Are there gender-specific preferences for location factors?
A Grouped Conditional Logit-Model of interregional migration flows in Germany

Schneider L., Kubis A. 1

January, 2007

Abstract.

Many European regions show large imbalances in terms of sex ratios of young adults. Gender-specific valuations of regional characteristics might be one explanation for this phenomenon. Our empirical study analyzes the question whether women and men differ in their tastes for location factors. The question is answered by quantifying the impact of location characteristics on interregional migration behaviour. Our main interest lies on four dimensions which – as prior research has shown – determine migration decisions: labour market, education, family and amenities.

The analysis is based on the micro-econometric concept of a conditional logit model which can be derived from a standard random utility framework. The model estimates the probability of choosing a certain destination, conditional only on choice characteristics, i.e. regional endowment. Furthermore, we augment the framework by controlling for violation of the independence of irrelevant alternatives assumption (IIA) and for overdispersion. Migration data stem from a comprehensive migration statistics for German NUTS-3 regions covering the age- and gender-specific migration flows for 2005.

As basic result of our gender-specific regressions we find no general differences in terms of direction, i.e. location factors that attract young men also attract their female counterparts. However, a closer look reveals gender differences in terms of intensity. Women seem to appreciate regional characteristics typically located in agglomerations. Particularly the educational institutions and cultural offerings play an essential role in explaining women’s migration behaviour. Men’s migration activities are primarily determined by job related issues. But even after controlling for regional endowment men seem to show lower probabilities to migrate than women.

Keywords: Regional Migration, Gender Economics, Discrete Choice Model

JEL-classification: R23, J61, C25

---

1 Halle Institute for Economic Research, P.O. Box 11 03 61, 06017 Halle (Saale), Germany, Phone: +49 (0)345 7753 851, Fax: +49 (0)345 7753 820, Email: lsr@iwh-halle.de.
Are there gender-specific preferences for location factors?
A Grouped Conditional Logit-Model of interregional migration flows in Germany

Schneider L., Kubis A.

1 Introduction

At the beginning of statistical migration analysis Ravenstein noted that women are more mobile than men (Ravenstein 1885). Since then, migration research has widely neglected the gender dimension of spatial mobility – especially with respect to interregional migration. While the empirical literature dealing with international migration recently turned to gender issues (Parrenas 2001; Ehrenreich and Hochschild 2002; Dumont et al. 2007), only a few studies can be found that explicitly focus on gender-specific internal migration patterns and determinants.2 This research gap is surprising since differing mobility rates usually translate into imbalances between regional in- and out-migration and, thus, into unequal regional net migration rates in terms of gender. In the long run regions mostly left by (or attracting) one sex will show substantially unbalanced sex ratios. Regarding younger age groups such trends can be observed in many European countries in particular within transition regions. High correlations between regional net migration rates, sex ratios, agglomeration levels and growth rates show up. A lot of rural low growth areas face strong female out-migration followed by remarkable disequilibria in terms of sex ratios.3 Due to the regional dualism between Eastern and Western regions this trend is particularly prevalent in Germany. Against this background our study tries to answer two questions: i) Are women more mobile than men? ii) Which regional location factors are women and men attracted by, i.e. what are the gender specific valuations of regional characteristics?

The analysis is based on a comprehensive migration data set covering the age- and gender-specific migration flows between German NUTS-3 regions. The econometric approach draws heavily on Davies et al. (2001) who estimate the impact of regional wage levels and unemployment rates on interstate migration in the US. Like Davies et al. we apply the micro-econometric framework of a Conditional Logit model derived from a Random utility model on aggregate migration flow data. We augment the framework by controlling for violation of the independence of irrelevant alternatives assumption (IIA) and for overdispersion using a method developed by Guimaraes and Lindrooth (2007).

The remainder of the paper is organized as follows: the next section outlines the recent empirical literature dealing with internal migration. Section three describes the econometric model, section four introduces the explanatory variables. The dataset is character-

---

2 For a brief overview see Chapter 2.

3 The analysis of internal migration within 10 European countries done by Rees and Kupiszewski (1999) confirms this trend for the post-transition countries of Poland and Romania. Other case studies detect such developments in Finland and Sweden (Kupiszewski et al. 2000, 2001).
ized in section five. Descriptive evidence and estimation results are presented in section six. A conclusion completes the paper.

2 Empirical Literature

The empirical literature analyzing the determinants of interregional migration can be divided into micro level and aggregate approaches. The micro concepts focus on the migration decision or intention of individuals or households whereas the interregional migration flows, i.e. the outcome of individuals’ behaviour, are explained in aggregate level studies (Cushing and Poot 2004). Due to the improved availability of survey data during the last decades the literature more and more turned to the micro concepts as on this level the actual migration motives of individuals or households can be captured most closely (Cushing and Poot 2004). However, these survey based studies generally rely on a limited sample size, so, the impact of location factors at a small scale regional level cannot be examined. Thus, with respect to internal migration in Germany these micro studies usually focus only on the migration from East to West Germany (Burda 1993; Burda et al. 1998; Hunt 2000; Bruecker and Truebswetter 2007). Solely Hunt (2004) analyzes the determinants of regional migration on a smaller regional scale but still on NUTS-1 level.

Due to our focus on the small scale regional dimension of internal migration, the paper is closer related to the empirical literature dealing with migration on the aggregate than on the micro level. The empirical literature analyzing the regional determinants of aggregate migration flows was traditionally focussed on the role of regional labour market conditions (Greenwood 1997). From a human capital perspective (Sjaastad 1962; Harris and Todaro 1970) the regional wage level as well as the unemployment rate were supposed to affect the regional migration balance. The third variable usually implemented in this type of analysis is distance. It is referred to as proxy for migration costs, thus, it should discourage migration. Many country studies confirm the positive effect of income levels and the negative impact of distance. In terms of regional unemployment rates the results are rather mixed.4 For Germany the results are very similar (Burda and Hunt 2001, Parikh and Van Leuvensteijn 2003, Arntz 2006).

