

## Measures of labour market accessibility. What can we learn from observed commuting patterns?\*

Arnstein Gjestland<sup>1</sup>, Liv Osland<sup>1</sup> and Inge Thorsen<sup>1</sup>

<sup>1</sup> Western Norway University of Applied Sciences

Received: 13 February 2019/Accepted: 28 February 2020

**Abstract.** It is well known that measures of labour market accessibility explain spatial variation in housing prices, even in markets with polycentric labour market structures. This paper examines whether data on observed commuting patterns can replace or supplement gravity-based measures representing the commuting potential at specific locations. We use data from a region in Western Norway, and we find that measures based on observed commuting flows and commuting time cannot replace a gravity-based measure of labour market accessibility. Based on, inter alia, the spatial Durbin estimator we find that measures of observed commuting flows increase the explanatory power of a hedonic house price model.

### 1 Introduction

The relationship between house prices and access to workplaces is a central theme in both theoretical and empirical housing market research. There are many reasons why this relationship is important. Travelling to work is a regular and bounded trip. According to [Vågane et al. \(2011\)](#), travelling to work constitutes 18 per cent of all travels in Norway. During workdays, approximately 25 per cent of the travels are journey to work. Most commuters travel at the same point in time every day, which also creates congestion. Road transportation infrastructure is often given a capacity to deal with such traffic peaks. At the same time, investments in the road network affect labour market accessibility, which is in turn capitalized into house prices. In order to reduce many of the transportation problems related to commuting, planners may seek to locate houses in areas where job accessibility is assumed to be high.

According to [Handy, Niemeier \(1997\)](#), there is no consensus in the literature on a good measure of accessibility. In explaining housing prices, gravity-based accessibility measures have been suggested as a generalization of modern polycentric labour market structures. Although more recent research has shown that gravity-based accessibility measures explain significant spatial variation in housing prices, [Handy, Niemeier \(1997\)](#) show that the gravity-based measures are not the only weighted measures that can be used to capture the job opportunity density of a given area.

Gravity-based measures of labour market accessibility reflect the potential for commuting from a specific residential area. In this paper we introduce three other measures, based on actual commuting patterns. One measure is origin specific: the percentage of the total working population living in the zone and working in a different zone. The second measure

\*We are grateful to two anonymous reviewers for very helpful comments.

is destination specific: the percentage of people working in the zone that are living in another zone. The third is based on the calculation of actual commuting time in each zone. The proposed measures are related to labour market accessibility. Our ambition has been to test empirically whether the measures can replace a gravity-based measure of labour market accessibility, or whether they should supplement such a potential measure, adding relevant information on spatial characteristics in order to explain housing prices.

The measures based on observed commuting flows have some advantages relative to the potential measures of job opportunities. The new measures can be more easily explained to non-experts in the field, and they are computationally simpler than the gravity-based measures, which may involve non-linear methods of estimation. Some of the measures of observed commuting flows are also less demanding in terms of data requirements, for example, travelling times and the transportation network.

Simplicity and data requirements are not, however, the most important issues in favour of incorporating measures based on observed commuting flows. Such measures potentially offer a kind of market-based evaluation of characteristics relevant for explaining housing prices. The values of a gravity-based measure represent the potential of making favourable labour market decisions, in terms of the traveling time between the residential location and the job location. In contrast, the other measures we consider are based on labour market decisions that have actually been made. Our basic hypothesis is that locations offering favourable labour market opportunities capitalize into the housing market, see e.g. [Gjestland et al. \(2014\)](#). A main motivation of this paper is to study whether observation-based information can substitute, or maybe supplement, the gravity-based potential measure in explaining spatial variation of housing prices. According to [Handy, Niemeier \(1997\)](#), “The fundamental issue is that an accessibility measure is only appropriate as a performance measure if it is consistent with how residents perceive and evaluate their community. In other words, a practical definition of accessibility must come from the residents themselves.” (p. 1176). In view of this citation, estimation of hedonic house prices can be a useful tool. Assuming market equilibrium, this method can be characterized as a revealed preference approach. It enables the measurement of the implicit prices of goods and amenities that are not directly traded in any markets. As such, this approach can be an appropriate framework for evaluating how alternative measures of potential and observed labour market interaction contribute to explain house prices.

The paper is outlined as follows. Section 2 contains a brief literature review. Thereafter, in Section 3, we present the study area and formulate explicit hypotheses to be tested. In Section 4 we present the data while the empirical results and the analyses are presented in Sections 5–7. Finally, conclusions based on our findings follows in Section 8.

## 2 A brief literature review

The most widely accepted theory that links residential location to the price of housing is given by urban economic theory represented by the monocentric city model. The relevant prediction of this model is that households living far from the centre of employment are compensated for higher costs of commuting by way of a lower price for housing.

In the housing market literature, accessibility has traditionally been accounted for by the simple measure of distance to the central business district (CBD) (see e.g., [Ball, Kirwan 1977](#), [Dubin 1992](#)). It is, however, well acknowledged in the literature that the monocentric model frequently has not been supported by empirical evidence. Many reasons have been suggested for the disparity between theory and empirical results. One obvious suggestion is the polycentric pattern of employment ([Anas et al. 1998](#)). In spite of this, there are in fact relatively few papers that focus on how polycentrism may affect property values. One natural suggestion to cope with polycentrism is found in [Waddell et al. \(1993\)](#), who include both the distance to the CBD and the distances to secondary employment centres. One potential problem with this approach is that the researcher has to choose which employment centres to include. Because of problems with spatial multicollinearity and interpretation of partial effects, it may not be straightforward to include distances to many employment nodes as separate variables in an empirical hedonic

house price model. See, however, [Heikkilä et al. \(1989\)](#) for possible ways of dealing with this issue.

The potential, gravity-based, measures of accessibility ([Handy, Niemeier 1997](#)) are frequently used in the literature. To cite [Anselin \(2002, p. 250\)](#), these variables are specified so that “the potential for interaction between an origin  $i$  and all destinations  $j$  was formulated as a sum of ‘mass’ terms in the destination, suitably downscaled by a distance decay function”. For a useful general discussion on the use of the accessibility concept in spatial analysis, see [Kwan et al. \(2003\)](#). [Farber et al. \(2013\)](#) discuss metrics based on the time-geographical concept of joint accessibility for measuring the spatial interaction potential of a region. However, in this paper, we focus on the use of gravity-based accessibility measures that have been suggested as a generalization of modern polycentric labour market structures ([Heikkilä et al. 1989](#)). Nevertheless, there are not many papers that relate gravity-based accessibility measures to housing prices.

The evaluation of gravity-based measures differs in the literature. [Jackson \(1979\)](#) does not find evidence of the dominance of either the CBD-gradient or the gravity-based employment index. [Adair et al. \(2000\)](#) find heterogeneous results. In the overall Belfast Urban Area, the gravity-based accessibility had small or negligible effect, while stronger effects were found by repeating the analysis at the sub-market level. By using Norwegian housing price data from a wider labour market area, the accessibility measure was clearly significant in [Osland, Thorsen \(2008\)](#). [Osland, Pryce \(2012\)](#) use housing price data from Glasgow. The employment data were from all Scottish data zones. They found a highly significant non-monotonic relationship between house prices and access to employment. According to this research, house prices would fall as we move very close to an employment node if there are significant negative externalities from the firms located at the employment node. The result that there are negative externalities related to high levels of accessibility is in line with results found in [Li, Brown \(1980\)](#), although this paper measures access to employment by way of distance to the CBD. [Ahlfeldt \(2011\)](#) studies land prices and finds that a gravity-based accessibility measure can explain residential land prices. According to this paper, the measure is able to disentangle positive accessibility effects from negative congestion effects related to transportation infrastructure.

### 3 Study area and hypotheses to be tested

Our study area is situated in the south-west of Norway. The population is approximately 230 000, most of it concentrated in the north-western corner in the twin cities of Stavanger and Sandnes. Because of natural barriers, the study area is clearly delimited from neighbouring markets. This is also reflected in data on commuting flows, and contributes to making the market appropriate for an empirical analysis of the relationship between observed commuting flows, labour market accessibility, and spatial variation in housing prices.

This paper presents results from a regression model where the price of homogenous housing at a given location is related to a range of variables. These variables are either structural variables related to the house itself, or to its specific location in the geography as follows:

$$P_{it} = f(z_{sit}, z_{lit}) \quad (1)$$

where  $P_{it}$  is the price of house  $i$  in year  $t$ ,  $z_{sit}$  is the value of structural dwelling-specific attributes, and  $z_{lit}$  represents location-specific attributes.

In this paper, focus is in particular on location-specific labour market attributes. The ambition is to introduce alternative measures reflecting the prospects of finding favourable combinations of residential location and job location. The first type of measure to be considered focuses on the spatial dimension: short commuting trips are preferred to longer distance commuting, *ceteris paribus*, since commuting involves both time costs and pecuniary, distance dependent, costs. Jobs are not distinguished in terms of e.g. positions, career opportunities or wages. Both jobs and workers are considered to be homogeneous in such respects. The commuting literature offers solid support for the use of a labour market accessibility measure to explain commuting flows between different

zones of a geography. One frequently used measure is the Hansen measure (Hansen 1959). Incorporating such an accessibility measure into a doubly constrained gravity modelling framework gives a so called competing destinations model (Fotheringham 1983, Gitlesen, Thorsen 2000). Osland, Thorsen (2008) introduced this measure as an attribute of a specific location in a housing market study. In Osland, Thorsen (2008) the specific formulation of the labour market accessibility measure was given by:

$$S_j = \sum_{k=1}^{98} E_k \exp(\sigma d_{jk}) \quad (2)$$

In this expression  $E_k$  represents employment in postal zone  $k$ ,  $d_{jk}$  represents minutes driving time between zones  $j$  and  $k$ ,  $\sigma$  is a parameter estimated by maximum likelihood estimation. In this way the distance deterrence parameter is estimated simultaneously with the other parameters in the models to be presented in Section 5. There are 98 postal delivery zones in the region.

