University of Neuchatel

Institute of Economic Research

IRENE, Working paper 12-02

Energy endowments, barriers to trade and industry location in Chinese provinces

*Irina Hotz**

** University of Neuchatel*

UNIVERSITÉ DE NEUCHÂTEL

Institut de recherches économiques

ENERGY ENDOWMENTS, BARRIERS TO TRADE AND INDUSTRY LOCATION IN CHINESE PROVINCES*

IRINA HOTZ†

First and actual Version : **10. August 2012**

Abstract

This study analyzes the determinants of manufacturing activity across Chinese provinces with emphasis on the important role of energy endowments on the location of industries. The data set consists of a panel of 28 Chinese provinces and 13 manufacturing industries for the years 1994,1997 and 1999-2009. A model of production location is estimated, including comparative advantage and economic geography dynamics. The effects of trade impediments and changes in economic policies on the distribution of economic activity are also considered. Results validate energy as a driving force of industry location. They further show that provincial protectionism poses a problem to market mechanisms, but indicate an increase of concentration of economic activity through time.

Key Words China, factor endowments, industry location, energy, externalities, protectionism

JEL Classification Numbers F18, P2, R14

* *I would like to thank Prof. Jean-Marie Grether and Nicole Mathys for their support and helpful comments.*
† *University of Neuchâtel, e-mail : irina.hotz@unine.ch*

1. INTRODUCTION

Manifold studies have tested the Heckscher-Ohlin (HO) model in empirical estimations of factor content of trade and production. The explanatory power of the model has shown to improve when the analysis is limited to a region or a country and when economic geography dynamics are included. These findings have conditioned my approach in this current study. The regional focus on China and more specifically on provincial energy endowments is motivated by China's growing international importance with regard to energy consumption and the country's ongoing economic reform process. China's energy consumption has surged since the beginning of the 2000's, recently becoming the world's largest energy consumer. Energy intensive industries contribute substantially to the overall consumption being the origin of half of the yearly 12% energy consumption increase.[1] Only slowly however has the central government begun implementing environmental policies, introducing for example special tariffs on the most energy intensive industries.[2] Moreover, the implementation of different policies relating to environmental concerns, have proven to be very weak. In parallel, firms and goods have started to move more freely across provinces and international trade has exponentially increased since the initiated economic reforms in 1978. Market mechanisms have therefore been able to determine to a larger extent firm's location choices. The fact that energy intensive firms optimize their costs and relocate to where energy is abundant may consequently have an impact on the efficiency of the environmental regulations established by the government.

Applying the economic geography augmented HO model approach, as proposed by many of the recent studies in the field, to Chinese provincial data, I test the following hypotheses:

H1: Comparative advantages and new economic geography mechanisms determine industry location across Chinese provinces. Consequently the activity of energy intensive industries will be higher in energy abundant regions.

H2: The comparative advantage effect is larger for immobile energy production factors that are subject to higher transport costs.

H3: The continuing transition from a planned economy to an open economy will show a net increase in the importance of factor endowments and new economic geography in determining industry location.

H4: As coastal regions have benefited from preferential treatment and open market policies since the beginning of the reform, the influence of economic geography and comparative advantage forces is larger in these provinces.

H5: Interprovincial barriers to trade exist between provinces, implying that the impact of market mechanisms on industry location is reduced.

This study contributes to the further understanding of the subject in two main aspects. To my knowledge this is the first paper to analyze the specific contribution of energy endowments on industry location in China, whereby it additionally includes measures of internal trade barriers. Also, I compare the effect of New Economic Geography (NEG) and comparative advantage over a time span of 15 years, separating the data into six different time

---

[1] *China Daily Newspaper, 07.09.2011. « http://www.chinadaily.com.cn/bizchina/2011-07/09/content_12869411.htm »*

[2] *The first major step towards environmental regulations was the establishment of the Environmental Action Plan of China 1990-2001.*

periods including data as recent as 2009, to evaluate and compare developments in economic activity as well as assess a part of the influence of governmental reforms and the changing economic context on industrial activity.

This study is structured as follows: The next section gives an overview of the framework literature of the model applied as well as a country-specific background that defined and conditioned my approach. I give the theoretical and methodological background in section 3, specifying the estimated equation. Section four exposes the data sources, variable definitions and a brief descriptive analysis. The hypotheses are tested econometrically and the results are evaluated in section 5. I conclude in section 6.

2. LITERATURE REVIEW

2.1 Heckscher-Ohlin (HO) Model Overview

The Heckscher-Ohlin model seeks to illustrate the role of factor endowments as trade determinants. In its simplest form, it is a basic general equilibrium model (2 countries, 2 factors, 2 goods) explaining patterns of trade between countries. According to the model, trade is explained by relative differences in factor endowments, which is in contrast to Ricardian theory, where trade is based on technological differences. Very restrictive assumptions such as perfect competition, identical technology, homothetic preferences, mobile goods and immobile production factors, full employment, zero transport costs and constant returns to scale are the framework of the model. Four main theorems are derived from this context:

1. According to the HO theorem, countries relatively abundant in a production factor will export the good that uses this factor intensively. i.e. if the factor intensity of a good rises, the production/trade shares of countries abundant in this endowment will increase.
2. The Rybzcynski theorem observes changes in the entities' factor endowments. An increase of a factor endowment (with the output price remaining constant) will lead to an increase of the production of the good that uses the factor intensively and a reduction of the output of the other good.
3. The Stolper Samuelson theorem relates output prices to factor prices. Moving from autarky to free trade, the output price of the exported good will rise, engendering a growth in production and therefore also the demand of the producing (abundant) factor. The increase in demand of the production factor implies an increase of its price and simultaneously a reduction of the price of the scarce factor as production shifts to the former.
4. The abovementioned eventually concludes in factor price equalization (FPE). On the grounds of free trade, output prices will equalize. The price of the abundant production factor in one country will rise together with the output price of the exported good. Hence, with free trade and zero transport costs and provided countries remain diversified, factor prices of the trading countries converge.

Jaroslav Vanek further extended the interpretation of the HO model (HOV model). International trade doesn't so much consist of trade in goods, but of an indirect trade of production factors embodied in the traded goods. With identical production technologies, goods embody the same amount of services of the productions factors, no matter where they are produced. The resulting HOV theorem therefore predicts the "factor content of trade". This allows going beyond the 2-2-2 setting of the general HO model, to a framework where the amount of goods is much larger than the amount of production factors, which would otherwise make the prediction of the corresponding production vector difficult under the basic model.

Over the past 50 years the use of the Heckscher-Ohlin model to explain regional and international patterns of trade and production has been lively debated. Especially the empirical implementation of the model has proven to be difficult and often failed to produce significant results in explaining industry location and factor content of trade. First empirical analyses applied the HO model to estimate the effect of factor intensities on trade, running cross-industry regressions for a single country, or on the effect of countries' factor endowments on trade, running cross-country regressions estimating industry by industry. Leontief (1953) cast first doubts on the empirical validity of the model by relating the capital embodied in exports and labour embodied in imports of the US, concluding that the US capital endowment was actually scarce relative to labour - the so called Leontief paradox. As Bowen, Leamer and Sveikauskas (1987) note, this previous work fails to analyse simultaneously all three elements of the HOV - trade, endowments and intensities and propose a more global approach including all elements. Their rank and sign test results on factors absorbed in consumption and factors absorbed in production however are sobering. Even when allowing for neutral technology differences across countries and measurement errors in factor contents and endowments, the measured parameters of many countries seem implausible. The authors conclude: "The HO model does poorly, but we do not have anything that does better. It is easy to find hypotheses that do as well or better in a statistical sense, but these alternatives yield economically unsatisfying parameter estimates." (p. 805)

Trefler (1995) in a thorough theoretical approach explains the failure of the model with differences of the factors' efficiency in the considered countries, reconciling the Heckscher-Ohlin model with the failure of FPE. The HO model predictions clearly overestimate trade in his analysis; he calls this divergence from actual trade "the mystery of the missing trade". Including the home bias effect - which describes individuals' preferences in consuming local products - and neutral technology differences, the predictions of the model are greatly improved.

Following the different empirical difficulties the HO model encountered, a growing number of researchers turned from a cross-country analysis to a within country or regional analysis, featuring countries and/or regions with similar levels of development (for example within the EU), benefiting from more homogeneous preferences and technology levels. Davis et al. (1997) apply the HOV to Japanese data arguing that running the analysis with regional data should improve the results, for the abovementioned reasons. Nevertheless this didn't improve their model as much as they would have expected, only to reproduce the so-called mystery of the missing trade. In consequence, only the abandoning of the FPE lead to acceptable predictions. Davis and Weinstein (2001) further explain that the HOV actually performs quite well when some of the assumptions are relaxed, allowing for Hicks-neutral technology differences, departures from identical homothetic preferences and accepting the break-down of the FPE and therefore recognizing that industry input is correlated with factor abundance. More precisely, this states a departure from considering the world economy as being an integrated equilibrium. Kim (1999) specifies that the regional approach additionally facilitates the analysis, as the regional factor endowments are in the same cone of diversification and trade barriers and tariffs don't have to be taken into account (as supposedly goods flow freely between regions within a country). In their research applying the HOV model on Spanish regions, Requeña, Castillo and Artal (2008) on the contrary find that the model performs rather poorly in conducted sign and rank tests even when considering world FPE breakdown, homothetic preferences and Hicks-neutral technological differences across the included regions. The authors in conclusion argue, as expansion of their study, in favour of the complementation of the HO model with new economic geography.

2.2 New Economic Geography (NEG)

One of the main drawbacks of the basic HO model is its failure to explain trade between countries with similar factor endowments as well as not accounting for intra-industry trade. Krugman (1980) confirms: "Neither the extensive trade among the industrial countries, nor the prevalence in this trade of two-way exchanges of differentiated products, make much sense in standard theory." He proposes a simple model being part of a new framework including economies of scale, imperfect competition and differentiated products, thereby explaining intra-industry trade and the importance of the domestic market for the export industry. Imperfect competition allows for monopoly power but entry results in zero monopoly profits. Therefore gains from trade (due to a greater goods variety) can arise between imperfectly competitive economies even though they have the same technology, factor endowments and homothetic tastes. He puts special focus on the home market effect demonstrating that increasing returns to scale will lead a country to export the good for which it has the relative larger domestic demand. Krugman (1991) further explores generalized external economies for the manufacturing industry as a whole, developing a model of geographical concentration by interacting transport costs and economies of scale, setting the context of the new economic geography framework. Technology spillovers, an often cited positive effect of industry agglomeration, is not the driving force of concentration in Krugman's model, but pecuniary effects of demand and/or supply linkages. Myrdal (1957, as cited in Krugman 1991) describes the phenomenon as circular causation, where industries, thanks to the mobility of manufacturing labour, tend to locate near the largest market potential, which is in itself result of industry agglomeration. Even though this might explain growth in industry concentration, it does not explain the origin of it. Krugman explains: "With lower transportation costs, a higher manufacturing share, or stronger economies of scale, circular causation sets in, and manufacturing will concentrate in whichever region gets a head start." In Krugman and Venables (1995, as cited in Redding 2009) the model is extended to the use of intermediate inputs in manufacturing allowing explaining concentration without assuming factor mobility. This is of importance when combined with the HO framework, which includes the restriction of factors being immobile.

The problem of the indeterminate origin of industry agglomeration was in the following subject of many studies such as Kim (1999), Ellison and Glaeser (1999) and Davis and Weinstein (1999). The researchers explain concentration in a first step through comparative advantage mechanisms with firms locating in areas with natural cost advantages. Ellison and Glaeser (1999) find their natural advantage interactions to actually explain 20 % of the variation in employment shares across the US. In a second step in Kim (1999) and in Davis and Weinstein (1998) the spillover effects are considered, refining the basic model of trade. In combing HOV and agglomeration effects, Smith (1999) confirms the importance of geographic scale economies in explaining missing trade shares in the HOV model. She focuses on scale economies that *augment* HO trade: efficiency gains through spillover effects that are a result of industry concentration, which in turn arose through relative abundance of certain factor endowments.

2.3 HO and NEG

Many studies have been published over the past decades testing the empirical viability of the HO model, relaxing one or the other restrictive assumption of the model, improving with more or less success its predictions. The NEG approach at the same time also found increasing support from economists. Davis and Weinstein (1999) argue that the downside of NEG is that it can't include input composition and demand structure and that it therefore should be combined with the factor endowments analysis. Hence, studies increasingly include NEG considerations in addition to factor endowments in their trade models.

