CHINA'S ENERGY INTENSITY AND ITS DETERMINANTS AT THE PROVINCIAL LEVEL

By

Xin Zhang

Bachelor of Science in Electronics Engineering, Fudan University, 2003
Master of Science in Management, Ecole des Mines de Paris, 2006
Master of Science in Real Estate Development, Massachusetts Institute of Technology, 2008

Submitted to the Department of Civil and Environmental Engineering and Department of Urban Studies and Planning in partial fulfillment of the requirements for the degrees of

Master of Science in Transportation
Master of City Planning at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2009

Author______________________________
Department of Civil and Environmental Engineering
Department of Urban Studies and Planning
February 3, 2009

Certified by______________________________
Karen R. Polenske
Professor of Regional Political Economy and Planning
Thesis Supervisor

Certified by______________________________
Moshe Ben-Akiva
Professor of Civil and Environmental Engineering
Thesis Reader

Accepted by______________________________
Joseph Ferreira
Chair, MCP Committee
Department of Urban Studies and Planning

Accepted by______________________________
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

CHINA'S ENERGY INTENSITY
AND ITS DETERMINANTS AT THE PROVINCIAL LEVEL

By
Xin Zhang

Submitted to the Department of Civil and Environmental Engineering and Department of Urban Studies and Planning on February 3, 2009 in partial fulfillment of the requirements for the degrees of
Master of Science in Transportation and Master of City Planning

Abstract

Energy intensity is defined as the amount of energy consumed per dollar of GDP (Gross Domestic Product). The People's Republic of China's (China's) energy intensity has been declining significantly since the late 1970s. The first part of this thesis is a direct descriptive statistical analysis at both the national level and provincial level. Regional variation in terms of energy production and consumption is pronounced especially between the inland provinces and coastal provinces. The role of railway transportation in moving coal from the inland regions to the coastal regions is studied. I find that the capacity limit of railways has indirectly affected the decline of China's energy intensity.

The second part adopts methodology similar to that used by Sue Wing (2008), as well as Metcalf (2008) paper. I have created two indexes to decompose changes in energy intensity into intra-province (efficiency) and inter-province (structural) effects. Efficiency change refers to the energy-intensity reduction within a particular province due to factors such as fuel prices, temperature, economic sector shift, infrastructure investment, etc. The structural change refers to the change of energy intensity due to the growth of the share of provincial output in total GDP, such as when less energy-intensive provinces increase their share of output in total GDP. I find that the efficiency change has outperformed the structural change over the sample period of 1988-2006.

The third part identifies and tests the potential factors that may positively or negatively contribute to the reduction of energy intensity within each province. As stated above, I collected a panel dataset of 29 provinces from 1988 to 2006 (= 551 observations) for analysis. I present results from the fixed-effect regression model of the energy intensity on economic- and temperature-related variables, namely, fuel prices, per capita income, heating degree days, cooling degree days, time trend, capital-labor ratio, and investment-capital ratio. The provincial analysis shows that the increases in per capita income, time trend and capital-labor ratio have played an important role in the decline of China's energy intensity. I further separated the 29 provinces into three major economic regions and conducted the same analysis. I found that regional-specific characteristics and regional variance in response to the energy use have been magnified.

Thesis Supervisor: Karen R. Polenske
Title: Professor of Regional Political Economy and Planning

Acknowledgements

I would like to extend my sincere thanks to Professor Karen R. Polenske for supporting me as her research assistant for the two and half years' study. She opened the door for me to pursue this subject and I'm amazed at how she can always give such constructive advice from just a few minutes' talk. She is very devoted to the field of energy study and very rigorous with the research. Being one of her students is something I will be proud of for a long time.

I own a great debt of gratitude to my reader, Professor Ian Sue Wing. He has impressed me with his extensive knowledge and his passion for the energy field. His advice and comments have constantly helped me to improve and I sincerely hope that this thesis will meet his standards.

Table of Contents

List of Figures

List of Tables

Chapter I

Introduction

Energy intensity is defined for this study as the energy consumption per unit of gross domestic product (GDP). Polenske (2005) finds that the energy intensity of developed countries, such as the United States, has been declining steadily in the past decade, while that of the developing countries, such as Brazil and India has either fluctuated or slightly increased, as illustrated in Figure 1.

As noted below, literature reviewers identify two underlying factors that contribute to such changes: energy efficiency change and economic structural change. The service sector tends to be less energy intensive than the manufacturing and construction sectors, which are, in turn, more energy intensive than the agriculture and other primary sectors. The shifts in sectoral composition from agriculture to industry to services that accompany economic growth and development would therefore lead one to expect that energy intensity first increases and then decreases as countries grow richer (Hofman and Labar, 2007). The trend in China's energy intensity has exhibited a different pattern, however, falling by about 67% between 1978 and 2003 (Polenske, 2007). The aim of this study is to reconcile this fact with the theory.

What are the primary factors that contributed to this significant reduction of China's energy intensity? And further, what caused the national aggregate energy intensity to rebound since year 2002? Should this decade-long critical improvement be attributed to technology innovation, government environmental policies, state-level controls, industrial concentration, or more broadly, to the effect of China's gradual switching from a state-controlled economy to a market economy? If we can identify these factors, we can go one step further and analyze and focus on the specific roles that government policy, technology and innovation, energy prices, etc. have played and their impacts. The policy implications however will not be the focus of this study.

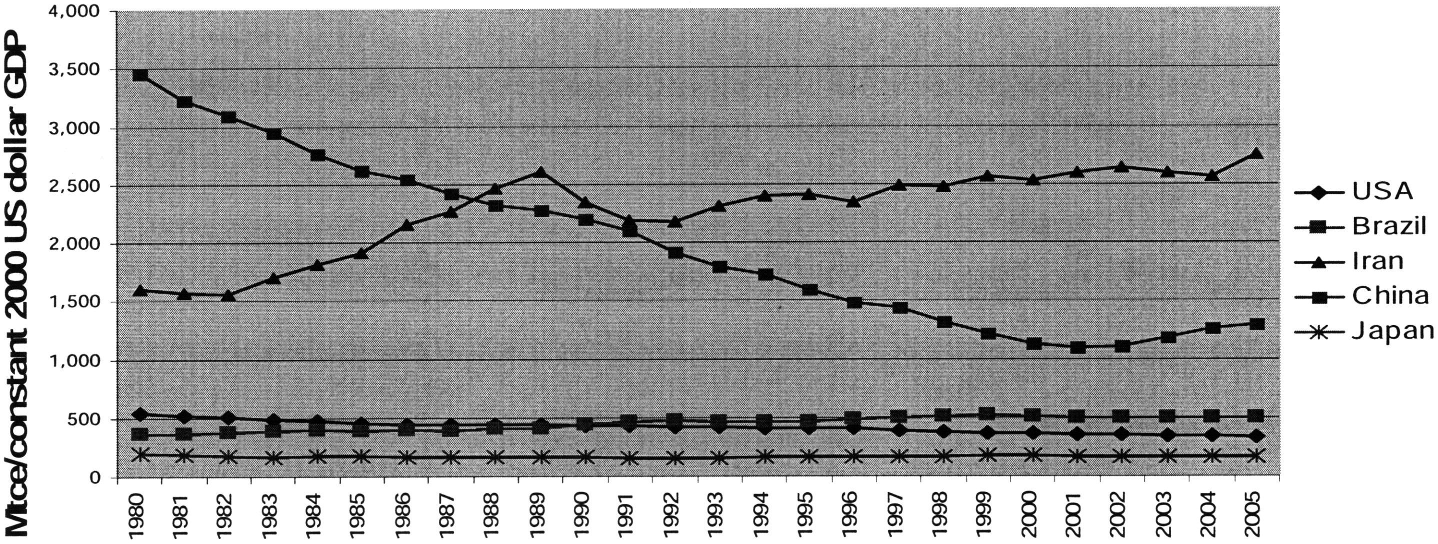


Figure 1: Energy Intensity of Selected Countries, 1980-2005

Notes: (1) GDP = gross domestic product, measured in 2000 US dollars using market exchange rates
(2) Mtce = Million (metric) tons of coal equivalent.
(3) Units of energy consumption were converted from British thermal units to Mtce (1 Btu = .036 Mtce)
Source: Compiled by the Multiregional Planning Research Team at MIT, from Energy Information Administration (2007) data.

1.1 The Focus of this Study

My goal in this study is to contribute to our understanding of China's energy intensity change by undertaking a statistical analysis using provincial level data. First, regional variation in terms of energy production and consumption has been pronounced especially between the inland provinces and coastal provinces. I analyze the transportation of coal by the national railway system from inland coal-producing regions to coastal coal-consuming regions. I find a constraint of improving energy intensity posed by the lack of transportation infrastructure investment.

Second, by adopting a similar methodology to the one used by Sue Wing (2008), I create two indexes to decompose changes in energy intensity into intra-province (efficiency) and inter-province (structural) effects. The former effect is referred to as the *within* index, while latter effect is referred to as the *between* index. Within index captures the change of aggregate energy intensity caused by the provincial energy-intensity change, weighted by the share of provincial energy use to national energy use. This change is normally attributed to factors such as fuel prices, weather, local infrastructure investment, economic sector shift, etc. The between index captures the change of aggregate energy intensity caused by the share of provincial output change in total GDP, also weighted by the share of provincial energy use to national energy use. The major difference from previous literature (Metcalf, 2008) is that I conduct the decomposition analysis using provincial (Beijing, Jilin, Guangdong, etc) economic factors, rather than sectoral (agriculture, industry, service, etc) economic factors.

Finally, I identify and test the various independent variables that may positively or negatively contribute to the reduction of energy intensity. In this study, I apply the decomposition analysis, instead of other methods (e.g., shift-share analysis), in order to be consistent and comparable with other research. I collect a panel dataset of 29 provinces[1] from 1988 to 2006 (551 observations) for analysis. I present results from the fixed-effect model regression of the energy

---

[1] These are Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia, Liaoning, Jilin, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Hainan, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang. Tibet, Hong Kong, and Macau are not part of this study due to data limitations.

intensity on economic- and weather-related variables, such as fuel prices, per capita income, heating degree days, etc. With provincial data, my analysis shows that the increase of per capita income and the time trend have played an important role in the decline of China's energy intensity. Heating degree days, capital-labor ratio and investment-capital ratio are positively related to energy intensity. Through an aggregation of the 29 provinces into three economic regions (north, south and west), I find that fuel price responds negatively. For per capita income, capital-labor ratio and investment-capital ratio, I include the quadratic terms in the regression to capture the non-linear effect.

1.2 Literature Review

The intensity with which countries use energy has been the subject of extensive scrutiny. Looking first at the United States, Sue Wing (2008) decomposed changes in energy intensity into shifts in the structure of sectoral composition and adjustments in the efficiency of energy use within individual industries. He found that interindustry structural change was the principal driver of the decline at the aggregate level, while the intraindustry efficiency improvements have played a more important role in the post-1980 period. Boyd and Roop (2004) first used the Fisher Ideal Index as the basis for the decomposition of energy-intensity changes into changes in energy efficiency and economic activity. Metcalf (2008) further constructed and analyzed U.S. indexes at both national and state levels over a much longer period, between 1970 and 2003. He identified weather- and economic-related variables and used regression analysis to measure their effects on the energy-intensity change. He found that rising per capita income contributes to improvements in energy efficiency and intensity and that price changes also play a key role.

For China, Lin and Polenske (1991, 1995) applied the shift-share analysis method to decompose changes in aggregate energy intensity into structural change and efficiency change, and found that although both factors reduce energy intensity, the latter predominates. Shi (2005) also found that final demand's structural shifts, which correspond to structural shifts in energy consumption decline, did not account for as large reductions as efficiency improvements. Zhang (2003) investigated the change in energy consumption in China's industrial sector based on a dataset for the 29 industrial subsectors. Through decomposition analysis, he found that the efficiency change plays a key role. Fisher-Vanden et al. (2004) found that changes in industrial composition and gains in firm-level productivity both explain the continuous decline in China's energy intensity until year 2000, while the latter is the more important driver. Lewis et al. (2003) concluded that government industrial policy was a primary factor in the energy decline, supported by ongoing programs to increase energy efficiency.

But this discussion would not be complete if it did not study concerns regarding the integrity of the data used to generate the foregoing results. Some analysts have cast doubt on China's statistical gross domestic product (GDP) and energy-consumption data. Sinton (2001) points out that the energy data since the mid-1990s are likely to be misleading, induced by policies of the late 1990s that aimed at closing small coal mines. It is more likely that, rather than being closed, the mines disappeared from the statistics (World Bank, 2006), with the consequence of biasing calculations of aggregate energy intensity downward dramatically. As well, China's GDP data for each year has recently been modified, with an increase varying from 0.6% in year 1980 to 16% in year 2003, based on a national poll of China's economy carried out by the National Bureau of Statistics of China.

However, all these extensive China energy-related studies are based on aggregated national data. China has over one-fifth of the world's population. One province in China is almost as large as any one of the countries in Europe. China's energy resources are unevenly dispersed, and energy intensities across the regions have significant regional variations. Although using regional data will not solve the problems with national data, it, at least, will enable us to examine more closely what is occurring with the national energy-intensity trend. Exploiting interregional heterogeneity can serve as an additional source of insights, as well as a robustness check on the results generated with aggregated statistics. Polenske's (2006) Technology-Energy-Environment-Health (TEEH) team conducted a case study for Shanxi Province, the largest coal and coke production area in China, and they found that its energy intensity is twice that of all China. On the other hand, they also found that Liaoning Province, despite serving as the largest base for the steel-making industry, has an energy-intensity about the same level as in all of China.

There are only a few analyses using China's regional data. Auffhammer and Carson (2007), based on a provincial panel data set, exploited the considerable heterogeneity that exists across China's provinces and clearly reject the static environmental Kuznets-curve specification. Hofman and Labar (2007) found that beyond structural change, energy intensity is negatively correlated with energy prices, the efficiency within the energy sector, and the share of light industry in provincial GDP, and positively with the share of state enterprises in industrial output. Hu and Wang (2005), by using regional data, found that the central area of China has the worst energy efficiency and its total adjustment of the energy consumption amount is over half of China's total. They also found a U-shaped relationship between each area's total-factor energy efficiency and per capita income in that area, suggesting that energy efficiency does in fact improve with economic growth.

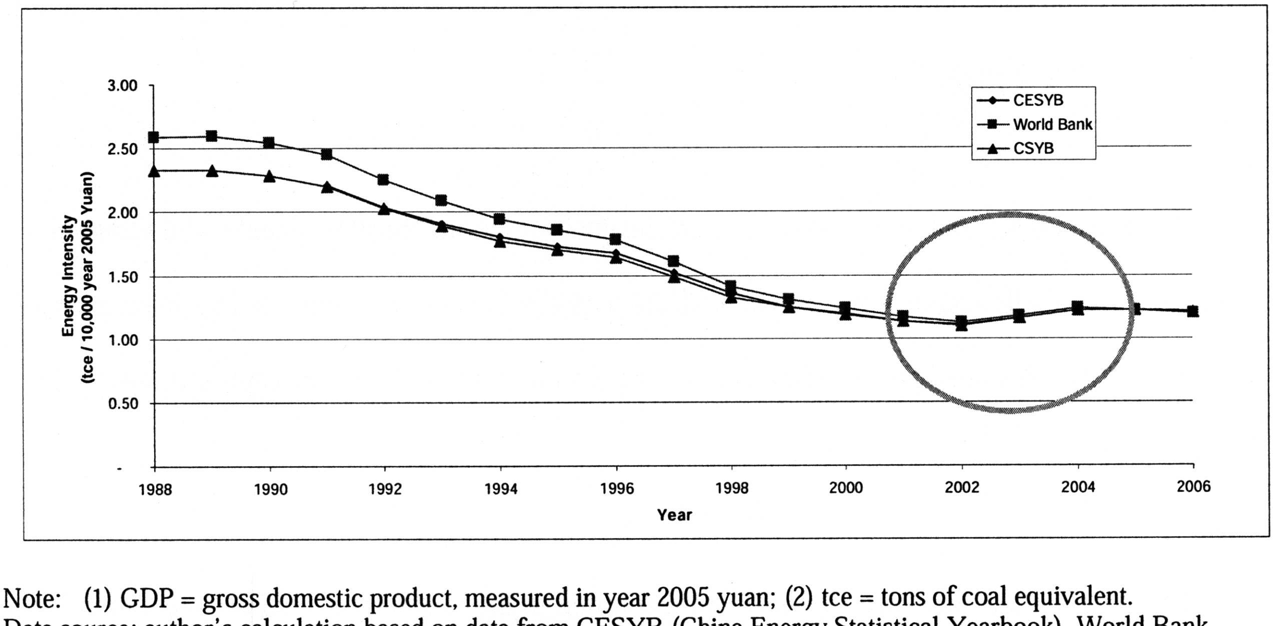
Chapter II

China's Aggregate and Provincial Energy Intensity

As I will show, China's aggregate energy intensity has been declining steadily since 1988 and then it started to increase after 2002, while China's provincial energy intensity has shown a much more volatile trend. I selectively identify the five most- and least-energy-intensive provinces. The provincial variance is pronounced.

