CESifo WORKING PAPERS

7979
2019

November 2019

Estimating the Local Effectiveness of Institutions: A Latent-Variable Approach

Antonia Reinecke, Hans-Jörg Schmerer

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7979*
Category 12: Empirical and Theoretical Models

Estimating the Local Effectiveness of Institutions: A Latent-Variable Approach

Abstract

This paper develops an approach that allows constructing regional proxies of government effectiveness at a highly dis-aggregated level. Our idea builds on the well documented interdependence between institutions and exports, which allows estimating the latent government effectiveness using methods of structural equation modeling. Unobserved institutional quality for each individual region is predicted using the regression outcomes. The impact of this novel identification strategy is tested using various panel applications. Results show that the magnitude of the effect of institutional quality can be biased in estimates that neglect regional differences in the effectiveness of institutions.

JEL-Codes: D020, F100, P160.

Keywords: trade, institutions, Chinese plant data, latent variable, structural equation modelling.

Antonia Reinecke
FernUniversität in Hagen / Germany
antonia.reinecke@fernuni-hagen.de

Hans-Jörg Schmerer
FernUniversität in Hagen / Germany
hans-joerg.schmerer@fernuni-hagen.de

1. Introduction

Legal security is one important determinant of firm performance but the enforceability of institutions varies across regions. This regional disparity in government efficiency is supposed to be more pronounced in developing countries where bureaucrats' earnings are lower and where corruption in general is more difficult to control.

Most empirical applications can deal with these systematic and unobserved differences in institutional quality or government efficiency by using region or firm fixed effects. However, this solution is inappropriate in the analysis of institutions itself. This latter type of analysis must build upon proxies that capture the differences in government efficiency across regions but no such data is waiting in wings.

We propose using the well known link between exports and firm productivity to uncover a detailed proxy for local effectiveness of institutions on a highly dis-aggregated level of zip-code regions. More productive firms are more inclined to export, which is reflected in a higher export intensity. This intuitive explanation can be rationalized by an asymmetric Melitz (2003) model featuring multiple export destinations. If exporting fixed costs are heterogeneous across different destinations, heterogeneous market entry costs and sorting into those markets according to productivity helps explaining why more productive firms have higher export intensities. They are able to serve a larger number of foreign destinations at the extensive margin. A higher government efficiency should aggravate or boost firm productivity associated with a positive link between government efficiency and the respective location's firms export intensity. There is a large and emerging literature on this hypothesized link[1] but exiting studies are similar with respect to the level of aggregation: Government efficiency, the degree of anti-corruption legislation, and contract enforcement are usually measured at a

[1] Anderson and Marcouiller (2002) estimate the effect of institutional quality on trade. Uncertainty induced by unpredictable government actions is a hidden tax for the firms operating in the respective market. Reinecke and Schmerer (2017) find a negative effect of low government efficiency, as a determinant of institutional quality, on the positive relation of firm size and export volume. De Groot et al. (2004) show that the degree of homogeneity in institutions introduces some security among trading partners associated with soaring export volumes between them. Similarities in business proceedings can reduce transaction cost, which has a stimulating effect on trade. Ranjan and Young Lee (2007) show that the quality of contract enforcement has a positive effect on trade volumes. Gani (2017) identifies a significant and positive relation between contract enforcement improvement and trade volumes by analyzing time and financial expenditures to enforce contracts.

highly aggregated country- or province-level. We are using these measures to capture differences in institutional quality across broader regions. Moreover, we argue that these measures mainly reflect the situation in the core area around the capital of a province, which is defined as political hub in our study.

A higher level of government efficiency in one province benefits all firms located within the respective location but the effect likely deteriorates with distance to the political and/or economic hub of the province. The better infrastructure and the higher well being of inhabitants in the capital likewise attracts more eloquent political leaders and more productive firms. The opposite argument applies for regions more remote to the capital of the province, which gives rise to assortative matching between firms and local officials. The attractiveness of a destination deteriorates with the distance to the capital as access to the facilities provided in the more densely populated capital are more and more difficult to access. Moreover, the higher remoteness gives less powerful officials more room for corruption. The distance to the capital elevates the incentive to take bribes through a lower probability of being detected. Surveillance in more remote places is difficult and more costly and wages are usually lower. Distance to the capital is the crucial variable that allows estimating a discount factor of the effectiveness of a government. The intuitive explanation is taken from the "New Economic Geography" literature. Economies of scale, transportation cost and market size are important determinants of agglomeration (e. g. Krugman (1991), Ottaviano et al. (2002), Henkel et al. (2018)). While low transportation cost promotes agglomeration, low economies of scale and/or high product differentiation tend to be forces of dispersion.[2] Moreover, the proximity to local party leaders, and thereby to the policy-makers of the respective region, may be an agglomeration force in itself. Political connections can be an important source of success. Therefore, firms may want to move closer to the

[2]Housing markets and agglomeration are also related to each other. Allen and Arkolakis (2014) introduce a share of income paid for fixed local factors. These cost of living can differ across regions and are affected by agglomeration: if a region is preferred over another region and therefore attract migrants, the price for the immobile factor increases. Henkel et al. (2018) additionally implement transfers between regions and find that if a region is recipient of transfer payments, this region becomes more attractive due to a higher provision of public goods. This in turn leads to migration from the donor to the recipient region. In the Chinese case it can be expected that cluster policies favoring particular regions attract migrants which in turn increases prices for housing and land.

political hubs in order to maintain their network to stronger policy makers.

For the research question at hand, we focus on the role of variable transportation and fixed living cost. In compliance with the previous discussion, low transportation cost in an area accelerates agglomeration, which provides several advantages for firms and workers: a large local market, a high labor supply and a high concentration of human capital as well as a rich set on information about policy measures due to the proximity to the local government.

However, agglomeration also provides disadvantages, such as ecological damage, shortage of public goods, high competition, missing capacity of space and high land and housing prices. Especially, the last three issues should work against the agglomeration forces. Agglomeration pulls the firms closer to the economic hubs of a province but the higher fix cost should push them towards the periphery. The optimal choice of the location should be close but not too close to the important hubs of the respective location. This consideration gives rise to a hump shaped relationship between distance from the respective hubs and export intensity. Thus, the effect of distance on government efficiency is expected to be quadratic.

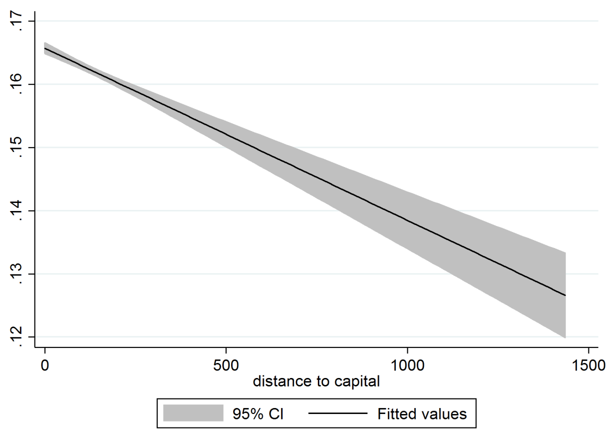
First Glimpse at the data. Our novel approach is implemented for China, where regional heterogeneity matters a lot. The differences in regional development prevail across and within regions.[3] The political landscape of China is characterized by policy clusters across narrowly defined areas and those particular policies are strongly determined by local party leaders, albeit the huge efforts that were made to reduce inequality across regions. Political reforms were usually implemented in specific regions and particular industries. For example, in Guangdong the *Zujiang Delta* (Pearl River Delta) describes a spatially delimited area where export-processing and equipment supply were favored by the government. The economic performance in this region is mainly driven by foreign investors. In contrast, the *Wenzhou Model* in Zhejiang promotes light industries dominated by local, private owned enterprises. In opposition

[3]Jefferson et al. (2008) provide empirical evidence that especially productivity differentials still exists between the span 1998 - 2005. Even if the authors are able to identify a substantial catching-up of the central with the coastal regions, west and northeast productivity levels still lagging far behind those of the coastal region.

to those contrasting development strategies, the *Sunan Model* in Jiangsu emphasizes the role of local governments investing into state owned enterprises located in Jiangsu. This clustered political landscape indicates that local party leaders have a high impact on shaping the institutional environment within a region. The academic literature confirms that these diverging development strategies in different areas led to high economic and institutional disparities across and within provinces (e. g. Xinyue and Wei (2005), Wei and Kim (2002), Gu et al. (2001)). The choice to support certain regions is also driven by the comparative advantage determined by the proximity to certain economic hubs (e. g. Hong Kong, Shanghai, or the respective provincial capital). To attract further private and foreign capital since the 1990's and shortly before China's WTO entry in 2001, regional governments had an incentive to improve institutional quality at locations with higher density of strong firms. Good institutions promote trade, and especially exports were one important driver of growth during the reforms pushed by the government.