Regarding the topic of gender-specific migration patterns the literature is rare. A lot of micro studies implement a gender dummy, a lot of macro analysis calculate gender-specific mobility rates. But explanations of gender mobility gaps and analysis of gender-specific migration determinants are almost never provided so far. A remarkable exception is the work of Detang-Dessendre and Molho (2000) as well as the analysis of Faggian et al. (2007).5 The French authors investigate the migration patterns of young


5 Furthermore, in the context of developing countries some articles can be found dealing with rural-urban migration (Yang and Guo (1999) for China, Kanaiaupuni (2000) for Mexico). The relevance of these studies for our purpose is limited since gender roles in modern societies greatly differ from those in developing countries.
women in rural France after completing their education. They conclude that women might be more migratory than man since the woman usually moves to the man when a couple is formed. Women seem to be tied movers, in the sense of Mincer (1978). They move to regions where the male partner maximizes his income. Faggian et al. (2007) explore migration behaviour of university graduates in the UK. They also found higher mobility rates of young women. However, the higher mobility rates are not explained by partnership motives but by labour market factors. Faggian et al. derive the higher mobility of women in UK from the fact that migration is used as compensation mechanism for discrimination in the labour market. So, not only partnership consideration but also regional labour markets might be relevant for female migration decisions.

Aside from these two micro studies no econometric analysis can be found dealing with the gender-specific determinants of internal migration on the aggregate level. Our approach to identify gender-specific valuations of location factors and, thus, to explain the gender patterns of internal migration flows seems to be quite new and nonetheless of vital interest. We expand the framework of Davies et al. (2001) by implementing the gender issue, controlling for violations of basic assumption and adding explanatory variables. The last aspect is especially crucial since we do not solely focus on labour mobility but on educational and partnership-oriented migration too.

### 3 Empirical Approach

**3.1 Grouped Conditional Logit Model (GCL)**

Our empirical analysis is based on the micro econometric approach of McFadden (1974) known as Conditional Logit Model. Within the framework of a random utility model a probability function is derived which represents the likelihood of rational agents to choose a certain discrete alternative. As we will see, the concept is transferable to aggregate data of migration flows between regions which substantially lowers the computational effort. Otherwise, an estimation for millions of individuals and a choice set consisting of 439 alternatives (regions) could hardly be done. Our analysis focuses solely on regional characteristics as determinants of gender-specific migration. Since our data set does not contain information on individuals beyond age and gender we abstract from additional individual attributes. Thus our estimation assumes homogenous agents at the regional level, i.e. the groups – men as well as women – only differ with respect to their origin region.

---

6 A similar argument can be found in Zaiceva (2007). By analyzing the outcome of interregional migration in Germany she provides evidence that women reduce their labour supply after migration, but do not give it up entirely in case of migration in East-West-direction. Hence, a tied mover interpretation of a large part of female migration flows seems quite natural.

7 Alternatively, we could estimate a count data model to overcome computational problems. However, these models lack a sound micro-econometric foundation and, thus, a straightforward interpretation of coefficients.
To describe the applied concept in more detail, we have to consider a one stage decision of individual \( i \) between \( J \) alternatives. The decision is categorized as one stage process since choice set \( J \) also contains the source region. In other words, the individual simultaneously decides if she moves and where she moves. Staying in the source region is no qualitatively different phenomenon than moving to a different destination. Strictly speaking, the model assumes that the decision to stay can be seen as decision to move from the source to the source region.\(^8\)

Expected utility \( U \) of a (representative) individual \( i \) moving to region \( j \) is given by the equation:

\[
U_{ij} = \beta'X_{ij} + \varepsilon_{ij} \quad j \in J, i \in N
\]  

(1)

Vector \( X \) contains attributes of destination \( j \) as well as attributes of individual \( i \). All individuals face the same choice set \( J \). According to the rationality condition, an individual chooses the region that maximizes her utility. Thus, the probability to move to region \( j \) is given by:

\[
P(c_i = j) = P(U_{ij} > U_{ik}) \quad \forall k \neq j
\]  

(2)

Given the statistical properties described in McFadden (1974) the probability of individual \( i \) to move to region \( j \) can be expressed as:

\[
P(c_i = j) = p_{ij} = \frac{e^{\beta'x_{ij}}}{\sum_{j=1}^{J} e^{\beta'x_{ij}}}
\]  

(3)

According to Guimaraes and Lindrooth (2007), an indicator variable \( d_{ij} \) is defined which is set to 1 if individual \( i \) chooses option \( j \) or 0 otherwise. Then, the likelihood function of the migration decision is given by:

\[
L = \prod_{i=1}^{N} \prod_{j=1}^{J} P_{ij}^{d_{ij}}
\]  

(4)

In the case of grouped data, the likelihood can be concentrated if a group of individuals \( i \) can be treated equally and the choice set is the same for all individuals. The so called Grouped conditional logit model is formulated as (Guimaraes and Lindrooth (2007)):

\[
L = \prod_{g=1}^{G} \prod_{j=1}^{J} P_{gj}^{n_{gj}}
\]  

(5)

\(^8\) As we will see, in the model a dummy for non-movers is implemented to capture a possible qualitative difference between the decision to stay and the decision where to go. But the basic model itself does not make this distinction.
The exponent \( n_{ig} \) represents the number of individuals belonging to group \( g \) choosing region \( j \). The probability \( p \) to move to region \( j \) depends solely on destination attributes of region \( j \) and group characteristics \( g \). Individual heterogeneity of the members of a group is neglected in the model. The utility of individual \( i \) in group \( g \) deciding on destination \( j \) is given by:

\[
U_{igj} = \beta'X_{igj} + \nu_{igj}
\]

In our analysis we assign individuals to groups on the basis of their origin region. Therefore we obtain 439 groups and 439 potential choices. Thus, the data set consists of a 439x439 matrix. The log-likelihood function is:

\[
\ln L = \sum_{g=1}^{439} \sum_{j=1}^{439} n_{igj} \ln \left( \frac{e^{\beta'x_{igj}}}{\sum_{j=1}^{439} e^{\beta'x_{igj}}} \right)
\]

The factor \( n_{igj} \) refers to the number of women (or men) moving from region \( g \) to region \( j \). Since the groups are generated according to the origin region, choice probabilities are solely determined by regional attributes captured by the vector \( X \). Coefficient \( \beta \) can be interpreted as implicit price of the corresponding attribute \( X \) (Maddala 1983). Differences of coefficients between men and women indicate gender-specific valuations regarding a certain location factor.

