The Determinants of Joint Residential and Job Location Choices: A Mixed Logit Approach

Alexander Ebertz
(Ifo Institute)†

Work in progress: Preliminary version, February 2009

Abstract:

This paper empirically investigates the individual decision of where to live and where to work. If wages and housing costs at the metropolitan area are higher relative to the surrounding non-metropolitan area, the location decision amounts to trading off wages, housing costs, and commuting time. A mixed logit model is employed to quantify the interaction effects of these factors in the joint residential and job location choice. The approach does not rely on the restrictive IIA assumption and allows for arbitrary correlation patterns between coefficients. Using data from a recent survey of more than half a million German households, the elasticities of individual location choice with respect to wages, housing costs, and commuting time are estimated. The results show that individual tastes for these factors are of the expected sign but vary substantially in the population. Changes in the spatial distribution of households induced by political measures that subsidize commuting are calculated based on the empirical estimates.

Keywords: Location Choice; Commuting; Metropolitan Area; Discrete Choice Models; Mixed Logit; Simulation based Estimation

JEL Classification: R23; R12; C15; C25

†Address: Ifo Institute for Economic Research
Poschingerstr. 5
D-81679 Munich
Germany

Phone: +49 89 9224 1394
Fax: +49 89 985369
E-mail: ebertz@ifo.de

1 Introduction

It is a common fact that most people in the workforce travel from their homes to their place of work and back every day. This implies that at some point people make choices of where to live and where to work, respectively. These two non-separate choices implicitly include a decision on the daily commute. Reality shows that the extent of commuting may thereby range from a three minutes walk down the street to a three hour trip. It may take place within the community of residence or across the borders of communities, counties, federal states, or even countries.

The typical picture one bears in mind when thinking about the issue is people living in the suburbs and traveling to the urban center, where all the work is located in the central business district (CBD). This picture of the "monocentric city" has been formally described and analyzed in the seminal works of Alonso (1964), Muth (1969), and Mills (1972). Regardless of wether work is located only in the city center or also at other points in space (the "polycentric city"), the individual location choice is determined by trade offs between wages, housing costs, and the economic cost of commuting. Besides these true economic factors, economic literature emphasizes the role of amenities and the local quality of life in the household's location choice (prominent examples are the works of Rosen (1979) and Roback (1982)).

The objective of this paper is to empirically quantify the roles that the economic factors play in the joint residential and job location choice of households. I apply discrete choice methods to a large set of microdata that contains information on individual places of residence and work, individual commuting time, and the usual personal characteristics. The data is combined with county specific figures on wages and housing costs to allow for a comprehensive empirical analysis of all three economic determinants of location choice. The analysis thereby follows the idea of So, Orazem, and Otto (2001) to focus on the relationship between the metropolitan area and its surrounding nonmetropolitan area. In their analysis, the choice setting boils down to four alternatives: To live and work in the CBD, to live and work in the nonmetropolitan area, to live in the suburbs and work in the CBD, or to live in the center and work in the nonmetropolitan area, respectively. The object of their study is the metropolitan area of Des Moines, Iowa, and the surrounding nonmetropolitan area. While adopting their choice setting, I apply it to a dataset that is richer in a few ways. First, the data provides the relevant information on all labor market regions in Germany instead of only one, of which I use the three largest cities[1] at the moment[2]. Second, the counties of residence or work are identified in the data, such that regional figures can be assigned to the alternatives that have not been chosen. Thus, it is not necessary to estimate the counterfactual values.

For purposes of estimation, a mixed logit approach is employed where coefficients are allowed to vary randomly, instead of being constant over decision makers. This strategy also sidesteps issues involved with the rather restrictive IIA assumption that is implicit in classic multinomial logit estimation. Moreover, as there are serious concerns about possible correlation between individual tastes for the four alternatives in the choice setting described above, I allow for arbitrary correlation patterns in the estimation of coefficients.

The estimation results are converted to elasticities to show how changes in wages, housing costs, or commuting time affect the distribution of people between metropolitan and nonmetropolitan areas. A particularly interesting example in this respect are the effects of an increase in the German subsidy to commuters ("Pendlerpauschale"), which has recently come into effect. Further results include the calculation of changes in consumer surplus induced by political actions that reduce commuting time or affect the local cost of housing.

The remainder of the paper proceeds as follows. The next section briefly outlines the underlying random utility model. In section 3 the empirical setting and the data are described before the econometric specification is illustrated. Section 4 presents the results from the mixed logit estimation and the associated elasticities. The last part is concerned with consumer surplus. Section 5 discusses political implications of the results. Section 6 concludes.