Our analysis builds on a province-level index of government efficiency, which is one of the main determinants regarding the allocation of legal systems, public services, public infrastructure and reasonable decision making. Thus, it covers important aspects of institutional quality in China. Based on this information, we use the proximity to the provincial capital as discount factor for local government efficiency: Decisions about certain policy measures are made in the capital of a province and we argue that areas closer to the capital tend to be characterized by higher government efficiency than areas located in more rural areas, far away from the provincial capital. City-specific efficiency of local governments decreases in distance between the respective firm and the provincial capital due to a higher probability of ill-functioned control mechanisms (Huang et al. (2017)). Secondly, we argue that firms' export behavior is determined by the government's efficiency and therefore comprises some information on institutions at the firm's location. The higher the proximity between a firm's location and the provincial capital, the higher the expected government efficiency associated with this area and, ceteris paribus, the higher is the expected export share of the observed firm. This variable is endogenous as the export share likely determines the effort of the central government to promote good institutions through export promoting policies. This endogeneity bias can be uncovered using firm-level data containing highly detailed information on exports, location and other firm characteristics. Figure 1 illustrates the correlation between the export share of a firm and the distance to the capital of the province.

Figure 1: *Correlation between export volume and distance to provincial capital*



The stylized fact presented in Figure 1 supports the theoretical considerations sketched in the introduction. Firms located closer to the provincial capital tend to export more than firms located at places more remote from the capital of the province. Distance to the capital of a province matters for the export intensity. The remainder of this paper is structured as follows: section two presents the empirical strategy, section three describes the data used for the empirical analysis. Results are presented and discussed in section four and section five concludes.

2. Empirical Strategy

Suppose the following model, which explains a dependent variable on the firm-level using a constant, β_2, government-efficiency, GE, on the zip-code-level denoted by index z, and an additional vector of control variables on the firm-level, CV_i.

$$y_i = \beta_2 + \gamma_1 GE_z + \gamma_2 CV_i + \epsilon_{i1} \quad (1)$$

All variables other than GE_z are observables for which data is available in our empirical analysis. The unobserved variable GE_z is treated as latent variable in our setting. We assume that local government efficiency is a function of provincial government efficiency, GE_p, as well as the distance between the firm and the provincial capital, $ln(dist)$:

$$GE_z = \beta_1 + \lambda_1 GE_p + \lambda_2 ln(dist) + \epsilon_{i2} \tag{2}$$

The model translates the cross-province differences in government efficiency into more detailed measures at the zip-code level by taking a distance discount/premium into consideration.[4] Coefficients are estimated by fitting equation 2 into 1. The resulting equation solely depends on observable variables. Broader measures of government efficiency at the province level are available and can be merged to our enterprise data using the province identifier. This approach generates a multi-level data set with information on both the firm- and the province-level. The available address of the firm adds useful information on the firms' exact zip-code location, which can be used to identify the distance between the firm and the important economic and political hubs.

Substituting GE_z in equation (1) yields

$$y_i = \beta_2 + \gamma_1(\beta_1 + \lambda_1 GE_p + \lambda_2 ln(dist) + \epsilon_{i2}) + \gamma_2 CV_i + \epsilon_{i1} \tag{3}$$

The firms' export share is used as dependent variable for reasons discussed in the introduction. Among many other factors, the export intensity should also contain some useful information about the institutional environment at the firm's residence. However, the coefficients in equation (3) are interacted with each other, which prevents separate identification of both coefficients in one equation. A structural equation modeling approach is needed to solve the following three different equations simultaneously

[4] Depending on the sign of λ_2, distance adds a discount or a premium.

$$Sales_{i(z)t} = \beta_0 + \alpha_1 \times Controls + \alpha_2 \times GE_{(i)zt} + a_{it} \quad (4)$$

$$(exp/output)_{i(z)t} = \beta_2 + \gamma_2 \times Controls + \gamma_1 \times GE_{(i)zt} + c_{it} \quad (5)$$

$$GE_{(i)zt} = \beta_1 + \lambda_2 \times Distance + \lambda_1 \times GE_{(i)pt} + b_{it} \quad (6)$$

The first equation captures the link between government efficiency and sales. We argue that firms located in areas with higher efficiency have better access to national markets and therefore higher sales. This equation is important for identification as it is necessary to define starting values for the latent variable. Therefore, the coefficient α_2 is normalized to unity in this first equation and all other coefficients must be interpreted relative to α_2.

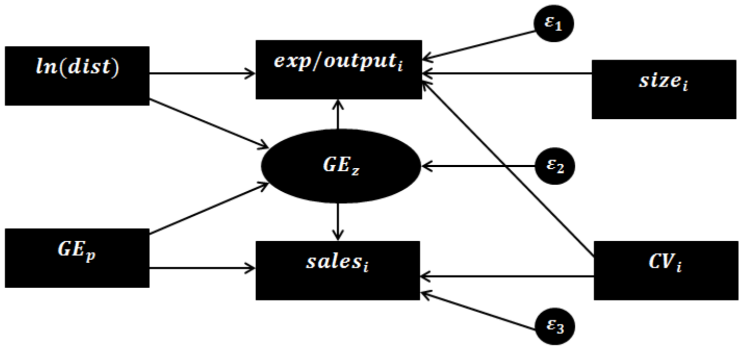
The second equation is the main equation containing valuable information about the latent variable, which is isolated from other effects by putting as many relevant controls as possible. Thus, the variables on the right clean up the error term by purging it from potential effects that are not related to government efficiency. Various distance measures are computed by geo-coding the data. The coefficients associated with those measures are indirect measures of transportation costs. Included distance measures comprise the distance to the capital of the province, the distance to the coast, all distances to the relevant harbor cities and other important destinations for Chinese exporters.

The third equation predicts the latent variable based on information for overall government efficiency at the province level and the distance times the distance discount/premium factor. Indices i, z and p reflect the nested structure of the estimation strategy that consists of firms, i, that are located in zip-code areas, z, that in turn are located in a specific province, p. The first index without parentheses indicates level of data availability. Index t specifies the time period. Each equation has its own independent error term: a_{it}, b_{it} and c_{it}.

The identification strategy can be summarized by the following path diagram:

Government Efficiency at province level, GE_p, distance between the zip-code area of the firm's location and the provincial capital, $ln(dist)$, and the remaining control

Figure 2: *Path diagram - structural equation modeling*



variables are treated as given. Values of the predicted latent variable are computed conditional on the explanatory variables and their coefficients. Hence, the assumption of joint normality conditional on the explanatory variables is sufficient. To handle non-normality of estimated standard errors those parameters are estimated using quasi-maximum likelihood.

Application. The gains of using the more detailed measure of government efficiency are demonstrated in a comparison of the old and the new index in a panel data regression analysis. We investigate the role of government efficiency for Total Factor Productivity (TFP) in our Chinese firm level data. The coefficients associated with the more common proxy of government efficiency on the province level are compared to the coefficients obtained from regressions that include the more detailed GE variable. The estimated model reads

$$y_{i(r)t} = \alpha + \beta_1 GE_{rt} + \beta_2 CV_{i(r)t} + \nu_t + \mu_i + \eta_p + \epsilon_{it} \quad . \tag{7}$$

The dependent variable $y_{i(r)t}$ is firm-level TFP observed for firm i located in region r at time t. All variables other than government efficiency are uniquely identified by the firm identifier i, which is the reason why we put the region index r into parentheses. GE_{rt} denotes the respective government efficiency measure, where index r distinguishes between the province and the zip-code region using p for province and z for zip-code level. Moreover, we include a vector of control variables, CV_{it}. Controls for the time trend are included by inserting year dummies, ν_t, in each specification. Provincial or firm-specific unobserved heterogeneity is controlled for by including province (η_p) or firm (μ_i) fixed effects when indicated in the output table. Robust standard errors are clustered at the province or zip-code level depending on the level of aggregation of the government efficiency measure. The error term is captured by ϵ_{it}.