One weakness of the approach is its treatment of non-movers. The non-moving option is regarded as equal to the alternative to move to any of the 438 remaining destinations. As Davies et al. (2001) argue, there might exist unobserved (fixed) costs of moving leading to a qualitative difference between migration and non-migration. As we estimate a one stage model this difference could bias the results. We follow Davies et al. who implement a dummy variable that indicates, if source and destination region are identical. This dummy is implemented to capture the effect of non-moving.

### 3.2 Two problems: Independence of irrelevant alternatives and overdispersion

Despite its sound microeconomic foundation the GCL model has two basic shortcomings: i) The model implies the independence of irrelevant alternatives (IIA), i.e. relative choice probabilities between two option are independent from existence and characteristics of other options (Maddala 1983; Dahlberg and Eklöf 2003). ii) Due to unobserved group specific heterogeneity a correlation between decisions of group members might deflate the variance-covariance matrix and inflate z-statistics – a problem referred to as overdispersion (Giumaraes and Lindrooth 2007).

i) Independence of irrelevant alternatives. The IIA assumption of the GCL model is rather idealistic in an interregional migration context where a lot of destinations are indistinguishable from the individuals perspective (Cushing and Cushing 2007). Thus, the corresponding Hausman test rejects the IIA hypothesis by comparing the coefficients of
estimations for different subsets of destination choices. The weakness can be remedied at least partly if the lack of IIA is seen as omitted variable problem (Guimaraes, Figueiredo and Woodward 2004). Then, the inclusion of an additional variable \( \gamma \) measuring the unobserved heterogeneity of destination regions avoids estimation biases to a large extent. A consistent way of constructing \( \gamma \) as fixed effects would require the implementation of dummy variables for every destination regions. Unfortunately the estimation containing such a large set of 439 dummy variables does not converge. Thus, instead of using dummies for every district, we implement a dummy on the federal state level. Then, the IIA property is implied only between region within a federal state but not between them. Since unobserved heterogeneity seems to be primarily relevant at the state level, the problem could be tackled by this approach.

\( i i ) \) Overdispersion. To avoid overdispersion caused by unobserved group-specific effects Guimaraes and Lindrooth (2007) propose the implementation of a random variable capturing the ignored group heterogeneity. The modified utility equation (6) is:

\[
U_{ij} = \beta' X_{ij} + \eta_{ij} + \nu_{ij}
\]  

The random effect \( \eta \) is supposed to be gamma distributed with parameters \((\delta_g, \lambda_g, \delta^*_g, \lambda^*_g)\) where \( \delta_g \) represents a group-specific parameter. The authors show that choice probabilities \( p' \) derived from (8) follow a Dirichlet distribution. The model can be estimated by ML-technique; the Likelihood function follows a Dirichlet-Multinomial multivariate distribution. Guimaraes and Lindrooth propose different methods to parameterize the random variable. We chose the option to treat \( \delta_g \) as constant.

4. Identification of explanatory variables

The log likelihood function (5) has to be maximized with respect to the parameter vector \( \beta \) which measures the implicit prices of the choice-specific attributes \( x \in X \). We model regional attributes as origin-destination relation. Therefore, the estimation does not include separate variables for origin and destination but only ratios or differences between them. With respect to these origin-destination-specific characteristics we distinguish between four groups of factors which we believe drive the migration behaviour of young adults:

i) labour market

ii) education

---

9 However, coefficients between estimations based on the full choice set and restricted estimations are equal in terms of direction and magnitude. Due to the small standard errors the Hausman test nevertheless rejects the IIA hypothesis.

10 The estimation was performed via a panel Possion Model with fixed effects using STATA. Guimaraes, Figueiredo and Woodward (2004) show the equivalence between a GCL and a Poisson estimator with group-specific effects.

11 The alternative procedure of determining the coefficient of correlation within the groups was also performed. The estimation results do not change.
iii) partnership & family 

iv) amenities.

i) Labour market. As shown in the literature review, the labour market conditions are part of nearly all studies analyzing the determinants of interregional migration flows. Following the models of Sjaastad (1962), Todaro (1969), and Harris and Todaro (1970), migration is described as investment in human capital. In the standard Sjaastad model of human capital investment a person moves to another region if the expected and discounted income differential between destination and origin region exceeds the monetary and non-monetary migration costs. Typically, the regional wage level and the unemployment rate are used as a measure of the expected income differentials. The distance variable acts as proxy for migration costs. We follow this approach and implement the regional average wage levels – computed as gross wages per employee – the unemployment rates as well as the of share of gender-typical occupations in our estimation. Whereas unemployment rates and gender-typical occupations can be disaggregated in terms of gender, we are not able to calculate gender-specific regional wage levels. However, it can be presumed that these gender wage differences only vary slightly between regions. To cover not only nominal income differences the implementation of the regional price level is necessary. Since appropriate regional price level data do not exist we include the building land prices. This variable represents at least partly regional rents which seems to be the main source for purchasing power disparities. The employed distance variable is defined as the time in minutes required when going by car from the administrative centre of the source to the administrative centre of the destination district. Since there might exist a structural break between short-distance and long-distance moves – short-distance moves might only affect residence and not jobs –, we estimate a separate regression where only long-distance moves are considered as migration. The frontier between short- and long-distance moves is set to 75 minutes travelling time since a duration below 75 minutes is officially regarded as reasonable daily commuting distance (Section 121 (4) German Social Security Code Book III).

