UNIVERSITÀ DEGLI STUDI DI MILANO

Department of Economics, Management and Quantitative Methods (DEMM)

LOMBARDY ADVANCED SCHOOL OF ECONOMIC RESEARCH

Lombardy Advanced School of Economic Research (LASER)

Ph.D. COURSE IN ECONOMICS

XVIII cycle

"ESSAYS ON ECONOMIC ASPECTS OF COMMUTING IN SWEDEN"

Ph.D. CANDIDATE: Sergii TROSHCHENKOV

SUPERVISOR: PROF. Massimiliano BRATTI
COORDINATORS: PROF. Paolo GARELLA
PROF. Alessandro MISSALE

A.Y. 2016-2017

“*In the great memory of my father,*

Oleksandr Troshchenkov,

because the end of life is not the end of love”

“В память о моем отце, Александре Трощенкове,

потому что конец жизни не означает конец любви”

Acknowledgements

Today, before the submission deadline, I would like to express my deepest gratitude to all the people who supported me on this way. Fists and foremost I would like to thank to my supervisor Professor Massimiliano Bratti for the leading me all this time, for his patience, knowledge and enthusiasm. Your passion and encouragement were extremely helpful during all ups and downs, extreme "turns" that could draw me out from my way of completion this dissertation. Thank you for conveying me with important advices and suggestions, every time I needed your help. I think, this dissertation would never be the same without your valuable comments and inspiration on all the stages of my research.

I would also like to show gratitude to my friend and coauthor, Professor Olle Westerlund from the University of Umea for his passionate attention and as desire to work with me during the writing of this dissertation and especially for the opportunity to complete visiting research at the University of Umea. Thank you, Olle, for the warm and friendly meetings during cold Scandinavian days. I am extremely grateful to you for the introducing to me the world of "labor economics" and "commuting". I am always impressed by your interest and dedication to the work, you do. Separately, I would like to acknowledge the assistance of the Professor Urban Lingren from the Department of Geography and Economic History of Umea University for the help in the access the dataset and his hospitality at the department. Generally, I would like to thank to all the people from the University of Umea for the support and friendly work environment.

Apart from that I am very grateful to the former head of L.A.S.E.R Professor Paolo Garella and current head of our PhD program Professor Alessandro Missale for the being nice and helpful persons that did all possible and impossible to facilitate students with the completion of PhD studies and writing dissertation. Also, I would like to thank Stefania Scuderi and Mikael Lindback for their immense help with tackling various administrative and other issues during this time.

I would like to mention a significant contribution to my work from Daniele Checchi, Tomasso Frattini, Marco Leonardi, Carlo Fiorio and Giovanni Pica who provided me with extremely useful criticism and valuable suggestion during the seminars at the University of Milan.

A special thanks goes to Catia Cialani, Joanna Tyrowicz, Erika Sandow and Amin Karimu − people who were an example to me all this time. Also, I wish to show gratitude to my friends: Andrey Falileev, Marko Ledic, Aleksey Klechikov, Alexander Babanskiy and Ruslan Huseynov for their support, motivation and inspiration; my colleagues: Pawel Sobstyl and Lukasz Sawka for the understanding and help in the many difficult situations.

All my gratitude and love goes to my parents: Oleksandr and Galina Troshchenkov and my wife Lesia. Mom, thank you for being always on my side regardless the situation, for teaching me to work hard and to be responsible to take care of other, for believing in me all this time. Thank you, Lesia, for helping and cheering me on my downs. I needed it, for motivating me to go further and do not give up. Dad, thank you, for believing and motivating me to study and for being proud of me. Wherever I have been, I felt your support and unwavering belief in my success. The completion of my PhD was your greatest wish. Even if you are not with us anymore, you gave me so much enthusiasm and inspiration. I will never forget it.

Contents

Conjoint work

Chapter 1 is the result of joint work with Professor Olle Westerlund from Department of Economics at the University of Umeå, Sweden. Chapters 2 and 3 are my own work, as is, overall the responsibility for any errors. The access to dataset was kindly offered by Professor Urban Lingren from the Department of Geography and Economic History at the University of Umeå, Sweden.

Summary of chapters

This dissertation studies economic aspects of commuting. It explores, on one hand the mechanism of self-selection into long-distance commuting, return to commuting and, on the other hand, the factors that determine exits from commuting. After investigation of the main bulk of literature about commuting, the research addresses the selectivity of commuters from ex-ante earnings and ability distributions (Chapter 1), monetary return to the commuting distance (Chapter 2) and factors that affects the probability of various exits from commuting spells (Chapter 3) with particular focus on the role of commuting distance and earnings. The analysis uses extensive longitudinal dataset with the precise geocoded information on the individuals' places of work and residence which is based on the administrative registers of Statistics Sweden.

The first research paper, titled "Self-selection into long-distance commuting on earnings and latent characteristics", focuses on understanding the nature of selectivity, as it is important factor in interpretation the results of empirical research. In our study we consider two potential dimensions of self-selection: the selection based on latent characteristics and the selection based on the measured earnings before starting long distance commuting. Both dimensions are captured using single model allowing identification of testable hypothesis about the simultaneous selection based on the previous earnings and latent characteristics. In order to conduct our analysis, we apply extensive administrative geocoded dataset with precise individual information including the coordinates of the places of residence and work. We demonstrate the negative selection of commuters from the ex-ante earning distribution. In the same time, our results indicate that the individuals with unobserved traits associated with higher earnings are also more likely engage into the long distance commuting.

The second research paper, titled "Return to commuting distance in Sweden", aims to estimate the magnitude of the economic return to commuting and compare the relative returns received by men and women. We apply fixed effect models to deal with individual unobserved heterogeneity that could potentially generate an endogeneity issue. We use a large dataset based on Administrative Registers for Sweden, which gathers detailed information on residential and job location, and indirectly on commuting. Results indicate that individuals receive relatively small compensations for commuting, with higher returns in agglomerations. Moreover, the relative return as a fraction of hourly wage is approximately similar across genders. This last finding provides evidence of similar bargaining powers for both men and women.

In our third paper, titles “Hazard from commuting: the role of earnings and distance. The case of Sweden”, we estimate the effect of earnings and commuting distance on the probability of exiting from a duration spell of commuting using a discrete time competing risk model. The data set, used in analysis, is based on the Swedish administrative registers from Statistics Sweden and the Swedish Tax Board and covers the period between 2000 and 2009. The problem of endogeneity of individual earnings and commuting distance in determining the length of work-related commuting spells is addressed using two-stage residual inclusion (2SRI). The estimates reveal that the earnings paid by firms have a positive impact on the probability of migration and a negative impact on the probability of job separation. At the same time, greater distance increases the probabilities of migrating closer to the place of work, re-employment closer to the place of residence and separation to non-employment while decreasing the probabilities of migration further away from the place of work and re-employment further away from the place of residence. The results are revealed to be robust in the samples of married and unmarried individuals.

Chapter 1

SELF-SELECTION INTO LONG-DISTANCE COMMUTING ON EARNINGS AND LATENT CHARACTERISTICS

Sergii Troshchenkov[*]

Olle Westerlund[+]

Abstract: Understanding the nature of selectivity is important factor in interpretation the results of empirical research. In our study we consider two potential dimensions of self-selection: the selection based on latent characteristics and the selection based on the measured earnings before starting long distance commuting. The both dimensions are captured using single model allowing to identify the testable hypothesis about the simultaneous selection based on the previous earnings and latent characteristics. In order to conduct our analysis we apply extensive administrative geocoded dataset with precise individual information including the coordinates of the places of residence and work. We demonstrate the negative selection of commuters from the ex-ante earning distribution. In the same time our results indicate that the individuals with unobserved traits associated with higher earnings are also more likely engage into the long distance commuting.

JEL codes: R 40, J61, J24

[*] University of Milan GRAPE (Group for Research in APplied Economics), email:sergii.troshchenkov@unimi.it

[+] Department of EconomicsUmeå University, Sweden, email: olle.westerlund@econ.umu.se

1.1 Introduction

Long distance commuting to work (LDC) has become a significant phenomenon in most developed countries, partially as a substitute for migration (Sultana & Weber, 2007; Sandow & Westin, 2010a). Previous studies have demonstrated that commuting options positively affect the job matching process, mitigate regional disparities, and satisfy labor demand of growing agglomerations and "grease wheels" of local economies (Edwardsson, 2000; Hazans, 2004; Lundholm, 2010). There is substantial literature on various socioeconomic and health outcomes of commuting (Koslowsky et al., 1996; Sandow, 2008; Gottholseder et al., 2009; Hansson et al., 2011; Lingren et al., 2014). A small but growing literature also study the effects of commuting on income and labor supply (Rouwendahl, 1998; Rouwendahl, 1999; Manning, 2003; Rouwendal, 2004; Mulalic & Pilegaard, 2010; Rupert et al., 2012).

Findings show that migrants as well as long distance commuters are non-randomly drawn from the total population and also from the total work force (e.g. Greenwood, 1985; Tunali, 2000; Eliasson et al., 2003). Increased availability of highly informative longitudinal micro data sets has improved the ability of researchers to control for confounders, but unmeasured characteristics of individuals may still lead to misleading results. Latent characteristics associated with selection into LDC and correlated with the outcome of interest is (e.g. earnings) is likely to cause bias in estimated effects. Therefore, studies on observational data usually use various econometric techniques to address this problem. In most cases, the issues of main interest are whether and how selectivity on unobserved characteristics into the group of migrants or commuters emerges and how this is related to ex-post income or other outcomes. Research on urban commuting indicates substantial non-random selectivity in the residential market, i.e. self-selectivity in observed residential location (Wasmer & Zenou 2002; Murata & Thisse, 2005; Wasmer & Zenou, 2006; Borck et al., 2008;. Rupert & Wasmer, 2009).

Knowledge on the nature of selectivity is important for interpretation of empirical results. This study contributes to the previous literature on selection into commuting in three ways. First, the main interest lies with selection on latent characteristics correlated with pre-LDC earnings instead of earnings after commuting. The rationale for this is that the decision to engage in LDC is taken concurrently or before commuting is initiated. Second, two potential dimensions of selection are considered: 1) one, based on unmeasured traits associated with pre-LDC earnings,

2) one, based on measured earnings before start of long-distance commuting. Following Heckman (1979), self-selection is based on unmeasured traits of commuters. The selection on unobserved characteristics is positive when individuals who exhibit latent propensities to engage in commuting are characterized by unobserved attributes that result in higher earnings. By contrast, negative self-selection is present when latent characteristics are associated with unexpected higher probability of LDC and lower earnings conditional on observed attributes, (e.g. Ihlanfelt, 1988; Zaiceva, 2006).

Another line of research has studied selection in terms of measured earnings previous to mobility in terms of commuting or migration; i.e positive selection means that mobile workers come from the upper part of the ex-ante earnings distribution. Negative selection occurs if long distance commuters (or migrants) are predominantly drawn from the lower part of the earnings distribution (e.g Öhman et al., 2003; Gabriel and Schmitz, 1995; Finnie, 2001). In the present study, both these aspects of selection are captured within a single model allowing testable hypotheses about selection based on both observed earnings and unmeasured attributes of potential long-distance commuters. Another contribution is that we use highly informative longitudinal population register data covering the whole population in Sweden. It includes detailed socioeconomic information and geocoding in 100 meter squares of individual's workplaces and places of residence.

Our main results indicate that men who commute over longer distances are negatively selected from the ex-ante earnings distribution, i.e. LDCs tend to have lower earnings than expected conditional on *observed* characteristics as measured the year before engaging in long-distance commuting. But at the same time, our findings indicate that individuals with *unobserved* traits associated with higher earnings are also more likely to engage in long-distance commuting. A possible interpretation is that the lower earnings before commuting reflects labour market mismatch, and unobserved traits associated with higher incomes also correlate with higher expected return of spatially extended job-search. The results for females are similar, although the estimated negative selection on earnings before start of LDC is not statistically significant different from zero.

The next section presents the econometric model and the parameters of main interest. Section 3 presents data, model specifications and descriptive statistics. Sections 4 and 5 give the results and robustness analysis followed by summary and conclusions in Section 6.

1.2 Earnings, commuting and self-selectivity

Following the modelling of migration in Tunali (2000) and Nakosteen et al. (2008), self-selection into long-distance commuting is assumed to be based on observed and unobserved (latent) characteristics. These may affect earnings as well as the probability of long-distance commuting. Here, for a sample of employees, LDC is defined as a change of work place involving transition from short-distance commute to long-distance commuting. Similar to most studies on migration, long-distance commuting is measured as a dichotomous outcome based on explicit criteria, here a commuting distance of at least 50 km. This will be discussed further in Section 4.

Individuals are observed at consecutive points in time. Selectivity is assumed to be manifested through two sources. One is observations of individual's earnings at the time the decision to start long-distance commuting or not is taken (first period). The other is the correlation between unobserved heterogeneity affecting first period earnings and unobserved heterogeneity impacting the probability of engage in long-distance commuting in the second period.

In the first period individuals are employed at an initial work place in a specified initial geographical location and consider employment options at other work places. Earnings of individual i in the first period are:

$$y_i = \beta' x_i + \varepsilon_i \tag{1}$$

where y_i denotes earnings, x_i is a vector of explanatory variables, and β is a vector of unknown coefficients to be estimated. The random error term ε is normally distributed with zero mean and variance σ_ε^2.

In the second period, the individual chose to become a long-distance commuter or not. The two outcomes are $c_i = 1$ if long-distance commuting is chosen, and $c_i = 0$, otherwise.

Back in the first period, the individual evaluates future employment options and the expected income during the second period is:

$$y_i' = y_i + \omega_i \tag{2}$$

The term ω_i is adjustment for influence of latent characteristics, from the individual's point of view representing the expected increase in earnings.

The self-selection mechanism can be expressed in terms of the difference in expected outcomes of alternative choices, here the difference between expected earnings if long-distance commuting is chosen, and expected earnings in the alternative case:

$$E(y_i'|c_i = 1) - E(y_i'|c_i = 0) = E(\omega_i|c_i = 1) - E(\omega_i|c_i = 0) \tag{3}$$

Individuals who start long-distance commuting are then self-selected on unobserved characteristics that influence expected earnings as evaluated in the first period. For example, unobserved ability may be associated with systematic "positive" or "negative selection", depending on the level of ability of long distance commuters. A "positive" self-selection on unobserved traits may be due to high ability associated with high earnings in any location of work places, but also associated with additional earnings premium in case of accepting a job offer involving long-distance commuting. A "negative" selection could stem from low ability associated with low earnings at any work place but at the same time combined with relative high expected pay off if accepting long-commutes,

Long-distance commuters are also systematically self-selected on observed attributes traits measured in available data. The nature of selectivity is an empirical question. High earnings in the first period may signal good job matches, high opportunity costs for jobs search and therefore low job search intensity and low incentives for job mobility. The probability of selecting long-distance commuting would then be negatively associated with earnings. On the other hand, high earnings could be associated with specific skills to be matched to few job openings on regionally "thin" labor markets (e.g. Manning, 2003). Long-distance commuters to a new work place can therefore be positively selected with respect to observed earnings in the first period.

The econometric model to be estimated considers systematic self-selection on observable characteristics as well as on unobserved heterogeneity. Let c_i^* indicate the latent propensity of individual's to engage in long-distance commuting. This option is chosen when $c_i^* > 0$. The joint model of earnings (eq. 1) and LDC is

$$\begin{aligned} y_i &= \beta' x_i + \varepsilon_i \\ c_i^* &= \alpha y_i + \delta' z + \omega_i \end{aligned} \tag{4}$$

where z_i is a vector of measured characteristics, δ a vector of coefficient parameters , and ω_i is the individual specific error term in the equation for expected second period earnings (eq. 2). [1] All variables with the exception of LDC status are measured in the initial period assuming that potential self-selection is reflected by the decision to start LDC and not necessarily in subsequent events.

A dichotomous variable (c_i) indicate whether the individual is observed as a long-distance commuter in the second period or not, and relate to the unobserved latent propensity for LDC (c_i*) as:

$$c_i = \begin{cases} 1 \; if \; c_i^* < \; 0 \\ 0 \; if \; c_i^* \; \geq \; 0 \end{cases} \tag{5}$$

The error terms ε_i and ω_i are assumed to be bivariate normally distributed with zero means, variances σ^2 and 1, respectively, and covariance $\sigma_{\varepsilon\omega}$.

The hypotheses of self-selectivity on pre-commuting earnings can be tested through estimates of α in eq 3. Conditional on other characteristics, long distance commuters represent a positive selection on measured earnings if $\alpha > 0$ and a negative selection if $\alpha < 0$. Self-selectivity on

[1] Following the adoption of Hausman and Wise (1979) used in Nakosteen et al. (2008).

unobserved heterogeneity is reflected by the estimated covariance $\sigma_{\varepsilon\omega}$. Higher ability or other unobserved traits associated with higher earnings may also be associated with higher propensity for LDC, i.e. $\sigma_{\varepsilon\omega} > 0$,an indication of a positive selection on unobserved traits. A negative covariance reflects negative selection due to latent characteristics from the conditional earnings distribution (initial period) into LDC.