3. Data

The analysis is based on a panel of firms covering the years 2001 to 2006. All firm-level information comes from the NBS, which surveys the universe of manufacturing firms with minimum yearly turnover of 5 Million RMB and all state owned enterprises (SOEs). The original data set is a repeated cross-section covering the years 1998 to 2006. Earlier years drop out of the sample because of missing information for government efficiency at the province level. Brandt et al. (2014) note working with Chinese data can be challenging, especially, the comparability of the data over time. The authors identify four main issues: first, unique identification of firms to build a consistent panel. Indeed, firms receive a distinct firm ID when they enter the sample but if firms are restructured, the firm is assigned to a new ID. Particularly, in the observed time period a significant volume of firms were restructured and privatized. Second, constructing detailed price deflators to obtain comparability of nominal variables over time. Brandt et al. (2012) provide detailed information how they solve the issue of constructing price deflators.[5] The capital stock in the Chinese data is measured based on three different fixed assets, which is not very common, and represents the third identified issue.[6] In Brandt et al. (2012) the authors show how they estimate the real value of the capital stock in the first year a firm reports capital. Based on this estimation, the authors construct the real capital stock by combining information on firms' age and the investment deflator provided by Brandt et al. (2008). Issue four discussed by Brandt et al. (2014) is the consistency of variable definition over time. For instance, the authors

[5] Brandt et al. (2012) build a price index by calculating the ratio between nominal and real prices. The at a four-digit industry level averaged changes between two years are used to calculate weighted averages for every sector, whereby the weight equals the current firm's output. Linking these changes over time leads to a consistent output deflator for 423 sectors. The data for the years 2004-2006 is deflated by using the ex-factory price index provided by the 2007 Chinese Statistical Year Book.

[6] These three assets are: original fix assets, net fix assets and total fix assets.

evolve a harmonized industry classification due to changes in the Chinese Industry Classification (CIC) in 1994, 2002 and 2011. Moreover, five consistent ownership types are defined by Brandt et al. (2012) (the original data set distinguishes 23 types of ownership). In addition to these adjustments we use the total factor productivity (TFP) measure constructed by Brandt et al. (2012). The benchmark TFP is constructed as Törnquist index number which allows technology heterogeneity in input elasticities across firms.[7] Overall, we have 1,728,740 observations during the whole period, where the number of firms covered by our panel varies between 146,106 in 1999 and 278,739 in 2006. Due to some inconsistency in the data, we drop duplicates and firms with birth year earlier than 1850 or later than 2006.

The *Chinese provincial government efficiency index* is taken from Tang et al. (2014). Their data cover 31 Chinese provinces[8] in a time span ranging from 2001 to 2010. The index is composed of four sub-indexes. First, government public service, containing information about public expenditures in different public sectors, such as education, healthcare and social security. Second, government public infrastructure that is composed of investment and efficiency measures concerning public infrastructure. Third, size of the government which includes in particular relations to non-public population and/or employment and spending. Last component is the residents' economic welfare comprising information of per capita income, consumer price index and the Engel coefficient differentiated by urban and rural residents.[9] Hence, the index covers a wide range of characteristics determining government efficiency. The raw data that was used to calculate the index stems from diverse Chinese Statistical Year Books. Tang et al. (2014) adopt the calculation methods of the International Institution for Management (IMD) by computing averages and standard deviations of the respective indexes. The results are weighted and standardized. Therefore, the index ranges from -1 to $+1$,

[7]For more detailed information about the data calculations we refer to Brandt et al. (2012) and Brandt et al. (2014).

[8]Anhui, Beijing, Chongqing, Fujian, Gansu, Guangdong, Guangxi, Guizhou, Hainan, Hebei, Heilongjiang, Henan, Hubei, Hunan, Inner Mongolia, Jiangsu, Jiangxi, Jilin, Liaoning, Ningxia, Qinghai, Shaanxi, Shandong, Shanghai, Shanxi, Sichuan, Tianjin, Tibet, Xinjiang, Yunnan, Zhejiang.

[9]Overall, the four main indexes consists of several sub-indexes: Government public services is composed of 24 indexes, government public infrastructure is calculated based on 11 indexes. Residents' economic welfare is premised on 5 indexes and size of the government is calculated based on 7 indexes. For more detailed information see Appendix I.

where a value of -1 is associated with the lowest level and $+1$ with the highest level of government efficiency.

The zip-codes reported in the NBS firm-level data are transformed into coordinates of the respective firm's residence area by using an online program to geocode the zip-codes given in the data. Second, we use the software QGIS, a Geographic Information System (GIS), to create different point layers where each pair of coordinates represents a firm location, and thereby provide the necessary information to calculate a distance matrix. We calculate linear distances between the corresponding firms' coordinate and the capital of the province or other economic/political hubs grounded on a coordinate reference system with EPSG Code 3857.[10] As output we get a point vector layer comprising the distance calculation for each input feature which we are able to merge to our enterprise and government efficiency data. Descriptive statistics can be found in Tables 7 and 8 reported in the Appendix of this paper.

4. Results

Column (4) of Table 1 presents the benchmark results. The first panel reports the coefficients estimated by fitting the first equation to the data. Firm sales are explained by some firm controls and the latent variable. As argued before, the purpose of this exercise is normalizing the coefficient associated with the latent variable in this first equation.

The second panel reports coefficients associated with the main equation. The share of export is explained by a large set of control variables and the latent variable GE_z. Distance controls pick up the effects associated with transportation costs. We control for the distance to the capital for reasons discussed in the introduction. The capital of the province is an important domestic market and transport hub to the rest of China. Moreover, Chinese firms export significant shares of their output to Taiwan, Hong Kong and Japan. The respective distances are included in the main equation as well. Distances to the main harbors[11] in China are controlled for using interactions with a

[10] This kind of reference coordinate system is also used by online maps suppliers, such as google maps or OpenStreetMap.

[11] These harbors are: Port of Shanghai, port of Shenzen, port of Ningbo, port of Guangzhou and port of

catchment dummy that takes the value one if the respective port is relevant for the firm and zero if the distance exceeds a certain threshold of the average distance between the respective harbors. The same procedure is applied for the distance control to the east coast: if the shortest distance exceeds 250km, the catchment area dummy takes a value of zero, otherwise it equals one.

We control for firm characteristics as the capital to labor ratio, the birth date of the firm, ownership dummies and firm size category dummies.

The latent variable GE_z enters the first and the main equation through a third equation that models the detailed government efficiency index as a function of GE_p and distance. The functional form is expected to be a linear transformation of the province level measure and some discount factor. The estimated coefficients that characterize the relationship between the latent variable GE_z, provincial government efficiency and distance are reported in the last panel of the outcome table.

Columns (1) to (3) show other specifications with less controls. The results are robust.

However, the third panel shows that government efficiency at the province level is significantly correlated with GE_z but distance is insignificant. One interpretation of this result is that the government efficiency at the province level is the best predictor of government efficiency at the zip-code level. The effect of distance is zero. As discussed in the introduction, we argue that the relationship may be non-linear.

Coefficients of the control variables. Our results indicate that larger firms tend to export more, represented by the highly significant and positive estimates of the *size group* dummies. The impact of a relative capital intensive production technology on *sales* is positive but insignificant. Similarly, the effect of the capital-labor ratio on a firm's export share is also insignificant but negative. This result is counter intuitive: Firms producing with relatively more capital are expected to be the more productive ones characterized by higher sales and higher export volumes but China is labor abundant. Exporting firms should be most efficient using a labor intensive technology.

Quingdao.