ii) Education. Micro approaches stemming from a more socio-geographic field of migration research have emphasized the relevance of life cycle aspects for mobility decisions (Rossi 1980; McAuley and Nutty 1982). Different age groups seem to have specific needs as well as specific ties driving residential mobility. Since our analysis focuses on the age group of 18-30 years educational migration motives should be of particular relevance. Young adults of about 20 years usually start their tertiary education and might choose a location depending on its educational institutions. Generally speaking, it can be presumed that persons with high educational potential move to regions with appropriate third level educational institutions. Since these organizations are regionally concentrated, the strongest education-driven migration effects might be caused by this parameter. In contrast, the existence of vocational training opportunities might be less influen-

12 The literature linked to the life cycle theory primarily focuses on family migration and the related concept of housing careers (Winstanley et al. 2002).
tial for migration behaviour since institutions offering such opportunities are allocated more equally amongst the regions.

In our analysis, migration effects of third level education as well as of vocational training are considered. The first aspect is reflected by gender-specific college availability, defined as the share of students per high school graduates. It is a measure for the regional capacity to absorb school graduates by the regional academic opportunities. The effects of vocational training are implemented by vocational training availability which represents the number of vacancies and mediated positions per person seeking for vocational training.

Beyond these considerations, a further life cycle migration motive is related to the transition from education to employment. This transition typically happens until the age of 30 years. Spatial mobility seems to be a crucial requirement to find adequate jobs for university graduates whereas person with vocational training usually stay within their firm after completing their education. Thus, the regional labour market capacity to absorb university graduates is implemented as explanatory variable. The capacity is measured as employment gap, i.e. as gender-specific ratio of students per high qualified employees.

iii) Partnership & family. From a life cycle perspective the period between 18 and 30 years is also crucial for founding a partnership or family. Factors related to these more private spheres of life might be important for the residential choice of young adults. With respect to partnership motives it is well documented that persons choose partners with at least the same level of education (Blossfeld and Timm 2003). Since high qualified persons are usually supposed to be more mobile, regions with a high share of potential partners for high qualified migrants will exhibit more in-migration. This aspect of partnership matching is considered by two measures. On the one hand, the regional share of potential high qualification partners is calculated (high qualified partners). On the other hand, the regional sex ratio of high school graduates is implemented.

Regarding family issues we presume two exemplary concepts: Either, a reconciliation of work and family is aimed at, i.e. both partners have a job and externalize significant parts of child care. Or tasks are split in employment and household production which means that the sole earner has to generate necessary monetary resources.13 Egalitarian families should be attracted by regions with appropriate child care facilities. Traditional families have to maximize the income of the sole earner but are not affected by external child care. In addition a mixed family model is supposed to appreciate an adequate offer of part time jobs. To take these aspects into account we implement i) the number of child care places per children under 6 years ii) the share of gender-specific part time

13 Juerges (2006) shows the relevance of the distinction between sole earner (“traditional”) and double income (“egalitarian”) couples for their migrations decisions. Furthermore, Zaiceva (2007) provides evidence that women reduce their work supply after migration but do not give it up entirely in case of migration in East-West-direction.
jobs iii) the gender-specific participation rate. Which family model predominates can be answered by means of the estimated coefficients.

iv) Amenities. Due to their impact on life quality, the utility of residential choice is also affected by natural and cultural amenities. The cultural endowment is considered via the number of concert halls. Even if concerts are only a small part of cultural life it seems to be an appropriate proxy for the entire culture of a region. The extent of park areas as well as near-nature areas – defined in square kilometre per inhabitant – measures the natural component of amenities.

In addition to these four categories of explanatory variables we control for regional population size of the relevant gender and age group and implement the gender-specific population ratio between source and destination region. This variable measures the potential stock of in- and out-migrants. A more populous region is supposed to attract more young adults and vice versa more young adults should leave such a region. The effect on net migration is ambiguous.
5. Data

In our analysis, we use the migration data set for 2005 stemming from the migration statistics of the Federal Statistical Office. The data are based on the official register of residence and comprise all residential movements across district borders within Germany. It enables us to analyze migration flows at the small scale level of NUTS-3 regions. Since the migration data are laid down as a 439 origin-destination matrix, we know where the migrants come from and where they go. An individual is classified as migrant if she transfers her first residence from one NUTS-3 region to another during the year 2005. With respect to age and gender the data set differentiates between migration flows of men and women as well as certain age groups. Since our analysis focuses on young adults, we explore migration flow data for individuals aged from 18 to 30 years. A drawback of the official register of residence is its lack of information about crucial individual attributes, e.g. the educational status of movers cannot be observed. So, our analysis mainly has to focus on regional characteristics. We have to abstract from the impact of individual attributes as well as interactions between individual and regional level.

To avoid endogeneity bias, the explanatory variables in general refer to the year 2004. They are taken from different sources. Regional wage levels come from the German National Accounting of the Federal States (VGR der Länder); unemployment rates are provided by the German Federal Employment Office. The distance variable measuring the travelling time between two regions is computed in ArcGIS on the basis of a detailed German road map. The educational variables, child care availability, sex ratios of high school graduates and amenity variables are taken from the INKAR statistics of the Federal Office for Building and Regional Planning (BBR). Information on part-time-, gender- and qualification-specific occupations within a region stem from a comprehensive data set of the German Federal Employment Office. These data contain records for every employee registered in National Security System, i.e. for ca. 2/3 of total employment. Every record includes information on employee’s job location, gender, employment status (part-time/full-time), occupation and qualification. So, this data gives a very detailed description of regional labour markets. Table 1 presents a short illustration of all explanatory variables. Note that summary statistics refer to values of NUTS-3 regions itself while the estimation uses the computed ratios or differences between source and destination region.