2.1 Energy Intensity at the National Level

The recent rebound of China's energy intensity has raised many analysts' concerns. In order to deal with the issue of data quality, I calculate China's energy intensity by using data from three different sources. The results plotted in Figure 2 suggest that there are slight differences in terms



Note: (1) GDP = gross domestic product, measured in year 2005 yuan; (2) tce = tons of coal equivalent.
Data source: author's calculation based on data from CESYB (China Energy Statistical Yearbook), World Bank (www.econstats.com), CSYB (China Statistical Yearbook)

Figure 2: Energy Intensity of China with Multiple Data Sources, 1988-2006

of GDP and energy consumption data. However, the GDP deflators are measured differently in the *China Statistical Yearbook* and by the World Bank.

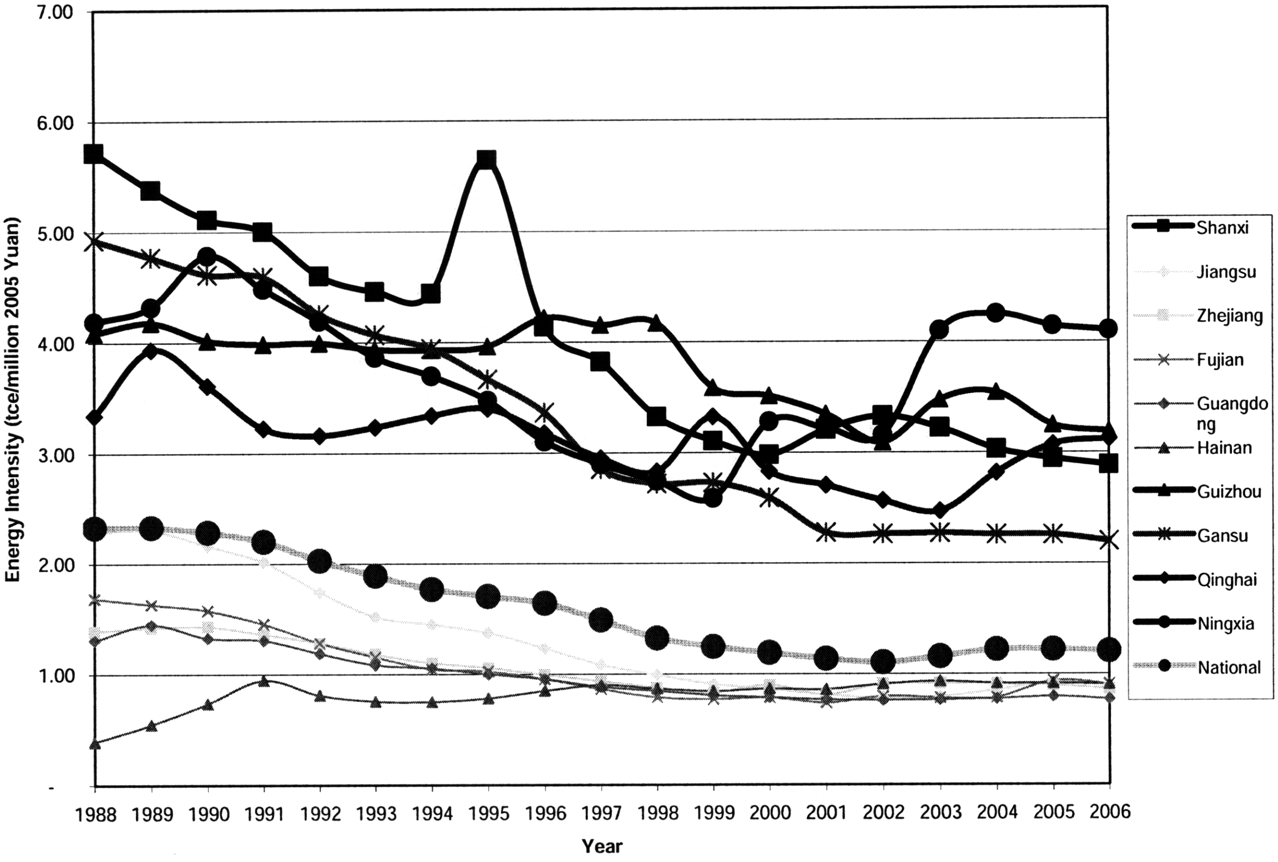
Despite the small discrepancies in the absolute values, all the results clearly show a trend of energy intensity rising slightly since 2002. There are different explanations for this rebound. Hofman and Labar (2007) claim that it is, in part, ascribed to a sectoral shift towards industry in the majority of provinces, but this is offset by continued efficiency gains within industry and other sectors. Jiang and O'Neill (2005) argue that the accelerating speed of urbanization and the residential energy consumption may play an important role.

2.2 Energy Intensity at the Provincial Level

The energy intensity of China's 29 provinces shows more variety when I plot the data into one chart. The data and chart are presented in the appendix. In Figure 3, I include the five most and least energy-intensive provinces in China, ranked by the arithmetic mean of energy intensity for the period 1988 through 2006.

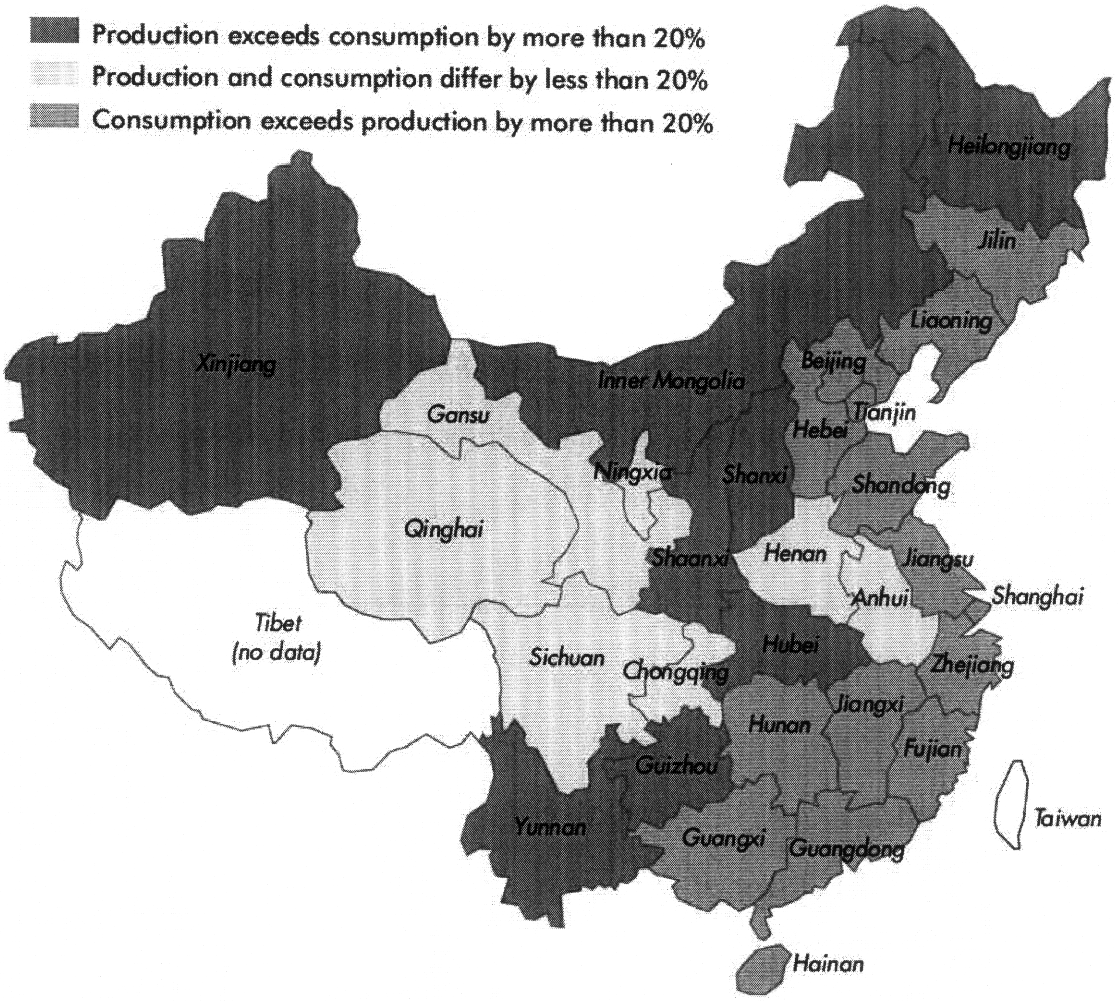
The five most energy-intensive provinces are Shanxi, Guizhou, Ningxia, Gansu, and Qinghai, and their total GDP accounts for only 5% of the nation's total GDP, while the five least energy-intensive provinces are Jiangsu, Zhejiang, Fujian, Guangdong and Hainan, representing 35% of GDP.

The least energy-intensive provinces are concentrated in the southern coastal area of China, starting from Jiangsu to Hainan and all of the five are border connected, forming an area that is famous for agriculture, service, and commerce industries. The most energy-intensive provinces are landlocked, and concentrated in the inland part of China. If I include the sixth most energy-intensive province – Sichuan (including Chongqing) –, these six provinces also appear to form a continuous region. If I match the location of the identified provinces to the energy production and consumption map (Figure 4), I find a rather interesting result. The least energy-intensive regions are regions where consumption exceeds production by more than 20%. However, the most energy-intensive regions are not those where production exceeds consumption by more than 20%, but rather production and consumption differ by less than 20%.



Note: GDP is measured in 2005 Chinese Renminbi (Yuan)
Source: author's calculations based on China Statistical Yearbooks, China Energy Statistical Yearbook, and China provincial statistical yearbook (1989-2007)

Figure 3: China's Five Most and Least Energy-Intensive Provinces, 1988-2006



Sources: *World Energy Outlook 2007*; National Bureau of Statistics (2007); IEA

Figure 4: China's Energy Production and Consumption by Province, 2005

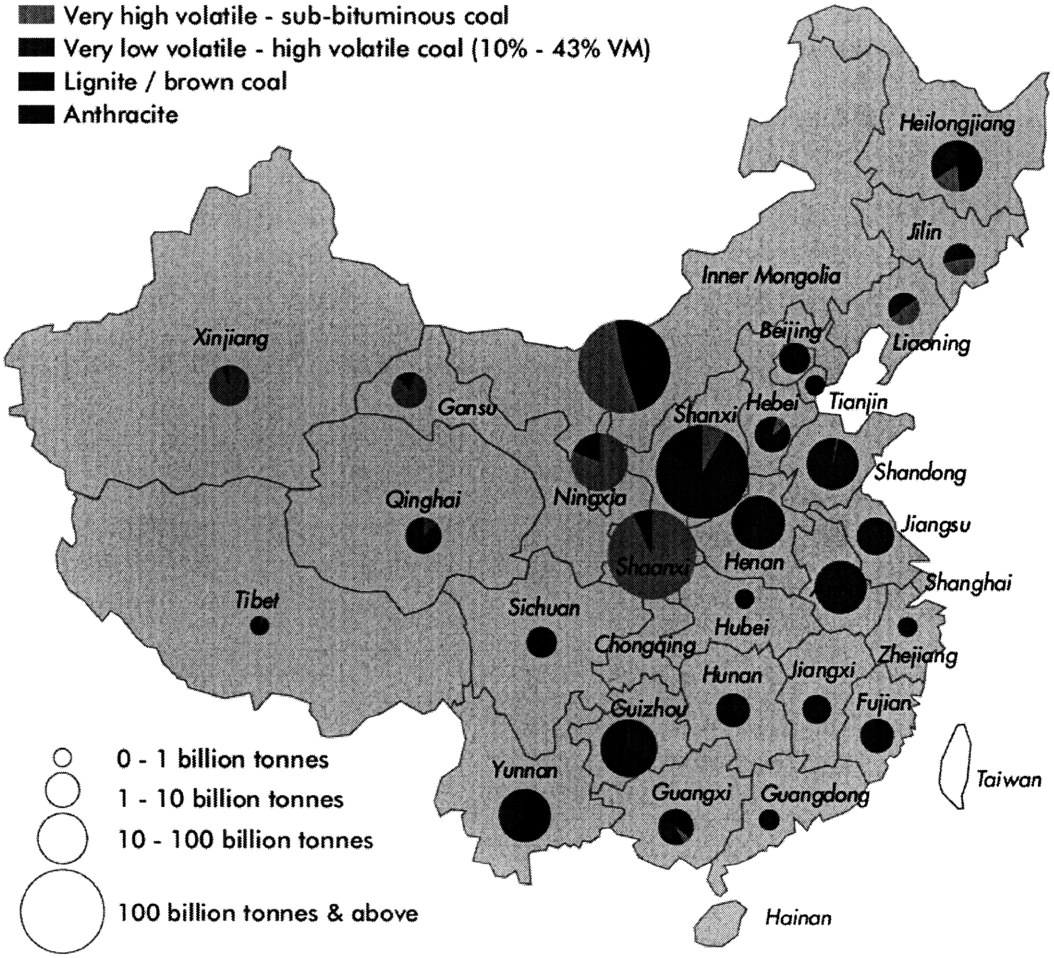
In the recent ten years, the most energy-intensive provinces show more volatility than the least energy-intensive provinces[2]. I suspect it is due to the fact that the extent of energy-producing dominated inland regions, affected by national policy, and the capacity of transportation infrastructure, can vary significantly from year to year, while the energy-consuming dominated pattern in coastal regions can be stabilized over the years.

---

[2] The volatility is calculated by the coefficient of variation (= standard deviation / mean)

2.3 Energy Imbalance and Regional Variance

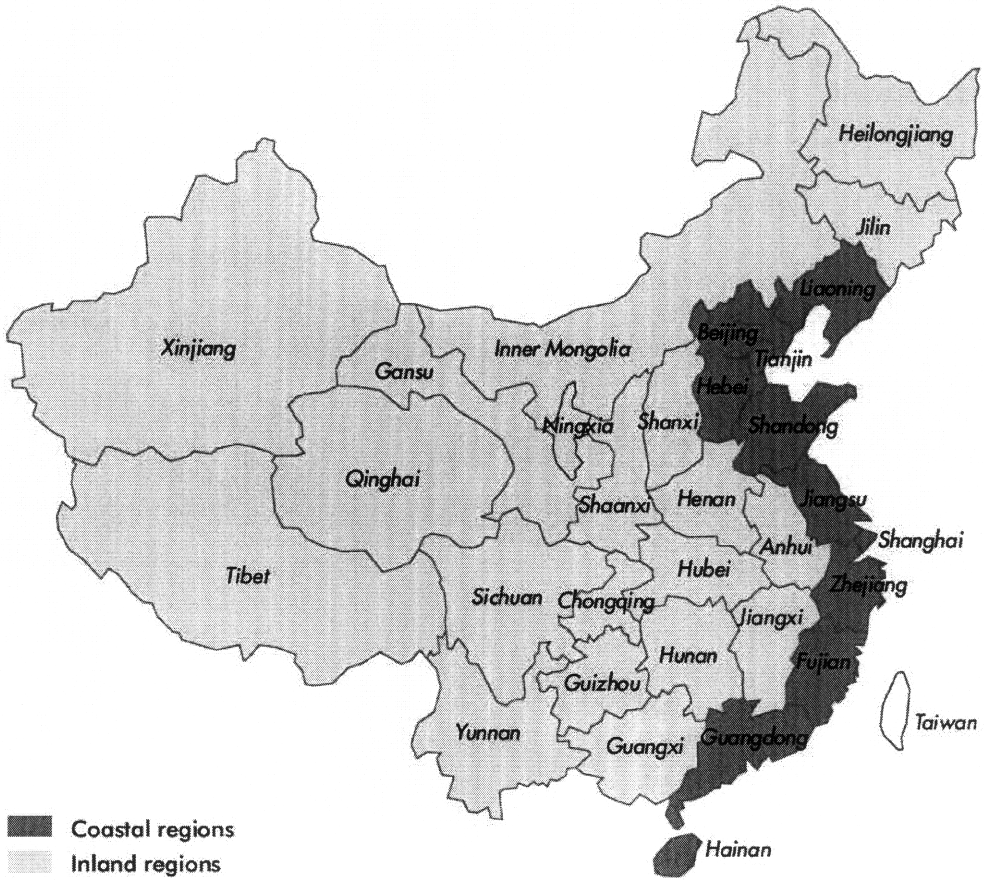
There exists a geographical imbalance in terms of China's energy production and consumption, as we can observe from Figure 5. This has raised the issues of inter-regional energy import, export, inventory, and energy transportation. While overall energy supply and demand are evening out in China, some areas continue to suffer periodic imbalances. Most of China's coal and coke is produced in a few inland provinces, such as Shanxi, Ningxia, and Inner Mongolia, while the largest centers of demand for coal and for the electricity that is generated by coal and coke, are in the coastal provinces.



Sources: *World Energy Outlook 2007*; Beijing HL consulting (2006); IEA

Figure 5: China's Coal Resources

Inland provinces are the main energy-producing regions and are traditionally more energy intensive than energy-consuming coastal provinces. Many analysts have claimed that there is a shift from an energy-intensive economic sector towards a non-energy-intensive economic sector. With the same concept, I suspect that such energy imbalance and reverse pattern may contribute to the aggregate energy-intensity change. In other words, one possibility for why China's energy intensity has gone down over the years is that the energy-intensive inland regions have been exporting more and more energy to the less energy-intensive coastal regions. I first define coastal and inland regions, as shown in Figure 6. I then calculate the GDP, energy consumption and energy intensity for both regions, as shown in Table 1.