Table 1: *GSEM benchmark (collapsed)*

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
sales				
capital/labor	0.006 (0.01)	0.007 (0.01)	0.007 (0.01)	0.007 (0.01)
GE_z	1.000	1.000	1.000	1.000
constant	9.892***(0.03)	9.907***(0.03)	9.907***(0.03)	9.907***(0.03)
export/output				
ln(dist)	0.011***(0.00)	0.016***(0.00)	0.014***(0.00)	0.016***(0.00)
ln(dist_Taiwan)	-0.091***(0.01)	-0.074***(0.01)	-0.091***(0.01)	-0.074***(0.01)
ln(dist_Japan)	-0.093***(0.01)	-0.048***(0.01)	-0.085***(0.01)	-0.048***(0.01)
ln(dist_Hong Kong)	-0.076***(0.01)	-0.044***(0.01)	-0.076***(0.01)	-0.043***(0.01)
D ×ln(dist_Quingdao)	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)
D ×ln(dist_Shanghai)	0.003 (0.00)	0.010***(0.00)	0.003 (0.00)	0.010***(0.00)
D ×ln(dist_Shenzen)	-0.006 (0.00)	-0.008***(0.00)	-0.006 (0.00)	-0.008***(0.00)
D ×ln(dist_Ningbo)	-0.006***(0.00)	-0.009***(0.00)	-0.007***(0.00)	-0.010***(0.00)
D ×ln(dist_Guangzhou)	0.014***(0.00)	0.015***(0.00)	0.013***(0.00)	0.015***(0.00)
D ×ln(dist_coastline)	-0.001 (0.00)	-0.001 (0.00)	-0.002 (0.00)	-0.001 (0.00)
capital/labor	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)
birth date	0.001***(0.00)	-0.000***(0.00)	0.001***(0.00)	-0.000***(0.00)
GE_z	0.356***(0.06)	0.234***(0.04)	0.552***(0.07)	0.227***(0.04)
Private		0.015***(0.00)		0.014***(0.00)
Foreign		0.303***(0.00)		0.303***(0.00)
size group 2			0.014***(0.00)	0.016***(0.00)
size group 3			0.028***(0.00)	0.024***(0.00)
size group 4			0.038***(0.00)	0.027***(0.00)
size group 5			0.045***(0.00)	0.021***(0.00)
constant	0.490***(0.16)	2.094***(0.13)	0.436***(0.16)	2.142***(0.13)
GE_z				
ln(dist)	-0.006 (0.01)	-0.008 (0.01)	-0.008 (0.01)	-0.008 (0.01)
GE_p	0.309***(0.03)	0.243***(0.03)	0.243***(0.03)	0.243***(0.03)
var(e.GE_z)	0.090***(0.01)	0.041***(0.01)	0.032***(0.01)	0.011 (0.01)
var(e.sales)	1.747***(0.02)	1.798***(0.02)	1.807***(0.02)	1.828***(0.02)
var(e.export/output)	0.094***(0.00)	0.090***(0.00)	0.095***(0.00)	0.091***(0.00)
Number of obs.	1,047,989	1,047,989	1,047,989	1,047,989
AIC	4220383.212	4083310.557	4218044.249	4082743.181
BIC	4220632.322	4083583.392	4218340.808	4083063.465

Standard errors are clustered at zip-code level and reported in parentheses. Coefficients are significant at the 10 percent (* p<0.10), 5 percent (** p<0.05) or 1 percent (*** p<0.01) level. Specification is estimated by generalized simulation equation modeling. Dependent variables are **sales**, export share, **exp/output** and government efficiency at zip-code level, **GE_z**, which is simultaneously a latent variable. GE_p is observed government efficiency at province level. *capital/labor* specifies the capital-labor ratio. *ln(dist)* measures the distance between the zip-code area the firm is located and the provincial capital city. *ln(dist_Quingdao)*, *ln(dist_Shanghai)*, *ln(dist_Shenzen)*, *ln(dist_Ningbo)*, *ln(dist_Guangzhou)* are the distances between firm location and the respective harbor. *ln(dist_coastline)* represents the shortest distance between coastline and firm location. *birth date* identifies the year of establishment. Firms are classified into *size groups* defined by output quintiles. *SOE*, *Private* and *Foreign* specify firm ownership and stand for state owned enterprise (including collective owned enterprises), private owned firms as well as foreign owned firms (including firms of Hong Kong and Taiwan). *SOE* serves as base category. *AIC* specifies Akaike's information criterion, while *BIC* represents Bayesian information criterion.

The estimates of *birth date* are ambiguous: Including ownership dummies in column (2) and (4) yield significantly negative coefficients. The effect in column (1) and (3) is highly significant and positive suggesting that older firms export more on average.[12]

The export share is negatively affected by the distance between location of a firm and important trading partners; Taiwan, Japan and Hong Kong, illustrated by the highly significant and negative coefficients $ln(dist\_Taiwan)$, $ln(dist\_Japan)$, and $ln(dist\_HongKong)$. Similarly, the export share of a firm is also positively affected by higher proximity to the coastline but this effect turns insignificant. The estimation results concerning the distances between firm location and harbors are ambiguous. The distance of a firm to the port of Ningbo and Shenzen is related to a lower share of exports. In contrast, the estimated coefficient of $D \times ln(dist\_Guangzhou)$ is statistically significant and positive, suggesting that a higher distance is associated with a higher export share of a firm. The estimates concerning the distance to the port of Shanghai are ambiguous as well: the effect is insignificant in column (1) and (3), whereas the effect is significantly positive in column (2) and (4). The effect of the distance to the port of Quingdao is insignificant.

Agglomeration and Dispersion Forces

The results in Table 1 suggest that there is no significant linear relationship between zip-code level efficiency of a government and the distance between political hub and firm location. However, the theoretical considerations presented in the introduction suggest a hump shaped relationship between distance and institutional quality. Exporting firms prefer locations close to the political and/or economic hub but rising fixed costs push them out of the city center. Export promoting policies determine the regional government efficiency: A higher density of exporters can be associated with higher government spending for good institutions. Put differently, the efficiency may be higher in areas located slightly off the core of the province but it is decreasing for distances higher than a certain distance-threshold.

[12] According to Reinecke and Schmerer (2017) especially shortly before China's WTO entry in 2001 the number of private and foreign-owned establishments, that are characterized by higher export shares than SOEs, increased a lot. Thereby, private and foreign-owned firms are younger by trend.

Table 2: *GSEM including quadratic distance*

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
sales				
capital/labor	0.006 (0.01)	0.007 (0.01)	0.007 (0.01)	0.007 (0.01)
GE_z	1.000	1.000	1.000	1.000
constant	9.426***(0.10)	9.433***(0.10)	9.433***(0.10)	9.433***(0.10)
export/output				
ln(dist)	-0.099***(0.03)	-0.062***(0.02)	-0.176***(0.05)	-0.061***(0.02)
$ln(dist)^2$	0.015***(0.00)	0.010***(0.00)	0.025***(0.01)	0.010***(0.00)
ln(dist_Taiwan)	-0.090***(0.01)	-0.074***(0.01)	-0.090***(0.01)	-0.074***(0.01)
ln(dist_Japan)	-0.094***(0.01)	-0.048***(0.01)	-0.085***(0.01)	-0.048***(0.01)
ln(dist_Hong Kong)	-0.078***(0.01)	-0.044***(0.01)	-0.078***(0.01)	-0.044***(0.01)
D× ln(dist_Quingdao)	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)
D× ln(dist_Shanghai)	0.003 (0.00)	0.010***(0.00)	0.003 (0.00)	0.010***(0.00)
D× ln(dist_Shenzen)	-0.006 (0.00)	-0.008***(0.00)	-0.007 (0.00)	-0.008***(0.00)
D× ln(dist_Ningbo)	-0.006***(0.00)	-0.009***(0.00)	-0.007***(0.00)	-0.010***(0.00)
D× ln(dist_Guangzhou)	0.013***(0.00)	0.015***(0.00)	0.012***(0.00)	0.015***(0.00)
D× ln(dist_coastline)	-0.001 (0.00)	-0.001 (0.00)	-0.002 (0.00)	-0.001 (0.00)
capital/labor	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)	-0.001 (0.00)
birth date	0.001***(0.00)	-0.000***(0.00)	0.001***(0.00)	-0.000***(0.00)
GE_z	0.374***(0.06)	0.265***(0.05)	0.628***(0.09)	0.258***(0.05)
Private		0.015***(0.00)		0.015***(0.00)
Foreign		0.303***(0.00)		0.303***(0.00)
size group 2			0.014***(0.00)	0.017***(0.00)
size group 3			0.028***(0.00)	0.025***(0.00)
size group 4			0.038***(0.00)	0.027***(0.00)
size group 5			0.045***(0.00)	0.021***(0.00)
constant	0.477***(0.16)	2.091***(0.13)	0.424***(0.16)	2.141***(0.13)
GE_z				
ln(dist)	0.277***(0.06)	0.282***(0.06)	0.282***(0.06)	0.282***(0.06)
$ln(dist)^2$	-0.038***(0.01)	-0.038***(0.01)	-0.038***(0.01)	-0.038***(0.01)
GE_p	0.291***(0.03)	0.215***(0.03)	0.215***(0.03)	0.215***(0.03)
var(e.GE_z)	0.085***(0.01)	0.036***(0.01)	0.028***(0.01)	0.010 (0.01)
var(e.sales)	1.746***(0.02)	1.798***(0.02)	1.806***(0.02)	1.824***(0.02)
var(e.export/output)	0.093***(0.00)	0.089***(0.00)	0.093***(0.00)	0.091***(0.00)
Number of obs.	1,047,981	1,047,981	1,047,981	1,047,981
AIC	4217438.125	4080280.410	4214948.743	4079710.595
BIC	4217710.960	4080576.969	4215269.027	4080054.604