---

14 Self-employed workers, civil servants and people working in liberal professions (e.g. lawyer, doctors, artists) are not covered by the data. However, there should be a high correlation between the characteristics of the included and not-included workforce. Furthermore, this issue concerns only few and less relevant industries so our analysis should be widely unaffected by this problem.
Table 1: Description of explanatory variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Men Mean</th>
<th>Men Min</th>
<th>Men Max</th>
<th>Women Mean</th>
<th>Women Min</th>
<th>Women Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average wage level</td>
<td>Average regional gross wage per employee and year (in Euro)</td>
<td>24611</td>
<td>18251</td>
<td>37932</td>
<td>24611</td>
<td>18251</td>
<td>37932</td>
</tr>
<tr>
<td>Unemployment rate ♂♀</td>
<td>Gender-specific unemployment rate (in per cent)</td>
<td>13.53</td>
<td>4.22</td>
<td>30.70</td>
<td>13.14</td>
<td>4.64</td>
<td>29.19</td>
</tr>
<tr>
<td>Typical occupations ♂♀</td>
<td>Share of gender-specific jobs (defined as jobs filled with at least 70% men or women in Germany)</td>
<td>0.43</td>
<td>0.25</td>
<td>0.67</td>
<td>0.38</td>
<td>0.16</td>
<td>0.52</td>
</tr>
<tr>
<td>Land price</td>
<td>Price of building land (Euro/m²)</td>
<td>94.08</td>
<td>5.41</td>
<td>886.21</td>
<td>94.08</td>
<td>5.41</td>
<td>886.21</td>
</tr>
<tr>
<td>Distance</td>
<td>Distance between centres of two regions required when travelling by car (in minutes)</td>
<td>261.3</td>
<td>0</td>
<td>703.3</td>
<td>261.3</td>
<td>0</td>
<td>703.3</td>
</tr>
<tr>
<td>Population</td>
<td>Number of gender-specific population aged 18 to 30 years</td>
<td>13457</td>
<td>2599</td>
<td>273986</td>
<td>13049</td>
<td>2419</td>
<td>277408</td>
</tr>
<tr>
<td>College availability ♂♀</td>
<td>Gender-specific number of students per high school graduates</td>
<td>3.44</td>
<td>0.00</td>
<td>31.85</td>
<td>1.32</td>
<td>0.00</td>
<td>14.71</td>
</tr>
<tr>
<td>Vocational training</td>
<td>Number of persons looking for training per offered training positions (in percent)</td>
<td>94.98</td>
<td>80.00</td>
<td>104.30</td>
<td>94.98</td>
<td>80.00</td>
<td>104.30</td>
</tr>
<tr>
<td>Employment gap ♂♀</td>
<td>Gender-specific number of students per high qualified job b</td>
<td>0.23</td>
<td>0.00</td>
<td>4.05</td>
<td>0.33</td>
<td>0.00</td>
<td>4.38</td>
</tr>
<tr>
<td>Participation rate ♂♀</td>
<td>Gender-specific number of employees per population at working age</td>
<td>0.53</td>
<td>0.22</td>
<td>1.94</td>
<td>0.46</td>
<td>0.21</td>
<td>1.04</td>
</tr>
<tr>
<td>Part-time jobs ♂♀</td>
<td>Gender-specific number of high qualified part-time jobs per high qualified jobs b</td>
<td>0.01</td>
<td>0.00</td>
<td>0.08</td>
<td>0.05</td>
<td>0.02</td>
<td>0.16</td>
</tr>
<tr>
<td>Child care</td>
<td>Number of kindergarten places per children under six years (in percent)</td>
<td>73.1</td>
<td>40.4</td>
<td>139.9</td>
<td>73.1</td>
<td>40.4</td>
<td>139.9</td>
</tr>
<tr>
<td>High qualified Partner ♂♀</td>
<td>Share of high qualified employees of opposite gender b</td>
<td>0.06</td>
<td>0.02</td>
<td>0.18</td>
<td>0.09</td>
<td>0.03</td>
<td>0.33</td>
</tr>
<tr>
<td>High school sex ratio</td>
<td>Number of female high school graduates per male graduates</td>
<td>1.35</td>
<td>0.54</td>
<td>3.59</td>
<td>1.35</td>
<td>0.54</td>
<td>3.59</td>
</tr>
<tr>
<td>Near-nature area</td>
<td>Near-nature area (m² per inhabitant)</td>
<td>51.73</td>
<td>2</td>
<td>864</td>
<td>51.73</td>
<td>2</td>
<td>864</td>
</tr>
<tr>
<td>Park area</td>
<td>Recreation area (m² per inhabitant)</td>
<td>43.81</td>
<td>7</td>
<td>222</td>
<td>43.81</td>
<td>7</td>
<td>222</td>
</tr>
<tr>
<td>Concert halls</td>
<td>Number of concert halls and opera</td>
<td>0.24</td>
<td>0</td>
<td>4</td>
<td>0.24</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

a In the estimations, the variables are implemented as origin-destination relation. Instead, table values refer to values of region itself not to the relation of regions. Otherwise an adequate interpretation would be difficult. b High qualification jobs are defined as jobs filled by high qualified employees = employees with academic degree.

Source: Own calculation.
6 Results

6.1 Descriptive evidence

Descriptive analysis of interregional migration flows within Germany in 2005 reveal a higher mobility of young women in comparison to their male counterparts (Table 2). For entire Germany women exhibit emigration rates of 9.62% whereas only 7.99% of the male population aged 18-30 years left their origin region. Only 7 of 439 regions show higher emigration rates of men than women. However, the gender mobility gap reduces significantly if we focus solely on long-distance migration.