For a theoretical interpretation of this measure in a commuting context, see Gitlesen, Thorsen (2000), where the rationale of a labour market accessibility measure is argued to follow from a two-stage household decision-making process. The first step involves the selection of a set of relevant location alternatives. In a search theoretical framework, distance appears as an information filter, increasing the probability of choosing combinations with short distance between job and residence. The accessibility measure is capturing relevant information on the spatial distribution of jobs. A location of high labour market accessibility is attractive, for instance because it increases the likelihood that household members can coordinate their journeys-to-work. In this context, the labour market accessibility measure is interpreted as a job opportunity density measure, and it also makes good sense to introduce such a measure in hedonic housing market studies. It is according to standard urban economic theory that houses for sale in highly accessible labour market locations can be expected to attract high bids, reflecting a high willingness to pay for residential locations involving low expected commuting costs. Another possibility would be to apply a network modelling approach to measuring accessibility. Xiao et al. (2016) demonstrate that this approach adds explanatory power in an urban setting. However, in our slightly more macroscopic framework we proceed with a more transparent and easily available information to measure accessibility.

As an alternative, or supplement, to the measure  $S_j$ , we suggest the following intuitive indicator of spatial labour market interaction in an explanation of housing prices. Let  $\mathbf{X}$  be a commuting flow matrix where a typical element  $x_{ij}$  denotes the number of people living in zone  $i$  and working in zone  $j$ . The variable *OUT-COM* is then defined as the proportion of people living in zone  $i$  and working in another zone in the study area as follows:

$$OUT-COM_i = \frac{\sum_{j=1, j \neq i}^N x_{ij}}{\sum_{j=1}^N x_{ij}} 100 \quad (3)$$

The variable *IN-COM* is defined as the proportion of people working in zone  $j$  and living in one of the other zones:

$$IN-COM_j = \frac{\sum_{i=1, i \neq j}^N x_{ij}}{\sum_{i=1}^N x_{ij}} 100 \quad (4)$$

These measures represent a computationally simpler way to account for spatial labour market interaction than the non-linear accessibility measure. In addition, they do not require data on distances, or travelling times, between all the zones.

The two measures provide information on the spatial structure in the region. Labour market accessibility, represented by  $S_j$ , can be interpreted as a potential measure, representing the job opportunity density of a residential location. The measure reflects the degree to which a worker is able to take advantage of spatial variations in wage offers and the supply of career-enhancing jobs, within a reasonable commuting time. A reasonable hypothesis is that *OUT-COM* and *IN-COM* measure to what degree the workers actually

takes advantage of a favourable labour market accessibility, being based on observed rather than potential labour market behaviour. Individual heterogeneities in qualifications and preferences may result in a high level of spatial labour market interaction, with a correspondingly high level of excess commuting in densely populated urban areas. This further can be expected to correspond to high observed levels of out- and in-commuting in the centrally located zones of the region, representing a rationale for incorporating the measures *OUT-COM* and *IN-COM* in a model focusing on the relationship between the housing market and labour market interaction.

Despite a relatively wide scaling, Figure 1 indicates that there is a significant and positive correlation between *OUT-COM* and *IN-COM*. For the 98 postal delivery zones, the correlation coefficient between the two measures is 0.772. Both *OUT-COM* and *IN-COM* are further positively related to  $S_j$ , represented by correlation coefficients of 0.717 and 0.837, respectively. To the degree that the observation-based measure *IN-COM* is presupposed to represent labour market accessibility, it should a priori be expected to have a positive impact on housing prices. In particular for centrally located zones, the situation is similar for *OUT-COM*. The high level of out-commuting might result from a matching process, where heterogenous workers take advantage of attractive opportunities and job offers in a reasonable commuting distance outside the residential zone. In general, however, a high value of *OUT-COM* might also indicate that few jobs are available within the zone, contributing to a low level of labour market accessibility and low housing prices. Hence, it is not obvious what sign should be expected for *OUT-COM* in a hedonic regression model of housing prices.

In addition to this labour market accessibility perspective, it is important to account for the fact that there is a key difference between  $S_j$  and the other two measures.  $S_j$  is essentially capturing the existing spatial distribution of jobs, whereas *OUT-COM* and *IN-COM* in addition reflect the residential location choices of people. This means that the two observation-based measures are reflecting the (general) attractiveness of a place, including other perspective than the potential for labour market interaction. A high local value of *OUT-COM* might for instance reflect local amenities and/or attractive neighbourhood characteristics, making the zone appealing as a residential location. Hence, a positive estimate of the parameter attached to *OUT-COM* can be interpreted to capture positive neighbourhood externalities, in addition to the somewhat ambiguous effect of variations in labour market accessibility.

On the other hand, a substantial level of commuting into an area could produce congestion and other negative externalities that might have a significant effect on housing prices, see for instance Hughes, Sirmans (1992). The fact that job concentration and traffic in itself can be connected with negative externalities is also a major point in, for instance, Li, Brown (1980), Wilhelmsson (2000) and Osland, Pryce (2012). Hence, the parameter attached to *IN-COM* can be influenced by negative externalities, in addition to the positive effect stemming from labour market accessibility. This means that expected estimated sign of this parameter is also ambiguous, as it is a result of two counteracting effects. The possibility that *OUT-COM* and *IN-COM* capture different kinds of externalities is an argument in favour of including both measures in the model formulation, in addition to the labour market accessibility measure,  $S_j$ .

Following standard urban economic theory commuting time should reflect the actual commuting costs for households. It is also to be expected that commuting time rather than distance is a proper measure of commuting costs (Ma, Banister 2006). The third measure is, hence, based on actual mean commuting time in each zone:

$$MCT_i = \frac{\sum_{j=1}^J x_{ij} d_{ij}}{\sum_{j=1}^J x_{ij}} \quad (5)$$

This measure is calculated by first computing the total commuting time (*TCT*) from zone  $i$  as:

$$TCT_i = \sum_{j=1}^J x_{ij} d_{ij}$$

where  $i$  is the residential zone,  $j$  is the destination zone or job-zone,  $x_{ij}$  is the number of people residing in zone  $i$  and working in zone  $j$ , and  $d_{ij}$  is the travelling time between zone  $i$  and  $j$ . If  $i = j$ , individuals live and work in the same zone. In these cases, internal travelling time have been calculated as half of the travelling time to the nearest zone. In this way, we adjust for the fact that the areal size of the zones varies. The  $TCT$  for zone  $i$  is, finally, divided by the total size of the workforce living in zone  $i$ , given by  $\sum_{j=1}^J x_{ij}$ .

Assuming perfectly competitive housing market and following standard urban economic theory, the impact of commuting time on housing prices should be negative. However, for various reasons, workers do not minimize commuting distances (Hamilton 1982). Residential decision making is not merely about minimizing transportation costs and different structures of the urban or regional spatial structure could give different results regarding the extent of excess commuting (Ma, Banister 2006). A priori, it is therefore not obvious what sign should be expected for the impact of variations in  $MCT$  on housing prices. A high average commuting time might result for peripheral locations, where long distance commuting is the only relevant option for many workers. Such cases pull in the direction of a negative impact of  $MCT$  on housing prices. On the other hand, a high  $MCT$  can be observed in very centrally located zones, with a high level of spatial interaction, and significant labour market opportunities in many industries. As mentioned above, the high level of spatial interaction might reflect a situation with highly heterogeneous jobs and workers, where workers take advantage of attractive job opportunities in neighbouring zones. If such cases are dominating, then  $MCT$  should be expected to have a positive effect on housing prices.

