We would like to submit this paper to:
The special issue of JCEBS
Or JCEBS as individual paper
The Best PhD Paper competition (Yi Zhang, PhD student at Utrecht University, supervisors Harry Garretsen and Hein Roelfsema)

Regional Determinants of FDI in China: A Factor-based Approach

Martijn A. Boermans and Yi Zhang[1]

Abstract

We empirically investigate the factors that drive the uneven regional distribution of foreign direct investment (FDI) across Chinese provinces from 1995 to 2006. We first perform a factor analysis to summarize information embodied in around forty variables and derive four FDI determinant factors: 'institutional quality', 'labor costs', 'market size', and 'geography'. Applying these estimated factors, we then employ IV estimation to account for endogeneity in panel estimation. In line with theoretical predictions we find that foreign investors invest in provinces with good institutions, low labor costs, and large market size. The Arellano-Bond dynamic panel GMM estimation results show strong agglomeration effects that multinationals tend to invest in provinces which attract other foreign firms, consistent with the economic geography literature. Several robustness tests indicate that low labor costs combined with improvements in institutions are the key for attracting FDI.

Key Words: China, FDI, regional distribution, factor analysis
JEL-codes: F21, F23, O18, O53, R11

[1] Corresponding author, email: y.zhang@uu.nl. Yi Zhang is a PhD student at the Utrecht University School of Economics, Janskerkhof 12, 3512 BL Utrecht, The Netherlands. Martijn A. Boermans is a PhD student at the HU Business School of the University of Applied Sciences Utrecht, Research Group for International Business and Innovation, Padualaan 101, 3508 AA Utrecht, the Netherlands.

1. Introduction

Over the last decades, foreign direct investment (FDI) has become an important engine for Chinese growth. However, there are large differences in FDI patterns across Chinese regions. For example, in 2000 the East geo-economic region has received more than 85 percent of the country's total FDI inflows, whereas the Central and Western provinces have only attracted less than 15 percent (National Bureau of Statistics of China). Regions also differ in the type of FDI they attract (Huang, 2003). Urban growth centres increasingly develop into magnates for market seeking FDI, whereas other regions are function as the factory of the world.

Most papers that study Chinese FDI distribution take a traditional route of analyzing FDI from a specific theoretical angle and therefore focus on a limited number of determinants to explain the variation across regions, such as geographical factors, agglomeration effects, labor costs, or institutional quality (Chen, 1996; Wei et al., 1999; Cheng & Kwan, 2000; Sun et al., 2002; Cole et al., 2006; Li & Park, 2006; Du et al., 2008). Further, as is often stressed in factor analysis, traditional empirical methods often use proxies for the underlying more general concepts, which hampers causal inference due to measurement errors and omitted information. Besides, there are other identification problems in most of the existing papers that deal with FDI in China. One obvious frustration is reverse causality arising from the plausible relation that FDI inflows affect regional characteristics. Clearly, panel analysis is one candidate to effectively cope with this issue, but such methodology is difficult with for example firm level data. Given these restrictions in focus and method, evidence on what explains the variation in FDI across Chinese regions is still incomplete.

We aim to provide a more eclectic approach to analyze FDI patterns and to handle endogeneity problems by combining conventional estimation methods and (less conventional) factor analysis. Let us briefly explain our line of thinking, without claiming that it solves all the problems mentioned above. We use data on FDI at the provincial level for the period 1995-2006. This is a period in which FDI spread from highly concentrated Pearl River Delta (PRD), and hence Guangdong province, towards other coastal regions as well as recently the Western and Central provinces (Chan et al., 2008). Before we identify provincial characteristics for which we have theoretical priors that they are correlated with FDI, we first explore to what extent provinces actually differ in their economic and social characteristics. To this end, we perform a factor analysis where we include 42 variables to show which factors (clusters of variables) explain a large proportion of regional variance. [2] Certainly we hope that these estimated factors can be related to economic theory: new economic geography, regional comparative advantage, new institutional economics and the like. As we have included many variables to explain a significant part of regional variance, we can be confident in controlling for many potentially omitted variables with extensive information used.

After that, we run traditional panel regressions where we account for endogeneity by using instruments. Broadly speaking, the following results stand out. First, institutions, comparative advantage, and market size all matter, but there are important differences with respect to coastal and inner provinces and with respect to interaction effects among these factors. However, as a single factor, differences in comparative advantage and especially labor costs seem to matter most in explaining the FDI distribution between 1995 and 2006. Hence, from a policy perspective one may argue that the efforts to spread investments towards regions with lower labor costs have succeeded. Second, the results of Arellano-Bond dynamic panel GMM estimation present strong agglomeration effect. Foreign investors tend to locate in the same province to benefit from potential positive externalities such as knowledge spillovers. Third, although governance and infrastructure cluster into one factor, especially infrastructure seems a precondition for labor costs and market size to have a sizeable effect on inward FDI. This calls for support of policies that promote (massive) infrastructural projects in regions where FDI is low, such as the Western and Central provinces. Lastly, we find positive effect of good governance on FDI, besides its connection with an increased supply of public goods.

The paper commences as follows. The next section discusses related literature with the aim of providing a theoretical foundation for our empirical research. Section 3 introduces the data and empirical strategy in more detail, with a special emphasis on the

---

[2] One has to keep in mind that these factors are clusters of variables that change over time, although some of the variables are rather static.

role of factor analysis in this paper. Following that, section 4 presents the core results. Then, section 5 performs robustness checks on the main findings. Section 6 concludes.

2. Related literature

FDI inflows into China are a widely studied subject for several reasons. For evidence on the determinants of FDI, studying China attracts great interest because flows are high and regionally dispersed. Further, by focusing on a single large country one accounts for many variables that could have been omitted (or at least imperfectly captured), which reduces selection bias when variables are evaluated at the end of a period without controlling for unobserved initial conditions. In addition, FDI inflows have created much policy debate within China because of its close links to the diversion in economic growth rate (Chan et al., 2008; Fleisher et al., 2010; Huang et al., 2003). The aim of this section is to cement the claim that there are many rival theories that explain the distribution of FDI across Chinese provinces. Hence, an empirical identification strategy which has the ability to make these theories 'compete' against each other is valuable, so as to shed light on the predictive power of each of them.

It is well known that the wave of setting up of production facilities for assembly in China at least initially has been driven by a comparative advantage in labor intensive production (Liu et al., 1997; Wang & Swain, 1997). From this observation alone it is not clear whether differences in labor costs also explain the regional distribution of FDI *across* Chinese provinces, however, there is some (rather old) evidence that they do. Using an error correction model, Wei et al. (1999) analyze the long-run relationship between inward FDI and regional characteristics. They conclude that provinces with lower wage rates attract more FDI. Fung et al. (2002) study the location choice of the Japanese and US multinationals and find a lagged negative impact of wage rates on FDI from both countries for the period 1991-1997. In the business literature, there are many studies that stress the importance of differences in labor costs across provinces. For example, based on a questionnaire survey among Hong Kong multinationals in the manufacturing industry, Zhang & Yuk (1998) indicate that - on top of the closeness to Hong Kong headquarters – low labor cost is one of the most important reasons for establishing subsidiaries in Guangdong rather than in other regions. Further, a new literature stresses the growing importance of labor quality. Cheng & Kwan (2000) test a partial dynamic panel adjustment model using province level data 1985-1995. Though they identify low wage rate as an important determinant of FDI, this effect is mitigated by high human capital levels. Sun et al. (2002) argue that there is a nonlinear relation between wage rates and FDI for the period 1986-1998, positive before 1991 and negative thereafter, while labor quality is a positive attractor of FDI throughout the sample period.

Clearly, with a vast number of potential consumers and high growth rate, the increase in domestic demand has attracted FDI to sell in Chinese markets (Zhang, 2005). Again, there is evidence that such motive explains variation of FDI across provinces. In an early cross-sectional study, Broadman & Sun (1997) show that a province's FDI stock increases with its market size. To control for initial conditions, Chen (1996) uses a conditional logit model and shows a similar positive relation between market size and FDI for a yearly panel that runs from 1987 to 1991. Buckley & Meng (2005) examine the horizontal and vertical FDI motives in Chinese manufacturing sector. Despite the coexistence of both motivations, they argue that for the period 1992-2002 the market-oriented FDI dominates. The significant impact of market size on regional distribution of FDI is also addressed in standard econometric settings by Wei et al. (1999), Fung et al. (2002), Sun et al. (2002), and Hong & Chin (2007). Notice that Sethi et al. (2003) explore Dunning's OLI model for China by using a factor analysis. Although they do not address causality, the principal components in this paper show that 'regional characteristics' and 'market attractiveness' are highly correlated with FDI.

