Tjalling C. Koopmans Research Institute

Discussion Paper Series nr: 05-30

Trade, Location, and Wages in the United States

Thijs Knaap

Universiteit Utrecht

Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Vredenburg 138
3511 BG Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics. It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to M.Damhuis@econ.uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Thijs Knaap
Utrecht University
Utrecht School of Economics
Vredenburg 138
3511 BG Utrecht
The Netherlands.
E-mail: t.knaap@econ.uu.nl

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 05-30

Trade, location, and wages in the United States

Thijs Knaap

Utrecht School of Economics
Utrecht University

July 2005

Abstract
This paper estimates a spatial wage structure for the United States. I employ the market-access and supplier-access method of Redding and Venables (2004), where access is determined using interstate trade data. Economic geography models predict that state-level wages are correlated to this measure, owing to higher levels of demand and better availability of intermediate goods in easily accessible regions. After correcting for omitted-variable bias with exogenous 'first nature' regressors and using the appropriate instruments, I find that the explanatory power of access-variables is weak in this dataset.

Keywords: Spatial wage structure, United States, Economic Geography

JEL classification: R30, J30, F12

Acknowledgements
This paper was partly written while the author was at Groningen University. I thank Steven Brakman, Harry Garretsen, Vernon Henderson, Catrinus Jepma, Joachim Möller, Peter Neary, and two anonymous referees for their constructive comments on earlier drafts. Any mistakes are the author's sole responsibility.

1 Introduction

Models of economic geography (Fujita et al., 1999; Brakman et al., 2001) attempt to explain the spatial concentration of economic activities. The explanation involves increasing returns to scale at the firm- and industrial level and costs of transport that impede the flow of trade. Depending on the parameters, concentration of activity may arise as a result of pecuniary externalities.

But while such concentrations of economic activity surely exist in practice, as many real-world examples will testify, it may be difficult to determine if the purported mechanisms of economic geography theory have anything to do with them. To show that this is the case, you would need to distinguish this cause for agglomeration from other possible explanations, such as natural geographical circumstances. Ellison and Glaeser (1999) claim that natural advantages, such as the presence of a natural harbor or a particular climate, can be used to explain "at least half of observed geographic concentration" (p. 316). Setting these causes, which are often called 'first nature,' apart from economic incentives to agglomerate (or, 'second nature') is a methodological challenge.

In this paper, I use data on shipments between American states (the 1997 Commodity Flow Survey from the Bureau of Transportation Statistics) to estimate the effect that a state's accessibility has on its wage level. Models of economic geography predict that easily accessible states should have higher prices of local factors, including local wages. The estimation is done using the method of Redding and Venables (2004), who conduct a similar estimation on worldwide country data. Testing the model on data from the United States has the advantage that firms in the continental U.S. are more likely to choose an optimal location, based on the considerations that play a role in economic geography models. With institutions and available factors similar across the U.S., the spatial organization of production is bound to be more amenable to economic analysis than that across the world.

The paper is organized as follows: Section 2 below discusses related literature and presents the main equations, which follow from the model in Venables (1996). That model uses vertical relations between firms as the agglomerating force and explains biregional flows of trade by relating them to the number of firms and the number of consumers in each of the trading regions. Both the number of firms and the number of consumers have a positive effect on trade. Because the model also involves transport costs that increase with distance, it predicts large trade flows between states that are big and close. Taken together, this specification is reminiscent of the gravity trade model, which says that trade is proportional to the size of the trading economies, and is inversely proportional to their distance.

Estimation is done in Section 3, largely using the methods of Redding and Venables (2004). I find that the correlation between market access and wages is strong, as the model would predict. However, when the (econometrically suspect) effect of the own state on its wages is taken out, a severely weakened relationship remains. Especially large states like California have wage levels that cannot be explained by economic activity in neighboring states. Finally, Section 4 concludes.

2 Related literature

Regional variation in wages, even within the same country, can be substantial. The bureau of labor statistics of the United States reports that the average annual pay in Connecticut is about 80% higher than that in South Dakota.[1] There can be many possible causes for these differences between regional wages. Both the presence of complementary factors such as physical capital and natural resources and the quality of the labor force in terms of human capital can be expected to play a role in the determination of the local wage. Recently however, a number of studies have linked spatial variation in wages to another factor entirely: the degree to which large markets are easily accessible from a region.

There are good economic reasons why easy access to major markets should be reflected in high factor prices (Head and Mayer, 2004). Accessibility, which is usually measured by a form of market potential (Harris, 1954), should affect local factor prices in two ways: firstly, firms in easily accessible regions have low costs of transport to their customers. Secondly, when intermediate products must first be shipped from major markets, marginal costs should also be lower in easily accessible regions. These two effects cause regions with high market potential to have a relatively good export position and low costs of intermediates, allowing a generous compensation of local factors. Reversely, immobile factors in isolated regions are squeezed from two sides: demand is low while complementary intermediate factors are relatively expensive.

To quantify the effect of accessibility on local factor prices, I use the formal model of economic geography by Venables (1996). This model describes a manufacturing sector in which firms produce unique products for both intermediate and final consumption. This leads to a quantitative expression for regional wages, through the use of a number of properties of monopolistically competitive markets. Prices in these models are equal to marginal costs with two markups: the standard Dixit and Stiglitz (1977) profit-maximizing markup and an iceberg-type markup that depends positively on distance to the customer. Marginal costs themselves are a weighed geometric average of the costs of the two factors, intermediate products and labor.

As the model's main results are well known, I will abstain from discussing their derivation, referring the interested reader to the original article instead. Per-firm demand from each region has the usual CES-form with a price-elasticity of $-\sigma$, the elasticity of substitution. For instance, in region r the demand for a product

[1] Figures for 1997, see appendix A for details.

from region s is[2]

$$D_{rs} = E_r p_s^{-\sigma} T_{sr}^{1-\sigma} G_r^{\sigma-1}. \tag{1}$$

Here, E_r are expenditures on manufactured goods for region r, G_r is that region's price index and T_{sr} is the transport markup between regions s and r. If there are n_s firms in region s, each producing a unique differentiated product with the same price, the total demand in region r for products from region s is n_s times the expression in (1). The value of this stream of goods is

$$X_{rs} = n_s p_s^{1-\sigma} T_{sr}^{1-\sigma} E_r G_r^{\sigma-1}, \tag{2}$$

which is the flow of trade from s to r.

The number of firms in region s is determined by the zero-profit condition, which fixes per-firm production at some level $\bar{Y}$. Production must equal demand, so that in each region s,

$$\begin{aligned}\sum_{r=1}^{N} E_r G_r^{\sigma-1} T_{sr}^{1-\sigma} &= \bar{Y} p_s^{\sigma} \\ &= \bar{Y} \left[\frac{\sigma}{\sigma-1} w_s^{\alpha} G_s^{1-\alpha} \right]^{\sigma} \end{aligned} \tag{3}$$

where N is the total number of regions, using (1) above. On the right-hand side, I use the fact that the f.o.b. price is a markup over an average of regional wage w_s and the price index (of intermediate goods) G_s. Formula (3) shows that there exists a relationship between a region's accessibility and its wage level. For small regions which are far away from the large markets, the left-hand side of this equation will take on relatively small values, due to high transport markups. On the same grounds the local price index G_s, which depends negatively on costs of transport, will be higher than average. Both these effects will lower local wages w_s, compared to regions close to the industrial core.