In this paper we will test the hypotheses that:

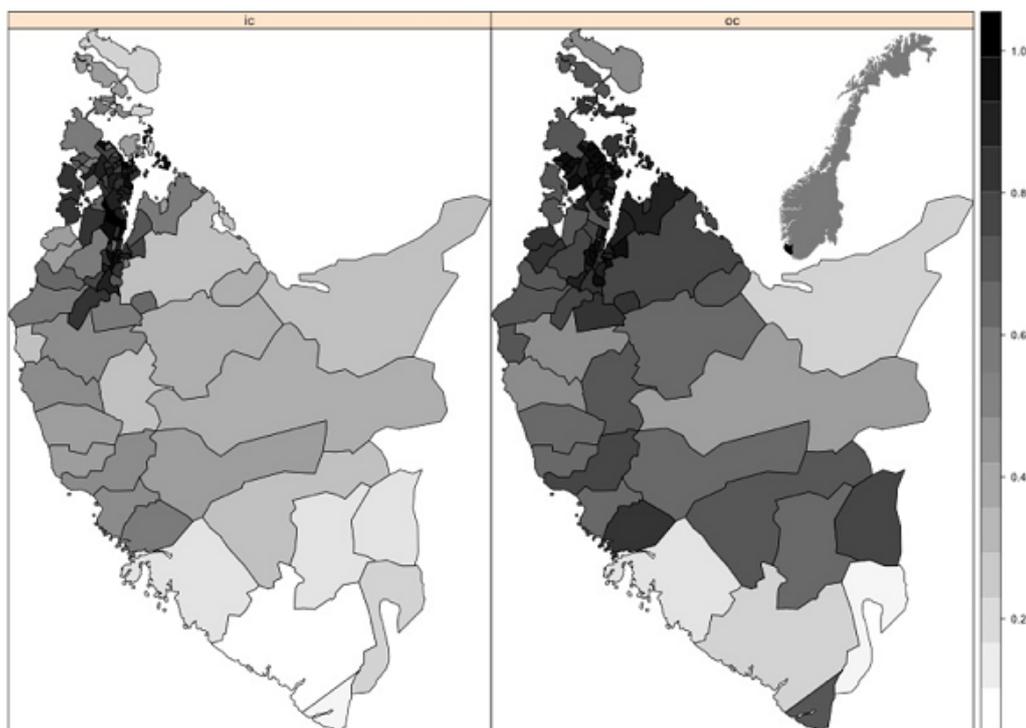
- $H_0^A$ : *OUT-COM* and *IN-COM* can replace the accessibility measure  $S_j$  in explaining spatial variation in housing prices
- $H_0^B$ : *OUT-COM* and *IN-COM* supplement the accessibility measure, and contribute with additional relevant information in explaining spatial variation in housing prices.
- $H_0^C$ :  $MCT$  can replace the accessibility measure, and contribute with additional relevant information in explaining spatial variation in housing prices.
- $H_0^D$ :  $MCT$  supplement the accessibility measure, and contribute with additional relevant information in explaining spatial variation in housing prices.

An illustration of the observed variation of the pattern of in- and out-commuting in the study area is given in Figure 1. The general tendency is a high degree of in-commuting in the central municipality of Stavanger and surrounding zones in the north. The further south and away from the CBD we move, the lower the degree of in-commuting. The percentage of out-commuting is also high in the most central cities, and is at a smaller but still high level in the postal zones surrounding these cities. One has to move to the most eastern and southern postal zones to find low levels of out-commuting.

It should be noted that the zones vary in size, and the largest postal code zones are located in rural, sparsely populated, areas. In such zones, it is to be expected that the percentage of commuting will be lower because significant distances must, on average, be travelled before the boundaries of these zones are crossed.

#### 4 Overview of data

The results to be presented are based on housing price data from the second half of 2003 to 2007. The sample consists of 4392 observations from 13 municipalities, and 98 postal delivery zones. Only privately-owned single-family houses are included. One important reason for this restriction is that this is about the only house-type available on the market in rural areas. The housing data comes from two sources: Finn.no (Finn), a web-based service used by the main real-estate franchises in Norway, and GAB, the National Building Register. The data from Finn are used to compute the national housing price index. It



Note: The map of all the postal delivery zones in the study area. Darker areas signify high levels of commuting. In the top right-hand corner, the study area is indicated on a map of Norway.

Figure 1: The 2006-values of *IN-COM* (left) and *OUT-COM* (right)

includes the actual selling price, the year and month of sale, a measure for the size of the house, the type of house, the year the house was built, and an identification code for each property.

Statistics Norway started collecting this data in 2002 and the completeness of the information available for each observation improves over time. Complete identification codes are, for instance, missing from all observations in 2002 and the first half of 2003. By 2007 the identification codes are nearly complete. According to Statistics Norway, the Finn data covers about 40% of the house sales in Norway. From 2004 this percentage is probably higher, because data from one of the largest real-estate franchises (Notar AS) was added.

The GAB register is a combination of three registers: the official land property register, an address register, and a building register. A lot of information is missing on

Table 1: Descriptive statistics

Variable	n	Mean	Median	Std. dev.	Min	Max	Source
Price (in 1000 NOK)	4392	2630.7	2450	1188.1	280	16100	Finn and GAB
Age	4392	41.1	34	32.2	0	307	Finn (GAB)
LivingArea	4392	166.8	160	56.9	33	714	Finn
LotSize	4392	656.4	573	679	8.9	24700	GAB
Garage dummy	4392	0.355	-	-	-	-	GAB
DistCBD (minutes by car)	4392	19.05	12.74	18.40	0	104.02	-
IN-COM	4392	0.68	-	0.22	0.10	0.99	-
OUT-COM	4392	0.79	-	0.20	0.12	0.97	-
MCT (Mean Travelling Time)	4392	10.00	-	3.30	5.97	29.47	-

buildings constructed before 1983. For buildings and additions constructed after 1983 information is quite extensive and accurate. The GAB register plays a central part in the formal registration of a property transaction in Norway. This formal registration is not compulsory. Statistics Norway reports that about 94% of the house sales that are formally registered are registered within six months. GAB and Finn contain different information about housing characteristics. In order to obtain as much information as possible on housing attributes, we have combined the two data sources. Matching is based on property identification code and selling price. The implication is that the prices are available from both sources. In GAB, only the last selling price of a house appears in the register. In cases where a house has been sold more than once in the study period, we have to rely on price information from only the Finn database. Houses on leased lots (about 2-3% of the houses) have been excluded.

The hedonic or micro variables we were able to obtain are presented in Table 1, which also shows quite a big dispersion in lot size. Observations with lot size equal to zero have been excluded. We have also excluded observations with a useful floor space below 30 square meters. Observations with missing exogenous variables are not included in the regressions.

Statistics Norway states the following about the Finn sample: “The statistics (...) cover a majority of all used dwellings sold in Norway. Nonetheless it is possible that to some degree there is systematic sample skewness with regard to geography.” To be more specific, the sample seems to be relatively smaller in most rural areas. This is in addition to the fact that the population of sold houses is smaller in the rural zones. In spite of this, the Finn data is used to compute the official national price index for used dwellings. Accordingly, we do not believe that this issue will have any impact on estimated results. To our knowledge this data is the best information available in Norway.

In addition to the micro data described above, we use some variables that are grouped according to postal zone. These zones vary greatly in areal size, and the urbanized zones in northwest are smaller than the most rural parts of the areas located in the south and east. In addition, there are topological differences. The terrain is far more mountainous in the east and south with a more limited road network. In these zones some of the habitation is concentrated in small hamlets in the valleys, but a substantial part is more dispersed stemming from small farms and holdings no longer used for agricultural production.

For each zone we have defined travel time to the CBD, in addition to travelling distances between all zones in the area and the number of jobs in each zone. The matrices of travelling times were calculated by the Norwegian Mapping Authority. The estimations were based on the specification of the road network into separate links, with known distances and speed limits existing in 2006. Information on speed limits and road categories is converted into travelling times through instructions from the Institute of Transport Economics. The centre of each (postal delivery) zone is found through detailed information on residential densities and the road network. Finally, the matrix of travelling times is constructed from a shortest-route algorithm.

The study area is markedly different from metropolitan areas in other countries. The region we are studying is one of the most affluent in Norway. The crime rate is relatively low, and the variation in the quality of public schools is small. The last point is due to an extensive egalitarian regional policy in Norway. However, some amenities such as provision of a range of services, closeness to open land and nature, etc. is expected to change when moving towards the CBD. Thus, the variable, distance to CBD, is important and is interpreted as urban attraction.

## 5 Alternative empirical model specifications

There are many examples in the literature where the hedonic methodology is used in empirical studies of housing markets. A review of some contributions can be found in [Anselin, Lozano-Gracia \(2009\)](#). There is no agreement, however, on what is the correct specification of a hedonic house price model, and the question of functional form remains an empirical problem that must be determined for each market under scrutiny. Hence we start by finding a parsimonious base model (M0). We then use the RESET test and

semi-parametric regression as aids to determine the correct functional form. The resulting base model is then tested for spatial effects. In the next section, the results from tests of the hypotheses from Section 3 and corresponding models is presented. We also illustrate how the predicted house price will vary with the relevant variables using the technique of a standard house. Finally, we explore the possibility of verification of our results by the spillover impacts from a spatial Durbin model.

Ignore first the possibility that labour market accessibility and commuting flows contribute to explaining housing prices. Based on previous empirical research from the study area (Osland et al. 2007, Osland, Thorsen 2008), we start with the following formulation of an empirical hedonic price model:

$$\ln(P) = a_0 + a_1 \ln(\text{DistCBD}) + a_2(\ln(\text{DistCBD}))^2 + \mathbf{bA} + c\text{YearDummy}_t + \epsilon_t \quad (6)$$

where  $P$  is the observed real selling price of house  $i$  (1998 is the base year),  $\mathbf{A}$  is a vector of the dwelling attributes listed in Table 1,  $\text{DistCBD}$  is the travelling time to the CBD, measured in minutes of car driving, and  $t$  represents the year of sale. All variables appear in logarithmic form except for the dummy variables. In the following discussion, equation (6) represents our base model, M0.

In order to estimate the housing price gradient, it is necessary to identify the centre of the geography. Following Plaut, Plaut (1998), much of the empirical literature in the field assumes that the location of the centre is known in advance. In our study the zone representing the CBD is found endogenously. We have experimented with different centrally located zones and used the descriptive measures of  $R^2$  and SRMSE/APE (defined in Table 2) to find the zone appearing as the CBD of this geography. The result corresponds to a priori knowledge of the city of Stavanger. The inclusion of a quadratic term of the CBD account for the fact that the CBD-house price gradient are more elastic with increased distance to CBD (Osland et al. 2007). The variable  $\text{DistCBD}$  is interpreted as accounting for the effect of urban attraction, and reflects that households value urban amenities found in the city centre of the region. The inclusion of a gravity-based accessibility measure (equation (2)) can be interpreted as representing a more general labour market accessibility effect on housing prices (Osland, Thorsen 2008).