Any combination of positive or negative selection indicated by the two parameters are possible, e.g. a negative selection on initial period earnings does not rule out a positive or a negative selection on unobserved traits. For example, individuals in the lower part of the earnings distribution may have unobserved traits reflected in unexpected high earnings (conditional on observed characteristics) and at the same time unexplained high propensity for LDC.

Equations (1) and (3) are estimated jointly stating the model as two reduced form equations. The reduced form commuting equation is obtained by substitution for y_i in eq (3)

$$c_i^* = \alpha\beta' X_i + \delta' z_i + \widetilde{\omega}_i \tag{6}$$

where

$$\widetilde{\omega}_i = \alpha\varepsilon_i + \omega_i$$

The error terms $\widetilde{\omega}_i$ and ε_i are assumed to follow a bivariate normal distribution with zero means and covariance Cov $(\varepsilon_i, \widetilde{\omega}_i) = \alpha\sigma_\varepsilon^2 + \sigma_{\varepsilon\omega}$

Let $g(\varepsilon_i)$ denote the unconditional density of the error term in equation (1) for income in the decision period prior to realization of LDC or not (a function of $y_i - \beta' x_i$), and $f(\widetilde{\omega}_i)$ the conditional density function of long-distance commuting (a function of $\alpha, \beta' X_i, \delta' z_i, \sigma_{\varepsilon,} \sigma_{\varepsilon w,}$).[2] The likelihood function for the sample of C long-distance commuters and N individuals in the non-LDC group is:

$$L = \prod_{i=1}^{C} f(\widetilde{\omega}_i)|\varepsilon_i, c_i = 1) \cdot g(\varepsilon_i) \cdot \prod_{i=1}^{N} f(\widetilde{\omega}_i)|\varepsilon_i, c_i = 0) \cdot g(\varepsilon_i) \tag{7}$$

[2] See Nakosteen et al 2008, p 773 and 774 for details.

Estimates of equations (1) and (3) together with the variance and covariance parameters are obtained by maximising L.

As stated previously, the present model of selection into long-distance commuting on observed and unobserved characteristics is an application of the migration model in Nakosteen et al. (2008) which in turn is a partial adaption of Tunali (2000). A basic presumption is that individuals self-select for mobility based on expected outcomes in terms of earnings. These expectations are partially dependent on latent attributes unobserved of the researcher. In Tunali's model, selection on unobserved heterogeneity is present if, conditional on migrant status, the means of those attributes differ between movers and stayers. Here, self-selection on unobserved traits is captured by the covariance between random error terms in equations (1) and (3), which carry information on endogenous selection in addition to the selection on measured earnings.

1.3 Data and empirical model

We use longitudinal data from Swedish population registers administered by Statistics Sweden. Apart from the precise geocoded information on place of residence and work place, data provide detailed information on the individual's personal characteristics and labour market outcomes.

The main interest of this study lies with the determinants of spatial labor mobility in terms of changing location of individual's employment. Commuting distances of the stock of employed convey very limited information on labour mobility in terms of *changing* spatial allocation of labour supply. The overwhelming share of the total stock of employed (commuters) does not change work place from one year to another. Therefore, we sample from the inflow of new employees to all work places in Sweden. The sample consists of individuals who in 2007 were of age 20-64, employed or registered as unemployed in 2007, and who became employed at a new work place in 2008. They represents the major share of total flow of external hirings to all work places in the economy as measured between two time points, November 2007 and November 2008.[3] Individuals who change location of work places within the same firm are also included. Cases with missing information on their place of residence in 2007 or 2008 and cases with missing information on work place in 2008 are excluded. Students and individuals staying at

[3] It represents an understatement of total matches because multiple changes of individual's work place is not observed.

home with parental benefits are also excluded because of uncertainties regarding labour force status and identification of change of work place.[4] The sample includes 392 818 individuals, 206 281 men and 186 537 women.

Table 1. Sample means and standard deviations, long-distance commuters and short-distance commuters.

Variable	*Males*		*Females*	
	LDC	*Non-LDC*	*LDC*	*Non-LDC*
Log Earnings	*7.94*	*7.957*	*7.745*	*7.756*
	(0.555)	*(0.646)*	*(0.551)*	*(0.643)*
LDC	*0.077*		*0.053*	
	(0.267)		*(0.220)*	
Age	*42.312*	*40.476*	*44.952*	*41.397*
	(12.916)	*(12.814)*	*(12.452)*	*(13.011)*
Age squared	*1957.214*	*1802.561*	*2175.756*	*1883.033*
	(1095.073)	*(1071.235)*	*(1070.456)*	*(1089.432)*
Foreign	*0.122*	*0.093*	*0.125*	*0.099*
	(0.327)	*(0.291)*	*(0.331)*	*(0.299)*
Accessibility	*369123.2*	*245467.9*	*392355.1*	*270879.5*
	(451154.3)	*(359408.3)*	*(452774.9)*	*(379375.4)*
Post-secondary	*0.066*	*0.074*	*0.041*	*0.051*
education <2	*(0.249)*	*(0.262)*	*(0.199)*	*(0.221)*
Post-secondary	*0.231*	*0.278*	*0.384*	*0.426*
education >2	*(0.421)*	*(0.448)*	*(0.486)*	*(0.494)*
MSc or PhD	*0.015*	*0.021*	*0.009*	*0.021*
	(0.122)	*(0.143)*	*(0.098)*	*(0.146)*
Married	*0.405*	*0.376*	*0.461*	*0.358*
	(0.49)	*(0.484)*	*(0.498)*	*(0.479)*
Single	*0.029*	*0.027*	*0.1*	*0.085*
mother/father	*(0.17)*	*(0.162)*	*(0.3)*	*(0.279)*
Living with parents	*0.069*	*0.105*	*0.032*	*0.079*
	(0.253)	*(0.306)*	*(0.176)*	*(0.271)*
Children	*0.052*	*0.05*	*0.064*	*0.0526*
	(0.223)	*(0.218)*	*(0.244)*	*(0.223)*
Manufacture	*0.243*	*0.14*	*0.08*	*0.063*
	(0.429)	*(0.347)*	*(0.271)*	*(0.243)*
Construction	*0.105*	*0.121*	*0.014*	*0.015*
	(0.307)	*(0.327)*	*(0.118)*	*(0.122)*
Retail	*0.22*	*0.259*	*0.159*	*0.214*
	(0.414)	*(0.438)*	*(0.366)*	*(0.41)*
Private services	*0.086*	*0.091*	*0.056*	*0.073*
	(0.281)	*(0.288)*	*(0.23)*	*(0.261)*
Log Median of	*7.967*	*7.965*	*7.97*	*7.966*
earnings in LA	*(0.052)*	*(0.054)*	*(0.051)*	*(0.055)*
n	*17449*	*207863*	*10580*	*187837*

[4] We also excluded individuals with commuting distances over 50 km in 2007, because of uncertainties regarding discrimination between new and old work places. Robustness checks using a sample without these restrictions does not affect our main results and our conclusions. Results are available on request from the authors.

Long distance commuting (LDC) is defined as commuting over 50 km between place of residence and work place. Distance is measured as Euclidian distance based on co-ordinates (100 square meters precision). Individuals starting long-distance commutes amounted to 17 449 men (7.7 percent) and 10 580 women (5.3 percent).

In the earnings equation we control for age, education, foreign citizenship, sector of employment, and median of earnings in the region of residence. Covariates in the commuting equation measure individual's earnings in the first period (2007), marital status, age, education, foreign citizenship, and distance weighted access to jobs. Descriptive statistics by gender and commuting status are presented in Table 1.

Comparison of the means of previous earnings between LDC:s and non-LDC:s suggest a selection into LDC from the upper part of the unconditional earnings distribution. In line with expectations, Table 1 indicates that earnings in the previous year are higher for non-commuters than commuters.

The labor force participation rate of females in Sweden is high and previous studies suggest different commuting and earnings patterns for males and females (Albrecht et al., 2001; Sandow 2008; Lundholm, 2010; Sandow & Westin, 2010a). We therefore estimate the earnings and commuting equations separately by gender. Earnings and commuting are assumed to be determined by individual and regional characteristics. The variables in the earnings equation includes age, educational attainment, sector of employment, nationality and regional wage level. The age variable captures individual experience, productivity and life course effects. Educational dummy variables are additional indicators of human capital affecting earnings and pay-off from commuting. The reference category is educational attainment of secondary school of three years or less. Variation in regional wage levels is controlled for by a variable measuring the median of earnings in the local labour market area where the work place is located. A set of dummy variables captures earnings differences by sector of employment. Nationality is a dummy variable indicating individuals of Swedish origin. Table 2 gives the specifications of the empirical counterparts to equations 1 and 3.

The commuting equation includes covariates measuring previous earnings, age, education, marital status, presence of children, sector of employment, nationality, regional wage level, and regional accessibility to jobs. Previous studies show systematic influence of age, education, marital status and family characteristics on commuting. (Bartel & Lichtenberg, 1987; Borsch-Supan, 1990; van Ham et al., 2001; Sandow & Westin, 2010b; Lingren et al., 2014).

Table 2. Covariates in the earnings and commuting equations

Earnings equation	*Commuting equation*
Nationality	*Nationality*
Age	*Age*
	Age squared
Age squared	*Accessibility*
Education	*Education*
	Family status
	Presence of children
	Previous earning
Regional wage level	*Regional wage level*
Sector of employment	*Sector of employment*

Regional labour market tightness and regions attractiveness for commuters is captured by the median wage level in the region where the workplace is located. Conditional on place of residence (ex-ante), spatial accessibility to jobs will affect the probability of long distance commuting. Following Eliasson et al. (2003), the accessibility measure was defined as a discounted sum of all jobs discounted by the distance between population centers of labor markets and individuals place of residence.[5] Systematic differences in commuting distances by industry are captured by a set of dummy variables indicating individuals sector of employment.

1.4 Results

The joint estimation of the earnings and commuting equation is carried out using maximum likelihood (MLE). The parameters of main interest are α as indicator of selectivity on (observed) earnings, and the covariance σ_{ew}. The latter indicates association between unobserved heterogeneity that correlates with earnings and unobserved heterogeneity correlating with the probability of long-distance commuting.

[5] Access is measured as $\sum Ejdij^{-\alpha}$ where E_j is all jobs in region j and $dij^{-\alpha}$ is the distance decay function with distance measured as distance between labour market region of residence (i) and labour market region of work place and α is the distance decay parameter.

Males

The estimation results for males are given in Table 3. The earnings equation estimates are (qualitative) in line with expectations. Age, education and the regional wage level are positively correlated with earnings, the indicated concave age/earnings profile is also as expected. The point estimates suggest that earnings increase by age up to a turning point at about 50 years of age. The results also signal a significant premium of education. Relative to the baseline category of individuals with an educational attainment at the secondary level, the increase in earnings ranges from 21% for individuals with tertiary level education shorter that two years and up to 65% for males with a Master or PhD degree. Individuals of non-Swedish origin have lower estimated earnings, about -25% relatively to Swedish natives.

The coefficient estimates of the commuting equation indicate a negative and statistically significant selection into long-distance commuting on earnings ($\hat{\alpha}$ = - 0,6345, $|t|$= 6,73). The estimate of the covariance parameter σ_{ew} is positive and significant ($\hat{\sigma}_{\varepsilon\omega}$=0,8753, $|t|$= 8,27). Thus, while the LDC:s are systematically selected from the lower part of the (unconditional) earnings distribution, there is an indication of a positive selection into long-distance commuting on latent characteristics affecting earnings (a positive selection from the conditional earnings distribution).

Coefficients on linear and squared terms of age suggest a concave profile of age with an estimated turning point between 45 and 50 years. The probability of commuting increases also with the level of education. Presence of a spouse/partner and/or children decreases probability of long-distance commuting in comparison to single men. Curiously enough, individuals living with parents tend to be more mobile than single individuals without children. Moreover, presence of children is not associated with lower probability of LDC among men. Sector of new employment correlates with the probability of commuting, possibly reflecting variation in spatial workplace distribution across industries. The reference category is individuals who received a job in the public sector. The estimated suggest higher probability of LDC for employed in the other sectors except manufacturing. Turning to the covariates measuring regional attributes, the estimates are indicative of a positive association between regional wages and commuting. A higher regional wage level in the work place region increases the attractiveness of work places as commuting destinations.

Table 3. MLE-results for the earning and commuting equations. Sample of males. No exclusion restrictions on sample selection

Variables	*Earning equation*		*Commuting equation*	
	Coefficient	*Std. error*	*Coefficient*	*Std. error*
Previous earning (α)			*-0,6345****	*0,0943*
Age	*0,0591****	*0,0012*	*0,0530****	*0,0075*
Squared Age	*-0,0006****	*0,00001*	*-0,0006****	*0,00008*
Nationality	*-0,2553****	*0,0074*	*-0,3352****	*0,035*
Post-gymnasium level of education<2	*0,2150****	*0,0088*	*0,4243****	*0,0354*
Post-gymnasium level of education>2	*0,3198****	*0,0057*	*0,5556****	*0,034*
University level of education	*0,6526****	*0,0165*	*0,9651****	*0,0803*
Married			*-0,0262*	*0,018*
Single mother/father			*-0,0063*	*0,0453*
Living with parents			*0,355****	*0,0266*
Children			*-0,0011*	*0,0334*
Regional wage	*1,2079****	*0,0482*	*4,6128****	*0,2808*
Manufacture	*0,2014****	*0,0059*	*-0,5228****	*0,0267*
Construction	*0,1422****	*0,0066*	*0,1155****	*0,0267*
Retail	*0,0556***	*0,0066*	*0,1486****	*0,0201*
Private services	*0,2054****	*0,0121*	*0,091****	*0,0292*
Accessibility			*-1,13e-06****	*3,39e-08*
Constant	*-3,088****	*0,0165*	*-35,0406****	*2,0934*
Sigma (σ_{ew})		*0,8753****	*0,1059*	
Number of observations			*225312*	

Asterisks indicate significance level
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
The standard errors are heteroscedasticity robust and clustered at the individual level

Accessibility is (unexpectedly) negatively associated with LDC. One possible explanation is that the attractiveness of neighboring labor markets is offset by the spatial distance, or that regional access is higher in densely populated areas with higher density of jobs within shorter commuting distances.

Females

The estimates for females (Table 4) indicate similar selectivity on observed and latent characteristics as for males.

Table 4. MLE-results for the earning and commuting equations. Sample of females. No exclusion restrictions on sample selection

Variables	*Earning equation*		*Commuting equation*	
	Coefficient	*Std. error*	*Coefficient*	*Std. error*
Previous earning (α)			*-0,4388****	*0,1184*
Age	*0,0379****	*0,0012*	*0,0119*	*0,0078*
Squared Age	*-0,0003****	*0,00001*	*-0,0002***	*0,00008*
Nationality	*-0,1269****	*0,0074*	*-0,1845****	*0,0343*
Post-gymnasium level of education<2	*0,1868****	*0,0116*	*0,4645****	*0,0496*
Post-gymnasium level of education>2	*0,3125****	*0,0051*	*0,5079****	*0,0414*
University level of education	*0,6977****	*0,0266*	*1,4693****	*0,1077*
Married			*-0,2918****	*0,0235*
Single mother/father			*-0,1725****	*0,0367*
Living with parents			*0,6219****	*0,0416*
Children			*-0,0408****	*0,0420*
Regional wage	*1,3491****	*0,0493*	*3,4008****	*0,3948*
Manufacture	*0,1886****	*0,0084*	*-0,1179***	*0,0423*
Construction	*0,1623****	*0,01*	*0,2149***	*0,0776*
Retail	*0,0328****	*0,0074*	*0,3334****	*0,0261*
Private services	*0,2238****	*0,0125*	*0,4442****	*0,0413*
Accessibility			*-1,02e-06****	*4,52e-08*
Constant	*-3,9639*	*0,3898*	*-26,768****	*2,9132*
Sigma (σ_{ew})		*0,6537****	*0,1366*	
Number of observations		*198417*		

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
The standard errors are heteroskedasticity robust and clustered at the individual level

Selection into long-distance commuting on observed earnings is negative $\hat{\alpha}$= - 0,4388, |t= 3,71|) and the estimated covariance parameter is positive (σ_{ew} =0.6537). As for the sample of men, the results for women show systematic influence of latent characteristics associated with lower earnings, but at the same time a higher probability of LDC for individuals with higher than expected earnings conditional on observed traits.

Estimated coefficients of linear and squared term of age suggest a concave age-earning profile of female workers and earnings increases with education. The estimated difference in earning between individuals having a Masters or a PhD degree vis-à-vis the reference category of individuals with less than 12 years of schooling is 69%. Female workers of foreign origin receive on average 12% less than natives. Sector of employment also plays significant role in determining the level of earnings. Relatively to the public sector, the results suggest higher earnings in all other sectors: manufacturing 18%; construction 16%; retailing 3%; and private services 22%. Again, the regional wage level at the place of work is associated with higher earnings of individuals.