Other parts of the economy are left implicit in this model. Real wages are allowed to differ across regions, indicating a low propensity to move in response to wage differences. While the manufacturing sector presumably does not comprise the entire economy, it is assumed that other sectors do not interact with it other than competing on the same labor. Possibly, these other sectors produce untraded 'local products' with labor and a fixed factor, thus setting local wages.

The empirical literature has tried to measure regional accessibility in order to study the link between accessibility and local wages. In addition to naive market

[2] Throughout, I use the index s for the *sending* region and r for the *receiving* region. The different coefficients of price and transport markup are the result of the iceberg assumption that transport costs are incurred in the product itself: for an amount 1 to arrive at r, T_{sr} must actually be produced, accounting for the goods that 'melt' in transit. However, iceberg transport costs are a convenient fiction and these extra goods are not observed in the data; it is therefore defendable to leave the extra T_{sr} out. For consistency with other work, I maintain it here.

potential functions in the spirit of Harris (1954), two main approaches exist: the first builds on Helpman's (1998) economic geography model to augment the simple market potential function with data on stocks of nontradables and wages (Hanson, 2004; Mion, 2004). In this approach, intermediate products do not play a role. The second, used in this paper, estimates accessibility in two steps: first the flow of trade is fitted to (2), after which the results are manipulated for use in (3). This approach is due to Redding and Venables (2004). Estimating the gravity relation (2) involves issues regarding the proper specification of trading costs T_{sr}.

Using the former approach, Hanson (2004) looks at the relationship between market potential and factor prices in US counties. Estimation is done using first differences of the data, to account for (unchanging) external qualities of the land, and finds significant coefficients of the expected sign. The current work differs from Hanson's analysis in that it takes the second approach, using trade data to proxy for accessibility, and takes account of both final and intermediate demand. As trade data is only available on the state level, the spatial units are much coarser than those in Hanson's work.

Brakman et al. (2001) criticize Hanson for failing to recognize the importance of international trade in his estimations. They offer an alternative study into the spatial wage structure in East and West Germany. Their initial estimates are similar to Hanson's (2004), and do not change very much when trade with the rest of Europe is factored into the model. The German case is also studied by Roos (2001), who finds that a naive market potential function is a better indicator for regional wages than the function that follows from formal economic geography models. More recently, other spatial wage structures have been estimated for the European Union (Combes and Overman, 2004, present an overview). Niebuhr (2004) repeats Hanson's (2004) analysis for 158 European regions. She finds that there is some evidence for (backward) linkages, which operate over rather large distances. Mion (2004) applies the Helpman (1998) model to Italian provinces, introducing a proxy for local wages to correct for labor market rigidities. His estimation is based on panel data and instrumental variables to take account of spatial and temporal correlation. He finds that accessibility has an impact on local wages, and concludes that linkages influence the concentration of activity. Fingleton (2005) tests the influence of accessibility against the alternative force of (non-market) externalities on European wages. He finds that the latter dominates as an explanation for wage variation.

Using the second approach, Redding and Venables (2004) estimate a spatial wage structure in two steps, computing accessibility from the observed pattern of trade. Their data concerns 101 countries worldwide, and includes multilateral trade flows and shipping distances, a border dummy and local wages (approximated by GDP per capita). They find that accessibility and location are important determinants of the local wage level. The current work uses their methodology on a dataset of American states.

In order to estimate the gravity[3] relation (2), an operational form of the trade costs term T_{sr} is needed. It is an empirical regularity that transport technology exhibits increased efficiency at larger distances (Overman et al., 2004). And while distance gradually takes its toll on trade, crossing a (state) border seems to have a large negative effect (McCallum, 1995; Wolf, 2000). The size of the border effect may be due to mismeasurement of within-region distance, however (Head and Mayer, 2002).

I employ the usual iceberg specification, taking into account the effects of adjacency and borders with dummy variables, and allowing for increasing efficiency at longer distances. The amount that needs to be shipped to get one unit of the product to arrive from location s in location r, T_{sr}, corresponds to the distance travelled as

$$T_{sr} = (\xi d_{sr})^{\tau_1} \exp\left(\tau_2 \cdot \mathrm{bord}_{rs} + \tau_3 \cdot \mathrm{own}_{rs}\right) \tag{4}$$

where ξ is positive and τ_1 lies between 0 and 1; bord_{rs} is the border dummy and own_{rs} indicates trade within a state.[4]

In the remainder of this paper, I will apply the methodology of Redding and Venables (2004) to data on trade and wages of American states. This will allow an assessment of the role that accessibility plays for local wages in the US.

3 Estimation

The data used in this estimation concern the 50 American states; they are described in detail in appendix A. I make use of the Commodity Flow Survey 1997, a dataset compiled by the US Bureau of Transportation Statistics. It contains estimates of inter- and intrastate trade based on a survey among firms that ship traded goods within the US, which asks for origin and destination addresses, value and weight. The Commodity Flow Survey is a relatively underexploited dataset. The 1993 version of the survey was used by Wolf (2000) to estimate gravity equations and examine the home market effect in US states. The 2002 version of the survey has been conducted, but the results were not yet available at the time of writing.

Other data concerns wage levels in each state and the distances between different states. The latter are computed as the length of a straight line between the state centers, which are found by weighing county employment. Intra-state distance is derived from the state's area. Wage levels are taken from the average annual pay statistics from the BEA, see appendix A.

[3]In a gravity trade equation, trade increases with economic size (n_s, E_r) and decreases with distance (T_{sr}); the price index G_r indicates region r's remoteness (Deardorff, 1995; Wolf, 2000).

[4]Both τ_2 and τ_3 are expected to be negative, *i.e.* both dummies indicate a trade-enhancing phenomenon. This could conceivably push T_{sr} under unity for some within-state trade flows. However, in light of the parameter estimates of paragraph 3, this seems unlikely to happen in practice.

3.1 First stage estimation: gravity

For the estimation of a gravity trade equation, rewrite the trade relationship in (2) as

$$X_{rs} = \phi_s T_{sr}^{1-\sigma} \psi_r. \tag{5}$$

Redding and Venables (2004) call ϕ_s region s's *supply capacity* and ψ_r region r's *market capacity*. Each of these two terms contains information on a region's trade characteristics that is the same towards all its trading partners. Market capacity $\psi_r = E_r G_r^{\sigma-1}$ indicates the quantity of imports absorbed by region r. It increases when the region spends more on imported goods, or when it is (on average) far away from its trading partners.[5] Supply capacity $\phi_s = n_s p_s^{1-\sigma}$ varies with the number of firms in region s, and hence with its total production of tradeables.