Midelfart-Knarvik, Overman and Venables (2001, from now on referred to as MKOV) follow a broad approach with a panel of industries and 14 EU countries letting both country and industry characteristics interact. In contrast to the abovementioned study by Davis and Weinstein (1999), they explicitly include NEG and factor endowments effects on the same level, to analyse how these forces simultaneously determine firms' location choices. The authors do not include imperfect competition in their theoretical model, as this could lead to a multiplicity of equilibria, making a unique determination of location impossible. They observe industry location in EU countries during the time span of 1980-1997 through a set of interaction variables, with growing economic integration positively affecting NEG and factor endowment dynamics.

Romalis (2004) uses a simple extendable two-factor model, combining Krugman's (1980) model of monopolistic competition and transport costs with the HO framework. He states a deviation from FPE caused by transport costs, which implies that locally abundant factors are relatively cheap. He then explicitly makes the link between factor prices and a country's factor abundance. Therefore, relative industry output prices, which are built upon the interaction of industry and country characteristics (factor intensity and factor price) define a country's share in total production.

Considering the various findings of empirical and theoretical research on trade and industry activity with respect to varying technology matrices, economies of scale and market externalities, I will hence follow the recent literature and combine the HO and NEG models, restraining the study to a within country analysis of China.

2.4 Energy

In the context of environmental awareness and failing green house gas reduction policies, research has only slowly been enlarging the scope of HO models by focusing on the specific contribution of energy endowments on industry structure and location. Hillman and Bullard (1978) address the Leontief paradox by including energy as a production factor and not only intermediate input. Using energy-capital as composite factor, the authors argue that the US is actually abundant in labour and capital and scarce in energy, which "distorts" the basic capital-labour relation. Much later, in the framework of combined comparative advantage and NEG models, Ellison and Glaeser (1999) include electricity, natural gas and coal interactions amongst others, in the study already mentioned above. The authors find that the effects of the energy interactions are quite large, explaining a substantial part of the total comparative advantage. Gerlagh and Mathys (2010) assess the effect of energy endowments on trade using a panel on 14 OECD countries. They construct energy abundance from the country-sector input data dividing energy per labour use into a country specific and industry-specific component and find that energy abundant countries have higher activity in energy-intensive industries, which is accentuated in trade exposed sectors. Michielsen (2011) performs a within country analysis on the location of industry in the US, regressing value added on 6 different energy carrier interactions (electricity intensity with coal, natural gas and hydropower abundance and fuel intensity with oil and natural gas abundance) in addition to the main factor endowments, skill and capital, and NEG interactions (market potential with economies of scale and intermediate input intensity). The author shows that the HO model holds also with respect to energy carriers and that energy endowments are found to play a significant role in determining US industry activity. Energy proving to determine industry location, I similarly decide to emphasize the role of energy endowments in this study in extension to the basic HO model, being of special interest as China's overall energy consumption surges.

2.5 China

Generally, literature analyzing industry location in China's provinces is not yet very developed. This could be explained with important data constraints as well as data viability. Even though a lot of progress has been made, especially regarding earlier periods and those provided by rural provinces, one cannot be entirely confident of the viability of the statistical data furnished.

Also, the benefits of conducting a regional study, as already mentioned above, are not necessarily realized in China which shows a large heterogeneity among regions with respect to their levels of development and technology. I will hence briefly highlight in the following some of the characteristics of the economic and political context, conditioning and potentially impeding regional specialization mechanisms.

After three decades of centralized planned economy, China generally reported a very weak degree of specialization across provinces, due to the central government's efforts to generate a widespread supply security in case of foreign invasion. Young (2000) argues: "With material supplies only ensured when one actually produced them oneself, and with the central regime actively encouraging and funding the local development of industries, each province, county, city and locality tried to develop its own duplicate set of industries." (p. 7) Herein lies the particular interest of a regional study on China; The opening reforms initiated in 1978 and which are still under development, have allowed market mechanisms to exercise a growing influence on economic activity. Nevertheless, the country continues to maintain high barriers of trade on the international as well as the regional level, also impeding domestic trade. Protectionism not only limits geographic concentration and the benefits of specialization but also the free flow of goods across provinces, thus interfering with comparative advantage mechanisms.

Economic reforms launched in 1978 reduced central control over factor allocation with the introduction of fiscal decentralization. At the same time the grip on price determination was loosened. According to Young (2000) this principally lead to a transfer of power to rent-seeking provincial governments, eventually engendering the fragmentation of the domestic markets through barriers and subsidies (Young mentions interregional trade "wars"), with actual diversion of production, moving away from patterns of comparative advantage. Poncet (2003) describes: "The magnitude of Chinese provinces' border effects turns out to be closer to that existing between independent sovereign countries than to that measured inside individual countries such as the US of Canada." (p. 16) She concludes that while barriers to international trade decreased between 1987 and 1997, interregional border effects almost doubled during the same time span. Coastal provinces essentially being more open to international trade as well as interprovincial trade since the beginning of the economic reforms, the author finds on the other hand a substitution effect in less open, central provinces in favour of international products.

Bai, Du and Tao (2003) on the other hand argue that the overall time trend from 1980-1997 indicates an increase of regional specialization. The authors add “tax+profits” of each industry as an indirect measure of protectionism, as provincial governments tend to protect industries with a high “tax+profit” level and the variable should therefore have a negative effect on regional specialization. They mention that high percentages (although decreasing through time) of state-owned enterprises (SOEs) are an additional element affecting negatively the specialization of production. In 1997, the average share of SOEs over their industry sample was at 40.92%, with petrol processing, smelting and chemical industries revealing high SOE shares.

Most closely related to this study, Batisse and Poncet (2004) apply the economic geography augmented HO model to China, additionally accounting for regional protectionism. They criticize the indirect measure of protectionism used by Bai et al. (2003) and compute a variable based on a gravity equation using domestic trade flows (imports and exports to and from “the rest of China”). The authors show that both comparative advantage and economic geography play a significant role in the location of the industry across provinces for the period 1992/1997 but that protectionist measures have increased at the same time, especially in less “outward-looking”, central provinces. Amiti and Javorcik (2008) assess internal impediments to trade in China by evaluating the effect of provincial internal and external components of demand and supply linkages (i.e. market potential and supplier access) on the entry of foreign firms. They find that the overall supplier access and market potential play an important role in determining entry of foreign industry but that the internal dominate clearly over the external linkages, indicating the existence of barriers to trade.

In this sense I further extend the HO and NEG framework with focus on energy considerations, to a model that takes into account the important trade impediments within China, which interfere with the basic theoretical conditions of the HO model. Even though the work mentioned above demonstrates increasing internal protectionist tendencies between provinces, I also propose to enlarge the analysis with a dynamic component, assessing potential changes in HO and NEG influences through time, induced by gradual opening economic reforms of the central government.

Finally, turning to the energy framework in China, Rosen and Houser (2007) give a thorough descriptive overview of the current political and economical situation: In 2005, China’s total energy needs were covered by 69.6% with coal (over 80% of the electricity is generated from coal). The Chinese energy market is strongly controlled and regulated by the central government. According to their paper, energy prices for producers have been to a large extent liberalized while downstream prices remain fixed, leading to general supply expansion rather than demand management. They note that, contrary to central government efforts, investment and production in energy intensive industries such as in smelting, chemical and paper industries have been on the rise for the last decade. Even though the energy efficiency has been rising, the authors argue that the structural shift away from the heavy industry that was taking place until the late 1990’s has seen reversal in the meantime. These industries often show a large share of SOE’s (which don’t have to distribute their profits) whose profits after tax ratios have been increasing, leading local governments to promote them locally to raise government revenue. In consequence these heavy industries have had the tendency to further fragmentation rather than specialization. In 2006, the heavy industry accounted for 54% of China’s total energy consumption.

In 2004 a national differential power pricing policy ran into effect, introducing specific tariffs with the objective to discourage further investment in energy intensive industries. Nevertheless, even accounting for these tariffs, the energy prices in China are, according to Lin and Liu (2010), lower than both the social and private equilibrium. Additionally, the provinces have significant possibilities to intervene and counterturn the policies of the central government. As Lin and Liu argue, the heavy industry in some remote provinces is a major source of taxation and is an important consumer of other industrial products. Following the 2008 crisis, many regional governments lowered the tariffs for energy-intensive industries and gave other incentives through tax reductions and subsidies. These facts indicate that the resource prices do not just reflect the abundance of energy endowments but are also, in the Chinese case to an important part, the result of regional energy and fiscal policies.

In summary, based on the literature discussed above, I will implement an NEG augmented HO model, extending the analysis to a specific focus on the importance of energy and trade barriers as well as the evolution of market mechanisms through time, by splitting the data into 6 time periods between 1994-2009.

3. METHODOLOGY

The methodology and theory is based on MKOV and follows the approach of Michielsen (2011) and Batisse and Poncet (2004). The industry output is thereby explained by the interactions of provincial and sectoral characteristics, i.e. resource abundance and sectoral factor intensity. I explicitly include both a series of factor endowment interactions and economic geography effects to determine industry localization. Augmenting the basic comparative advantage model with NEG mechanisms such as economies of scale and transportation costs have, as already mentioned, generally shown to improve significantly the basic HO model, showing that industries tend to locate near demand and supply to reduce transportation costs.

The equations and explications described below follow the model given in Batisse and Poncet (2004) described as a canonical economic geography model[3], where provincial endowments and industrial intensities interact under inclusion of transport costs and barriers to trade (goods are tradable but subject to transport costs and impediments to trade). The model consists of R provinces denoted p and m and I industries denoted i. The transport costs $\boldsymbol{t^i_{pm}}$ are adopted in the form of an iceberg mark-up on shipping, with the costs proxied by the distance between p and m. A number of varieties of differentiated goods are produced by each industry, $\boldsymbol{n^i_p}$, and assumed to be determined exogenously. MKOV explain: "More generally the presence of product differentiation means that factor prices are not independent of endowments, that there is trade in all goods (despite trade costs), and that the structure of production is determinate (even if there are more industries than factors)." Batisse and Poncet (2004) additionally assume constant ad valorem barriers $\boldsymbol{B^i_{pm}}$ for all inter-provincial trade.
The different varieties are aggregated on the basis of a CES function; constituting the demand side/ price index for industry i in province m:

$$\boldsymbol{G^i_m = \left[\sum_{p=1}^{R} n^i_p \left(p^i_p t^i_{pm} B^i_{pm}\right)^{(1-\sigma)}\right]^{1/(1-\sigma)}} \qquad (1)$$

[3] Defined as core-periphery model in Krugman (1991a, as cited in Redding, 2009) with the tension of the two agglomeration forces (home-market and price index effect) and the two dispersion forces (market crowding effect and the immobility of agricultural labour).

$\boldsymbol{p_p^i}$ is the f.o.b price of the goods produced in p. The elasticity of substitution σ between varieties is assumed identical for all industries.

Using Shepard's lemma on the price index, the value of sales, $\boldsymbol{x_{pm}^i}$, can be derived, with $\boldsymbol{E_m^i}$ defined as total expenditure on the goods of industry i in province m:

$$\boldsymbol{x_{pm}^i = (p_p^i)^{-\sigma}(B_{pm}^i t_{pm}^i)^{1-\sigma} E_m^i (G_m^i)^{\sigma-1}} \tag{2}$$

When aggregating over all varieties and all provinces, the following equation follows for the total output value $\boldsymbol{y_p^i}$ of industry i in province p:

$$\boldsymbol{y_p^i = n_p^i p_p^i \sum_m x_{pm}^i = n_p^i p_p^i x_p^i = n_p^i (p_p^i)^{1-\sigma} \sum_{m=1}^{R} (B_{pm}^i)^{1-\sigma} (t_{pm}^i)^{1-\sigma} E_m^i (G_m^i)^{\sigma-1}} \tag{3}$$

The prices equal unit cost on the supply side, with input prices in province p denoted by the vector $\boldsymbol{\nu_p}=[\boldsymbol{\omega_{p,}q_p}]$:

$$\boldsymbol{p_p^i = \theta^i c(\nu_p : \phi^i)} \tag{4}$$

$\boldsymbol{\omega_p}$ is the price of primary factors and $\boldsymbol{q_p}$ of the composite intermediate good. $\boldsymbol{\phi^i}$ refers to the costs of industry i with regard to primary and intermediary production factors.