As mentioned above, the modelling procedure was motivated by previous estimation results from the same study area. The new data used in this paper are, however, from a more recent time period, with less information on housing attributes than was the case in Osland et al. (2007) and Osland, Thorsen (2008). To avoid model mis-specifications we therefore initially apply a semi-parametric approach, the RESET test (Ramsey 1969), and tests for spatial effects (Anselin 1988). The chosen modelling procedures are advocated and applied in Osland (2010).

The RESET test is a mis-specification test related to the functional form of the variables included in the model. In this case the test is based on powers of the fitted values and the fourth power is the highest. We test the null hypothesis that the model has no omitted variables. The alternative hypothesis is that the model is mis-specified.

The estimation was mainly performed in Stata, but we also use the program R combined with related packages (see Bivand et al. 2008).

Semi-parametric regression analysis is a flexible approach that is used as an exploratory tool to detect non-linearity in the data. There exist some hedonic studies that use similar approaches (see for instance Coulson 1992, Pace 1998, Bao, Wan 2004). In this paper a variant of the generalized additive models based on Hastie, Tibshirani (1990) is applied in combination with iterative penalized regression-smoothing splines. The method is explored in detail in Wood (2006). We estimate the model represented by equation (6) and include each continuous variable in turn into the smooth function  $s(z)$ , so that  $z$  is a variable vector not included in  $\mathbf{A}$ . The estimations have been made by using the `mgcv` (multiple generalized cross-validations) package (version 1.7-28) in R.

Consider for example the inclusion of lot size as a variable in the hedonic regression model, represented by the variable  $\ln(\text{LotSize})$  in the smooth function. The graphical result is illustrated in the right-hand side of Figure 2. This graph is based on a thin plate regression spline. The values on the caption of the  $y$ -axis denote efficient degrees of

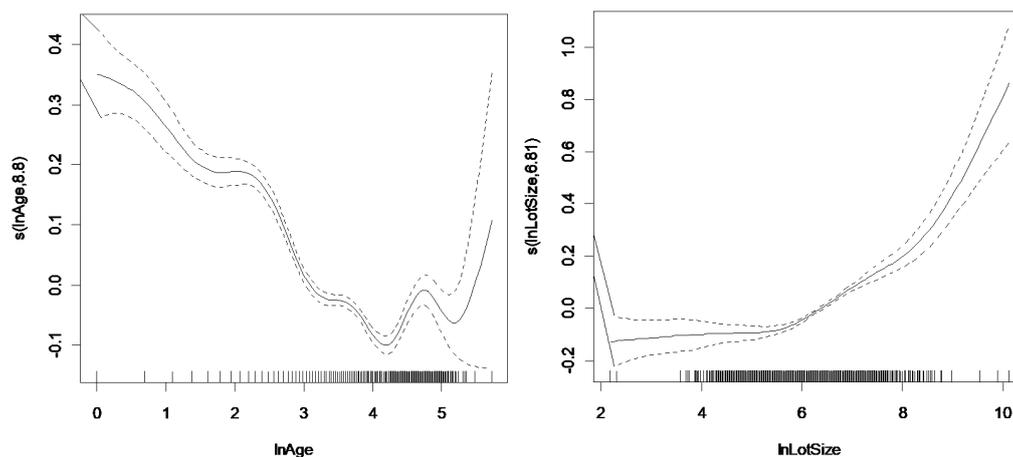


Figure 2: Age (left) and LotSize (right) are in turn included in the semi-parametric smooth function

freedom of the plotted term. The interpretation is that the equivalent of 6.81 degrees of freedom is used in estimating the smooth function (see Wood 2006, p. 170-172). The solid line in the figure represents the variation around the mean predicted value of the dependent variable. The dashed lines represent the approximately 95% confidence regions of the predicted values. The figure illustrates that the square of the variable  $\ln(LotSize)$  should be included in the model. Finally, we have also included the square of  $\ln(Age)$  in the hedonic regression model. A priori it is to be expected that these variables should be included as a non-linear relationship with housing prices.

The most important question in this paper is how to represent labour market accessibility and the characteristics of commuting flows in the model. The evaluation is based on traditional specification tests, such as Wald tests and log-likelihood ratio tests, in addition to the other measures described above.

Finally, a range of descriptive measures is included; see Table 1. Starting with a relatively parsimonious model formulation, more comprehensive model specifications are based on the results of the documented tests and descriptive measures.

Table 2 offers results of the following model specifications:

- M1:** M0 extended by the gravity-based accessibility indicator defined by equation (2).
- M2:** M0 extended by observed in and out-commuting, defined by equations (3) and (4).
- M3:** M2 extended by the gravity-based accessibility indicator defined by equation (2).
- M4:** M0 extended by mean commuting time defined by equation (5).
- M5:** M4 extended by the gravity-based accessibility indicator (2).

All the model specifications have been tested for spatial effects. We use the `spdep` (spatial dependence) package, from the R statistical programming environment. The robust Lagrange-multiplier (RLM) tests (see Florax, Nijkamp 2003) are reported in Table 2. The RLM tests asymptotically follows a chi-squared (1) distribution. The RLM-error statistics test the null hypothesis of no significant spatial error correlation, correcting for the presence of local spatial lag dependence in the dependent variable. Similarly, the RLM-lag statistics test the null hypothesis of no spatial autocorrelation in the dependent variable, correcting for the presence of local spatial error dependence. The used row-standardized spatial weight matrices, is based on a  $k$ -nearest neighbour structure. The  $k$ -nearest neighbour is chosen on the basis of distances in meters. Based on the log-likelihood values we use  $k = 3$  for the spatial error model, so that each observation have the minimum of three neighbours.

For all model specifications, the null hypothesis of no spatial autocorrelation had to be rejected. When the RLM-error test statistic is the largest, a spatial error model with the above-mentioned weight structure would normally, remedy the problem (Anselin 1988). An important consequence of this is that the ordinary least squares estimator is consistent (Anselin 1988). Based on M0, the spatial error model is formulated as follows:

$$\begin{aligned}\ln(P) &= a_0 + a_1 \ln(\text{DistCBD} + a_2(\ln(\text{DistCBD}))^2) + \mathbf{bA} + c \text{YearDummy}_t + \tilde{\epsilon}_t \quad (7) \\ \tilde{\epsilon} &= \lambda \mathbf{W}\tilde{\epsilon} + u\end{aligned}$$

where  $\mathbf{W}$  is the weight matrix, and  $\lambda$  is the spatial autoregressive parameter (Bivand et al. 2008, p. 284). The estimation of the spatial error model variants of M0–M5 does not change any of the results. The results from the spatial error models are presented in Appendix A.

## 6 Results on potential and observed measures of commuting pattern

Notice first from Table 2 that labour market accessibility (*ACCESS*) has a significantly positive impact on housing prices in all the models where it is taken into account (M1, M3 and M5). This is consistent with previous empirical analysis of the housing market in the region (Osland, Thorsen 2008). According to Table 2 the results are also consistent with the findings in Osland, Thorsen (2008) that spatial variation in housing prices is explained by a labour market accessibility effect and an urban attraction effect (*DistCBD*).

As mentioned in Section 3, it would be convenient, both from a computational and from a data collection point of view, if the variables *IN-COM* and *OUT-COM* could replace the accessibility measure  $S_j$ . By comparing the results from the model specifications M1 and M2 in Table 2, however, the hypothesis  $H_0^A$  has to be rejected. The negative estimate of the coefficient attached to *IN-COM* in M2 means that this variable cannot be interpreted as representing labour market accessibility. According to model specification M2, the variable *OUT-COM* has no significant impact on housing prices. In addition, the value of the log-likelihood function is clearly higher in M1 than in M2. Most of the other reported descriptive statistics favour M1.

As labour market accessibility cannot be replaced by the observed characteristics of commuting flows, M2 has an important spatially defined variable omitted from the model specification. Hence, the parameter estimates related to *IN-COM* and *OUT-COM* will be biased. M3 accounts for both labour market accessibility and the relevant characteristics of observed commuting flows.

The results based on M3 support the hypothesis  $H_0^B$  that the variables *IN-COM* and *OUT-COM* contribute additional relevant information in explaining spatial variation in housing prices. According to Table 2, both variables appear to be significant in the hedonic regression model M3. The value of the likelihood ratio test statistic is approximately 40 when M3 is compared with M1. This value clearly exceeds the critical value of the chi-square distribution ( $\chi_{0.05}^2(2) = 5.991$ ). The  $p$  value of the Wald test is 0.000, given a null hypothesis of no joint significance of these two variables.

How should the results related to *IN-COM* and *OUT-COM* be interpreted? Consider, for instance, a zone located a short distance from the CBD, with a high value of the gravity-based labour market accessibility measure. It follows from the results in Table 3, that housing prices are predicted to be high in this zone; the labour market accessibility effect and the urban attraction effect operate in the same direction. As mentioned in Section 3, *IN-COM* at least to some degree captures the effect of variations in labour market accessibility. However, the result from M2 means that this is not the dominating effect of variations in *IN-COM*. The estimate of the coefficient attached to *IN-COM* is significantly negative. This means that negative externalities related to job concentrations, stemming for example from traffic, is dominating the effect explained by spatial variation in labour market accessibility. This conclusion is supported by the results following from M3, where labour market accessibility is explicitly accounted for by the variable  $S_j$ .