I estimate (5) with the fixed effects panel data method, using (4) to rewrite it as

$$\log(X_{rs}) = \delta_0 + \phi' \iota_s + \delta_1 \log(\mathrm{dist}_{rs}) + \delta_2 \mathrm{bord}_{rs} + \delta_3 \iota_s' \iota_r + \psi' \iota_r + u_{rs}. \tag{6}$$

The $N \times 1$ vector ι_i is filled with zeros, except at the ith position, where it is one. Thus, the $N \times 1$ vectors ϕ and ψ contain the supply and market capacities of all regions.

The dependency between distance and trade is captured by the δ-parameters.[6] The first, δ_0, is a scaling factor. Distance (in miles) has a coefficient of δ_1, which is expected to be negative. The influence of the spatial characteristics of the two regions is further captured by two dummy-variables: bord_{rs} is one if the two regions r and s share a border[7] and the product $\iota_s' \iota_r$ is one only if the sending and receiving state are the same.

Using the data about the distances between US states, the size of their bilateral trade and the existence of borders, I estimate the parameters in relation (6). For comparison, the outcomes of similar estimations in Redding and Venables (2004), who use 1994 bilateral trade flows between 101 countries, are reported. The distance between two countries is that between the capital cities. Trade within a region is not taken into account in their estimations, so the regressor $\iota_s' \iota_r$ is left out. In the present dataset, data on trade within a state is available; I estimate both with and without it.

The results of the estimation are in table 1. I report the outcome of similar regressions on world data from Redding and Venables (2004) in the same table for easy comparison. the table contains the values of $\hat{\delta}_1$, $\hat{\delta}_2$ and $\hat{\delta}_3$, leaving the (large) vectors $\hat{\phi}$ and $\hat{\psi}$ out. These coefficients will be used later on, however.

In the first estimation (in the first two columns), the full sample is used. This includes pairs of regions for which no trade is recorded. For both datasets, this

[5] Regions with a small home market that are far away from trading partners will have a high value for G; this means that they will be less daunted by high import prices, since *all* of their import comes from far away.

[6] This follows from the form of T_{sr} that was assumed in formula 4. For instance, the estimate for δ_1 is $(1-\sigma)\tau_1$ in that expression. The scaling parameter ξ vanishes in δ_0, and it is not pursued in

means that the actual trade between the two regions is probably very small. I substitute a zero for (the logarithm of) these unmeasured flows. Distance has the expected negative sign, the border-dummy has a positive coefficient. Both are highly significant. The coefficient of the variable $\iota_s'\iota_r$, called own_{rs} in the table, is also positive and significant.[8]

In the second estimation, pairs of regions between which no trade is recorded are taken out of the sample. This leads to smaller, but more significant coefficient estimates. For the World dataset, the R^2 does not increase; leaving out the zeros does not improve the performance of the model. The R^2 does increase, markedly, for the US dataset. This is caused by the fact that many unobserved pairs involve either Hawaii or Alaska, two states which turn out to be outliers in this dataset.

In the third estimation, pairs with unobserved trade are reintroduced and treated as left-censored observations. The model parameters are estimated using the Tobit method. This increases the coefficient on distance and decreases the border dummy. Standard errors are slightly worse, though.

The final two columns pertain only to the US dataset. The fourth estimation uses only contiguous states, eliminating Alaska and Hawaii from the sample. These two states suffer from many missing observations, whereas those that are available act as outliers. The District of Columbia is also struck from the sample, as the model also performs relatively badly for this region. This is probably due to its small size and atypical sectoral makeup. The fifth estimation eliminates the remaining 49 observations of in-state trade data. This hardly affects any parameters, showing that the use of an in-state dummy adequately captures the special nature of trade within the same state.

Comparing the parameter estimates for trade within the United States to those for world trade, at first glance the results are rather similar. All corresponding parameters have the same sign and the order of magnitude is the same for similar parameters. The differences do amount to several times the standard error, though: the effect of distance and the effect of a shared border are greater for world trade data. The explanatory power of the model is greater for US data, however. This is expected, given the absence of administrative and physical barriers in the US.

What should have been expected about the differences between the two estimations? Given that the methodology is exactly the same, variations in outcomes must be caused by differences between the two datasets. Firstly, the dataset of Redding and Venables is larger by a factor of four; *ceteris paribus*, this leads to smaller estimation errors. However, their data pertains to the whole world and is probably more heterogenous than that measured within the United States alone. For instance, the distance between two countries is likely to include a stretch of ocean, whereas

this paper.

[7]In the United States, some states share a border of size zero as their corners just touch each other. This is the case for Arizona and Colorado, for instance. In spite of this tangential relationship, the border-dummy is set to one for these pairs of states.

[8]For distance, I measure $\delta_1 = (1-\sigma)\tau_1$, with $\sigma > 1$ and $0 < \tau_1 < 1$. I find that sharing a border increases trade, as does shipping within the region.

this rarely happens between two US states. Given the different nature of trade over sea, there is little reason the expect similar trade costs in the two datasets. Also, two countries sharing a border is a more unlikely event than two states sharing one. This could make the effect of borders more significant in the World dataset. Finally, trade between countries may or may not be hampered by formal restrictions such as tariffs, or by soft barriers such as cultural differences. Given the relative homogeneity of the US states, and the protected status of interstate commerce, I would expect less unexplained variation in the latter sample.

3.2 Second stage estimation: Wages

The results of the previous exercise can be transformed into a measure of regional accessibility, which should be correlated with local wage levels. I construct two new variables: *Market Access* for a region s is defined as

$$\mathrm{MA}_s = \sum_{r=1}^{N} E_r G_r^{\sigma-1} T_{sr}^{1-\sigma} = \sum_{r=1}^{N} \phi_r T_{sr}^{1-\sigma} \tag{7}$$

and *Supplier Access* for region r is

$$\mathrm{SA}_r = G_r = \sum_{s=1}^{N} n_s \left(p_s T_{sr}\right)^{1-\sigma} = \sum_{s=1}^{N} \psi_r T_{sr}^{1-\sigma}. \tag{8}$$

These variables were first used by Redding and Venables (2004). Market access is a weighted average of the expenditures on differentiated goods by the region's potential trading partners. The weights contain distance to the region with a negative sign and the relative isolation of the potential trading partner (as indexed by their price index G) with a positive sign. As such, the measure is reminiscent of the market potential function suggested by Harris (1954). Supplier access in (8) is inversely proportional to the regional price index G_r. It is an index of the ease with which firms in the region can get intermediate goods, and with which consumers can get final goods.

The two variables defined above share two desirable traits: firstly, their values can be computed using only the results of the first-stage estimation. Secondly, formula (3) above implies that they are related to the level of wages in their region. Since the actual wages are observable, this offers us a way to test the theory.