Also in line with the results for males, probability of LDC is concave in age and increases with level of education. Non-natives have lower probability of long-distance commuting and the regional wage level in the local labor market area of the work place seems to attract a larger share of LDC:s and the coefficient on access is negative. The estimates indicating relationship between sector of employment and LDC also show a similar pattern for females as for the corresponding results for the sample of men. In contrast to the results for men, having children is associated with a lower likelihood of commuting over longer distance among women.

In sum, the results demonstrate that long-distance commuters systematically self-selects from the lower part of the income distribution. It is in line with the prediction of our model which suggests that past income negatively affects the probability of commuting. Also, latent characteristics that affect earnings are positively correlated with the probability of LDC.

1.5 Robustness checks

To verify the robustness of estimated parameters of main interest (α and σ_{ew}), the empirical model was re-estimated using different definitions of long-distance commuting, a less restrictive sampling criteria, and by using an extended set of covariates.

Our definition of LDC > 50 km Euclidian distance, approximately corresponding to > [55-70] km road distance and at least 45 minutes one way travel, is meant to define a subsample of commuters who experience significant monetary and non-monetary losses associated with commuting. Following previous studies that analyse commuting behavior of individuals, we test 40, 50 and 60 kilometers as the threshold for definition of LDC (Mulalic & Pilegaard 2010; Sandow & Westin 2010b; Eliasson et al.; 2003, Manning 2003).

Based on our baseline specification of the model, Table 5 gives the results from stability checks with respect to different cutoffs for commuting distance defining LDC. The indicated negative selection on observed earnings into commuting is confirmed and the general pattern is that the negative selection increases with commuting distance. The results confirm our previous findings of a positive correlation between latent characteristics affecting earnings and the probability of LDC.

Table 5. Estimates by alternative criteria for definition of LDC

Variable	*Male sample*			*Female sample*		
	40 km	*50 km*	*60 km*	*40 km*	*50 km*	*60 km*
Previous	*-0.5169****	*-0.6345****	*-0.4279****	*-0.2886****	*-0.4388****	*-0.5507****
earning (α)	*(0.0848)*	*(0.0943)*	*(0.0596)*	*(0.1131)*	*(0.1184)*	*(0.1205)*
Sigma σ_{ew}	*0.7017****	*0.8753****	*0.7563****	*0.4354****	*0.6537****	*0.7881****
	(0.0951)	*(0.1059)*	*(0.0753)*	*(0.1289)*	*(0.1365)*	*(0.1407)*

Robust standard errors within parenthesis.
*Significance levels: * p<0.05, **p<0.01, *** p<0.001.*
The standard errors are clustered at the individual level.

The extended sample includes all individuals who find a full time job in 2008 regardless of previous commuting experience. The estimation results using our baseline model specification and alternative definitions of LDC are presented in Table 6. The point estimates α remain negative and increases with distance defining LDC. But they are now smaller in magnitude and not significantly different from zero for the sample of females. The estimates of the covariance parameter are positive statistically significant as before although they are smaller in magnitude.

Table 6. Estimates using alternative sampling criteria and alternative definitions of LDC

Variable	*Male sample*			*Female sample*		
	40 km	*50 km*	*60 km*	*40 km*	*50 km*	*60 km*
Previous	*-0.053**	*-0.068***	*-0.071***	*-0.072*	*-0.063*	*-0.007*
earning (α)	*(0.029)*	*(0.032)*	*(0.036)*	*(0.052)*	*(0.059)*	*(0.059)*
Sigma (σ_{ew})	*0.178***	*0.206****	*0.213****	*0.216****	*0.216****	*0.140**
	(0.039)	*(0.044)*	*(0.048)*	*(0.068)*	*(0.077)*	*(0.085)*

Robust standard errors within parenthesis.
*Significance levels: * p<0.05, **p<0.01, *** p<0.001.*
The standard errors are clustered at the individual level.

Regarding the extended model specification, we included the family patterns characteristics such as marital status and presence of children in order to capture variation in labor supply in the earning equation. In addition, regional dummies were included into the earnings and commuting equations to control for different regional specific characteristics not reflected by the variables measuring regional wages and regional access.[6,7] Using the baseline sample criteria, and allowing for different thresholds of distance defining LDC, the estimates from the extended model specification are presented in Table 7. Again, the point estimates of alpha are negative but statistically significant only in two cases out of five. The covariance parameter is positive and statistically significant in most cases.

Table 7. Estimates for the extended model specification and alternative definitions of LDC

Variable	*Male sample*			*Female sample*		
	40 km	*50 km*	*60 km*	*40*	*50*	*60*
Previous earning (α)	*-0.0212*	*-0.2109**	*-0.2889****	*-0.0216*	*-0,0926*	*N/A*
	(0.0837)	*(0.0927)*	*(0.1029)*	*(0.0558)*	*(0,1655)*	
Sigma σ_{ew}	*0.1152*	*0.3659****	*0.4994****	*0.1608**	*0,2658**	*N/A*
	(0.092)	*(0.1026)*	*(0.114)*	*(0.0731)*	*(0,1865)*	

Standard errors are in parentheses below the main coefficients
Asterisks indicate significance level
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
The standard errors are heteroskedasticity robust and clustered at the individual level

The general impression of the robustness checks is that the estimated parameters of main interest are relatively robust for the sample of males as compared with the sample of females. The signs of estimated parameters remain the same and they are statistically significant in most cases. Selection into LDC on observed earnings is negative and unobserved heterogeneity affecting earnings is positively correlated with unobserved factors affecting the probability of LDC. The magnitude of estimated parameters varies, moderately by different definitions of LDC but decreases substantially when using less restrictive sampling criteria. The results for females are more sensitive, especially for using different sampling criteria and extended model specification.

1.6 Summary and discussion

This study deals with non-random selection into long-distance commuting to work on observed and unobserved individual characteristics.

Using Swedish population register data we estimate self-selection into long-distance commuting on earnings and selection on latent characteristics affecting earnings and the probability of

[6] The regional dummies were aggregated on the NUTS2 level according to Nomenclature des Unites Territoriale Statistique (NUTS) classification of European Union.
[7] The results for the extended specification do not include the estimates on 60 km threshold determining LDC in the female sample due to the difficulties with convergence of the female sample.

matching with a job involving a long commute. Earnings and latent characteristics affecting earnings are measured the year before commuting is observed, i.e. approximately at the time when the decision to commute is made.

Our findings indicate that long-distance commuters are negatively selected on earnings the year before they start to commute to their new work places. However, individuals with latent characteristics associated with higher than expected income have also a higher than expected probability of engage in long-distance commuting. Selection on earnings and latent characteristics show the same pattern for both women and men. However, the results for women are considerably less robust than the results for men.

The negative association between earnings and propensity for long-distance commuting may reflect that commuting is preferred to migration because of spatial variation in costs for housing. Commuting of high income specialists facing thin regional labor markets seems to be of less importance quantitatively. However, recent entrants to the labor market with high education may be found in the lower part of the income distribution and, because of thin regional labor markets for specialists, commuting may be necessary for matching their skills with higher paid jobs.

Conditional on observed characteristics, the positive correlation between unobserved traits affecting earnings and probability of long-distance commuting speaks against job mismatch as an explanation for commuting to a new work place. A more plausible explanation is that heterogeneity in unobserved traits reflects individual ability associated with job search conducted with higher intensity, efficiency and over larger geographical areas.

To identify exact mechanisms for the observed positive selection on latent characteristics, further research using more direct measures of individual heterogeneity in cognitive and non-cognitive skills and measurement of different aspects of the job matching process is needed. For example, latent characteristics associated with higher than expected earnings (ex-ante) and higher probability of long-distance commuting, can yield even higher earnings (ex-post). Comparisons of how different dimensions of unobserved heterogeneity are associated with earnings (measured ex-ante and ex-post) may perhaps provide evidence on whether job mismatch on latent characteristics is a major explanation to long-distance commuting or not.

References

Albrecht, J., Bjorklund, A. and Vroman, S., (2003). Is there a glass ceiling in Sweden? *Journal of Labor Economics* 2003; 21(1), 145-177.

Bartel A. P. and Lichtenberg F. R. (1987).The comparative advantage of educated workers in implementing new technologies. *Review of Economics and Statistics* 69. pp 1-11

Borck, R. and Wrede, M., 2008. Commuting subsidies with two transport modes. *Journal of Urban Economics*, *63*(3), pp.841-848.

Borsch-Supan A. (1990) Education and its double-edged impact on mobility. *Economics of Education. Review* 9. pp.39-53

Edvardsson, I.R., Heikkila, E., Johansson, M., Persson, L.O., Stambol, L.S., (2000). Competitive Capital. Performance of Local Labor Markets – An International Comparison Based on Gross-stream Data. *Working Paper 2000:7. Nordregio, Stockholm*.

Finnie, Ross. (2001). "The Effects of Interprovincial Migration on Individuals' Earnings: Panel Model Estimates for Canada," *Research Paper No. 163. Business and Labor Market Analysis Division, Analytical Studies Branch. Statistics, Canada (October).*

Gabriel, P. E., & Schmitz, S. (1995). Favorable self-selection and the internal migration of young white males in the United States. *Journal of Human Resources*, 460-471.

Greenwood, M. J. (1985). Human migration: Theory, models, and empirical studies. *Journal of regional Science*, *25*(4), 521-544.

Gottholmseder, G., Nowotny, K., Pruckner, G. J. and Theurl, E. (2009). Stress perception and commuting, *Health Economics*, 18, pp. 559–576

Ham, M. Van, Mulder, C.H. & Hooimeijer, P., (2001). Spatial flexibility in job mobility macrolevel opportunities and microlevel restrictions. *Environment and Planning A*, 33(5), pp.921–940.

Hansson, E., Mattisson, K., Björk, J., Östergren, P.O. and Jakobsson, K., (2011). Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden. *BMC public health*, *11*(1), p.1.

Hausman, J. A., & Wise, D. A. (1979). Attrition bias in experimental and panel data: the Gary income maintenance experiment. *Econometrica: Journal of the Econometric Society*, 455-473.

Eliasson, K., Lindgren, U. & Westerlund, O., (2003) Geographical Labor Mobility: Migration or Commuting? *Regional Studies*, 37(8), pp.827–837.

Hazans, M., (2004). Does Commuting Reduce Wage Disparities? *Growth and Change*, 35(3), pp.360–390.

Ihlanfeldt, K. R., (1988). "Intra-metropolitan variation in earnings and labor market discrimination." *Southern Economic Journal*, 55: 123-140

Koslowsky, M., Aizer, A. and Krausz, M., (1996). Stressor and personal variables in the commuting experience. *International Journal of Manpower*, 17(3), pp.4-14.

Lundholm, E., (2010) Interregional Migration Propensity and Labor Market Size in Sweden, 1970–2001. *Regional Studies*, 44(4), pp.455–464.

Manning, A., (2003) The real thin theory: monopsony in modern labor markets. *Labor Economics*, 10, pp.105–131.

Mulalic, I. & Pilegaard, N., (2010). Wages and Commuting : Quasi-Natural Experiments ' Evidence from Firms that relocate.

Murata, Y. and Thisse, J.F., (2005). A simple model of economic geography à la Helpman–Tabuchi. *Journal of Urban Economics*, *58*(1), pp.137-155.

Nakosteen, R.A., Westerlund, O. and Zimmer, M., (2008). Migration and self-selection: measured earnings and latent cheracteristics. *Journal of Regional Science*, *48*(4), pp.769-788.

Rouwendahl, J., (1998). Search theory, Spatial Labor Markets and commuting. *Journal of urban Economics*, 43, pp.1–22.

Rouwendahl, J., (1999). Spatial job search and commuting distances. *Regional sciense and urban economics*, 29 (491-517).

Rouwendal, J., (2004). Search Theory and Commuting Behavior. *Growth and Change*, 35(3), pp.391–418.

Rupert, P. & Wasmer, E., (2009). Housing and labor market: time to move and aggregate unemployment.

Ruppert, P., Stancanelli, E. and Wasmer, E., (2009). Commuting, wages and bargaining power. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, pp.201-220.

Sandow, E., (2008). Commuting behaviour in sparsely populated areas: evidences from northern Sweden. *Journal of Transport Geography*, 16, pp.14–27.

Sandow, E., Westerlund, O. and Lindgren, U., (2014). Is your commute killing you? On the mortality risks of long-distance commuting. *Environment and planning A*, *46*(6), pp.1496-1516.

Sandow, E. & Westin, K., (2010). Preferences for commuting in sparsely populated areas.

Sandow, E. & Westin, K., (2010b). The persevering commuter - Duration of long-distance commuting. *Transportation Research Part A: Policy and Practice*, 44(6), pp.433–445.

Tunali, I., (2000). Rationality of migration. *International Economic Review,41*(4), pp.893-920.

Wasmer, E. & Zenou, Y., (2002). Does City Structure Affect Job Search and Welfare? *Journal of Urban Economics*, 51(3), pp.515–541.

Wasmer, E. & Zenou, Y., (2006). Equilibrium search unemployment with explicit spatial frictions B. , 13, pp.143–165.

Weber, J., & Sultana, S. (2007). Journey-to-Work Patterns in the Age of Sprawl: Evidence from Two Midsize Southern Metropolitan Areas*. *The Professional Geographer*, *59*(2), 193-208.

Westerlund, O. (1997). Employment opportunities, wages and international migration in Sweden 1970–1989. *Journal of regional science*, *37*(1), 55-73.

Zaiceva, A. (2006). Reconciling the estimates of potential migration into the enlarged European Union.

Öhman, M. & Lindgren U., (2003). "Who is the long-distance commuter? Patterns and driving forces in Sweden." *Cybergeo: European Journal of Geography*.

Chapter 2

Return to commuting distance in Sweden

Sergii Troshchenkov[+]

Abstract

The aim of this paper is to estimate the magnitude of the economic return to commuting and compare the relative returns received by men and women. We apply fixed effect models to deal with individual unobserved heterogeneity that could potentially generate an endogeneity issue. We use a large dataset based on Administrative Registers for Sweden , which gathers detailed information on residential and job location, and indirectly on commuting. Results indicate that individuals receive relatively small compensations for commuting, with higher returns in agglomerations. Moreover, the relative return as a fraction of hourly wage is approximately similar across genders. This last finding provides evidence of similar bargaining powers for both men and women.

[+] University of Milan, and GRAPE (Group for Research in APplied Economics)

2.1 Introduction

Commuting distances significantly increased during the last decades owing to the decrease of transportation costs, facilitation of accessibility to remote areas and improvements in infrastructures. Evidence from Western European countries suggests a relatively significant increase in commuting flows on daily and weekly basis (Sultana and Weber, 2007; Lyons and Chatterjee, 2008). The importance of matching demand and supply, and of ensuring the equilibrium in local labor markets in the context of sustainable development is acknowledged by policymakers of many European countries.. Sweden is not a peculiar country in this respect. The main aim of the Swedish Transport Policy Act of 2009, for instance, was to provide sustainable and efficient transport provision for population and industrial production in the whole country (Swedish Government, 2008; Sandow and Westin, 2010). In Sweden, major investments were made in the infrastructure with the aim of increasing commuting streams between regions (Nutek, 2000; 2001).

From a policy view, commuting is considered to be a solution to many issues generated by geographically separated labor markets such as mismatch between demand and supply.. The increase in both internal and external mobility flows might "grease wheels" for regional economies and speed up convergence between regions. Moreover, commuting mitigates earnings disparities and promote economic equality between regions (Lundholm, 2010; Nutek, 2000; Hazans, 2004). From the individual perspective, costs of commuting can be compensated by better carrier opportunities, increase in labour income or differences in terms of prices and amenities in the housing market (Renkow and Hower, 2000); Shuai, 2012). Commuters may enjoy the advantages of both better amenities where they live and higher wage in urban centers where they work (Fu and Ross, 2010; Hover and Renkow, 2000). Although commuting is seen as a solution to a variety of regional labor markets problems, it could also have negative consequences such as: the decrease of individual productivity due to absenteeism and psychological stress caused by commuting, the increase of traffic congestion, negative effects on environment and the social life of individuals (e.g., owing to reduced leisure time). Commuting also reinforces traditional family relationships ("male breadwinner" model) due to predominantly male commuting over longer distances. This, in turn, leads to the reinforcement of gender inequality within the household. Sandow, (2012) predicts, for instance, the increase of the probability of divorce for long-distance commuters. Lingren et al. (2014) find evidence of positive effects of commuting experiences on mortality.

The current literature provides two main reasons which might explain positive returns to commuting. The first maintains that firms possess some monopsony power over workers. This, in turn, allows employers to compensate commuting expenses incurred by workers (under competitive labor markets, workers are not compensated for commuting and wage levels are equivalent to the marginal product.) Therefore workers are able to re-negotiate wages with the employer for the fraction of commuting expenses in bilateral bargaining. An alternative theory suggests that employers might have different marginal productivity due to sectoral differences or agglomeration effects of urban centers.

