An application of the Helpman (1998) model to the Oresund-region

Kristian Mangor
Jonkoping International Business School, Sweden

Tutors:
M. Andersson, M. Backman and V. Jienwatcharamongkhol.

August, 2011

Abstract

This paper investigates the effect of the reduction of transport cost Oresundsbron has caused between the two subregions of the Oresund-region. The paper utilizes the the Helpman (1998) model and the procedure used by Hanson (2005) for estimating this equation. The paper brings a clear inside into the method and thus provides an excellent starting position, with evaluation of possible pitfalls.

Contents

1 Introduction 2
2 Background 4
3 Models and the derivation of the final equation 6
4 Data and method 12
5 Results 14
6 Conclusion 15
A OxMetrics 6.2 code used to estimate 19
1 Introduction

An empirical test of the Helpman (1998) model is presented in this paper with the testing ground being the Oresund-region in Scandinavia. This area has been divided by a strait up until June 2000, where Oresundsbron was completed. Thus an ex-ante and ex-post application of the Helpman (1998) model would yield interesting results in regards to the effect of a reduction in transport cost that the bridge has caused between the two sub-regions. The Helpman (1998) model is part of the economic sub-disciplin called New Economic Geography, which is gaining more and more ground and recognition. This should not come as an surprise since it incorporates the fact that the world is not a homogeneous, flat plain. Unlike e.g. Neoclassical Trade Theory, where trade is a product of comparative advantage, new economic geography produces trade flows due to a love-for-variety effect first introduced by Dixit & Stiglitz (1977) and later utilized by Paul Krugman in his paper from 1979, which is a very powerful simplification that allows for monopolistic competition with many firms. Furthermore, models based on comparative advantage (such as
Ricardo’s, Heckscher-Ohlin model) assumes no transport cost in an attempt to simplify the models. This is arguably both a strong assumption, but also a necessary pre-computer assumption so that the model could be solved analytically and thus possible to evaluate. Today, in the computer era, an analytical solution is neat, but not necessary, and with the rediscovery (and mathematical additions) of iceberg cost by Paul Samuelson in 1954 of von Thünen’s ‘grain eating ox’, transport cost can now be handled rather simple (Samuelson 1983) and the non-linear equation can be calculated.

By the use of the Helpman (1998) model and its equality conditions, this paper looks into the effect of a sudden, but expected, change in transport cost between two sub-regions in the Oresund-region in Scandinavia due to the completion of the Oresund bridge in July 2000, thus attempting to add further empirical research of the Helpman (1998) model. The paper will in total estimate two equations - one before and one after the completion of the bridge. The two equations will be estimated and compared. The reason for using data starting 3.5 year after the completion of the bridge is to let the population and business environment adapt to the new situation, since it was found that the initial effect of the bridge was less than expected (Matthiessen 2004). In the period investigated, the region have had steady economic development without any shocks of significance, except for the completion of Oresundbro. Thus it will be possible to investigate the effect in somewhat isolated circumstances, which ideally would provide new insights into the application of the Helpman (1998) model.

The empirical arena, the Oresund-region The Oresund-region consists of Skåne in southern Sweden and eastern Denmark including Zealand, Lolland-Falster, Møn and Bornholm. Of large cities it includes the Danish capital, Copenhagen, and Sweden’s third largest city, Malmö. A total of 3.732.000 people live in the region (January 2010) of which two-thirds live in the Danish part and the rest in the Swedish part of the region. It is possible to travel without passport between the Scandinavian countries and education and health care is free for all citizens in all the Scandinavian countries, no matter which of these countries they come from. Furthermore, the language are, though different, quite alike and it is of no greater difficulty for a Danish person to understand a Swedish person, or vice versa. The Oresund-region has always been divided by a strait and it was only possible to cross by boat or airplane up until July 2000. This natural barrier combined with the small differences of the two sub-regions, has allowed each sub-region to create it’s own agglomeration and not make a complete agglomeration together. That the opening of the bridge has had some effect is evident from the increased traffic over the strait. Throughout the 1990’s around 2 million vehicles passed the strait per year. In the first couple of years after the opening of the bridge, traffic didn’t match the expectations. In 2010, though, the vehicle traffic had increased to around 9 million per year (Oresund 2010). The above map is of the Oresund-region.
Structure The paper is structured with background in section 2. In section 3 the background models are presented and the equation for estimation is derived. In section 4 the data used and the methodology are discussed and in section 5 the results are presented. Section 6 brings the conclusion.