According to results for M3, a high value of *OUT-COM* is predicted to have a positive impact on house prices, adding, for instance, to the effects of the variables representing

labour market accessibility effect and urban attraction. The parameter estimates reported in Table 2 is small, however, and the positive impact of variations in *OUT-COM* is relatively marginal. *OUT-COM* is not found to have a significant impact on housing prices when  $S_j$  is not accounted for, in the model formulation M2. Recall from Section 3 that a possible negative impact of *OUT-COM* on housing prices can be argued to reflect a situation where few jobs are available within the zone, corresponding to a low level of labour market accessibility. Altogether, the results reported in Table 2 mean that this effect is dominated by the combined effect of positive neighbourhood externalities and a generally high spatial labour market interaction in centrally located areas.

Another hypothesis is that observed characteristics of in- and out-commuting may in particular be relevant for rural areas. To test this hypothesis, the variables representing commuting flows were interacted with a dummy variable, taking the value 1 if the zone is a rural zone, otherwise taking the value 0. This model extension did not alter significantly the parameter estimates related to the variables *IN-COM* and *OUT-COM*.

The labour market accessibility measure  $S_j$  offers information on the spatial distribution of jobs. It does not take into account the residential location pattern which would reflect the number of competing workers. Similarly, it does not account for the possibility that spatial labour market interaction is influenced by heterogeneities both in the working force and in the supply of jobs. Jobs for different categories of workers may for instance be clustered in specific zones of the geography, and this may influence commuting flows, residential location choices, and the willingness to pay for houses in different locations. Hence, information on the spatial distribution of different categories of jobs and workers may prove relevant in studying both commuting flows and house prices, but such data are not in general available at a sufficiently disaggregate subdivision into zones. In such a scenario, observation-based measures can, to some degree, capture the effect of labour market heterogeneities and characteristics of the residential location pattern. These issues are definitely not captured by the potential measure of labour market accessibility,  $S_j$ .

The Figures 3 and 4 illustrate how the predicted price of a so-called standard house varies along with variation in *IN-COM*, *OUT-COM*, and the gravity-based accessibility variable. A standard house is defined as a house that was sold in 2007, has a garage, and has not been sold in a rural area. Except for this, all the continuous variables are set to their average values for the sample (Osland, Thorsen 2008). The dependent variable has been transformed from its logarithmic form to prices in accordance with the following transformation rule:

$$P = \exp\left(\widehat{\ln(P)}\right) \exp\left(\frac{\widehat{\sigma^2}}{2}\right) \quad (8)$$

Here,  $\sigma^2$  denotes an unbiased estimator of the residual variance (see Wooldridge 2003, p. 208).

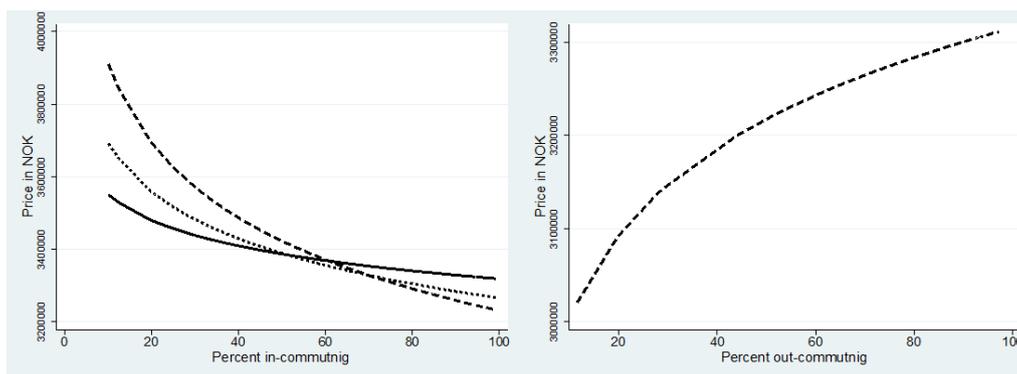
Notice from the left part of Figure 3 that incorporating the gravity-based accessibility measure, in a more adequate model formulation, contributes to increase the partial effect on housing prices of variations in the variable *IN-COM*. The model formulation M2 gives a biased, undervalued, estimate of the negative externalities associated with *IN-COM*. Figure 4 shows that the results related to the gravity-based accessibility measure are not sensitive to whether we include the commuting variables or not. As the value of the accessibility measure increases, so does the price of a standard house, albeit at a decreasing rate.

We have also experimented by introducing other variables related to observed commuting flows. The results on the impact of average commuting time are reported in Table 2. Figure 5 indicates that the relationship between the mean commuting time (*MCT*) and house prices is nonlinear. Experiments proved that the nonlinearities are satisfactorily represented by a quadratic term in a simple polynomial regression. A Wald test of the joint significance of the inclusion of the variable *MCT* and  $MCT^2$  clearly has to be rejected in M5. The relevant  $p$ -value is 0.000.

Figure 6 provides an illustration of how different model formulations estimate the impact of variations in *MCT* on housing prices. At a first glance, the estimates resulting from the models M4 and M5, might seem relatively similar in Table 2. According to

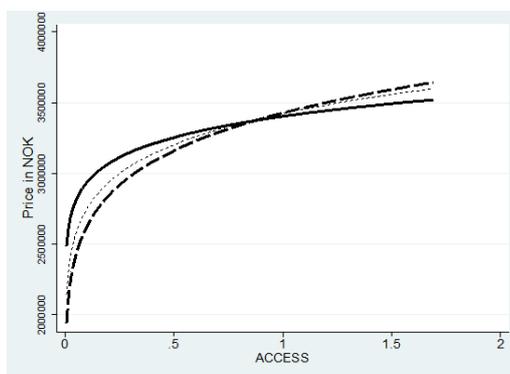
Table 2: Estimated results from alternative hedonic house price models

Variable name	M0	M1	M2	M3	M4	M5
Constant	12.970 (71.89)	12.417 (63.52)	12.942 (71.75)	12.085 (59.50)	12.543 (49.53)	11.430 (38.87)
LotSize	-0.173 (-3.00)	-0.204 (-3.62)	-0.167 (-2.91)	-0.203 (-3.55)	-0.177 (-3.1)	-0.221 (-3.92)
LotSize <sup>2</sup>	0.022 (4.62)	0.025 (5.27)	0.022 (4.55)	0.025 (5.22)	0.023 (4.72)	0.026 (5.59)
RurLotSize	-0.025 (-10.28)	-0.021 (-8.83)	-0.027 (-10.15)	-0.023 (-9.64)	-0.025 (-10.11)	-0.019 (-8.36)
Age	-0.2017 (-11.75)	-0.192 (-11.38)	-0.205 (-11.84)	-0.199 (-11.72)	-0.204 (-11.75)	-0.196 (-11.47)
Age <sup>2</sup>	0.017 (5.75)	0.016 (5.23)	0.018 (5.92)	0.017 (5.65)	0.018 (5.85)	0.016 (5.38)
Garage	0.042 (7.21)	0.043 (7.44)	0.040 (6.93)	0.042 (7.16)	0.042 (7.19)	0.040 (7.06)
LivingArea	0.510 (42.77)	0.505 (42.69)	0.510 (42.80)	0.503 (42.66)	0.511 (42.99)	0.504 (42.88)
YearDum04	0.097 (8.89)	0.098 (9.03)	0.097 (8.90)	0.097 (9.05)	0.097 (8.89)	0.097 (8.98)
YearDum05	0.206 (18.75)	0.207 (18.92)	0.205 (18.68)	0.205 (18.90)	0.205 (18.68)	0.205 (18.84)
YearDum06	0.370 (33.55)	0.371 (33.67)	0.370 (33.49)	0.370 (33.82)	0.370 (33.51)	0.370 (33.82)
YearDum07	0.558 (53.95)	0.558 (54.17)	0.558 (53.85)	0.557 (54.44)	0.558 (53.93)	0.555 (54.39)
DistCBD	-0.056 (-3.14)	-0.104 (-5.36)	-0.043 (-2.28)	-0.114 (-5.57)	-0.076 (-3.90)	-0.171 (-7.40)
DistCBD <sup>2</sup>	-0.046 (-13.18)	-0.027 (-5.84)	-0.051 (-12.22)	-0.024 (-4.53)	-0.043 (-10.86)	-0.0150 (-2.85)
ACCESS		0.064 (6.48)		0.096 (7.79)		0.114 (7.06)
IN-COM			-0.030 (-2.06)	-0.083 (-4.90)		
OUT-COM			-0.043 (0.03)	0.001 (2.31)		
MCT					0.393 (2.45)	0.399 (2.29)
MCT <sup>2</sup>					-0.081 (-2.41)	-0.054 (-1.45)
n	4392	4392	4392	4392	4392	4392
R <sup>2</sup>	0.822	0.824	0.823	0.826	0.823	0.826
R <sup>2</sup> (adj)	0.822	0.824	0.822	0.825	0.822	0.825
Log-likelihood	1184.02	1207.05	1187.48	1226.90	1187.60	1228.14
VIF	16.36	17.56	16.06	17.60	45.87	46.20
Ramsey reset	0.767	0.760	0.464	0.316	0.8003	0.5364
APE	510405	508469	510366	507978	510486	508223
SRMSE	0.289	0.288	0.290	0.288	0.289	0.288
RLM-lag	1.70	0.86	1.38	0.17	1.70	0.39
RLM-error	323.59	313.09	324.15	307.78	320.36	313.01



Note: The dashed lines refer to M3, while the solid line refers to a corresponding model specification without a gravity-based accessibility measure (M2). Predicted housing prices for variation in *OUT-COM* based on M2 is not shown in the Figure to the right, given that this variable does not contribute to explain the variation in housing prices in these two models.