[1]Namely Berlin, Hamburg, and Munich

[2]Current work is to extend the analysis to more cities

2 Theory

This section briefly illustrates the well known random utility model that underlies the empirical analysis. Households choose a residential location i that is characterized by a bundle of local attributes, A_i, and the local housing cost, h_i. They also choose a job location j, where the prevailing wage w_j is earned. If $i \neq j$, the household has to commute and commuting costs of t_{ij} have to be incurred. As the focus of the theoretical and empirical work is on centralized labor market regions, locations i and j are assumed to be located either in the metropolitan area (the center) or in the surrounding nonmetropolitan area. Residential and job locations are chosen simultaneously to maximize utility. The indirect utility obtained from residing at i and working at location j is therefore given by

$$V_{i,j} = V\left(w_j, h_i, t_{ij}, A_i\right) \quad i, j = M, N, \tag{1}$$

where M and N refer to the metropolitan and the nonmetropolitan area, respectively. Standard models of households' location choice suggest that $\frac{\partial V}{\partial w} > 0$, $\frac{\partial V}{\partial h} < 0$, and $\frac{\partial V}{\partial t} < 0$. Local attributes may have a positive or a negative effect on utility, depending on their nature as an amenity or disamenity: $\frac{\partial V}{\partial A} \lesseqgtr 0$.

Utility maximization requires that the chosen alternative is at least as good as all other possible alternatives:

$$V\left(w_{j^*}, h_{i^*}, t_{i^*j^*}, A_{i^*}\right) \geq V\left(w_j, h_i, t_{ij}, A_i\right) \quad \forall i \neq i^* \; j \neq j^*. \tag{2}$$

Comparing commuters to non-commuters, this implies that commuters require a wage premium over their residential wages. This wage premium rises with commuting cost. Furthermore, suppose that housing costs are higher in the metropolitan area, $h_M > h_N$, because prices are bid up due to the higher population density. In this case, average wages earned in the center have to exceed average wages earned in the suburbs as long as differences in amenities are not too big to entirely compensate for the housing price differential.

3 Empirical Approach

Setting and Data

Given the focus on the relation between the metropolitan area and the surrounding nonmetropolitan area, individuals select one of four possible alternatives to maximize utility:

- MM: Live in the metropolitan area, work in the metropolitan area
- MN: Live in the metropolitan area, work in the nonmetropolitan area
- NN: Live in the nonmetropolitan area, work in the nonmetropolitan area
- NM: Live in the nonmetropolitan area, work in the metropolitan area

The objects of investigation are 3 labor market regions ("Raumordnungsregionen") in Germany, which all exhibit a clearly defined urban county as the metropolitan center region and various surrounding counties that constitute the associated nonmetropolitan area. In particular, the study so far focusses on the labor markets of Berlin, Hamburg, and Munich.

Individual data on commuting time, location of residence and job, age, education, children, and household income are taken from the "Perspektive Deutschland" study 2004, a large survey among more than half a million Germans. It reports opinions and valuations of German residents concerning a variety of aspects of life in Germany and the German regions, respectively. Representativeness is ensured by sampling weights drawn from a parallel field-survey with more than 10,000 participants.[3] Monthly net household income in € is reported net of taxes and including transfers. In order to ensure a certain degree of homogeneity of decision makers and the driving forces behind their decisions, I focus only on full-time employed individuals in the analysis.

Commuting cost is proxied for by commuting time. However, since individual commuting time is only reported for the chosen alternative, the respective values for the other three alternatives are

[3] See Fassbender and Kluge (2006) for an overview of the project.

missing. Estimating commuting time for each of the four alternatives solves this problem and in addition sidesteps the issue that wages, commuting time and housing cost are simultaneously chosen in the location decision. Individual commuting time (in minutes, one way) for each alternative a in each labor-market region l is therefore predicted by OLS, according to:

$$t_{n,l,a} = \delta_0 + \delta_1 Age_{n,l,a} + \delta_2 Education_{n,l,a} + \delta_3 Sex_{n,l,a} + \delta_4 Married_{n,l,a} + \delta_5 Children_{n,l,a} + e_{n,l,a}, \quad (3)$$

with n indexing households, l indexing labor regions, and $a = MM, MN, NN, NM$ indexing alternatives. Children is a dummy variable indicating if the household has children aged between 1 and 16 years. Education gives the years of schooling associated with the highest degree achieved. Age is reported in categories that each subsume five years.