The model assumes perfect competition, which, according to MKOV, is necessary to have a precise link between the theory and the empirics[4] and implies that the producer mark-up $\boldsymbol{\theta^i}$ is set to 1. The authors also argue it being convenient to take the total value of production as a numeraire, so that, $\Sigma_p \Sigma_i \boldsymbol{y_p^i} = 1$. It follows that instead of varieties being determined endogenously by free entry and exit in the competitive monopolistic framework, the number of varieties produced in each province are exogenously fixed in proportion to industry and province size up to an error term $\boldsymbol{\epsilon_p^i}$:

$$\boldsymbol{n_p^i = \sum_i y_p^i \sum_p y_p^i exp[\epsilon_p^i] = s_p s^i exp[\epsilon_p^i]} \tag{5}$$

Where s_p is the share of province p and s_i the share industry i in the country's total output.

When substituting (4) and (5) into equation (3) one obtains:

$$\boldsymbol{\frac{y_p^i}{s_p s^i} = c([\omega_p, q_p] : \phi^i)^{1-\sigma} \sum_{m=1}^{R} (B_{pm}^i)^{1-\sigma} (t_{pm}^i)^{1-\sigma} E_m^i (G_m^i)^{\sigma-1} exp(\epsilon_p^i)} \tag{6}$$

The explained variable is therefore expressed as a double relative measure of industry output with respect to industry and province size. The first term on the right hand side of (6) is "an aggregate version of an industry-specific technology efficiency parameter whose determinants at the regional level are brought back to factor endowments" (Batisse and Poncet, 2004)[5]. More

---

[4] The authors argue that imperfect competition could imply that there are multiple equilibria and therefore impede a unique determination of industrial location through industry and country characteristics.

[5] After the log-linearization around reference points, the terms, according to MKOV, refer to the elasticity of industry costs with respect to industry prices and the elasticity of country market potential with respect to industry characteristics.

precisely, MKOV specify that $\boldsymbol{\omega_p}$ relates to factor endowments, rather than factor prices for primary factors, as the latter are determined endogenously.

The sum captures the demand effects, which MKOV refer to as market potential of industry i in province p. According to Batisse and Poncet (2004) the market potential corresponds to the term developed by Harris (1954), with demand arising from local and external consumers. The market potential is decomposed into two parts: regional protectionism and the mentioned traditional market potential.

In the following, the economic implementation and estimation of the model will be done according to this log-linearized equation on the relative output:

$$ln \frac{y_{p,t}^{i}}{s_{p,t} * s_{t}^{i}} = \alpha_{p,t} + \beta_{i,t} + \gamma_1 \theta_{skill_{p,t}} * \pi_{skill^i} + \gamma_2 \theta_{capital_{p,t}} * \pi_{capital_{p,t}^{i}} + \gamma_3 \theta_{energy_{p,t}} * \pi_{energy^{i,t}} + \delta_{mpi}\left(mpi_p^i + \delta_{mpo} mpo_p^i\right) + \delta_{sai}\left(sai_p^i + \delta_{sao} sao_p^i\right) + \varepsilon_{p,t}^{i} \quad (7)$$

where total output per industry i in province p, $y_{p,t}^{i}$, is expressed relative to the share of province p in national output, s_p and the share of industry i in national output, s^i. This double relative measure follows the approach of MKOV and Batisse and Poncet (2004) and allows to additionally account for industry and province size.

Equation Nomenclature

y	Industry Output	γ	HO interaction coefficients
p	Province subscript	θ	Provincial factor endowments
i	Industry subscript	π	Industry factor intensities
t	Time subscript	δ	Economic geography variable coefficient

The dependent variable is explained by the skill, capital and energy interaction terms of comparative advantage nature and market potential (mpi and mpo)[6] and supplier access (sai and sao)[5] of economic geography nature. $\alpha_{p,t}$ and $\beta_{i,t}$ are the province-year and industry-year fixed effects that control for heterogeneity of the industries and provinces in each time period. Additionally I add a dummy variable for coastal provinces. The detailed definition and construction of the variables can be found in table 1 p. 13 and in appendix III.

The HO effects include, apart from skill and capital, four different energy carrier interactions, comprising the different combinations coal, petrol and natural gas on a disaggregate level and also a composite total energy interaction to assess the importance of energy endowments on industry location. Pursuant to the predictions of the HO model, the coefficients γ are expected to be positive and larger for relatively immobile factors.

In contrast to MKOV I include a capital interaction in my empirical estimation. The authors argue that as capital is highly mobile, it does not comply with the condition of factor immobility of the HO model, is therefore not expected to have an effect on industry location and should be excluded. I follow Batisse and Poncet (2004), given that the Chinese capital market is far from being liberalized, by arguing that capital cannot be considered completely mobile and should therefore show an effect on production location.

---

[6] *«i » refers to internal and « o » to outer or external.*

The economic geography variables measure the proximity to demand and to supply in intermediate inputs. The existence of forward linkages, here defined as supplier access, and backward linkages, here defined as market potential, will result in positive δ coefficients. [7]

The analysis within one country brings advantages regarding the empirical implementation, as tariffs, different technological matrices and non-identical homothetic preferences do not pose a large problem compared to cross-country analysis. As seen in the literature review, China has large interior barriers to trade and exhibits quite a heterogeneous development across provinces. I therefore split up the market potential as proposed by Batisse and Poncet and include a protectionist measure. The division of the market potential into an internal mpi and external mpo effect allows accounting for protectionism by evaluating their relative contribution to total market potential (and to supplier access equivalently). The construction and interpretation of the variables are given in appendix III and paragraph "Economic Geography variables" on p. 14.

4. DATA

The China Industrial Economic Yearbooks include detailed information for all 31 provinces in 25 manufacturing industries, some of which I aggregated to correspond to the equivalent input-output table definitions. After aggregation and exclusion of mining and extraction industries, whose location is principally defined by its resources, I remain with the data of 13 industries for the years 1994[8], 1997 and 1999-2009[9]. Throughout this study I use two different data sets: On the one hand, my principal set of data is a balanced panel for the period 1999-2009 and to which I will refer to as "panel data set". Furthermore, given the incomplete panel regarding the period between 1994-1999, I construct a second data set to which I'll refer to as "cross-section data set". It includes 6 periods of time: 1994, 1997, 2000, 2003, 2006 and 2009, whereby the periods between 2000-2009 have been constructed using the panel data set and constitute a three year average (with exception of 2009 which is the average of 2008 and 2009). This has the advantage of enabling the analysis of the estimated model over a slightly larger time period as wells as accounting for small fluctuations and measurement errors.

After a first analysis of the data I decide to exclude petroleum-processing, which figures as outlier with strong leverage and an industry that is often located near extraction facilities. Of the 31 provinces I exclude Tibet in a first step, for its specific political situation most likely strongly affecting the production structure and also because the statistical information is mostly incomplete. In a second step I decide to further exclude Xinjiang and Hainan, because of their geographic location, Xianjiang's economy's reliance on petrol and Hainan's general low industrial activity. This implies the reduction of outliers in the data set, thereby showing a potential weakness in the capability of the model to explain irregularities, which does however not correspond to the scope of this paper and therefore not further treated.[10] Even though the

---

[7] *Forward linkages refer to the situation where the increase of intermediate goods leads to the growth of the industries that us them. Backward linkages are when an increase of final consumption leads to the growth of its suppliers.*

[8] *The centrally controlled province of Chongqing was split from Sichuan in 1997. I deducted the 1994 values from the 1997 proportions between the two provinces.*

[9] *I linearly interpolated the values for 2004 for which no industry information was available.*

[10] *When including the two provinces, the coefficients are strongly reduced, with only energy and NEG variables entering significantly and positively the basic specification, nevertheless confirming my main hypotheses.*

heterogeneity across Chinese provinces is high, which would favour the use of nuanced technology matrices, I had to calculate nationwide industry intensities because of the unavailability of the disaggregated data. Province fixed effects should however account for eventual technology differences between the provinces.

4.1 Interaction Variables

TABLE 1: *Variable Definition**

Variable		Definition	Dimension
Explained Variable	$ln \frac{y^i_{p,t}}{s_{p,t} * s^i_t}$	Log of relative output measure: output per industry i in province p relative to the share of industry (s^i) and the share of province p (s_p) in total national output	province-industry-year
Abundance (ab) Variables (Country characteristics) θ	Capital ab	log of total capital stock per capita in province p	province-year
	Skill ab	log of population with university degree	province-year
	Energy ab	log of total energy production per capita (log of energy reserves per capita)	province-year (province)
	Coal ab	log of coal production per capita (log of coal reserves per capita)	province-year (province)
	Petrol ab	log of petrol production per capita	province-year
	Natural gas ab	log of natural gas production per capita	province-year
Intensity (int) variables (Industry characteristics) π	Capital int	log of industry capital stock per worker	province-industry-year
	Skill int	log of average wage rate	industry
	Energy Int	log of total energy consumption per worker	industry-year
	Electricity int	log of electricity consumption per worker	industry-year
	Petrol int	log of petrol consumption per worker	industry-year
	Natural gas int	log of natural gas consumption per worker	industry-year
NEG Variables**	mpo & mpi	log of market potential at the industry-province level weighted by distance -mpo: log of market potential from all other provinces -mpi: log of market potential within province p	province-industry
	sao & sai	log of supplier access at the industry-province level weighted by distance -sao: log of supplier access from all other provinces -sai: log of supplier access within province p	province-industry

* *Details on variable source and construction are in appendix III*

** *Details can be found on p. 14 "NEG variables" and in appendix III*

Linear intensity and abundance variables have, after taking the logs, additionally been standardized per year over province and industry, facilitating the interpretation of the coefficients of the resulting interaction terms. Also, standardizing separately per year allows accounting for a general time trend (see growth rates in appendix IV, table A5) and gives a comparable value of relative abundance and relative intensity in each time period.

As measure of energy abundance, instead of using natural energy reserves in each province as proposed in Michielsen (2011), I opt for using energy production per capita as this gives a better indication of actual availability of the factors and includes the effects of eventual distorting policies on abundance. Gerlagh and Mathys (2010) argue that the effective price of energy is relevant for firms, which is the result of both endowments and energy and fiscal policies. Because of the lack of provincial-sectoral energy use data, I rely on provincial energy production as proxy for abundance. I interact electricity intensity, not with electricity abundance - as this is more likely endogenous to the provincial industry structure and population[11]- but with coal production per capita, as 80% of the electricity in China is generated in coal power stations. Nevertheless a potential endogeneity problem due to reverse causality arises for all natural resource abundance measures, because abundance could actually be the result and not the cause of industry structure, as also suggested by Gerlagh and Mathys (2010). In a robustness test I use self-sufficiency (indigenous provincial energy production divided by final consumption on the provincial level in 2006) and the 2005 natural reserves level as instrumental variables to control that the causality doesn't run from industry structure to production (ex. energy intensive industry structure incites provincial government to subsidize the production of the energy source in question). [12]

4.2 Economic Geography Variables

The overall market potential is derived from Harris' (1954, as cited in Batisse and Poncet, 2004) original formulation where the final demand of industry i is expressed relative to distance. The supplier access, also used by Batisse and Poncet (2004), is the one originally proposed by Davis and Weinstein (2001). Following Amiti and Javorcik (2008) and their analysis of economic geography determinants of foreign industry location in China, I divide the backward linkage MP and forward linkage SA in an internal and external component respectively representing supplier access and market potential in the province itself and in all other provinces:

$$lnMP_p^i = ln(MPI_p^i + \beta_{MAO} MPO_p^i) \quad (8)$$

$$lnSA_p^i = \ln(SAI_p^i + \beta_{SAO} SAO_p^i) \quad (9)$$

This division allows controlling for potential provincial barriers that may affect the results of the HO model by preventing free flow of goods inside the country. As both the internal and external measures have been adjusted for distance, they indicate the relative importance of the two effects. β_{MAO}<1 suggests that there are barriers to trade or differences in transport infrastructure. The authors use a nonlinear model to estimate β_{SA} and β_{SAO} and β_{MP} and β_{MPO}. Instead I choose to include all four variables in the linear regression and perform a nonlinear Wald test on $\beta_{MPO}/\beta_{MPI} < 1$ to verify the existence of barriers to trade. The detailed construction of these variables can be found in Appendix III. The economic geography

[11] Electricity production per capita is highly correlated with regional GDP per capita, implying that production is potentially linked to generally higher consumption in developed provinces.

[12] The log of the average coal production per capita has a correlation of 0.7246 and 0.7802 with coal reserves and coal sufficiency respectively.