To see how these measures of access interact with the wage level, write (3) as

$$\alpha\sigma \log(w_i) = \zeta + \log\left(\mathrm{MA}_i\right) + (1-\alpha)\frac{\sigma}{\sigma-1}\log\left(\mathrm{SA}_i\right) + \epsilon_i \tag{9}$$

for a region i. Notice that both market and supplier access have a positive coefficient in this equation. Products from a region with low market access incur large transport costs before they reach their customers. As these products have to compete with other, cheaper products, this limits the wages that can be paid in their production. Similarly, low supplier access means that intermediate goods are expensive: this squeezes the value that can be added in a region from the other side.

Computing the values of MA_s and SA_r involves using the estimated values of ϕ and ψ that were obtained earlier, as well as the estimate of the costs of transport. Construct

$$\begin{aligned} \widehat{\mathrm{MA}_r} &= \exp(\phi_r) \cdot \mathrm{dist}_{r,r}^{\delta_1} \cdot \exp(\delta_3) + \\ &\quad \sum_{s \neq r} \exp(\phi_s) \cdot \mathrm{dist}_{s,r}^{\delta_1} \cdot \exp(\delta_2 \cdot \mathrm{bord}_{s,r}) \\ &\equiv \mathrm{DMA}_r + \mathrm{FMA}_r \end{aligned} \tag{10}$$

and

$$\begin{aligned} \widehat{\mathrm{SA}_r} &= \exp(\psi_r) \cdot \mathrm{dist}_{r,r}^{\delta_1} \cdot \exp(\delta_3) + \\ &\quad \sum_{s \neq r} \exp(\psi_s) \cdot \mathrm{dist}_{s,r}^{\delta_1} \cdot \exp(\delta_2 \cdot \mathrm{bord}_{s,r}) \\ &\equiv \mathrm{DSA}_r + \mathrm{FSA}_r \end{aligned} \tag{11}$$

using the estimate of $T_{r,s}^{1-\sigma}$ from the previous section. These formulas implicitly define four *access*-variables by splitting off access to the own region from access to other regions. DMA and DSA are domestic market- and supplier access, and FMA and FSA their foreign equivalents. Separating these terms will allow us to test them separately later on.

I will estimate equation (9) using generated values for MA and SA. These are computed as in (10) and (11), using predicted values for ϕ and ψ. This procedure renders OLS standard errors unusable: the stochastic errors in the gravity equation (6) turn up in the predicted values of MA and SA, which affect the stochastic behavior of ϵ_i in (9), violating the assumptions that underlie standard OLS analysis.

To estimate the standard error in spite of these difficulties, bootstrap methods are available (see Efron and Tibshirani, 1993, for instance). For the gravity equation, I construct a new sample of the same size by drawing random observations (each observation a flow of trade and its regressors) from the original sample. This sample is a bootstrap-replication, in which original observations may be absent, or appear more than once. From the bootstrap-replication I then re-estimate the trade-equation (6) and use the outcome to generate $\widehat{\mathrm{MA}}$ and $\widehat{\mathrm{SA}}$ as usual, which together with observations on the wage make up a sample for equation (9). I generate 200 samples in this way, the conventional number of bootstrap-replications according to Efron and Tibshirani. For each of these samples, I use the same procedure to generate 200 bootstrap-replications. Estimating equation (9) on the resulting data gives forty thousand estimates, from which the standard error of the regressors can be directly observed.[9]

[9] As it turns out, bootstrap standard errors lie between one and two times the (invalid) OLS-standard errors, indicating that the extra variability due to generated regressors is reasonably small. I report only bootstrap-standard errors.

Several other problems potentially plague this estimation. As Redding and Venables remark, a contemporaneous shock to a region that affects both the independent variable and the regressors could introduce a bias into the results. To reduce the possibility of a contemporaneous shock, I estimate using wages from 1999 with regressors from 1997. This does not eliminate another class of third variables, a possible time-invariant region-specific effect that plays into both a region's wage and its market- and supplier access. To correct for this possibility, I report regressions on total access as well as 'foreign access,' as defined in (10) and (11). In the latter regressor, data from the own region does not play a role. Below, I will experiment with adding a number exogenous regressors that proxy for a region's time-invariant attractiveness.

I will estimate the wage equation using the results from regression US 4 in table 1, which is the best-fitting trade equation. This estimate uses the sample of all contiguous states, with trade flows including those to the sending state itself. As it turns out, the Market Access and Supplier Access regressors are highly colinear; the correlation between the two series is 0.95. This means that estimating (9) directly would be problematic. I proceed by using just Market Access as a regressor. At the end of the paragraph I compare the results to those obtained with Supplier Access.

The results of the estimation are in table 2. Once again, the results obtained on the world dataset by Redding and Venables (2004, table 2) are reported alongside the US estimates. A scatterplot of the first two regressions for the United States is in figures 1 and 2. Each point in those plots represents a state, indicated by its two-letter abbreviation. The horizontal axis in figure 1 shows predicted market access according to formula (10). On the vertical axis the log of that state's average annual wage is plotted. Figure 2 is similar, only this time the variable on the horizontal axis is foreign market access.

From the first column of table 2, note that the relation between foreign market access and the level of wages is much weaker in this estimation than in the *World* dataset. Both the explained variation and the statistical significance of the coefficient are smaller. The coefficient is not significantly different from zero. The reasons for this weak performance are evident from the scatterplot in figure 2: while here is a clear positive relationship between FMA and wages for small states, such as Delaware (DE) and Vermont (VT), there are a number of outliers that spoil the correlation. These outliers consist of large states, whose own market is by definition excluded from the foreign market access variable. Especially those that are surrounded by (economically) smaller states fall outside the usual relationship, *e.g.* California (CA) and Texas (TX). This makes sense: explaining the wage levels in California by its proximity to Nevada and Arizona is bound to be problematic, but New Jersey's wage levels certainly have something to do with its wealthy neighbors. Relatively large states disturbing these measurements may be an explanation for the fact that this estimation works better for worldwide data, where the dominance of large states is perhaps less of an issue.[10]

These problems disappear when full market access (MA) is used as a regressor, in the second column. The explained variance is about the same as in the World dataset, as is the statistical significance. This points to a large role for domestic market access, which is confirmed by the final estimation in column three. Even though both coefficients have the correct sign, DMA clearly trumps the insignificant FMA as a regressor for wages.