Wages are taken from the regional sample of employees (*Beschäftigtenstichprobe*) of the Institute for Employment Research (IAB), which constitutes a two percent random sample of all German employees subject to social security contributions and reports individual daily wages in € .[4] I merge this detailed wage information with the survey data by county level and by education. Thus, I assign the average wage of people of the same education level who chose the same alternative within the respective labor-market region to each of the 4 alternatives one individual faces.

The third important determinant of the location decision is housing cost. The regional statistical offices provide data on the average prices for land in 2001-2004 in € per sqm at the county level, which serve as excellent indicators for the local cost of housing. However, in this context it is important to consider not only the price but also the quantity of housing space consumed. Since I do not have information on the individual demand for living space, I use the reported number of adults and children in the household along with the respective figures on average housing demand of one- (two-, three-or-more-) person households in sqm from the German Statistical Office to proxy for the desired housing space of a household.[5] Since theory predicts that the demand for housing varies with income, the predicted housing cost is divided by the household income that is reported

[4]See Drews (2008) for a detailed description of the data

[5]Accordingly, the average housing space consumed by a one-person household in Germany in 2004 is 67.5 sqm (93.2 sqm for a two-person household, and 113.4 sqm for three-or-more-person households.

Table 1: Sample Means by Residential and Job Location

Variable	*MM*	*MN*	*NN*	*NM*	*M*	*N*
Location specific						
Commuting Time	29.8	39.6	25.9	46.3	30.6	33.9
Wage	103	99.9	94.0	104	103	97.9
Housing Cost	13.5	19.9	7.36	5.97	14.0	6.81
Individual						
Age	40-44	35-39	40-44	40-44	40-44	40-44
(Age Category)	(6.57)	(6.32)	(6.66)	(6.76)	(6.55)	(6.70)
Children	.234	.219	.310	.347	.233	.324
Education	13.4	13.8	12.6	12.9	13.4	12.7
Observations	21026	1893	8793	5710	22919	14503

in the "Perspektive Deutschland" data. Thus, individual housing cost is calculated as

$$h_{n,l,a} = \frac{L_{l,a} q_n}{y_n}, \tag{4}$$

where $L_{l,a}$ denotes the average land price for alternative a in labor-market region l, q_n denotes the demand for housing of household n, and y_n is the reported net household income net of taxes and including transfers. Note, that the individual income variable includes capital income and is different from the wage variable, $w_{l,a}$, which is given at the county level. This way of constructing the housing cost variable ensures that the empirical analysis measures the valuation of housing cost of people having roughly the same demand for housing and the same level of income.

Table 1 reports summary statistics of the sample by residential and job location. The facts are as expected. Wages are higher for people working in metropolitan areas and housing costs in the center widely exceed those in nonmetropolitan regions, although the latter are corrected for individual housing demand and income. Interestingly, the theoretical predictions seem to hold only partly as average wages are higher for those who commute from the suburbs to the center compared to wages of non-commuters who reside in nonmetropolitan areas. However, commuters living in the metropolitan area do not exhibit such mark-ups over the wages of their non-commuting neighbors. Commuting time is, of course, much higher in the alternatives that involve commuting. On average, people in the sample who chose different alternatives do not differ in age, except for those who live in the center and commute to the periphery. They tend to be slightly younger on average, being in the category of people aged $35-39$ years, while those who choose the other alternatives are all in the group aged $40-44$. Furthermore, people who live in the center exhibit a higher amount of years of schooling and are less likely to have children.

Econometric Specification

To estimate the influence of wages, housing costs, and commuting time on the household's simultaneous choice of residential and work location, I employ the well known discrete-choice framework that has its foundations in the work of McFadden (1973). More precisely, I adopt the so called "mixed logit" approach where coefficients are allowed to vary over decision makers.[6] In this case, a household chooses a combination of living and working place among the four possible alternatives a, $a = MM, MN, NN, NM$, which lie in one of the labor market regions. The utility that household n derives from choosing alternative a is

$$U_{n,a} = \alpha X_{n,a} + \beta_n Z_{n,a} + \varepsilon_{n,a}, \tag{5}$$

where $\varepsilon_{n,a}$ is an unobserved random term that is identically and independently drawn from an extreme value type I distribution.