Standard errors are clustered at zip-code level and reported in parentheses. Coefficients are significant at the 10 percent (* p<0.10), 5 percent (** p<0.05) or 1 percent (*** p<0.01) level. Specification is estimated by generalized simulation equation modeling. Dependent variables are **sales**, export share, **exp/output** and government efficiency at zip-code level, **GE_z**, which is simultaneously a latent variable. GE_p is observed government efficiency at province level. *capital/labor* specifies the capital-labor ratio. $ln(dist)$ measures the distance between the zip-code area the firm is located and the provincial capital city. $ln(dist\_Quingdao)$, $ln(dist\_Shanghai)$, $ln(dist\_Shenzen)$, $ln(dist\_Ningbo)$, $ln(dist\_Guangzhou)$ are the distances between firm location and the respective harbor. $ln(dist\_coastline)$ represents the shortest distence between coastline and firm location. *birth date* identifies the year of establishment. Firms are classified into *size groups* defined by output quintiles. *SOE*, *Private* and *Foreign* specify firm ownership and stand for state owned enterprise (including collective owned enterprises), private owned firms as well as foreign owned firms (including firms of Hong Kong and Taiwan). *SOE* serves as base category. *AIC* specifies Akaike's information criterion, while *BIC* represents Bayesian information criterion.

This threshold is estimated by including $ln(dist)^2$ as additional control variable in our benchmark specification.

Sales again depend on government efficiency at the zip-code level and the capital-labor ratio. Similarly, the export regression is specified as the regression displayed in Table 1.

However, the latent dependent variable government efficiency at the zip-code level is a function of the provincial level efficiency, the distance between firm location and provincial capital city as well as the squared capital distance. Standard errors are again clustered at the zip-code level.

The estimated signs of the other control variables are basically the same as discussed for the benchmark regression table.

The latent dependent variable GE_z is positively affected by provincial government efficiency, which is similar to the outcome presented in the benchmark regression in Table 1. The effect of distance to the capital on government efficiency at the zip-code level becomes significant: government efficiency is affected by distance as indicated by the highly significant coefficient of $ln(dist)$. The squared distance $ln(dist)^2$ is significant and negative. These two coefficients taken together suggest a humped shaped relation between government efficiency and distance to the capital: First, government efficiency increases with distance but the effect turns from positive to negative at a value for distance equal to $ln(3.71) \approx 40.88\ km$.

This result supports the hypothesis concerning the trade off between low transportation and high fix cost in the agglomeration center, which leads to an adjustment in the location choice of a firm into the suburban area.

We choose the model with the lowest AIC/BIC statistic as preferred model when calculating the predicted latent variable. The test statistic is based upon the Akaike's and Bayesian information criterion. Table 3 reports several moments of the predicted outcome for GE_z. Moreover, the second row replicates these moments for the standard government efficiency at the province level.

Figure 3 illustrates both measures. The upper panel represents the more aggregated provincial level measure of government efficiency. In contrast, the panel at the bottom

Table 3: *Descriptive statistics for predicted and observed government efficiency*

	Obs.	Mean	Stand. Dev.	Min	Max
GE_z	1,068,695	0.589	0.102	-1.027	1.027
GE_p	1,284,417	0.147	0.237	-0.88	0.77

depicts the novel government efficiency index at the county-level, which is constructed by collapsing the data from the zip-code to the county-level. Darker areas describe a higher value of the government efficiency index.

It appears that regional heterogeneity tends to be important analyzing more dis-aggregated research questions concerning the impact of institutions on divers outcome variables. Within provinces we can observe high differences regarding government efficiency that if neglected may lead to biased empirical results.

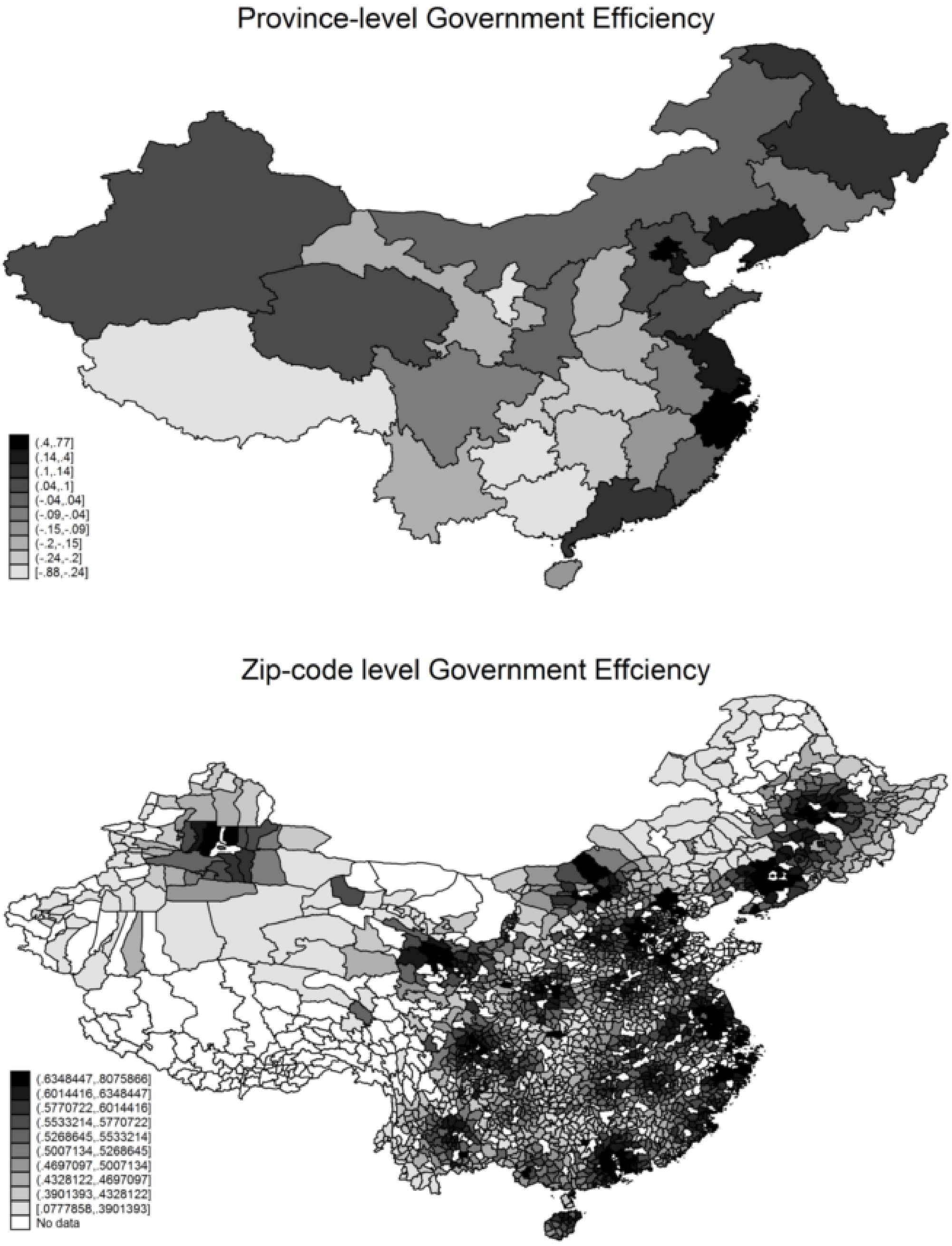
4.1. Application results

We test the performance of our new index using standard firm-level regressions to fit government efficiency to a measure of total factor productivity. All setups include either the provincial government efficiency index or our novel GE_z measure. Results for the TFP regressions are reported in Table 4. Regressions reported in columns (1) to (3) are based upon the government efficiency index at the provincial level. Columns (4) to (6) represent the estimates including the more dis-aggregated measure of government efficiency at the zip-code level. Time fixed effects are included in all specifications. Column (2) and (5) additionally put province fixed effects into the regression. Column (3) and (6) control for firm fixed effects. Standard errors are estimated robust and are clustered at province or zip-code level depending on the employed index.