Source: *World Energy Outlook 2007*

Figure 6: China's Coastal Regions versus Inland Regions

In Table 1, I find that the share of coastal regions' aggregate GDP has increased steadily over the years. The share of the coastal region's energy consumption is increasing, with a dip in 2001 and then decrease from 2002 to 2004. This happens to be the same period where China's national energy intensity augments. At the same time, however, the relative energy intensity has experienced first a decline, then a fluctuating period, starting from the year 2000.

Table 1: GDP, Energy Consumption, and Energy Intensity for Inland and Coastal Regions

	GDP		Energy Consumption		Energy Intensity	
Year	Coastal % of total	Inland % of total	Coastal % of total	Inland % of total	Coastal % of Inland	Inland % of Inland
1988	52.7%	47.3%	44.3%	55.7%	71.5%	100%
1989	52.6%	47.4%	44.7%	55.3%	72.7%	100%
1990	52.7%	47.3%	44.5%	55.5%	72.0%	100%
1991	54.0%	46.0%	44.8%	55.2%	69.4%	100%
1992	55.2%	44.8%	45.0%	55.0%	66.3%	100%
1993	56.4%	43.6%	45.8%	54.2%	65.4%	100%
1994	57.3%	42.7%	46.5%	53.5%	64.8%	100%
1995	57.8%	42.2%	45.3%	54.7%	60.5%	100%
1996	57.8%	42.2%	45.7%	54.3%	61.6%	100%
1997	58.0%	42.0%	45.5%	54.5%	60.5%	100%
1998	58.3%	41.7%	45.8%	54.2%	60.3%	100%
1999	58.8%	41.2%	46.6%	53.4%	61.1%	100%
2000	59.2%	40.8%	48.7%	51.3%	65.5%	100%
2001	59.4%	40.6%	47.6%	52.4%	62.1%	100%
2002	59.7%	40.3%	48.8%	51.2%	64.1%	100%
2003	60.2%	39.8%	48.3%	51.7%	61.7%	100%
2004	60.5%	39.5%	48.1%	51.9%	60.5%	100%
2005	60.6%	39.4%	49.3%	50.7%	63.2%	100%
2006	60.8%	39.2%	49.3%	50.7%	62.6%	100%

Note: energy intensity for costal regions is the relative percentage to that of the same year inland regions' energy intensity.
Source: author; *China Statistical Yearbooks*

I further assume that the recent rebound of energy intensity is partly due to the fact that the channel of the inland-coastal energy transportation corridor has reached its capacity due to new infrastructure investment. According to the World Energy Outlook (WEO 2007), the projections of coal transport imply a continuing need to expand China's inland coal transport infrastructure. Shipments from inland to coastal provinces will need to increase from 507 Mtce in 2005 to 1060 Mtce in 2030 for steam coal and from 117 to 182 Mtce for coking coal.

Chapter III

China's Coal Supply, Demand, and Transportation

As one of the largest developing countries undergoing massive industrialization, China has experienced a typical increase in energy consumption to keep up with its rapid economic growth. China is already the world's second largest user of energy and the largest emitter of greenhouse gases. Official statistics show that China's energy consumption has risen from 930 million to 2,463 million metric tonnes of coal equivalent (tce) between 1988 and 2006, an average increase of 5.6% per year. During the same period, China's GDP has grown from 4,000 billion to 20,400 billion Yuan in year 2005 currency, an average annual increase of 9.5%.

3.1 Coal Production and Consumption

Coal is China's most important and abundant fuel, accounting for about two-thirds of its primary energy supply. China's coal output rose from 1.3 billion tonnes in 2000 to 2.2 billion tonnes in 2005, making it the largest global coal producer. In 2005, electricity generation accounted for 69% of all coal consumption in China, while cokemaking accounted for 21%. Through the study of historical data, I find that China's dependence on coal steadily decreased since the early 1990s, while the energy provided by natural gas and hydro-power increased over the same time period (Table 2).

Shanxi province is the largest coal- and coke-producing region in China. According to the official statistical yearbook, in 2005, Shanxi Province produced 80 million metric tonnes of coke,

Table 2: Energy Consumption by Type in China, 1978-2006

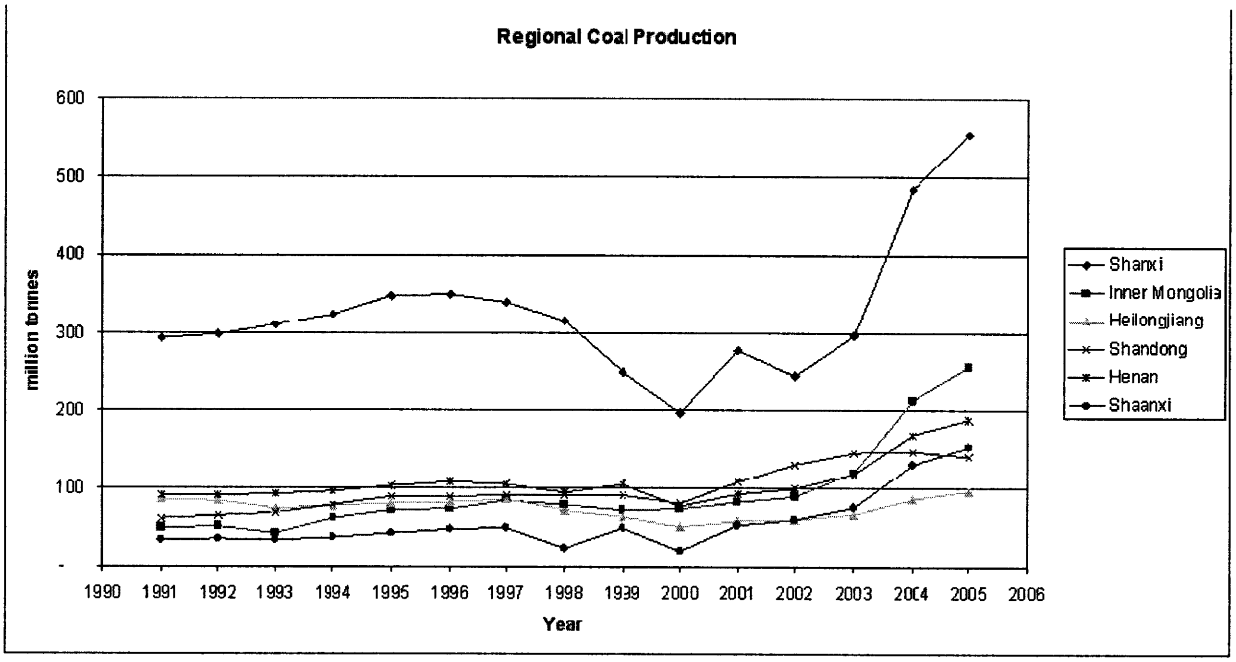
Year	Energy Consumption (million tonnes SCE)	Percentage of Total Energy Consumption				
		Coal	Crude Oil	Natural Gas	Hydro	Total
1978	571	70.7	22.7	3.2	3.4	100.0
1980	603	72.2	20.7	3.1	4.0	100.0
1985	767	75.8	17.1	2.2	4.9	100.0
1990	987	76.2	16.6	2.1	5.1	100.0
1995	1,312	74.6	17.5	1.8	6.1	100.0
2000	1,386	67.8	23.2	2.4	6.7	100.0
2005	2,247	69.1	21.0	2.8	7.1	100.0
2006	2,463	69.4	20.4	3.0	7.2	100.0

Source: author, *China Statistical Yearbook 2007*

accounting for about 31% of China's total coke output versus 45% in 1998. The same year, Shanxi Province exported 5.96 million metric tonnes of coke, accounting for 47% of China's total 12.76 million tonnes of coke export.

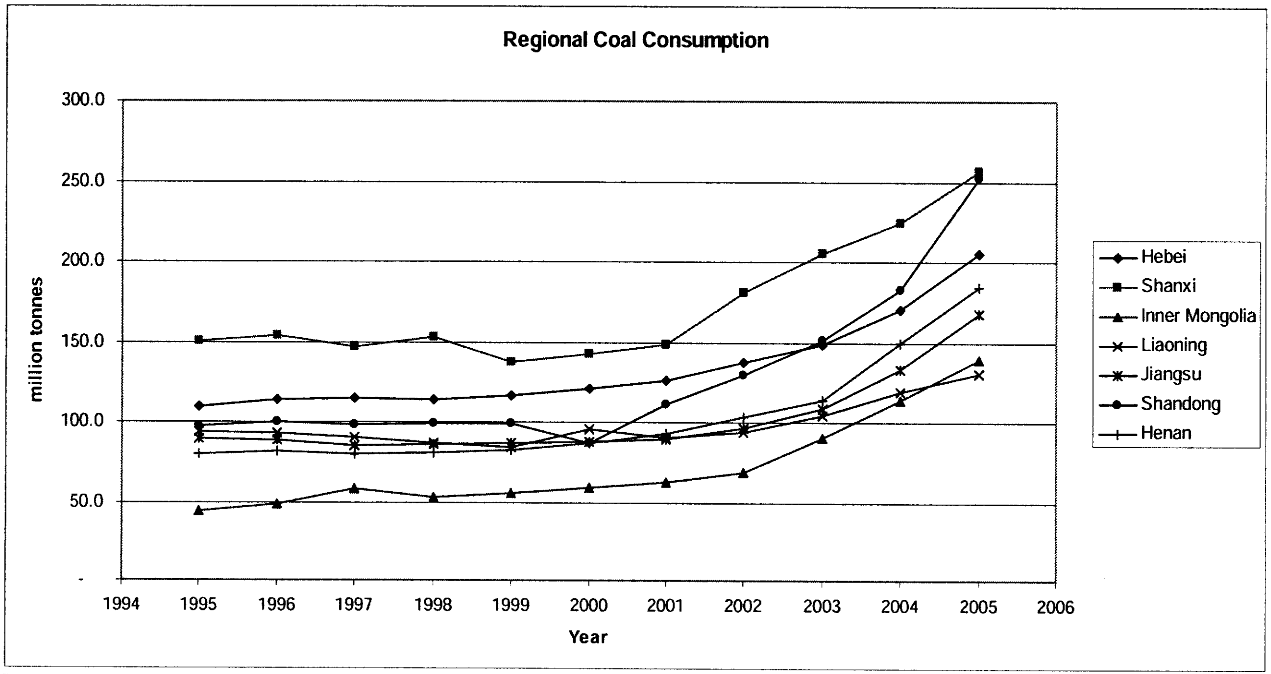
I have not yet been able to obtain the official published data of aggregate energy production for each province. However, since most of China's energy is from coal, I collected the coal production and consumption data for all the provinces from 1991 to 2005. The top coal producing provinces and consuming provinces are shown in Figures 7 and 8. I find that the highest coal-producing provinces also tend to be the highest coal-consuming provinces. In this sense, the assumption that most coal is transported from inland regions to coastal regions is not correct. However, China's coal-fired power plants increasingly have been built around coal mines or the railway network to handle the increased amount of coal that needs to be transported

(WEO, 2007). As a result, the coal transportation issue is not a simple cross-regional raw coal-transportation issue, but rather a transmission of its secondary product--electricity.



Source: *China Energy Statistical Yearbook* 1991-1996, 1997-1999, 2005, 2006

Figure 7: Provincial Coal Production 1991-2005

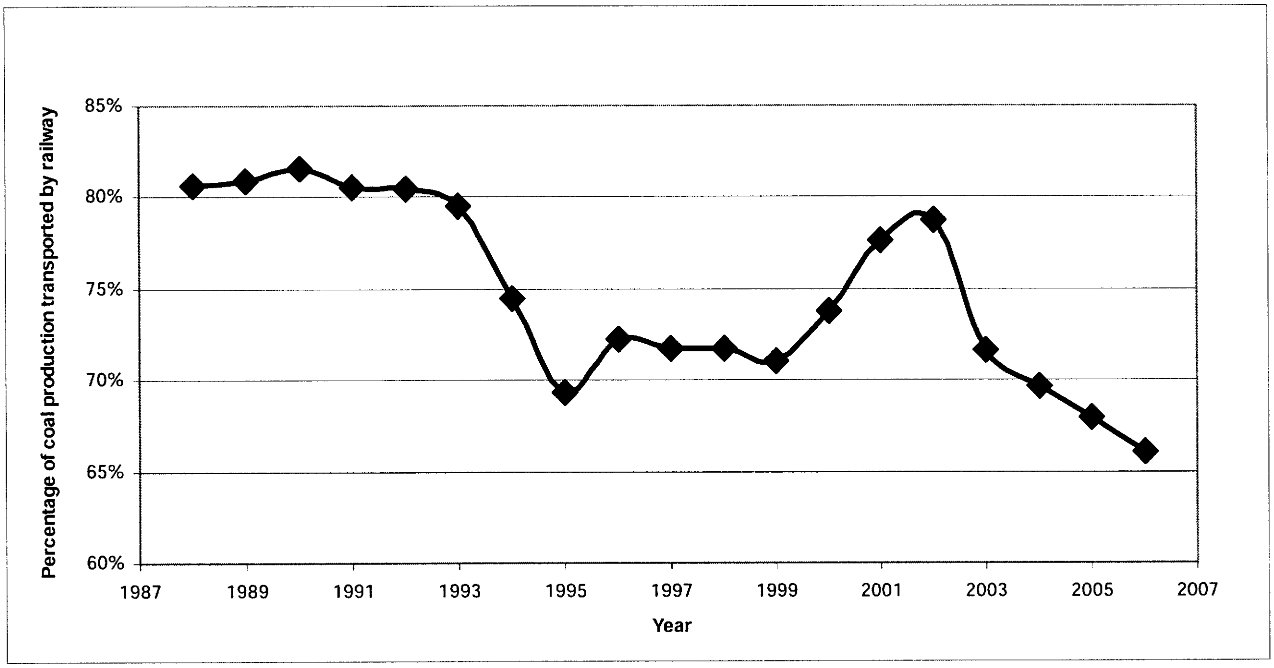


Source: *China Energy Statistical Yearbook* 1991-1996, 1997-1999, 2005, 2006

Figure 8: Provincial Coal Consumption 1991-2005

3.2 Coal Transportation by Railway Systems

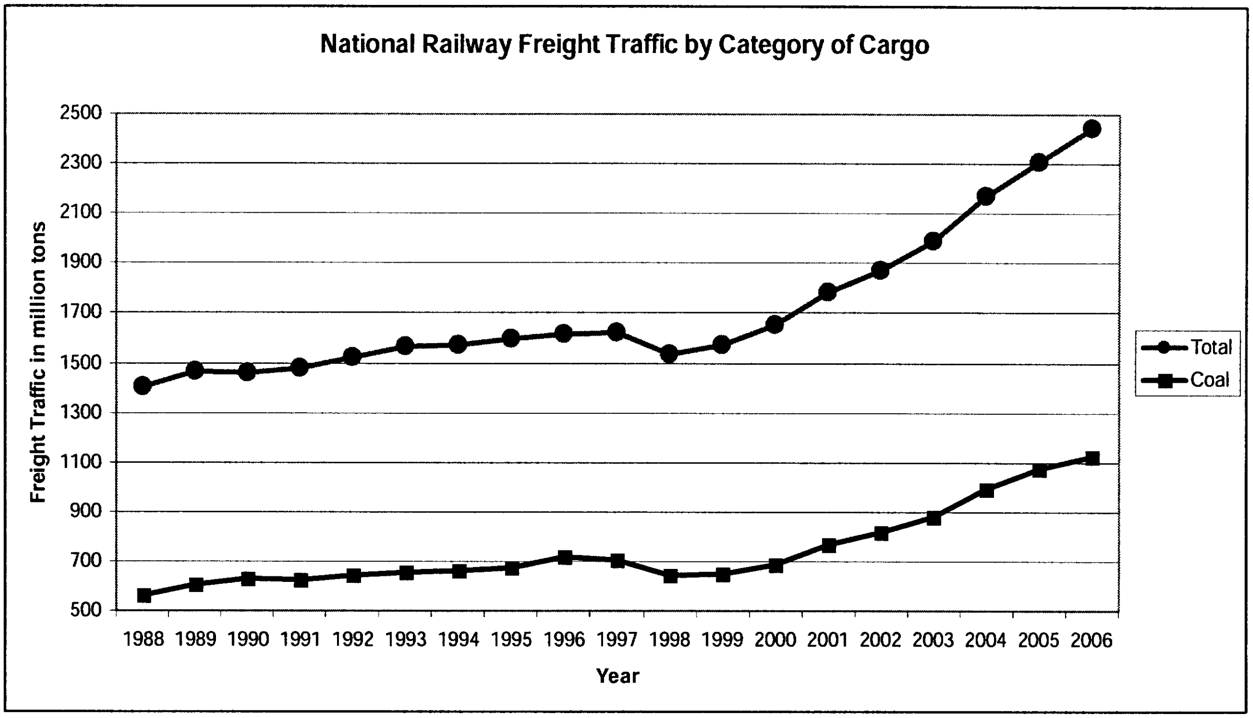
At present, China's coal transportation from producing regions to consumers, such as power plants, relies on rail, road, inland waterways and coastal vessels, among which railways are the most important mode. According to the statistical yearbook, in 2006, around 66% of China's coal production was transported by means of the national railway system, as shown in Figure 9. It is also observed that although railways always remain the most important mode, the percentage has been decreasing rapidly since 2002.