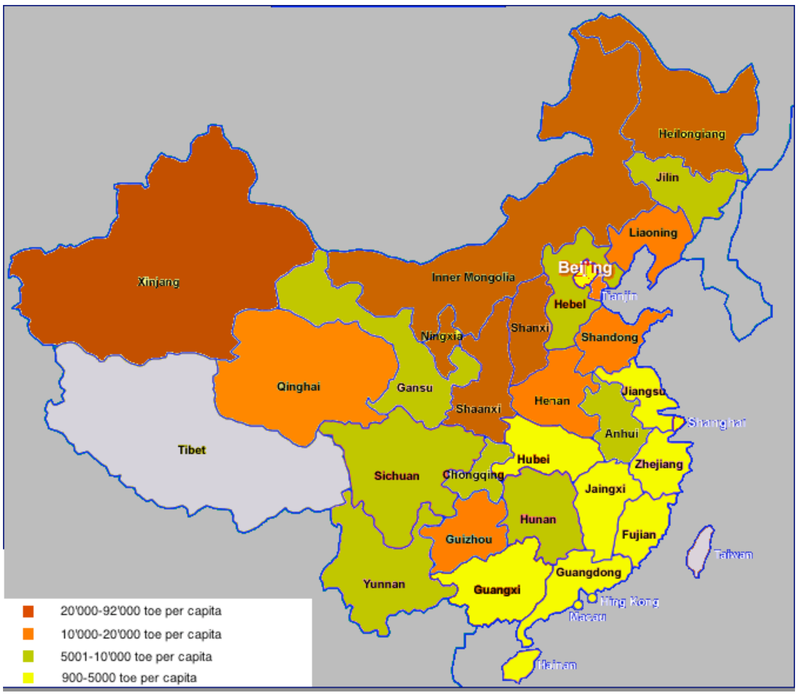
components are highly correlated with each other and will therefore require special attention during the empirical estimation.

4.3 Descriptive Analysis

Regarding energy abundance we can see in Figure 1 that energy production is strongly concentrated in the northern provinces. Xinjiang, Inner Mongolia, Shanxi, Heilongjiang and Ningxia are origin of 60% (with Shanxi province alone accounting for 20%) of the whole country's total energy production. The coastal, relatively capital abundant provinces, in comparison, are energy scarce and one can actually observe a spatial grouping with respect to energy abundance.

Regarding factor intensities on the industry level, table 2 shows what one might expect: the paper, chemical, smelting and non-metallic mineral industries are the most intensive in energy use for all energy carriers. On the right hand side of figure 2, the relation between the different average factor intensities are graphed. We can see that there is a very clear positive relation between energy and capital intensity, the variation in energy being almost double the variation of capital. The left-hand side of graphs 2a and 2b show, similar to the findings for intensity, that the distribution of energy production i.e. energy abundance is also very uneven, the variance being two times bigger than for skill or capital. Capital and skill abundance are hence more evenly distributed across provinces, although we can observe a strong and positive link between capital and skill endowments. This larger variance of energy abundance and energy intensity between provinces and industries suggests the effect of the interaction to be more important with respect to capital and skill.

FIGURE 1: *Energy abundance distribution (log of average energy production per capita in toe over sample period)*[13]



[13] *The definition and unit specification of energy can be found in appendix III.*

TABLE 2*: Average factor intensities per industry per labour**

	Capital	Skill	Electricity	Coal	Petrol	Natural Gas	Total Energy
Measuring instruments	0.883	12’720	0.508	0.378	0.191	0.096	8.727
Electrical machinery	1.171	12’405	0.608	0.611	0.222	0.398	11.723
Communication equipment	1.687	17’636	0.731	0.303	0.251	1.437	11.668
Machinery	1.009	10’537	0.778	1.454	0.245	0.721	19.478
Transport equipment	1.521	14’409	0.869	2.230	0.287	1.210	26.455
Food (manufacture/process.)	1.841	9’842	1.026	6.931	0.381	0.324	60.858
Textile	0.977	7’268	1.264	3.517	0.230	0.170	37.248
Metal products	1.033	10’075	1.923	1.314	0.382	0.459	28.888
Paper	2.005	9’766	2.722	21.015	0.485	0.352	174.17
Non-metallic mineral products	1.553	8’123	2.985	36.75	1.604	3.845	300.42
Chemical industry	2.453	11’275	4.327	23.691	1.293	0.243	213.84
Petro-processing	7.353	17’357	4.744	232.11	8.263	27.042	1767.3
Smelting	3.311	15’032	8.761	47.571	1.281	2.407	418.34

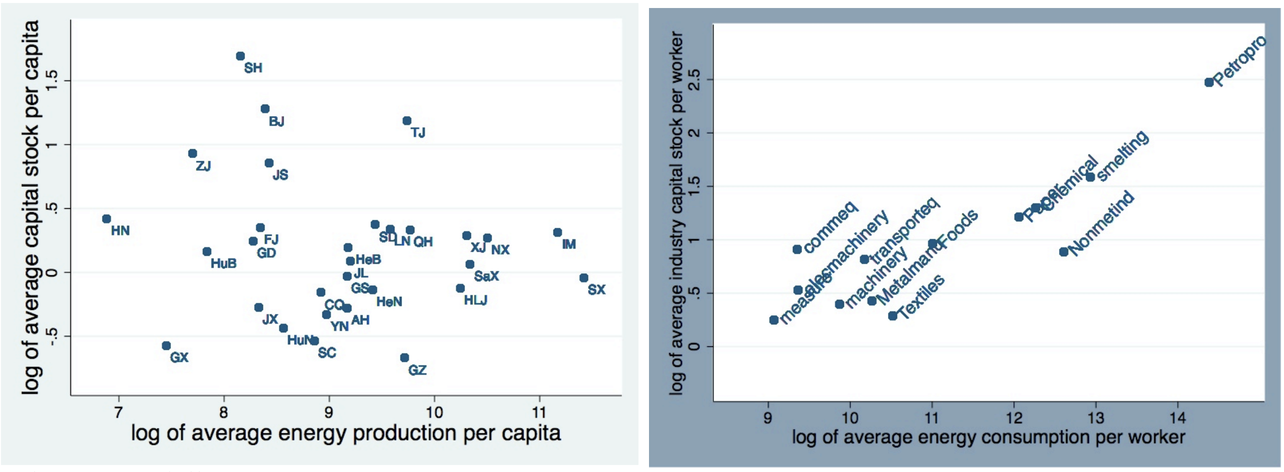
**For construction and unit specification of the variables see p. 13, table 1 and appendix III.*
Sorted from least to most electricity-intensive industry. Natural gas per labour is multiplied by 100. Energy is expressed in toe.

Generally, in the HO framework, factor intensities and endowments are considered constant through time and the issue is therefore often not further addressed in the statistical analysis. Table A5 in appendix IV however shows that for the time period between 1999 and 2009 the average abundance across provinces and the average intensity across industries are actually subject to important annual growth in China (with exception of petroleum). This is somewhat contrary to other findings, for example for energy intensity in Rosen and Houser (2007), where the energy consumption is expressed relative to output and not labour, stating a decrease rather than increase in energy intensity. These diverging statements might be explained by the construction of the here used industry intensities calculated as “per-labour ratios”. In this sense, labour employment in the industrial sector might be for example subject to some overall structural changes stating a decreasing tendency with respect to capital or energy. Also, as I measure energy abundance with production, the growth in abundance is actually quite in line with the overall growth of manufacturing and the surge in energy demand.

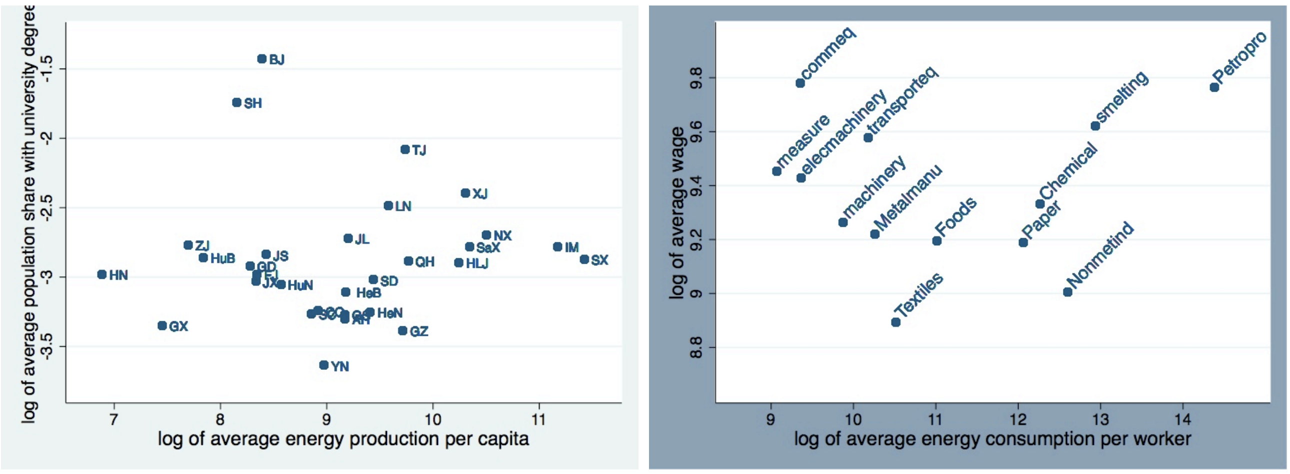
FIGURE 2*: Pairwise comparison of factor endowments and intensities*
(average over sample period)[14]

ENDOWMENTS *INTENSITIES*

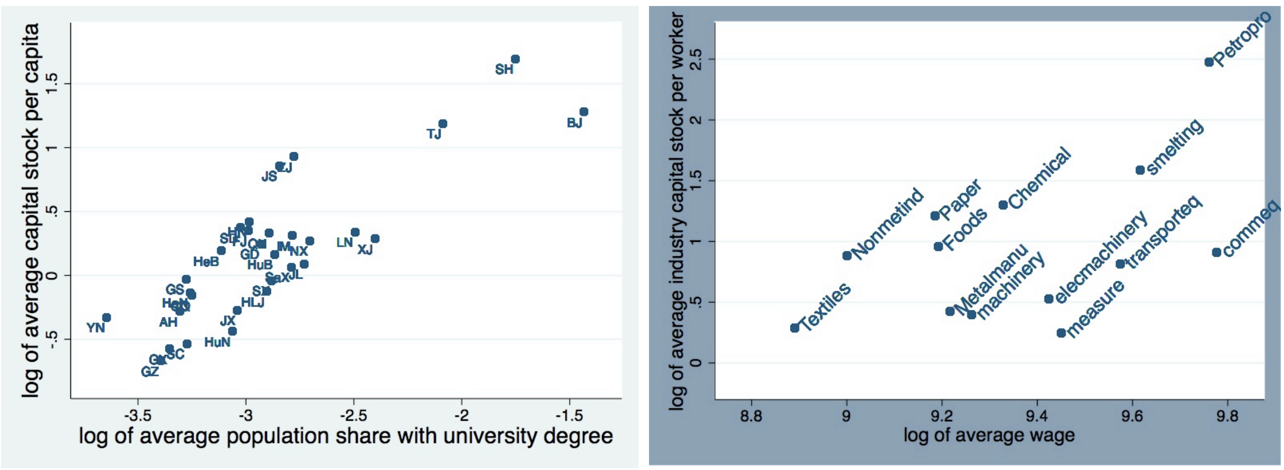
2a. Energy-capital



2b. Energy-skill



2c. Capital-skill



---

[14] *Table of abbreviations can be found in appendix IV*

A closer look on income and development across provinces reveals an astonishingly clear division between provinces on the coast and inland. Figure 3 exhibits that out of the 10 provinces with the highest GDP per capita, 9 are situated on the coast.[15] This is not only due to the locational advantage, but more specifically related to economic policies of the Chinese government, which biased the allocation of domestic and foreign capital towards the largest cities on the east coast. Also, the establishment of Special Economic Zones, starting in the 1970s, was uniquely concentrated in coastal areas and only slowly extended inwards at the

FIGURE 3*: GDP per capita in 2009 (GDP in 100 million RMB, population in 10'000)*