In this study, we examine the role of commuting distance on individual earnings. We tested both hypotheses of re-bargaining the fraction of commuting costs applying an approach pioneered by Gutiérrez-i-Puigarnau and van Ommeren, (2010) and Mulalic et al., (2014), and differences in the productivity of employees due to the agglomeration effects presented by Ross and Fu (2010). Taking into account significant differences in gender commuting patterns, we stratified our analysis by gender. A significant part of previous studies (Mulalic et al., 2014; Ruppert et al., 2012; van Ommeren and Gutiérrez-i-Puigarnau, 2011), suggests concavity of the wage-commuting profile, i.e. earnings increase with commuting distance at a decreasing rate. We have experimented with different samples within economically reasonable rage of distances in the Euclidian space, and have found a confirmation of this fact in our data. Moreover, the fixed effects estimation, applied in our analysis, addresses the individual heterogeneity that poses a problem of endogeneity.

This study contributes to the existing literature in several ways. Firstly, we analyze a significantly richer register-based longitudinal dataset. Mulalic et al., (2014) similarly with Gutiérrez-i-Puigarnau & van Ommeren, (2010) focus their analyses on employer-level data. At the same time, Manning, (2003) and Ruppert et al., (2012) base their analyses on survey data. Another feature, that makes our analysis different, is the way of dealing with endogeneity. Mulalic et al., (2014) and Gutiérrez-i-Puigarnau & van Ommeren, (2010) conduct their analysis in a setting of exogenous reallocation of employers and measure the compensating return for the related difference in commuting distances. The other approach (Manning, 2003) aimed to analyze the association between commuting time and hourly wage considering commuting time being exogenous. Finally, some studies apply econometric methodologies requiring the availability of an instrument for commuting (Ruppert et al., 2012; Oswald, 1999). We ground our analysis on the assumption that the main source of endogeneity is individual time invariant heterogeneity that affects simultaneously commuting and earnings behavior. On one hand individuals with higher ability may have faster growing career or higher bargaining power that would allow them to re-bargain a higher fraction of commuting expenses. On the other hand, workers with lower levels of ability will have higher net commuting expenses, leading to accept jobs over shorter commuting distance. Hence commuting distance and earnings should be positively correlated. Support for this prediction was previously reported by the study of Ruppert et al. (2012). Under this identifying assumption, we obtain consistent estimates for the return to commuting by applying the fixed effects estimation procedure.

The structure of the paper is as follows. The next section provides an overview of the existing literature on returns to commuting, along with a description of commuting patterns and the wage formation mechanism in Sweden. The description of the econometric model is presented in Section 3. Section 4 described data and selection criteria for our main sample, analyzed in this paper. Sections 5 and 6 report the main results from regression analysis of the male and female samples together with a comparison with the findings of previous studies. Section 7 denotes the results of estimating a wage growth model. The comparison between returns to commuting patterns of males and females is described in Section 8. The heterogeneity in return to commuting is explored in Section 9. The conclusions are laid out in the Section 10.

2.2 Literature review: Commuting patterns and the consequences of commuting

There is a broad variety of theoretical and empirical studies that attempt to explain the positive return to commuting, and its causes and consequences. In this section, we briefly review the main papers which analyzed both theoretically and empirically these patterns, and the determinants and consequences of commuting.

2.2.1 Determinants and patterns of commuting

The importance of the spatial dimension of the labor market was recognized long time ago by Simpson, (1980, 1992), Rouwendal and Rietveld, (1994) and van Ommeren et al. (2000), among others. The main approach was to view commuting as the result of individuals' optimizing behavior during the job search process in spatial labor markets. This approach was pioneered and developed by Rouwendahl (1998, 2004) who suggested the existence of equilibrium in the labor market with spatial characteristics. The model also attempted to explain such phenomenon as "excess commuting"[81] which was considered as a significant issue in studies of commuting. Moreover, the model presented by Rouwendal (2004) suggested the existence of critical values for the maximally acceptable daily commuting distances for individuals. This approach was further developed by van Ommeren et al. (2000) in the context of job and residence choices, since the authors suggested these choices to be simultaneous. They concluded that factors that cause imperfections in the housing market have an ambiguous impact on the job-search process. In a similar fashion, Wasmer and Zenou (2002, 2006) developed a urban equilibrium where individuals work and reside in different locations, with employed and unemployed people perfectly segregated. Moreover, they introduced a land market which in turn leads to positive costs of reallocation. This allowed them to demonstrate the existence of zones in the city where employed and unemployed labor coexist. Further on, Rupert and Wasmer (2009) suggested that with high commuting costs, high frictions in the residential market play an important role in the decline of individual mobility. Together with that, Manning (2003) proved that commuting appears to be a result of the "virtually existing" monopsony power of employers that affects job

[8] The concept of "excess commuting" describes the difference between the actual commuting and equilibrium commuting within a monocentric urban model.

search through the spatial dimension. Along with all the above mentioned theoretical studies that developed the self-selection mechanism into commuting, it is worth mentioning the studies that are theoretical in nature, although contributing significantly to shaping the empirical analysis. Van Ommeren (2004) analyzed the commuting distribution. He showed the heterogeneity of vacancies or job search with the spatial component under the labor market rigidities. Moreover, this study suggested that the residential mobility does not contribute to explaining the shape of commuting distribution. The final important conclusion of his work is that the shape of commuting density function is similar for countries with different spatial structures.

While theoretical studies provide clear background theory for the selection mechanism, empirical studies have shown ambiguous results. The analysis carried out by Rouwendahl (1999) showed the importance of the spatial component during job search suggesting the fact that around 50 % of jobs originate from the nearest 25 kilometers area. The existence of such a boundary distance was further investigated by Lundholm (2010) who suggested the idea of “narrow labor markets” – the labor market with commuting distances that have an impact on daily life of individuals but that is tolerable for most of them. Furthermore, the concept of “extensive labor markets” was also developed, which are related to labor markets with commuting distances entailing pecuniary and non-pecuniary costs unacceptable for some individuals but still tolerable for others. Eliasson et al. (2003) found clear evidences of the significant impact of the labor market characteristics of surrounding areas on labor mobility, although without clear answers to the question of the impact of the characteristics of surrounding labor markets on commuting decision. In addition, the study conducted by Sandow, (2008) focused particularly on the various impacts of commuting in the “extensive labor market” using the administrative longitudinal data for northern Sweden where this issue is particularly relevant due to low population density and high commuting distances. Among other findings, the author reported that the commuting patterns of the population are significantly affected by the geographical structure. Moreover, clear evidences of significant gender differences in commuting patterns were also demonstrated by the previously mentioned authors.

2.2.2 Determinants of commuting

There exist some stylized facts about employees' commuting patterns. One being that the commuting distance decreases with *age and experience* (van Ham et al., 2001). The previous studies, such as Booth et al., (1999) showed that young people are more prone to commute than older people. The explanation could be that older people obtain more firm-specific human capital, and the subsequent return from job-to-job changes is lower than for younger people. Since the previous studies revealed substantial fixed costs of commuting, the expenses induced by long-distance commuting could be unacceptably high for them. On the other hand, older people have more experience than younger colleagues in the same educational category, so they have more career opportunities and, as a result, higher return from commuting. Due to the fact that commuting becomes more costly with age, more skilled employees are able to commute due to higher earnings (Osth, 2007). It is reasonable to assume that there is an age threshold for commuting. Before achieving this threshold age, commuting increases but after passing the threshold, it decreases.

It is shown by Dargay and Clark (2012) that the length of commuting distance is reasonably affected by the *population density* in the particular region of residence. Therefore people who live in rural areas travel more than those who live in metropolitan areas. This is likely to be related to the lack of employment opportunities in the place of residence. Indeed, rural areas are characterized by relatively lower population density and higher unemployment rate. Another important factor affecting commuting intensity is the concentration of firms and enterprises in a region. Ham et al. (2001) proved that the accessibility of employment is an important characteristic which affects the probability of job acceptance over a greater distance.

The effect of *education* on commuting distance is unambiguous. Previous studies such as Bartel and Lichtenberg, (1987) argued that more educated people have a faster developing career and, as a result, are willing to commute more. Borsch-Supan (1990) supported this finding by explaining it with the decreasing effect of transaction costs. Since higher education is assumed to lead to higher returns, the burden of fixed commuting cost will be lower for individuals with higher potential return. Better-educated individuals are able to carry the job-search process more efficiently, also due to their job-searching skills and network obtained during the years of education. It is also worth mentioning that jobs requiring higher education are often more specialized and less spatially dispersed than those that require a lower qualification.

Gender and *household characteristics* are also important determinants in the choice of commuting distance. It is a stylized fact that women on average commute less than men. Young single women approximately accept the same commuting distance as single men. The evidence proposed by van Ham et al. (2001) shows that, a highly educated unmarried woman has a higher probability of accepting jobs over a greater distance than men with the same characteristics. The age effect has a more significant impact on the probability of being a long-distance commuter for women. Having a partner who works has no effect on the commuting distance for men, however decreases this distance for women. The explanation of this result could be an additional workload on the woman in household production and supports the theory of the "traditional family" with one working spouse. As expected, the presence of children has an impact on both partners by making them less spatially mobile than single or unmarried people. The likelihood of commuting for a long distance is negatively related to the number of children in the family (McQuaid and Chen, 2012). Contrary to all these findings, Camstra (1996) showed that gender effects are almost absent for the "modern groups"[9] of the population.

The *sector of employment* has also an important effect on commuting distances. Workers employed in the financial, business, and construction sectors commute more than those who are employed in health care or education sectors (van Ham et al. 2001). This can be explained by the fact that jobs in the financial, industrial, and banking sectors are relatively spatially concentrated, while vacancies in social services are more evenly geographically dispersed. Another important factor which increases the probability of commuting for a long distance is the effectiveness of transportation.

2.2.3 Consequences of commuting

Van Ommeren (2002) focused on the consequences of commuting applying the equilibrium job-searching model. The author demonstrated that in the presence of imperfections such as

[9] Groups with characteristics attributed to the "modern lifestyle" such as late marriage or high labour market mobility.

searching costs and bargaining between employers and workers, the presence of market power dictates the extent to which workers can be compensated for commuting. Yet, a surprising evidence was that the workers with stronger market power are compensated less than workers with weaker market power. This stream of research found further development in the study of van Ommeren & Rietveld (2005) of the "commuting time paradox". The authors suggested that under the conditions of constant labor market tightness, the ratio between commuting expenses (pecuniary and non-pecuniary) and wages remains constant over time. It is explained, on the one side, by the increase in productivity in the long run that leads to the increase in wages, and that leads, on the other side, to the shift of the preferences in the transportation mode. The shift in preferences leads to a decrease in non-pecuniary costs but to a rise of pecuniary costs (for example individuals can choose faster but more expensive transport modes such as fast trains or private cars instead of ordinary public transport).

Many studies are focused on the consequences of commuting on various socio-economic aspects. Rouwendahl (1999) estimated the willingness to accept a lower wage of 0.12 Gulden which was approximately 1% of the hourly wage in order to avoid an additional kilometer of commuting for the Netherlands. Further on, Manning (2003) suggested that workers are not fully compensated for long-distance commuting. Moreover, it was found that the job separation rate for commuters is higher than for stayers. The evaluation of the compensation for commuting was further developed by Fu and Ross, (2010) who showed clear agglomeration effects for the wages of commuters. Ruppert et al. (2012) reported the significant impact of commuting time as well as vacancy characteristics on job-acceptance decisions and future wages rates. The authors documented the evidence of wage increasing with distance at decreasing rates. Mulalic et al., (2014) estimated the bargaining power of employees through the estimation of wage increases owing to long-distance commuting in case of exogenous firm reallocation. They suggested that individuals are able to re-bargain ex-post around 0.5% of the salary for every kilometer increase in commuting distance. Van Ommeren & Fosgerau, (2009) analyzed the workers' daily marginal cost of commuting and suggested it to be about 17 Euro per hour of daily commuting time. Gutiérrez-i-Puigarnau & van Ommeren, (2010) suggested that commuting increases the daily and weekly labor supply of individuals, whereas the subsequent study (van Ommeren & Gutiérrez-i-Puigarnau, 2011) demonstrated that commuting positively affects the rate of absenteeism and job separation of individuals. It is also worth mentioning the study of Hazans, (2004) who showed in his analysis that commuting decreases the regional urban-rural wage and employment disparities, reducing inequality, between the capital city and surrounding regions, and positively affects national output.

2.2.4 Commuting in Sweden

Previous studies conducted in the field of labor mobility in Sweden such as Lundholm, (2010); Sandow and Westin (2010); and Sandow (2008) suggests that 50% of men commute less than 8 kilometers to their job whereas 50% of women commute for less than 6 kilometers. Such a difference in commuting distances can be explained by many factors: such as the role of women in household production, individual heterogeneity towards commuting or the different industry chosen by individuals. Male are typically employed in the construction, manufacturing and retail sectors, while women mainly work in private and public services.

Results from the previous studies carried by Lundholm, (2010) and Sandow, (2008) suggest that commuters receive significantly lower income in comparison to stayers in the male and female subsamples of the population. Male, commuters over 30 kilometers earn 2300 hundreds of annual income (approximately 25555 EUR) compared to 3,369 hundreds for stayers (37,433 EUR). The difference in earnings between commuters in the female population is even stronger: 1,351 hundred SEK (15,011 EUR) for commuters and 2,492 hundred SEK for stayers (26,263 EUR). Therefore, the return to commuting becomes ambiguous in comparison with the predictions of theoretical models, at least when considering the raw data.

All above mentioned studies give a clear theoretical framework and allows us to proceed further in our empirical analysis.

2.4 Description of the econometric model

This section contains the description of our empirical model together with the motivations behind our choices. The formal identification strategy will be presented later on in Section 4.

In this analysis, the relationship between *annual earnings* as a dependent variable and the distance of commuting as the main independent variable, including a set of various socio-economic and geographic control variables is studied through the application of the fixed effect model. In our setting, fixed effect estimation allows us to estimate models with longitudinal data accounting for individual heterogeneity, addressing potential endogeneity issues (individual self-selection into commuting) generated by time-invariant unobservable characteristics. Previous studies suggest, indeed, that there might be unobserved individual time invariant features influencing simultaneously commuting distance and individual earnings (Ruppert et al., 2012; Mulalic et al., 2014). Our analysis is carried out on a 7-year longitudinal panel dataset using Ordinary Least Squares (OLS) and Fixed Effect (FE) estimation. It has borrowed some features from the previous studies conducted by Manning, (2003), Mulalic et al., (2014) and Ruppert et al. (2012) together with the set-up and selection of variables made for Sweden by Elliason et al., (2008) together with Sandow and Westin, (2010).

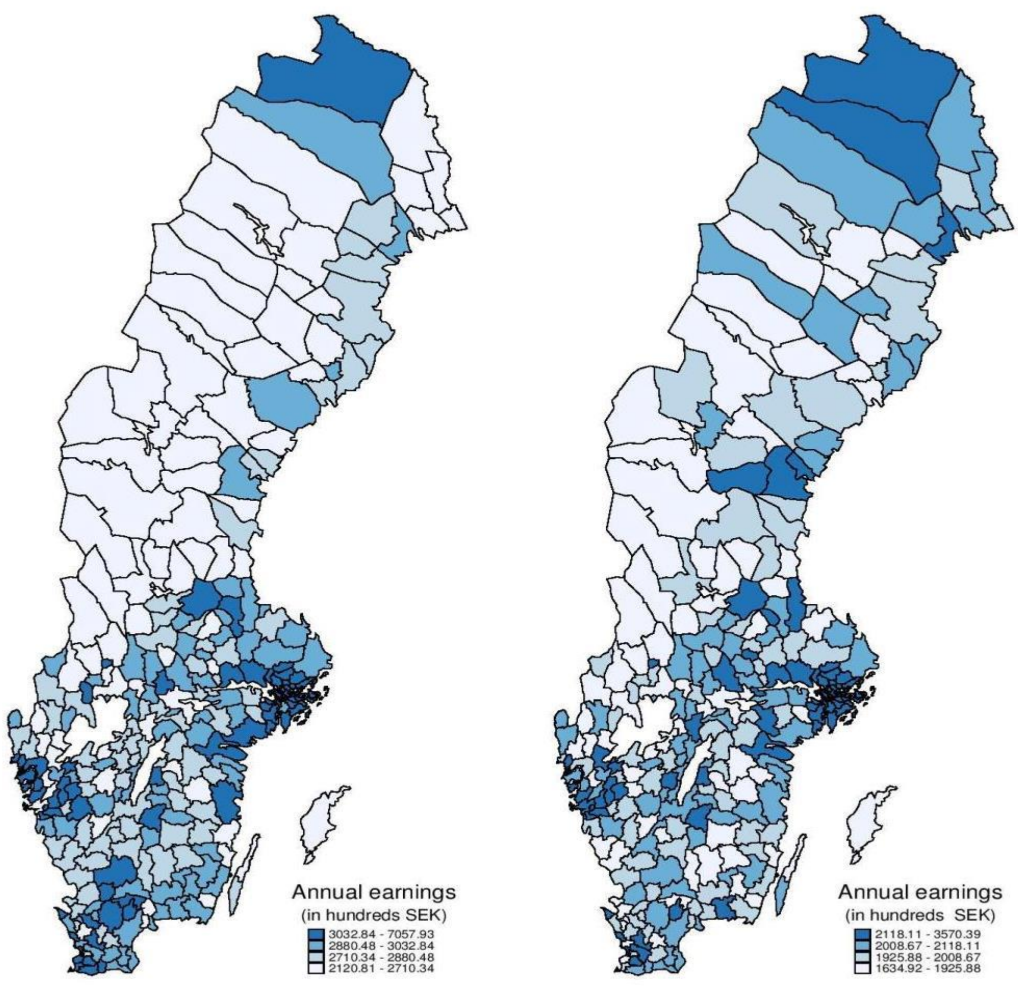
Taking into account the underlying theory and the results of analysis of the descriptive statistics, the model for estimation of effect of commuting distance on annual earnings takes the form:

$$Log\ (Annual\ earnings)_{it} = \alpha_{it} + \gamma_1 Distance_{it} + \gamma_2 (Squared\ Distance)_{it} + \beta_1 X_{it} + z_i + \varepsilon_{it} \qquad (1)$$

where $i=1 \dots T$ stands for cross-section units (individuals) and $t=1 \dots K$ indicates time, whereas α, γ and β are coefficients to be estimated and X is a generic vector of additional explanatory variables that captures individuals' lifecycle events and labor market conditions at the place of work. z_i is the individual fixed effect and ε_{it} an error term. The list of variables and their definition is presented in Table 1.