There may be a problem with the use of full market access as a regressor, though. As local demand in a state is included in this variable, local shocks that affect productivity in a state show up in the regressors as well as in the dependent variable. This causes simultaneity bias in the estimation. Another detrimental effect of including local market access can be seen in the last two rows of table 2. These contain the results of Moran's I test on the residuals of the estimated wage equation. Moran's statistic tests for spatial autocorrelation (see Cliff and Ord, 1973; van Oort, 2002, chapter 4) using a weight matrix to indicate which regions are close to each other. The weight matrix is B, in which entries are equal to one if the two states share a border.[11] Moran's I statistic is computed as

$$I = \frac{N}{\iota' B \iota} \frac{\epsilon' B \epsilon}{\epsilon' \epsilon} \tag{12}$$

with N the number of observations, ι an $N \times 1$ vector of ones and ϵ the $N \times 1$ vector of errors. Table 2 also reports the place of each Moran's I in the distribution of this statistic (under the hypothesis of no spatial autocorrelation).[12] All realizations of the statistic allow us to reject zero spatial autocorrelation at the 1,5% level, indicating that a high realization of the wage in one state makes a higher than expected wage in the bordering states more likely. However, the estimations which include local market access as a regressor show by far the most significant realizations of this statistic.

Are things any different with supplier access instead of market access as an explanatory variable? The model tells us that SA and MA each determine part of the variation in wages, as can be seen in equation (9). However, the pair of regressors suffers from severe multicollinearity so that only one of them can be used in the regression. By the same token, supplier access could have been the ony

[10]According to the BLS (see appendix for data sources), at the end of 1997 California, Texas and New York together accounted for 25% of employment in the USA.

[11]The choice of the weight matrix is, to a degree, arbitrary and its impact should be measured. Alternative statistics have been computed using a matrix B' where $b'_{ij} = \exp(-.001 \cdot \mathrm{dist}_{ij})$ (with dist_{ij} the distance between states i and j). Their level of significance was very close to the values obtained with B.

[12]The expectation of Morans I is $-1/(N-1)$, with N the number of observations. I bootstrap the distribution of I by generating 100,000 vectors ϵ^*, where each ϵ^* is a random permutation of ϵ (in the usual terminology of spatial autocorrelation, I use *nonfree sampling*). I compute the corresponding values of I, and indicate the percentage of outcomes *higher* than the recorded statistic. An asymptotic distribution for the statistic is known (Cliff and Ord, 1973, chapter 2) but its small-sample behavior inspires more confidence in bootstrap methods (see Anselin and Florax, 1995).

regressor. The results of this estimation are in table 3. Comparing the results with those in Redding and Venables (2004), a similar pattern as in table 2 emerges: the regression using only foreign access gives a lower, and less significant, value of the coefficient and a lower R^2 compared to the World data set. Using a full measure of supplier access improves the estimation but leads to higher spatial autocorrelation in the residuals.

The results in this section are less than satisfactory, due to two problems with the estimation. Firstly, unaccounted third factors may play a role in the explanation of wages and introduce bias in the estimates, as Moran's statistics indicate. Secondly, there is reason to believe that simultaneity bias is present. Trying to minimize bias by using only foreign access variables leads to insignificant results, but that may be due to the coarseness of the method. Below, I will adress both problems by adding proxies for first-nature variables in Section 3.3 and by estimating with instrumental variables in Section 3.4.

3.3 Exogenous amenities

When estimating state-level wages as a function of market- and supplier access, all other factors that may also have a bearing on those wages are neglected. In as much as these factors correlate with the regressors, they can cause a bias in the estimation. It is easy to think of a situation in which this may happen.

It is well known that so-called first-nature causes of geographic concentration may play a role in the determination of wages: the physical features of the area, its climate and natural infrastructure all have an effect on productivity. Imagine, for instance, that a predominantly warm climate opens up economic opportunities (*e.g.*, tourism) in a state. This may raise the general level of wages. If a number of neighboring states share the same climate, this third factor will increase wages in all of them. Being close together, market- and supplier access for each of these states will probably be at a comparable level. Suppose it is lower than average; in that case, the unobserved regressor 'climate' causes a downward bias in the present estimates.

In order to test the robustness of the initial estimates against the influence of third factors, this section presents the results of a number of regressions similar to those above, but including a number of possible third factors as regressors. I use the following exogenous amenities:

- **Climate**. In order to control for an exceptionally warm or cold climate I introduce two regressors, normal yearly heating degree days (*nrmhdd*) and normal yearly cooling degree days (*nrmcdd*). The former is defined as the cumulative number of (Fahrenheit) degrees in a year by which the mean temperature of each day falls below 65°F, the latter as the cumulative number of degrees in a year by which the mean temperature lies above 65°F.[13] The idea

[13]Somewhat counterintuitively, cooling degree days measure warmth and heating degree days measure coldness. An example may clarify: if the mean temperature in a state is 67°F all year long,

is that an exceptionally warm or cold climate may account for differences in productivity. For reasons of scale, I divide these regressors by 1000 in the actual regression.

- **Geology**. Special economic opportunities may arise from the presence of precious minerals in a state. To proxy for these opportunities, I use the value of nonfuel mineral production per square kilometer in 1997, as reported by Smith (1997), in thousands of dollars.

- **Access to sea**. Finally, I include a dummy variable that indicates if there exists a deep sea port in the state. Access to sea may proxy for the possibility of international trade.

I expect heating- and cooling degree days, regressors that indicate an unpleasant climate, to have a negative impact on productivity. The presence of minerals is likely good for wages, as is the presence of a port. I first regress wages on these exogenous amenities alone, and then include measures of market access. The results are in table 4.

From the first column of this table, notice that the four exogenous regressors have the expected sign and succeed in explaining about half the variation in wages. However, Moran's I is rather high (higher than all but 2.8% of the distribution under H_0) and indicates that not all region-specific exogenous amenities are present in the set of regressors. The inclusion of Foreign Market Access in the regression hardly changes the values of the earlier coefficients. However, the value of the coefficient for FMA is about a third of the earlier measure (table 2) and lies below one standard error. Explained variation hardly improves, showing that accessibility adds little to the explanation of the first-nature regressors. Including FMA does improve Moran's statistic to a point where it is no longer possible to reject the hypothesis of no spatial autocorrelation at the 5% level.

Things turn out differently when measures of (Domestic) Market Access are included. The coefficients of the exogenous regressors change substantially (more than one standard error in all cases) and Moran's statistic again increases to a significant level. This result again points to problems with the inclusion of Domestic Market access.

3.4 IV estimation

The estimations above may suffer from the occurrence of simultaneity bias, which occurs when the error term from an estimation is correlated with one (or more) regressors. In this matter the underlying model is clearly the culprit, as it indeed allows the error terms to influence the market access regressors. In this Section, I assess the size of the problem and eliminate the bias using instrumental variable estimation.

the yearly cooling degree days are $(67 - 65) \times 365 = 730$ and the yearly heating degree days are zero.

The error terms in the regression imply that observed wages are, to a degree, inconsistent with the model, either because of measurement error or because of misspecification. Where a wage w_i^* is expected in state i, we actually find $w_i = w_i^* + \epsilon_i$. That w_i is the dependent variable in our estimation, but it also makes its way into the regressors; prices are a function of the local wage, and serve as an input into all the price indices G_r. The regressors, MA and SA, are again a function of prices and price indices (*cf.* formulas 7 and 8). This puts the error ϵ_i in the (supposedly) exogenous variables. The question is, whether the weight that ϵ_i receives in MA_i and SA_i is large enough to influence the estimation.