Table 2: Emigration rates of interregional migration
- Average, Germany, 2005, 18-30 years -

<table>
<thead>
<tr>
<th>Emigration rate</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total migration</td>
<td>9.62%</td>
<td>7.99%</td>
</tr>
<tr>
<td>Long-distance migration (&gt;75 min)</td>
<td>4.18%</td>
<td>3.63%</td>
</tr>
</tbody>
</table>

* Göttingen is omitted since the biasing impact of the reception camp.

Source: Own calculation.

From a regional policy point of view not emigration rates but rather net migration is of particular interest (Table 3). Generally, round about two thirds of German districts face negative net migration with respect to the age groups under focus. The direction of migration goes from rural areas to agglomerations. Peripheral regions of East Germany are particularly affected by massive net outflows of young adults.

Table 3: Net migration rate of interregional migration
- Quantiles, 438 German districts, 2005, 18-30 years -

<table>
<thead>
<tr>
<th>Regional net migration rate (Total migration)</th>
<th>Regional net migration rate (Long-distance migration &gt;75 min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
</tr>
<tr>
<td>10%-Quantile</td>
<td>-5.10%</td>
</tr>
<tr>
<td>25%-Quantile</td>
<td>-2.79%</td>
</tr>
<tr>
<td>50%-Quantile</td>
<td>-1.45%</td>
</tr>
<tr>
<td>75%-Quantile</td>
<td>-0.14%</td>
</tr>
<tr>
<td>90%-Quantile</td>
<td>1.94%</td>
</tr>
</tbody>
</table>

* Göttingen is omitted since the biasing impact of the reception camp.

Source: Own calculation.

In terms of gender a high correlation between net migration flows of men and women can be observed: the coefficient of correlation of female and male net migration rates is 0.958. Nonetheless, a closer look reveals gender-specific differences. Female net migration rates of regions with strong migration losses are remarkably below the rates of men. Thus, regions not attractive to young adults primarily loose young women. In regions with strong net migration inflows the opposite tendency holds but to a lower extent.

To sum up, descriptive findings support the theses that i) young women are more mobile then young men ii) young women are more responsive to inappropriate location factors
than young men. As a consequence, outflow regions are characterized by remarkable imbalances in terms of sex ratios.

6.2 Econometric results

Table 4 presents estimation results of the GCL method controlling for overdispersion.\textsuperscript{15} The left part of the table displays coefficients and standard errors for gender-specific regressions without dummy variables for destination regions. To control for violation of the IIA implication the regressions on the right side of table 4 contain dummies for destination regions, although only on federal state level. To a large extent, results are very similar. However, interpretation of coefficients and gender-specific differences should be based on regressions results presented on the right side. In the last column of every regression regime (≠) comparison of gender-specific regressions are displayed. If the 95\% confidence intervals do not overlap, it seems to be justified to assume differences between men and women in their valuation of the corresponding location factor.

Before turning to the particular findings, the straightforward interpretation of coefficients in the conditional logit model is to mention. As Davies et al. (2001) show, a change in probability \( P \) of moving from \( g \) to \( j \) can be expressed as:

\[
\frac{\partial P_{gj}}{\partial x_{gj}} \bigg/ P_{gj} = (1 - P_{gj}) \beta \tag{9}
\]

Since \((1-P_{gj})\) is practically 1 for migration \((g\neq j)\), beta can be interpreted as proportional change of migration probability if the corresponding variable \( x \), i.e. the relation between destination and origin, alters by one percent.\textsuperscript{16} Another interesting aspect concerns the interpretation of the stay dummy (Davies et al. 2001). Suppose a hypothetic destination region indistinguishable from origin region then, the only difference of staying at origin or “jumping” to destination is expressed by the stay dummy. Therefore, the stay dummy measures all unobservable costs associated with migration itself. Analytically, the difference in probability of migrating compared to non-migration is given by:

\[
\frac{P_{\text{migration}} - P_{\text{stay}}}{P_{\text{stay}}} = e^{-\beta_{\text{stay dummy}}} - 1 \tag{10}
\]

Finally, to compare coefficients, it might be interesting which monetary compensation must be paid to offset the effect of a certain factor. For instance, we can calculate “how

\[\text{\textsuperscript{15} If the GCL Model is estimated without consideration of overdispersion, the standard errors are much more smaller and all coefficients are highly significant. Since confidence intervals become narrower, nearly all factors exhibit gender-specific differences.}\]

\[\text{\textsuperscript{16} Since the quadratic term, interpretation for distance is somewhat more difficult.}\]
many” unemployment one is willing to accept for an increase of one euro in wage level to stay indifferent between moving and non-moving.17

If we now turn to results, the estimations are not in favour of gender-specific impacts of location factors in terms of sign. Focusing firstly to the labour market and structural factors, high wages and low unemployment attract young men and young women. An increase of wage level ratio between destination and origin by 1% rises migration probability by about 0.6 to 0.7%. If destination unemployment goes up and the relation between destination and origin doubles, migration probability reduces by almost 2%. Also the price level variable shows the same sign for males and females – but in a surprising direction: young adults move to regions with high land prices. Since individuals care about real wages, this unexpected sign can be consistent with utility maximization when price levels are high where wages are high. Then, the wage effect might overweight higher price levels and, consequently, individuals move to high price level regions. The impact of distance on migration behaviour is u-shaped for both sexes. This finding is firstly driven by the large propensity of moving to adjacent regions. Secondly, the farer people move the lesser the binding impact of proximity becomes.18 Moreover, estimations are in favour of an obvious agglomeration effect. Young adults – men as well as women – are attracted by more populous regions. If a region is two times larger than another in terms of population, the probability to move to the larger region is about 3% higher. And, not surprisingly, the option not to migrate exhibits an exceptionally high probability. The probability to migrate is 99.7% lower than to stay. To compensate people for moving instead of staying, one would have to pay over 200,000 €.19 A gender-specific difference in terms of sign occurs with respect to the availability of gender-typical occupations. While the female coefficient shows the expected sign young men seem to be attracted by regions with low shares of occupations typically filled by men. However, to a large extent male-typical occupations only require low or medium qualifications. Given a substantial share of high qualified movers, the variable seems to be not very important for a lot of migrants.