The importance of agglomeration effects to explain FDI is related to the emergence of the new economic geography literature. The central thinking is that firm location choice involves a trade-off between making use of positive externalities that come from agglomeration and the negative effects that agglomeration has on factor costs. Given that China just recently opened up to foreign capital, it provided for an ideal study ground to observe the *dynamics* of FDI location choice. The seminal paper in this approach is Head & Ries (1996) who, controlling for other geographical factors, find strong agglomeration effects in FDI decisions, concentrated in the coastal areas' export processing zones. Many would follow in their footsteps. For example, recently Amiti &

Javorcik (2008) use firm level data to show effects of agglomeration and costs advantages on FDI decisions.[3] Ng & Tuan (2006) study mainland investment decisions of Hong Kong firms and find agglomeration effects, also outside the nearby PRD region. Focusing on 98 Chinese inland cities, Luo et al. (2008) conclude that industrial agglomeration and policy are the most important factors for Western regions to attract FDI based on pooled OLS estimation. Coughlin & Segev (2000) apply a spatial error model to control for special autocorrelation among Chinese provinces. Their results indicate that foreign investors tend to choose provinces with FDI intensive neighbours. In contrast to these studies, Sun et al. (2002) account for the shift in the nature of FDI and show that there is a movement to increasingly choose to invest in provinces with relatively *few* other foreign firms during 1986-1998.

The new institutional economics literature stresses the role of 'rules of the game' in economic development. It has been noticed that there are large differences in institutional quality across Chinese provinces, for example in controlling corruption (Cole et al., 2006; Li & Park, 2006), property right protection (Cheung & Lin, 2004), and local absorption capacity (Fu, 2008). Previous studies show that institutional variables such as control of corruption and legal development have a positive impact on attracting FDI. Du et al. (2008a) study the impact of economic institutions on the inward FDI distribution. Using a sample of US firms from 1993 to 2001, they find that economic institutions influence the location choice of FDI in China. Their conditional logit results indicate that strong contract enforcement, protection of intellectual property rights, weak government intervention, and low corruption all positively affect FDI in a province. Accounting for FDI from US, EU, Japan, and Korea, Du et al. (2008b) further find an interesting interaction between the degree of horizontal agglomeration and institutional quality as this agglomeration effect can mitigate the negative impact of weak public institutions on FDI inflows. Moreover, institutions also refer to local government effectiveness and public goods (La Porta et al., 1999). For example, Cheng & Kwan (2000) claim that foreign investment flows to regions with high road and railway density. Based on an industrial census dataset, Li & Park (2006) show that multinationals favour

[3] With firm level data it is important to note that often they restrict the analysis to cross section only, since there is no investment pattern at the firm level recorded over time. But clearly reverse causality is a limited problem when using firm level data.

provinces with better infrastructure in the form of electricity provision, communication facilities, as well as roads.

A current wave is to put more emphasis on heterogeneity across firms engaged in FDI to China, which may have an effect on its spatial distribution. Zhao & Zhu (2005) study heterogeneous macro motives for 50 source countries to invest in China. Their results indicate that foreign investors with different countries-of-origin have divergent reactions to location determinants like labor costs and infrastructure. Hu & Owen (2005) find that firms from Hong Kong, Macau, and Taiwan (HMT) have different FDI motives compared to firms from OECD countries. More specifically, agglomeration effects are especially important for firms from OECD countries, while labor costs attract FDI from HMT firms. For our results it is important to keep in mind that over time FDI flows are driven by the fact that firms from OECD countries enter later, existing firms become more acquainted in doing business in China and may therefore create a resource base comparable to firms from HMT, and that increasingly China is 'discovered' by medium sized firms. Belderbos & Carree (2002) analyze investment behaviour of Japanese firms in China and conclude that small firms are interested in benefits from agglomeration, whereas large firms pay more attention to cost advantages. In spite of working with aggregated data, Shapiro et al. (2007) contribute to this literature by studying the entry-mode-specific FDI determinants across Chinese provinces. They show that agglomeration effects mainly exist for high control modes of entry and FDI-favouring policies help to attract low control modes of entry.

Most mentioned studies have tackled the massive FDI inflows from an a priori background and highlighted a limited number of determinants to explain the variation across regions. We summarize some important analyses on Chinese FDI determinants in Table 1. Clearly, various studies present some different results. In the next section we show that using this traditional route requires proxies to test the broad theoretical constructs that as such can hampers causal inference due to measurement errors and omitted information. Given the relatively narrow focus of prior research we take no ex ante stance with regard to the discussed literature and hope to see that the empirical findings can fit in with any theoretical background.

[Table 1 here]

3. Data and methodology

The China Statistical Yearbook (1995-2007), published by the National Bureau of Statistics of China, provides data on FDI for 31 Chinese provinces from 1995 to 2006. From this dataset, we take the number of foreign funded enterprises (FFE) and their investment as the dependent variables to measure the extensive and intensive scales of FDI. Navaretti & Venables (2004) argue that it is important to make a distinction between the location choice and the amount invested by a multinational, which essentially are different decisions of the firm. The number of foreign funded enterprises marks the binary decision whether or not to invest in a province, while the amount of investment of foreign funded enterprises shows how - after the location choice - firms set production levels. We provide explanations for variables in Appendix.

As for explanatory variables, we derive latent factors that capture the variation of economic conditions across Chinese provinces by using factor analysis discussed in more detail below. Factor analysis concentrates variation of a large number of observed variables into (far) fewer aggregated dimensions and is widely regarded as a strong data reduction tool (Forni et al., 2001; Bernanke et al., 2006). As a result, we endogenously derive the factors ‘institutions’, ‘labor costs’, ‘geography’, and ‘market size’. We have to bear in mind that we are ‘lucky’ that the analysis translates into factors that closely match theory. Certainly, labelling is subjective (Comrey & Lee, 1992). Although some of the labels are relatively straightforward, it is important to note upfront that we deploy a wide concept of institutions, which covers infrastructure (hard institutions) as well as quality of government and rule of law (soft institutions). This is not a modelling choice: factor analysis is neutral in determining the composition of each factor, leaving us no room at this point to split hard and soft institutions as the loadings of observed variables cluster them into a single factor. The interesting interpretation is that there are no good hard institutions without good soft institutions.[4]

[4] This interpretation is supported by La Porta et al. (1999), who show that infrastructural quality is highly correlated to government performance.

3.1 Estimation methods

After running the factor analysis, we follow a more conventional approach. With the reduced number but high powered explanatory variables from the factor analysis, we first apply the fixed effects estimation to get rid of province-specific effects.[5] Given potential reverse causality, we then employ IV estimation to alleviate the problem of endogeneity. As an example of such reverse causality, foreign investment may raise labor costs by driving up wages. In that case, the simple fixed effects estimation is likely to underestimate the impact of labor costs on FDI. Given the first-order autocorrelation in our data, we use the lagged two years variables as internal instruments. In addition, to account for dynamic patterns in FDI distribution and well-known agglomeration effect we use a standard Arellano-Bond GMM model. Finally, we perform various robustness checks by using sub-samples for Eastern and Western China, forcing different factors, and regressing on alternative dependent variables.

3.2 Factor analysis

The standard regression-based approach in economics uses variable selection and model specification informed by theory. By contrast, factor analysis selects variables not by using theory, but by reducing a large number of variables into a set of clusters (Jöreskog, 2007). Essentially, factor analysis asks which variables belong to one another in a separate class. Hence, instead of taking up a specific variable implied by theory, one hopes to generate a factor close to the theory which includes more information. To some extent, the method is related to making indexes comprising of various variables, common in economics. Factor analysis takes account of the statistically 'correct' way of creating such an index by using information on the underlying correlation of the variables and their proper weighting in the index (Velicer & Jackson, 1990; Wall & Amemiya, 2007).

Factor analysis uncovers patterns of association in the dataset and describes the variation among observed variables in terms of fewer factors, so that a complete set of

[5] Clearly, first test for panel unit root and panel cointegration: all the series are I (1) and cointegrated in the long run. As a check, the random effects estimation failed to pass the Hausman test.

interdependent relationships is examined. A natural question to ask is under which conditions a factor-based approach may be superior to a hand-picked selection of explanatory variables informed by theory. Clearly, this depends on some trade-offs. The first trade-off is between the efficient use of information and the imprecise matching of factors to theories. Factor analysis may make better use of information for the following reasons. First, because a large number of variables can be reduced to a small set of factors which account for most of the variation in the initial dataset, a reduction in the degrees of freedom can be avoided without losing information. Such a feature is especially attractive in small samples. In this case, the classic trade-off between good fit and parsimony means that additional information cannot be freely exploited by adding variables. Second, the identification problem of using highly correlated variables can be mitigated by only employing orthogonal factors in regressions. Normally, economic indicators are interdependent, causing empirical difficulties to show the individual effect of each single variable. Again, such an identification problem is severe in small samples, for limited information obfuscates the connection between explanatory variables. Third, factor analysis allows the efficient inclusion of variables that relate to more than one theory. For example, since wage levels give information on both labor costs as well as market demand, taking up wage levels as an explanatory variable does not identify theories that explain the inflow FDI. For wage rates may enter as an individual item along with others in a factor 'labor costs' as well as in a factor 'market demand', in this way such 'undetermined' variables are used efficiently.

However, factor analysis may reduce efficiency because the matching of factors to theories may be less precise than matching hand-picked variables to theories. As factors summarize information of all observed variables, generated factors contain variation that may be distant from the theory to which it is (subjectively) matched. Indeed, the generated factors only make sense when we can attach economic meaning to them and link them to theory. The benefits of the factor-based approach outweigh the costs when rival theories are tested for which a large number of variables is needed, but where the sample is small. For example, it is impossible to include 50 highly correlated variables into a single regression if we only have 200 observations. Clearly, a limited sample also restricts the application of the factor-based methodology.[6] To be precise, the factor-based approach is efficient when the sample size is large enough to provide a stable correlation matrix for factor analysis, but is too small to efficiently run regressions with a large number of highly correlated observed variables.