With this problem in mind I used two regressors above, MA and FMA, where the former excludes market data from the own state. The use of local market capacity in the regressor MA will probably introduce ϵ in MA with a large weight. Indeed, I find that the regressions where FMA is used instead of MA show lower spatial autocorrelation of the errors.

However, ϵ can be eliminated from the regressors entirely if instrumental variable estimation is employed. This idea is used in Mion (2004), who takes a panel-approach on Italian data. Brakman et al. (2004) use it to isolate the effect of one particular disturbance in a spatial growth process and Ciccone and Hall (1996) employ four "deep historical" instruments that proxy for the innate attractiveness of American states as places of residence.

For IV, one needs instruments that correlate with the regressors MA and SA, but not with the errors ϵ. I use distance from major economic centers as instruments, in particular the distance from New York City and from Los Angeles.[14] (Redding and Venables, 2004, employ a similar strategy.)

The results are in table 5. The first two columns use only Market Access variables as regressors, and can be compared to the results in table 2. Both coefficients are of comparable size, while standard errors have increased. Foreign market access is insignificant, but the coefficient on (full) market access should now also be free from bias. It is just significant at the 5% level. However, adding the exogenous amenities that were introduced above ruins significance once again.

3.5 Discussion

I have estimated a relationship that explains the levels of wages in the United States by the degree of market access. The variable that indicates market access is itself a construct from the results of a regression, which resembles a gravity-type relationship. To construct the measures of access, heavy use was made of the theoretical model of economic geography.

The estimations mimic those of Redding and Venables (2004), but the results are less satisfying. To a certain extent, this can be explained by the nature of our dataset: it is smaller and possibly more dominated by large regions. The main difference between the two studies is that Redding and Venables (2004) use a world-

[14] As usual, distance is measure from the (employment-weighed) center of the state so that New York and California each have positive distances to these economic centers.

wide dataset, while the present study concentrates on US states. To the extent that (exogenous) productivity varies with geography, it is to be expected that the estimation will work better on worldwide data. This does not necessarily validate the underlying model. However, using only data on US states also brings some advantages, which fail to realize. The dependent variable is actual recorded wages instead of a proxy.[15] Also, institutions are bound to be more similar inside the USA than worldwide. This means that institutional differences (and, for that matter, international frictions such as tariffs) are no longer a factor. These differences were proxied for by distance, but supposedly less than perfectly. In spite of these advantages, the explanatory power of the model, especially when it relies on foreign market access, is much less than that measured on a worldwide scale. The same hold when supplier access is used as a regressor.

The initial estimations suffer from an omitted variable bias that results in spatial autocorrelation of the errors. This problem is addressed by the introduction of an extra set of regressors that proxy for exogenous qualities of each state, and by estimating with IV. This ensures that the active element in the Market Access variables is indeed the access to markets in other states. In these estimations, foreign market access is insignificant, while full market access only contributes when first-nature regressors are left out.

A potential problem with the methodology used above is the fact that the estimated relationships are not neccesarily consistent with the general equilibrium solution of the model. For instance: when estimating the gravity equation in (5), the relation in (2) is parametrized. The latter shows that each region's supply capacity is directly related to the number of firms and the price, both of which are in turn determined by other variables in the model. The same goes for market capacity. However, this relationship is not used in the procedure until much later: only when regional wages are regressed on the access variables do we observe that in fact, the relationships of the model do not hold: if they did, the regression would have had to give us a perfect fit. The variables that were kept constant would not, had they been subjected to the rules of the model, have stayed so.

This leaves us with the question of how to interpret the findings in this section. Economic geography models predict a correlation between wages and accessibility, but the estimation of this correlation may be subject to bias. Different methods have been used to correct for this, but in all the regressions where the issues of bias have been dealt with, access is insignificant. At present, this result points to a lack of evidence for economic geography models in this dataset.

There are many plausible reasons why, even if the real world were governed by this model, this would not result in a perfect final regression. Measurement error, for instance, or the imperfect approximation for transport costs. It remains slightly unsatisfying, however, that the numbers that are used in the estimation are not necessarily an equilibrium outcome of the model. This is especially true in the

[15]Redding and Venables (2004) use GDP per capita for their main estimations, altough they do estimate the relation with wage data for a smaller sample.

class of economic geography models, where for certain parameters a distributed outcome is infeasible, and agglomeration the only stable solution. The problems that are raised in this discussion point to an alternative method for estimating the parameters in a general equilibrium model, one in which all the relationships in the model receive equal weight. I have undertaken such an estimation in related work (see chapter 5 of Knaap, 2004).

4 Conclusions

In this paper, I use intra-American trade data to estimate the effect of a state's accessibility on its wage, using a method previously employed by Redding and Venables (2004). The method assumes certain parts of the model constant and measures the correlation between market access and wage. The dataset covers the US states in 1997 and 1999 and is described below, in appendix A.

I find that the first-stage, gravity, equation gives a reasonable description of the trade between US states. The relationship between wages and the resulting measure of accessibility is problematic, however. After dealing with bias in the estimates, access measures become insignificant. For foreign market access regressors, this is caused by the fact that a few large states dominate the U.S. economy, and their wage level cannot be explained in terms of the economic size of their neighbors. When using instrumental variables and full market access, significance disappears when first-nature regressors are added. These findings point to a very limited role for economic geography models in the explanation of state-level wages in the USA.

A Data

The dataset used in this paper concern the 50 US states in the year 1997. The complete set can be found on the internet, at `http://knaap.com/gdata`. Data was collected from a variety of sources. They are listed here, together with a download address.

- **Employment.** Total nonfarm employment per state, from the Bureau of Labor Statistics. Available at `http://146.142.4.24/cgi-bin/srgate`. Request series SAS*xx*00000000001, where *xx* is the state number.

- **Wages.** Average annual pay for 1999, from the Bureau of Labor Statistics. Available at `http://stats.bls.gov/news.release/annpay.t01.htm`.

- **Interstate flow of commodities.** Bureau of Transportation Statistics 1997 State-to-state commodity flows in millions of US$. Available at `http://www.bts.gov/cfs/cfs97od.html`.

- **Distance between states.** Duncan Black kindly supplied a computer file with the latitude and longitude of each US county. I averaged these into state coordinates, weighing them with county employment. The distance between two states is then computed in miles using the great circle formula. For the distance within a state, I obtained the state area A_i and computed the quasi-radius as $\sqrt{A_i/\pi}$. This number approximates the average distance travelled within a state. State areas may be found at `http://www.census.gov/population/censusdata/90den_stco.txt`.

- **Weather data.** National Climatic Data Center, Asheville, NC. Tables can be found at `http://ols.nndc.noaa.gov/plolstore/plsql/olstore.prodspecific?prodnum=C00095-PUB-A0001`.