2 Background

Agglomeration and Transport cost This paper focuses on the effect of change in transport costs, which is shown by Krugman (1991) (among others) to affect the degree of agglomeration. If there was no transport cost, there would be no agglomeration since everything (products, ideas, knowledge) could and would be transported at zero cost and thus no justification for difference in rent. The different forces that cause agglomeration are named centripetal forces, and forces which cause dispersion are called centrifugal forces. In the review article of Krugman (1998) six of these forces are discussed (three in each category):

Centripetal forces
- Market-size effects
- Thick labour markets
- Pure external economies

Centrifugal forces
- Immobile factors
- Land rents
- Pure external diseconomies

The market-size effects are the benefits of a large market, such as how easy it is to find a local supplier or a distributor. Thick labour markets refer to the increased possibility of a good match between firm and employee, thus increasing synergy effects. Pure external economies refer to positive spill overs that agglomeration causes. This includes the concept of ideas, which is a non-rivalrous good, but also the increased local demand for intermediate goods. Of the centrifugal forces, immobile factors include natural resources and land. Land rents are obvious, but are also strongly related to immobile factors. A typical example of pure external diseconomies is congestion, which is a problem seen in almost every mayor city in the world. Congestion is actually a kind of transport cost, but normally the literature refers to transport costs as the cost of transporting between agglomerated areas and not within the different agglomerated areas. Paul A. Samuelson revolutionized transport costs in 1954 by introducing the "iceberg" kind of transport cost, even though it was introduced by von Thünen (1826) who noted that the oxen, who moves the grain, must eat some of it to move it one mile (Samuelson 1983). This was later on formalized in Samuelson (1954). Before its formalization, the most preferred technique to treat transport costs were either to ignore it or to create an entire market for transport costs, which
made models unnecessarily complicated. With iceberg cost a certain part of the traded goods simply disappear - i.e. when sending $y$ products to a trading partner, only $y-x$ arrives (with $0 < x < y$). In the extreme cases where $x=0$ and $x=y$ there is respectively no transport cost and no trade possible.

**Market Potential**  The market potential is the total potential of a given market based on itself and its surrounding markets demand, but negative related to the distance or cost of transport to these surrounding markets. This was first captured in the market potential function of Harris (1954). Market potential is defined as the sum of demand by all $n$ locations, each divided by their respective distance to the $i$'th location:

$$MP_i \equiv \sum_{j=1}^{n} \left( \frac{M_i}{D_{ij}} \right)$$

, where $MP_i$ is the market potential in region $i$, $M_i$ is demand by region $j$ for goods from region $i$, and $D_{ij}$ is the distance between regions $i$ and $j$. The market potential of region $i$ thus depends postively on demand from the other regions, but negatively on distance. The market potential of a 'far away' region in this context thus approaches zero.

**Market Access**  Market access is directly related to market potential and is of crucial importance. However, market access is only build into the models with the distance parameter, $D_{ij}$. This parameter is important, but it does not capture everything. This paper will, however, only use the distance parameter to capture transport cost due to the difficulty of measure other variables and to keep consistency with the models, which is discussed later. The Oresund-region is a border region and transactions across borders are more difficult than in a streamlined region (i.e. a national region), while these differences also can produce transactions of different kind (Schack 1999). Besides the physical transaction cost, other costs could be language differences, law differences and cultural differences. All these are important, but in the Oresund-region the differences are minor. The two languages, Swedish and Danish, are very similar and the cultures are alike, with an understanding of the differences. In regards to the law difference there is obvious great differences, but much has been done in the recent years to streamline important laws, such as tax laws (Matthiessen 2004). However, the cost for crossing the bridge has been found to be a great obstacle for further integration and thus agglomeration in the region (OECD 2003). For the sake of simplicity, this paper will only include the distance parameter. The model which the estimation equations are build upon are that of the Helpman (1998) extension of the Krugman model.

**Reference paper**  In the Hanson (2005) paper a similar equation and approach is utilized. Other papers, such as e.g. Redding & Venables 2004,
Head & Mayer 2006, uses a different approach with bilateral trade data. That however is not possible in this context, which is one of the main reasons that the Helpman (1998) model has been chosen. Another reason for that is to have an actual model to back up the estimation equation. A quick summary of the Hanson (2005) paper follows here as to use later for reference.

In Hanson (2005) the data is based on 3075 counties in the United States from 1970 to 1990. To begin with, a simple market-potential function is estimated, based on Harris (1954). This function is then expanded to an augmented market-potential function, similar to the one in this thesis. Hanson (2005) estimates with non-linear least squares and GMM and all estimations yields plausible and exact results. It is found that distance does affect negatively and that a larger housing sector, higher wages and higher personal income all affect a region positively. The augmented market potential equation is found to have higher explanatory power than the simple market potential function, although relative low. By using the equilibrium conditions from the Helpman (1998) model, Hanson finds that $\sigma$, $\mu$ and $\tau$ all have plausible results.