Figure 3: Predicted house prices of a standard house



Note: The solid line refers to M1 where the commuting variables are excluded. The dotted line refers to M3, and the dashed line refers to M5. The values of the accessibility variables is mean-normalized.

Figure 4: Predicted house prices of a standard house

Figure 6, however, there is a substantial difference between the model M5, and the more parsimonious model M4. Technically, this is due to the different parameter values estimated for the quadratic term, but it can also be argued that the results reflect changes in the characteristics of the urban structure.

As pointed out in Section 3, a high *MCT* may reflect either a peripheral location, or a centrally located area with heterogeneous agents and considerable excess commuting. For the model M5, the effects of labour market heterogeneity and a high level of spatial interaction in the central parts of the urban area seem to dominate, since *MCT* is estimated to have a positive impact on housing prices. Over time, the job growth in the Stavanger urban area has come in areas that used to be the outskirts of the city, mainly in suburban industrial parks. As a consequence, workers living in residential areas close to the city centre, see Figure 1, no longer have on average shorter commuting times than workers living in some of the suburban areas. Still, some of the residential areas close to the city centre have very high housing prices. These areas have traditionally been fashionable residential locations, with neighbourhoods that are popular, beyond the pure urban attraction effect. This is one possible explanation why housing prices are predicted to be an increasing function of *MCT*.

The first, rising part of the curve resulting from model M4 in Figure 6 can be explained from the same line of reasoning as above. In this model, however, the labour market accessibility measure is not incorporated. Since *MCT* is closely and negatively correlated to labour market accessibility (Pearson's correlation coefficient is  $-0.8677$ ), the labour

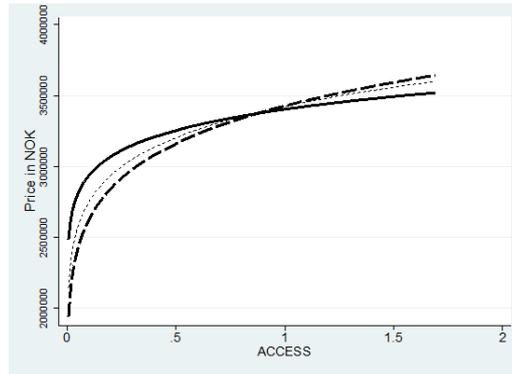
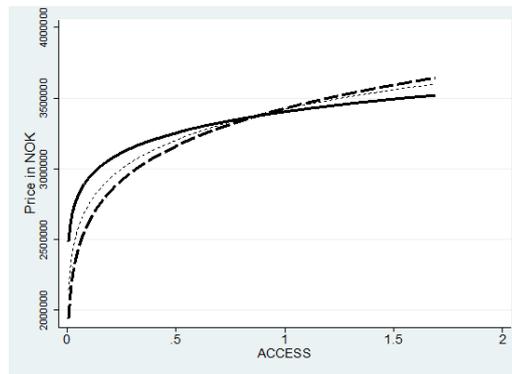


Figure 5: The variable  $\ln MCT$  is included in the semi-parametric smooth function



Note: The solid line refers to M4, which includes the variable mean commuting time. The dotted line refers to M5, which also includes the gravity based accessibility measure in addition to mean commuting time.

Figure 6: Predicted house prices of a standard house

market accessibility effect dominates for variations in  $MCT$  higher than 10 minutes. The effect of  $MCT$  is, hence, negatively biased in this model formulation.

## 7 Issues of endogeneity and robustness checks

Empirical studies using the hedonic house prices model necessitates considerations of endogeneity, which may bias the estimated implicit prices. First, there should not be an omitted variable bias. This problem is discussed in Section 7.1. Another potential source of endogeneity bias is reverse causality, which is discussed in Section 7.2.

### 7.1 Omitted variable bias

An important example of left out spatially related missing characteristic is negative and positive externalities such as noise, local air quality or the physical and social surroundings. We do not control for these types of variables in our model specifications. The reason for not including the variables is lack of data. If these variables have a significant impact on housing prices, and if they correlate with the studied variables, they may, create a bias of the studied coefficients. However, the bias may go in many directions, given the potential of a large number of missing spatial characteristics.

We have performed two robustness checks in order to study if an important omitted variable bias is present. Given that we do not have information on specific environmental variables, we use the average value of houses sold in an area as control for a range of left-out characteristics. We define a neighbourhood either at the postal code level or at the municipality level. There are 98 postal codes and 11 municipalities in the study area. The inclusion of the control variable does not change any of the conclusions regarding

sign of coefficients for the studied variables in any of the models. However, the impact on the coefficients vary according to which definition of neighbourhood we use. Using the most disaggregate neighbourhood level reduces the absolute value of the relevant coefficients, whereas the inclusion of the most aggregate control variable increases the absolute value of the coefficients. The difference in estimated coefficients is largest when using a control variable at the disaggregated level.

It is difficult to get information about the direction of a potential bias based on this robustness check. Given the changes in the coefficients, the bias is relatively small, still most of the results show that the estimated coefficients of the accessibility variables documented in Table 2, are outside a 95 % confidence region of the models which include a control variable.

The second approach is to use an alternative estimator, the spatial Durbin model. This estimator is robust to omitted variables reflecting spatial characteristics (LeSage, Pace 2009). Description of estimation procedures and general interpretations of the spatial Durbin model will follow the presentations found in these publications. According to LeSage, Pace (2009), this model-estimator is robust to omitted variables reflecting spatial characteristics. The results presented in Table 2 are based on relatively parsimonious model specifications, and most empirical hedonic house price models are encumbered by omitted variables. Hence, the spatial Durbin model could reveal additional information in this respect. The spatial Durbin variant of the model is specified as follows:

$$P = \rho \mathbf{W} \beta_0 + \rho \mathbf{W} \mathbf{X} \beta_1 + \epsilon \quad (9)$$

In equation (9),  $P$  is a vector of observed prices,  $\mathbf{X}$  is a matrix of observations on independent variables, and  $\mathbf{W}$  is the  $n \times n$  matrix of exogenous spatial weights. This model allows a spatial lagging of the dependent variable, in addition to a spatial lagging of the independent variables (see Bivand 1984, LeSage, Fischer 2008). The spatial Durbin model has been estimated by using the same weight matrix as described for the spatial error model in Section 5, except that we use  $k = 4$  in the weights, because this gives the highest log-likelihood values in this case.

According to LeSage, Fischer (2008) the estimated parameters related to the spatial Durbin model have no straightforward interpretation. For this reason, we only report the spillover impacts, estimated by the procedure in LeSage, Pace (2009, p. 38). The covariance matrix of the coefficients has been calculated by using numerical methods (LeSage, Pace 2009, p. 56-59). This matrix and traces of powers series of the weights matrix, estimated by Monte Carlo approximations, were used to derive impact measures, and tests of significance (LeSage, Pace 2009, p. 96-104 and 114-115).

In the spatial Durbin model, represented by equation (9), the price of a house  $i$  is a function of the neighbouring house prices through the lagged dependent variable. Neighbouring house prices are a function of the values of the houses' own attributes. Changing these attributes has an effect on its own price, and hence also on the price of house  $i$ . In addition, the price of house  $i$  depends on the attribute values of its neighbours, as expressed through the spatially lagged independent variables. The dimension of the spillover effects depends upon the size of the estimated spatial autocorrelation parameters and the specification of the neighbourhood matrix (see LeSage, Fischer 2008, LeSage, Pace 2009, Kirby, LeSage 2009). Even when the lagged independent variables are statistically not significant, there may still exist some significant spillover effects occurring through a spatially autocorrelated dependent variable.

The estimated average impacts from the spatial Durbin model are presented in Table 3. The results are based on M3 and M5. Excluding the polynomial variants of all the commuting variables that we are studying makes the interpretation of the results easier. In our case the direct impacts are calculated as the average effect on a house price  $i$  of a change in each of the explanatory variables related to that house. The average total impact is the estimated effect on the price, followed by a change in each of the variables, respectively, over all observations. The indirect impact is represented by the difference between the total and direct impacts. In this way the indirect impact captures the average effects on the price of house  $i$  from the change in the variables of other houses.