The second trade-off is between reducing some errors while creating some others. Factor analysis reduces three types of errors which hamper finding causal relations in regressions. First, the selection bias is smaller because the factor-based approach takes up many variables within factors, so as to be more likely to also include variables that cause selection when left out, or which are correlated to unobserved variables that may cause selection bias. This is closely related to endogeneity issues in making causal inferences due to omitted variables. Especially in small samples the general-to-specific modelling approach cannot fully overcome the selection bias (Sala-i-Martin, 1997). Second, imprecise representation of general economic concepts by specific variables may lead to weak predictive power of theories. For instance, the choice of a specific data series for the concept economic activity is often arbitrary to some degree (Bernanke et al., 2006). As another example, researchers normally use a proxy which can be correlated with omitted variables, which in turn hampers causal inference. By summing up information among a large number of observable variables, factor analysis identifies groups of inter-related variables and covers "un-measurable" or latent dimensions which single variables cannot capture (Jöreskog, 2007). Last, the factor-based approach mitigates measurement error ex post, as clustering of variables smoothes out such errors in individual variables. Especially, when both dependent and independent variables suffer from measurement errors, spurious regressions may result. Hence, when measurement errors are prevalent but randomly distributed, the diffusion property of aggregated factors attributes to the reduction of measurement biases and identification of causal relations.

However, the factor-based approach constructs two new errors. The first one is a "generated error" when factors are created, since factors are estimated variables based on

[6] Though strict theoretical rules with respect to the sample size in factor analysis are still missing, some practical suggestions either on the absolute sample size or functions of sample size are available in shaping stable patterns. Comrey & Lee (1992) view a sample size of 100 to be poor, 200 to be fair, 300 to be good, 500 to be very good, and 1000 or more as excellent. Other guidance provides minimum ratio of the sample size to the number of observed variables. It ranges from 2:1 to 20:1 and 10:1 is commonly used as a rule of thumb.

correlation matrices of observed variables and factor loadings. The second one is a new specification bias since only a limited number of factors are retained by choice. Since we are agnostic about the true number of factors, improper determination of factors may yield model selection bias. Therefore, the factor-based approach can be a better way to identify causality in the situation where traditional errors are more predominant than the new errors. Specifically, in the case of a relatively small sample, selection bias may be large because the process of from general to specific is hard to be realized given limited observations and highly correlated variables; representation bias is hard to be solved by including all relevant variables; measurement errors are severe due to the lack of sufficient variations.

Reviewing these trade-offs, the case of regional distribution of FDI in China fits well into the framework of factor analysis. First, the large number of hypotheses on FDI determinants (we use 42 indicators common in FDI literature) result in a hassle in model specification and efficiency conservation. Second, many observed variables are highly correlated (e.g. gross regional products and wage rates) and cover different rival theories, which cause identification problems. Third, representation errors are nontrivial as some potentially important FDI determinants are theoretically defined and hard to be captured by a specific single variable. For example, the factor institutions covers various dimensions in politics, economics, and culture, which cannot be represented by only including a variable like 'government efficiency' or 'rule of law'. Fourth, both dependent and independent variables are subject to measurement errors, because of different quality in provincial accounting practices, revisions and redefinitions, and the mismatch between national and international statistical rules. All these problems are especially serious in small sample. Moreover, the costs of the factor-based approach seem not to be severe in our study. First, although with around 300 observations and 42 highly correlated variables it is hard to get reliable regression results, such a sample size suffices for factor analysis. Second, as we will see, variables load rather straightforward to various factors and as such we can attach meaningful economic interpretations while we link the endogenous factors to theory. Finally, in our study it seems that retaining correct number of factors is not a big worry, since the first four factors explain around 70 percent of all variation in the dataset.

4. Results

4.1 Factor analysis results

Focusing on the partial correlations among variables, we first test sampling adequacy with the Kaiser-Meyer-Olkin (KMO) measure. We find that the KMO measure is 0.859 in our case, greater than the critical value 0.5 for a satisfactory factor analysis to proceed. Table 2 then reports the rotated factor loadings by using the principle-component method to analyze the correlation matrix.[7] We apply an orthogonal varimax rotation to maximize the variance of the squared loadings within factors to obtain independent factors. In this table the higher the loading, the stronger is the correlation between the item and the underlying factor. Regarding loadings lower than 0.4 as 'low' correlations as suggested by Hair et al. (1998), we obtain a factor structure close to a 'simple factor structure' with most main loadings greater than 0.6 and few cross-loadings. In this factor structure, four factors are retained based on a scree test showing the plot of the eigenvalues (Jennrich, 2007). The proportion of variation explained is 42 percent (factor 1), 17 percent (factor 2), 10 percent (factor 3), and 7 percent (factor 4), respectively. Factor 1 consists of infrastructure (transportation and communication) and governance indices (for example, the NERI institutional quality, property rights protection, and the preferential policy). Since we have named infrastructure 'hard institutions' and governance 'soft institutions', we interpret factor 1 as provincial 'institutional quality'. Factor 2 shows items related to 'labor costs', such as wage rates, education variables, and labor endowment variables.[8] Factor 3 captures 'first nature' geography such as local climate conditions and natural resource. Factor 4 takes up variables common in economic geography theory and relates

[7] We also apply maximum-likelihood method to derive latent factors using the same dataset. This method assumes multivariate normal observations and maximizes the determinant of the partial correlation matrix. Results indicate that different methods of factor analysis yield similar latent variables.

[8] In Table 2 educational variables (especially primary and junior high school education) are positively loaded to factor 2. This implies that wage rates increase faster than productivity with the improvement in basic education. In fact, low skilled workers may exert marginal impact on innovation and they tend to move out of non-stated-owned firms since low-efficient state-owned firms pay more to workers with low education (Zheng & Hu, 2004).

specially to 'market size' with high loadings for GDP per capita, household consumption, and the level of market transactions.

[Table 2 here]

We then test the reliability of factors by applying Cronbach's alpha which shows how well a set of variables measures the same underlying construct. Table 3 illustrates whether to delete a single item can get a higher alpha value for each factor. Across panels we find no significant contribution of eliminating any item. The test scale results reveal acceptable factor reliability with average Cronbach's alpha above 0.7.[9] This indicates that the variables clustered to each category have relatively high inter-item correlations and therefore the factors obtained are consistent. We also identify some important variables (e.g. government expenditure for factor 1 and import for factor 3) as removing these items would result in significantly lower values of alpha. We later leave these variables out of the factor generation process and use them as control variables in robustness check.

[Table 3 here]

4.2 Estimation results

Table 4 presents the fixed effects estimation and the cross-sectional 'between' estimation results. Across provinces, all factors except for the 'labor costs' factor are correlated in the expected way to the number of foreign investors and the level of foreign investment. One explanation for the insignificance of the 'labor costs' factor in the between estimation is that indeed foreign firms expect labor costs to be low in all provinces relative to their home market. Hence, foreign investors are not so much concerned about the differences in labor costs across regions. Again, however, we point out that the

[9] The widely-accepted social science cut-off value for alpha is 0.7 (some use 0.6). We admit that our factor 2 is not satisfying in this sense. However, the standard of reliability required varies between different fields. For example, a value of 0.8 or higher is appropriate for psychology. Hence, it is not surprising to obtain relatively low alpha values as we study a loose geo-economic structure. Results not reported here show that if we exclude primary and junior high school enrolment the alpha value for factor 2 goes up to 0.6031, close to the cut-off value. We therefore reconstruct factors by dropping these items in robustness checks.

impact of labor costs is likely to be underestimated since there is the possibility of reversed causality. Not only labor costs can affect FDI, location choice of foreign firms may also affect local labor costs. Without controlling for this, using endogenous labor costs gives biased results. Certainly, the downward bias also affects the magnitude and significance of the factors ‘market size’ and ‘institutions’.

[Table 4 here]

To take care of endogeneity, we obtain the IV estimation results in Table 5 with the lagged two years explanatory variables as internal instruments. All regressions control for time and province-specific effects. The significant negative impact of the ‘labor costs’ factor and positive impact of the ‘market size’ factor in Columns (1) and (3) provide evidence of the coexistence of vertical FDI and horizontal FDI in China. The magnitude of the coefficients indicates that labor costs are the most important determinant of FDI across China – note that factors are standardized. The small but significant interaction effects imply that good institutions magnify the effects of low labor costs and market size on FDI. So, the conclusion is that, since vertical FDI is the most important to China, low labor costs wins the first round of the horse race among the factors.[10]

[Table 5 here]

Clearly, path dependence in FDI may be an important factor, especially if agglomeration effects are at work. To take account of panel dynamics we apply Arellano-Bond dynamic panel GMM estimation. With some good will, we can interpret the coefficients of the lagged dependent variables as agglomeration effects. In Table 6 the positive lagged dependent variables imply that multinationals tend to invest in provinces that have attracted other foreign firms in the past. When we control for agglomeration, the effects of other factors on the number of firms engaged in FDI remain similar in size and significance to the results of the static model in Table 5. But clearly, for the level of FDI,

---

[10] As evidence for the prevalence of vertical FDI in China, according to the National Bureau of Statistics of China, exports by foreign funded enterprises (FFE) occupy around 60 percent in total exports from China in 2006, with an increasing trend over the past 15 years.

including agglomeration effects reduces the significance of all factors. Hence, the factors concerned have an impact on attracting new firms but these factors do not significantly affect the overall level of FDI when agglomeration is included. One reason is that FDI levels in the past are determined by the lags of the factors, so that these to some extent are included in the lagged dependent variable. This in turn reduces the significance of the present levels of the explanatory factors. In a more qualitative way, when local conditions improve, it is relatively easy to set up a new firm, however, it takes time for these firms to grow, so as to significantly change the overall level of FDI in a province.