Note, that the vector of coefficients β is subscripted with n while α is not. The influence of the determinants of choice contained in $X_{n,a}$ is assumed to be constant across households. $X_{n,a}$ contains characteristics that vary with the alternatives and the households, like the above described housing cost variable. In this particular case, the specification amounts to assuming that individual tastes

[6] This approach is also known as "random coefficients logit" or "error commponents logit" and has been applied in many studies, e.g. Bhat (1998), Brownstone and Train (1999), or Train (1998).

for housing cost are identical for households with the same income and the same demand for living space. Further variables in $X_{n,a}$ are designed to capture that households' tastes for living in the center or for commuting might systematically vary with individual characteristics like age, children, and education.

In contrast, the coefficients of the variables in $Z_{n,a}$ are assumed to vary randomly over households. For example, the individual tastes for commuting costs in the population, expressed by the coefficient β_n^t, are assumed to follow a normal distribution with its parameters θ to be estimated. This distribution is also called the mixing distribution. Adopting this specification takes into account that there might be substantial unobserved variation in personal tastes for commuting, beyond the systematic variation with age, education, and children. Imagine, for example, people who travel to work by public transport: Some might gladly use the time to read a book or newspaper, while others explicitly dislike the crowded busses or trains. Similarly, some of the commuters who travel by car might be more fond of driving as such than others. With respect to wages, one can easily think of arguments in favor and against the hypothesis that tastes are identically distributed in the population. I therefore estimate different specifications, where the wage variable enters $X_{n,a}$ or $Z_{n,a}$, respectively.

Obviously, the adopted setting raises strong concerns with respect to the well known IIA assumption, that holds in classic multinomial logit models. As there are always two respective alternatives that are somehow similar (e.g. two commuting alternatives: MN, NM or two alternatives that imply living in the center: MM, MN.), it is not reasonable to expect tastes for them to be independent from each other a priori. Therefore, $Z_{n,a}$ contains indicator variables that identify the average taste for each alternative, assuming that their coefficients are normally distributed in the population. When estimating the mean and variance-covariance matrix of this distribution, I explicitly allow for arbitrary correlation patterns between the variables. Adopting this specification explicitly addresses the conjecture that utility is correlated over alternatives in this setting as it does not rely on the IIA assumption.[7] Note, that the alternative specific constants also capture the average tastes for local characteristics of the alternatives, that have been labeled amenities in the theoretical model of the previous section.

[7] See Train (2003) for a discussion.

4 Results

Estimation

The results of various specifications of the mixed logit model are reported in table 2. The coefficients on the alternative specific constants are estimated together with their standard deviations in all specifications. It is assumed that they follow a normal distribution with the reported means and standard deviations.

In general, the model performs quite well as all coefficients show the signs that one would expect from theory. Higher wages attract people, while higher commuting time and higher housing costs make an alternative less likely to be chosen. The estimated standard deviations of the random coefficients are all highly significant, indicating that tastes for commuting time and wages indeed vary considerably in the population. Note, that this variation may be due to unobservable characteristics as well as observable ones, which are not included in the model. In the specification reported in column 1 of table 2 only the coefficient on commuting time is allowed to vary over decision makers, while wages and housing cost are modeled to be valued identically in the population. The estimation reported in column 2 treats wages as additional random coefficients, all else being equal. While the coefficient and standard deviation on commuting time and housing cost remain remarkably constant, the coefficient on wages is clearly higher in the second specification. Its standard deviation is quite high compared to the coefficient, which points to a high variation in tastes for wages in the population. In fact, this result suggests that some decision makers even dislike higher wages. Similarly, the tastes for commuting time show a relatively high variation, even including positive valuations. This result is not surprising, however, as the individual cost per minute of commuting time is very likely to differ greatly with observable characteristics like age or income, as well as with unobservable tastes for or against circumstances involved with commuting (e.g. driving a car or using means of public transport).

The specification reported in column 3 further includes individual characteristics as fixed coefficients. While the coefficients on wages and housing costs turn out to be smaller in this specification, commuting time has no significant, although a negative, coefficient. However, its standard devia-

