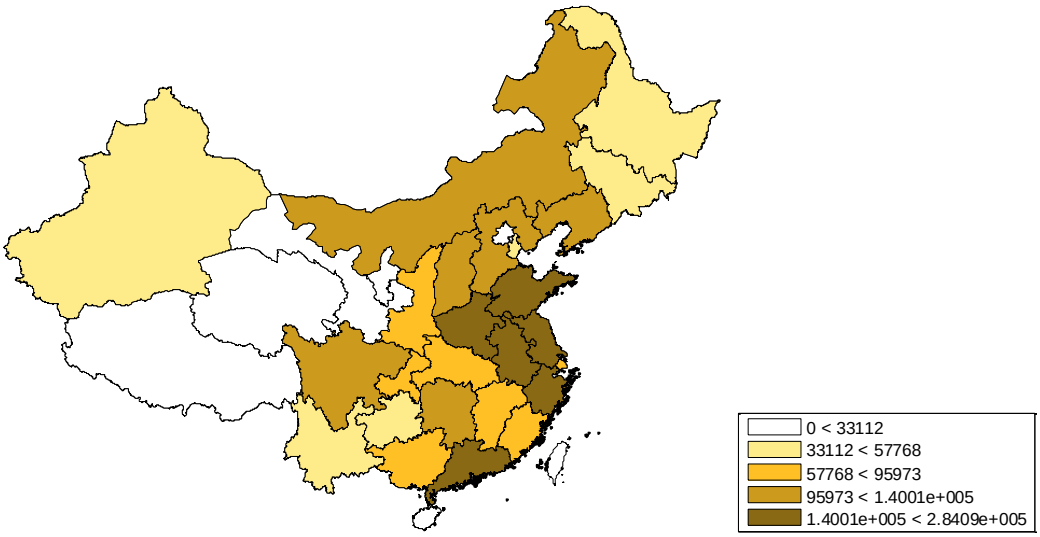


Figure 8: Freight Traffic (10,000 tons)

1997



2009



In brief:

- Investment and production seem to have strengthened the North-South nexus that stretches from Inner Mongolia to Guangdong. The West and Northwest provinces remain relatively under-developed.
- Inflation remains a problem for the West all along. Price pressure intensifies in the North and in Central China. But Guangdong has the smallest price pressure nationwide despite its dominance in production.
- Real wages have moved eastwards with the North/Northeast witnessing the fastest growth in wage pressure.
- Higher education graduates as ratio to population – a proxy of human capital in our study – shows a mild centripetal pattern during the sample period. They concentrate on areas between Beijing and Shanghai and along the Shaanxi-Hubei region as at 2009.
- There is little change in the railway network where the North and Northeast have a relative dominance.
- The highways network has clustered in Central and Southwest China.
- The freight traffic shares a similar pattern with real output and is dominated by Guangdong and the coastal regions in the East.

3. The Fixed Effects Spatial Lag Model

Next, we explore the set of determinants of FAI by doing a panel data analysis. Instead of an ordinary panel data model, we specify one with a spatial lag term. That is, there is a structure in the model that governs how investment in nearby provinces would affect FAI in the subject region. Spatial panel data models are relatively recent econometric methods, and useful references can be found in Elhorst (2003, 2010).

The Fixed Effects Spatial Lag Model (SAR-FE) takes the form:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij}\, y_{jt} + \alpha_i + \gamma_t + X_{it}\beta + \varepsilon_{it}, \qquad (1)$$

$$E(\varepsilon_{it}) = 0,$$

$$E(\varepsilon_{it}\, \varepsilon_{it}{}') = \sigma^2 I_N.$$

The major features of the SAR-FE model are summarized as follow:

- Without the first term on the RHS of equation (1), the model is just an ordinary fixed effects panel data model[1].

[1] The estimation results of the non-spatial version of equation (1) can be found in the Appendix.

- The terms w_{ij} are elements of the spatial contiguity weight matrix, compiled according to the procedures described in the Appendix. The weight matrix is row-standardized to give proportional weights to units having different amount of neighbors.
- The model has a built-in contemporaneous feedback between neighboring y_j and y_i. There are also not-so-obvious feedbacks of other variables contained in X to y_i (to be discussed below).
- α_i and γ_t are the fixed effects pertinent to the units i and time t, respectively.