Hanson also estimates these parameters using the more strict Krugman (1991) model and finds that $\sigma$ and $\tau$ are consistent, but that the estimates of $\mu$ are too large.

3 Models and the derivation of the final equation

The Krugman model New Economic Geography has it origin in the paper of Krugman (1979). The paper introduced a one factor model which differed from other existing theories of trade by having internal economies of scale as the driving force. With the assumption of free entry and exit this allows for monopolistic competition with firm $i$ having the basic production function $l_i = \alpha + \beta x_i$ with $(\alpha, \beta > 0)$. Each firm does therefore only produce one product. The consumers are assumed to prefer all products equally much, which is an approach borrowed from Dixit and Stiglitz (1977) work on monopolistic competition. The utility function is $U = \sum_{i=1}^{n} \tau(c_i)$, with $c_i$ being the consumption of the $i$'th good. The function $\tau(c_i)$ is assumed to be concave with a global maximum, thus allowing an optimum consumption (excluding infinite consumption of good $i$). In equilibrium the model has an optimum number of firms, each earning zero profit and a maximum amount of utility. Krugman showed that technology difference or difference in factor endowments are not needed to provide initiatives for trade - the only thing that is needed is love for variety. However, Krugman (1979) does not incorporate transport costs in any real way, since this would create full agglomeration in one region due to the factor endowment, labour, is perfectly mobile. This was changed in his paper from 1980.
In Krugman (1980) basically the same model is used\(^1\), but the paper investigates the effect of transport costs more thoroughly. These are of the before mentioned iceberg type. When trade opens and there is no transport cost, the only change that occurs is an increase in total utility due to the larger variety of products, as inhabitants will divide their expenditure on products from home and abroad - increasing their utility while real wage stays unchanged. The firm sizes are also unchanged since they are determined by the amount produced of each product, which has the following equation:

\[
 x_i = \frac{\alpha}{(\frac{P}{w} - \beta)}
\]  

(2)

When transport cost is introduced, the only thing that changes is the relative wage rate between the two regions: The relative larger region "A" in regards to region "B", has the higher relative wage (called the home market effect). The paper ends with introducing a variant of the model by including two different kinds of product groups which is demanded by two different population groups. This is done to prove the home market effect.

In Krugman (1991) a model is introduced with two sectors - a manufacturing sector and an agricultural sector. It builds upon the earlier work, but with great modifications. An agricultural sector has been introduced to keep income in each region and so as to be used as numeraire. The production function of the industrial sector is the same as in the previous papers and the basics of the utility function for manufactured products are also similar. Furthermore, the iceberg transportation cost is also applied. In this model, output per firm is fixed due to the increasing returns to scale production and free entry and exit of firms. The model thus describes the effect of transport costs upon agglomeration and what other underlying factors that might influence it. There is no real dynamics in the model, but it yields some end results given different transport costs, which is the outcome of the model: For high transport cost (and other parameters stable at a specific value) the outcome of the model is that manufacturing workers are divided between the two regions in an equal way to agriculture. When transport costs are low, we have agglomeration of manufacturing workers in one region.

The model utilizes the Dixit & Stiglitz (1977) method of modeling monopolistic competition. The utility curve is a basic function, where utility of the two products (manufacturers and agricultural products) each yields some utility, given the value of \(\mu\):

\[
 U = C_M^\mu C_A^{1-\mu}
\]

(3), which is a shared utility function for the total population of the economy. The total consumption of manufacturers are determined with the following

---

\(^1\)The utility function \(U = \sum_{i=1}^{n} \tau(c_i)\) is replaced by \(U = \sum_{i=1}^{n} \theta\) with \(0 < \theta < 1\), which basically produce the same result.
Figure 1: Marginal utility falls with increased consumption of one product.

\[ C_M = \left( \sum_{i=1}^{N} c_i^{(\varepsilon-1)/\varepsilon} \right)^{\varepsilon/(\varepsilon-1)} \]  

, with \( \varepsilon > 1 \). This secures that all products are consumed, since consumption has concave utility but the summing is linear (for illustration, see figure 1).