The signs of the control variables are in line with our expectations: younger firms tend to be characterized by higher TFP. This result is supported by the positive and significant coefficients of the ownership dummies *Private* and *Foreign*. Private- and foreign-owned firms are characterized by higher TFP compared to SOEs.[13] Moreover,

[13]Brandt et al. (2017) estimate the effect of the WTO accession on Chinese firm performance. They show that on average reduction in tariffs increases productivity. Additionally, the authors are able to provide evidence that especially firms entering the market boost productivity because they are more productive

Figure 3: *Comparison of province and zip-code level government efficiency*



TFP seems to be positively affected by firm size: The coefficient of $ln(labor)$ is highly significant and positive. The dummy variable indicating a firm as exporter or non-exporter is significant and positive controlling for firm-level unobserved heterogeneity

on average and these firms benefit from input-tariff reduction. Referring to the data, particularly private and foreign owned firms enter the market during the observed time span 2001-2006.

Table 4: *Application of the new GE-Index - Relation on TFP*

Dependent variable: Total Factor Productivity	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
GE_p	0.060	-0.196	-0.301**			
	(0.19)	(0.17)	(0.14)			
GE_z				0.596***	0.581***	0.981***
				(0.06)	(0.07)	(0.09)
birth date	0.016***	0.016***	-0.000	0.016***	0.015***	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
exporting firm	-0.044	-0.017	0.073*	-0.059***	-0.024***	0.066***
	(0.05)	(0.03)	(0.04)	(0.01)	(0.01)	(0.01)
ln(labor)	0.121***	0.118***	0.080***	0.119***	0.115***	0.078***
	(0.02)	(0.02)	(0.02)	(0.00)	(0.00)	(0.01)
ln(capital)	-0.297***	-0.301***	-0.468***	-0.297***	-0.301***	-0.469***
	(0.02)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
private	0.112**	0.093**	0.013	0.102***	0.086***	0.015*
	(0.04)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
foreign	0.127*	0.122*	0.050***	0.117***	0.122***	0.065***
	(0.07)	(0.06)	(0.02)	(0.01)	(0.01)	(0.02)
constant	-32.019***	-31.369***	1.803*	-31.671***	-31.193***	0.673
	(3.83)	(3.50)	(0.98)	(0.68)	(0.67)	(0.75)
Provincial FE		yes			yes	
Firm FE			yes			yes
Number of obs.	985,212	985,212	985,212	985,212	985,212	985,212
R-sq. within	0.162	0.183	0.778	0.164	0.184	0.778
adj. R-sq.	0.162	0.183	0.660	0.164	0.184	0.660

Standard errors are reported in parentheses. In specification (1) to (3) they are clustered at province level, in (4) to (5) standard errors are clustered at zip-code level. Coefficients are significant at the 10 percent (* p<0.10), 5 percent (** p<0.05) or 1 percent (*** p<0.01) level. The dependent variable is Total Factor Productivity (TFP). GE_p measures of government efficiency at the province level, while GE_Z is our predicted government efficiency index at the zip-code level. *birth date* specifies the firm's year of establishment. $exp/output$ defines the export share on total output. *SOE*, *Private* and *Foreign* identify firm ownership and stand for state owned enterprises (including collectively owned enterprises), private owned firms as well as foreign owned firms (including firms of Hong Kong and Taiwan). *SOE* is serves as base category. *labor* and *capital* are controls for employment and capital stock. Column (2) and (5) report estimations including province fixed effects, while column (3) and (6) represent results with firm fixed effects. All regressions include year dummies.

in column (3) and (6). Consistently, exporting firms are associated with a higher level of total factor productivity. Less intuitive is the effect of the capital stock on TFP. This result may be driven by state owned enterprises. SOEs have better access to the financial market, and hence, have a higher capital stock compared to the more productive private or foreign owned firms. At the same time, they are characterized by lower TFP and export levels than foreign and private firms. Brandt et al. (2017) show

SOEs displaying lower growth rates than private firms despite increased competition after WTO accession of China.

The first row presents the coefficient of provincial government efficiency. Estimations in column (1) and (2) are insignificant. Including firm fixed effects in column (3) leads to a negative coefficient of provincial government efficiency: a higher government efficiency reduces total factor productivity. This result is counter intuitive. The second row presents the highly dis-aggregated zip-code level index of government efficiency. Our novel measure yields much more plausible results indicating that a higher government efficiency spurs total factor productivity. This result is robust and survives putting firm fixed effects. We conclude that the effect of government efficiency is downwards biased when regional heterogeneity is not taken into account.

5. Robustness check - consider area size in prediction of GE_z

The non-monotonic relationship between distance and GE_z may stem from the huge differences in size of the provinces. Provinces at the east coast are small compared to provinces in the hinterland. We test this hypothesis by including the interaction between the size of the province and distance in addition to the direct measure of distance. We expect that the direct effect is positive but the interaction should be negative. A higher distance to the core of the capital is associated with a higher government efficiency but the effect turns negative in bigger provinces. This approach allows identifying the size at which the distance premium turns into a distance discount. Surveillance in smaller provinces should be much easier than monitoring the hinterland in huge provinces. Moreover, the infrastructure is much better in smaller provinces, which reduces the transportation costs channel mentioned before. The *GSEM* results including distance, province size as well as an interaction term between those two determinants are represented in table 5. As in the regressions represented in Table 1 and 2 the three equation are estimated simultaneously.

Table 5: *GSEM considering province size in $ln(km)^2$ - 2*

	(1) b/se	(2) b/se
sales		
capital/labor	0.007 (0.01)	0.007 (0.01)
GE_z	1.000	1.000
constant	7.510***(0.28)	7.522***(0.29)
export/output		
ln(dist)	0.010* (0.01)	0.010* (0.01)
$ln(dist)^2$	0.001 (0.00)	0.001 (0.00)
ln(dist_Taiwan)	-0.075***(0.01)	-0.075***(0.01)
ln(dist_Japan)	-0.053***(0.01)	-0.055***(0.01)
ln(dist_Hong Kong)	-0.041***(0.01)	-0.041***(0.01)
D ×ln(dist_Quingdao)	-0.001 (0.00)	-0.001 (0.00)
D ×ln(dist_Shanghai)	0.007***(0.00)	0.007***(0.00)
(D ×ln(dist_Shenzen)	-0.009***(0.00)	-0.009***(0.00)
(D ×ln(dist_Ningbo)	-0.007***(0.00)	-0.007***(0.00)
D ×ln(dist_Guangzhou)	0.016***(0.00)	0.016***(0.00)
D ×ln(dist_coastline)		-0.001 (0.00)
capital/labor	-0.001 (0.00)	-0.001 (0.00)
birth date	-0.000***(0.00)	-0.000***(0.00)
SOE	-0.305***(0.00)	-0.305***(0.00)
private	-0.290***(0.00)	-0.291***(0.00)
GE_z	0.121***(0.02)	0.119***(0.02)
size group=2		0.017***(0.00)
size group=3		0.025***(0.00)
size group=4		0.027***(0.00)
size group=5		0.022***(0.00)
constant	2.146***(0.14)	2.208***(0.14)
GE_z		
ln(dist)	0.521***(0.07)	0.518***(0.07)
GE_p	0.391***(0.03)	0.392***(0.03)
$ln(prov\_size)$	0.210***(0.02)	0.209***(0.02)
$ln(prov\_size) \times ln(dist)$	-0.046***(0.01)	-0.046***(0.01)
var(e.GE_z)	0.075***(0.02)	0.014 (0.02)
var(e.sales)	1.758***(0.02)	1.820***(0.02)
var(e.export/ouput)	0.091***(0.00)	0.091***(0.00)
Number of obs.	1,047,991	1,047,991
AIC	4079827.056	4079247.629
BIC	4080123.615	4079603.501

Standard errors are clustered at zip-code level and reported in parentheses. Coefficients are significant at the 10 percent (* p<0.10), 5 percent (** p<0.05) or 1 percent (*** p<0.01) level. Specification is estimated by generalized simulation equation modeling. Dependent variables are **sales**, export share, **exp/output** and government efficiency at zip-code level, **GE_z**, which is simultaneously a latent variable. GE_p is observed government efficiency at province level. *capital/labor* specifies the capital-labor ratio. $ln(dist)$ measures the distance between the zip-code area the firm is located and the provincial capital city. $ln(prov\_size)$ describes province size in logarithmized square kilometers. $ln(dist\_Quingdao)$, $ln(dist\_Shanghai)$, $ln(dist\_Shenzen)$, $ln(dist\_Ningbo)$, $ln(dist\_Guangzhou)$ are the distances between firm location and the respective harbor. $ln(dist\_coastline)$ represents the shortest distance between coastline and firm location. *birth date* identifies the year of establishment. Firms are classified into *size groups* defined by output quintiles. *SOE*, *Private* and *Foreign* specify firm ownership and stand for state owned enterprise (including collective owned enterprises), private owned firms as well as foreign owned firms (including firms of Hong Kong and Taiwan). Foreign owned firms serve as base category. *AIC* specifies Akaike's information criterion, while *BIC* represents Bayesian information criterion.