The SAR-FE model is estimated with maximum likelihood methods, details of which are in Elhorst (2003, 2010). The dependent variable is Real FAI, and the explanatory variables include all those discussed in the previous section. Table 1 shows the results of the estimation which turn out to be comparable to those of an ordinary fixed effects model (see Appendix).

Table 1: Estimation Results of Fixed Effects Spatial Lag Model

Model Type	SAR-FE		
Dependent Variable	Real Fixed Asset Inv.		
R^2	0.9484		
Squared Corr. Coef.	0.7421		
Log-likelihood	-1184.50		
No. of N	31		
No. of T	13		
Variable	**Coefficient**	**Asym t-stat**	**Probability**
RGDP per Worker	2.7566	13.8523	0.0000 **
CPI	-0.2115	-1.8713	0.0613 *
Real Wage	-0.0008	-4.1332	0.0000 **
Human Capital	0.0964	2.1375	0.0326 **
Infra-Rail	0.0021	2.2705	0.0232 **
Infra-Road	0.0001	7.3957	0.0000 **
Freight	0.0002	9.2431	0.0000 **
Spatial Lag	0.1216	3.0855	0.0020 **

Remarks: One (Two) arterisk indicates significance at 0.10 (0.05) level.

The SAR-FE model is highly significant and gives very good fit to the Real FAI data. The R^2 is close to 0.95 and all variables, except CPI, are significant at the 5% level. The signs of the coefficient are consistent with what one might expect. **Real FAI increases with real output per worker, the amount of human capital in the region, the size of infrastructural networks, and the freight traffic in the region. Inflation and wage pressure, on the other hand, have a deterring effect on FAI. The coefficient of the spatial lag term (ρ) is positive indicating that (i) investments in nearby regions are similar rather than dissimilar, and (ii) there are spillovers in investments across regions**.

Unlike non-spatial fixed effects model, the (partial) impact of a change in variable k usually does not equal to the coefficient of that variable ($\hat{\beta}_k$). To see this, we can vectorize (1) and rewrite the equation as

$$Y_t = \rho W Y_t + \alpha + \gamma_t + X_t\beta + \varepsilon_t$$
$$\Rightarrow Y_t = (1 - \rho W)^{-1} X_t \beta + (1 - \rho W)^{-1} \tilde{\varepsilon}_t$$

where the adjusted error term $\tilde{\varepsilon}_t$ captures the unit and time fixed effects for notation convenience. So instead of β, it is $(1 - \rho W)^{-1}\beta$ that is of relevance. The relationship between X and Y can be summed up by:

(1) Average Direct Impact:
This is the average (over all regions i) impact of a change in variable k in region i on the real FAI of region i.

(2) Average Indirect Impact:
This is the impact of a change in variable k in regions $j \neq i$ on the real FAI of region i (averaged over all regions).

(3) Average Total Impact:
The sum of average direct impact and average indirect impact.

The breakdown of these effects for our SAR-FE model is stated in Table 2 below. As we can see, the average indirect impact account for approximately 12% of the total impact for most of the explanatory variables.

Table 2: Breakdown of Average Total, Direct and Indirect Effects

Variable	Average Direct	Average Indirect	Average Total
RGDP per Worker	2.7664	0.3938	3.1602
CPI	-0.2065	-0.0286	-0.2351
Real Wage	-0.0008	-0.0001	-0.0010
Human Capital	0.0955	0.0133	0.1087
Infra-Rail	0.0022	0.0003	0.0025
Infra-Road	0.0001	0.0000	0.0001
Freight	0.0002	0.0000	0.0002

We wrap up our discussion with a survey of the status of various provinces in the West/Northwest and consider the implication of the SAR-FE model on their prospects in attracting investments.