The production of agricultural products is assumed to be linear and there is no transport cost connected to this sector (therefore good to use as numeraire). The model assumes that population is one and that peasants (employees in the agricultural sector) is equal to \((1 - \mu)\) divided equally to both regions. Peasants are also assumed immobile. This ensures, as mentioned earlier, that all regions always have an income and thus meaning in the model. Employees in the second sector, manufacturers, are called workers and are assumed to be mobile. Their total number adds up to \( \mu : \mu = L_1 + L_2 \). The production function in the manufacturing sector is assumed to be of increasing returns to scale:

\[ L_{Mi} = a + bx_i \]  

, where \( L_{Mi} \) is the amount of labour needed to produce \( x \) amount of product \( i \). Since there is free entry and exit in the model, and given the utility curve,
there will be zero profit. The transport cost is positive in this sector and is simplified to the inverse of the beforementioned iceberg kind, so that $\tau < 1$. The price in region $j$, $p_j$, is therefore equal to:

$$p_j = \left( \frac{\varepsilon}{\varepsilon - 1} \right) \beta w_j.$$  \hfill (6)

Given the assumption of free entry and exit, the output per firm must be the same in all regions, with the output function looking like this:

$$x_1 = x_2 = \frac{\alpha(\varepsilon - 1)}{\beta}$$  \hfill (7)

When estimating the market-potential function this paper will, as Hanson (2005), utilize much of the above model, but with the modification made by Helpman (1998). It is done since Hanson (2005) estimates equations based on both models and finds the Helpman (1998) model to be more correct. Thus the estimation procedure will utilize a lot of Krugman’s work, but with the modification made by Helpman (1998), where the agricultural sector is replaced with a housing sector to improve the model.

**The Helpman (1998) extension** The housing sector in Helpman (1998) is introduced as a non-tradable factor that works as a centrifugal force instead of the agricultural sector in Krugman (1991). People are assumed to freely be able to move between regions. It is also assumed that all activities of the person is done in the region which it resides, i.e. all purchasing (manufacturers, housing) and working is done in the region of residence. The larger a region, the greater the variety of locally produced goods due to the same monopolistic competition framework used in Krugman (1991). This produces the two opposing forces of the Helpman (1998) model. The centrifugal force is the cost of housing. The larger the region compared to housing stock, the more expensive housing is. The centripetal force is the greater variety of goods. The utility function of these two kind of goods is basically the same as Krugman (1991), but with housing replacing agricultural products.

With its foundation in the above theory, the following equilibrium equations will be used. The equilibrium condition for the housing market:

$$P_k H_k = (1 - \delta) Y_k$$  \hfill (8)

where $P_k$ is the price of housing in region $k$ and $\delta$ is the share of income spent on manufacturers. The assumption that real wage will equalize between the regions are also used:

$$\frac{W_r}{(P_r^{1-\delta} I_r^\delta)} = \frac{W_k}{(P_k^{1-\delta} I_k^\delta)}$$  \hfill (9)

where $I$, which is the price index, is measured by: $I = \sum_{j=1}^{N} p_j$. 


Furthermore, the consumers maximizes' their utility of manufacturers subject to the income constraint; the income spent on manufacturers $\delta Y$:

$$C_M = \left( \sum_{i=1}^{N} x_i^{(\varepsilon-1)/\varepsilon} \right)^{\varepsilon/(\varepsilon-1)}$$

subject to $\delta Y = \sum_{i=1}^{N} p_i x_i$

which yields the first order condition (F.O.C.):

$$\left( \sum_{i=1}^{N} x_i^{(\varepsilon-1)/\varepsilon} \right)^{-1/(\varepsilon-1)} x_i^{1/\varepsilon} - \mu p_i = 0$$

And the second order condition (S.O.C.) verifies a maximum; S.O.C. < 0. By taking the ratio of F.O.C. for product $i$ to the F.O.C. of product 1, we gain the equation for the amount of product $i$, given price and quantity of product 1: $x_i = p_i^{-\varepsilon} \hat{p}_i x_1$. Inserting this into the budget constraint and using the definition of the price index, the demand for product 1 is found:

$$x_1 = p_1^{-\varepsilon} I^{\varepsilon-1} \delta Y$$

Product $x_1$ is demanded in its own region and the surrounding regions, but the demand from the surrounding regions are reduced by the transport cost. The general demand for product $s$ can thus be stated as following:

$$x_s = p_s^{-\varepsilon} \delta \left( \sum_{r=1}^{R} Y_r I_r^{\varepsilon-1} T^{D_{rs}} \right)$$

where $T^{D_{rs}}$ is equal to 1 when $r = s$.