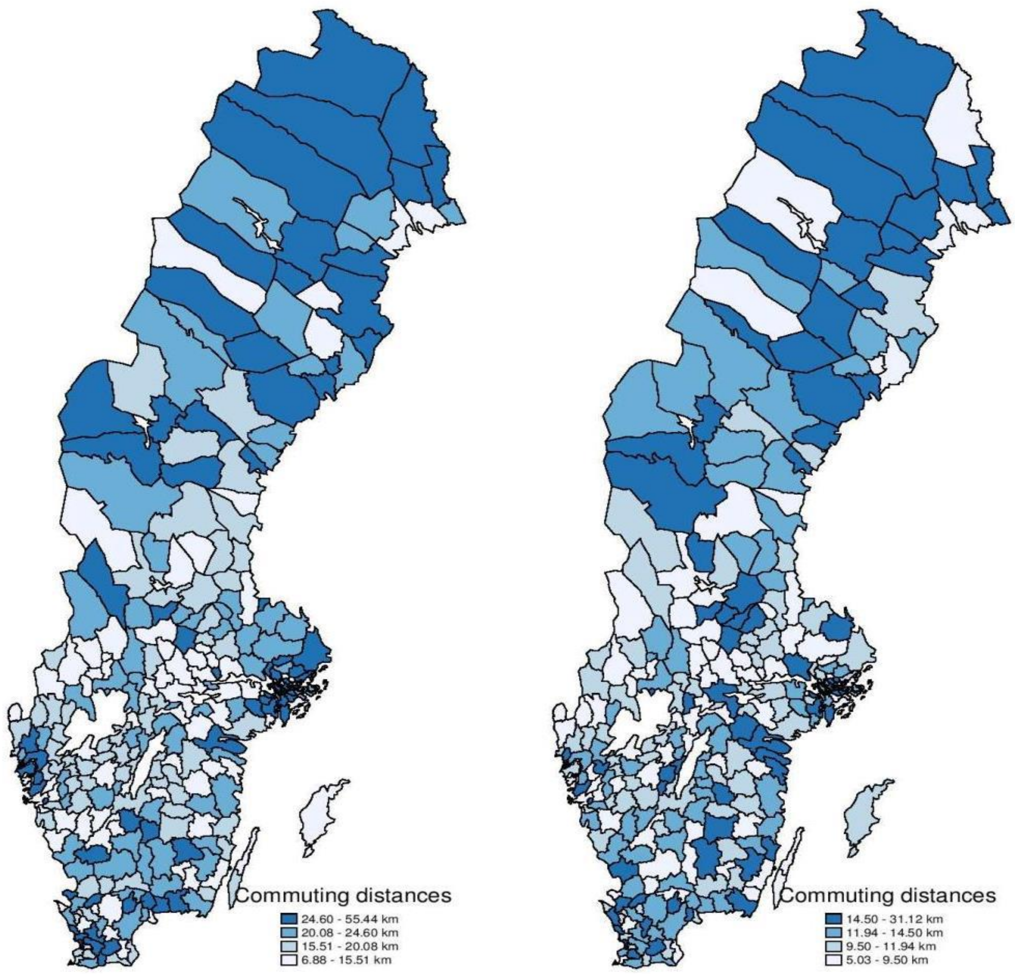
The main dependent variable *Annual earnings* indicates the annual earnings that individuals received from employment in natural logarithms. Since the main assumption of the model is that individuals should work full time, *Annual earnings* was considered in the interval 1,500 hundreds SEK and 8,518 hundreds SEK of gross annual income. The rationale behind imposing the lower threshold, which constitutes the lowest quartile of the earnings distribution, is explained by the need to eliminate the bias generated by the inclusion of part-time employment. Individuals with part-time employment might possess more spare time for commuting. The introduction of the upper bound is explained by the exclusion of individuals who are more likely to work overtime (Isacsson and Swärdh, 2007).

Figure 1. Average earnings by municipality in Sweden. Male and female samples



The spatial distribution of average annual earnings in the male and female samples is presented in Figure 1. The municipalities with lowest earnings are concentrated in the middle and west part of the country. At the same time the highest earnings are shown in the three biggest urban agglomerations: Goteborg, Stockholm and Malmo, and in municipalities along the coastal line. The county of Norrbotten also shows high earnings. It can be viewed as an outlier due to the significant fraction of people employed in the extractive industry.

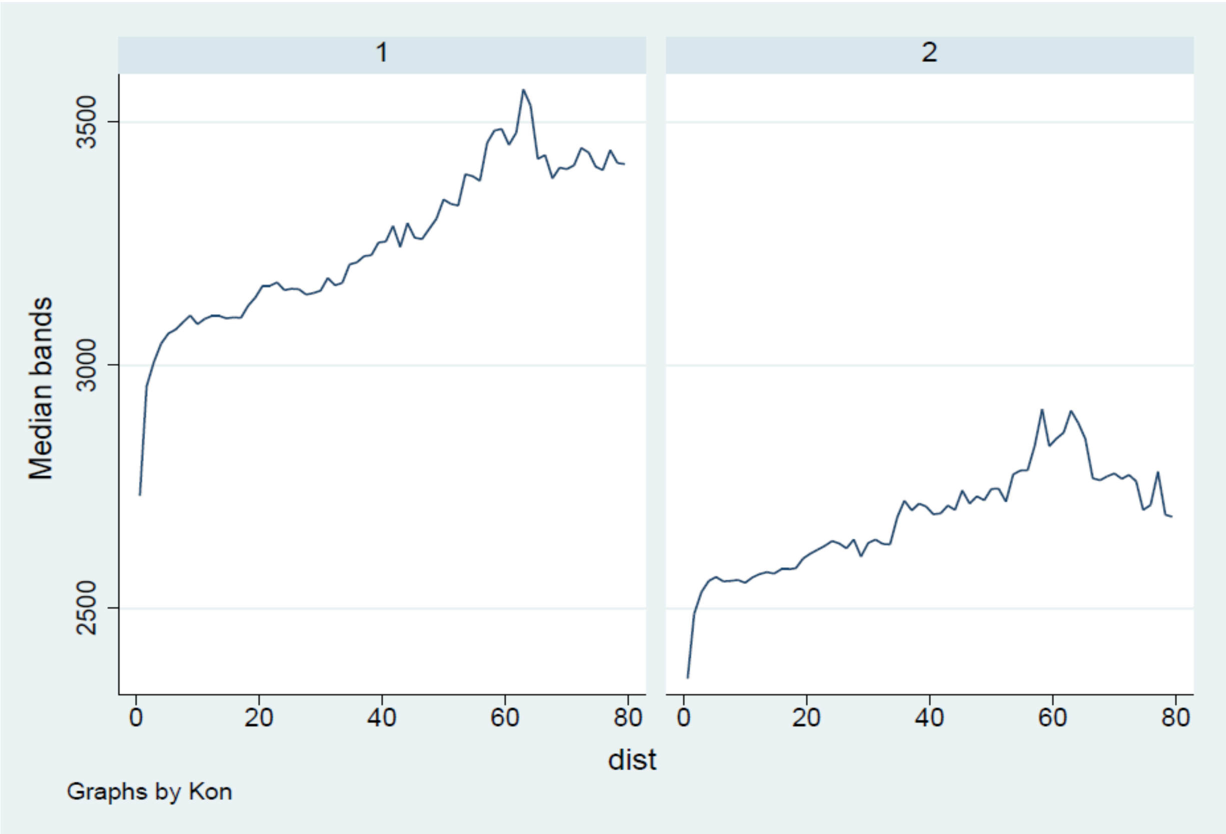
Figure 2. Average commuting distances by municipality in Sweden. Male and female samples



Distance is commuting distance in kilometers for every individual calculated using Pythagoras formula. A unique future of our dataset is the availability of geographical coordinates on individuals' place of residence and place of work on 100 meters span. Therefore, we are able to calculate a very good proxy of the daily commuting distance of individuals. The main assumption behind this variable is that the individuals commute on a daily basis. Moreover individuals' place of residence should be geographically separated from their place of work. Therefore, individuals who work at home are excluded from our analysis.

The variable *Squared Distance* specifies the square of commuting distance covered by individuals on a daily basis. The inclusion of the squared form of commuting distance is justified by the possible non-linearity in the relationship between earnings and distance as it is suggested by both theory (Lundholm, 2010) and descriptive statistics. In particular Figure 3 represents median-band plot of annual income against commuting distances. The possible concavity can be traced from the graph especially for the female sample.

Figure 3. Cross-plot of earning against commuting distance



Since the return for commuting distance in the fixed effect setting is identified only for those individuals who changed the commuting distance over the length of study period, the variation in commuting distance will be discussed in more details. As unemployed people are omitted from the analysis, the main source of variation in commuting distance is explained by the on-the-job mobility. There are two types of changes in the labor market status: voluntary and involuntary. The voluntary change of the labor market status is associated with future increases in return to experience; the involuntary change is related to decrease of earnings or job loss. In both cases a significant role is played by individual time invariant heterogeneity (e.g., mathematical competences, manual skills, etc.). The unobserved ability traits can affect spatial labor mobility in both directions: on one hand individuals with higher ability may have faster growing career associated with spatial mobility; one the other hand workers with lower ability can be affected by the hidden or explicit unemployment. The elimination of ability bias may be addressed through application of the fixed effects estimation approach (Manning, 2013; Winter-Ebmer and Zweimuller, 1999; Adowd et al., 1999; Winter-Ebmer, 1996) .

In order to capture the age profile of commuting, a set of dummy variables for intervals was introduced into the model. The dummies for age intervals allow capturing the concavity of the age profile as a proxy for experience. The age dummy variables *Age between 20 and 25, Age between 25 and 30, Age between 30 and 35, Age between 35 and 40, Age between 40 and 45, Age between 45 and 50, Age between 50 and 55, Age between 55 and 60* are introduced in the model as a proxy for individual experience. The reference category was selected to be a individuals in the age between 20 and 25 years (*Age between 20 and 25*).

The variables *Married, Single mother/father, Living with parents* and *Single* control for the marital status of individuals. The variable *Married* is a dummy variable that denotes the presence of wife or husband or formally recognized partner in civil cohabitation i.e "sambo"[10]. The *Living with parents* is dummy controlling single individuals who lives with parents. *Single mother/father* is a dummy variable for being a single mother or a single father. *Single* indicates single individuals (the reference category). Previous studies suggest that the presence of a partner and/or children should have a significant impact on the decision to commute. It is explained by individuals accepting job offers at the household level rather that individually (i.e., collective household model). Moreover, this effect might alter the commuting distance for females due to additional tasks in household production.

[10] Groups with characteristics attributed to the "modern lifestyle" such as late marriage or high labour market mobility.

The variables *Pre-gymnasium education, Gymnasium level of education, Post- gymnasium level of education <2 years, Post-gymnasium level of education >2 years* and *University level of education* are dummy variables that specify the educational attainments of an individual. The lowest level is *Pre-gymnasium education* which corresponds to the pre-gymnasium level of education, whereas the highest one −*University level of education* corresponds to the possession of the PhD or Licentiate degree.[11] The reference category was selected to be *Pre-gymnasium education* which is equivalent to completion of basic level of education school. The previous studies suggest that education has a significantly positive impact both on earnings and mobility.

The set of controls for the sectors of employment are *Manufacture, Construction, Retailing, Private services* and *Public services.* The *Public services* variable was selected as the reference group due to the highest spatial dispersion of this sector.

Apart from that, a set of local labor market characteristics was included in the estimation such as: logarithm of the median of wages prevailing at the local labor markets (*Log of median of wage in the region of residence),* the unemployment rate prevailing in the labor market (*Unemployment rate)*, and the number of employed people at the local labor market (*Size of the labor force).* A set of year dummies was included into the analysis in order to capture the effect of business cycles over the estimated period.[12]

[11] In Sweden, Licentiate degree is a pre-doctoral degree which requires completion of all courses and academic research equivalent to half dissertation.

[12] Apart from that, a specification also including a set of labor market dummies was also estimated. We do not report these results since the introduction of labor market fixed effects did not produce any relevant change in the coefficients on commuting distances (linear and squared).

2.5 Data description and sample selection

This section contains a data description together with the criteria and the motivation for the selection of the particular estimation sample.

2.5.1 Description of the data

The data used in the analysis is collected from the administrative registers of Labor Market Board (HANDEL) and Statistics Sweden (LOUISE). LOUISE provides information about family conditions, presence of children, education, employment status, sector of employment, branch of employment together with geographical coordinates for the place of residence and the place of work. Data from Labor Market Board reflects information for income from employment and non-employment activities. The merged dataset is a longitudinal geocoded panel that contains information about all individuals in the age range 20-64 living in Sweden for the time period 2003—2009. The analysis is carried out at the individual level. The presence of exact coordinates of the places of work and residence in the UTM (United Transverses Mercator) system allows defining the commuting distance using the Pythagoras formula. The advantage of this system lies in the simplification of the calculation of the commuting distance. On the other side, it is the shortest geographical distance between place of work and residence without taking into account the nonlinearity in the construction of the road system. Nevertheless, the distance calculated in this fashion can be considered as a good proxy for actual distance of commuting and, therefore, commuting expenses. The analysis focuses on the individuals who are in the age between 20 and 60 and are employed. One potential source of bias arises from the fact that individuals who experience difficulties in finding a job in the narrow labor market might shift to the extensive labor market[13] during the job-search process during the year. This leads to the systematic misreporting of annual income and subsequent underestimation of the role of the commuting distance in the wage formation due to the possibility of working less than full time,

[13] [13] Narrow labor market—the labor market with commuting distances impacting daily life of individuals but tolerable for the most of them (0-30 km). Extensive labor market—the labor market with commuting distances that entail pecuniary and non-pecuniary losses unacceptable for most individuals (30- 120 km).

and therefore the availability of additional time for commuting. To reduce this concern, individuals who possessed alternative sources of income from the welfare system such as unemployment benefits were excluded from the analysis. Together with that, individuals who carried entrepreneurship activities were not taken into account in our analysis. The motivation behind this exclusion lies in the fact that entrepreneurs have places of work rather than jobs, and income that is independent of commuting distance (van Ommeren and van der Straaten 2008). Moreover, the commuting distance was constrained to the maximum border of the extensive labor market which is 120 kilometers of one way distance as proposed by Lundholm, (2010). This distance can be approximated to 3.5 hours of commuting taking into account road complexity and traffic congestion. This sample cut allows excluding individuals who commute on a weekly basis.

The analysis was constrained to those individuals who have both coordinates of the place of residence and of the place of work. One of the main assumptions of the analysis is the observability of commuting distance. Also, individuals employed at home are very different in their characteristics. Therefore, those individuals who work at home were excluded from the analysis. Moreover, we further imposed an assumption about the existence of an economically meaningful distance for commuting i.e. individuals should incur pecuniary and/or non-pecuniary losses. That is why the individuals with commuting distances less than 500 meters one way are excluded from the sample as well. Together with the information on earnings and commuting distance the dataset contains information about the age, gender, sector of employment, education, marital status, and presence of children and characteristics of the labor market of residence such as: unemployment rate, employment and median of wage prevailing at the local labor market. The sample is split by gender. The motivation behind lies in the different commuting patterns for males and females together with the difficulties experienced by females in finding jobs, and family constrains on the long distance commuting. Although we ran the analysis by the gender, labor market variables were calculated for the whole sample (pooled genders) with the purpose of capturing mutual substitution of male and female workers in the labor market.