- **Mining.** Data from Smith (1997) available at `http://minerals.er.usgs.gov/minerals/pubs/commodity/statistical_summary/871497.pdf`

References

Anselin, L. and Florax, R. J.: 1995, Small sample properties of tests for spatial dependence in regression models: Some further results, *in* L. Anselin and R. J. Florax (eds), *New directions in Spatial Econometrics*, Springer Verlag, Berlin, GE, chapter 2, pp. 21–74.

Brakman, S., Garretsen, H. and Schramm, M.: 2004, The strategic bombing of German cities during World War II and its impact on city growth, *Journal of Economic Geography* **4**(2), 1–19.

Brakman, S., Garretsen, H. and van Marrewijk, C.: 2001, *An Introduction to Geographical Economics*, Cambridge University Press, Cambridge, UK.

Ciccone, A. and Hall, R. E.: 1996, Productivity and the density of economic activity, *American Economic Review* **86**(1), 54–70.

Cliff, A. and Ord, J.: 1973, *Spatial autocorrelation*, Pion Limited, London, UK.

Combes, P.-P. and Overman, H. G.: 2004, The spatial distribution of economic activities in the European Union, *in* V. Henderson and J.-F. Thisse (eds), *Handbook of Regional and Urban Economics, 4*, Elsevier, Amsterdam, NL, chapter 64.

Deardorff, A. V.: 1995, Determinants of bilateral trade: Does gravity work in a neoclassical world?, NBER Working Paper # 5377, Cambridge, MA.

Dixit, A. K. and Stiglitz, J. E.: 1977, Monopolistic competition and optimum product diversity, *American Economic Review* **67**(3), 297–308.

Efron, B. and Tibshirani, R. J.: 1993, *An Introduction to the Bootstrap*, Vol. 57 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York, NY.

Ellison, G. and Glaeser, E. L.: 1999, The geographic concentration of industry: Does natural advantage explain agglomeration?, *American Economic Review-Papers and Proceedings* **89**(2), 311–316.

Fingleton, B.: 2005, Testing the 'new economic geography': a comparative analysis based on EU regional data, Internet: `http://ideas.repec.org/p/wpa/wuwpur/0504003.html`.

Fujita, M., Krugman, P. R. and Venables, A. J.: 1999, *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, Cambridge, MA.

Hanson, G. H.: 2004, Market potential, increasing returns, and geographic concentration, Forthcoming in the *Journal of International Economics*. Mimeo, UCSD. Internet: `http://irpshome.ucsd.edu/faculty/gohanson/market_potential.pdf`.

Harris, C.: 1954, The market as a factor in the localization of industry in the United States, *Annals of the Association of American Geographers* **44**, 315–348.

Head, K. and Mayer, T.: 2002, Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade, CEPII working paper 2002-01. Internet: `http://www.cepii.fr/anglaisgraph/workpap/pdf/2002/wp02-01.pdf`.

Head, K. and Mayer, T.: 2004, the empirics of agglomeration and trade, *in* V. Henderson and J.-F. Thisse (eds), *Handbook of Regional and Urban Economics, 4*, Elsevier, Amsterdam, NL, chapter 59.

Helpman, E.: 1998, The size of regions, *in* D. Pines, E. Sadka and I. Zilcha (eds), *Topics in Public Economics: Theoretical and Applied Analysis*, Cambridge University Press, Cambridge, MA., chapter 2, pp. 33–54.

Knaap, T.: 2004, Models of Economic Geography, Dissertation, SOM, Groningen University. Available at `http://irs.ub.rug.nl/ppn/263135411`.

McCallum, J.: 1995, National borders matter: Canada-U.S. regional trade patterns, *American Economic Review* **85**(3), 615–23.

Mion, G.: 2004, Spatial externalities and empirical analysis: The case of Italy, *Journal of Urban Economics* **56**(1), 97–118.

Niebuhr, A.: 2004, Market access and regional disparities, HWWA discussion paper 269. Internet: `http://www.hwwa.de/Publikationen/Discussion_Paper/2004/269.pdf`.

Overman, H. G., Redding, S. and Venables, A. J.: 2004, The economic geography of trade, production and income: A survey of the empirics, *in* E. K. Choi and J. Harrigan (eds), *Handbook of International Trade, Volume I*, Basil Blackwell, London, pp. 353–387.

Redding, S. and Venables, A. J.: 2004, Economic geography and international inequality, *Journal of International Economics* **62**(1), 53–82.

Roos, M.: 2001, Wages and market potential in Germany, *Jahrbuch für Regionalwissenschaft* **21**, 171–195.

Smith, S. D.: 1997, *Statistical summary, Minerals Yearbook (II)*, US Geological Survey, Dept. of the Interior, Reston, VA.

van Oort, F. G.: 2002, Agglomeration, economic growth and innovation. spatial analysis of growth- and R&D externalities in the Netherlands, Dissertation, Tinbergen Institute, Erasmus University Rotterdam.

Venables, A. J.: 1996, Equilibrium locations of vertically linked industries, *International Economic Review* **37**(2), 341–359.

Wolf, H. C.: 2000, Intranational home bias in trade, *Review of Economics and Statistics* **82**(4), 555–563.

$\log(X_{rs})$	US 1	World 1	US 2	World 2	US 3	World 3	US 4	US 5
Obs.	2601	10100	2201	8079	2601	10100	2091	2042
Year	1997	1994	1997	1994	1997	1994	1997	1997
Estimation	OLS	OLS	OLS	OLS	Tobit	Tobit	OLS	OLS
$\log(\mathrm{dist}_{rs})$	-1.181	-1.538	-1.044	-1.353	-1.330	-1.738	-0.983	-0.987
	[0.056]	[0.041]	[0.025]	[0.032]	[0.063]	[0.043]	[0.024]	[0.023]
bord_{rs}	0.774	0.976	0.492	1.042	0.658	0.917	0.554	0.554
	[0.126]	[0.195]	[0.052]	[0.141]	[0.140]	[0.179]	[0.049]	[0.048]
own_{rs}	2.462	-	2.210	-	2.335	-	2.232	-
	[0.232]		[0.095]		[0.257]		[0.090]	
R^2	0.779	0.789	0.921	0.786	-	-	0.924	0.921
log L	-	-	-	-	-4422	-20306	-	-

Standard errors in parentheses. *World* columns are from Table 1 in Redding and Venables (2004), *US* columns are own computations. Estimation 1 uses the full sample, including zeros. Estimation 2 uses a censored sample, from which the zeros have been eliminated. Estimation 3 again uses the full sample, taking care of the left-censored observations by using a Tobit estimation. Estimation 4 uses only the contiguous states, eliminating Hawaii and Alaska, as well as the District of Columbia. Estimation 5, finally, uses that sample without the within-state flows.