Source: author's calculation based on data from *China Statistical Yearbooks*

Figure 9: Percentage of Coal Production Transported by National Railway

Coal transportation also accounted for about 46% of the total national railway freight transportation in 2006, as shown in Figure 10. This percentage has been relatively constant since 1990 (43%). This raises an interesting question: as the national railway system expands, although



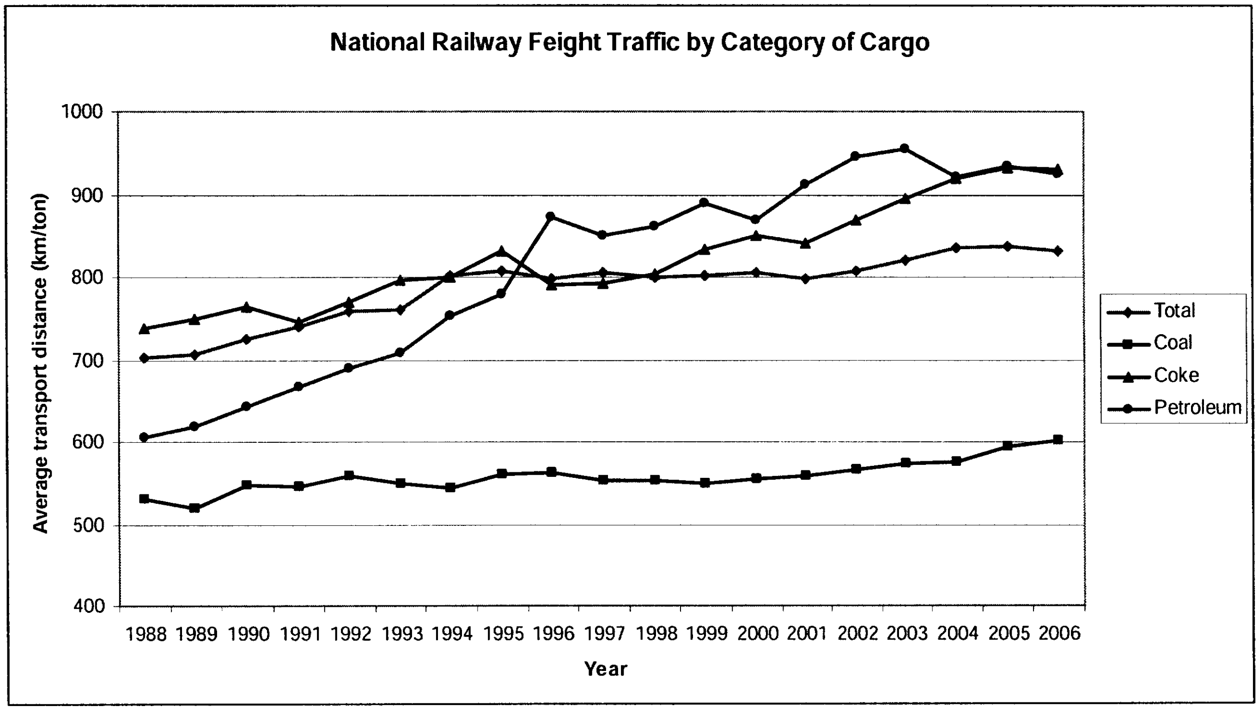
Source: *China Statistical Yearbooks*

Figure 10: National Railway Freight Traffic Coal versus Total Cargo

the volume of coal transportation increases, it cannot keep up with the increasing demand for coal that needs to be shipped from producing areas to consuming areas.

There are three possible explanations for the decreasing role of the railway mode for coal transportation. First, there may be an increasing use of other transportation modes, such as trucks and waterways. However, this should be an involuntary result as the railway rate is around 0.12 yuan per tonne-km, while the truck rate ranges from 0.5 to 0.8 yuan per tonne-km (WEO, 2007).

Second, national railway infrastructure investment has not been able to keep up with the needs of coal transportation. As *World Energy Outlook 2007* points out, much coal is transported on older trains over routes shared with passenger and the other freight trains, and there are only two modern rail links dedicated to coal: the 600-km Daqin line and the 588-km line from Shuozhou to Huanghua. As Figure 11 shows, the average transport distance of coal, although steadily increasing, remains around 600 km/tonne in recent years. This number does not differ much from 550 km/tonne in 1990, almost 15 years ago, and it is far below that of coke and petroleum, 931 and 926 km/tonne, respectively. The lack of new investment to expand rail capacity adds constraints as coal is increasingly transported over longer distances from the deep inland production regions. However, coal-fired power plants increasingly have been built around coal mines for the railway network to handle the increased amount of coal that needs to be transported (WEO, 2007). And once such coal-powered plant systems are established, the average transport distance of coal should not vary much.

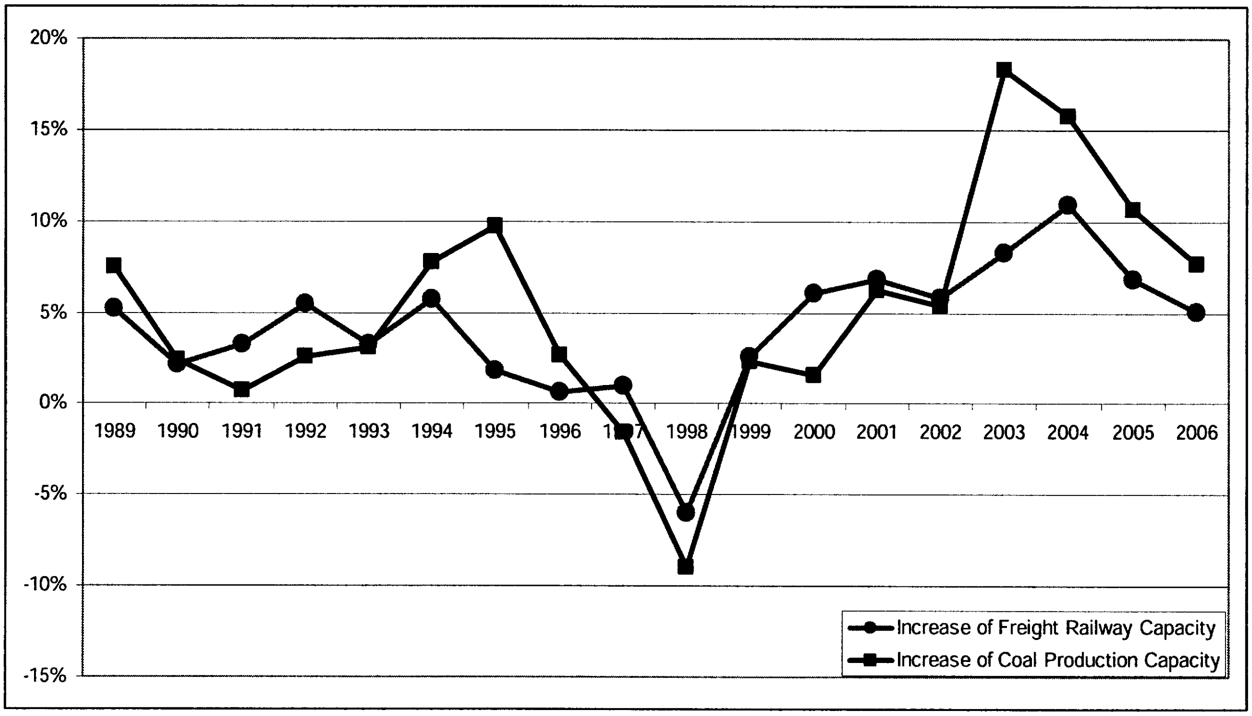


Source: author ,*China Statistical Yearbooks*

Figure 11: National Railway Freight Traffic: Average Transport Distance

3.3 The Constraints of Railway Capacity and Energy Intensity

In order to test if the most important transportation mode–the national railway system–has imposed a constraint on coal transportation and further affected energy-intensity trends, I compare the increase of the capacity of both railway freight transportation and coal production. I assume that the railway system has been fully used each year, and I used its total freight tonne-kilometers to measure the overall capacity. I used the coal production data as a measurement of its production capacity. The year-by-year percentage increase is shown in Figure 12.



Source: *China Statistical Yearbooks*

Figure 12: Increase of Freight Rail Capacity and Coal Production Capacity

I find that the two periods that the increased capacity of freight railway transport systems has not kept up with that of coal production are 1994-1996 and 2002 onward. When I matched the two periods with China's energy-intensity chart (Figure 2), I find that 1994-1996 China's energy intensity has the smallest slope of decrease; while since 2002, the energy intensity has been increasing.

What if China's energy loop is as follows: (1) coastal regions are the main energy-consuming areas, while the inland regions are the main energy-producing areas; (2) inland regions produce and transport coal through the railway system either directly or electricity is processed locally to meet coastal regions' energy demand; (3) the coastal regions are more energy efficient than inland regions. Such a system established from (1) through (3) can actually decrease aggregate energy intensity. The limited capacity of the railway transportation system has affected (2) above. I believe that this has posed a negative effect and hindered the further decline of energy intensity. To analyze this issue in more depth, provincial data of energy import, export and inventory should be collected. However, at this stage, such data are not readily available.

Chapter IV
Decomposing Energy Intensity into an Intra- and Inter-Provincial Index

Having discussed the difference of energy production and consumption patterns between inland regions and coastal regions, I now turn to the second feature of this study. It is a method used by Sue Wing (2008): an approach to decompose aggregate energy-intensity change into provincial energy-intensity change and inter-provincial structural change.

Energy intensity (ey_t) can be written as a function of various components in order to capture different effects. With the decomposition results, indexes of different components can be created. Metcalf (2008) constructed an intensity index at the state level from a decomposition of the U.S. intensity index in which he disentangles changes in energy use within a sector and changes in sectoral activity over time.

4.1 Decomposition Analysis

In this study, I write the decomposition as a function of intra-provincial and inter-provincial components. The intra-provincial component captures the energy efficiency effects that are influenced by prices, technology, weather, etc, which I refer to as the "within index." The inter-provincial component captures the structural change effects, such as the increase of the share of a less-energy-intensive province's output in national GDP, which I refer to as the "between index."

$$(1) \qquad ey_t = \frac{E_t}{Y_t} = \frac{\sum_i e_{it}}{Y_t} = \sum_i \frac{e_{it}}{y_{it}} \times \frac{y_{it}}{Y_t}$$

Where ey_t is the aggregate energy intensity in year t, E_t is aggregate energy consumption in year t, Y_t is the national GDP in year t, e_{it} is the energy consumption in province i in year t, y_{it} is the province i's output in year t.

(2) $$\frac{\partial \log ey_t}{\partial t} = \sum_i w_{it} \frac{\partial \log \frac{e_{it}}{y_{it}}}{\partial t} + \sum_i w_{it} \frac{\partial \log \frac{y_{it}}{Y_t}}{\partial t}$$

where each weight (w_{it}) is the ratio of province i's share of energy consumption to the national energy consumption in year t. The left-hand side of Equation (2) is interpreted as the rate of change in aggregate energy intensity in logarithmic form. It is decomposed into two effects on the right-hand side: the change attributed to the energy-intensity change within provinces–the efficiency-change effect, and the change attributed to the change of the share of provincial output in total GDP–the structural effect.

The right-hand side of Equation (2) can be further developed into two separate indexes, I derive separate equations to make an approximate calculation of the value of $\frac{\partial \log \frac{e_{it}}{y_{it}}}{\partial t}$, $\frac{\partial \log \frac{y_{it}}{Y_t}}{\partial t}$ and w_{it}. w_{it} can be calculated by using,

(3) $$w_{it} = \frac{e_{it}}{E_t} \approx \frac{1}{2} \times \left(\frac{e_{i,t}}{E_t} + \frac{e_{i,t-1}}{E_{t-1}}\right)$$

And further with the Central Difference Derivative Approximation, $\frac{\partial \log \frac{e_{it}}{y_{it}}}{\partial t}, \frac{\partial \log \frac{y_{it}}{Y_t}}{\partial t}$ can be approximately calculated by Equations (4) and (5),

(4) $$\frac{\partial \log \frac{e_{it}}{y_{it}}}{\partial t} = \frac{\Delta \frac{e_{it}}{y_{it}}}{\frac{e_{it}}{y_{it}}} = \frac{\frac{e_{i,t}}{y_{i,t}} - \frac{e_{i,t-1}}{y_{i,t-1}}}{\frac{1}{2} \times (\frac{e_{i,t-1}}{y_{i,t-1}} + \frac{e_{i,t}}{y_{i,t}})}$$

(5) $$\frac{\partial \log \frac{y_{it}}{Y_t}}{\partial t} = \frac{\Delta \frac{y_{it}}{Y_t}}{\frac{y_{it}}{Y_t}} = \frac{\frac{y_{i,t}}{Y_t} - \frac{y_{i,t-1}}{Y_{t-1}}}{\frac{1}{2} \times (\frac{y_{i,t}}{Y_t} + \frac{y_{i,t-1}}{Y_{t-1}})}$$

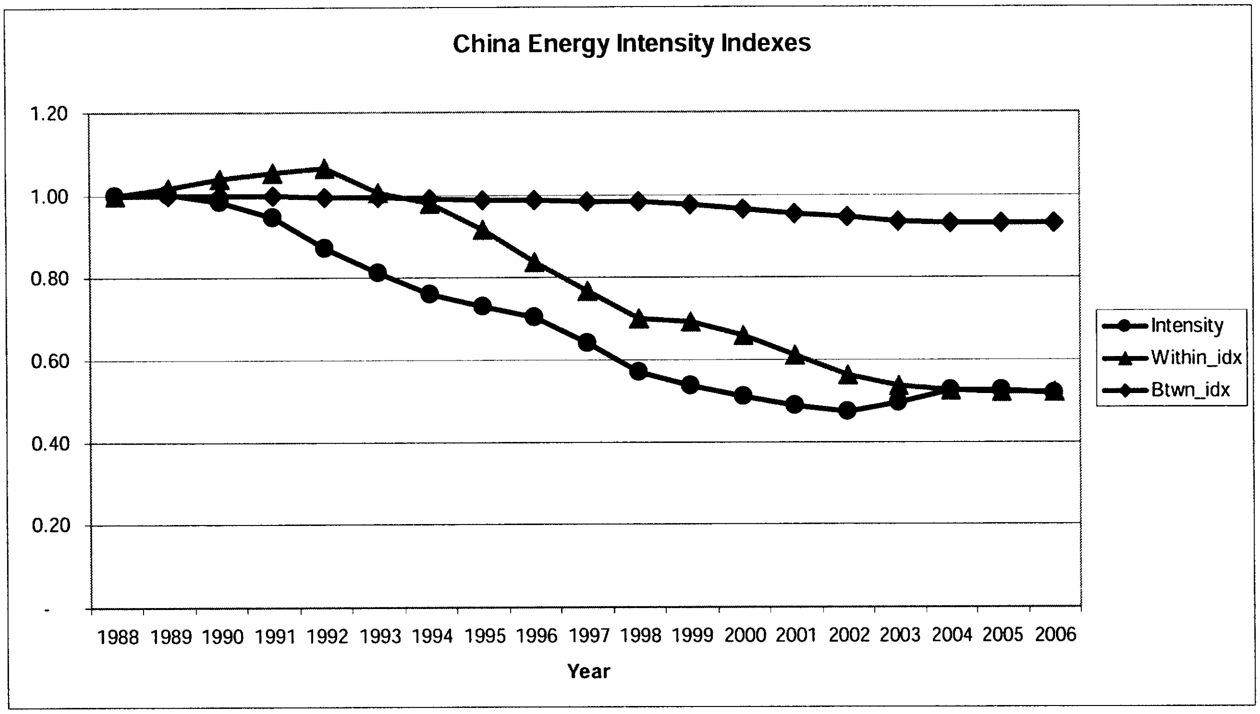
Until this step, all the unknown components in Equation (2) have been expressed as functions of existing data variables that I have collected, namely, ey_t, E_t, Y_t, e_{it}, y_{it} and t. I numerically evaluate Equations (1) through (5) using provincial GDP and energy consumption from 1988 through 2006 in Matlab and express the result as a series of index numbers.[3] For the programming code, please refer to the appendix.

4.2 The Within Index and Between Index

The index result is presented in Figure 13. I find that the within index declines at a much faster rate than the between index. The improvement of energy efficiency within each province has contributed more to the change of energy intensity, than the growth of its share of provincial

---

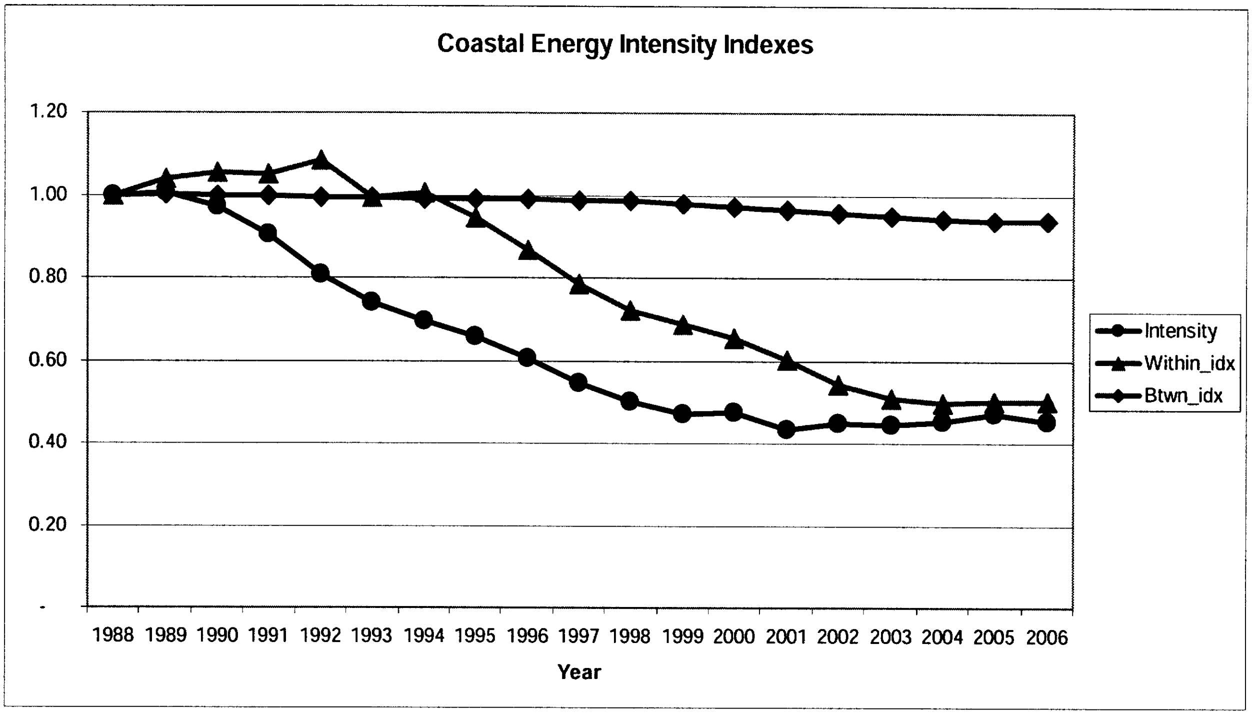
[3] Programming code is attached in the appendix.