2.5.2 Description of variables

The variables of main interest are individual gross annual labor income *Log (Annual earnings)* and commuting distance *Distance.* Income variable *Annual earnings* was transformed into the logarithm form whereas distance variables represent linear term *Distance* and *Square distance/100* introduced in order to capture the nonlinearities in of distance profile in the model. The descriptive statistics suggests that the average earning of males is 3.69 hundreds SEK (40,079 EUR) while the annual earning of females is 3.82 hundreds SEK (33,468 EUR).

Table 1. Descriptive statistics of the male sample

Variable	Obs	Mean	Std. Dev	Min	Max
Log (Annual earnings)	*2862864*	*7.9601*	*0.3626*	*7.4237*	*9.0543*
Distance (Km)	*2862864*	*28.2963*	*35.5864*	*1.0198*	*80*
Square distance/100	*2862864*	*20.6707*	*56.8389*	*0.0104*	*576.9828*
Married	*2862864*	*0.4990*	*0.4999*	*0*	*1*
Single mother/father	*2862864*	*0.0314*	*0.1745*	*0*	*1*
Living with parents	*2862864*	*0.0468*	*0.2112*	*0*	*1*
Nationality	*2862864*	*0.0890*	*0.2848*	*0*	*1*
Gymnasium level of education	*2862864*	*0.4860*	*0.4998*	*0*	*1*
Post- gymnasium level of education <2 years	*2862864*	*0.0808*	*0.2754*	*0*	*1*
Post-gymnasium level of education >2 years	*2862864*	*0.2621*	*0.4398*	*0*	*1*
University level of education	*2862864*	*0.2053*	*0.1418*	*0*	*1*
Age between 25 and 30	*2862864*	0.0985	*0.2980*	*0*	*1*
Age between 30 and 35	*2862864*	0.1028	*0.3715*	*0*	*1*
Age between 35 and 40	*2862864*	0.1428	*0.3499*	*0*	*1*
Age between 40 and 45	*2862864*	0.1723	*0.3777*	*0*	*1*
Age between 45 and 50	*2862864*	0.1874	*0.3902*	*0*	*1*
Age between 50 and 55	*2862864*	0.1893	*0.3917*	*0*	*1*
Manufacture	*2862864*	*0.2788*	*0.4484*	*0*	*1*
Construction	*2862864*	*0.1054*	*0.3071*	*0*	*1*
Retailing	*2862864*	*0.2012*	*0.4009*	*0*	*1*
Private services	*2862864*	*0.0989*	*0.2986*	*0*	*1*
Log of median of wage in the region of residence	*2862864*	*7.4763*	*0.3540*	*5.3602*	*7.9412*
Unemployment rate 0	*2862864*	*0.1515*	*.0520*	*0.02*	*0.4952*
Size of the labor force	*2862864*	*307947.4*	*297732.9*	*790*	*785363*

The main independent variable distance was transformed into kilometers. The descriptive statistics suggest that the average commuting distance for the commuters within the labor market is 28 kilometers for males and 22 kilometers for females. Together with that, 50 % of the male population commutes within 6.53 kilometers and 50 % of females for 4.85 kilometers. These facts are consistent with the previous studies on labor mobility in Sweden (Lundholm, 2010). This result suggests that males and females are employed in positions that require different levels of commuting. Alternatively, assuming different spatial dispersion of industries across the city, it can be an evidence of the self-selection or gender segregation by sector. The higher percentage of males employed in manufacturing supports this hypothesis.

The return to commuting is identifiable in the fixed effect setting only for those individuals who have changed the distance to workplace. Therefore, Table 3 and Table 4 report the numbers of individuals who experienced changes in commuting distance over the study period, divided by gender.

Table 3. Number of movers by the type of exit. Male sample

Type of exit	*Stayers*	*Movers*
Individuals with no change in employer or workplace	*406,290*	*362,194*
Individuals who changed workplace (within same employer)	*167,928*	*92,856*
Individuals who changed employers and workplace	*357,159*	*542,851*

Table 3 and Table 4 describe a number of "movers"[14], i.e. the number of individuals who have changed either a place of work or place of residence during study period. The figures in Tables 3 and 4 suggest that individuals with stable place of work and residence represent about 20% of the overall size of the panel. At the same time we observe significant number of people who have changed workplace within the same employer, changed employer and workplace and changed a place of residence. Therefore, the remaining number of switchers is sufficient to identify the returns to commuting distance on earnings in the fixed effect setting.

[14] An individual is considered as a "mover" if he/she at least one changed the place of work or residence. Workers employed at firms which were subjected to merges or change of owners are considered to be stayers if they had not moved to another place of work.

The results from Table 3 and 4 indicate the higher number of residential movers among males. However, women relatively more often experience a move to a different workplace within the same employer, which can be explained by the higher workplace attachment among females.

Table 4. Number of switchers by the type of exit. Female sample

Type of exit	*Stayers*	*Movers*
Individuals with no change in employer or workplace	*443,372*	*325,546*
Individuals who changed workplace (within same employer)	*198,452*	*103,024*
Individuals who changed employers and workplace	*302,935*	*419,602*

Other facts which can be observed from the descriptive statistics are that on average there is a higher fraction of females with high education, but a higher percentage of males with PhD or a Licentiate degree. The comparison suggests that the fraction of single parents is higher for females. The main results from the descriptive statistics for in the male sample are presented in Table 2 and for the female sample in Table 5.

Table 5. Descriptive statistics of the female sample

Variable	*Obs*	*Mean*	*Std. Dev*	*Min*	*Max*
Log (Annual earnings)	*2008503*	*7.8524*	*0.3279*	*7.4237*	*9.0543*
Distance	*2008503*	*22.0928*	*28.9074*	*1.0198*	*239.9921*
Square distance/100	*2008503*	*13.2373*	*41.8717*	*0.0104*	*575.962*
Married	*2008503*	*0.5338*	*0.4988*	*0*	*1*
Single mother/father	*2008503*	*0.0999*	*0.2928*	*0*	*1*
Living with parents	*2008503*	*0.0181*	*0.1333*	*0*	*1*
Nationality	*2008503*	*0.1038*	*0.3050*	*0*	*1*
Gymnasium level of education	*2008503*	*04087*	*0.4916*	*0*	*1*
Post- gymnasium level of education <2 years	*2008503*	*0.0438*	*0.2048*	*0*	*1*
Post-gymnasium level of education >2 years	*2008503*	*0.4495*	*0.4974*	*0*	*1*

Table5. Continued

Variable	*Obs*	*Mean*	*Std. Dev*	*Min*	*Max*
University level of education	*2008503*	*0.0133*	*0.1145*	*0*	*1*
Age between 25 and 30	*2008503*	*0.0677*	*0.2513*	*0*	*1*
Age between 30 and 35	*2008503*	*0.0744*	*0.2625*	*0*	*1*
Age between 35 and 40	*2008503*	*0.1401*	*0.3470*	*0*	*1*
Age between 40 and 45	*2008503*	*0.1985*	*0.3988*	*0*	*1*
Age between 45 and 50	*2008503*	*0.2216*	*0.4153*	*0*	*1*
Age between 50 and 55	*2008503*	*0.2147*	*0.4106*	*0*	*1*
Manufacture	*2008503*	*0.1038*	*0.3550*	*0*	*1*
Construction	*2008503*	*0.0156*	*0.1242*	*0*	*1*
Retailing	*2008503*	*0.1318*	*0.3383*	*0*	*1*
Private services	*2008503*	*0.0.775*	*0.2673*	*0*	*1*
Log of median of wage in the region of residence	*2008503*	*7.5037*	*0.3418*	*5.3690*	*7.9412*
Unemployment rate 0	*2008503*	*0.1534*	*0.0511*	*0.02*	*0.4955*
Size of the labor force	*2008503*	*335310.1*	*304788.1*	*790*	*785363*

2.5 Analysis of the male sample

The estimated results from the male sample are presented in Table 6. Our estimates document the concavity of the earnings profile with respect to commuting distance. Individuals receive an increasing return from commuting at a decreasing rate. The coefficient of the commuting distance derived with OLS is 0.000566 whereas the estimated coefficients of the squared term of the commuting distance divided by 100 is -0.0000129. It is worth mentioning that the squared term is insignificant.

Back of the envelope calculations, assuming one hour of commuting time to be approximately equal to 35 kilometers, and a hourly wage of 199 SEK (22.11 EUR) suggest that individuals receive a compensation of 31 SEK (3.44 EUR) per one hour of daily commuting,[15] which constitutes 16 % of hourly wage.

The fixed effect (FE, hereafter) estimation suggests somewhat lower point estimates of distance (0.000305). At the same time, the coefficient of the quadratic term is much higher and more significant (-0.000156). It implies that individuals receive 17SEK (2 EUR) of reward for one hour of daily commuting. It is approximately 8.5% of the individual’s hourly wage.[16]

The age dummies in OLS and FE significantly affect earnings with a clear evidence of concavity with a turning point between 40 and 45 years. Earnings increase for the age groups up to 45 years, and show a decrease for older age categories.

Individuals experience 8.3% increase in their earnings if they are married when using OLS, and 1.6% increase using the FE estimation compared to the baseline category of single individuals. Single fathers increases earnings by 4% with OLS and 0.3% with FE. One of the explanations of this phenomenon is a redistribution of family duties and economies of scale in household production that affect labor supply. Males living with their parents experience a wage penalty of 4.5% with OLS, and 3.1% with FE estimation likely due to the presence of unobserved attributes that affects the choice of living with parents and wage bargaining.

[15] [15] This proxy does not include the congestion effect. Moreover, it might significantly vary due to the differences in the place of location, development of local infrastructure and selection of transport mode.
[16] [16] These results are likely to underestimate the real amount due to the business travels and absenteeism

Table 6. Estimation results using OLS and FE. Male sample

Variables	*Male sample OLS estimation*		*Male sample Fixed effect estimation*	
	coefficient	*t-values*	*coefficient*	*t-values*
Distance variables				
Distance	*0.000566******	*29.53*	*0.000305******	*9.12*
Square distance/100	*-0.0000129*	*-0.77*	*-0.000156******	*-5.80*
Age variables				
Age between 25 and 30	*-0.0186******	*-17.58*	*0.10599******	*82.15*
Age between 30 and 35	*0.0126******	*11.65*	*0.1854******	*106.42*
Age between 35 and 40	*0.0525******	*49.08*	*0.2149******	*98.15*
Age between 40 and 45	*0.0717******	*68.79*	*0.2184******	*86.04*
Age between 45 and 50	*0.0736******	*71.35*	*0.2003******	*70.18*
Age between 50 and 55	*0.0727******	*70.57*	*0.1756******	*55.57*
Age between 55 and 60	*0.0767******	*74.21*	*0.1468******	*42.33*
Family status variables				
Married	*0.0838******	*178.98*	*0.0165******	*16.69*
Single mother/father	*0.0461******	*40.80*	*0.00363**	*2.45*
Living with parents	*-0.0457******	*-41.67*	*-0.0318******	*-16.99*
Education level				
Gymnasium level of education	*0.0654******	*122.13*	*0.00293*	*0.39*
Post- gymnasium level of education <2 years	*0.218******	*253.34*	*0.0676******	*8.09*
Post-gymnasium level of education >2 years	*0.263******	*378.41*	*0.145******	*19.04*
University level of education	*0.423******	*237.20*	*0.221******	*25.65*
Nationality	*-0.0635******	*-86.30*	.	.
Sector of employment				
Manufacture	*0.0705******	*127.50*	*-0.00794******	*-5.50*
Construction	*0.0546******	*76.49*	*0.0217******	*11.63*
Retailing	*0.0480******	*75.96*	*0.000801*	*0.56*
Private services	*0.131******	*163.61*	*0.00371**	*2.55*
Macroeconomic variables				
Log of median of wage in the region of residence	*0.0702******	*11.55*	*0.0528******	*9.66*
Unemployment rate	*0.0595******	*8.97*	*0.0107*	*-0.660*
Size of the labor force	*0.0000001******	*138.95*	*3.94e-08******	*9.77*
Time period dummies	*Yes*		*Yes*	
Constant	*8.242******	*177.06*	*8.337******	*197.32*
Adjusted R^2	*0.272*		*0.219*	
Number of observations	*2445423*		*2445423*	

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the log of annual earnings.

The standard errors are heteroskedasticity robust and clustered at the individual level

Education clearly demonstrates increasing returns to the level of education. In this case the reference category is selected to be education below the gymnasium level. OLS estimation shows that individuals with gymnasium have 6.5% higher earnings while FE estimation suggests that earnings increase only for 0.29%. Possession of less than 2 years of the post-gymnasium level of education leads to an earnings increase of 21.8% applying OLS. FE reports 6.7% increase in earnings for this category of individuals. More than 2 years of post-gymnasium education increases earnings by 26.3 % with OLS by 14.5% with FE estimation. Finally, OLS estimation shows that Master degrees or PhD increase the earnings by 42% whereas FE shows a 22.1% increase. Individuals of foreign origin receive on average 6% lower earnings than natives.

Sector of employment significantly influences individuals' earnings. In this setting, individuals employed in the public sector such as health care, defense, public order and social administration were selected to be the baseline category. OLS estimation suggests that individuals employed in manufacturing receive 7% higher wages than in public services. On the contrary, results obtained by using FE shows that individuals employed in manufacturing experience a penalty of 0.7%. This shows that importance of controlling for individual heterogeneity, and in particular that individuals may self-select into sector of employment. The explanation of these results might lie in the fact that in manufacturing the percentage of workers covered by collective agreements is less than the corresponding percentage in public services. Individuals employed in construction receive 5.4% higher annual income according to the results from OLS estimation and 2.1% higher with FE. Such an evident increase can be explained by the high riskiness of the job, and by the working conditions of workers employed in construction. Employment in the retail sector is associated with a 4.8% increase using OLS and 0.08% increase using FE. OLS suggest that individuals employed in the private service sector receive 13.1% higher annual earnings than individuals in the reference category. FE estimation shows only a 0.3% increase in annual earnings for this category of workers. In general, it should be noted that these results are consistent with a quite compressed earnings distribution across economic sectors. Macroeconomic variables behave in the expected fashion. The logarithm of median wage in the labor market of work positively and significantly affects earnings.

Curiously enough, we found that the unemployment rate is positively correlated with wage level, which contradicts the existence of a wage curve in Sweden. A possible explanation is that high unemployment benefits and immigration owing to the attractiveness (amenities) of a certain destination together with the level of wages established during bargaining between cartels of employers and labor unions generate such a result. In the same time, the size of the labor market has a positive impact on earnings [17]

Generally speaking, it is possible to conclude when individual heterogeneity is not controlled for, the magnitude of most coefficients, and their significance levels, are severely inflated.

[17] The set of regional fixed effects was included in order to test for labor market heterogeneity. Results indicate that the introduction of Labor market fixed effects does not change significantly the coefficients of interest. Therefore, we do not report them in the main text.

2.6 Analysis of the female sample

The outcome from the female sample with the OLS and FE estimation methods are presented in Table 7. The magnitude and significance levels of the coefficients of the distance variables support the concavity assumption of the commuting distance profile in the earnings equation. The coefficient of the linear term of commuting distance is 0.000684 and the quadratic term is -0.000076 using the OLS, and 0.000301 and -0.000217 using FE estimation. These results suggest that female individuals, with an average hourly wage of 174 SEK (19.33 EUR), receive a financial compensation for 1 hour of commuting equal to 33SEK (3.66EUR) which constitutes 17% of their hourly wage when using OLS, and 14 SEK (1.5 EUR) which is 8.4% according to FE estimation procedure.

The OLS estimated coefficients on the age dummies suggest that earnings steadily increase with the age up to 55 years with a subsequent fall. The results from the FE estimation suggest that the turning point occurs somewhat later than for males: approximately in the 40-45 age group.

Marital status significantly affects the wage of female workers. The results from OLS estimation suggest that married women earn 0.6% less than single women, whereas FE results indicated that married women have a 0.9 % higher earnings that single ones. These results can be also explained by the redistribution of the tasks in household production. Single mother show an annual earning premium of 0.5% according to OLS, and a 0.7% premium according to FE estimation. Individuals living with their parents experience a significant penalty which is 3% in OLS and almost 4% in FE estimation. Education plays an important role in the wage formation of female individuals. As before, the reference category was selected to be females with education lower than gymnasium level. The result from OLS estimation indicate that the individuals with gymnasium level of education receives 4% higher earnings than the baseline category. FE suggests that the increase is around 6%. Post-gymnasium level of education shorter than 2 years increases earnings by 17.9% with OLS, and 13.3% with FE estimation. Individuals with the post-gymnasium level of education receive on average 21% more than the reference category. A university degree accounts for a 42.2% earnings' increase according to OLS estimation, and 30.6% increase according to FE estimation.