Given the free entry and exit of firms, profit is zero and supply is equal to demand. Thus, replacing $x_s$ with $\frac{\alpha(\varepsilon-1)}{\beta}$, the following equilibrium equation is obtained:

$$\frac{\alpha(\varepsilon-1)}{\beta} = p_s^{-\varepsilon} \delta \left( \sum_{r=1}^{R} Y_r I_r^{\varepsilon-1} T^{D_{rs}} \right)$$

To find the wage in one region the mark-up pricing rule (eq. 6) is used. Isolate for $p$ in eq. 6 and use this to replace $p$ in eq. 14 such that:

$$\frac{\alpha(\varepsilon-1)}{\beta} = \left( \frac{w \beta}{1-\varepsilon} \right)^{-\varepsilon} \delta \left( \sum_{r=1}^{R} Y_r I_r^{\varepsilon-1} T^{D_{rs}} \right)$$

Solving for $w$ we get:
\[ w = \left( \beta^{\frac{\varepsilon+\varepsilon}{b-1}} - \varepsilon \right) \left( \frac{\delta}{\alpha(\varepsilon-1)} \right)^{1/\varepsilon} \left( \sum_{r=1}^{R} Y_r I_r^{\varepsilon - 1} T^{D_{r,s}} \right)^{1/\varepsilon} \]  

(16)

The first parenthesis is normalized\(^2\) and the estimation equation is produced:

\[ w = \left( \sum_{r=1}^{R} Y_r I_r^{\varepsilon - 1} T^{D_{r,s}} \right)^{1/\varepsilon} \]  

(17)

The next step is to get rid of the price index and introduce the price of the housing markets and the nominal wages for the \( R \) regions. This is done by first inserting the equilibrium condition for the housing market (eq. 8) into the real wage equalization condition (eq. 9), where the right hand side is replaced by \( \bar{w} \)\(^3\) and:

\[ \frac{W_r}{\left( \frac{(1-\delta)Y_r}{H_r} \right)^{1-\delta} I_r^\delta} = \bar{w} \]  

(18)

Isolating for \( I_r \) and inserting in the wage equation (10) so that the equation can be estimated, the following is done:

\[ W_j = \left( \sum_{r=1}^{R} Y_r \left( \frac{W_r}{\bar{w} \left( \frac{(1-\delta)Y_r}{H_r} \right)^{1-\delta}} \right)^{\varepsilon-1} T^{D_{jr}(1-\varepsilon)} \right)^{\frac{1}{\varepsilon}} \]  

(19)

Next step is to get all the variables with \( r \) as an index out:

\[ W_j = \left( \sum_{r=1}^{R} \bar{w}^{\frac{1-\delta}{b}} \left[ 1 - \delta \right] \varepsilon Y_r^{\frac{1-\delta}{b}} W_r^{\frac{\varepsilon-1}{b}} \frac{H_r^{(1-\delta)(\varepsilon-1)}}{T^{D_{jr}(1-\varepsilon)}} \right)^{\frac{1}{\varepsilon}} \]  

(20)

and the rest are constants, so that:

\[ W_j = C \left( \sum_{r=1}^{R} Y_r^{\frac{1-\delta}{b}} W_r^{\frac{\varepsilon-1}{b}} \frac{H_r^{(1-\delta)(\varepsilon-1)}}{T^{D_{jr}(1-\varepsilon)}} \right)^{\frac{1}{\varepsilon}} \]  

(21)

The constants are taken out of the paratases and then log is taken, which yields our augmented estimation equation:

\[ \log(W_j) = c + \varepsilon^{-1} \log \left( \sum_{r=1}^{R} Y_r^{\frac{1-\delta}{b}} W_r^{\frac{\varepsilon-1}{b}} \frac{H_r^{(1-\delta)(\varepsilon-1)}}{T^{D_{jr}(1-\varepsilon)}} \right) \]  

(22)

---

\(^2\) See Brakman et al. page 118.

\(^3\) To simplify the derivation.
, which will be estimated in the reduced form of equation (23) - see below. With equation (23), the \( \alpha \) parameters can be estimated and then solved in a linear system to obtain the values for \( \mu, \varepsilon \) and \( \tau \). This is done to check if the parameters show plausible values. The term \( e \) is used instead of \( T \) to have an actual value. It will yield qualitatively the same, since distance cannot be negative and zero distance also makes \( e = 1 \).

**Summation** In the above text the models has been outlined, the equation for estimation as been derived, the area in context has been described and supporting theory presented. The final estimation equation, which tests the hypothesis, is based on these models.

**Hypothesis** "Has Oresundsbron, in the settings of the Helpman (1998) model, increased the wages and thus the market potential in the Oresund-region, Scandinavia?"

To support the hypothesis the estimated parameters in the estimation equation:

\[
\log(W_j) = \alpha_0 + \alpha_1 \log \left( \sum_{r} Y_r^{\alpha_2} W_r^{\alpha_3} H_r^{\alpha_4} e^{D_r, \alpha_5} \right)
\]

should have the following attributes: \( \alpha_1 \) is expected to be between 0 and 1, \( \alpha_2 \) is unknown, \( \alpha_3 \) is expected to be positive, \( \alpha_4 \) is expected to be positive and the distance parameter \( \alpha_5 \) is expected to be negative.