[Table 6 here]

5. Robustness

There are some worries with respect to the basic results presented in the previous section. First, there are large differences between FDI flows towards the Eastern and Western provinces. Hence, we should check whether the drivers of FDI play the same role in the East as they do in the West of China. We further check province sensitivity by excluding the four municipalities directly administrated by the central government – Beijing, Shanghai, Tianjin, and Chongqing. Second, we have seen that unforced principal components result in clustering of hard- and soft institutions into a single factor. We are interested to see what happens when we force a split, so as to say more on the connection between political institutions (soft) and FDI. Third, as discussed before, there are trade-offs between the factor-based approach and the regression-based approach. Using some hand-picked variables as direct measures in benchmark models, we would like to know whether estimation results differ significantly by using the factors and the observed variables. Moreover, since some variables are relatively important in getting reliable factors, dropping these variables may change the significance levels of the factors in regressions. We generate new factor structures by kicking out some key variables and estimate the effects of the new factors using these important items as control variables. Last, we check whether the basic results are robust to alternative dependent variables.

5.1 Province sensitivity

We group 31 Chinese provinces into the East (13 provinces: Beijing, Fujian, Guangdong, Hainan, Hebei, Heilongjiang, Jiangsu, Jilin, Liaoning, Shandong, Shanghai, Tianjin, Zhejiang) and the West (Anhui, Chongqing, Gansu, Guangxi, Guizhou, Henan, Hubei, Hunan, Inner Mongolia, Jiangxi, Ningxia, Qinghai, Shaanxi, Shanxi, Sichuan, Tibet, Xinjiang, Yunnan).[11] Table 7 indicates that the 'labor costs' factor is the most important FDI determinant for both regions, restating the overall importance of vertical FDI in China. Yet, there are other different reasons for investing in the East and the West. Factors 'Institutional quality' and 'market size' significantly affect FDI distribution across the Eastern provinces. By contrast, on top of labor costs, the factor 'first nature' geography plays by far the most important role in explaining the variation of FDI across the Western provinces. This may imply that FDI in the West is to generate natural resource inputs for (export) production in the Eastern provinces. An interesting finding is the significant negative effect of the 'institutional quality' factor on the amount of investment in Column (3). One explanation for this result may be that FDI is a substitute for the absence of sufficient local public goods. Table 8 shows that the interaction effects of FDI determinants in Table 5 are robust in the East. Moreover, Column (2) implies that in the West where foreign investors produce for intra-country trade, in the absence of good institutions, low labor costs have no impact on the number of FFEs. The reason may be that the competitive advantage across the Western provinces relies more on infrastructure that enables transportation to the East than on low labor costs.

[Table 7 and Table 8 here]

We further drop the four direct-controlled municipalities from our estimation. As the highest-ranked cities, these municipalities are treated as provinces but differ from other provinces in many aspects like population, government expenditure, and policy. Though we control for such disparities in our factor analysis, a subsample study could show a

[11] Although the East and the West are loaded with the common factors (institutions, labor costs, market, and geography), factor structures differ for these two groups. Hence, we generate factors for the East and the West, respectively. However, using the original factors does not change the results.

clearer picture on this issue. Table 9 presents the IV estimation and dynamic panel estimation results for the other provinces. Comparing with our basic results in Table 5, we find that in Columns (1) and (2) the static effects of the 'institutional quality' and 'market size' factors on attracting FDI (both the number of firms and investment amount) are larger for the rest provinces than for the municipalities. The differences may result from the relatively well-developed institutions and limited size of those provincial-level cities. Given the fact that labor costs may be crucial for the FDI location choice in these high wage rates cities, it is not surprising to see a smaller impact of the 'labor cost' factor after excluding the municipalities. However, once we account for the past FDI, only the 'labor costs' factor matters for increasing the number of foreign invested enterprises in Column (3) except for the lagged dependent variable. This implies that local conditions on institutional quality and market size in these provinces improve slower than in the municipalities which are directed authorized by the central government, so that the lagged dependent variable captures the significance of these two factors.

[Table 9 here]

5.2 Hard and soft institutions

To separate hard- and soft institutions, we run a factor analysis for only the variables correlated with the overall factor 'institutional quality', which provides us two sub-factors: 'infrastructure' (hard institutions) and 'governance' (soft institutions), see Table 10 for rotated factor loadings. Table 11 reports the IV regression results for using these two sub-factors. With respect to FDI location choice (the number of FFEs), due to the vertical nature of FDI, foreign investors are mainly concerned with quality of transportation and communication. Once the investment location has been chosen, soft institutions, such as property rights protection, matter for the amount of investments. A negative interaction between the 'market size' and 'governance' factors in Column (4) suggests that when market has sufficient power to tackle with the hassles in contracts, the importance of external enforcement is reduced.

[Table 10 and Table 11 here]

5.3 Alternative FDI variables

Table 12 shows IV estimation results of using various FDI dependent variables which are FDI inflows, registered capital of foreign funded firms, the number of people employed by FDI firms, and a factor based on all FDI related variables, respectively. Effects of ‘labor costs’ and ‘market size’ factors are consistent across all panels, while the impact of the factor ‘institutional quality’ depends on the dependent variable selected. One remark is that dependent variables like FDI inflows can not distinguish between the location choice stage and the investment stage of FDI, and therefore miss the significance of the ‘institutional quality’ factor.

[Table 12 here]

5.4 Benchmark models

To compare the factor-based approach and the regression-based approach, we report the static IV estimation results of using some selected observed variables as direct measures for infrastructure (government expenditure and/or ways), labor costs (wage and/or GRP per capita), market size (GRP and/or household consumption), and geography (natural resource) in Table 13. We also incorporate primary school enrolment, higher education enrolment, *NERI* institutional quality index, as well as provincial preferential policy to capture human capital and soft institutions. Across panels, the estimation results are highly sensitive to model specification. For example, government expenditure variable loses its significance (though with an unexpected negative sign) in Column (2) once we substitute GRP per capita for wage to measure labor costs in Column (4). The same puzzle appears if we measure infrastructure by local transportation condition in Column (3) rather than the overall government expenditure in Column (2). The mixing results of these benchmark models imply that the hand-picked variables may be highly correlated and steal each others' significance, as shown in Column (6). Furthermore, the omitted variable problem can be severe and cause biased regression results.

[Table 13 here]

5.5 Variables as controls

We have known from Table 3 that some items are quite important for generating reliable factors, for example government expenditure, imports, exports, and junior high school enrolment, in the sense that excluding them results in substantial drop in alpha values. Further, as we can see in Table 2, the loadings of soft institutions are debatable. There is a worry whether factor analysis still provides us with informative factors once we leave these items out. We also notice that Cronbach's alpha of our factor 2 'labor costs' increases to the cut-off value 0.6 once we drop primary school and junior high school enrolment. Therefore, to check the reliability of our main results, we create a number of new factors without certain observed variables using factor analysis and report the estimation results in Table 14. We first eliminate the variables in question and construct new factors with a smaller number of items. Then we report the effects of these new factors on the number of foreign funded enterprises, together with dropped variables as controls. We proceed by keeping out government expenditure in Columns (1) and (2), international trade in Columns (3) and (4), the three together in Columns (5) and (6), the key variables on soft institutions in Columns (7) and (8), as well as primary and junior high school enrolment in the last two. Though in each pair we use a new list of factors, the resulted factor structures are similar to what we have in Table 2 and four new factors are retained. Across panels, the results for the factors are comparable to our basic findings in magnitude and significance level. All control variables are significant with the expected signs (exports and imports jointly significant).

[Table 14 here]

6. Concluding remarks

In this paper we have analyzed recent FDI inflows in China at the provincial level. Our approach is eclectic as we combine factor analysis with conventional regression methods. Informed by a large literature that stresses many variables which are correlated with FDI flows, we first run a factor analysis to establish latent regressors for which Chinese provinces differ. Broadly speaking, on top of geographical fixed factors, regions differ in labor costs, market potential, and hard- and soft institutions. Using panel estimation methods, the fixed effects estimation and IV estimation, we then perform a 'horse race' among these factors to see which factors matter most. We show that for the 1995-2006 period, labor costs and infrastructure (and especially when combined) are the key for attracting FDI. In addition, our Arellano-Bond dynamic panel GMM estimation results indicate strong agglomeration effects in FDI distribution across Chinese provinces.