The first equation considers the relation between sales, the capital-labor ratio as well as government efficiency at the zip-code level where the coefficients are set to one. The second equation depicts the relation between a firm's export share and government efficiency at the zip-code level as well as diverse firm-level and geographic control variables. The last equation captures the link between the latent variable government efficiency at the zip-code level distance, province size and the interaction between those two variables. Standard errors are clustered at zip-code level. The first column presents results excluding size classes, while the specification in column 2 includes these control variables.

The signs of the coefficients of equation (1) and (2), with sales and export share as dependent variable, respectively, remain unchanged compared to the estimation results shown in Table 1 and 2.

The capital-labor ratio is estimated to has neither a significant effect on sales nor on a firm's export share. Furthermore, firms associated with higher sales tend to be characterized by a higher export share indicated by the highly significant coefficients of size group 2 to 5. Also ownership is predicted to have a significant effect on the export behavior of a firm: SOEs as well as private owned firms are associated with lower export shares than foreign owned firms well established in global supply chains.

As in the regressions before, government efficiency at the zip-code level, GE_z, is positively affected by provincial government efficiency, GE_p. Similarly, distance to the capital city, $ln(dist)$, and province size, $ln(prov\_size)$, are positively related to zip-code level efficiency. However, these positive effects are mitigated or even reversed by the negative interaction effect between size and distance, $ln(prov\_size) \times ln(dist)$. We are interested in the marginal effect of distance. Calculating the marginal effect of distance under consideration of the interaction term, $\frac{\delta GE_z}{\delta ln(dist)} = 0.518 - 0.046 \times ln(prov\_size)$, the direct positive effect of distance turns into negative if the province size exceeds approximately 77,720 km^2. This means the effect of distance is negative for all provinces larger in size than Ningxia, which applies for 26 out of 30 provinces.[14] Hence, the

[14] A List of provinces by size can be found in the Appendix. The horizontal line visualizes the threshold when the positive effect of distance becomes negative.

hypothesis that smaller provinces can be governed more efficiently than larger provinces due to information issues tend to be confirmed. According to the Akaike's and Bayesian information criterion the model in column 2 is preferred over the model in column 1. Therefore, the latent variable government efficiency at the zip-code level, GE_z, is predicted based on the second specification.

To test the robustness of the zip-code level government efficiency index predicted on the basis of the results shown in Table 5, the index again finds application in the panel data regression fitting total factor productivity.

Table 6: *Application of the new GE_z including area size in prediction - Relation on TFP*

Dependent Variable: TFP						
	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
GE_p	0.061	-0.196	-0.300**			
	(0.19)	(0.17)	(0.14)			
GE_{z2}				0.659***	1.423***	0.691***
				(0.06)	(0.08)	(0.07)
Provincial FE		yes			yes	
Firm FE			yes			yes
Number of obs.	985,287	985,287	985,287	985,287	985,287	985,287
R-sq. within	0.162	0.183	0.778	0.164	0.188	0.778
adj. R-sq.	0.162	0.183	0.660	0.164	0.188	0.660

Standard errors are reported in parentheses. Coefficients for birth date, export dummy, size, capital, ownership dummies and a constant are estimated but not reported. In specification (1) to (3) they are clustered at province level, in (4) to (5) standard errors are clustered at zip-code level. Coefficients are significant at the 10 percent (* p<0.10), 5 percent (** p<0.05) or 1 percent (*** p<0.01) level. The dependent variable is total factor productivity (TFP). GE_p measures of government efficiency at the province level, while GE_Z is our predicted government efficiency index at the zip-code level. *birth date* specifies the firm's year of establishment. The dummy variable $exporting\ firm$ indicates whether a firm is an exporter ($exporting\ firm = 1$), or not ($exporting\ firm = 0$). *SOE*, *Private* and *Foreign* identify firm ownership and stand for state owned enterprises (including collectively owned enterprises), private owned firms as well as foreign owned firms (including firms of Hong Kong and Taiwan). *SOE* is serves as base category. $ln(labor)$ and $ln(capital)$ are logarithmized controls for employment and capital stock. Column (2) and (5) report estimations including province fixed effects, while column (3) and (6) represent results with firm fixed effects. All regressions include year dummies.

Table 6, column (1) to (3) show the results including provincial-level efficiency, whereas column (4) to (6) represent the results applying the newly and highly disaggregated efficiency index at the zip-code level. The setup is identical to the application discussed for the benchmark results above. The results confirm the outcomes discussed before. Our novel index yields more plausible results in line with the common perception and in line with the findings documented in other studies. Government efficiency spurs total factor productivity.

6. Conclusion

In this paper we investigated in how far measures of institutional quality at the national level are appropriated measures to analyze the relation between firm-level outcomes and institutional quality. Our analysis builds on the hypothesis that the efficiency of institutions decreases with increasing distance to the regional capital city. We argue that the proximity of a firm location to the regional capital city is an important factor determining local government efficiency. Political decisions are made in capital cities and political measures were implemented in core areas first. Moreover, we argue that less productive firms are more concentrated in remote areas governed by local officials who are also more difficult to control. This endogenous sorting of firms and local officials according to individual performance may yield systematic but unobserved differences in government efficiency across and within provinces. Based on this considerations, we expected areas close to political hubs being characterized by a higher level of government efficiency. To identify a highly dis-aggregated measure of local government efficiency, we additionally adopt the well analyzed link between high institutional quality and export performance of firms: legal security, provided by strong institutions is one important determinant of a firm's export behavior, by reducing transaction cost. Adapted from both these arguments and well established results from academic literature, we argue that the eased efficiency in institutions is associated with a negative effect on a firm's export share. Put differently, the closer a firm is located to the regional capital city the higher is the efficiency of the government and the higher the expected export share of the observed firm. This information expected in the export-government efficiency relation as well as the assumption of a negative relation between capital distance and government efficiency is used to predict an efficiency measure at the zip-code level applying a structural equation modeling approach.

Based on Chinese firm-level data and the *Chinese provincial government efficiency index*, that we use to approximate institutional quality, the latent variable zip-code level government efficiency is predicted by estimating three equations simultaneously. Thereby, unobserved government efficiency at the zip-code level is a function of provincial government efficiency and the distance between firm location and provincial capital city. According to our estimates, the relation between local government efficiency and distance is not linear, but quadratic. This result is explained by agglomeration and dispersion forces: A firm is faced a trade-off between low transportation but high housing cost in the agglomeration center. High fixed cost push firms out of the agglomeration center into suburban areas. In these areas firms benefit from lower housing prices and reasonable transportation cost. This phenomenon gives an incentive to local policy makers to invest in the effectiveness of institutions to attract further firms. Consistently, the relation between government efficiency and distance from capital city to firm location is hump-shaped. Based on our predictions, we are able to show that government efficiency varies at a high level within provinces.

To show validity of the newly predicted index, it finds application in a panel data regression analysis. The analysis shows: First, government efficiency positively affects firms' total factor productivity at a high significance level. Second, the effect of government efficiency on the mentioned firm-level outcomes is highly underestimated neglecting regional heterogeneity concerning institutional quality.

To sum up: we provide a new method to predict a highly dis-aggregated measure of institutional quality using a structural equation modeling approach. Additionally, we give empirical evidence of the necessity to consider regional heterogeneity concerning institutional quality and show that neglecting this heterogeneity leads to highly biased estimation results analyzing the effect of institutional quality on diverse firm-level outcomes.

References

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, 129:1085–1140.

Anderson, J. E. and Marcouiller, D. (2002). Insecurity and the pattern of trade: An empirical investigation. *The Review of Economics and Statistics*, 84(2):342–352.