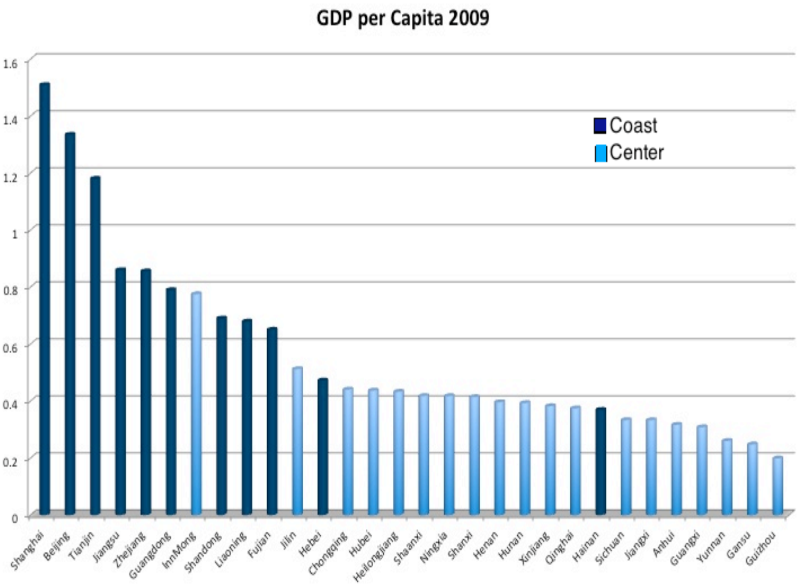
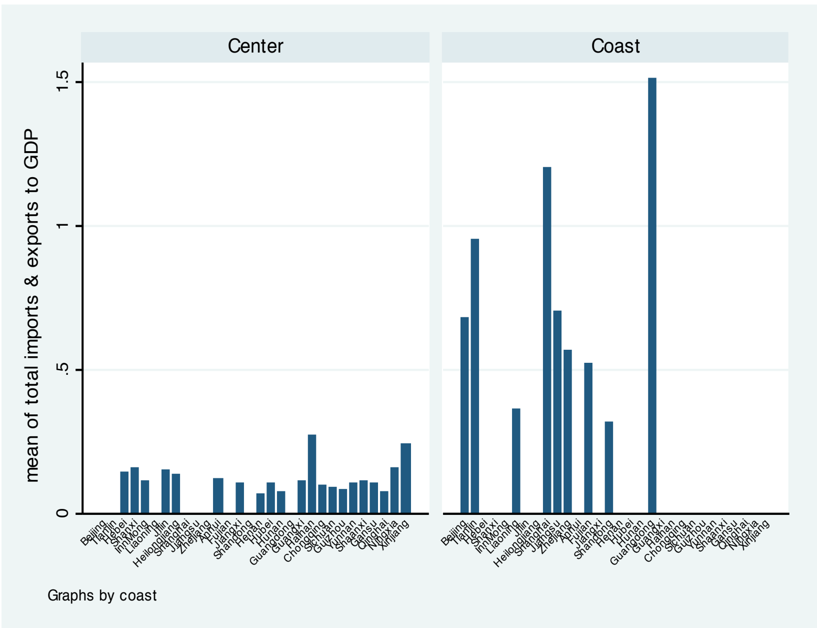


Figure 4 : Comparison of gross imports and exports relative to GDP in center and coastal provinces (average over sample period, in %)



[15] *See table A4 in appendix IV for the list of coastal provinces according to my definition.*

beginning of the 1990s. These economic zones are concentrated areas (often cities and their surroundings), where international trade and foreign investment is facilitated and economic activity is subject to more market-oriented policies (Yueng, Lee and Kee, 2009). Consequently, figure 4 shows the enormous difference between coast and center relative to internationally oriented imports and exports. These circumstances suggest a generally more liberalized economy in the coastal provinces enabling to a larger degree industry specialization, given the "free reign" of market forces even though the origin of concentration has been determined exogenously.

5. RESULTS

5.1 Framework

For my basic estimations of which the results are listed in table 3, I run a two-way fixed-effects regression with industry-year and province-year fixed effects (number of dummies in parenthesis), controlling for heterogeneity of industries and provinces for each period in time. Following Michielsen (2011) and Batisse and Poncet (2004), in a first step I assume that relative abundance and intensity remain constant through time and therefore do not include province-industry fixed effects. Wald tests indicate that the included fixed effects are jointly significantly different from zero. I have auto-correlation issues and use clustered robust standard errors, allowing for autocorrelation in each province-industry group. As the formulation and interpretation of the dependent variable can be quite difficult to assess, I ran a regression on the basic specification in column 1 using the log of industry output instead of the calculated relative abundancy measure. The results are robust, exposing a larger R^2 (0.89). In table 3 regressions 1-3, the basic results are shown for total energy, and in columns 4 and 5 the disaggregated interaction variables; coal, petrol, natural gas. The explanatory power of the model is quite good, with an R^2 of at least 65% across columns. I perform a robustness check using natural resource reserves per capita as alternative measure for abundance: in column 6, I replace the disaggregated production, in 7 only the coal and in 8 the total energy interaction variables with their respective natural resource reserves value. The estimated coefficients are robust, even if slightly larger when using natural reserves as abundance measure. However, it is important to point out that the natural reserve variables might also be subject to certain endogeneity bias as they are constructed from the data of the ensured reserves in 2005 and don't account for changes in the reserves due to new discoveries made.[16] This could explain the larger elasticity of energy reserves in comparison to production which is rather surprising, as one would expect the endogenous factors affecting production to have the tendency to accentuate the positive effects of energy endowments on relative output. Even though natural reserves are potentially endogenous, due to the lack of valid alternatives, I am constrained to rely on natural reserves as control and as the instrumental variable of an additional IV regression performed to which the results are shown in table A3 in appendix I. Furthermore, as there might be contemporaneous

[16] *The Chinese government has been encouraging the exploration of reserves and Chinese newspaper articles announcing new discoveries of coal deposits of 20 billion tons have been published almost annually.*

correlation issues i.e. correlation between the clusters[17], I perform a robustness test with Discroll Kraay[18] standard errors in table A2 in appendix I, with persistent results.

5.2 Validation of the Hypotheses

<u>Hypothesis 1 (table 3):</u> *Comparative advantages and NEG determine industry location across Chinese provinces. Consequently the activity of energy intensive industries will be higher in energy abundant regions.*

Results in table 3 demonstrate that skill, market potential, supplier access and the different specifications of energy abundance determine positively and significantly the location of production, as predicted by theory, emphasizing that market forces are actually at work across Chinese provinces.

The hypothesis that energy intensive industries locate in energy abundant regions is clearly validated in all specifications in table 3, with special attention paid to total energy, which enters the regression significantly at the 1%-level and with a large coefficient. This positive and significant relationship has shown to be very robust throughout the various estimations performed and consistently reproduced as shown in all regression tables.

Surprising are the significant, negative and relatively large coefficients for the capital interaction variable. As for the energy carriers above, capital is considered more mobile than labour and should therefore expose a smaller (but nevertheless positive) coefficient than the skill interaction. Intrigued by these results I evaluated, and in the following rejected, the use of the share of gross capital formation in total output per industry instead of the value of the net fixed assets.[19] A possible explanation of the contradictory results is the phenomenon put forward by Rosen and Houser (2007) and briefly mentioned above, stating that provincial governments, especially in relatively energy abundant and inland provinces, strongly intervene in the localization of heavy industry which is not only energy but also capital intensive, leading to a strong distortion of industry localisation. Hence, provinces relatively scarce in capital, actually attract capital-intensive industries and exhibit a larger than proportional activity in those industries. However this is only a presumption and while calling for further investigation, exceeds the framework of this thesis and will therefore not be subject to detailed discussion.

The impact of demand and supply linkages are specifically large with respect to the other interactions, indicating that pecuniary spillover effects are relatively dominant in determining industry location. In column 3, I exclude internal and external supplier access because demand and supply linkages are highly correlated and have a tendency to crowd each other out; the coefficient for market potential becomes larger when supplier access is dropped. Both of the external linkage effects, sao and mpo are insignificant, so in column 2 I list a separate regression replacing the four economic geography variables with their equivalent aggregate value (the sum of the internal and external measure): total market potential

---

[17] *A source of contemporaneous correlation could be the existence of correlation between neighbouring provinces in the abundance of a production factor. In this data set for example, one might depict potential spatial correlation in energy production (see figure 1 p. 15).*

[18] *Discroll Kraay standard errors account for heteroskedasticity, autocorrelation and contemporaneous correlation. However, the xtscc command in stata indicates that as the estimator is based on asymptotic theory, the results should be interpreted with care, given that the cross-section in this study is fairly large and T relatively small.*

[19] *See appendix II for a more detailed discussion.*

MP=ln(mpi + mpo) and supplier access SA=ln(sai+sao). Their coefficients remain significant and are even accentuated. The indicator for backward linkages is consistently higher than for forward linkages, suggesting that market potential plays a more important role in determining relative output.

Hence, not only comparative advantage and factor endowments but also market externalities condition firms' location choices, substantiating the predictions of both the HO model and NEG. Energy proves to be a driving force in determining industry location next to the two principle comparative advantage indicators, skill and capital.

TABLE 3: *Industry location using 1999-2009 panel data set,*
Dependent variable=log of relative output

		Total energy production			Disaggregate energy		Energy reserves		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Comparative Advantage	Skill int x skill ab	0.077**	0.119***	0.089**	0.069*	0.068**	0.062*	0.063*	0.078**
		(0.035)	(0.040)	(0.040)	(0.036)	(0.036)	(0.037)	(0.036)	(0.035)
	Capital int x capital ab	-0.204*	-0.261**	-0.269**	-0.156	-0.155	-0.106	-0.108	-0.188*
		(0.114)	(0.125)	(0.118)	(0.130)	(0.129)	(0.135)	(0.136)	(0.113)
	Total energy int x total energy ab	0.126***	0.102**	0.188***					0.142***
		(0.042)	(0.046)	(0.046)					(0.037)
	Electricity int x Coal ab				0.050*	0.052**	0.087**	0.085**	
					(0.028)	(0.026)	(0.037)	(0.036)	
	Petrol int x petro ab				-0.024		-0.009		
					(0.044)		(0.041)		
	Natural gas int x natural gas ab				0.029		-0.008		
					(0.038)		(0.035)		
Economic Geography Variables	lnmpo	-0.442		-0.326	-0.203	-0.227	-0.233	-0.260	-0.463
		(0.585)		(0.520)	(0.580)	(0.583)	(0.589)	(0.584)	(0.572)
	lnmpi	1.173***		1.518***	1.114***	1.117***	1.137***	1.131***	1.159***
		(0.238)		(0.172)	(0.234)	(0.235)	(0.238)	(0.237)	(0.234)
	lnsao	0.609			0.610	0.599	0.587	0.579	0.631
		(0.554)			(0.578)	(0.576)	(0.562)	(0.563)	(0.542)
	lnsai	0.560**			0.638***	0.638***	0.611***	0.614***	0.521**
		(0.220)			(0.215)	(0.216)	(0.218)	(0.218)	(0.220)
	lnmp =ln(mpi +mpo)		2.725***						
			(0.554)						
	lnsa =ln(sai+sao)		1.318**						
			(0.515)						
Industry-year fe (131)		Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
Province-year fe (307)		Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
Coast dummy		Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
Clustered std. errors		Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
Non linear Wald test	$H_0 : b_{mpo}/b_{mpi}<1$	p=.9977				p=.9902	p=.9908		p=.9982
	$H_0 : b_{sao}/b_{sai}<1$	p=.4717				p=.5225	p=.5134		p=.4384
N Cluster		336	336	336	336	336	336	336	336
N		3685	3685	3685	3685	3685	3685	3685	3685
r2		0.689	0.653	0.664	0.685	0.685	0.687	0.687	0.695

*Standard errors in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01*

<u>Hypothesis 2 (table 3):</u> *The comparative advantage effect is larger for immobile energy production factors that are subject to higher transport costs.*

In regression 4, of the three included energy carriers only the coal x electricity interaction enters significantly and its coefficient exceeds the other two clearly.[20] This is in line with the theory's predictions, given that coal is considered relatively immobile and as the most important energy resource in China.[21] In column 5, I exclude petrol and natural gas, as they are not significant to control for any influence on the coal interaction, which turns out to be very small. When accounting for contemporaneous correlation in regression 3 in table A2 (appendix I), the coefficients for natural gas and petrol become significant but do not cease to be inferior to the coal-electricity interaction coefficient, validating the larger impact of immobile factors on industrial activity.

<u>Hypothesis 3 (table 4):</u> *The continuing transition from a planned economy to an open economy will show a net increase in the importance of factor endowments and NEG in determining industry location.*

To assess the effects of changes in the economic and political environment on the importance of comparative advantage and economic geography forces, I run, after a Wald test confirmed the joint significance of the year dummies, a separate regression for 1994, 1997, 2000, 2003, 2006 and 2009 on my cross-section data set. I include industry and province fixed effects and Huber-White heteroskedasticity robust standard errors. We can observe an increasing trend between 1994 and 2009 in the elasticity of market potential and energy, both variables entering significantly at the 1%-level. The fluctuations in the energy coefficient, especially the dip in 2006, however do not completely validate the hypothesis of a continuing process of a growing importance of market mechanisms in industry location. Also, when performing an additional Wald test comparing the coefficients of the different regressions, the results fail to demonstrate that the energy interaction coefficients are different in 1994 and 2009. Regarding the market potential variable however, the test clearly validates the assumption of an increase in the coefficients, thereby statistically confirming a growth in the importance of demand linkages on industrial activity.