Table 2: Results of the Mixed Logit Estimation

	1			2			3		
Variable	*Coefficient*		*SE*	*Coefficient*		*SE*	*Coefficient*		*SE*
MM	.438	***	(.140)	3.04	***	(.289)	5.85	***	(.507)
MM (SD)	.273		(.309)	5.20	***	(.447)	5.46	***	(.445)
NN	-3.53	***	(.548)	1.29	***	(.292)	5.01	***	(.515)
NN (SD)	7.01	***	(.696)	7.04	***	(.566)	6.32	***	(.396)
NM	-3.63	***	(.544)	-.586		(.388)	1.97	***	(.465)
NM (SD)	9.54	***	(1.06)	7.05	***	(.757)	5.24	***	(.505)
Commuting Time	-.133	***	(.015)	-.119	***	(.013)	-.004		(.007)
Commuting Time (SD)	.219	***	(.019)	.292	***	(.027)	.245	***	(.015)
Wage	.099	***	(.012)	.175	***	(.027)	.028	***	(.004)
Wage (SD)				.296	***	(.040)			
Housing Cost	-.031	***	(.004)	-.038	***	(.003)	-.016	***	(.002)
Commute X Age							-.032	**	(.013)
Commute X Children							-.017		(.069)
Commute X Education							.074	***	(.009)
M X Age							-.097	***	(.008)
M X Children							-.663	***	(.041)
M X Education							.098	***	(.006)
N	149688			149688			149688		
SLL	-40351.84			-40351.69			-39912.96		

Dependent variable: Choice of alternative a, with $a = MM, MN, NN, NM$. *** **denotes significance at 1% level (** at 5%). 149688 observations correspond to 37422 individuals. 50 Halton draws used for simulation.**

tion is still highly significant and of comparable magnitude to the other results. This might indicate that after controlling for some individual characteristics the tastes balance out in the population, such that the mean of the distribution cannot be estimated precisely. The estimates on the personal characteristics give some interesting insights. Older people are less likely to live in the metropolitan area and to be commuters, while people with higher education clearly prefer the city centers and commute more. Having children does not affect the probability to commute significantly, but it raises the probability to live in nonmetropolitan areas.

Elasticities

This section is current work in progress.

Consumer Surplus

This section is current work in progress.

5 Implications

This section is current work in progress.

6 Conclusion

This paper empirically quantifies the effects of wages, housing costs, and commuting time on the joint residential and job location choice of households. Applying discrete choice methods to a large set of microdata allows a comprehensive empirical analysis of the three most important economic determinants of location choice. The analysis thereby focusses on the relationship between the urban centers and their surrounding nonmetropolitan areas of the largest German cities of Berlin, Hamburg, and Munich.

A mixed logit approach is employed where coefficients are allowed to vary randomly instead of being constant over decision makers. This estimation strategy avoids the rather restrictive IIA assumption that is implicit in classic multinomial logit estimation. Moreover, arbitrary correlation patterns of coefficients are explicitly allowed for as correlation between tastes for the alternatives is very likely in the adopted choice setting. The estimates suggest that tastes for commuting time and wages indeed vary substantially in the population.

The estimation results are converted to elasticities to show how changes in wages, housing costs, or commuting time affect the distribution of people between metropolitan and nonmetropolitan areas. Furthermore, changes in consumer surplus induced by political actions that reduce commuting time or affect the local cost of housing are calculated.

References

Alonso, W. (1964), *Location and Land Use: Toward a General Theory of Land Rent*, Cambridge, MA: Harvard University Press

Bhat, C. (1998), Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling, *Transportation Research Part A: Policy and Practice*, 32(7), 495 – 507

Brownstone, D. and K. Train (1999), Forecasting New Product Penetration with Flexible Substitution Patterns, *Journal of Econometrics*, 89, 109 – 129

Drews, N. (2008), Das Regionalfile der IAB Beschäftigtenstichprobe 1975-2004, Handbuch-Version 1.0.0, Nuremberg

Fassbender, H. and J. Kluge (2006), *Perspektive Deutschland: Was die Deutschen wirklich wollen*, Berlin

McFadden, D. (1973), Conditional Logit Analysis of Qualitative Choice Behavior, in: P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press

Mills, E. S. (1972), *Studies in the Structure of the Urban Economy*, Baltimore, MD: Johns Hopkins University Press

Muth, R. (1969), *Cities and Housing*, Chicago: University of Chicago Press

Roback, J. (1982), Wages, Rents, and the Quality of Life, *The Journal of Political Economy*, 90(6), 1257 – 1278

Rosen, S. (1979), Wages-based Indexes of Urban Quality of Life, in: P. Mieszkowski and M. Straszheim, eds., *Current Issues in Urban Economics*, Baltimore, MD

So, K. S., P. F. Orazem, and D. M. Otto (2001), The Effects of Housing Prices, Wages and Commuting Time on Joint Residential and Job Location Choices, *Amer. J. Agr. Econ.*, 83(4), 1036 – 1048

Train, K. (1998), Recreation Demand Models with Taste Variation, *Land Economics*, 74, 230 – 239

Train, K. (2003), *Discrete Choice Models with Simulation*, Cambridge, MA: Cambridge University Press