Regarding the educational motives the expected signs predominate. Young adults are attracted by adequate facilities for college education and vocational training. Likewise, the absorption capacity of regional labour markets for university graduates stimulates in-migration. Some unpredicted effects have to be noticed on the field of partnership and family. Results for the part-time variable is not robust; participation rates and child care facilities have negative impacts on net migration. Despite the public ideology of the egalitarian family model, these findings might indicate, that migrants choose regions

17 See Davies et al. (2001) for a derivation of “iso-probability curves”.

18 Surprisingly, after a distance of 370 minutes the impact turns to a positive direction. However, only 5% of migrants and 0.4% of the total sample move over 370 minutes. So, the right tail of the distribution may be not well identified.

19 This calculation is comparable to the estimations of Davies et al. (2001) who quantifies compensation for certain states on 171,100$ or 238,000$. The concrete value is depending on the origin region. Our calculation is based on a region with an income of 24,611 Euro, i.e. the mean of our sample.
where child care is organized within the family, i.e. the traditional sole earner model dominates. However, our estimation does not support the thesis, that mostly women (have to) choose option of household production. With respect to the last dimension of amenities the estimated coefficients confirm the importance of cultural infrastructure – both for men and women. Finally, recreation areas seem to operate as pull factor.

Even if our analysis does not reveal substantial differences between men and women regarding their preferences for location factors, some dissimilarities can be observed at least in terms of intensity. Robust differences can be seen with respect to gender typical occupations, distance, population, non-migration dummy, college availability, employment gap and high school sex ratio. However, most partnership, family and amenity aspects do not exhibit gender dissimilarities. Compared to men young women are supposed to be more migratory and to choose rather nearby regions and agglomerations. To a larger extent, they are attracted by regions with adequate availability of universities and female jobs. Finally, women move to regions with high shares of male high school graduates. Men migrate to regions with appropriate capacities of high qualification positions. And they choose location with low share of female high school graduates – an unpredicted result since male adults also move to regions with high shares of partners with university degree.

Form a regional policy perspective it might be interesting if preferences change when only long-distance movers are considered. Table 5 displays the estimation results where only relocation with at least 75 minutes distance are counted as migration. Thus, effects caused by suburbanization trends or by arbitrarily fixed administrative borders are filtered out. And migrants seem to be less bounded by social networks and private loyalties when decisions have to be made solely between distant regions.

However, findings remain unchanged, with some exceptions. Child care now operates as pull factor as well as availability of part-time jobs. One may conclude that traditional family models dominate the short-distance migration while otherwise egalitarian concepts prevail. However, low participation rates also work as attracting location factor. Moreover, natural amenities seem to act as valued location factor. Regarding gender effects some variations can be noticed. On the one hand men respond more strongly than women to regional wage levels. An increase of 1% in destination-to-origin wage level causes a higher migration probability of 0,7% for male individuals whereas the female probability increases only by 0,45%. On the other hand, women appreciate factors facilitating work life balance (child care, participation rate). Finally, women are more attracted by cultural infrastructure.

\[20\] Nonetheless, due to their higher wage elasticity, men demand less than women \((207,550 \text{ €} < 227,320 \text{ €})\) to compensate non-movers hypothetically for migration.
Table 4: GCL-Regression for interregional migration
- 2005, 18-30 years -

<table>
<thead>
<tr>
<th></th>
<th>Without Dummy variables for destination region</th>
<th></th>
<th>With Dummy variables for destination region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour market &amp; structural characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average wage level</td>
<td>0.629*</td>
<td>0.032</td>
<td>0.783*</td>
</tr>
<tr>
<td>Unemployment rate ♀♂</td>
<td>-0.011*</td>
<td>0.001</td>
<td>-0.009*</td>
</tr>
<tr>
<td>Typical occupations ♀♂</td>
<td>0.313*</td>
<td>0.067</td>
<td>0.058</td>
</tr>
<tr>
<td>Land price</td>
<td>0.017*</td>
<td>0.001</td>
<td>0.018*</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.020*</td>
<td>0.000</td>
<td>-0.020*</td>
</tr>
<tr>
<td>Population</td>
<td>0.040*</td>
<td>0.001</td>
<td>0.046*</td>
</tr>
<tr>
<td>Stay Dummy</td>
<td>5.666*</td>
<td>0.011</td>
<td>5.944*</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College availability ♀♂</td>
<td>0.050*</td>
<td>0.002</td>
<td>0.018*</td>
</tr>
<tr>
<td>Vocational training</td>
<td>0.005*</td>
<td>0.001</td>
<td>0.007*</td>
</tr>
<tr>
<td>Employment gap ♀♂</td>
<td>-0.120*</td>
<td>0.008</td>
<td>-0.167*</td>
</tr>
<tr>
<td>Family &amp; partnership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation rate ♀♂</td>
<td>-0.467*</td>
<td>0.023</td>
<td>-0.358*</td>
</tr>
<tr>
<td>Part-time jobs ♀♂</td>
<td>0.835*</td>
<td>0.179</td>
<td>1.347*</td>
</tr>
<tr>
<td>Child care</td>
<td>-0.001*</td>
<td>0.000</td>
<td>-0.001*</td>
</tr>
<tr>
<td>High qualified Partner ♀♂</td>
<td>0.829*</td>
<td>0.093</td>
<td>1.630*</td>
</tr>
<tr>
<td>High school sex ratio</td>
<td>-0.076*</td>
<td>0.009</td>
<td>-0.080*</td>
</tr>
<tr>
<td>Amenities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Near-nature area</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Park area</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Concert halls</td>
<td>0.202*</td>
<td>0.007</td>
<td>0.170*</td>
</tr>
<tr>
<td>Observations</td>
<td>192721 (439x439)</td>
<td>192721 (439x439)</td>
<td>192721 (439x439)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-290752*</td>
<td>-278462*</td>
<td>-289461*</td>
</tr>
</tbody>
</table>

*a 1% significance level; ** 5% significance level; italics not significant at 5% level; ● represents gender-specific differences in coefficients based on comparison of 95% confidence intervals.