With Poncet's (2003) and Batisse and Poncet's (2004) findings in mind, of protectionism between provinces actually increasing, the results observed in table 4 are surprising. The increasing barriers to the free flow of goods between regions would hinder firms from concentrating and specializing across provincial borders and the importance of HO and NEG forces should decline. When we consider augmenting or constant internal barriers, the increase in the importance of economic geography forces would be largely attributable to the reduction of external barriers, i.e. impediments to international trade, or lower transport costs due to improving infrastructure. The former corresponds to the economic policy developments made on the international level during the period regarding free trade with other countries as China reduced considerably the impediments to international trade preceding the accession of the country to the WTO in 2001.

---

[20] The previous normalization of the variables allows for direct comparison of the resulting regression coefficients.

[21] The world coal association indicates that the transportation costs account for a large share of the delivered coal price.
(http://www.worldcoal.org/coal/market-amp-transportation/, 13.06.2012)

In short, based on the evolution of the elasticity of the demand linkage variable, the data clearly supports the prediction of a liberalizing tendency, giving rise to increased specialization and concentration of manufacturing, accentuating the importance of proximity to markets.

Hypothesis 4 (table 5): *As coastal regions have benefited from preferential treatment and open market policies since the beginning of the reform, the influence of economic geography and comparative advantage forces is larger in these provinces.*

Instead of running separate regressions for coastal and internal provinces that feature a different number of observations and would therefore not be directly comparable, I interact the independent variables with the coastal dummy and include them in the model. The results for the OLS (columns 2-4) and IV (columns 5-6) estimations can be observed in table 5. In columns 2-4 the coast-skill interaction enters positively and significantly into the estimation of the equation, to the detriment of the overall skill interaction, implying that the coastal region is the main force behind the skill's positive effect. The elasticity of the energy interaction is generally higher in coastal provinces even if insignificant, but turns significant when the economic geography variables are excluded.

Interestingly the dummy variable for coast in the basic specification is generally negative and significant as shown in column 1, which is contrary to expectations. A possible explanation is the high correlation of the coastal dummy with the market potential and supplier access indicators mpi, mpo and sai, sao for which, when excluded, the coast dummy coefficient becomes positive.[22] This suggests that the higher market potential and supplier access of the coastal provinces already cover a large part of their advantage or positive effect with respect to relative output. This could also explain to some extent why their interactions with coast are insignificant in 2-3.

The exogenous instruments in the IV regression in columns 5 and 6 are constructed as interactions of the alternative measures of abundance (sufficiency and reserves) with energy consumption per labour. Applying the IV to the coast interaction specification would imply instrumenting triple interactions. To avoid these complications I nevertheless run a separate regression for coast and center. The results are not conclusive as the energy interaction is not found to be endogenous and the overidentification restrictions are rejected for the "center regression", implying a possible correlation of the instruments with the error term. [23]

While single results allow for varying interpretations, they nevertheless support the overall conclusion of a larger elasticity of HO and NEG effects in coastal provinces, most likely attributed to the further advanced liberalization of said region.

---

[22] *When excluding lnmpi, lnmpo lnsai and lnsao from the specification the coefficient for the coastal dummy is 1.483 (0.482) and enters significantly at the 1%-level.*

[23] *When replacing total energy with the coal x electricity interaction the instruments are valid only in the regression for coastal provinces, rendering the two equations incomparable (p-value of endog 0.808, Hansen J p-value 0.016 for the central provinces)*

TABLE 4*: Industry location using the cross-section data set*
Dependent variable=log of relative output

	pooled 1994-2009		1994	1997	2000	2003	2006	2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Skill int x skill ab	0.095*** (0.033)	0.098*** (0.038)	0.074** (0.034)	0.113*** (0.038)	0.134*** (0.046)	0.105** (0.052)	0.122** (0.047)	0.094** (0.046)
Capital int x capital ab	-0.173*** (0.039)	-0.195*** (0.038)	-0.086 (0.076)	-0.315*** (0.113)	-0.213*** (0.041)	-0.326*** (0.086)	-0.527*** (0.104)	-0.143*** (0.042)
Total energy int x total energy ab	0.137*** (0.040)	0.195*** (0.043)	0.176*** (0.039)	0.137*** (0.039)	0.208*** (0.057)	0.233*** (0.063)	0.168*** (0.047)	0.225*** (0.063)
lnmpo	-0.586 (0.576)	-0.406 (0.490)	-0.468 (0.475)	-0.298 (0.485)	-0.958* (0.577)	-0.455 (0.629)	-0.403 (0.597)	-0.149 (0.634)
lnmpi	1.113*** (0.239)	1.372*** (0.164)	0.983*** (0.156)	1.165*** (0.169)	1.347*** (0.191)	1.547*** (0.184)	1.600*** (0.181)	1.513*** (0.172)
lnsao	0.684 (0.513)							
lnsai	0.434* (0.221)							
industry-year fe (131)	Ok	Ok	-	-	-	-	-	-
province-year fe (307)	Ok	Ok	-	-	-	-	-	-
province (27)	-	-	Ok	Ok	Ok	Ok	Ok	Ok
industry (11)	-	-	Ok	Ok	Ok	Ok	Ok	Ok
Coast	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
Std. errors	clustered	clustered	Huber-White	Huber-White	Huber-White	Huber-White	Huber-White	Huber-White
N Cluster	336	336						
N	2014	2014	336	336	336	336	334	336
r2	0.724	0.703	0.542	0.579	0.619	0.659	0.669	0.655

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

<u>Hypothesis 5 (table 3):</u> *Interprovincial barriers to trade exist between provinces, implying that the impact of market forces on industry location is reduced.*

Back to table 3, nonlinear Wald tests verify the existence of barriers to trade. A value of β_{mpo}/β_{mpi} (or the equivalent for sai and sao) smaller than 1 indicates that the contribution of distance-weighted external market potential is smaller than the contribution of the internal to the overall market potential effect and therefore implies the existence of trade barriers. The tests on both demand and supply linkages cannot reject the H_0 of β_{mpo}/β_{mpi} being smaller than 1, the results validating presumed provincial protectionism. With protectionism impeding the free flow of goods, the possibilities of specialization across provinces and industry concentration are reduced, impacting negatively NEG and HO forces.

In addition to the Wald tests in table 3, I perform a nonlinear least squares estimation as proposed by Amiti and Javorcik (2008) in table A1, using the economic geography indicators as defined in Appendix III (equation 8 and 9), as an alternative way to measure the existence of protectionism. The coefficients are estimated on a basic model without fixed effects but including the linear variables of the interactions to account for province and industry differences. In columns 2 and 3, where I consider MP and SA separately, the results from table 3 are confirmed. The overall market potential and supplier access effects are lower than in previous regressions but highly significant and positive. The coefficients for the external market potential and supplier access enter insignificantly into the regression with a value close to 0, implying that the internal market externalities indicator determines almost the totality of the overall demand and supply effect. In this sense we have seen that both estimations confirm the existence of important barriers to trade between Chinese provinces.

5.3 Additional Robustness Tests

In table A3, I run an IV regression on the basic specification given in table 3 as well as on an alternative specification including the linear variables (the components of the interaction terms) instead of fixed effects. Column 1, 3 and 4 show the IV regression for the different energy interactions, whereby their coefficients enter significantly and are larger than in the basic specification. However as the instrumented and instrumental variables refer to interaction terms, the direction of the endogeneity bias is not necessarily clear and the difference in coefficients is to be interpreted with care. The elasticities of the other explanatory variables prove to be consistent, with the exception of the capital interaction, which turns insignificant when the coal x electricity interaction is instrumentalized. The robustness test performed in regressions 2 and 5, including the linear variables show that the energy interaction coefficients are robust. An additional IV regression performed on the cross-section data set and the specification of table 4, is not included in this paper, as the exogeneity of the instrumented variables could not be rejected.

In appendix I, table A2, columns 5, I add province-industry to the province-year and industry-year fixed effects as proposed by Gerlagh and Mathys (2010). This allows capturing the dynamics of the panel, accounting for the variation of the country and industry characteristics through time. The coefficients are expected to be lower as relative abundance and intensity are assumed to be constant. The economic geography indicators are dropped because, as per construction, they do not vary through time. The whole set of fixed effects explains almost all of the variation with an R-squared of 96.5%. The coefficients of the interaction variables show the same sign as before and are very much reduced as was expected and remain significant only when additionally excluding the province Qinghai.[24] This shows that only in a limited manner the interactions for skill, capital and total energy continue exerting a significant impact on the dependent variable.

[24] Qinghai's relative abundance measures show a very large variation across the time period analyzed, therefore not complying with the long term aspect of the HO framework.

TABLE 5: *Industry location accounting for coast-center differences*
Dependent variable=log of relative output

	Coastal dummy interaction				IV	
	(1)	(2)	(3)	(4)	(5)	(6)
					Coast	Center
Skill int x skill ab	0.077** (0.035)	-0.092 (0.088)	-0.114 (0.090)	-0.036 (0.093)	0.150*** (0.037)	-0.143 (0.090)
Capital int x capital ab	-0.204* (0.114)	-0.134 (0.218)	-0.174 (0.223)	-0.246 (0.239)	-0.043 (0.082)	-0.096 (0.191)
Total energy int x total energy ab	0.126*** (0.042)	0.083* (0.050)	0.132** (0.054)	0.127** (0.056)	0.093 (0.059)	0.149** (0.062)
lnmpo	-0.442 (0.585)	-0.103 (0.817)	0.154 (0.790)		0.169 (0.380)	-0.169 (0.801)
lnmpi	1.173*** (0.238)	1.226*** (0.293)	1.513*** (0.234)		1.169*** (0.199)	1.445*** (0.249)
lnsao	0.609 (0.554)	0.510 (0.617)				
lnsai	0.560** (0.220)	0.492* (0.262)				
Coast*skill interaction		0.271** (0.107)	0.332*** (0.114)	0.375*** (0.122)		
Coast* capital interaction		-0.224 (0.273)	-0.309 (0.273)	-0.332 (0.300)		
Coast*energy interaction		0.113 (0.086)	0.126 (0.091)	0.192* (0.111)		
Coast*lnmpo		-0.654 (0.976)	-1.107 (0.897)			
Coast*lnmpi		-0.348 (0.443)	-0.146 (0.358)			
Coast*lnsao		-0.076 (0.929)				
Coast*lnsai		0.318 (0.453)				
Coast	-1.99*** (0.664)					
Industry-year fe (131)	Ok	Ok	Ok	Ok	Ok	Ok
Province-year fe (307)	Ok	Ok	Ok	Ok	Ok	Ok
Clustered std errors	Ok	Ok	Ok	Ok	Ok	Ok
Underidentifica-tion (F-stat)					35.949	76.919
Weak instrument (F-stat)					28.364	213.309
Hansen J (p-value)					0.468	0.04
Endog (p-value)					0.537	0.290
N Cluster	336	336	336	336	336	336
N	3685	3685	3685	3685	1188	2497
r2	0.689	0.698	0.676	0.600	0.553	0.290

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6. CONCLUSION

In the course of this research, the assessment of past literature with respect to the empirical implementation of the HO model has led to the specification and estimation of an NEG augmented HO model, restricted to the analysis of Chinese provinces under inclusion of regional protectionism and with special focus on energy endowments. I analyze the determinants of industry location in China between 1994 and 2009 in the context of the growing importance of energy resources and the rapid evolution of the political and economic framework in the country.

In the results displayed, we have seen that throughout Chinese provinces comparative advantage and economic geography forces significantly determine industry location, thereby demonstrating the importance of market mechanisms even in the context specific to China, which has otherwise shown to not quite meet the ideal framework of a regional HO study. More specifically, energy proves to be a driving force of the specialization of industrial activity. This suggests that within China, relatively energy abundant provinces attract energy intensive industries and energy scarce provinces attract industries with relative low energy intensity respectively. The analysis of the disaggregated energy carriers confirms the prevailing importance of coal endowments, knowing that coal features the highest transport costs and that it is the principle source of energy in China. Furthermore, demand and supply linkages play a dominant role in explaining industry output, implying the existence of a concentration process.