Table 7. Estimation results using OLS and FE. Female sample

Variables	*Female sample OLS estimation*		*Female sample Fixed effect estimation*	
	coefficient	*t-value*	*coefficient*	*t-value*
Distance variables				
Distance	*0.000684****	*30.22*	*0.000301****	*7.80*
Square distance/100	*-0.000076****	*-3.46*	*-0.00021****	*-6.69*
Age variables				
Age between 25 and 30	*0.0185****	*11.47*	*0.1285****	*65.30*
Age between 30 and 35	*0.0565****	*34.61*	*0.2001****	*81.23*
Age between 35 and 40	*0.0901****	*56.54*	*0.2348****	*77.79*
Age between 40 and 45	*0.0947****	*62.72*	*0.2443****	*72.22*
Age between 45 and 50	*0.0936****	*63.38*	*0.2367****	*64.25*
Age between 50 and 55	*0.0972****	*66.42*	*0.2233****	*56.14*
Age between 55 and 60	*0.0986*	*67.64*	*0.2078****	*48.49*
Family status variables				
Married	*-0.00646****	*-12.21*	*0.00901****	*7.70*
Single mother/father	*0.00598****	*7.60*	*0.00784****	*6.43*
Living with parents	*-0.0321****	*-15.64*	*-0.0379****	*-12.56*
Education level				
Gymnasium level of education	*0.0433****	*58.26*	*0.0641****	*8.03*
Post- gymnasium level of education <2 years	*0.179****	*137.12*	*0.133****	*15.33*
Post-gymnasium level of education >2 years	*0.210****	*262.03*	*0.213****	*24.66*
University level of education	*0.422****	*158.96*	*0.306****	*31.12*
Nationality	*-0.0177****	*-23.80*	.	.
Sector of employment				
Manufacture	*0.109****	*145.54*	*-0.0146****	*-7.20*
Construction	*0.0969****	*54.26*	*0.0179****	*5.14*
Retailing	*0.0604****	*84.86*	*-0.0136****	*-7.64*
Private services	*0.138****	*151.16*	*-0.00211*	*-1.16*
Log of median of wage in the region of work	*0.142****	*21.70*	*0.165****	*26.03*
Unemployment rate	*0.1091****	*15,00*	*0.0713****	*3,87*
Size of the labor force	*0.00000018****	*133.38*	*4.36e-08****	*9.87*
Time period dummies	*Yes*		*Yes*	
Constant	*8.669****	*172.34*	*8.965****	*182.55*
Adjusted R^2	*0.269*		*0.192*	
Number of observations	*1761593*		*1761593*	

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the log of annual earning.
The standard errors are heteroskedastisity robust and clustered at the individual level

As it was stated before, the distribution of individuals' earnings is quite compressed across sectors of employment in Sweden. Moreover, the earnings in different sectors are seriously affected by individual's heterogeneity. OLS reports the increase in earnings to be 11% in manufacturing whereas FE shows earnings decrease by 1.4%. Construction shows an increase in earnings of 9.6% in case of OLS estimation and 1.7% in case of FE estimation. Individuals employed in retailing earn 6% more when using OLS, but according to the FE estimates these individuals experience a penalty of 1.3%. Working in private services leads to an increase in earnings of 13.8% derived using OLS, and a decrease in earnings of 0.2% with FE. One of the explanations of the clear penalty in employment in sectors other than the public sector is a higher protection against gender discrimination in the latter.

Macroeconomic variables behave in the expected manner. Logarithm of median wage prevailing in the local labor market affects positively wages using both OLS and FE estimation, as do unemployment rate and size of the labor market.

2.7 Estimation of the model using firm level fixed effect

We have also estimated the earning model including firm level fixed effects in order to verify the existence of return to commuting for different commuting distances while controlling form firm-specific heterogeneity.

Table 8. Estimation results using firm level FE. Male and female sample.

Variables	*Male sample Fixed effect estimation*		*Female sample Fixed effect estimation*	
	coefficient	*t-values*	*coefficient*	*t-values*
Distance variables				
Distance	*0.00036****	*27.69*	*0.00033****	*20.54*
Square distance/100	*0.000031****	*3.67*	*0.00007****	*6.79*
Age variables				
Age between 20 and 25	*-0.031002****	*-30.13*	*-0.00092*	*-0.63*
Age between 25 and 30	*0.00223**	*-2.09*	*0.03438****	*22.46*
Age between 30 and 35	*0.04649****	*43.56*	*0.07629****	*50.12*
Age between 35 and 40	*0.07051****	*67.72*	*0.08726****	*59.99*
Age between 40 and 45	*0.07736****	*74.67*	*0.09222****	*64.30*
Age between 45 and 50	*0.08097****	*77.99*	*0.09874****	*69.19*
Age between 50 and 55	*0.08713****	*83.45*	*0.10206****	*71.50*
Age between 55 and 60	*0.08015****	*69.27*	*0.09621****	*63.47*
Family status variables				
Married	*0.076****	*174.74*	*-0.00202****	*-4.10*
Single mother/father	*0.04468****	*42.08*	*0.01178****	*15.64*
Living with parents	*-0.03618****	*-38.41*	*-0.02335****	*-14.34*
Education level				
Gymnasium level of education	*0.05503****	*98.91*	*0.03612****	*0.00079*
Post- gymnasium level of education <2 years	*0.16741****	*202.62*	*0.15574****	*127.13*
Post-gymnasium level of education >2 years	*0.23229****	*336.76*	*0.22154****	*265.93*
University level of education	*0.45256****	*285.71*	*0.47628****	*263.10*
Nationality	*-0.04545****	*-68.58*	*.-0.00886*	*-13.07.*
Sector of employment				
Manufacture	*-0.02986****	*-12.20*	*-0.02311****	*-6.18*
Construction	*0.03121****	*16.78*	*0.02202****	*5.71*
Retailing	*0.00614***	*2.65*	*0.00379*	*1.25*
Private services	*0.01087*	*4.97*	*0.01377****	*4.89*
Macroeconomic variables				
Log of median of wage in the region of residence	*0.20477****	*35.86*	*0.26985****	*43.19*
Unemployment rate	*-0.17902****	*-23.55*	*-0.00968***	*-1.11*
Size of the labor force	*7.66e-07****	*107.89*	*1.79e-07****	*98.17*
Constant	*9.39906****	*214.64*	*9.6939****	*202.38*

Table 8. Continued.

Firm fixed effects	*Yes*	*Yes*
Time fixed effects	*Yes*	*Yes*
Adjusted R^2	*0.2700*	*0.2518*
Number of observations	*2720145*	*1905038*

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the log of annual earnings.
The standard errors are heteroskedasticity robust and clustered at the individual level

The results in Table 8 indicate that the return to commuting is similar to the one calculated in the presence of individual heterogeneity. β-coefficient for the linear term of commuting distance is 0.00036 and quadratic term is 0.000013 in the male sample. In the same time, the result from the estimation of the female sample β-coefficient for the linear term is 0.00033 while the coefficient on the quadratic term is 0.00007. Therefore, the return to commuting increases with the distance at an increasing rate, with the increasing rate being higher for the female sample. Again, assuming that males receive an hourly wage of 199 SEK (22 EUR) and females 174 SEK (19.33 EUR), back of envelope calculations suggest that males gain 18.54 SEK (1.95 EUR) for one hour of commuting (35 kilometers) which constitutes around 9.34% of the hourly wage while females receives 17.53 SEK (1.83 EUR) for one hour of commuting which is around 10.8% of the hourly wage. Therefore we can conclude that *within-firms* individuals with higher commuting distance tend also to have higher wage. This fact can be seen as an explicit evidence of presence of individual's bargaining power.

2.8 Estimation of the wage growth model

One possibility is that individuals might accept lower current earnings over greater commuting distances if they expect a faster increase in their earnings over subsequent years. To analyze this possibility, we have estimated a model where the dependent variable is the difference in the earnings of individuals between 2003 and 2010. The analysis was focused on individuals who neither changed the place of residence nor the place of work during the time window considered (and of course does not include individual FEs) The results on the estimation for the male and female samples are shown in Table 9.

Table 9. Results on estimation of difference in log earnings using OLS in 2003-2010

Variable	*Male sample*		*Female sample*	
	B coefficient	*t-value*	*B coefficient*	*t-value*
Distance in 2003	*0.0000759**	*2.52*	*0.0000326*	*0.72*
Δ marriage 2010-2003	*0.00924****	*4.91*	*-0.00666****	*-3.56*
Δ cohabitation 2010-2003	*0.0332****	*9.21*	*0.0228****	*5.16*
Difference single parents 2010-2003	*0.0311****	*15.08*	*0.0638****	*25.57*
Δ education gymnasium 2010-2003	*0.313****	*15.98*	*0.161****	*9.07*
Δ education <2 years 2010-2003	*0.235****	*9.01*	*0.147****	*6.58*
Δ education >2 years 2010-2003	*0.265****	*11.01*	*0.249****	*10.32*
Δ university 2010-2003	*0.379****	*13.10*	*0.399****	*12.98*
Δ age 2010-2003	*0.0934****	*33.46*	*0.0918****	*29.95*
Δ squared age 2010-2003	*-0.000734****	*-93.96*	*-0.000633****	*-6373*
Δ median wage 2010-2003	*0.119**	*1.96*	*0.0772*	*1.19*
Δ employment rate 2010-2003	*-0.195**	*-1.93*	*0.122*	*1.08*
Δ size of labor market 2010-2003	*0.0000002**	*2.22*	*-0.0000002**	*-2.42*
Adjusted R^2	*0.247*		*0.0213*	
Number of observations	*152522*		*146552*	

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the difference between logs of earnings in the initial period (2003) and final period (2010).

The standard errors are heteroskedasticity robust and clustered at the individual level

The β-coefficient on earnings 0.0000759 indicates that the males on average experience higher earnings growth rate associated with commuting. It suggests that individuals who commute 60 km per day for round way distance would experience a 0. 4% higher earnings growth rate over time. The point estimate for females is 0.0000326, but is not statistically significant.

The results in this Section show that commuting on top of ensuring higher earnings immediately also contributes to a faster earnings' increase in the medium run, at least for men.

2.9 Heterogeneity in the return to commuting

To check the robustness of the obtained coefficients, a set of additional sample restrictions were tested. Previous studies suggest that age is positively correlated with mobility patterns, with the peak of mobility being between 20 and 25 years. Moreover, individuals in this age group receive lower income due to the lack of experience and low social capital. Therefore, we expected to observe a higher commuting premium for individuals which are above the peak of the commuting age threshold. Apart from that, many urban economics studies such as Fu and Ross (2010), DiAddario and Potacchini (2008) or Fu (2007) suggest that enterprises located in larger agglomerations offer higher wages or give higher bargaining power to their employees due to higher productivity. The results are presented in the Table 10

Table 10. Estimation results using FE. Alternative male sample specifications

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample including individuals with commuting distances >240 km	*2485651*	*0.000233**** *(9.80)*	*-0.0000828**** *(-6.10)*
Sample including commuting distances>240 older than 25 years	*2398402*	*0.000315**** *(9.29)*	*-0.000162**** *(-5.93)*
Sample including individuals with commuting distance >240 km who does work in agglomeration	*1198665*	*0.000317**** *(10.32)*	*-0.000106**** *(-6.41)*
Sample including individuals with commuting distance >160 km who does work in agglomeration	*1168319*	*0.000377**** *(8.48)*	*-0.000166*** *(-4.67)*
Sample including individuals with commuting distance >160 km who does work in agglomeration and older than 25 years	*114618*	*0.000379**** *(8.42)*	*-0.000164**** *(-4.55)*

t-values are in parenthesis
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note: The dependent variable is the logarithm of annual earning of individual.
The standard errors are heteroskedasticity robust and clustered at the individual level

The additional sample restrictions were tested for the female sample as well. The results from the estimation are presented in Table 11. Results clearly support the thesis we just postulated. The return to commuting is higher for individuals older than 25 years. Moreover, individuals who work in urban agglomerations experience a higher return to commuting, likely due to agglomeration effects on productivity which affects bargaining power of individuals.

Table 11. Estimation results using FE. Alternative female sample specifications

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample including individuals with commuting distances >240 km	*1775856*	*0.000142*** (5.02)*	*-0.0000576*** (-3.35)*
Sample including commuting distances>240 older than 25 years	*1738885*	*0.000317*** (8.14)*	*-0.000227*** (-6.93)*
Sample including individuals with commuting distance >240 km who does work in agglomeration	*793418*	*0.000160*** (4.19)*	*-0.0000459*** (2.07)*
Sample including individuals with commuting distance >160 km who does work in agglomeration	*783525*	*0.000210*** (3.90)*	*-0.0000952* (-2.11)*
Sample including individuals with commuting distance >160 km who does work in agglomeration and older than 25 years	*775027*	*0.000212*** (3.92)*	*-0.0000958* (-2.11)*

Notes: The dependent variable is the logarithm of annual earning of individual; t-values are in parenthesis
Note: Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
The standard errors are heteroskedasticity robust and clustered at the individual level

The results are similar for the male and female samples. Although the magnitude of the return is much lower for females. Older individuals experience higher returns to commuting. Moreover, female individuals who work in urban agglomerations experience a higher return to commuting due to agglomeration effects on productivity or better developed infrastructures.

Also, empirical evidences, presented by Mulalic et al. (2014) and Rupert et al. (2009) suggest that individuals might have different return to commuting with respect to the commuting distance, position in the earning distribution or the level of attained education. In order to evaluate the difference in the return to commuting for individuals that originate from different quartiles of the earnings distribution we estimated individual fixed effect model for subsamples of individuals that belong to the different quartiles of earning distribution. The results of estimation of the male sample are presented in the Table 12.

Table 12. Estimated return to commuting for different quartiles of earning distribution. Male sample.

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample of individuals from first quartile of income distribution	*405326*	*-0.0000611* *(-0.41)*	*-0.0000387* *(-0.16)*
Sample of individuals from second quartile of income distribution	*476340*	*-0.000111** *(-2.02)*	*0.0000379* *-0.41*
Sample of individuals from third quartile of income distribution	*524104*	*-0.0000818** *(-2.02)*	*0.000179*** *-2.68*
Sample of individuals from first quartile of income distribution	*564108*	*0.000186*** *-2.73*	*-0.000038* *(-0.36)*

t-values are in parenthesis
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note: The dependent variable is the logarithm of annual earning of individual.
The standard errors are heteroskedasticity robust and clustered at the individual level

The results of estimation of the female part of sample are presented in Table 13.

Table 13. Estimated return to commuting for different quartiles of earning distribution. Female sample.

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample of individuals from first quartile of income distribution	*461938*	*-0.00000625* *(-0.05)*	*-0.0000585* *(-0.26)*
Sample of individuals from second quartile of income distribution	*565835*	*-0.000223*** *(-2.83)*	*0.000325*** *-2.87*
Sample of individuals from third quartile of income distribution	*634859*	*-0.000107** *(-2.51)*	*0.000132* *-1.76*
Sample of individuals from fourth quartile of income distribution	*696879*	*0.000278**** *-4.92*	*-0.000190** *(-2.02)*

t-values are in parenthesis
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note: The dependent variable is the logarithm of annual earning of individual.
The standard errors are heteroskedasticity robust and clustered at the individual level

Overall, these results indicate that individuals from fourth quartile of earnings distribution (i.e. with highest earnings) receive the highest return to commuting. Moreover, Mulalic et al. (2014) suggests that the return to commuting is higher for long distance commuters due to the tax refund schemes. We test this hypothesis by estimating the individual fixed effect model on the subsamples of individuals in different quartiles of the commuting distance distribution. The results from the estimation are presented in the Table 14 for the male sample and in the Table 15 for female sample.

Table 14. Estimated return to commuting for different quartiles of commuting distance distribution. Male sample

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample of individuals from first quartile of distance distribution	*488621*	*0.000612* *-0.09*	*0.000816* *-0.39*
Sample of individuals from second quartile of distance distribution	*503233*	*-0.00309* *(-0.67)*	*0.000364* *-0.8*
Sample of individuals from third quartile of distance distribution	*504956*	*0.00285*** *-2.87*	*-0.000109* *(-1.08)*
Sample of individuals from first quartile of distance distribution	*473068*	*-0.000513* *(-1.54)*	*0.00000512* *-1.37*

t-values are in parenthesis
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note: The dependent variable is the logarithm of annual earning of individual.
The standard errors are heteroskedasticity robust and clustered at the individual level

The results from the estimation of the female sample are presented in the Table 15. They indicate that the highest significant return to commuting experience male commuters with one way commuting distance in interval between 7.57 and 18.13 kilometers.

Table 15. Estimated return to commuting for different quartiles of commuting distance distribution. Female sample

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample of individuals from first quartile of distance distribution	*748526*	*0.00452* *-0.85*	*-0.000572* *(-0.35)*
Sample of individuals from second quartile of distance distribution	*626727*	*0.0121*** *-2.82*	*-0.00118*** *(-2.78)*
Sample of individuals from third quartile of distance distribution	*580629*	*-0.00169* *(-0.67)*	*0.000084* *-0.84*
Sample of individuals from first quartile of distance distribution	*403629*	*0.000133* *-0.32*	*-0.00000338* *(-0.72)*

t-values are in parenthesis
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note: The dependent variable is the logarithm of annual earning of individual.
The standard errors are heteroskedasticity robust and clustered at the individual level

The estimated coefficients from the female sample suggest that the highest return to commuting is received by individuals who have one-way commuting distance between 2.73 and 7.53 kilometers.