Source: Own calculation.
Table 5: GCL-Regressions for interregional long-distance migration a
- 2005, 18-30 years, migration over 75 min. -

<table>
<thead>
<tr>
<th></th>
<th>Without Dummy variables for destination region</th>
<th>With Dummy variables for destination region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td><strong>Labour market &amp; structural characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average wage level</td>
<td>0.884* 0.032</td>
<td>1.067* 0.034</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>-0.015* 0.001</td>
<td>-0.012* 0.001</td>
</tr>
<tr>
<td>Typical occupations</td>
<td>0.351* 0.069</td>
<td>0.387* 0.061</td>
</tr>
<tr>
<td>Land price</td>
<td>0.020* 0.001</td>
<td>0.020* 0.001</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.014* 0.000</td>
<td>-0.013* 0.000</td>
</tr>
<tr>
<td>Distance²</td>
<td>0.000* 0.000</td>
<td>0.000* 0.000</td>
</tr>
<tr>
<td>Population</td>
<td>0.037* 0.000</td>
<td>0.044* 0.001</td>
</tr>
<tr>
<td>Stay Dummy</td>
<td>7.066* 0.016</td>
<td>7.319* 0.017</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College availability</td>
<td>0.058* 0.003</td>
<td>0.017* 0.001</td>
</tr>
<tr>
<td>Vocational training</td>
<td>0.006* 0.001</td>
<td>0.007* 0.001</td>
</tr>
<tr>
<td>Employment gap</td>
<td>-0.166* 0.009</td>
<td>-0.199* 0.012</td>
</tr>
<tr>
<td><strong>Family &amp; partnership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation rate</td>
<td>-0.782* 0.024</td>
<td>-0.535* 0.020</td>
</tr>
<tr>
<td>Part-time jobs</td>
<td>1.534* 0.190</td>
<td>3.570* 0.483</td>
</tr>
<tr>
<td>Child care</td>
<td>0.000 0.000</td>
<td>-0.002* 0.000</td>
</tr>
<tr>
<td>High qualified Partner</td>
<td>1.269* 0.099</td>
<td>2.605* 0.144</td>
</tr>
<tr>
<td>High school sex ratio</td>
<td>-0.066* 0.010</td>
<td>-0.072* 0.011</td>
</tr>
<tr>
<td><strong>Amenities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Near-nature area</td>
<td>0.000 0.000</td>
<td>0.000* 0.000</td>
</tr>
<tr>
<td>Park area</td>
<td>-0.000* 0.000</td>
<td>0.000 0.000</td>
</tr>
<tr>
<td>Concert halls</td>
<td>0.331* 0.007</td>
<td>0.275* 0.008</td>
</tr>
<tr>
<td>Observations</td>
<td>192721 (439x439)</td>
<td>192721 (439x439)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-233040*</td>
<td>-224829*</td>
</tr>
</tbody>
</table>

a 1% significance level; 5% significance level; italics not significant at 5% level; ● represents gender-specific differences in coefficients based on comparison of 95% confidence intervals.

Source: Own calculation.
7 Discussion

What are the essential findings of our analysis? Generally, women and men seem to be attracted by the same location factors. Some characteristics are higher valued by men than by women and vice versa. But the direction of valuation is for the most part identical. Young adult prefer regions with high wage levels, low unemployment rates, appropriate jobs for university graduates and adequate educational institutions. Partnership potential as well as cultural infrastructure also act as pull factors. Additionally, even after controlling for main regional attributes, net migration goes from peripheral regions to more central districts with larger population.

Despite these general findings our analysis can at least partly contribute to the explanation of the observed gender mobility gap. Young women seem to be more mobile for their higher valuation of factors linked to agglomerations. Therefore they leave peripheral regions to a greater extent than men do. One main reason is their higher probability of university enrolment. But even after controlling for educational factors women tend to choose central regions more frequently than men. Our analysis also shows that women have a higher propensity to migrate than men even if the main location factors of their origin and potential destination regions are taken into account. So, location factors are only part of the explanation of the mobility gap. However, our estimations can not fully verify if these differences are finally attributable to gender. Educational motivated migration is not a gender phenomenon, but a consequence of educational potential. Women might be more migratory not because they are women, but for their better education.

As a final point, some limitations of our work have to be mentioned. Data stem from an aggregate migration statistics, hence, a lot of desirable information about migrating individuals as well as their life cycle position are not given and ecological fallacies cannot completely ruled out. An essential improvement of our analysis – which of course requires (longitudinal survey) data not available for Germany – would be the implementation of individual level variables and their interaction with regional characteristics. Of primary interest are education and life cycle attributes. Then, we could distinguish between education-, labour market- and family-related migrations and we could analyse if high qualified movers differ from low skilled migrants in their location preferences. Furthermore, discrimination between individual and household migration would be very useful. Otherwise, a huge share of tied movers within partnerships could bias results since their migration behaviour is driven by household and not individual preference. A man might choose a region not for loving its characteristics but for loving his female partner moving there.
Literature