Aggregating the data set into 6 different time periods have allowed to some extent to assess China's gradual shift away from the centralized planned economy structure through the economic reforms and opening policies initiated in 1978, whereas important steps were made towards the establishment of a market economy during the 1990's. Indeed, the results confirm that the importance of market proximity has risen significantly between 1994-2009 and that the energy endowment effect has a rising tendency over the whole period. Additionally, the data demonstrates to a certain degree that, as the opening policies were concentrated and exclusively directed towards coastal provinces until the end of the 1980's, market mechanisms have a larger influence in these regions.

The existence of internal impediments to trade has also been assessed. The impediments are generally estimated to be very high and comparable to the barriers between independent countries. Separating the access to demand and supply, into an internal and external effect, allows analyzing each respective contribution to the overall market externality effect. The importance of external linkages has shown to be of negligible size in comparison to the internal linkages, indicating the existence of trade barriers. This paper thereby confirms that inter-provincial trade impediments, which can include tariffs and restrictions as well as subsidies and specific regional policies, are omnipresent and constitute a major force counterbalancing HO and NEG mechanisms. The rising effects of market forces mentioned above, have therefore occurred in spite of constantly high internal barriers and are more likely associated with international opening policies and the reduction of transportation costs through improving infrastructure, constituting a possible further extension to this study.

The conditions specific to China cast, as already mentioned, certain doubts upon the empirical benefits of a within country analysis, given that neither free flow of goods, nor identical technology matrices are confirmed in the Chinese context. Also, as the applied specification does not allow for an analysis of the specific effects of regional and international protectionist measures on output nor its influence on the HO and NEG effects, room is left for a more profound analysis of interregional dynamics. Furthermore, with the continuing improvement and expansion of statistics published by the Chinese government, an extension of the model to different technology matrices would potentially be possible. Regarding the location and the rising importance of energy intensive industries, their relative large share of state-owned enterprises (SOEs) also recommend for further research on behalf of the SOEs' influence on industrial activity in heavy industries.

In conclusion, the abovementioned findings could have an important impact on current and future environmental policies launched by the Chinese government in two ways. On the one hand, having energy intensive industries concentrating in regions where energy endowments are abundant will most likely weaken the effects of national environmental policies directed towards energy consumption reduction, as the share of energy in production costs increases, thereby exerting a larger influence on firms' location decisions. On the other hand, the important influence and interventionist tendency of local governments, fostering and protecting certain industries, subsidizing heavy industry at which environmental policies are directed, prevents efficiency gains due to specialization and makes policy implementation more difficult. In this sense, the complexity of China's economic and political framework calls for further investigation on the dynamics of the relationship between these forces through time.

APPENDIX

I. Additional Regression tables

TABLE A1: *Industry location based on nonlinear regression, proposed by Amiti and Javorcik (2008) to assess the existence of provincial barriers to trade*
Dependent variable=log of relative output

		(1)	(2)	(3)
Interaction variables	Skill int x skill ab	0.176***	0.186***	0.174***
		(0.039)	(0.038)	(0.039)
	Capital int x capital ab	-0.397***	-0.416***	-0.352**
		(0.136)	(0.134)	(0.143)
	Total energy int x total energy ab	0.188***	0.217***	0.200***
		(0.052)	(0.056)	(0.056)
	lnSA	0.518**		0.205***
		(0.224)		(0.093)
	lnsao	0.014		0.001
		(0.030)		(0.025)
	lnMP	-0.863***	0.170**	
		(0.278)	(0.083)	
	lnmpo	1.877	-0.008	
		(2.180)	(0.007)	
	Linear variables	Ok	Ok	Ok
	Coast	Ok	Ok	Ok
	Population	Ok	Ok	Ok
	Industry-year fe	-	-	-
	Province-year fe	-	-	-
	Clustered std. errors	Ok	Ok	Ok
	N Cluster	336	336	336
	N	3685	3685	3685
	r2	0.992	0.992	0.992

*Standard errors in parentheses, * p < 0.10, ** p < 0.05, *** p < 0.01*

TABLE A2: *Robustness test with contemporaneous correlation and industry-province fixed effects*

Dependent variable=log of relative output

	basic	Discroll Kraay standard errors			industry-province fe
	(1)	(2)	(3)	(4)	(6)+
Skill int x skill ab	0.077**	0.077***	0.069***	0.079***	0.033**
	(0.035)	(0.012)	(0.013)	(0.012)	(0.014)
Capital int x	-0.204*	-0.204**	-0.156*	-0.206**	-0.050**
capital ab	(0.114)	(0.079)	(0.081)	(0.085)	(0.023)
Total energy int x	0.126***	0.126***			0.028*
total energy ab	(0.042)	(0.010)			(0.014)
Electricity int x			0.050***	0.080***	
Coal ab			(0.015)	(0.015)	
Petrol int x			-0.024***	-0.012	
Petro ab			(0.005)	(0.008)	
Natural gas int x			0.029**	0.024	
Natural gas ab			(0.013)	(0.016)	
lnmpo	-0.442	-0.442**	-0.203	-0.017	(omitted)
	(0.585)	(0.190)	(0.188)	(0.169)	
lnmpi	1.173***	1.173***	1.114***	1.509***	(omitted)
	(0.238)	(0.073)	(0.064)	(0.058)	
lnsao	0.609	0.609***	0.610***		
	(0.554)	(0.115)	(0.127)		
lnsai	0.560**	0.560***	0.638***		
	(0.220)	(0.030)	(0.024)		
Coast dummy	-1.991***	24.057***	25.798***	20.922***	(omitted)
	(0.664)	(1.108)	(1.108)	(1.368)	
Industry-year fe (131)	Ok	Ok	Ok	Ok	Ok
Province-year fe (307)	Ok	Ok	Ok	Ok	Ok
Province-industry fe (324)	-	-	-	-	Ok
Clustered std. errors	Ok	Ok	Ok	Ok	-
N Cluster	336	336	336	336	
N	3685	3685	3685	3685	3556
r2	0.689	0.689	0.685	0.652	0.965

*Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

+ Excluding the province Qinghai because of its strong variation in energy production through time: The province evolved from energy scarce to energy abundant in a short time span, strongly influencing the regression coefficients. An IV regression rejects the H_0 of the Sargan test of valid instruments (p-value of 0.000) and is therefore not listed.

TABLE A3 : *Industry location through IV regression using panel data set*
Dependent variable=log of relative output

		IV with total energy interaction variable[+]		IV with disaggregate energy variable interactions[++]		
		(1) total energy	(2-OLS°) linear variables	(3) coal	(4) coal,petrol,gas	(5) linear variables
Interaction variables	Skill int x skill ab	0.078**	0.141***	0.062*	0.062*	0.118***
		(0.034)	(0.035)	(0.034)	(0.035)	(0.038)
	Capital int x capital ab	-0.203*	-0.345***	-0.103	-0.105	-0.202
		(0.107)	(0.129)	(0.128)	(0.129)	(0.133)
	Total energy int x	0.203***	0.209***			
	total energy ab	(0.071)	(0.05)			
	Electricity int x			0.106***	0.115***	0.203***
	Coal ab			(0.043)	(0.047)	(0.058)
	Petrol int x				0.036	0.070
	Petro ab				(0.048)	(0.057)
	Natural gas int x				0.036	0.045
	Natural gas ab				(0.052)	(0.059)
	lnmpo	-0.591	0.115	-0.253	-0.312	0.118
		(0.553)	(0.182)	(0.550)	(0.5554)	(0.187)
	lnmpi	1.200***	0.663***	1.104***	1.090***	0.645***
		(0.228)	(0.105)	(0.220)	(0.218)	(0.109)
	lnsao	0.597		0.569	0.542	
		(0.506)		(0.536)	(0.541)	
	lnsai	0.504**		0.627***	0.634***	
		(0.219)		(0.203)	(0.200)	
Linear Variables	lneducation		-0.179*			-0.209*
			(0.106)			(0.111)
	lncapital per capita		-0.396**			-0.371*
			(0.1201)			(0.205)
	lnwage		-0.036			-0.026
			(0.120)			(0.205)
	lncapital intensity		0.178***			0.174***
			(0.048)			(0.041)
	lncoal consumption per labour		0.203***			0.208***
			(0.057)			(0.059)
	lnoil consumption per labour		0.085			0.087
			(0.071)			(0.072)
	lnelectricity consumption per		-0.177*			-0.189*
	labour		(0.096)			(0.101)
	lnnatural gas consumption per		-0.002			-0.003
	labour		(0.006)			(0.006)
	lncoal production per capita		0.004			0.006
			(0.005)			(0.006)
	lnpetrol production per capita		0.006*			0.008**
			(0.003)			(0.003)
	lngas production per capita		-0.006			-0.005
			(0.003)			(0.004)
	lnelectricity production per		0.011			-0.032
	capita		(0.102)			(0.109)
IV Statistics	Underidentification (F)	98.705	84.455	27.757	103.195	100.654
	Weak instrument (F)	16.14	21.907	162.980	96.693	99.762
	Hansen J (p-value)	0.107	0.096	0.3356	0.2430	0.1943
	Endog (p-value)	0.04	0.097	0.0087	0.0553	0.0051
	Coast dummy	Ok	Ok	Ok	Ok	Ok
	population dummy	-	Ok	-	-	Ok
	Year dummies	-	Ok	-	-	Ok
	Industry-year fe (131)	Ok	-	Ok	Ok	-
	Province-year fe (307)	Ok	-	Ok	Ok	-
	Clustered std. errors	Ok	Ok	Ok	Ok	Ok
	N Cluster	336	336	336	336	336
	N	3685	3685	3685	3685	3685
	r2	0.323	0.557	0.310	0.309	0.525

*Standard errors in parentheses, * p < 0.10, ** p < 0.05, *** p < 0.01*

[+]*The energy interaction is instrumented by the interactions of the sufficiency and reserves of coal and petrol with the energy carriers' respective consumption per labour.*

[++] *The instruments for the coal interaction are the interactions of coal sufficiency and coal reserves with the electricity consumption per labour.*

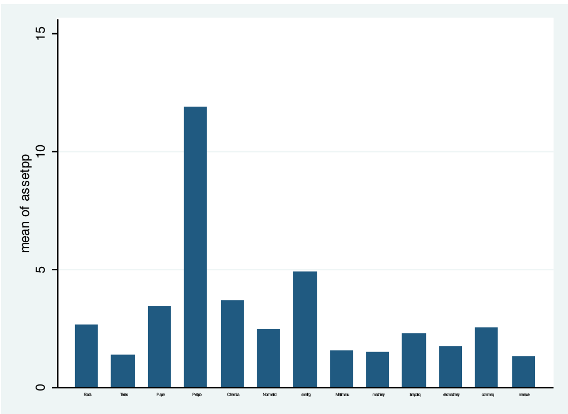
°The IV statistics given in this column reject the validity of the instruments with a Hansen J test at the 10%-level, therefore the coefficients given, are those of OLS.

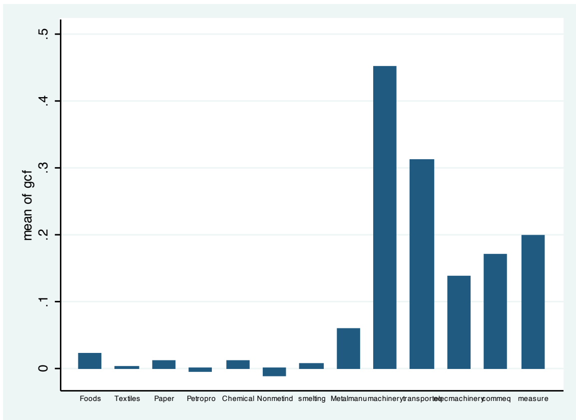
II. Capital Intensity Discussion

The fact that the coefficient for the capital interaction is persistently negative, relatively large and often significant at the 1% -level throughout my estimations implied a more profound study of the variables used in its construction. I was inclined to use a different measure of capital intensity such as proposed by Batisse and Poncet (2004) who used the share of gross capital formation (GCF) in the industries' total output. However as shown in Figure A1b, GCF gives quite a different picture. It is the complete opposite to capital stock in A1a regarding the conclusion on which industries turn out to be capital intensive. The capital stock per industry measure describes the industries according to what one would a priori expect: petrol refining, smelting, chemical and paper industries are the most (column 4, 7, 3 and 5) and measuring instruments, textiles and machinery (column 13, 2 and 9) among the least capital intensive industries.