Note: Indexes are normalized to one in 1988.
Source: author's calculation; data from China Statistical Yearbooks, and China Energy Statistical Yearbooks

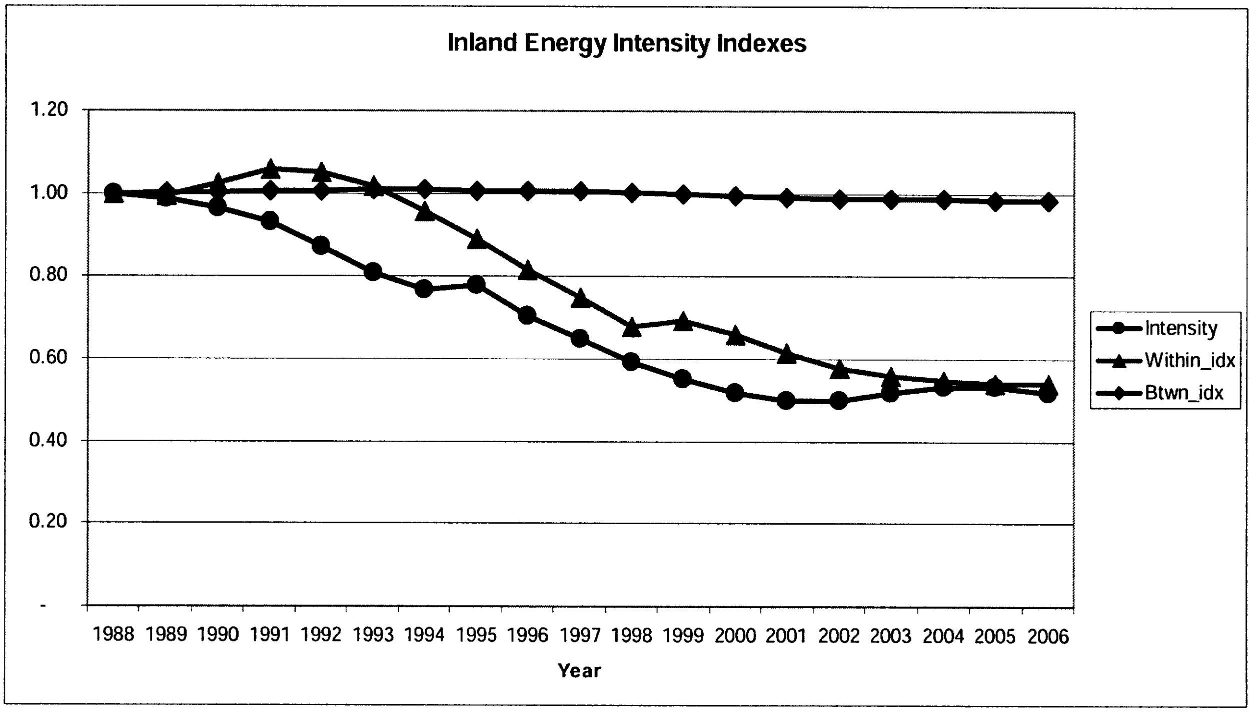
Figure 13: China Energy Intensity Indexes

output in total GDP. Let me illustrate this finding by using the data in Table 1. For simplicity, there are only two regions in China, coastal regions and inland regions. In each year, coastal regions have a higher w_{it}, defined in Equation (3) and they also have lower regional energy intensity and a continuously increasing share of regional output. On the other hand, the inland regions have a lower w_{it}, but they have a higher regional energy intensity and a continuously decreasing share of regional output. What Figure 13 indicates is that the overall declining trend of aggregate energy intensity is more attributed to the decreasing trend of regional energy intensity (in this case in both regions), than to the fact that coastal regions produce an increasing of total output using less energy. I also create the within and between index for inland regions and coastal regions separately. The results are shown in Figures 14 and 15.



Source: author's calculation; data from *China Statistical Yearbooks,* and *China Energy Statistical Yearbooks*

Figure 14: Coastal Region Energy-Intensity Indexes



Source: author's calculation; data from *China Statistical Yearbooks,* and *China Energy Statistical Yearbooks*

Figure 15: Inland Region Energy-Intensity Index

As we can see, both coastal and inland regions follow the same declining pattern. The coastal regions affect the overall between index more, while the inland regions affect the overall within index more.

When I compare the China energy-intensity index to the U.S. energy-intensity index (Metcalf 2008, see appendix), I find that the indexes follow the same trend. For both countries, the efficiency index outperforms the activity index. However, one thing to note is that for both efficiency and structure indexes, I am not comparing apples to apples due to the different ways of defining the "sectors" used for the decomposition analysis. In the U.S. case, Metcalf has isolated two key determinants of changes in energy intensity through separating the data by economic sectors (residential, commercial, industrial, and transportation). For him, the efficiency change can be the price-induced energy-efficiency change within an industrial sector; while the structural change can be that the economic sector has shifted X% from energy-intensive transportation sector into the less energy-intensive service sector. In the China case, I have separated the data by geographical units (provinces). For me, the efficiency change can be the GDP shift X% from the industrial sector to the agriculture sector within one province, thus including both efficiency and structural effects defined in Metcalf's analysis. The structural change in this study is a very different concept that is related to the share of provincial output change in total GDP. For example, Guangdong province, being very energy-efficient, thus increasing its total share of provincial output in GDP can contribute to the overall energy-intensity reduction.

Chapter V

Panel-Regression Analysis

There is considerable variation in all three indexes across time and provinces. In order to find the factors that drive changes in energy intensity, I ran regressions using provincial data between 1988 and 2006 of energy intensity (in log form) on various variables. Following the variable identified by Metcalf (2008), I collected the data for the following factors of each province: *fuel*, the fuel price; *pci*, per capita income; *pci2*, the square term of per capita income; *hdd*, heating degree days; *cdd*, cooling degree days; *time*, time trend; *time2*, the square term of time trend; k*l*, capital labor ratio; *kl2*, the square term of capital labor ratio; *ik*, investment-capital ratio; ik2, the square term of investment-capital ratio.

Table 3: List of Independent Variables

Variables	Description
fuel	the fuel price
pci	per capita income
pci2	the square term of per capita income
hdd	heating degree days
cdd	cooling degree days
time	time trend
time2	the square term of time trend
kl	capital labor ratio
kl2	the square term of capital labor ratio
ik	investment-capital ratio
ik2	the square term of investment capital ratio

Source: author

5.1 Dataset Variables: Expectations and Summary Statistics

The panel dataset I use contains a series of observations for each of the 29 provinces. Each province includes 19 observations (year 1988 through year 2006), with 2006 being the latest year for which data are currently (January 2009) available in the statistical yearbooks. Thus, the total number of observations is 29*19 = 551. I use a loglinear regression, aiming to find the best fit between the data and a loglinear model.

I include the fuel price derived from the retail index of fuel from the statistical yearbooks. As the fuel price goes up, the energy intensity (ei) should go down as residents and industry will use fuel more efficiently, e.g., drive less, improve technology, etc. On the other hand, the retail fuel prices do not reflect the price paid by some of the state-owned industrial enterprises, especially given the implied fact that the latter is largely controlled by the government. However, this price index is the only data that are readily available.

I then include the log of per capita income (pci) and its quadratic term to account for the non-linearity effect in the response of energy intensity to personal income. As resident's income goes up, they may switch from biomass energy use to fuel energy use, which leads to a decrease of ei at the early stage. If the income goes up further, the expansion of electrical appliance use may affect ei in the other direction. The overall effect between intensity and income may appear as a "U" shape. With the coefficients from the regression result, the value at which the combination of the linear and quadratic terms is extreme ("turning point") can be calculated. Depending on the location of the turning point relative to the range of sample per capita income, I can determine whether the total effect of pci is always positive or always negative or whether it changes sign at a meaningful value.

Weather can be a very important factor affecting energy use. Heating degree days (hdd) and cooling degree days (cdd) are quantitative indexes designed to reflect the demand for energy needed to heat or cool a home or business. However, I modified the calculation of hdd and cdd as follows. I multiplied a local population weight factor to reflect the weather impact on energy use. For example, one of the largest provinces--Qinghai province–has only 5 million residents (I reserve my opinion of using the term "only" here) while Beijing city alone has almost 16 million registered residents. Even if the weather pattern were same for the two provinces, the population factor should definitely be included to reflect the amount of energy use.

I include the time trend and its squared term, because as China is transitioning from an agriculture-based economy to industrial economy, and as innovation drives technology, energy efficiency should go up dramatically. In other words, the time-trend effect can be a substitute of the effect of technology improvement on energy use.

I further include the capital-labor ratio (k/l) and its quadratic term. I deriver the capital stock from the fixed-asset investment dataset by using the perpetual inventory method (PIM). The effect of k/l ratio can go either way, for example, with a positive sign indicating increasing energy intensity with a larger share of capital-intensive heavy industries (e.g., iron and steel, petrochemicals, etc), or a negative sign indicating a transition to high-tech industries like electronics and computers.

Finally, I include the fixed-asset investment-capital ratio (i/k). The total investment in fixed assets in the whole country refers to the volume of activities in construction and purchases of fixed assets of the whole country and related fees, expressed in monetary terms during the reference period. It is a comprehensive indicator that shows the size, structure, and growth of the investment in fixed assets, providing a basis for observing the progress of construction projects and evaluating the results of investment.[4] I discount the investment data by the price indexes of investment in fixed assets by region (in current price). The i/k ratio is expected to respond positively to the energy intensity change. With large investments, the country is adding larger quantities of equipment, machinery, construction, and infrastructure facilities, which embody more energy-consuming ability.

For detailed data sources, assumptions, equations, and data procession, refer to the appendix. Descriptive statistics for these variables are given below.

Table 4: Summary Statistics of the Aggregate Dataset[5]

Variable[6]	Description	*	Mean	Std. Dev.	Min	Max
Ln(ei)	Energy intensity	overall	0.63	0.51	-0.94	1.74
		between		0.43	-0.22	1.36
		within		0.27	-0.09	1.26
Ln(fuel)	Retail fuel price index	overall	-0.84	0.58	-2.16	0.17
		between		0.09	-1.01	-0.69
		within		0.57	-2.01	0.34
Ln(pci)	Per capita income	overall	1.87	0.74	0.41	4.03
		between		0.54	1.01	3.23
		within		0.51	0.83	3.00

---

[4] Definition from *China Statistical Yearbook, 2007*

[5] [Stata] xtsum ei fuel pci pci2 hdd cdd time time2 cl cl2 ik, i(id)

[6] Observations: N=551, n=29, T=19

Ln(hdd)	Heating degree days	overall	3.79	1.29	-2.30	5.19
		between		1.31	-1.64	5.07
		within		0.11	3.12	4.22
Ln(cdd)	Cooling degree days	overall	2.92	1.43	-2.30	5.00
		between		1.44	-2.23	4.73
		within		0.19	1.06	3.60
Ln(kl)	Capital-labor ratio	overall	1.11	0.76	-0.69	3.34
		between		0.59	0.13	2.60
		within		0.49	-0.46	2.41
Ln(ik)	Investment-capital ratio	overall	-1.91	0.29	-3.91	-0.64
		between		0.13	-2.13	-1.61
		within		0.26	-3.77	-0.93

*This table shows the min and max, std and mean in three ways that are of interest: (1) the overall sample; (2) the between sample, i.e,. x(bar)i; and (3) the within sample, i.e., xit-x(bar)i-x(global bar)
Source: author

The overall standard deviation (stdev) measures variation in the entire cross-province cross-time data set. The between standard deviation measures variation across the 29 provinces, while the within standard deviation measures variation from the province-specific data across time. For example, for the cooling degree days, the within stdev is 0.19, well below the between stdev, which is 1.44. It is because the weather does not vary much within a province from year to year but it does vary much from province to province in the same year. On the other hand, for the retail fuel price, the between stdev is well below the within stdev. It makes sense as fuel prices change much from year to year, but do not change much from province to province within the same year.

With 29 provinces and 19 years' data, I design the model as the "large N, small T" panel in which I use a large number of individuals to construct the large-sample approximations. The small T may put limits on what can be estimated. For all the variables, the within standard deviation is non-zero, which means none of them is a time-invariant variable. There may be time-invariant characteristics that will help to explain energy intensity and its components, but these will be subsumed within the provincial fixed effects.

5.2 Fixed-Effect Regression and its Result

A panel data set has multiple observations on the same economic units. In this study, I have collected multiple observations on energy intensity over time. In this set of panel data, each element has the characteristics of cross-regional and cross-time. As a result, panel data can have time effects, or group effects, or both.

Given the panel data, I define models that arise from the most general linear representation:

$$\begin{aligned} & y_{it} = \sum_{k=1}^{k} x_{kit}\beta_{kit} + \varepsilon_{it} \\ (6) \qquad & i = 1, \ldots\ldots, N \\ & t = 1, \ldots\ldots, T \end{aligned}$$

where N is the number of provinces (29), T is the number of periods (19 years), and k is the number of identified variables (e.g., pci, i/k ratio, etc). In this study, the linear representation is:

$$\begin{aligned} & \log(E/Y)_{it} = \\ (7) \qquad & \alpha_i + \beta_1 \log\ fuel_{it} + \beta_2 \log pci_{it} + \beta_3 (\log pci_{it})^2 \\ & + \beta_4 \log hdd_{it} + \beta_5 \log cdd_{it} + \beta_6 time_t + \beta_7 time_t^2 \\ & + \beta_8 \log kl_{it} + \beta_9 (\log kl_{it})^2 + \beta_{10} \log ik_{it} + \beta_{11} (\log ik_{it})^2 \end{aligned}$$

The effects can be analyzed by either the fixed-effect (FE) or random-effect (RE) models. I conduct the Hausman test to decide from which model to choose. Hausman test's null hypothesis--that the RE estimator is consistent--is rejected. The provincial-level individual effects do appear to be correlated with the regressors. I decided to start with the FE model.

Table 5: Panel Regression with FE model

```
Fixed-effects (within) regression               Number of obs      =       551
Group variable (i): id                          Number of groups   =        29

R-sq:  within  = 0.7990                         Obs per group: min =        19
       between = 0.3146                                        avg =      19.0
       overall = 0.3520                                        max =        19

                                                F(11,511)          =    184.62
corr(u_i, Xb)  = -0.8051                        Prob > F           =    0.0000

------------------------------------------------------------------------------
          ei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fuel |   1.043143   .2064475     5.05   0.000     .6375529    1.448734
         pci |  -1.687456   .2020048    -8.35   0.000    -2.084318   -1.290594
        pci2 |    .022531    .025861     0.87   0.384     -.028276    .0733381
         hdd |   .0748148   .0536582     1.39   0.164     -.030603    .1802325
         cdd |   .0017877   .0312503     0.06   0.954    -.0596071    .0631825
        time |  -.0664659   .0100647    -6.60   0.000    -.0862392   -.0466926
       time2 |   .0021448   .0002643     8.11   0.000     .0016255     .002664
          kl |    .569478    .057302     9.94   0.000     .4569015    .6820545
         kl2 |  -.1166289   .0223972    -5.21   0.000    -.1606308    -.072627
          ik |   .3335833   .1169662     2.85   0.005     .1037895    .5633771
         ik2 |   .0704051   .0272688     2.58   0.010     .0168324    .1239779
       _cons |   4.591293   .6828185     6.72   0.000     3.249816     5.93277
-------------+----------------------------------------------------------------
     sigma_u |  .66719932
     sigma_e |  .12688754
         rho |  .96509428   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(28, 511) =    54.76             Prob > F = 0.0000
```

Source: author

The within R-square term is 0.799, and it shows that the data are explaining a large amount of the variation in energy intensity. The estimate of rho suggests that almost all the variation in energy intensity (ei) is related to provincial differences in ei values. The F test following the regression indicates that there are significant individual (provincial level) effects. As I also observe that the corr(u_i, xb) is -0.80, standing for a high correlation between u_i and the regressors in the model, using FE model is correct. All the coefficients except for time trend represent the elasticity. For example, when per capita income increases by 1%, the energy intensity will reduce by 1.68%.