Our study certainly does not contradict the relations found in other papers. A main difference is that we focus on a time frame where the Chinese government has changed course to advocate a transfer of FDI towards the West. After setting up the export processing zones, the Chinese government in the 1990s has made great strides to diffuse FDI. Our results first fit against a background of FDI diffusion away from the Pearl River Delta to other Eastern provinces. For instance, increasingly Pan-Bohai Rim is able to capture a large share of FDI by effectively tapping into cheap labor from the inner provinces. On top of that, our empirical findings are consistent with a trend of shifting FDI towards inner provinces, especially by firms from Taiwan and Hong Kong. For these firms, cost advantages are important assets in competitive world markets, so that they shift to cheaper Western location when infrastructure is ready. According to the National Bureau of Statistics of China, the share of FDI occupied by the East decreases from 87.7 percent in 2000 to 75.1 percent in 2008, while for the same period the share of the Central and Western area increases from 12.3 percent to 24.9 percent.

Can we draw lessons for the ongoing policy debate on the relative importance of geography, big push development, and institutions? Clearly, we have to be cautious here. However, from our analysis it becomes clear that geography is not all important if big push efforts in infrastructure are made. In addition, in China soft institutions (such as differences in local corruption and education) do not seem to play an important role other that they tend to go together with 'hard' institutions such as infrastructural improvements. This calls into question to what extent institutional reform alone in China as well as in other parts of the world is able to create FDI flows.

However, the analysis may also point to a more critical observation, one that is shared in much of the management literature on investing in China. In the data, there is the suggestion that labor costs and logistics remain the most important driving for foreigners to invest in China. This may also be because higher valued activities are still seen as too risky. The obvious reason is a lack of property rights protection, so that assembly based on higher skills (and, hence, higher labor costs and more schooling) remains unprofitable for foreign firms in the long run. A second reason is a lack of local management skills to perform integrated system production processes. Lastly, there is a often heard complaint that in joint ventures, ailing domestic firms are pushed by local politicians for inclusion in joint venture production. All these issues suggest that the dominant strategy for foreign firms still is to make use of cheap and disciplined labor, so that the next step towards high value added production is jet to come.

References

Amiti, M. & Javorcik, B. S. (2008): Trade Costs and Location of Foreign Firms in China, *Journal of Development Economics,* 85(1-2), pp. 129-149.

Belderbos, R. & Carree, M. (2002): The Location of Japanese Investments in China: Agglomeration Effects, Keiretsu, and Firm Heterogeneity, *Journal of the Japanese and International Economies*, 16(2), pp. 194-211.

Bernanke, B. S., Boivin, J. & Eliasz, P. (2006): Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics*, 120(1), pp. 387-422.

Broadman, H. G. & Sun, X. (1997): The Distribution of Foreign Direct Investment in China, *The World Economy*, 20(3), pp. 339-361.

Buckley, P. J. & Meng, C. (2005): The Strategy of Foreign-invested Manufacturing Enterprises in China: Export-orientated and Market-orientated FDI Revisited, *Journal of Chinese Economic and Business Studies*, 3(2), pp. 111-131.

Chan, K. W. J., Henderson, V. & Tsui, K. Y. (2008): Spatial Dimensions of Chinese Economic Development, in China's Great Economic Transformation, ed. by Rawski, T. G. & Brandt, L. Cambridge: Cambridge University Press.

Chen, C. H. (1996): Regional Determinants of Foreign Direct Investment in Mainland China, *Journal of Economic Studies*, 23(2), pp. 18-30.

Cheng, L. K. & Kwan, Y. K. (2000): What are the Determinants of the Location of Foreign Direct Investment? The Chinese Experience, *Journal of International Economics*, 51(2), pp. 379-400.

Cheung, K. Y. & Lin, P. (2004): Spillover Effects of FDI on Innovation in China: Evidence from the Provincial Data, *China Economic Review*, 15(1), pp. 25-44.

Cole, M., Elliott, R. & Zhang, J. (2006): "Corruption, Governance and FDI Location in China: A Province-Level Analysis," Department of Economics, University of Birmingham.

Comrey, A. L. & Lee, H. B. (1992): *A First Course in Factor Analysis*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Coughlin, C. & Segev, E. (2000): Foreign Direct Investment in China: A Spatial Econometric Study, *The World Economy*, 23(1), pp. 1-23.

Du, J. L., Lu, Y. & Tao, Z. G. (2008): Economic Institutions and FDI Location Choice: Evidence from US Multinationals in China, *Journal of Comparative Economics*, 36(3), pp. 412-429.

Du, J. L., Lu, Y. & Tao, Z. G. (2008): FDI Location Choice: Agglomeration vs Institutions, *International Journal of Finance & Economics*, 13(1), pp. 92-107.

Fleisher, B., Li, H. Z. & Zhao, M. Q. (2010): Human capital, economic growth, and regional inequality in China, *Journal of Development Economics*, 92(2), pp. 215-231.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2001): Coincident and Leading Indicators for the Euro Area, *The Economic Journal*, 111(471), pp. C62-C85.

Fu, X. L. (2008): Foreign Direct Investment, Absorptive Capacity and Regional Innovation Capabilities: Evidence from China, *Oxford Development Studies*, 36(1), pp. 89-110.

Fung, K. C., Iizaka, H. & Parker, S. (2002): Determinants of U.S. and Japanese Direct Investment in China, *Journal of Comparative Economics*, 30(3), pp. 567-578.

Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998): *Multivariate Data Analysis with Readings*. NJ: Prentice-Hall: Englewood Cliffs.

Head, K. & Ries, J. (1996): Inter-City Competition for Foreign Investment: Static and Dynamic Effects of China's Incentive Areas, *Journal of Urban Economics*, 40(1), pp. 38-60.

Hong, J. & Chin, A. T. H. (2007): Modeling the Location Choices of Foreign Investments in Chinese Logistics Industry, *China Economic Review*, 18(4), pp. 425-437.

Hu, A. G. & Owen, R. F. (2005): Gravitation at Home and Abroad: Regional Distribution of FDI in China, Singapore: Department of Economics, National University of Singapore.

Huang, J. T., Kuo, C. C. & Kao, A. P. (2003): The Inequality of Regional Economic Development in China between 1991 and 2001, *Journal of Chinese Economic and Business Studies*, 1(3), pp. 273-285.

Huang, Y. S. (2003): *Selling China: Foreign Direct Investment during the Reform Era*. Cambridge, UK: Cambridge University Press.

Jennrich, R. I. (2007): Rotation Methods, Algorithms and Standard Errors, in Factor Analysis at 100: Historical Development and Future Directions, ed. by Cudeck, R. & MacCallum, R. C. New Jersey: Lawrence Erlbaum.

Jöreskog, K. G. (2007): Factor Analysis and Its Extensions, in Factor Analysis at 100: Historical Development and Future Directions, ed. by Cudeck, R. & MacCallum, R. C. New Jersey: Lawrence Erlbaum.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. & Vishny, R. (1999): The Quality of Government, *Journal of Law, Economics and Organization*, 15(1), pp. 222-279.

Li, S. & Park, S. H. (2006): Determinants of Locations of Foreign Direct Investment in China, *Management and Organization Review*, 2(1), pp. 95-119.

Liu, X., Song, H., Wei, Y. & Romilly, P. (1997): Country Characteristics and Foreign Direct Investment in China: A Panel Data Analysis, *Review of World Economics*, 133(2), pp. 313-329.

Luo, L. J., Brennan, L., Liu, C. & Luo, Y. Z. (2008): Factors Influencing FDI Location Choice in China's Inland Areas, *China & World Economy*, 16(2), pp. 93-108.

Navaretti, G. B. & Venables, A. J. (2004): *Multinational Firms in the World Economy*. Princeton, New Jersey: Princeton University Press.

Ng, L. F. Y. & Tuan, C. (2006): Spatial Agglomeration, FDI, and Regional Growth in China: Locality of Local and Foreign Manufacturing Investments, *Journal of Asian Economics*, 17(4), pp. 691-713.

Sala-I-Martin, X. (1997): I Just Ran Two Million Regressions, *The American Economic Review*, 87(2), pp. 178-183.

Sethi, D., Guisinger, S. E., Phelan, S. E. & Berg, D. M. (2003): Trends in Foreign Direct Investment Flows: A Theoretical and Empirical Analysis, *Journal of International Business Studies*, 34(4), pp. 315-326.

Shapiro, D., Tang, Y. & Ma, X. J. (2007): Mode of Entry and the Regional Distribution of Foreign Direct Investment in China, *Journal of Chinese Economic and Business Studies*, 5(3), pp. 261 - 277.

Sun, Q., Tong, W. & Yu, Q. (2002): Determinants of Foreign Direct Investment across China, *Journal of International Money and Finance*, 21(1), pp. 79-113.

Velicer, W. F. & Jackson, D. N. (1990): Component Analysis versus Common Factor Analysis: Some issues in Selecting an Appropriate Procedure, *Multivariate Behavioral Research*, 25(1), pp. 1-28.

Wall, M. M. & Amemiya, Y. (2007): A Review of Nonlinear Factor Analysis and Non Linear Structural Equation Modelling, in Factor Analysis at 100: Historical Development and Future Directions, ed. by Cudeck, R. & MacCallum, R. C. New Jersey: Lawrence Erlbaum.

Wang, Z. Q. & Swain, N. (1997): Determinants of Inflow of Foreign Direct Investment in Hungary and China: Time-Series Approach, *Journal of International Development*, 9(5), pp. 695-726.