Table 1: Panel estimates for the gravity trade equation

$\log(w_r)$	US	US	US	World	World	World
Obs.	48	48	47	101	101	101
Year	1999	1999	1999	1996	1996	1996
$\log(\text{FMA}_r)$	0.133	-	0.066	0.476	-	0.316
	[0.082]		[0.044]	[0.076]		[0.088]
$\log(\text{MA}_r)$	-	0.257	-	-	0.479	-
		[0.029]			[0.063]	
$\log(\text{DMA}_r)$	-	-	0.119	-	-	0.141
			[0.014]			[0.059]
R^2	0.079	0.601	0.613	0.346	0.610	0.584
Moran's I	0.197	0.317	0.404			
$1 - F(I)$	0.0138	0.0006	0.0000			

US columns are own computations, *World* columns are from Table 2 in Redding and Venables (2004). The dependent variable in World columns is GDP per capita. Bootstrapped standard errors are in parentheses (200 replications). First stage estimation is Tobit for the World columns, US 4 (see table 1) for US columns. Moran's I is computed on the residuals of the estimation, using a matrix of border-dummies as a weighing matrix. On the line below is the position of the statistic in a bootstrapped distribution function (100,000 replications).

Table 2: Market Access and wage levels

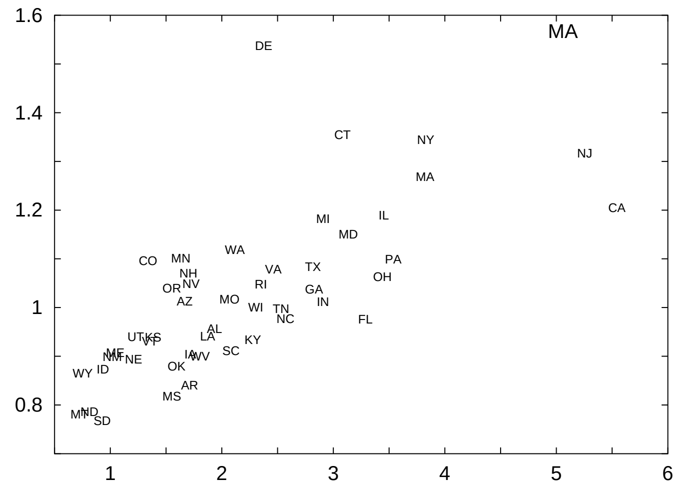

Figure 1: Predicted Market Access (horizontal, based on 1997 data) versus log wages (vertical, data from 1999, wages in ten thousands of dollars) for 48 states. MA regressors come from the US 4 estimation.

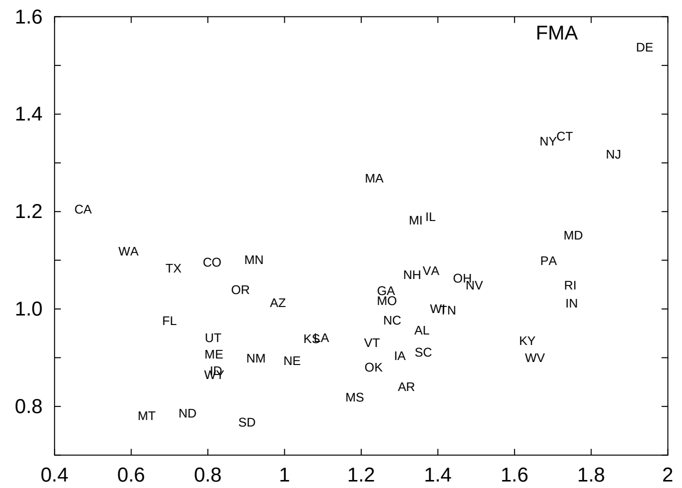

Figure 2: Predicted Foreign Market Access (horizontal) versus log wages (vertical) for 48 states. FMA regressors come from the US 4 estimation.

$\log(w_r)$	US	US	World	World
Obs.	48	48	101	101
Year	1999	1999	1996	1996
$\log(\mathrm{FSA}_r)$	0.118	-	0.532	-
	$[0.082]$		$[0.114]$	
$\log(\mathrm{SA}_r)$	-	0.229	-	0.345
		$[0.030]$		$[0.032]$
R^2	0.075	0.542	0.377	0.687
Moran's I	0.217	0.322		
$1-F(I)$	0.0091	0.0006		

US columns are own computations, *World* columns are from Table 9 in Redding and Venables (2004). See the note under table 2.

Table 3: Supplier Access and wage levels

$\log(w_r)$	US	US	US	US
Obs.	48	48	48	47
Year	1999	1999	1999	1999
$\log(\text{FMA}_r)$		0.042		0.044
		[0.063]		[0.046]
$\log(\text{MA}_r)$			0.234	
			[0.041]	
$\log(\text{DMA}_r)$				0.112
				[0.019]
nrmcdd	-0.103	-0.103	-0.049	-0.063
	[0.044]	[0.049]	[0.036]	[0.032]
nrmhdd	-0.010	-0.010	0.013	0.009
	[0.016]	[0.016]	[0.013]	[0.011]
minerals	0.022	0.021	0.006	0.005
	[0.004]	[0.004]	[0.005]	[0.005]
port	0.130	0.130	0.070	0.044
	[0.035]	[0.036]	[0.029]	[0.031]
R^2	0.545	0.551	0.756	0.776
Moran's I	0.164	0.128	0.230	0.205
$1 - F(I)$	0.0281	0.0586	0.0061	0.0127

Standard errors in parentheses. Except for the first column, these errors come from bootstrap methods (200 replications). First stage estimation for market access variables is US 4 (see table 1). Moran's I is computed on residuals, using a matrix of border-dummies. The position of the statistic in a bootstrapped distribution function is indicated below (100,000 replications).

Table 4: Exogenous amenities, Market Access and wage levels

$\log(w_r)$	US	US	US	US
Obs.	48	48	48	48
Year	1999	1999	1999	1999
$\log(\mathrm{FMA}_r)$	0.169		0.076	
	$[0.101]$		$[0.079]$	
$\log(\mathrm{MA}_r)$		0.232		0.144
		$[0.107]$		$[0.111]$
nrmcdd			-0.103	-0.070
			$[0.057]$	$[0.046]$
nrmhdd			-0.011	0.004
			$[0.019]$	$[0.017]$
minerals			0.020	0.012
			$[0.005]$	$[0.009]$
port			0.129	0.093
			$[0.036]$	$[0.037]$
R^2	0.073	0.599	0.547	0.724
Moran's I	0.163	0.291	0.105	0.145
$1 - F(I)$	0.0299	0.0013	0.0908	0.0423

Instrumental variables estimation. In the first two columns, instruments are the distance from New York and the distance from Los Angeles. In the third and fourth column, the four exogenous regressors are added to the set. Standard errors (in parentheses) come from bootstrap methods (200 replications). First stage estimation for market access variables is US 4 (see table 1). Moran's I is computed on residuals, using a matrix of border-dummies. The position of the statistic in a bootstrapped distribution function is indicated below (100,000 replications).

Table 5: Instrumental variables estimation