Table 3: Estimated direct, indirect, and total impact from variants of M3 and M5 using the spatial Durbin Estimator

Variable Name	M3			M5		
	Direct	Indirect	Total	Direct	Indirect	Total
Lotsize	-0.201 (-6.47)	0.150 (2.21)	-0.051 (-0.70)	-0.173 (-4.92)	0.012 (0.15)	-0.161 (-1.84)
LotSize <sup>2</sup>	0.025 (9.69)	-0.013 (-2.20)	0.013 (2.01)	0.024 (8.73)	-0.005 (-0.67)	0.020 (2.64)
RurLotsize	-0.028 (-8.89)	0.007 (1.76)	-0.021 (-8.10)	-0.041 (-3.92)	0.024 (2.21)	-0.017 (-5.24)
Age	-0.196 (-11.93)	0.038 (1.05)	-0.158 (-4.13)	-0.161 (-9.39)	-0.097 (-2.51)	-0.285 (-6.22)
Age <sup>2</sup>	0.016 (5.89)	-0.004 (-0.71)	0.012 (1.84)	0.009 (2.95)	0.020 (3.09)	0.028 (4.13)
Garage	0.040 (6.48)	0.026 (2.06)	0.066 (4.70)	0.039 (6.78)	0.009 (0.63)	0.048 (2.86)
LivingArea	0.503 (52.91)	-0.037 (-1.84)	0.466 (21.11)	0.487 (53.64)	0.094 (3.72)	0.581 (20.94)
YearDum04	0.099 (9.58)	-0.029 (-1.24)	0.070 (2.86)	0.097 (9.05)	-0.001 (-0.04)	0.096 (2.22)
YearDum05	0.209 (19.93)	-0.027 (-1.16)	0.182 (7.48)	0.207 (19.40)	0.038 (1.03)	0.245 (5.72)
YearDum06	0.373 (35.41)	-0.014 (-0.57)	0.359 (14.40)	0.371 (35.31)	-0.019 (-0.53)	0.353 (8.50)
YearDum07	0.560 (56.92)	-0.022 (-0.92)	0.538 (21.51)	0.557 (54.24)	0.001 (0.04)	0.558 (14.02)
DistCBD	-0.081 (-3.45)	-0.073 (-2.09)	-0.153 (-4.86)	-0.036 (0.40)	-0.103 (-0.97)	-0.139 (-4.23)
DistCBD <sup>2</sup>	-0.024 (-4.20)	0.008 (0.95)	-0.016 (-1.94)	0.014 (0.37)	-0.033 (-0.87)	-0.019 (-2.66)
ACCESS	0.084 (7.23)	0.035 (1.87)	0.120 (6.99)	0.161 (2.00)	-0.051 (-0.63)	0.109 (5.22)
IN-COM	-0.073 (-4.67)	-0.064 (-2.62)	-0.136 (-6.18)			
OUT-COM	0.040 (2.36)	0.033 (1.87)	0.073 (2.72)			
MCT				0.247 (1.80)	-0.122 (-0.88)	0.125 (3.43)

Note: Z-values in parentheses.

The spatial Durbin model yields the same sign on all the estimated parameters, as for the ordinary least squares estimator. The direct impacts yield results that are within the 95% confidence region of the ordinary least squares regression results for the model variant based on M3. For the specification based on M5, this is not always the case. For most of these variables the total impact is within the 95% confidence region. The impacts related to age variables is not within the mentioned 95% confidence region. For M1, not reported, the total impact of *ACCESS* is not within the 95% confidence region. The indirect impact is negative, but not significant.

Indirect impacts are significant at the 5% level for the variables *LivingArea* and *Age<sup>2</sup>* for the model based on M5. The same is the case for the spatial Durbin variant of M1. In the model based on M3, only *IN-COM* has a significant indirect impact at the 5% significance level. The indirect impact is negative. This means that an increase in *IN-COM* of the houses neighbouring a house *i*, will on average have a negative impact on the price of house *i*. Hence, there are negative spillover effects of having a house located in an area with a high relative level of in-commuting. This result is in line with our interpretations of regression results found in previous sections. The average indirect

impacts of *OUT-COM* and *ACCESS* are positive, but not significant at the 5% level. The results of this robustness check also show that all the direct impacts related to the studied variables have the same sign as the ones documented in Table 2. Given the fact that the correlation between the variables are relatively high, it is to be expected that the inclusion and exclusion of variables will have an impact on the value of the estimated coefficients. The direction of any existing bias is not possible to determine based on any of the robustness checks.

## 7.2 Reverse Causality

The second relevant type of endogeneity problem could be reverse causality. By way of example, neighbourhoods with higher housing prices could attract high-income earners, perhaps also with higher educational attainments. Locations close to these places could be relevant for specific firms in order to get access to an attractive pool of labour. This type of reasoning is based on the hypothesis that “jobs follow people”, which subsequently may improve labour market accessibility in places with higher housing prices. Hoogstra et al. (2017) provide a useful meta-analysis of the related literature regarding this question. Their main finding is that the evidence is inconclusive. The causality between jobs and people could run in different directions (pages 371-372). If this holds true, for our data, the resulting bias could be minor, and unpredictable.

If there exists reverse causality, identification of the impact of accessibility would necessitate an instrument. One possible instrument could be an exogenous change in measures of accessibility. However, we do not have access to any such instrument. Moreover, we focus on several variables, which are potentially endogenous. Testing for exogeneity is, hence, not straightforward according to e.g. Baum et al. (2007). We have to use several instruments, and in these cases, the traditional IV-estimators could be biased and inconsistent (see also Nordvik et al. 2019, for further discussions of this issue).

Finally, it is also possible to argue that the need for an instrument is less important in our case, given the results from the spatial Durbin model. This model accounts for indirect spill over impacts and the spatial Durbin model, in general, is robust to omitted systematic spatial variation of characteristics (LeSage, Pace 2009).

## 8 Conclusions

In explaining spatial variation in housing prices, gravity-based accessibility measures have been suggested as a generalization of modern polycentric labour market structures. From a computational and data collection perspective, it would be convenient if easily available information on actual commuting patterns could replace a more complex measure of commuting potential. According to our results, two of the hypotheses formulated in Section 3 have to be rejected, however. The labour market accessibility effect is not adequately represented by the proposed characteristics of observed commuting patterns. We used a wide range of different methods to obtain robust conclusions. Labour market accessibility in relation to housing prices is best captured by the gravity based potential variable.

Our results, on the other hand, provide support for the hypotheses  $H_0^B$  and  $H_0^D$ . Observed measures of commuting patterns are found to contribute with information that adds to the effect of a potential measure in explaining spatial variation in housing prices. In particular, the results from the ordinary least squares and the spatial Durbin estimator support the hypothesis that a relatively high level of commuting into a zone corresponds to negative externalities, such as noise, pollution, or other negative effects of heavy traffic and/or industrial activities. There is only weak support for the hypothesis that a relatively high level of out-commuting from a zone corresponds to attractive neighbourhood characteristics, which are positively related to housing prices. Average commuting time, a priori, reflect something about the actual commuting costs for households. However, our interpretation of the positive impact of this variable is that a high level of spatial interaction in the central parts of the urban area seem to dominate. Over time, the job growth in the Stavanger urban area has come in areas that used to be the outskirts of the city. Still, some of the residential areas close to the city centre have very high

housing prices. These areas have traditionally been fashionable residential locations, with neighbourhoods that are popular, beyond the pure urban attraction effect.

The results presented in this paper contribute to modifying predictions of how changes in labour market accessibility affect housing prices. Assume, as an example, that a number of jobs are relocated from a zone. This means that the labour market accessibility is reduced for this zone, and house prices decrease. On the other hand, the reduced number of jobs might lead to a reduced commuting flow into the zone, contributing to increased house prices. To some extent, this offsets the effect of a reduction in labour market accessibility. We will not be more specific on the overall effect in this paper. This depends, for instance, on where the jobs are relocated, and how the labour market accessibility is affected.

The introduction of alternatives to the potential measure of accessibility was not only motivated by considerations of simplicity and data requirements. We have been arguing that commuting flows, the residential location pattern, and house prices may result from a complex mixture of labour market heterogeneities, characteristics of the residential location pattern, and different kinds of externalities. Ideally, such heterogeneities, as well as the conditions causing negative and positive externalities should be explicitly controlled for in the model formulation. In most cases, however, relevant information is not available without a massive data collection effort. It is a useful result that observation-based measures of commuting flows to some degree capture the effect of heterogeneities and externalities. This information on commuting flows is often easily available. Referring to the main ambition and motivation of our analysis, we find that this information is adding to the explanatory power of the hedonic model of housing prices. Hence, we conclude that observation-based measures of commuting flows can supplement, but not substitute, the gravity-based potential measure of accessibility in explaining spatial variation of housing prices.

## References

- Adair A, McGreal S, Smyth A, Cooper J, Ryley T (2000) House prices and accessibility: The testing of relationships within the Belfast urban area. *Housing studies* 15: 699–716. [CrossRef](#).
- Ahlfeldt G (2011) If Alonso was right: modelling accessibility and explaining the residential land gradient. *Journal of Regional Science* 51: 318–338. [CrossRef](#).
- Anas A, Arnott R, Small K (1998) Urban spatial structure. *Journal of Economic Literature* 36: 1426–1464
- Anselin L (1988) *Spatial econometrics: Methods and models*. Kluwer, London. [CrossRef](#).
- Anselin L (2002) Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27: 247–267. [CrossRef](#).
- Anselin L, Lozano-Gracia N (2009) Spatial hedonic models. In: Mills T, Patterson K (eds), *Palgrave Handbook of Econometrics*, Volume 2. Palgrave Macmillan. [CrossRef](#).
- Ball M, Kirwan R (1977) Accessibility and supply constraints in the urban housing market. *Urban Studies* 14: 11–32. [CrossRef](#).
- Bao H, Wan A (2004) On the use of spline smoothing in estimating hedonic housing price models: Empirical evidence using Hong Kong data. *Real Estate Economics* 32: 487–507. [CrossRef](#).
- Baum CF, Schaffer ME, Stillman S (2007) Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal* 7: 465–506. [CrossRef](#).
- Bivand R (1984) Regression modelling with spatial dependence: An application of some class selection and estimation methods. *Geographical Analysis* 16: 25–37. [CrossRef](#).