4 Data and method

**The parameters** Distance is measured by travel time. This way of capturing distance is superior to that of concentric distance bands as Hanson (2005) uses because it takes differences in roads and natural barriers into account. In regards to time versus kilometers traveled, the differences must be suspected to be minor. Furthermore, the assumption that people choose the quickest route instead of the shortest seems plausible.

The times are calculated from the economic center (largest city) in each region to the economic center of another region. Google Maps have been used to calculate all time distances. In regards to travel time between Denmark and Sweden, the time from each region to where the bridge begins (on either side) has been calculated and either 60 or 10 minutes have been added for the data respectively before and after the bridge. Furthermore, travel between the northern parts of both the Danish and Swedish part of the Oresund-region has been assumed to be done by ferry between Helsingor and Helsingborg instead.

---

4Based on information from Matthiessen (2004).
Table 1: Data presentation - all monetary terms are in DKK and distance in minutes.

<table>
<thead>
<tr>
<th>Year 1999</th>
<th>Y</th>
<th>W</th>
<th>H</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean value:</td>
<td>4925754455</td>
<td>212209</td>
<td>21003</td>
<td>99.95</td>
</tr>
<tr>
<td>Median value:</td>
<td>2757500000</td>
<td>223086</td>
<td>13451</td>
<td>97</td>
</tr>
<tr>
<td>Std. deviation:</td>
<td>9868789256</td>
<td>49002</td>
<td>33900</td>
<td>58.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year 2004</th>
<th>Y</th>
<th>W</th>
<th>H</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean value:</td>
<td>5716867469</td>
<td>242810</td>
<td>21370</td>
<td>78.07</td>
</tr>
<tr>
<td>Median value:</td>
<td>3256000000</td>
<td>258807</td>
<td>13932</td>
<td>75</td>
</tr>
<tr>
<td>Std. deviation:</td>
<td>11580175735</td>
<td>57450</td>
<td>33864</td>
<td>40.84</td>
</tr>
</tbody>
</table>

and with more regions doing this before than after the bridge. The travel time assumed for this route is 30 minutes\(^5\).

Income (Y) is total income in a region. This varies substantially between regions. The wage variable (W) is the income variable divided by number of employed persons in a region. Housing stock is number of housing in a region. This could be either a small one-room apartment to a penthouse apartment or a villa. A better piece of data might have been total number of housing square meters in a given region. However, that data was not available. Furthermore, in the basic setting of this model it could be assumed that a place to sleep and eat is all an employee needs besides work. This might be seen as a strong assumption, but then again - if the model should include preferences for housing it would quickly become extremely complicated, since some people would like to live near water, others in the center of a city, others again in small house or large house and so forth. The Danish and Swedish housing data set are gathered respectively from Danmarks Statistik and Statistiska Centralbyrån.

**Estimation**  The estimation procedure is in principle simple: Minimize the errors while iterating on the 78 equations (due to the 78 regions) until the lowest possible residuals occur of the following equation:

\[
\log(W_j) = \alpha_0 + \alpha_1 \log \left( \sum^R W_r^{\alpha_2} H_r^{\alpha_3} e^{D_j \cdot \alpha_4} \right) + \sigma
\]

, with \(\sigma\) being the error term to minimize. This is, of course, done with the parameters being the same for all 78 equations and starting values are given (otherwise we would most likely see a non-optimal solution). The starting values were chosen based on the estimated values from the results from Hanson (2005), but with modifications along the way to improve the

\(^5\)The ferry trip is just over 20 minutes and there is 4 connections per hour in daytime (thus average waiting time 7.5 minute in daytime). For interested, their current homepage is: http://www.scandlines.com/en.
estimation. For example, it was found that the initial values for the distance parameter had to be positive in order to force the minimization process to find the optimal value (otherwise it would stick to the given initial value, no matter its size). The program for estimation was made in OxMetrics 6.2 and consisted basically of two loops, one outer for each equation and one inner for the summation of the products. The estimation was in other words done with programming the actual way the calculations would have taken place (see Appendix A for the code). The errors were tested in EViews 7 with a simple Jarque-Bera test to measure their normality.

(Dis)advantages As apparently most statistical software packages cannot compute the above equation themselves and programming therefore is needed, the initial work is relative larger. Furthermore, easily used tests in most statistical software packages are not easily applied here unless they are already programmed. Of this reason, I have chosen to add the OxMetrics 6.2 code in the appendix, so that other people who wish to continue the work in this empirical arena has a better starting position.