The coefficient of fuel price is positive and significant at the 1% level. It is counter intuitive as under normal market conditions, an increase in fuel prices should induce energy conservation, resulting in a lowering of energy intensity. A possible explanation is that the inventory management behavior may play a role. Also, as discussed before, the non-market effects in some state-owned enterprises may cause energy-consumption patterns to be insensitive to the fuel price.

I observe that per capita income is negatively related to the energy intensity and is significant. Its quadratic term is positive, showing a curve opening upward. The turning point is at 37.5, which is far beyond the sample range (max=4.03). As a result, the effect of pci is always negatively related to the energy intensity change. As mentioned before, as the personal income increases, people will tend to switch from biomass to fuel energy, thus leading to a decline of energy intensity. The effect of income is consistent with the finding of Metcalf (2008) with the US data. The difference is that, with the China data, the speed of the energy-intensity change with respect to income is decreasing; while with the US data, the speed is increasing. Such a contrast indicates that the two countries are at two different development phases. For example, switching from using coal to electricity will lead to reducing energy intensity, but not as effectively as switching from electricity to green technology, etc.

Similarly, the capital-labor ratio could go either way. Its coefficient is positive and its quadratic term is negative, and thus showing a downward curve. The turning point is at 2.43, which is well beyond the mean 1.11. The effect can be considered to be always positive to the energy- intensity change. China is a developing country and still in the middle of its industrialization process. Most of the fixed-asset investment takes place in the construction, expansion, and operation of the large utilities. When the ratio is increasing, it indicates that the accumulated fixed-asset capital is increasing, resulting in an increase of energy intensity.

The coefficients of heating degree days and cooling degree days are both positive as expected. When the heating degree days increase, the energy intensity goes up, although the t-score shows that cooling degree days have no significant impact. At the aggregate level, it makes sense because in most parts of China, heating systems, such as indoor furnaces, are available, while cooling systems (e.g., air conditioners) are not.[7] As a result, energy consumption responds to cold days more than to hot days.

The investment-capital ratio seems to be behaving as expected. The investment here is the total investment in fixed assets. The turning point is at -2.38, well beyond the overall mean -1.91.With increased investment, the country is adding larger quantities of equipment, machinery, construction, and infrastructure facilities, which embody more energy consumption.

The coefficient of time trends is negative, as expected. The effect of technology improvement and innovation along time is obvious. However, the fact that the t-score is large makes me

---

[7] China's rural population stood at 737 million--56 percent of the total population of more than 1.3 billion--at the end of 2006. The ownership of air conditioners is only 7 units per 100 rural households (China Statistical Yearbook 2007).

suspect that the regression is spurious. The classical regression techniques are invalid when applied to any time-series variables that behave in a "trend-like" manner, as such techniques are only designed for variables that are stationary. However, the data I use in this study are likely to be non-stationary time series. After the first run result, I dropped the cooling degree days as it is not significant, and I dropped the time trend and its quadratic term to avoid the spurious effect.

Table 6: Panel Regression with FE model (modified)

```
Fixed-effects (within) regression               Number of obs      =       551
Group variable (i): id                          Number of groups   =        29

R-sq:  within  = 0.7720                         Obs per group: min =        19
       between = 0.4563                                        avg =      19.0
       overall = 0.5357                                        max =        19

                                                F(8,514)           =    217.49
corr(u_i, Xb)  = -0.2771                        Prob > F           =    0.0000

------------------------------------------------------------------------------
          ei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fuel |    .146821   .1614133     0.91   0.363    -.1702899    .4639318
         pci |  -1.414077   .2082578    -6.79   0.000    -1.823218   -1.004935
        pci2 |   .1021214   .0230525     4.43   0.000     .0568327    .1474101
         hdd |   .1070913   .0556499     1.92   0.055    -.0022379    .2164205
          kl |   .7482551   .0560177    13.36   0.000     .6382032     .858307
         kl2 |  -.1422248   .0219676    -6.47   0.000    -.1853821   -.0990676
          ik |   .5041655   .1221325     4.13   0.000     .2642252    .7441058
         ik2 |   .0986575   .0287019     3.44   0.001       .04227     .155045
       _cons |   2.600188   .6165127     4.22   0.000     1.388993    3.811383
-------------+----------------------------------------------------------------
     sigma_u |  .33845615
     sigma_e |  .13474748
         rho |   .8631832   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(28, 514) =    69.92             Prob > F = 0.0000
```

Source: author

As shown in Table 6, the time trend may have caused the spurious effects. Comparing the current result to that of Table 5, I see the regression results are rather improved: the significance of fuel price has been weakened (I cast doubt on the previous result anyway); the effect of heating degree days is significant at the 5% level; the t-scores of all the other variables are significant at the 1% level. For those variables with a quadratic term, the turning points do not vary much from previous results. The interpretation for all the variables remains the same as before.

5.3 Regression with Three Economic and Geographical Regions

China is divided into provinces, autonomous regions, and municipalities directly under the Central Government. As shown in Table 4, four out of six variables have a larger between variance than within variance. In order to capture the economic and geographical variance, and taking into account factors such as the difference between north and south, west and east, I categorize the 29 provinces into three major regions.

Table 7: China's Three Major Regions: North, South and West

North	Region 1	Liaoning, Jilin, Heilongjiang, Beijing, Tianjin, Hebei, Shandong, Henan, Shanxi, Shaanxi, Inner Mogolia
South	Region 2	Shanghai, Jiangsu, Zhejiang, Fujian, Guangdong, Guangxi, Hainan, Hubei, Hunan, Jiangxi, Anhui
West	Region 3	Sichuan and Chongqing,[8] Guizhou, Yunan, Gansu, Qinghai, Ningxia, Xinjiang

Source: http://www.china.com.cn/chinese/zhuanti/2004xdh/505635.htm

In order to avoid the spurious effect, I again drop the time-trend factor and its quadratic term. The regression result by adding regional dummy variables (code refers to appendix) is presented in Table 8.

---

[8] Chongqing province used to be a sub-provincial city within Sichuan province until 1997. It has become the fourth municipality in China since then. In this research, I studied Chongqing and Sichuan as one combined region.

Table 8: Descriptive Statistics of the Regional Dataset

Variable	Mean	Std. Dev.	Min	Max
ln(ei)	0.63	0.51	(0.94)	1.74
ln(fuel_1)	(0.32)	0.54	(1.97)	0.17
ln(fuel_2)	(0.33)	0.58	(2.16)	0.14
ln(fuel_3)	(0.18)	0.40	(1.56)	0.13
ln(pci_1)	0.78	1.08	-	3.89
ln(pci_2)	0.74	1.07	-	4.03
ln(pci_3)	0.35	0.67	-	2.65
ln(hdd_1)	1.68	2.19	-	5.19
ln(hdd_2)	1.21	1.86	(2.30)	4.75
ln(hdd_3)	0.90	1.63	-	5.04
ln(cdd_1)	1.08	1.49	-	4.49
ln(cdd_2)	1.44	1.88	-	5.00
ln(cdd_3)	0.40	1.18	(2.30)	4.86
ln(kl_1)	0.49	0.76	-	3.05
ln(kl_2)	0.39	0.72	(0.69)	3.34
ln(kl_3)	0.22	0.50	(0.22)	2.31
ln(ik_1)	(0.74)	0.96	(2.53)	-
ln(ik_2)	(0.69)	0.90	(2.53)	-
ln(ik_3)	(0.48)	0.86	(3.91)	-

Source: author

Table 9: Panel Regression with Regional Data

```
Fixed-effects (within) regression               Number of obs      =       551
Group variable (i): id                          Number of groups   =        29

R-sq:  within  = 0.8386                         Obs per group: min =        19
       between = 0.2821                                        avg =      19.0
       overall = 0.1324                                        max =        19

                                                F(27,495)          =     95.22
corr(u_i, Xb)  = -0.9676                        Prob > F           =    0.0000

------------------------------------------------------------------------------
          ei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      fuel_1 |  -.8632356    .288094    -3.00   0.003    -1.429274   -.2971977
      fuel_2 |  -.3397393   .2462692    -1.38   0.168    -.8236011    .1441225
      fuel_3 |  -.1685225   .3153979    -0.53   0.593    -.7882062    .4511613
       pci_1 |  -.2897118   .3513865    -0.82   0.410    -.9801047    .4006812
       pci_2 |  -1.175431   .2898716    -4.06   0.000    -1.744961   -.6059002
       pci_3 |   -1.16389   .4829932    -2.41   0.016     -2.11286   -.2149203
      pci2_1 |   .1703398   .0459223     3.71   0.000     .0801131    .2605666
      pci2_2 |   .1908066   .0315468     6.05   0.000     .1288244    .2527887
      pci2_3 |   .3320958   .0958546     3.46   0.001     .1437637    .5204279
       hdd_1 |  -.1793491   .1345248    -1.33   0.183    -.4436591    .0849609
       hdd_2 |   .2021312   .0585645     3.45   0.001     .0870656    .3171968
       hdd_3 |   .0528829   .1338085     0.40   0.693    -.2100197    .3157855
       cdd_1 |   .0125239   .0495662     0.25   0.801    -.0848622      .10991
       cdd_2 |   .3187505   .1083348     2.94   0.003     .1058977    .5316033
       cdd_3 |   .0015345   .0422507     0.04   0.971    -.0814783    .0845474
        kl_1 |   .3666125    .136547     2.68   0.007     .0983294    .6348956
        kl_2 |    .849215   .0682593    12.44   0.000     .7151014    .9833286
        kl_3 |   .2684027   .1432815     1.87   0.062    -.0131121    .5499175
       kl2_1 |  -.1665902   .0438513    -3.80   0.000     -.252748   -.0804325
       kl2_2 |  -.2146099   .0280571    -7.65   0.000    -.2697356   -.1594842
       kl2_3 |  -.1399244   .0733498    -1.91   0.057    -.2840399     .004191
        ik_1 |   2.644737   .3785528     6.99   0.000     1.900968    3.388505
        ik_2 |   .0514471   .2206691     0.23   0.816    -.3821165    .4850106
        ik_3 |   .3996918   .2772349     1.44   0.150    -.1450105     .944394
       ik2_1 |   .6590502   .0990665     6.65   0.000     .4644074    .8536929
       ik2_2 |   .0059802   .0625944     0.10   0.924    -.1170033    .1289638
       ik2_3 |   .0687889    .054135     1.27   0.204    -.0375738    .1751516
       _cons |   1.321454   .7122471     1.86   0.064     -.077946    2.720854
-------------+----------------------------------------------------------------
     sigma_u |  1.8494588
     sigma_e |  .11553225
         rho |   .9961129   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(28, 495) =    42.52             Prob > F = 0.0000
```

Source: author

As opposed to using the aggregate data, when I use regional data, the coefficients of fuel price are all negative, although only Region 1 (north region) is significant at the 1% level and Region 2 is significant at the 20% level. The price data I use here are the retail fuel-price-index data. It has two major components: the industry fuel prices and the gas prices for transportation use. The north region has the largest concentration of industrial enterprises of China, such as the "Northeast Iron Triangle"–Jilin, Heilongjiang and Liaoning, as well as the largest coal production and consumption base–Shanxi. By contrast, Region 2 includes many southern provinces with agriculture, textiles, and commerce as major industries, while Region 3 is relatively underdeveloped with almost no presence of heavy industries. Further, Regions 1 and 2 include most of the privately owned and industrial vehicles. It is different from the western part of China which has the most underdeveloped provinces. As per the discussion above, it makes sense that the energy intensity of the northern region is sensitive to fuel price changes.

Energy intensity, for all three regions, exhibits a quadratic response to per capita income, first falling then rising. The coefficient of the quadratic term is positive, which indicates a convex relationship, meaning that the decreasing speed is also decreasing as time goes by. It is consistent with the finding at the aggregate level. China is still a developing country, where underdeveloped rural areas count for about 55%. The effect of the income factor has two phases. When personal income increases, residents start to switch from biomass fuel to electricity, thus being more efficient. At a certain point, per capita income will become positively related to energy intensity. This is because when incomes increase further, residents will increase expenditures on electrical appliances, thus resulting in more energy consumption. I believe that the total effect of the income factor is the combination of the two phases described above. For

Regions 2 and 3, Phase one dominates and that is why the statistical result is significant at the 1% level. However, for Region 1, it is a mixed result of both phases and that is why it is not statistically significant. It also has the smallest coefficient, which further proves my point.

As for the weather factor, it is only significant for Region 2, southern China. The coefficient of cooling degree days is higher than that of heating degree days, indicating that the elasticity of the former factor is greater. It makes sense as cooling systems are more needed for warm southern areas than heating systems. However, it is still hard for me to explain the significance of heating degree days, as heating systems are not widely used in southern China in real estate construction, where cold days are rare.

The capital-labor ratio (K/L) is positively related to energy intensity, and it is statistically significant for all three regions. The energy intensity exhibits a quadratic response, first rising then falling. The turning point for this ratio is well above the mean value in each region's sample so that I simply assume that energy intensity is rising with respect to K/L for all regions for all years. It is consistent with the findings using aggregate data.

Finally, energy intensity exhibits a positive response to the investment-capital ratio (I/K), with only Region 1 being statistically significant at the 1% level and Region 3 being significant at the 15% level. It can be interpreted as the northern part of China, where heavy industrial enterprises concentrate, attracts most of the national fixed asset investment. Such investment has both an efficiency effect and an activity effect, depending on which phase it is in. During the construction phase, the enterprises barely have any productivity, and thus the activity effect (construction) dominates. This will lead to an energy-intensive result. During the stabilized phase, newly added investment (e.g., new expansion of plants, new equipment) is more efficient than the old facilities. The efficiency effect dominates, thus resulting in less energy use. As a result, the coefficient in Table 8 can be a mixed result of both effects. Region 1 has a significantly larger coefficient because the activity effect is much larger than the efficiency effect; region 2 has the smallest coefficient as the efficiency effect is catching up with the activity effect. Similar to what I have described for the per capita income factor, when the mixed effect takes place for a region, the statistical significance will decrease.

In this chapter, through panel regression analysis by both aggregate data and regional data, I find a positive relationship between energy intensity and heating degree days, cooling degree days, investment capital ratio and capital labor ratio. I also find a negative relationship between energy intensity and per capita income, retail fuel prices and time trend. Whenever there are discrepancy between the aggregate result and regional result, I will follow the latter one as it captures more regional variance.

Chapter VI

Conclusion

This study has built upon a large body of research and literature on energy intensity for both the United States and China. I have made several contributions:

First, I study the relationship of the transportation of coal and national energy intensity, based on the regional energy-efficiency variance as well as the geographical imbalance of energy production and consumption in China. I find that the lack of expansion of China's railway system and the limits of its capacity has negatively affected the decline of energy intensity.

Second, I create two indexes by decomposing China's national energy intensity into the efficiency effect and structural effect. Other analysts have defined the two key determinants of changes in energy intensity through separating the data by economic sectors (residential, commercial, industrial, and transportation). In this China study, I separate the data by geographical units and find that the efficiency effect outperforms the structural effect. In other words, the decline of China's energy intensity largely relies on the efficiency change within each province, but it is not affected by the fact that the more energy-efficient provinces increase their shares of output in total GDP.

Finally, I conduct the regression analysis by using China's provincial weather- and economic-related factors and analyze the energy-intensity changes at the national and regional levels. I find that fuel prices, heating degree days, capital-labor ratio, and investment-capital ratio are positively related to energy intensity; cooling degree days has a minimum effect; while per capita income and time trends are negatively related to energy intensity. After dropping the time trend factor in order to avoid the spurious effect, I obtain an improved regression result. I further aggregate the 29 provinces into three regions (north, south and west), and the regional variance is pronounced. The fuel price responds negatively to energy-intensity changes. The effects of heating degree days and cooling degree days are only significant for the southern region. Finally, factors like per capita income and investment-capital ratio have different effects on energy intensity during different phases. A region can thus have a mixed effect as its provinces are in different phases leading to a mitigated coefficient and a weakened t-score. For future study, a more appropriate way to break down the 29 provinces may lead to a better regression result.

For the transportation study, further analysis can be done if the coal and coke import, export, and inventory data are collected. For example, when transportation infrastructure reaches its capacity, whether the undelivered coal is sitting as inventory, or it is used in a rather inefficient way can lead to different interpretations on energy-intensity change. The impact of transportation can be even further measured if the locations of coal-fired power plants and iron and steel production plants are indentified. However, these data are fragmented at provincial levels and are not readily available.

In future studies, analysts can also focus on the types of fuel energy that each province is using. The degree of reliance on coal varies from province to province. Through a further breakdown of provincial energy consumption by fuel type, greenhouse emissions can be measured. As a result, the links among carbon dioxide (CO_2) emissions, energy intensity, and economic- and weather-related factors can be established. That will serve as a basis for further environmental policy implications, which I did not discuss in this study.

Appendix I: Data Sources

Raw data

Gross Regional Product (GRP) and Indexes:

The GRP indexes are collected from China Statistical Yearbooks. The GRP is recorded at current prices, while the indexes are calculated at constant prices. Data from 2002 to 2003 were modified based on the First China Economic Census results in 2004.

Regional energy consumption:

The energy consumption data for each region are collected from the China Energy Statistical Yearbook. The unit is in 10,000 tons of coal equivalent. Prior to 1996, the consumption for Chongqing was included in Sichuan. Because each region adopts a different conversion factor, the sum of all regions' energy consumption is not equal to the national level data. Some provinces' energy consumption from 1991-1994 are estimated using a simple linear estimation, as data are not available.