FIGURE A1a: *Mean of capital stock by industry (average over sample period)*

FIGURE A1b: *Mean of gross capital formation by industry (average over sample period, %)*





The gross capital formation on the other hand gives a short-term image and reflects current capital needs. Nevertheless after a brief descriptive analysis of the data there are three reasons that speak against its use compared to capital stock: First, it indicates the opposite of what one would expect to be typical capital intensive industries, with smelting and chemicals for example showing almost no or even negative gross capital formation. Also, it does not correspond to the findings of Rosen and Houser (2007) who observe a general shift towards heavy industry and not away from it. Second, gross capital formation is negatively correlated to energy use per worker, which is not supported by general findings of energy intensive industries often also being capital intensive. Third, gross capital formation, as already mentioned, is a snapshot of capital use and does not support the long-term orientation of the HO model framework.

III. Variable Construction

III.A Linear abundance and intensity variables

All monetary values are in 100 million RMB and deflated with respect to 1971. In the regressions, I have taken the log and then standardized all variables per year over province and industry, implying that the abundance and intensity measures have means of 0 with a variance of 1.

Dependent variable: China Industrial Statistical Yearbook.
Output per industry and per province relative to the share of industry i and province p in total national output.

Energy: China Energy Statistical Yearbook.
Energy abundance is defined as coal (in 10'000 tons), petrol (in 10'000 tons), natural gas (100 million cubic meters) and electricity (in 100 million kwh) production per province expressed relative to the province's population (in 10'000).
The energy intensity is defined as consumption per industry expressed relative to the number of workers (in 10'000) per industry.
Total energy is the aggregate of abovementioned coal, electricity, petrol and natural gas values converted to toe (ton per oil equivalent, using the standard conversion factors given on www.iea.org/stats/unit.asp) and expressed relative to population for abundance and workers for intensity respectively (both in 10'000).

Capital: China Industrial Statistical Yearbook, China Statistical Yearbook and Wu (2009).
Provincial capital stock values are taken from Wu (2009) and complemented according to the perpetual inventory method using the depreciation rates given in his paper and provincial gross capital formation from the China Statistical Yearbook. Capital intensity is the value of net fixed assets per industry and province divided by the number of workers (in 10'000) taken from the China Industrial Statistical Yearbook.

Skill: China Statistical Yearbook and Banister (2005).
Skill intensity per industry is proxied by the average urban wage in RMB in 2001 per industry, given in Banister (2005). The abundance is measured by the share of the population older than 6[25], with college level or higher educational attainment.

[25] *As defined in The China Statistical Yearbook.*

III.B Market Potential/Supplier Access

Market potential:

(8) $lnMP_p^i = \beta_{MP} ln(MPI_p^i + \beta_{MAO} MPO_p^i)$

Internal market potential

(8a) $MPI_p^i = \frac{\sum_{i'=1}^{I'} b_p^{ii'} * \partial_p^{i'} + b_p^i * \lambda_p}{\sqrt{\frac{Area_p}{\pi}}}$

External market potential

(8b) $MPO_p^i = \sum_{m \neq p}^{R} \frac{\sum_{i'=1}^{I'} b_m^{ii'} * \partial_m^{i'} + b_m^i * \lambda_m}{distance\ mp}$

$b^{ii'}$: Share of industry i output absorbed by industry i'

b^i: Share of industry i output absorbed by final demand so that $b^i + \sum_{i'=1}^{I'} b^{ii'} = 1$

$\lambda_p = \frac{GDP_p}{GDP_{China}}$

$\partial_p^{i'} = \frac{Y_p^{i'}}{Y_{China}^{i'}}$: Share of industry i' in province p's output relative to the national output of industry i'.

Supplier Access:

(9) $lnSA_p^i = \beta_{SA} \ln(SAI_p^i + \beta_{SAO} SAO_p^i)$

Internal supplier access

(9a) $SAI_p^i = \frac{\sum_{i'=1}^{I'} a_p^{ii'} * \partial_p^{i'}}{\sqrt{\frac{Area_p}{\pi}}}$

External supplier access

(9b) $SAO_p^i = \sum_{m \neq p}^{R} \frac{(\sum_{i'=1}^{I'} a_m^{ii'} * \partial_m^{i'})}{distance\ mp}$

$\partial_m^{i'} = \frac{Y_m^{i'}}{Y_{China}^{i'}}$: Share of industry i' in province m's output relative to the national output of industry i'.

$Area_p$: Province size in km^2

Distance mp: road distance from province p's capital to province m's capital.

$a^{ii'}$: share of industry i' in the intermediate input of industry i

The elasticity with respect to distance is -1, which according to MKOV, corresponds to market potential estimates from gravity models of trade.

Data used is provided in the 2002 China provincial Input-Output tables. Distances between provinces were measured as the quickest route between provincial capitals on google maps.

IV. Additional Descriptive Tables and Graphs

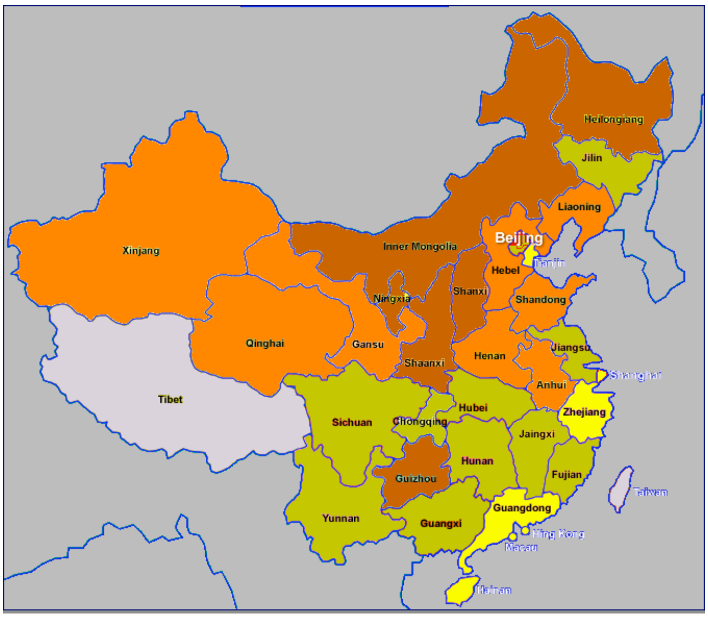
TABLE A4: *Coast-center regional division and province abbreviations*

Coast		Center			
Beijing	BJ	Shanxi	SX	Guangxi	GX
Tianjin	TJ	Hebei	HeB	Chongqing	CQ
Liaoning	LN	Inner Mongolia	IM	Sichuan	SC
Shanghai	SH	Jilin	JL	Guizhou	GZ
Jiangsu	JS	Heilongjiang	HLJ	Yunnan	YN
Zhejiang	ZJ	Anhui	AH	Shaanxi	SaX
Fujian	FJ	Jiangxi	JX	Gansu	GS
Shandong	SD	Henan	HeN	Qinghai	QH
Guangdong	GD	Hubei	HuB	Ningxia	NX
(Hainan)	HN	Hunan	HuN	Xinjiang	XJ

TABLE A5: *Average annual growth of industry intensity and provincial abundance in percentages*

	Intensity	Abundance
Capital	9.57	9.03
Skilled labour	NA	7.95
Energy	8.09	12.43
Electricity	8.99	10.80
Coal	8.54	12.26
Petroleum	-0.89	1.96
Natural Gas	10.02	16.78
Industry Output	20.45	

FIGURE A2: *Coal abundance distribution (log of average coal production over sample period in 10'000 tons)*



REFERENCES

Amiti, Mary and Javorcik Beata (2008). Trade Costs and Location of Foreign Firms in China. *Journal of Development Economics*, 85, 129-149.

Bai, Chong-En; Du, Yinguan; Tao Zhigang and Tong, Sarah Y (2003). Local Protectionism and Regional Specialization: Evidence from China's Industries. *Journal of International Economics*, 63(2), 397-417.

Banister, Judith (2005). Manufacturing Employment and Compensation in China. *Bureau of Labour Statistics.* Retrieved March 3, 2012, from www.bls.gov/fls/**china**report.pdf

Batisse, Cécile and Poncet, Sandra (2004). Protectionism and Industry Location in Chinese Provinces. *Journal of Chinese Economic Business Studies*, 2(2), 133-154.

Bowen, Harry P.; Leamer, Edward E. and Sveikauskas, Leo (1987). Multicountry, Multifactor Tests of the Factor Abundance Theory. *The American Economic Review*, 77(5), 791-809.

Davis, Donald R.; Weinstein, David E.; Bradford, Scott C. and Shimpo Kazushige (1997). Using International and Japanese Regional Data to Determine When the Factor Abundance Theory of Trade Works. *The American Economic Review*, 87(3), 421-446.

Davis, Donald R. and Weinstein David E. (1999). Economic Geography and Regional Production Structure: An Empirical Investigation. *European Economic Review*, 43, 379-407.

Davis, Donald R. and Weinstein David E. (2001). What Role for Empirics in International Trade. NBER Working Papers, No. 8543

Davis, Donald R. and Weinstein David E. (2003). International Trade as an "Integrated Equilibrium": New Perspectives. NBER Working Papers, No. W8637.

Ellison, Glenn and Glaeser, Edward L. (1999). The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *American Economic Review*, 89(2), 311-316.

Greene, William H. (2003). *Econometric Analysis*. (5th Ed.) Pearson Education, New Jersey.

Hillman, Arye L. and Bullard, Clark W. (1978). Energy, the Heckscher-Ohlin Theorem, and U.S. International Trade. *American Economic Review*, 68(1), 96–106.

Kim, Sukkoo (1999). Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987. *Regional Science and Urban Economics*, 29, 1-32.

Krugman, Paul R. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *The American Economic Review*, 70(5), 950-959.

Krugman, Paul R. (1991). *Geography and Trade*, MIT Press.

Krugman Paul R.; Obstfeld, Maurice and Melitz, Marc J. (2010). *International Economics : Theory and Policy*, (9th Ed.) Pearson Addison-Wesley, New York.

Leontief, Wassily (1954). Domestic Production and Foreign Trade: The American Capital Position Re-examined. *Economia Internazionale,* 7:3–32.

Lin, Boqiang and Liu, Jianghua (2010). Principles, Effects and Problems of Differential Power Pricing Policy for Energy Intensive Industries in China. *Energy*, 36.

Mathys, Nicole A. and Gerlagh Reyer (2010). Energy Abundance, Trade and Industry Location.

Michielsen, Thomas (2011). The Distribution of Energy-Intensive Sectors in the US. CentER Working Paper No. 2011-075.

Midelfart-Knarvik, Karen ; Overman, Henry and Venables, Anthony (2001). Comparative Advantage and Economic Geography: Estimating the Determinants of Industrial Location in the EU. CEPR Discussion Paper No. 2618.

Poncet, Sandra (2003). Measuring Chinese Domestic and International Integration. *China Economic Review*, 14(2003) 1-21.

Redding, Stephan J (2009). Economic Geography: A Review of the Theoretical and Empirical Literature. CEPR Discussion Paper No. 7126

Requena, Francisco; Castillo, Juana and Artal, Andrés (2008). Multiregion, Multifactor Tests of the Heckscher-Ohlin-Vanek Using Spanish Regional Data. *International Regional Science Review*, 31(2), 159-184.

Romalis, John (2004). Factor Proportions and the Structure of Commodity Trade. *The American Economic Review*, 94(1), 67-97.

Rosen, Daniel H. and Houser, Trevor (2007). China Energy: A Guide for the Perplexed. *China Balance Sheet*, Center for Strategic and International Studies and Peterson Institute for International Economics, Working Paper.

Smith, Pamela J. (1999). Do Geographic Scale Economies Explain Disturbances to Heckscher-Ohlin Trade? *Review of International Economics*, 7(1), 20-36.

State Statistical Bureau (various years). *China Industrial Statistical Yearbook*. Beijing Statistical Press

State Statistical Bureau (various years). *China Statistical Yearbook*. Beijing Statistical Press

State Statistical Bureau (various years). *China Input-Output tables*. Beijing Statistical Press

Suranovic, Steven M. *International Trade Theory and Policy,* 2007, internationalecon.com

Trefler, Daniel (1995). The Case of Missing Trade and Other Mysteries. *American Economic Review*, 85(5):1029–1046.

Wu, Yanrui (2009). China's Capital Stock Series by Region and Sector. The University of Western Australia, Discussion Paper 09.02.

Young, Alwyn (2000). The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China. *Quarterly Journal of Economics*, CXV, 1091-1136.

Yueng, Yue-man; Lee, Joanna and Kee, Gordon (2009). China's Special Economic Zones at 30. *Eurasian Geography and Economics*, 50(2).