Table 3: Model Implication on the Western Provinces

Region	Real FAI (spatial lag)	RGDP per worker	CPI	Real Wage	Human Capital	Railways	Highways	Freight
Signs of Coeff.		+	-	-	+	+	+	+
Percentiles (0 to 100) 2009								
Chongqing	40-60	20-40	0-20	60-80	40-60	20-40	40-60	40-60
Sichuan	60-80	40-60	80-100	20-40	20-40	60-80	80-100	60-80
Guizhou	0-20	0-20	80-100	0-20	0-20	20-40	40-60	20-40
Yunnan	20-40	20-40	80-100	0-20	0-20	40-60	80-100	20-40
Tibet	0-20	0-20	60-80	80-100	20-40	0-20	0-20	0-20
Shaanxi	40-60	40-60	40-60	60-80	80-100	60-80	40-60	40-60
Gansu	0-20	0-20	80-100	0-20	20-40	40-60	40-60	0-20
Qinghai	0-20	0-20	80-100	20-40	0-20	20-40	20-40	0-20
Ningxia	0-20	0-20	60-80	60-80	0-20	0-20	0-20	0-20
Xinjiang	20-40	20-40	80-100	20-40	0-20	60-80	60-80	20-40

The table summarizes the current states (as of 2009) of the provinces on a range of socio-economic factors, indicated as percentiles relative to national levels. For instance, Tibet's output per worker is within the 0-20 percentiles while her price pressure is within the 60-80 percentiles. Since the former encourage investment and the latter discourages investment, Tibet is among those regions with bleak prospect.

Considering the conducive and deterring factors, Chongqing, Sichuan and Shaanxi seem to be the regions with more potential in attracting investment. Tibet and Ningxia look particularly impotent in this respect.

4. Reference

Elhorst, J.P. (2003). Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review,* 26(3) 244-268.

Elhorst, J.P. (2010). Spatial Panel Data Models. in M.M. Fischer and A. Getis (eds.) *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications.* Springer-Verlag, Berlin.

5. Appendix

5.1 Code for Chinese Provinces

1. Beijing	2. Tianjin	3. Hebei
4. Shanxi	5. Inner Mongolia	6. Liaoning
7. Jilin	8. Heilongjiang	9. Shanghai
10. Jiangsu	11. Zhejiang	12. Anhui
13. Fujian	14. Jiangxi	15. Shandong
16. Henan	17. Hubei	18. Hunan
19. Guangdong	20. Guangxi	21. Hainan
22. Chongqing	23. Sichuan	24. Guizhou
25. Yunnan	26. Tibet	27. Shaanxi
28. Gansu	29. Qinghai	30. Ningxia
31. Xinjiang		

5.2 Spatial Contiguity Weight Matrix

The Spatial Weight Matrix is compiled using binary numbers (1) for neighbors defined as provinces with common boundaries or common corners (Queen's Neighbors) and (0) for non-neighbors. For example, the sub-matrix below gives the weights for 5 sample provinces/cities.

Queens Neighbors		Beijing 北 京	Tianjin 天 津	Hebei 河 北	Shanxi 山 西	Inner Mon 内蒙古
Beijing	北 京	0	1	1	0	0
Tianjin	天 津	1	0	1	0	0
Hebei	河 北	1	1	0	1	1
Shanxi	山 西	0	0	1	0	1
Inner Mon	内蒙古	0	0	1	1	0

The weight matrix can be row-standardized by replacing the entries w_{ij} with $w_{ij}^{*} = w_{ij} / \sum_{j} w_{ij}$.

5.3 Panel Data Analysis without Spatial Component

Table 4: Estimation Results of Fixed Effects Model with no Spatial Lag

Model Type	FEDV		
Dependent Variable	Real Fixed Asset Inv.		
R^2	0.9475		
Squared Corr. Coef.	n.a.		
Log-likelihood	-1187.20		
No. of N	31		
No. of T	13		
Variable	**Coefficient**	**t-stat**	**Probability**
RGDP per Worker	2.8130	13.900	0.0000 **
CPI	-0.2774	-2.459	0.0144 **
Real Wage	-0.0008	-4.087	0.0000 **
Human Capital	0.1166	2.549	0.0112 **
Infra-Rail	0.0017	1.831	0.0680 *
Infra-Road	0.0001	7.033	0.0000 **
Freight	0.0002	10.38	0.0000 **
Spatial Lag	n.a.	n.a.	n.a.

Remarks: One (Two) arterisk indicates significance at 0.10 (0.05) level.