Fuel index (in current price):

The fuel price index is recorded from the retail price indexes of the fuel commodity section, China Statistical yearbook. The data are recorded with the preceding year = 100. The data are recorded in nominal terms but not real terms and I have deflated the data to 2005 currency.

Population:

The data are recorded from the China Statistical Yearbook. According to the Statistical Yearbook, There was a National Sample Survey on population changes in 2006, and the national total population is adjusted on the basis of sampling errors and survey errors, but similar adjustments are not made to regional figures.

Fixed-asset investment (nominal):

The data are recorded from the China Statistical Yearbook. The total investment in fixed assets in the whole country refers to the volume of activities in construction and purchases of fixed assets of the whole country and related fees, expressed in monetary terms during the reference period. It is a comprehensive indicator, which shows the size, structure, and growth of the investment in fixed assets, providing a basis for observing the progress of construction projects and evaluating the results of investment (China Statistical Yearbook, 2007).

Price indexes of investment in fixed assets by region (in current prices):
The data are recorded from the China Statistical Yearbook, and it is a chain index recorded in current prices. It reflects the trend and degree of changes in prices of investment goods and projects in fixed assets during a given period. The investment in fixed assets consists of three components, namely the investment in construction and installation, the investment in purchases of equipment and instruments, and the investment in other items. Price indexes of investment in fixed assets are calculated as the weighted arithmetic mean of the price indexes of the three components of investment in fixed assets (China Statistical Yearbook, 2007). However, in the Yearbook, such indexes from 1988 to 1990 for all provinces are not recorded. I have thus estimated the indexes for each province during this period based on the average of the following five years. Similarly, I have estimated the indexes for Guangdong province from 1988 to 2000, and Hainan province from 1988 to 1999 by using a simple linear estimation.

Labor:
It is recorded from the section "number of employed persons at the year-end by three industries and regions" from the *China Statistical Yearbook.* According to the statistical yearbook, employed persons refer to persons aged 16 and over who are engaged in working and receive remuneration or earn business income.

Heating degree days (HDD) and Cooling degree days (CDD):
HDD and CDD are quantitative indexes designed to reflect the demand for energy needed to heat or cool a home or business. More specifically, the number of heating degrees in a day is defined as the difference between a reference value of 65°F (18°C) and the average outside temperature for that day. In this research, the monthly average temperature of the capital city of each province is collected from the *China Statistical Yearbook* and is used as a representative temperature for the whole province.

Processed data

Gross Regional Product (in 2005 currency):

The spreadsheet of gross regional product in 2005 currency is derived from the 2005 gross regional product and the index in constant prices, following the formula:

GRP year n = GRP year n+1 / GRP index year n+1 * 100

Energy intensity:

Energy intensity is calculated by dividing the energy consumption by the regional GDP measured in 2005 yuan. The unit is in (tce/million 2005 yuan).

Energy intensity = Energy consumption / GRP

Per capita income:

Per capita income is often used as a measure of the wealth of the population of a region, particularly in comparison to other regions. It is measured in units of 2005 currency.

Per capita income year n = gross regional product year n / regional population year n.

Fixed asset investment (in 2005 yuan):

The fixed-asset investment in 2005 prices is achieved by having the investment figure in nominal terms deflated by the fixed-asset investment indexes.

Capital stock:

The capital stock is calculated by using the perpetual inventory method (PIM). First, I estimate the starting value for capital K0 using the simplifying assumption that the economy is initially on a steady-state growth path, meaning that the growth rates of capital, labor, and output are the same. The formula is,

K0 = I0 / (g + delta)

I0 is the initial investment. Delta stands for the depreciation rate, and I use a value of 0.05. g as the average rate of growth of the economy over the entire period of the sample. I have used the average of all year-on-year growth rates of provincial GDP in real terms. Finally, each year's capital stock is calculated by

$K(t+1) = I(t) + (1-\text{delta}) * K(t)$

Investment-capital ratio:
This ratio, as the name suggests, is calculated by dividing each year's fixed asset investment data by that year's capital stock.

Capital-labor ratio:
This ratio, again as the name suggests, is calculated by dividing each year's capital stock by the total number of employers. The unit is in 10,000 yuan.

Weighted HDD and CDD:
When the NOAA (the US National Oceanographic and Atmospheric Administration) constructs the HDD and CDD data series, they use population weights. This is because it is not merely the duration of weather episodes above a comfortable temperature that determine service demands for heating and cooling, but also the size of the population that is exposed to these weather conditions. Consider the example of Death Valley, CA. It is one of the hottest places in the US, but almost nobody lives there. The procedure is as follows. First, for each year, compute the share of each province in China's total population. This is the time-varying weight. Then multiply HDD and CDD in each province and year by the corresponding weight.

Appendix II: The Main Industry and GDP of Each Province

Table 10: The Main Industry and GDP of Each Province

Province	Main industry	GDP (in 2006 100million yuan)	%
Beijing	administration	7,870	3.4%
Tianjin	heavy industry, major port	4,359	1.9%
Hebei	coal,oil	11,660	5.0%
Shanxi	agriculture, coal	4,753	2.1%
Inner Mongolia	agriculture (cattle, sheep)	4,791	2.1%
Liaoning	coal, iron, oil, engineering	9,251	4.0%
Jilin	agriculture, petrochemicals	4,275	1.9%
Heilongjiang	coal, timber, oil	6,189	2.7%
Shanghai	industry, major port	10,366	4.5%
Jiangsu	silk, cotton, industry	21,645	9.4%
Zhejiang	silk, rice, fish, industry	15,743	6.8%
Anhui	agriculture (rice, tea, timber), industry, coal	6,149	2.7%
Fujian	agriculture (rice, sugar, fruit)	7,615	3.3%
Jiangxi	agriculture (rice, sugar, tea), copper, tungsten	4,671	2.0%
Shandong	agriculture (tobacco, peanuts), oil, coal	22,077	9.6%
Henan	agriculture (grain), coal, machinery, textiles	12,496	5.4%
Hubei	agriculture (rice, cotton), iron, phosphates	7,581	3.3%
Hunan	agriculture	7,569	3.3%
Guangdong	light industry, commerce, agriculture	26,204	11.3%
Guangxi	agriculture, textiles, iron	4,829	2.1%
Hainan	agriculture	1,053	0.5%
Sichuan&Chongqing	agriculture, mining	12,129	5.4%
Guizhou	agriculture	2,282	1.0%
Yunnan	agriculture (rice), tourism	4,007	1.7%
Shaanxi	aviation, textiles	4,524	2.0%
Gansu	agriculture (grain), coal, mining	2,277	1.0%
Qinghai	agriculture (sheep), oil, mining	642	0.3%
Ningxia	coal, agriculture, metalworking	711	0.3%
Xinjiang	textiles, agriculture	3,045	1.3%

Source: author through searching various online resources.

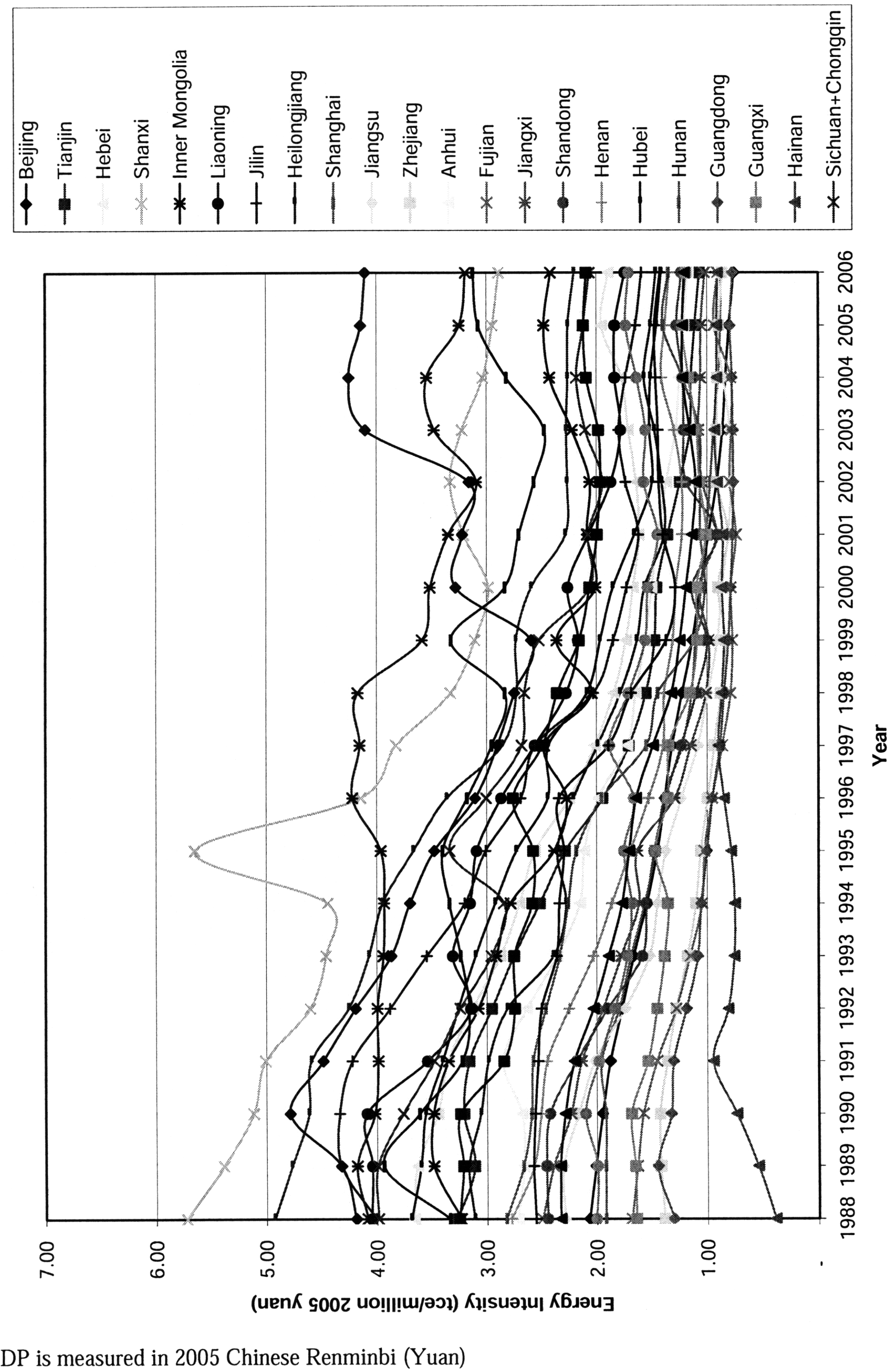
Appendix III: Provincial Energy Intensity of China (1990-2006)

Table 11: Provincial Energy Intensity from 1990 to 2006
(unit in tce/million 2005 yuan)

ID	Region	1990	1995	2000	2005	2006
1	Beijing	1.95	1.45	1.06	0.80	0.76
2	Tianjin	3.21	2.28	1.45	1.11	1.07
3	Hebei	3.45	2.56	1.67	1.96	1.89
4	Shanxi	5.11	5.65	2.98	2.95	2.89
5	Inner Mongolia	3.48	2.39	2.01	2.48	2.41
6	Liaoning	4.09	3.09	2.26	1.83	1.74
7	Jilin	4.33	3.01	1.73	1.65	1.59
8	Heilongjiang	3.56	2.73	1.85	1.46	1.41
9	Shanghai	1.92	1.46	1.05	0.88	0.87
10	Jiangsu	2.16	1.37	0.86	0.92	0.89
11	Zhejiang	1.43	1.06	0.90	0.90	0.86
12	Anhui	2.68	2.11	1.50	1.21	1.17
13	Fujian	1.58	1.03	0.79	0.94	0.91
14	Jiangxi	2.23	1.63	1.07	1.06	1.02
15	Shandong	2.42	1.48	1.14	1.28	1.23
16	Henan	2.51	1.69	1.28	1.38	1.34
17	Hubei	3.06	2.36	1.56	1.51	1.46
18	Hunan	2.59	2.18	1.02	1.40	1.35
19	Guangdong	1.33	1.00	0.79	0.79	0.77
20	Guangxi	1.68	1.47	1.09	1.22	1.19
21	Hainan	0.74	0.78	0.87	0.92	0.91
22	Sichuan&Chongqing	3.76	3.34	2.06	2.12	2.06
23	Guizhou	4.02	3.96	3.51	3.25	3.19
24	Yunnan	2.10	1.75	1.53	1.73	1.71
25	Shaanxi	2.56	2.29	1.29	1.48	1.42
26	Gansu	4.61	3.67	2.59	2.26	2.20
27	Qinghai	3.61	3.40	2.83	3.07	3.12
28	Ningxia	4.78	3.47	3.28	4.14	4.10
29	Xinjiang	3.24	2.58	2.06	2.11	2.09
	National	2.29	1.70	1.19	1.22	1.21

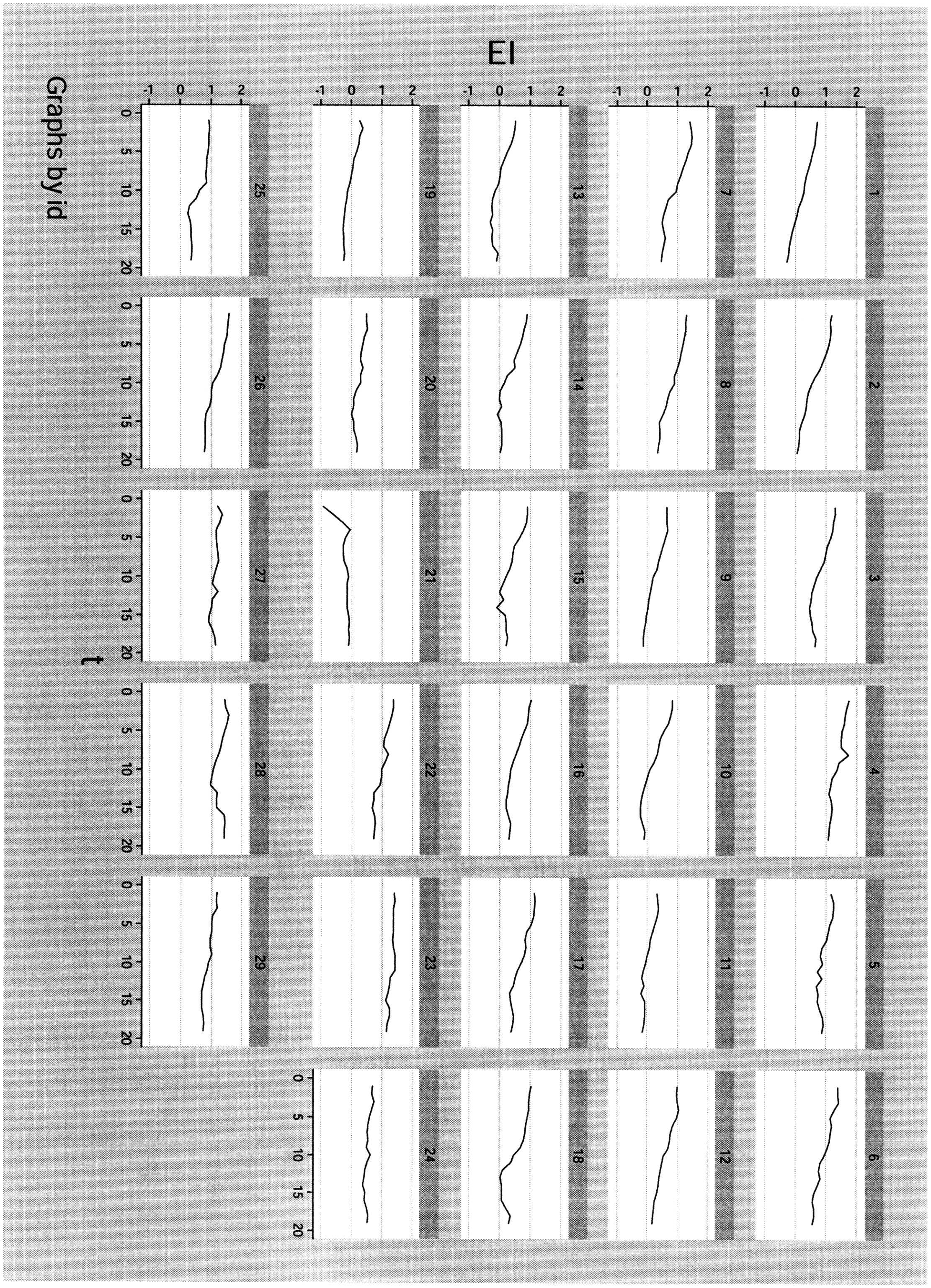
Source: author's calculations based on China Statistical Yearbooks, China Energy Statistical Yearbook, and various China provincial statistical yearbook (1989-2007)

Figure 16: Provincial Energy Intensity from 1988 to 2006



Note: GDP is measured in 2005 Chinese Renminbi (Yuan)
Source: author's calculations based on *China Statistical Yearbook*, *China Energy Statistical Yearbook*, and China provincial statistical yearbook (1989-2007)

Figure 17: Energy Intensity in Log Form of 29 Provinces in 19 Years



Note: The ID number of each province refer to Table 11, year 1=1988, year 19=2006
Source: author

Appendix IV: China Energy-Intensity Indexes (1988-2006)