In the context of the chosen model (Helpman 1998), there is several advantages. First of, the model of Helpman (1998) supplies the possibility to improve on something concrete. The Helpman (1998) model is thus a skeleton to change, add-on and improve. This is unlike unformulated models, where there is no center for the research. Secondly, the Helpman (1998) model exhibit the general patterns that are seen and it connects the variables in use. Non-model based equations which includes variables that seems likely to affect might be very good, but they demand a large amount of testing to be sure the explanatory power is correct or close to. The Helpman (1998) model allows for extra testing, since the basic parameters of \( \varepsilon \), \( \tau \) and \( \varepsilon \) can be found and evaluated if need be. An advantage which should be used if good results are obtained.

5 Results

The results can be seen in Table 2, with the standard errors in parentheses. For ease of reference the parameters will be restated: \( \alpha_0 \) is the constant, \( \alpha_1 \) is the inverse degree of substitution, \( \alpha_2 \) is the power of the income variable, \( \alpha_3 \) is the power of the wage variable, \( \alpha_4 \) is the power of the housing variable and \( \alpha_5 \) is the distance parameter. Their expected intervals are also shown. As can be seen the distance parameter (\( \alpha_5 \)) is indeed negative as expected, which is consistent with the theory and earlier findings. Compared to Hanson (2005) the distance parameter is 'less' negative (lower in absolute value). This might be due to a smaller area. In regards to the other parameters (disregarding the constant) all change signs from 1999 to 2004 and all are insignificant, which is problematic. The reason for that is that the parameter values for the year 1999 stays within its expected values, while that they do not for 2004. The distance parameter is insignificant, but here it must be noted that the
Table 2: The estimated values.

<table>
<thead>
<tr>
<th>Variable:</th>
<th>Constant</th>
<th>$\varepsilon^{-1}$</th>
<th>Y</th>
<th>W</th>
<th>H</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. Value:</td>
<td>-</td>
<td>[0; 1]</td>
<td>$\neq 0$</td>
<td>$&gt; 0$</td>
<td>$&gt; 0$</td>
<td>$&lt; 0$</td>
</tr>
<tr>
<td>Year 1999</td>
<td>0.057196</td>
<td>0.17352</td>
<td>-0.73146</td>
<td>6.6084</td>
<td>0.53949</td>
<td>-1.4992</td>
</tr>
<tr>
<td>(std. error)</td>
<td>(0.41542)</td>
<td>(0.89134)</td>
<td>(3.7587)</td>
<td>(33.946)</td>
<td>(2.7733)</td>
<td>(36.293)</td>
</tr>
<tr>
<td>Year 2004</td>
<td>0.43390</td>
<td>-0.63302</td>
<td>0.12703</td>
<td>-1.6990</td>
<td>-0.062791</td>
<td>-1.7832</td>
</tr>
<tr>
<td>(std. error)</td>
<td>(0.34775)</td>
<td>(9.1306)</td>
<td>(1.8330)</td>
<td>(24.507)</td>
<td>(0.90664)</td>
<td>(61.737)</td>
</tr>
</tbody>
</table>

non-linear estimation procedure yielded 'no convergence' when the initial value was too large (and positive). When $\alpha_5$ was a low, but positive value, it converged to a negative value with 'strong convergence' (below 8,500 iterations). Thus it must be extracted that in the given context, the distance parameter must be negative.

The residuals were tested in EViews 7 and for both years the Jarque-Bera test could not reject normality (see figure 2 and 3). As can be seen in the two figures, the Jarque-Bera tests yields a certainty above 0.05. Both have means which essentially is zero, as it should be. Both error-sets shows a larger degree of negative extreme values than positive, but within acceptable limits. Thus, the estimation procedure produced valid errors.

6 Conclusion

That the results are insignificant and in general inconsistent reduces the quality of the results of this study. The problem of insignificance can only be handled with a larger population, something that was never possible in this study due to its context. Furthermore, it does not change that the estimates are the most likely estimates based on the sample size. However, that the
estimates in general are inconsistent is a problem: $\alpha_1, \alpha_2, \alpha_3$ and $\alpha_4$ all change signs from one year to another, from being within the expected values in 1999 to being outside the expected values in 2004.

The $\alpha_5$ parameter, which is the weight of transport cost, is consistently negative and thus support the theory that transport cost negatively affects the wage in a region. However, it is also insignificant and it increases in absolute value - not decreases - as predicted by the theory. Thus the completion of the bridge has affected the market potential and wages negatively, which rejects my hypothesis. This rejection, though, is based on insignificant results. The best possible explanation for the rejection (beside that the sample being too small, yielding insignificant results) is that only a few regions have had increased wages while most have been affected less positively from the building of the bridge.