Brandt, L., Rawski, T., and Sutton, J. (2008). Industrial development ind china. In Brandt, L. and Rawski, T. G., editors, *China's Great Economic Transformationi*, pages 569–632. New York: Cambridge University Press.

Brandt, L., van Biesebroeck, J., Wang, L., and Zhang, Y. (2017). Wto accession and performance of chinese manufacturing firms. *American Economic Review*, 107:2784–2820.

Brandt, L., van Biesebroeck, J., and Zhang, Y. (2012). Creative accounting or creative destruction? firm-level productivity growth in chinese manufacturing. *Journal of Development Economics*, 97:339–351.

Brandt, L., van Biesebroeck, J., and Zhang, Y. (2014). Challenges of working with chinese nbs firm-level data. *China Economic Review*, 30:339–352.

De Groot, H. L. F., Linders, G.-J., Rietvield, P., and Sumbramanian, U. (2004). The institutional determinants of bilateral trade patterns. *KYKLOS*, 57(1):103–124.

Gani, A. (2017). Contract enforcement and trade. *Journal of Industry, Competition and Trade*, pages 1–13.

Gu, C., Shen, J., Wong, K.-y., and Zhen, F. (2001). Regional polarization under the socialist-market system since 1978: a case study of guangdong province in south china. *American Economic Review*, 98(5):1978–97.

Henkel, M., Seidel, T., and Suedekum, J. (2018). Fiscal transfers in the spatial economy. CESifo Working Papers 7012, SESifo.

Huang, Z., Li, L., Ma, G., and Xu, L. C. (2017). Hayek, local information and commanding heights: Decentralizing state-owned enterprises in china. *American Economic Review*, 107(8):2455–2478.

Jefferson, G. H., Rawski, T. G., and Zhang, Y. (2008). Productivity growth and convergence across china's industrial economy. *Journal of Chinese Economic and Business Studies*, 6(2):121–140.

Krugman, P. (1991). Cities in space: Three simple mooels. NBER Working paper Series 3607, National Bureau of Economic Research.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Ottaviano, G., Tabuchi, T., and Jaques-François, T. (2002). Agglomeration and trade revisited. *International Economic Review*, 43(2):409–435.

Ranjan, P. and Young Lee, J. (2007). Contract enforcement and international trade. *ECONOMICS & POLITICS*, 19(2):191–218.

Reinecke, A. and Schmerer, H.-J. (2017). Government efficiency and exports in china. *Journal of Chinese Economic and Business Studies*, 15(3):249–268.

Tang, R., Tang, T., and Lee, Z. (2014). The efficiency of provincial governments in china from 2001 to 2010: measurement and analysis. *Journal of Public Affairs*, 14(2):142–153.

Wei, Y. D. and Kim, S. (2002). Widening inter-county inequality in jiangsu province, china, 1950—95. *Journal of Development Studies*, 38:142–164.

Xinyue, Y. and Wei, Y. D. (2005). Geospatial analysis of regional development in china: The case of zhejiang province and the wenzhou model. *Eurasian Geography and Economics*, 46:895–921.

7. Appendix

7.1. Chinese Provincial Government Efficiency Index composition

Indicator	Sub-indicator	Index
Government public service	Service of science, education, culture and health	Expenditure on science per capita (new product trial fees, testing fees, and subsidies for major research projects, in yuan), qualified product rate (in %), patent application granted (invention patents, utility model patents, and design patents (per 100 000 people)), turnover per capita in technology market (in yuan), primary school students teachers ratio (reverse index), secondary school students teachers ratio (reverse index), functional illiteracy rate (in %, reverse index), state education budget (in % of GDP), literary publishing units (per 100 000 people), health beds (100 000 people), health personnel (100 000 people)
	Public security service	accident rate (traffic accidents, fires, and environmental pollution (one per 100 000 people, reverse index), loss of three accidents per capita (in yuan, reverse index), legislation (new legislation, amendments including laws, bills, and regulations), accepted case in court of first instance, concluded case in court of first instance, arrested suspect by the prosecutor's office, criminal case cracked or registered by the Public Security Bureau, criminal case rate (per thousand people, reverse index)
	Meterological service	Site for agricultural meteorology service (per 100 000 people), seismic monitoring stations (per 100 000 people)
	Social security Service	Employment agencies (per 100 000 people), urban community service facilities (per 100 000 people), rural social security network (per 100 000 people)
Government public infrastructure	Basic social infrastructure	basic infrastructure and renovation investment within state budget (RMB 100m), ratio of local projects and central government projects in basic infrastructure and renovation investment, basic infrastructure and renovation project completed rate (in %), industrial 'three wastes' treatment efficiency (waster, residue, and gas waste, in %), reservoir capacity per 10 000 people (100 million cubic meters per 10 000 people), ratio of nature reserve area (in %)
	Basic urban	city gas penetration rate (in %), urban public transport vehicles per 10 000 people, urban road area per capita (square meters), urban public green area per capita (square meters), urban public toilets per 10 000 people
Size of Government		proportion of administrative staff in the total population (per 10 000 people, reverse index), proportion of administrative employment in total employment (in %, reverse index), ratio of government consumption and final consumption (in %, reverse index), ratio of government consumption and GDP (in %, reverse index), ratio of income from confiscation and administration on total revenue (in %, reverse index)
Residents' economic welfare		per capita net income of rural households (in yuan), per capita urban disposable income of households (in yuan), Engel coefficient of rural residents (in %, reverse index), Engel coefficient of urban residents (in %, reverse index), consumer price index (previous year = 100, reverse index), per capita GDP (in yuan), ratio of subsidy expenditure and financial expenditure (in %)

Source: Tang, R., Tang, T. and Lee, Z. (2014): The efficiency of procincial governments in China from 2001 to 2010: measurement and analysis, p. 147/148

7.2. Summary statistics

Table 7: *Descriptive statistics - firm-level variables*

	Obs.	Mean	Stand. Dev.	Min	Max
$export/output$	1,689,972	.169	.343	0	26.0212
GE_p	1,284,366	.147	.237	-.88	.77
$sales$	1,689,972	9.819	1.418	0	18.878
age	1,723,523	10.851	12.270	0	155
$ln(labor)$	1,696,413	4.753	1.184	0	12.053
$ln(capital)$	1,700,289	3.846	1.671	-5.478	13.789
TFP	1,577,162	-.399	1.357	-15.907	9.522
SOE	1,723,523	.463	.497	0	1
$Private$	1,723,523	.331	.471	0	1
$Foreign$	1,723,523	.205	.404	0	1
N	1,723,523				

Table 8: *Descriptive statistics - distances*

	Obs.	Mean	Stand. Dev.	Min	Max
$dist$	20,447	174.358	132.216	0	1435.887
$dist\_Taiwan$	20,447	1322.120	625.930	179.505	4614.874
$dist\_Japan$	20,447	2437.356	689.148	1122.716	5597.583
$dist\_HongKong$	20,447	1295.922	736.330	13.517	4153.5
$dist\_Quingdao$	20,447	1026.206	574.997	74.728	3954.318
$dist\_Shanghai$	20,447	1034.217	596.353	.182	4272.414
$dist\_Shenzen$	20,447	1288.167	737.335	1.897	4131.983
$dist\_Ningbo$	20,447	1087.015	598.621	2.386	4366.164
$dist\_Guangzhou$	20,447	1234.62	730.988	.739	4037.704
$dist\_coastline$	20,447	439.699	408.937	.018	2359.78
N	20,447				

7.3. List of Chinese provinces by size

Table 9: *Chinese provinces by size*

Province	size in km^2
Shanghai	6,340
Tianjin	11,000
Beijing	16,807
Hainan	33,920
Ningxia	66,000
Chongqing	82,403
Zhejiang	101,800
Jiangsu	102,600
Fujian	121,400
Anhui	139,400
Liaoning	145,900
Shandong	156,700
Shanxi	156,800
Jiangxi	166,900
Henan	167,000
Guizhou	176,100
Guangdong	179,800
Jilin	187,400
Hubei	187,500
Hebei	187,700
Shaanxi	205,800
Hunan	210,800
Guangxi	236,700
Yunnan	394,000
Gansu	454,000
Heilongjiang	460,000
Sichuan	485,000
Qinghai	721,000
Inner Mongolia	1,183,000
Tibet	1,228,400
Xinjiang	1,660,000