Table 12: China Energy-Intensity Indexes--National, Coastal, and Inland

	National			Coastal			Inland		
Year	Intensity	Within index	Between index	Intensity	Within index	Between index	Intensity	Within index	Between index
1988	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1989	1.00	1.02	1.00	1.00	1.04	1.00	0.99	0.99	1.00
1990	0.98	1.04	1.00	0.97	1.05	1.00	0.96	1.03	1.00
1991	0.95	1.05	1.00	0.90	1.05	1.00	0.93	1.06	1.01
1992	0.87	1.07	1.00	0.81	1.08	1.00	0.87	1.05	1.01
1993	0.81	1.01	0.99	0.74	1.00	0.99	0.81	1.02	1.01
1994	0.76	0.98	0.99	0.69	1.01	0.99	0.77	0.96	1.01
1995	0.73	0.92	0.99	0.66	0.95	0.99	0.78	0.89	1.01
1996	0.70	0.84	0.99	0.61	0.87	0.99	0.70	0.82	1.01
1997	0.64	0.77	0.98	0.55	0.78	0.99	0.65	0.75	1.01
1998	0.57	0.70	0.98	0.50	0.72	0.99	0.59	0.68	1.00
1999	0.54	0.69	0.97	0.47	0.69	0.98	0.55	0.69	1.00
2000	0.51	0.66	0.97	0.47	0.65	0.97	0.52	0.66	1.00
2001	0.49	0.61	0.95	0.44	0.60	0.97	0.50	0.62	0.99
2002	0.47	0.56	0.94	0.45	0.54	0.96	0.50	0.58	0.99
2003	0.50	0.54	0.94	0.45	0.51	0.95	0.52	0.56	0.99
2004	0.52	0.52	0.93	0.45	0.50	0.94	0.53	0.55	0.99
2005	0.53	0.52	0.93	0.47	0.50	0.94	0.53	0.54	0.99
2006	0.52	0.52	0.93	0.45	0.50	0.94	0.52	0.54	0.99

Source: author's calculation; data from *China Statistical Yearbooks*, and *China Energy Statistical Yearbooks*

Table 13: U.S. Energy-Intensity Indexes

Year	Intensity	Efficiency	Structure
1970	1.00	1.00	1.00
1971	0.99	0.99	1.00
1972	0.98	0.98	1.01
1973	0.97	0.96	1.01
1974	0.95	0.95	1.00
1975	0.93	0.92	1.00
1976	0.93	0.92	1.01
1977	0.91	0.90	1.01
1978	0.89	0.88	1.01
1979	0.87	0.87	0.99
1980	0.84	0.85	0.99
1981	0.80	0.81	0.99
1982	0.78	0.79	1.00
1983	0.75	0.77	0.98
1984	0.73	0.75	0.98
1985	0.70	0.73	0.97
1986	0.68	0.72	0.95
1987	0.68	0.71	0.95
1988	0.68	0.71	0.96
1989	0.68	0.71	0.95
1990	0.66	0.70	0.94
1991	0.66	0.71	0.93
1992	0.65	0.71	0.92
1993	0.65	0.71	0.92
1994	0.63	0.69	0.92
1995	0.63	0.69	0.92
1996	0.63	0.69	0.92
1997	0.60	0.66	0.91
1998	0.58	0.67	0.88
1999	0.57	0.66	0.87
2000	0.56	0.65	0.87
2001	0.54	0.63	0.86
2002	0.54	0.63	0.86
2003	0.53	0.61	0.86

Source: Metcalf (2008, Appendix, Table A3)

Appendix V: Hausman Test

According to Baum (2006), for a given observation, an intercept varying over units results in the structure

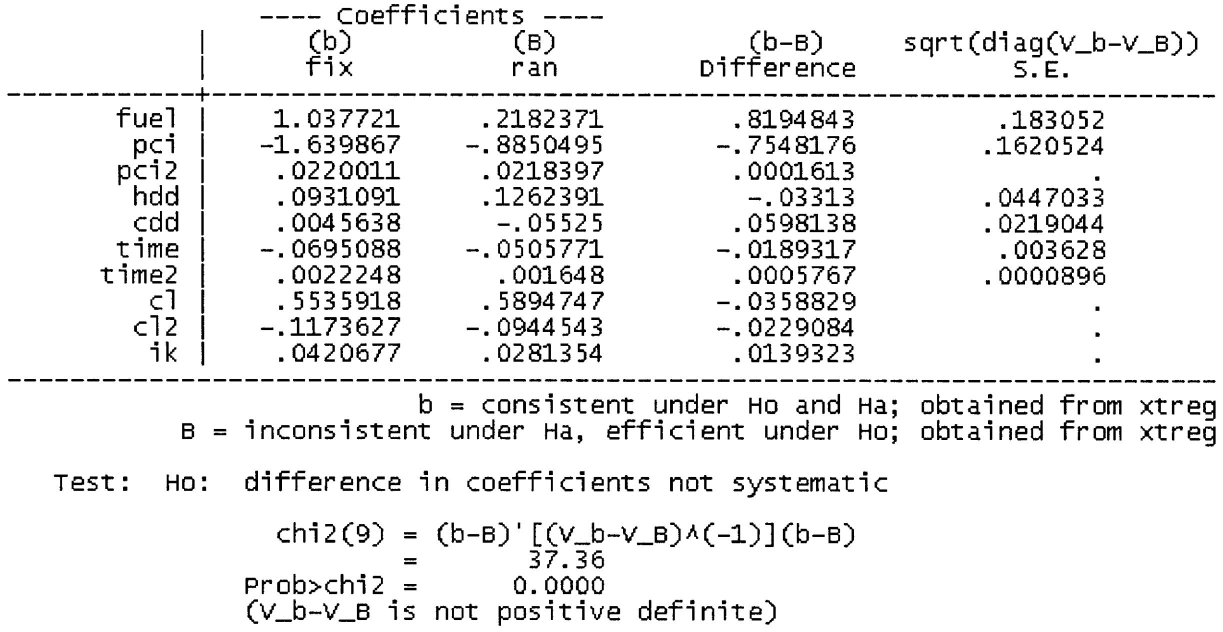
$$y_{it} = x_{it}\beta_k + z_i\delta + u_i + \varepsilon_{it}$$

Where x_{it} is a 1x k vector of variables that vary over individual and time, β is the k x 1 vector of coefficients on x, z_i is a 1x p vector of time-invariant variables that vary only over individuals, δ is the p x 1 vector of coefficients on z, u_i is the individual-level effect, and ε_{it} is the disturbance term. The u_i are either correlated or uncorrelated with the regressors in x_{it} and z_i and are always assumed to be uncorrelated with ε_{it}.

If the u_i are uncorrelated with the regressors, they are known as random effects (RE), but if the u_i are correlated with the regressors, they are known as fixed effects (FE). The origin of the term RE is clear: when u_i are uncorrelated with everything else in the model, the individual-level effects are simply parameterized as additional random disturbances. The sum $u_i + \varepsilon_{it}$ is sometimes referred to as the composite-error term and the model is sometimes known as an error-components model. The origin of the term FE is more elusive. When the u_i are correlated with some of the regressors in the model, one estimation strategy is to treat them like parameters or FE. The solution is to remove the u_i from the estimation problem by a transformation that still identifies some of the coefficients of interest. To decide which model to choose from, I first conducted the Hausman test.

As shown in Table 14, the Hausman test's null hypothesis- that the RE estimator is consistent- is rejected. The provincial-level individual effects do appear to be correlated with the regressors. I decided to start with the one-way FE model.

Table 14: Hausman Test



```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      fix          ran         Difference          S.E.
-------------+----------------------------------------------------------------
        fuel |   1.037721     .2182371        .8194843         .183052
         pci |  -1.639867    -.8850495       -.7548176        .1620524
        pci2 |   .0220011     .0218397        .0001613            .
         hdd |   .0931091     .1262391         -.03313        .0447033
         cdd |   .0045638      -.05525        .0598138        .0219044
        time |  -.0695088    -.0505771       -.0189317         .003628
       time2 |   .0022248      .001648        .0005767        .0000896
          cl |   .5535918     .5894747       -.0358829            .
         cl2 |  -.1173627    -.0944543       -.0229084            .
          ik |   .0420677     .0281354        .0139323            .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtreg
           B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       37.36
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```

As a matter of fact, even if I included here some unobserved province-specific but time-invariant variables into the regression, by choosing the FE model, the regression will automatically drop them as the within variance is zero.

Appendix VI: Programs and Codes

1. Indices Creation (GAMS)

```
sets
      i     regions /
               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     Sichuan
              23     Guizhou
              24     Yunnan
              25     Shaanxi
              26     Gansu
              27     Qinghai
              28     Ningxia
              29     Xinjiang/

      t     years  /1988 * 2006/

parameters
        e                regional energy consumption
        y                regional gdp
        e_agg            aggregate energy consumption
        y_agg            aggregate gdp
        ey               energy intensity
        yshr             provincial gdp share
```

```
        w               mean
        dlog_ey         dlog_ey
        dlog_yshr       dlog_yshr
        within          efficiency change
        between         structural change
        within_idx      within index
        between_idx     between index

;

table econs(*,*) energy consumption

{insert energy consumption table}
;

table gdp(*,*) real gdp

{insert real gdp table}
;

e(i,t)          =       econs(i,t);
y(i,t)          =       gdp(i,t);
e_agg(t)        =       sum(i, e(i,t));
y_agg(t)        =       sum(i, y(i,t));
ey(i,t)         =       e(i,t) / y(i,t);
yshr(i,t)       =       y(i,t) / y_agg(t);
w(i,t)          =       0.5 * (e(i,t)/e_agg(t) + e(i,t-1)/e_agg(t-1));
dlog_ey(i,t)$(ord(t) > 1)       =       2 * (ey(i,t) - ey(i,t-1)) / (ey(i,t) + ey(i,t-1));
dlog_yshr(i,t)$(ord(t) > 1)     =       2 * (yshr(i,t) - yshr(i,t-1)) / (yshr(i,t) + yshr(i,t-1));
within(t)                       =       sum(i, w(i,t) * dlog_ey(i,t));
between(t)                      =       sum(i, w(i,t) * dlog_yshr(i,t));
within_idx(t)$(ord(t) = 1)      =       1;
between_idx(t)$(ord(t) = 1)     =       1;

loop(t$(ord(t) > 1),

      within_idx(t) = within_idx(t-1) * (2 + within(t)) / (2 - within(t));

);

display within_idx, between_idx;
```

2. Indices Creation (Matlab)

```
clear all

gdp = load('gdp.txt');
econs = load('econs.txt');

e_agg = [];
y_agg = [];

for t=1:19
  temp_econ = 0;
  temp_gdp = 0;
  for i=1:29
    temp_econ = temp_econ + econs(i,t);
    temp_gdp = temp_gdp + gdp(i,t);
  end
  e_agg = [e_agg; temp_econ];
  y_agg = [y_agg; temp_gdp];
end

for t=1:19
  for i=1:29
    ey(i,t) = econs(i,t)/gdp(i,t);
    yshr(i,t) = gdp(i,t) / y_agg(t);
  end
end



for t=2:19
  for i=1:29
    w(i,t) = 0.5 * (econs(i,t)/e_agg(t) + econs(i,t-1)/e_agg(t-1));
    dlog_ey(i,t) = 2*(ey(i,t)-ey(i,t-1))/(ey(i,t)+ey(i,t-1));
    dlog_yshr(i,t) = 2*(yshr(i,t)-yshr(i,t-1))/(yshr(i,t)+yshr(i,t-1));
  end
end

within = [];
between = [];
within = [0;within];
between =  [0;between];

for t=2:19
  temp_within = 0;
  temp_between = 0;
  for i=1:29
    temp_within = temp_within + w(i,t)*dlog_ey(i,t);
    temp_between = temp_between + w(i,t)*dlog_yshr(i,t);
  end
  within = [temp_within; within];
  between = [temp_between; between];
end

within_idx(1) = 1;
between_idx(1) = 1;
```

```
for t=2:19
  within_idx(t) = within_idx(t-1)*(2+within(t))/(2-within(t));
  between_idx(t) = between_idx(t-1)*(2+between(t))/(2-between(t));
end

within_idx
between_idx
```

3. Panel Data Regression for Aggregate Regions (STATA 9)

```
tsset id t
xtreg ei fuel pci pci2 hdd cdd time time2 cl ik, fe

gen dr1=0

replace dr1=1 if (id==6) | (id==7) | (id==8) | (id==1) | (id==2) | (id==3) | (id==15) | (id==16) |
(id==4)| (id==25) | (id==5)

gen dr2=0

replace dr2=1 if (id==9) | (id==10) | (id==11) |(id==13) | (id==19) | (id==20) | (id==21) |
(id==17)| (id==18) | (id==14) | (id==12)

gen dr3=0

replace dr3=1 if (id==22) | (id==23) | (id==24) | (id==26)| (id==27) | (id==28) | (id==29)


gen fuel_1 = fuel * dr1
gen fuel_2 = fuel * dr2
gen fuel_3 = fuel * dr3



gen pci_1 = pci * dr1
gen pci_2 = pci * dr2
gen pci_3 = pci * dr3


gen pci2_1 = pci2 * dr1
gen pci2_2 = pci2 * dr2
gen pci2_3 = pci2 * dr3


gen hdd_1 = hdd * dr1
gen hdd_2 = hdd * dr2
gen hdd_3 = hdd * dr3


gen cdd_1 = cdd * dr1
gen cdd_2 = cdd * dr2
gen cdd_3 = cdd * dr3

gen time_1 = time * dr1
```

```
gen time_2 = time * dr2
gen time_3 = time * dr3

gen time2_1 = time2 * dr1
gen time2_2 = time2 * dr2
gen time2_3 = time2 * dr3

gen cl_1 = cl * dr1
gen cl_2 = cl * dr2
gen cl_3 = cl * dr3

gen cl2_1 = cl2 * dr1
gen cl2_2 = cl2 * dr2
gen cl2_3 = cl2 * dr3

gen ik_1 = ik * dr1
gen ik_2 = ik * dr2
gen ik_3 = ik * dr3



xi: xtreg ei fuel_1-fuel_3 pci_1-pci_3 pci2_1-pci2_3 hdd_1-hdd_3 cdd_1-cdd_3 time_1-time_3
time2_1-time2_3 cl_1-cl_3 cl2_1-cl2_3 ik_1-ik_3, fe
```

Reference:

Auffhammer, Maximilian and Richard T. Carson. 2007. "Forecasting the Path of China's CO_2 Emissions Using Province Level Information", *CUDARE working papers*, University of California, Berkeley.

Boyd, Gale A. and Joseph M. Roop. 2004. "A Note on the Fisher Ideal Index Decomposition for Structural Change in Energy Intensity," *The Energy Journal*, 25 (l): 87-101

Fisher-Vanden, Karen, Gary H. Jefferson, Ma Jingkui, and Xu Jianyi. 2006. "Technology Development and Energy Productivity in China," *Energy Economics*, 28 (5-6):690-705

Hofman, Bert and Louis Kuijs. 2007. "Rebalancing China's Growth," Paper presented at the conference on China's Exchange Rate Policy held at the Peterson Institute for International Economics, October 19.

Hofman, Bert and Kelly Labar. 2007. "Structural Change and Energy Use: Evidence from China's Provinces," *China Working paper Series* No. 6.

Hu Jin-Li, and Shin-Chuan Wang. 2006. "Total-factor Energy Efficiency of Regions in China", *Energy Policy*, 34 (2006) 3206-3217

International Energy Agency. 2007. World Energy Outlook 2007. "China and India Insights".

Lewis, Joanna I., David G. Fridley, Jonathan E. Sinton, Jieming Lin, and Lawrence Berkeley National Laboratory. 2003. "Sectoral and Geographic Analysis of the Decline in China's National Energy Consumption in the Late 1990s," proceedings of the ACEEE Summer Study on energy efficiency in industry, Rye Brook, New York, 2003

Lin, Xiannuan, 1991. "Declining Energy Intensity in China's Industrial Sector," *The Journal of Energy and Development*, 16: 195-218

Lin, Xiannuan, and Karen R, Polenske. 1995. "Input-Output Anatomy of China's Energy-Use Changes from 1981-1987," *Economic Systems Research*, 7 (1): 67-84

Metcalf, Gilbert E. 2008. An Empirical Analysis of Energy Intensity and Its Determinants at the State Level, *The Energy Journal*, 29 (3): 1-26.

MIT. 2007. The Future of Coal: An Interdisciplinary MIT Study, Cambridge, MA: Massachusetts Institute of Technology, (March), http://web.mit.edu/coal/

Polenske, Karen R., ed. 2006. The Technology-Energy-Environment-Health (TEEH) Chain in China: A Case Study of Cokemaking, Dordrecht, The Netherlands: Springer

Shi, Xiaoyu. 2005. "Energy Prices and Energy Intensity in China: A Structural Decomposition Analysis and Econometric study," MCP thesis, Cambridge, MA: Department of Urban Studies and Planning, MIT.

Sinton, Jonathan E. 2001. "Accounting and Reliability of China's Energy Statistics, " *China Economic Review*, 12: 373-383

Sue Wing, Ian. 2008. "Explaining the Declining Energy Intensity of the U.S. Economy"nsity of the U.S. Economy," Cambridge, MA: MIT Joint Program on the Science and Policy of Global Change.

Zhang, ZhongXiang. 2003. "Why Did the Energy Intensity Fall in China's Industrial Sector in the 1990s? The Relative Importance of Structural Change and Intensity Change, " *Energy Economics*, 25 (6), (November): 625-638.