As an end remark it should be noted that all iterations of the data set yielded negative $\alpha_5$ and that no positive value of $\alpha_5$ appeared with a response of 'good convergence' (thus all these had 'no convergence'). Here, this paper finds a small pocket of support for the theory that transport costs affects negatively, while there is no evidence opposing the theory.

As such, the results themselves is of minor use in the scientific world and this paper greatest contribution to that is the ready set application of the Helpman (1998) model. Furthermore, it does support the theory in that transport cost affects negatively and shows that the Helpman (1998) is of minor use in measuring a single transport improving project and should be used for larger areas. Also, this paper barks the notion and transport cost should be measured in time distance instead of concentric distance bands or other rough estimation methods. Time distance is what matters to people, not if they are within e.g. a 100 kilometer in bird flight.
Further research  This paper gives a starting position to expand the empirical testing ground for the Helpman (1998) model, which is recommended so the model can be improved and increasingly validated. Further research should apply a larger sample to insure statistical significant results. The Hanson (2005) paper had significant results by including 3075 counties, while this paper had insignificant results with 78 regions. Furthermore, the future research is encourage to use time distance instead of e.g. concentric distance bands - especially if they work in e.g. mountainous areas or others which would create significant different travel time compared to the flight distance. And lastly, the code below is to free use and editing.
References


A OxMetrics 6.2 code used to estimate

```c
#include <oxstd.h>
#include <oxfloat.h>
#import <maximize>

static decl s_vY, s_mX, s_iEval = 0, vEps;

Initialize(const mDat, const avY, const amX, const avP)
{
    decl iN, iK;
    iN = rows(mDat);
    iK = columns(mDat);

    avY[0] = mDat[][1];
    amX[0] = mDat[][5:8];

    avP[0] = < 0.1 ; 2.14980 ; -0.053677 ; 0.1795 ; 0.26530 ; 2.000>;

    println("Initial parameter values 
are: \z", "Xr", {"a0", "a1", "a2", "a3", "a4", "a5"}, avP[0]);
}

AvgLnLiklRegr(const vP, const adLnPdf, const avScore, const amHess)
{
    decl vLnL, iN, dAlpha, vBeta, dA0, dA1, dA2, dA3, dA4, dA5,
    dA0 = vP[0];
    dA1 = vP[1];
    dA2 = vP[2];
    dA3 = vP[3];
    dA4 = vP[4];
    dA5 = vP[5];

    dC = 0;
    dKLoop = 0;
    iN = 78;
    vw = s_vY;
    vY = s_mX[][2];
    vW = s_mX[][1];
    vH = s_mX[][0];
```
vD = s_mX[] [3];
vEps = constant(.NaN, iN, 1);

for(j = 0; j<iN; j++)
{
    dKLoop = 0;
    for(k = 0; k < iN; k++)
    {
        dKLoop = dKLoop + (vY[k].^dA2 .* vW[k].^dA3 .* vH[k].^dA4 .* exp(vD[dC] .*dA5 ));
        dC = dC + 1;
    }
    vEps[j] = log(vw[j*(iN-1)]) - (dA0 + dA1*log(dKLoop));
}

adLnPdf[0] = -vEps'*vEps;

Estimate(const avP, const adLnPdf)
{
    decl vS, ir;
    ir = MaxBFGS(AvgLnLiklRegr, avP, adLnPdf, 0, 1);
    return ir;
}

Output(const ir, const vP, const dLnPdf)
{
    decl mHess, mS2, dSigma, dStdError;
    println("The number of calls of the log likelihood function = ", s_iEval);
    println("MaxBFGS returns ", ir, ", meaning ", MaxConvergenceMsg(ir), ".");
    Num2Derivative(AvgLnLiklRegr, vP, &mHess);
    mS2 = invert(mHess*mHess')/(rows(vEps));
    dSigma = (1/rows(vEps))*vEps'*vEps;
    mS2 = dSigma*mS2;
    dStdError = sqrt(diagonal(mS2));
    println("Optimal parameter values are: \r", "\%r", {"a0", "a1", "a2", "a3", "a4", "a5"}, ", ";, "%c", {"Estimates", "Std Error"}, vP"dStdError");
}
println("The epsilon vector equals in this case: \r", vEps);
println(mHess);
}

main()
{
    decl mDat, vP, ir, dLnPdf, avScore, amHess, iK, vPTr;
    mDat = loadmat("C:\Documents and Settings\...\File.xls");
    Initialize(mDat, &s_vY, &s_mX, &vP);
    ir = Estimate(&vP, &dLnPdf);
    Output(ir,vP,dLnPdf);
}