- Bivand R, Pebesma E, Gómez-Rubio V (2008) *Applied spatial data analysis with R*. Springer. [CrossRef](#).
- Coulson N (1992) Semiparametric estimates of the marginal price of floorspace. *Journal of Real Estate Finance and Economics* 5: 73–83. [CrossRef](#).
- Dubin R (1992) Spatial autocorrelation and neighbourhood quality. *Regional Science and Urban Economics* 22: 433–452. [CrossRef](#).
- Farber S, Neutens T, Miller HJ, Li X (2013) The social interaction potential of metropolitan regions: A time-geographic measurement approach using joint accessibility. *Annals of the Association of American Geographers* 103: 483–504. [CrossRef](#).
- Florax R, Nijkamp P (2003) Misspecification in linear spatial regression models. *Technical report, Tinbergen Institute Discussion Paper, (TI 2003-081/3)*
- Fotheringham AS (1983) A new set of spatial interaction models: The theory of competing destinations. *Environment and Planning A* 15: 1121–1132. [CrossRef](#).
- Gitlesen JP, Thorsen I (2000) A competing destinations approach to modeling commuting flows: A theoretical interpretation and an empirical application of the model. *Environment and Planning A* 32: 2057–2074. [CrossRef](#).
- Gjestland A, McArthur D, Osland L, Thorsen I (2014) The suitability of hedonic models for cost-benefit analysis: Evidence from commuting flows. *Transportation Research Part A: Policy and Practice* 61: 136–151. [CrossRef](#).
- Hamilton B (1982) Wasteful commuting. *Journal of Political Economy* 90: 1497–1504. [CrossRef](#).
- Handy S, Niemeier D (1997) Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning A* 29: 1175–1194. [CrossRef](#).
- Hansen W (1959) How accessibility shapes land use. *Journal of the American Institute of Planners* 25: 73–76. [CrossRef](#).
- Hastie T, Tibshirani R (1990) *Generalized additive models*. Chapman and Hall, London. [CrossRef](#).
- Heikkilä E, Gordon P, Kim J, Peiser R, Richardson H (1989) What happened to the CBD-distance gradient? Land values in a polycentric city. *Environment and Planning A* 21: 221–232. [CrossRef](#).
- Hoogstra GJ, van Dijk J, Florax RJ (2017) Do jobs follow people or people follow jobs? A meta-analysis of Carlino-Mills studies. *Spatial Economic Analysis* 12: 357–378. [CrossRef](#).
- Hughes J, Sirmans C (1992) Traffic externalities and single-family house prices. *Journal of Regional Science* 32: 487–500. [CrossRef](#).
- Jackson J (1979) Intraurban variation in the price of housing. *Journal of Urban Economics* 6: 464–479. [CrossRef](#).
- Kirby D, LeSage J (2009) Changes in commuting to work times over the 1990 to 2000 period. *Regional Science and Urban Economics* 39: 460–471. [CrossRef](#).
- Kwan MP, Murray AT, O’Kelly ME, Tiefelsdorf M (2003) Recent advances in accessibility research: Representation, methodology, and applications. *Journal of Geographical Systems* 5: 129–138. [CrossRef](#).
- LeSage J, Fischer M (2008) Spatial growth regressions: Model specification, estimation and interpretation. *Spatial Economic Analysis* 3: 275–304. [CrossRef](#).
- LeSage J, Pace R (2009) Introduction to spatial econometrics. Chapman and hall/crc, boca raton. [CrossRef](#).

- Li M, Brown H (1980) Micro-neighbourhood externalities and hedonic prices. *Land Economics* 56: 125–140. [CrossRef](#).
- Ma K, Banister D (2006) Excess commuting: A critical review. *Transport Reviews* 26: 749–767. [CrossRef](#).
- Nordvik V, Osland L, Thorsen I, Thorsen IS (2019) Capitalization of neighbourhood diversity and segregation. *Environment and Planning A: Economy and Space* 51: 1775–1799. [CrossRef](#).
- Osland L (2010) An application of spatial econometrics in relation to hedonic house price modelling. *Journal of Real Estate Research* 32: 289–320
- Osland L, Pryce G (2012) Housing prices and multiple employment nodes: Is the relationship nonmonotonic? *Housing Studies* 27: 1182–1208. [CrossRef](#).
- Osland L, Thorsen I (2008) Effects on housing prices of urban attraction and labor market accessibility. *Environment and Planning A* 40: 2490–2509. [CrossRef](#).
- Osland L, Thorsen I, Gitlesen J (2007) Housing price gradients in a geography with one dominating center. *Journal of Real Estate Research* 29: 321–346
- Pace R (1998) Appraisal using generalized additive models. *Journal of Real Estate Research* 15: 77–99
- Plaut P, Plaut S (1998) Endogenous identification of multiple housing price centers in metropolitan areas. *Journal of Housing Economics* 7: 193–217. [CrossRef](#).
- Ramsey J (1969) Tests for specification errors in classical linear least squares regression analysis. *Journal of Royal Statistical Society B* 31: 350–371. [CrossRef](#).
- Vågane L, Brechan I, Hjorthol R (2011) Den nasjonale reisevaneundersøkelsen – 2009. The institute for transport economics, TØI report 1130
- Waddell P, Berry B, Hoch I (1993) Residential property values in a multinodal urban area: New evidence on the implicit price of location. *Journal of real estate finance and economics* 7: 117–141. [CrossRef](#).
- Wilhelmsson M (2000) The impact of traffic noise on the values of single-family houses. *Journal of Environmental Planning and Management* 43: 799–815. [CrossRef](#).
- Wood S (2006) *Generalized additive models. An introduction with R* (2nd ed.). Chapman and Hall/CRC. [CrossRef](#).
- Wooldridge J (2003) *Introductory econometrics. A modern approach*. South-Western, Mason, OH
- Xiao Y, Orford S, Webster CJ (2016) Urban configuration, accessibility, and property prices: a case study of Cardiff, Wales. *Environment and Planning B: Planning and Design* 43: 108–129. [CrossRef](#).

## A Appendix: Results of spatial error model estimations

Table A.1: Estimated results for the hedonic house price models based on the spatial error model formulation as specified in equation (9)

Variable Name	M0	M1	M2	M3	M4	M5
Constant	13.060 (120.88)	12.506 (75.10)	13.031 (119.33)	12.180 (66.87)	12.710 (48.96)	11.596 (37.09)
Lotsize	-0.187 (-6.11)	-0.204 (-6.61)	-0.183 (-5.94)	-0.202 (-6.56)	-0.189 (-6.15)	-0.214 (-6.93)
LotSize <sup>2</sup>	0.024 (9.55)	0.026 (9.98)	0.024 (9.39)	0.026 (9.99)	0.024 (9.58)	0.026 (10.30)
RurLotsize	-0.025 (-8.76)	-0.021 (-7.26)	-0.026 (-8.86)	-0.023 (-7.86)	-0.025 (-8.73)	-0.019 (6.47)
Age	-0.1809 (-10.71)	-0.178 (-10.51)	-0.183 (-10.78)	-0.182 (-10.82)	-0.182 (-10.78)	-0.180 (-10.67)
Age <sup>2</sup>	0.013 (4.64)	0.012 (4.43)	0.013 (4.75)	0.013 (4.77)	0.013 (4.75)	0.013 (4.60)
Garage	0.039 (6.71)	0.040 (6.82)	0.038 (6.59)	0.039 (6.70)	0.039 (6.68)	0.039 (6.67)
Size of house	0.493 (52.31)	0.491 (52.19)	0.493 (52.31)	0.490 (52.27)	0.493 (52.47)	0.491 (52.36)
YearDum04	0.0955 (9.87)	0.096 (9.91)	0.096 (9.87)	0.096 (9.93)	0.096 (9.88)	0.096 (9.90)
YearDum05	0.199 (20.41)	0.199 (20.45)	0.199 (20.39)	0.199 (20.46)	0.199 (20.41)	0.199 (20.43)
YearDum06	0.370 (38.33)	0.370 (38.34)	0.370 (38.34)	0.370 (38.36)	0.370 (38.32)	0.370 (38.35)
YearDum07	0.5573 (58.91)	0.557 (58.88)	0.557 (58.92)	0.557 (58.93)	0.557 (58.93)	0.556 (58.82)
DistCBD	-0.0574 (-2.58)	-0.101 (-4.14)	-0.040 (-1.58)	-0.105 (-3.93)	-0.074 (-2.96)	-0.163 (-5.71)
DistCBD <sup>2</sup>	-0.047 (-11.06)	-0.029 (-5.07)	-0.053 (-9.17)	-0.028 (-3.96)	-0.045 (-9.48)	-0.018 (-2.82)
ACCESS		0.059 (4.36)		0.088 (5.81)		0.109 (6.07)
IN-COM			-0.033 (-1.78)	-0.082 (-4.39)		
OUT-COM			-0.006 (-0.22)	0.034 (1.65)		
MCT					0.315 (1.48)	0.306 (1.47)
MCT <sup>2</sup>					-0.063 (-1.44)	-0.035 (-0.80)
$\lambda$	0.374 (23.67)	0.364 (22.85)	0.373 (23.60)	0.359 (22.30)	0.374 (23.60)	0.358 (22.24)
Log-likelihood	1420.26	1429.58	1422.56	1440.23	1422.69	1440.57

Note: Z-values in parentheses.



© 2020 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Louvain-la-Neuve, Belgium. This article is distributed under the terms and conditions of the Creative Commons Attribution, Non-Commercial (CC BY NC) license (<http://creativecommons.org/licenses/by-nc/4.0/>).