Wei, Y., Liu, X., Parker, D. & Vaidya, K. (1999): The Regional Distribution of Foreign Direct Investment in China, *Regional Studies*, 33(9), pp. 857-867.

Zhang, K. H. (2005): Why Does so Much FDI from Hong Kong and Taiwan Go to Mainland China?, *China Economic Review*, 16, pp. 293-307.

Zhang, X. & Yuk, H. P. (1998): Determinants of Hong Kong Manufacturing Investment in China: A Survey, *Marketing Intelligence & Planning*, 16(4), pp. 260-267.

Zhao, H. X. & Zhu, G. T. (2000): Location Factors and Country of Origin Differences: An Empirical Analysis of FDI in China, *Multinationals Business Review*, 8(1), pp. 60-73.

Zheng, J. H. & Hu, A. G. (2004): An Empirical Analysis of Provincial Productivity in China (1979-2001), Göteborg University, Department of Economics.

Table 1: The analysis of FDI determinants

Authors	Aggregation	Year	Labor Costs	Human Capital	Market Size	'First Nature' Geography	Hard Institution	Soft Institution	Agglomeration Effect
Liu et al. (1997)	Country	1983-1994	-/S		+/S	-/S distance		+/S	
Wang & Swain (1997)	Country	1978-1992	-/S		+/S			+/S	
Wei et al. (1999)	Province	1985-1995	-/S	+/S	+/S		+/S	+/S	+/S
Fung et al. (2002)	Province	1991-1997	-/S	+/S	+/S		+/S	+/S	
Zhang & Yuk (1998)	Guangdong	1997	-/S			-/S distance	+/S	+/S	
Cheng & Kwan (2000)	Province	1985-1995	-/S	+/I	+/S		+/S	+/S	+/S
Sun et al. (2002)	Province	1986-1998	-/S	+/S	+/S		+/S	+/S	-/S
Zhang (2005)	Country	1980-2001	-/S		+/S			+/S	
Broadman & Sun (1997)	Province	1985-1992	+/I	+/S	+/S	+/S coast	+/S		
Chen (1996)	Province	1987-1991	+/I	-/S		+/S resource	+/S		
Hong & Chin (2007)	City	1992-2001	-/S	+/S	+/S		+/S	-/I	+/S
Amiti & Javorcik (2008)	Province	1998-2001	-/S		+/S	-/S distance	+/S		+/S
Ng & Tuan (2006)	Guangdong	1998				-/S distance		+/S	+/S
Cole et al. (2006)	Province	1998-2003	-/S	+/S	+/S		+/S	+/S	+/S
Li & Park (2006)	Province	1995			+/I		+/S	+/S	+/S
Du et al. (2008 a and b)	Province	1993-2001	-/S	+/S			+/S	+/S	+/S
Hu & Owen (2005)	Province	1993-2003	-/S	-/S	+/S		-/S	+/S	-/S
Belderbos & Carree (2002)	Province	1990-1995	-/S		+/S	-/I distance	-/I		+/S
Shapiro et al. (2007)	Province	1997-2000	-/S	+/S	+/S	+/S coast	+/S	-/S	+/S
Luo et al. (2008)	City	1999-2005	-/I	+/S	+/S	-/I resource	+/S	+/S	+/S
Coughlin & Segev (2000)	Province	1990-1997	-/S	+/S	+/S	+/S coast	+/I		+/S

Note: '+'and '-' represent the sign of the effect; 'S' and 'I' denote for significant and insignificant, respectively.

Table 2: Rotated factor loadings

Variables	Factor1	Factor2	Factor3	Factor4	Uniqueness
Capital	**0.8434**	0.2703	0.0459	0.1813	0.0847
City road length	**0.7277**	**0.3637**	0.0375	0.1638	0.1491
City road area	**0.8298**	0.3185	0.0299	0.1246	0.0976
Civil vehicle	**0.8964**	0.2998	-0.0664	0.2053	0.0541
freight	**0.6493**	**0.4972**	-0.1093	0.0336	0.2225
Gov Expenditure	**0.9243**	0.1474	0.0463	0.2191	0.0400
Private vehicle	**0.9004**	0.1830	-0.0730	0.1739	0.0924
Ways (train, water, highway)	**0.5331**	**0.4195**	0.0440	-0.3019	0.2627
Exports	**0.8666**	0.0109	0.1786	0.1074	0.0566
Imports	**0.8010**	-0.0339	0.1232	**0.3547**	0.0712
Long telephone	**0.9185**	0.2223	0.1572	0.0110	0.0741
Local telephone	**0.9012**	0.2832	0.0736	0.0444	0.0447
Mobile	**0.9650**	0.0851	0.0903	0.0507	0.0415
Cable	**0.6272**	0.1851	-0.0852	-0.3405	0.1804
Patent registered	**0.8882**	0.0919	0.1469	0.1856	0.0678
Higher education enrolment	**0.7497**	0.2434	-0.0059	0.1332	0.0606
Higher education institutions	**0.6470**	**0.4633**	0.0305	0.2672	0.0961
Senior enrolment	**0.6481**	**0.5298**	0.0430	-0.0792	0.0622
Senior high school	**0.4405**	**0.8006**	0.1671	-0.0362	0.0837
Junior enrolment	0.3376	**0.8573**	0.1207	-0.1265	0.0906
Junior high school	0.1445	**0.9356**	0.0532	-0.1426	0.0562
Primary enrolment	0.1891	**0.9106**	0.2232	-0.1381	0.0539
Primary school	-0.1518	**0.8872**	0.0518	-0.1420	0.1078
Population	0.3394	**0.8953**	0.1759	-0.0718	0.0301
Workers	0.3422	**0.7798**	-0.0382	0.1225	0.0738
Humidity	-0.0293	0.1624	**0.9032**	-0.0523	0.1321
Sunshine	-0.0783	-0.3354	**-0.7836**	-0.0310	0.2208
Temperature	0.1785	0.1918	**0.8711**	0.0761	0.1514
Area	-0.0735	-0.1425	**-0.4013**	-0.2550	0.2192
Precipitation	0.2007	0.1313	**0.8404**	-0.0793	0.2164
Natural resource	-0.0015	0.1065	**-0.5645**	-0.1158	0.2023
Preferential Policy Index	**0.4026**	-0.2636	**0.4611**	0.1530	0.2464
Index government intervention	-0.1803	-0.0973	**0.4912**	0.0732	0.2344
Index contract enforcement	-0.0607	-0.0289	0.1441	0.0550	0.2229
Index corruption	-0.1055	-0.0792	0.0290	-0.1117	0.2553
Index property protection	**0.3781**	-0.1775	0.0762	**0.5757**	0.1521
NERI institutions index	**0.6980**	-0.0394	**0.3581**	**0.4004**	0.1099
GRP per capita	**0.5124**	-0.3228	0.0354	**0.7091**	0.0666
Wage	**0.6450**	**-0.3610**	0.0219	**0.5300**	0.0816
Consumption household	**0.5207**	-0.3316	0.0753	**0.6983**	0.0865
Tech market transaction	0.3244	-0.0611	-0.0714	**0.8662**	0.1183
Minority population	-0.0698	-0.0431	0.0149	-0.2814	0.3903

No. of observations: 309

Kaiser-Meyer-Olkin Measure of Sampling Adequacy. : 0.859

Note: loadings with absolute values larger than 0.35 are marked as bold.

Table 3: Cronbach's alpha for factor consistency

Variables	Factor1	Factor2	Factor3	Factor4	Overall
Capital	0.6916	0.5115	0.6809	0.6025	0.7288
City road length	0.6915	0.5115	0.6808	0.6025	0.7287
City road area	0.6914	0.5116	0.6807	0.6022	0.7285
Civil vehicle	0.6916	0.5116	0.6809	0.6025	0.7289
freight	0.6911	0.5103	0.6768	0.6015	0.7278
Gov Expenditure	**0.5827**	**0.4500**	**0.5747**	**0.5570**	0.6591
Private vehicle	0.6916	0.5116	0.6809	0.6026	0.7289
Ways (train, water, highway)	0.6906	0.5111	0.6803	0.6007	0.7277
Exports	**0.6046**	**0.4503**	**0.5948**	**0.4806**	0.6549
Imports	**0.5709**	0.5045	**0.6060**	**0.4655**	0.6620
Long telephone	0.6727	0.5041	0.6707	0.5770	0.7175
Local telephone	0.6916	0.5115	0.6809	0.6025	0.7288
Mobile	0.6915	0.5115	0.6809	0.6024	0.7288
Cable	0.6913	0.5111	0.6807	0.6019	0.7286
Patent registered	0.6913	0.5115	0.6808	0.6023	0.7287
Higher education enrolment	0.6911	0.5110	0.6804	0.6020	0.7289
Higher education institutions	0.6911	0.5110	0.6804	0.6020	0.7202
Senior enrolment	0.6911	0.5110	0.6804	0.6020	0.7289
Senior high school	0.6913	0.5113	0.6806	0.6022	0.7153
Junior enrolment	0.6837	**0.2962**	0.6722	0.5947	0.7289
Junior high school	0.6916	0.5113	0.6809	0.6025	0.6979
Primary enrolment	0.6807	0.5691	0.6816	0.5973	0.7289
Primary school	0.6916	0.5093	0.6807	0.6025	0.7198
Population	0.6916	0.5115	0.6809	0.6025	0.7287
Workers	0.6916	0.5111	0.6809	0.6025	0.7288
Humidity	0.6790	0.5140	0.6636	0.5907	0.7289
Sunshine	0.6916	0.5116	0.6809	0.6021	0.7289
Temperature	0.6916	0.5116	0.6809	0.6019	0.7289
Area	0.6915	0.5116	0.6809	0.6012	0.7289
Precipitation	0.6916	0.5115	0.6809	0.6025	0.7242
Natural resource	0.6876	0.5059	0.6780	0.5919	0.7284
Preferential Policy Index	0.6916	0.5116	0.6809	0.6025	0.7286
Index government intervention	0.6878	0.5059	0.6733	0.6022	0.7286
Index contract enforcement	0.6916	0.5116	0.6809	0.6025	0.7284
Index corruption	0.6916	0.5116	0.6808	0.6025	0.7284
Index property protection	0.6911	0.5110	0.6804	0.6020	0.7284
NERI institutions index	0.6916	0.5116	0.6809	0.6025	0.7284
GRP per capita	0.6904	0.5103	0.6697	0.5992	0.7287
Wage	0.6913	0.5112	0.6796	0.6022	0.7288
Consumption household	0.6916	0.5115	0.6807	0.6025	0.7288
Tech market transaction	0.6916	0.5116	0.6809	0.6025	0.7199
Minority population	0.6916	0.5116	0.6809	0.6025	0.7289
Cronbach's alpha	0.6911	0.5112	0.6804	0.6021	0.7284

Table 4: The fixed effects and between estimation results

	Number of FDI firms (log)		Amount of FDI (log)	
Variable	(1) FE	(2) Between	(3) FE	(4) Between
'Institution' factor	0.130	1.059***	0.015	1.136***
	(0.081)	(0.204)	(0.107)	(0.208)
'Labor costs' factor	-0.312	0.225	-0.328	0.217
	(0.302)	(0.179)	(0.396)	(0.183)
'Market' factor	0.148*	0.721***	0.102	0.791***
	(0.077)	(0.209)	(0.095)	(0.213)
'Geography' factor	-0.016	0.496***	0.172	0.449**
	(0.130)	(0.156)	(0.201)	(0.159)

Note: Number of observations is 309. In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level.

Table 5: 2SLS IV estimation results

	Number of FDI firms (log)		Amount of FDI (log)	
Variable	(1) Basic	(2) Interactions	(3) Basic	(4) Interactions
'Institution' factor	0.087***	0.134***	-0.081	-0.073
	(0.029)	(0.046)	(0.065)	(0.075)
'Labor costs' factor	-0.999***	-0.771***	-1.229***	-1.160***
	(0.211)	(0.279)	(0.479)	(0.447)
'Market' factor	0.278***	0.181***	0.218***	0.146
	(0.037)	(0.063)	(0.086)	(0.099)
*Inst*Costs*		-0.033**		-0.081***
		(0.017)		(0.031)
*Inst*Market*		0.031**		0.019
		(0.014)		(0.023)
'Geography' factor	-0.046	-0.048	0.002	-0.015
	(0.074)	(0.074)	(0.127)	(0.113)
Joint significance		'Labor costs': S 'Market': S 'Institution': S		'Labor costs': S 'Market': S 'Institution': S
H0: Exogeneity	p-val.: 0.0368	p-val.: 0.0773	p-val.: 0.0000	p-val.: 0.0000

Note: Number of observations is 241. In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level. In the joint significance test 'S' denotes for 'significant' and 'I' denotes for 'insignificant'.

Table 6: Arellano-Bond dynamic panel-data estimation results

	Number of FDI firms (log)		*Amount of FDI (log)*	
Variable	(1) Basic	(2) Interactions	(3) Basic	(4) Interactions
$Dependent_{t-1}$	0.167**	0.144**	0.231**	0.248***
	(0.080)	0.092	(0.100)	(0.097)
'Institution' factor	0.080*	0.158***	-0.024	-0.003
	(0.046)	0.059	(0.065)	(0.086)
'Labor costs' factor	-0.808**	-0.555**	-0.350	-0.352
	(0.334)	0.284	(0.345)	(0.249)
'Market' factor	0.257***	0.123	0.085	0.051
	(0.063)	0.083	(0.065)	(0.088)
*Inst*Costs*		-0.035		-0.027
		0.039		(0.044)
*Inst*Market*		0.045**		0.019
		0.020		(0.027)
'Geography' factor	-0.082	-0.028	0.154	0.165
	(0.111)	0.1144	(0.155)	(0.163)
Joint significance		'Labor costs': S 'Market': S 'Institution': S		'Labor costs': I 'Market': I 'Institution': I
No. of instruments	113	147	113	147
AR(2)	p-val.: 0.4595	p-val.: 0.3767	p-val.: 0.5327	p-val.: 0.6611
Sargan test	p-val.: 0.8774	p-val.: 0.7473	p-val.: 0.8519	p-val.: 0.7922

Note: Number of observations is 242. In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level. In the joint significance test 'S' denotes for 'significant' and 'I' denotes for 'insignificant'.

Table 7: Subsample IV estimation results

	Number of FDI firms (log)		Amount of FDI (log)	
Variable	(1) East	(2) West	(3) East	(4) West
'Institution' factor	0.481***	0.001	-0.179***	-0.023
	(0.115)	(0.057)	(0.065)	(0.076)
'Labor costs' factor	-1.267***	-0.933**	-1.798***	-1.022***
	(0.232)	(0.428)	(0.319)	(0.383)
'Market' factor	0.400***	0.102	0.365***	0.626**
	(0.064)	(0.241)	(0.091)	(0.275)
'Geography' factor	-0.008	-0.376***	0.192	-0.693***
	(0.087)	(0.112)	(0.133)	(0.164)

Note: Number of observations is 111 for the East and 130 for the West. In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level.

Table 8: Subsample IV estimation results with interactions

	No. of FDI firms (log)		Amount of FDI (log)			
Variable	(1) East	(2) West	(3) East	(4) East	(5) West	(6) West
'Institution' factor	0.073	0.599	0.083	-0.098	-0.085	0.257
	(0.092)	(0.460)	(0.087)	(0.120)	(0.092)	(0.180)
'Labor costs' factor	-1.149***	0.105	-1.123***	-1.598***	-1.011***	-0.372
	(0.328)	(0.898)	(0.355)	(0.459)	(0.343)	(0.585)
'Market' factor	0.210**	0.834	0.124	0.468***	0.455**	0.838**
	(0.104)	(0.603)	(0.092)	(0.106)	(0.230)	(0.345)
*Inst*Costs*	-0.088**	-0.129	-0.207***		0.057	
	(0.039)	(0.103)	(0.047)		(0.040)	
*Inst*Market*	0.054	-0.199		-0.064		-0.121
	(0.048)	(0.1500)		(0.060)		(0.074)
'Geography' factor	-0.076	-0.338**	0.099	0.225*	-0.679***	-0.651***
	(0.082)	(0.140)	(0.115)	(0.128)	(0.154)	(0.150)
Joint significance	'Labor costs': S 'Market': S 'Institution': S	'Labor costs': S 'Market': I 'Institution': I	'Labor costs': S 'Institution': S	'Market': S 'Institution': S	'Labor costs': S 'Institution': I	'Market': S 'Institution': I

Note: Number of observations is 111 for the East and 130 for the West. In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level. In the joint significance test 'S' denotes for 'significant' and 'I' denotes for 'insignificant'.

Table 9: Results without the four municipalities

Variable	IV estimation		Arellano-Bond dynamic	
	(1) Number FFE	*(2) Amount FDI*	*(3) Number FFE*	*(4) Amount FDI*
Dependent$_{t-1}$			0.169**	0.236**
			(0.080)	(0.099)
'Institution' factor	0.235**	0.151	0.051	-0.011
	(0.108)	(0.134)	(0.123)	(0.169)
'Labor costs' factor	-0.794**	-1.073***	-0.858**	-0.498
	(0.346)	(0.415)	(0.348)	(0.319)
'Market' factor	0.624***	0.811***	0.202	0.182
	(0.194)	(0.271)	(0.291)	(0.385)
'Geography' factor	0.022	0.139	-0.122	0.141
	(0.097)	(0.139)	(0.128)	(0.168)
No. of instruments			113	113
AR(2)			p-val.: 0.4732	p-val.: 0.6028
Sargan test			p-val.: 0.9301	p-val.: 0.9295
No. of obs.	216	216	217	217

Note: In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level. Dependent variables are in log form.

Table 10: Specified Factor 1 ('Institutional quality')

Variables	Factor 1 ('Infrastructure')	Factor 2 ('Governance')	Uniqueness
City road length	0.8459	0.2241	0.2343
City road area	0.9151		0.1328
Freight	0.7893		0.3644
Ways	0.6738	-0.5293	0.2658
Long telephone	0.9272		0.1395
Local telephone	0.9661		0.0625
Mobile	0.9367		0.1221
Cable	0.6854	-0.5204	0.2594
Patent	0.8509	0.2669	0.2047
NERI index	0.7276	0.4880	0.2324
Index property protection	0.3223	0.8060	0.2465
Index government intervention		-0.2028	0.9474
Index corruption		-0.3620	0.8376
Index contract enforcement		-0.3276	0.8921
PPI (Preferential Policy Index)	0.3601	0.5579	0.5590
Minority population		-0.6655	0.5433

Note: blanks represent loadings with absolute values below than 0.2; Number of observations is 309.

Table 11: IV estimation results of using sub-factors of institutions

	Number of FDI firms (log)		*Amount of FDI (log)*	
Variable	(1) Basic	(2) Interactions	(3) Basic	(4) Interactions
'Infrastructure' factor	0.157***	0.210**	-0.061	0.131
	(0.048)	(0.103)	(0.077)	(0.135)
'Governance' factor	0.137	-0.028	0.238	0.219
	(0.156)	(0.309)	(0.246)	(0.369)
'Labor costs' factor	-0.934***	-0.794**	-1.189***	-1.182***
	(0.231)	(0.324)	(0.348)	(0.407)
'Market' factor	0.244***	0.434**	0.161*	0.588**
	(0.056)	(0.193)	(0.094)	(0.253)
*Infra*Costs*		-0.123*		-0.125
		(0.074)		(0.087)
*Gov*Costs*		-0.200		-0.112
		(0.171)		(0.204)
*Infra*Market*		0.009		-0.011
		(0.023)		(0.034)
*Gov*Market*		-0.099		-0.208***
		(0.064)		(0.077)
'Geography' factor	-0.056	0.019	-0.036	0.053
	(0.092)	(0.126)	(0.138)	(0.159)
Joint significance		'Infrastructure' yes		'Governance' yes

Note: Number of observations is 241. In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level.

Table 12: Estimation results of using alternative dependent variables

Variable	(1) *FDI inflows*	(2) *Capital FFE*	(3) *Employees FFE*	(4) *FDI factor*
'Institution' factor	-0.205	-0.0611	0.1767***	0.489***
	(0.291)	(0.039)	(0.041)	(0.078)
'Labor costs' factor	-4.285**	-1.262***	-0.359*	-1.219***
	(1.953)	(0.293)	(0.215)	(0.449)
'Market' factor	0.3524*	0.245***	0.126***	0.443***
	(0.188)	(0.051)	(0.048)	(0.088)
'Geography' factor	-0.289	-0.055	-0.200**	0.209**
	(0 .254)	(0.096)	(0.083)	(0.098)
No. of obs.	162	241	234	241

Note: In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level.

Table 13: Benchmark IV results (dependent Variable: *log Number FFE*)

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Government expenditure (log)	-2.893*	-3.556		-4.021	-3.782	-1.548
	(1.626)	(2.309)		(3.406)	(3.290)	(2.731)
Ways (log)			-0.07			0.021
			(0.513)			(1.132)
Wage (log)	6.159*	5.558	2.334		7.262	-5.066
	(3.449)	(3.497)	(1.990)		(4.777)	(6.581)
GRP per capita (log)				15.174*		13.657
				(8.847)		(11.791)
GRP (log)	0.851	0.828	0.255	-9.14		-12.415
	(0.922)	(0.993)	(0.883)	(6.190)		(11.818)
Consumption household (log)					-0.008	5.722
					(1.145)	(6.642)
Natural resource	0.006	0.014	-0.000*	0.031	0.011	-0.005
	(0.007)	(0.012)	(0.000)	(0.024)	(0.013)	(0.012)
Primary school enrolment (log)		0.071	-0.127	1.652	0.187	2.143
		(0.319)	(0.266)	(1.188)	(0.333)	(2.322)
Higher education enrolment (log)		0.342	-0.013	-1.639	0.316	-2.141
		(0.288)	(0.197)	(1.253)	(0.357)	(2.079)
NERI index		0.201*	0.096	0.483*	0.217	0.324
		(0.109)	(0.085)	(0.261)	(0.142)	(0.278)
Preferential policy		0.286	-0.015	0.404	0.285	0.100
		(0.230)	(0.091)	(0.389)	(0.311)	(0.213)
No. of obs.	302	302	280	302	308	274

Note: In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level.

Table 14: IV results of using new factors and controls (dependent Variable: *log Number FFE*)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
'Institution' factor (new)	0.119***	0.105***	0.065**	0.064**	0.073**	0.057*	0.080**	0.068**	0.057*	0.075
	(0.021)	(0.020)	(0.033)	(0.032)	(0.036)	(0.036)	(0.031)	(0.031)	[0.030]	[0.047]
'Labor costs' factor (new)	-0.872***	-0.990***	-0.609***	-0.516***	-0.455***	-0.432***	-0.711***	-0.341**	-0.648***	-0.491**
	(0.137)	(0.152)	(0.146)	(0.139)	(0.137)	(0.144)	(0.199)	(0.164)	[0.124]	[0.247]
'Market' factor (new)	0.273***	0.286***	0.229***	0.203***	0.245***	0.220***	0.203***	0.093*	0.263***	0.220***
	(0.035)	(0.036)	(0.039)	(0.038)	(0.049)	(0.048)	(0.045)	(0.055)	[0.038]	[0.074]
'Geography' factor (new)	-0.024	-0.006	-0.111	-0.109*	-0.119*	-0.110	-0.158**	-0.236***	-0.029	-0.02
	(0.076)	(0.077)	(0.070)	(0.065)	(0.072)	(0.070)	(0.068)	(0.054)	[0.083]	[0.081]
Exports (log)				0.129		0.132				
				(0.098)		(0.096)				
Imports (log)				0.086		0.081				
				(0.068)		(0.067)				
Government exp (log)		0.545***				0.318**				
		(0.185)				(0.167)				
Primary enrolment (log)										-0.133
										[0.118]
Junior high enrolment (log)										-0.062
										[0.174]
Preferential policy								0.082*		
								(0.050)		
NERI index								0.150***		
								(0.050)		
Institutional index 2000								0.764***		
								(0.120)		
No. of Obs.	241	238	233	233	233	230	247	233	241	241

Note: In parentheses are standard errors. *** Statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level. *Institutional index 2000* is an aggregated factor based on *Index government intervention*, *Index contract enforcement*, *Index anti-corruption*, and *Index property protection* using principal-component method.

Appendix

Dependent variables	*Explanation*
Number of FDI firms	Number of Foreign Funded Enterprises (unit)
Amount of FDI	Total Investment by Foreign Funded Enterprises (100 million USD)
FDI inflows	Actually utilized FDI (10 000 USD) (until 2003)
Capital FFE	Registered capital of Foreign Funded Enterprises (100 million USD)
Employees FFE	Number of Employed person in Foreign Funded units (10 000persons)
Explanatory variables	
Exports	Total Exports (10 000 USD)
Imports	Total Imports (10 000 USD)
Ways (train, water, highway)	Length of Railways in operation (km) Length of Navigable Inland waterways (km) Total Length of Highways (km)
Capital	*Gross Capital Formation (100 million yuan)*
City road length and area	Length of Paved Roads (capital city year-end) (km) Area of Paved Roads (capital city year-end) (10 000 sq.m)
Civil vehicle	Possession Civil Vehicle (10 000 units)
Private vehicle	Possession of Private vehicles (10 000 units)
Electricity consumption	Electricity Consumption (100 million kwh)
Freight (train, water, highway)	Freight Railways (10 000 tons) Freight Highways (10 000 tons) Freight Waterways (10 000 tons)
Long telephone	Capacity of Long-distance Telephone Exchanges (circuit)
Local telephone	Capacity of Local Office Telephone Exchanges (10 000 line)
Mobile	Capacity of Mobile Telephone Exchanges (10 000 subscribers)
Cable	Length of Long Distance Optical Cable Lines (km)
Gov expenditure	Government Total Expenditure (10 000 yuan)
NERI index	NERI index of marketization (0-10) (Fan et al., 2007) NERI index dimension: government and market; development of non-state enterprises; development of commodity market; development of factor market; development of market intermediaries and legal environment
Patent	Number of Patent Application Granted (Piece)
Gross regional product	Gross Regional Production (100 million yuan)
Higher education institutions	Number of Higher Education School (units)

Senior high enrolment	Total enrolment of Senior High (person)
Senior high school	Number of Senior High School (unit)
Junior high enrolment	Number of Junior High School (unit)
Junior high school	Total Enrolment of Junior High School (person)
Primary enrolment	Total enrolments Primary School (person)
Primary school	Number of Primary School (unit)
Workers	Total Number of Staff and Workers (10 000 persons)
Population	Total Population (10 000 persons)
GRP per capita	Per capita Gross Regional Product (yuan)
Consumption household	Household consumption expenditure (yuan)
Tech market transaction	Transaction Value in Technical Market (10 000 yuan)
Wage	Average wage (yearly yuan)
Natural resource	petroleum (10 000 tons), natural Gas (100 million cu.m), coal (100 million tons), iron (Ore, 100 million tons)
Area	Area size of province (square kilometres)
Sunshine	Annual Total Sunshine hours of capital cities (hours)
Temperature	Annual Average Temperature of capital cities (C)
Humidity	Annual average relative humidity (capital city %)
Precipitation	Annual Total Precipitation of capital city (mm)
Institution index 2000	Institution Index 2000 (Du et al., 2008): intellectual property rights protection, government intervention in business operation, government anti-corruption, contract enforcement
PPI (Preferential Policy Index)	Index of Provincial Preferential Policy (0-4)