Table 16. Estimated return to commuting for different level of education. Male sample

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample of individuals with secondary education	*1049876*	*-0.0000488*	*0.000000844*
		(-0.57)	*-0.6*
Sample of individuals with post - secondary education ≤ 2	*134298*	*0.000853***	*-0.0000117***
		-3.28	*(-2.82)*
Sample of individuals with post - secondary education > 2	*437899*	*0.000358**	*-0.00000555**
		-2.41	*(-2.33)*
Sample of individuals with Msc or PhD degree	*17322*	*0.00250***	*-0.0000379***
		-2.7	*(-2.68)*

t-values are in parenthesis
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note: The dependent variable is the logarithm of annual earning of individual.
The standard errors are heteroskedasticity robust and clustered at the individual level

Finally, we estimated our model for different educational categories in order to observe heterogeneous return to commuting for people with different attained levels of education. The Table 16 reports the return to commuting for different educational categories in the male sample while Table 17 shows the return to commuting for the female subsample.

Table 17. Estimated return to commuting for different level of education. Female sample

Sample	*Number of observations*	*Commuting distance*	*Squared commuting distance/100*
Sample of individuals with secondary education	*1171451*	*0.0000209*	*-0.00000132*
		-0.23	*(-0.83)*
Sample of individuals with post - secondary education ≤ 2	*97537*	*0.000217*	*-0.00000556*
		-0.63	*(-0.97)*
Sample of individuals with post - secondary education > 2	*825813*	*0.000820****	*-0.0000127****
		-7.18	*(-6.67)*
Sample of individuals with Msc or PhD degree	*11828*	*0.00299**	*-0.0000394**
		-2.3	*(-1.98)*

t-values are in parenthesis
Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note: The dependent variable is the logarithm of annual earning of individual.
The standard errors are heteroskedasticity robust and clustered at the individual level

Bartel and Lichtenberg, (1987) suggest that individuals with high education are more spatially mobile due to changes in career path. Borsch-Supan, (1990) suggest that higher educated individuals are characterized with higher return than low educated workers which leads to the lower marginal cost of commuting. Finally, Cahuc et al., (2006) states that higher skilled individuals have more bargaining power which allows them to receive higher compensation for commuting.

The overall results suggest return to commuting increases with the level of education attained. The compensation is similar for men and women.

Summarizing the presented results, we can conclude that highest return to commuting is received by high educated people. Also, individuals with highest earnings also have highest return to commuting. Jointly these two facts are consistent bargaining power raising the compensation for commuting.

2.10 Comparison of the results of estimation

One of the most important conclusions of our work is that the returns to commuting for the male and female samples do not vary too much. So to say, OLS reports 16% of hourly wage compensation for 1 hour of commuting whereas females receive 17% more. FE estimation reports approximately similar returns for 1 hour of commuting across genders, although lower in magnitude: 8.5% for the male sample and 7.5% for the female sample. Results on estimation the model with inclusion of firm fixed effects suggests that women have slightly higher return to commuting (10.8%) comparing to men (9.34%), which are not however statistically different due to the overlapping confidence intervals. Also, commuting contributes to the earnings' growth rate of individuals in a medium run perspective. Males received an additional 0.4% of earnings' growth rate per one hour of commuting over the period 2003-2010. By contrast, the point estimates for the females are not significantly different from zero in this case.

The age dummies as a proxy of working experience suggests that the experience profile is more concave for women with a later turning point (approximately between 45 and 50). Variables indicating marital status suggest that married or cohabitating individuals of both genders have higher earnings than single individuals. Single parents of both genders experience approximately similar earnings penalties both in magnitude and significance. Education is more rewarding for females. At the same time, female individuals employed in sectors other than public service experience a decrease in earnings likely due to higher gender discrimination or "glass ceiling effects". Males employed in sectors different from public services do not experience significant increases in their earnings. It suggests the fact that the wage distribution across sectors is quite compressed. Macroeconomic variables affect the earnings of individuals of both genders in similar fashions and magnitudes.

To conclude, it is possible to say that males and females obtain approximately similar compensation for commuting. It can be viewed as an approximately similar bargaining power. Another explanation would be similar levels of efficiency during the spatial job-search process.

2.11 Conclusion

We provide evidence that the wage return to commuting is increasing in commuting distances within the borders of economically justified regions. Evidence of concavity, commonly reported by previous studies, is also found. We have also provided evidence that commuting induces a faster earnings' growth rate for males in the medium-run. Moreover, there is no significant evidence of a gender gap in compensation for commuting..

Our study addresses only one aspect of the reward from commuting, received through the job-search process under the form of higher compensation from the employer. We do not take into account the implicit compensation received by individuals from differences in housing prices, availability of natural and social amenities, differences in taxation, and the availability of public goods. Moreover, this study does not allow identifying net gains or losses due to commuting because of unavailability of information on commuting expenses and losses in social capital or health related to commuting.

Positive returns to commuting can be attributed to the bargaining power of individuals and the consequent thinness of the labor market, the efficiency of the job search process or differences in productivity across spatial units. Taking into account the wage formation process in Sweden, it is likely that the explanation provided by spatial differences in the employer monopoly power is also reasonable. Consistently with this explanation, we find returns to commuting to be higher for highly educated individuals and in the top quartile of the earnings distribution.

Our results provide some suggestions for further study. The current study, which focuses on individual heterogeneity of employees, can be usefully expanded to incorporate employer information with employer-employee matched data, in order to capture differences in productivity between employers and the consequent possibility to compensate workers differently for their commuting distances. Moreover, the availability of variables which better reflect commuting expenses and labor supply (i.e. considering wages rather than annual earnings, which mix information on wages and working hours) would significantly increase the precision of our estimates and provide a precise answer to answer a crucial question about the "wage return" to commuting.

References

Abowd, J. M., and Kramarz, F., 1999. The analysis of labor markets using matched employer-employee data. *Handbook of labor economics*,*3*, pp. 2629-2710.

Bartel A. P. and Lichtenberg F. R., 1987. The comparative advantage of educated workers in implementing new technologies. *Review of Economics and Statistics 69*. pp. 1-11(2 razy)

Benito, A. and Oswald, A.J., 2000. Commuting in Great Britain in the 1990s.

Booth A L, Francesconi M and Garcia-Serrano C., 1999. Job tenure and job mobility in Britain. *Industrial and Labour Relations Review 53*. pp. 43-70

Borsch-Supan A., 1990) Education and its double-edged impact on mobility. *Economics of Education. Review 9.* pp.39-53

Cahuc, P., Postel-Vinay, F. and Robin, J.M., 2006. Wage bargaining with on-the-job search: Theory and evidence.*Econometrica*, *74*(2), pp.323-364.

Camstra R. (1996). Commuting and Gender in a Lifestyle Perspective. *Urban Studies, Vol. 33*, No. 2, pp 283- 300

Dargay, J.M. and Clark, S., 2012. The determinants of long distance travel in Great Britain. *Transportation Research Part A: Policy and Practice*, 46(3), pp.576–587.

DiAddario, S. and Patacchini E., 2008. Wages in the Cities: Evidence from Italy. *Labour Economics,* 15, pp.1040-1061.

Eliasson, K., Lindgren, U. and Westerlund, O., 2003. Geographical Labour Mobility: Migration or Commuting? *Regional Studies*, 37(8), pp.827–837.

Fu, S. and Ross S.L., 2010. "Wage Premia in Employment Clusters: Agglomeration or Worker Heterogeneity?." *US Census Bureau Center for Economic Studies Paper No. CES-WP-10-04* .

Fu, S., 2007. Smart Café Cities: Testing Human Capital Externalities in the Boston Metropolitan Area. *Journal of Urban Economics*, 61, pp. 86-111.

Gutiérrez-i-Puigarnau, E. and van Ommeren, J.N., 2010. Labour supply and commuting. *Journal of Urban Economics*, 68(1), pp.82–89.

van Ham, M., Mulder, C.H. and Hooimeijer, P., 2001. Spatial flexibility in job mobility: macrolevel opportunities and microlevel restrictions. *Environment and Planning A*, 33(5), pp. 921–940.

Hazans, M., 2004. Does commuting reduce wage disparities? *Growth and Change* 35.3 pp.360-390.

Isacsson, G. and Swärdh J.E., 2007. An empirical on-the-job search model with preferences for relative earnings: How high is the value of commuting time?

Lyons, G. and Chatterjee, K. 2008. A human perspective on the daily commute: Costs, benefits and trade-offs. *Transport Reviews*, *28*(2), pp. 181-198.

Lundholm, E., 2010. Interregional Migration Propensity and Labour Market Size in Sweden, 1970–2001. *Regional Studies*, 44(4), pp.455–464.

Manning, A., 2013. Minimum wages: a view from the UK. *Perspektiven der Wirtschaftspolitik*, *14*(1-2), pp.57-66.

Manning, A., 2003. The real thin theory: monopsony in modern labour markets. *Labour Economics*, 10, pp.105–131.

McQuaid, R.W. and Chen, T., 2012. Commuting times – The role of gender, children and part-time work. *Research in Transportation Economics*, 34(1), pp.66–73.

Mulalic, I., Van Ommeren, J.N. and Pilegaard, N., 2014. Wages and Commuting: Quasi-natural Experiments' Evidence from Firms that Relocate.*The Economic Journal*, *124*(579), pp.1086-1105.

Nakosteen, R.A. and Westerlund, O., Michael Zimmer (2008)," Migration and self-selection: measured earnings and latent characteristics. *Journal of Regional Science*, *48*, pp.769-788.

NUTEK. 2000. Arbetskraftens rorlighet : ett smorjmedel for tillvaxt. Rapport 2000:15. NUTEK, Stockholm.

NUTEK (NARINGS-OCH TEKNIKUTVECKLINGSVERKET). 2001. Innovative Clusters in Sweden: Practical Lessons from Regional Cluster-building. Stockholm.

van Ommeren, J., 2004. The commuting distribution. No. 04-045/3. *Tinbergen Institute Discussion Paper*.

van Ommeren, J., Rietveld, P. and Nijkamp, P., 2000. Job mobility, residential mobility and commuting: A theoretical analysis using search theory. *The Annals of Regional Science*, 34(2), pp.213–232.

van Ommeren, J.N. and Gutiérrez-i-Puigarnau, E., 2011. Are workers with a long commute less productive? An empirical analysis of absenteeism. *Regional Science and Urban Economics*, 41(1), pp.1–8.

Van Ommeren, J.N. & van der Straaten, J.W., 2008. The effect of search imperfections on commuting behaviour: Evidence from employed and self-employed workers. *Regional Science and Urban Economics*, 38(2), pp.127–147.

Van Ommeren, J. and Rietveld P., 2002. *Commuting, spatial search and labour market bargaining*. No. 02-039/3. Tinbergen Institute Discussion Paper

Van Ommeren, J., 2004. *The commuting distribution* (No. 04-045/3). Tinbergen Institute Discussion Paper.

Van Ommeren, J. and Rietveld, P., 2005. The commuting time paradox.*Journal of Urban Economics*, *58*(3), pp.437-454.

Van Ommeren, J. and Fosgerau, M., 2009. Workers' marginal costs of commuting. *Journal of Urban Economics* 65.1, pp. 38-47.

Renkov, M.and Hoover, D., 2000, Commuting, Migration and Rural-Urban Population Dynamics. *Journal of Regional Science,* 40(2), pp. 261-287.

Rouwendal, J., Rietveld, P., 1994. Changes in commuting distances of Dutch households. Urban Studies

Rouwendahl, J., 1998. Search theory, Spatial Labour Markets and commuting. *Journal of urban Economics,* 43, pp.1–22.

Rouwendahl, J., 1999. Spatial job search and commuting distances. *Regional sciense and urban economics*, 29. pp. 491-517.

Rouwendal, J., 2004. Search Theory and Commuting Behavior. *Growth and Change*, 35(3), pp.391–418.

Rupert, P. and Wasmer, E., 2009. *Housing and labor market: time to move and aggregate unemployment,*

Ruppert, P., Stancanelli E. and Wasmer E., 2009 "Commuting, wages and bargaining power." *Annals of Economics and Statistics/Annales d'Économie et de Statistique,* pp. 201-220.

Sandow, E., 2008. Commuting behaviour in sparsely populated areas: evidence from northern Sweden. *Journal of Transport Geography*, 16(1), pp.14–27.

Sandow, E, and Westin K., 2010.Preferences for commuting in sparsely populated areas. *Journal of Transport and Land Use* 2.3/4 pp. 87-107.

Sandow, E., Westerlund, O. and Lindgren, U., 2014. Is your commute killing you? On the mortality risks of long-distance commuting. *Environment and planning A*, *46*(6), pp.1496-1516.

Sandow, E., 2013. Til work do us part: The social fallacy of long-distance commuting. *Urban Studies*, p.0042098013498280.

Shuai, X., 2012. Does Commuting Lead to Migration?. *Journal of Regional Analysis & Policy*, *42*(3), p.237.

Simpson, W., 1980. A simultaneous model of workplace and residential location incorporating job search. *Journal of Urban Economics*, *8*(3), pp.330-349.

Simpson, W., 1992. Urban structure and the labor market.

Swedish Government, 2008. Prop. 2008/09:93 Mål för framtidens resor och transporter (the aim for future travel and transport)

Wasmer, E. and Zenou, Y., 2002. Does City Structure Affect Job Search and Welfare? *Journal of Urban Economics*, 51(3), pp.515–541.

Wasmer, E. and Zenou, Y., 2006. Equilibrium search unemployment with explicit spatial frictions B. , 13, pp.143–165.

Weber, J., and Sultana, S. 2007. Journey-to-Work Patterns in the Age of Sprawl: Evidence from Two Midsize Southern Metropolitan Areas*. *The Professional Geographer*, *59*(2), pp. 193-208.

Winter-Ebmer, R., and Zweimüller, J., 1999. Firm-size wage differentials in Switzerland: Evidence from job-changers. *American Economic Review*, pp. 89-93.

Winter-Ebmer, R., 1996. Wage curve, unemployment duration and compensating differentials. *Labour Economics*, *3*(4), pp. 425-434.

Östh, J., 2007. Home, job and space: mapping and modeling the labor market.

Chapter 3

Hazard from commuting: the role of earnings and distance.

The case of Sweden

Sergii Troshchenkov*

Abstract. The aim of this paper is to estimate the effect of earnings and commuting distance on the probability of exiting from a duration spell of commuting using a discrete time competing risk model. The data set used is based on the Swedish administrative registers from Statistics Sweden and the Swedish Tax Board and covers the period between 2000 and 2009. The problem of endogeneity of individual earnings and commuting distance in determining the length of work-related commuting spells is addressed using two-stage residual inclusion (2SRI). The estimates reveal that the earnings paid by firms have a positive impact on the probability of migration and a negative impact on the probability of job separation. At the same time, greater distance increases the probabilities of migrating closer to the place of work, re-employment closer to the place of residence and separation to non-employment while decreasing the probabilities of migration further away from the place of work and re-employment further away from the place of residence. The results are revealed to be robust in the samples of married and unmarried individuals.

* University of Milan and GRAPE (Group for Research in APplied Economics), email:sergii.troshchenkov@unimi.it

3.1 Introduction

Many studies report a significant increase in the number of commuters and commuting distances during the past decades in most European countries. Lundholm (2010) reported that the average commuting distance has increased by almost half (from 10 km in 1970 to 15 km in 2001) during the last 30 years in Sweden. Meanwhile, the majority of people still tend to live close to their place of work: 50% of men reside within 8 km of their place of work, while 50% of women live within 6 km of their place of work. The distance to work has significant implications for both an individual's social life and his or her labour market performance. Rouwendal (1999, 2004) developed a theoretical model that suggests that individuals tend to accept commuting as long as the expected premium exceeds the transportation losses. In other words, individuals commute if the compensation from employers, price differentials on the housing market or benefits from the level of local amenities offset the losses associated with commuting (Potepan, 1992; Renkow & Hower, 2000; van Ommeren & Rietveld, 2005). Together with that, empirical studies have highlighted the significant negative impact of commuting on an individual's labour market performance and personal life. Ross and Zenoy (2008), along with Van Ommeren and Gutiérrez-i-Puigarnau (2010, 2011), showed that commuting time is positively associated with shirking and absenteeism, which in turn leads to a higher separation rate for blue-collar workers. The empirical relation with absenteeism has also been found to be a significant factor of the labour performance in Sweden (Hansson, 2011). Moreover, commuting has negative consequences for health. Various studies have reported a higher level of stress for long-distance commuters due to the stressfulness of the traffic situation and congestion associated with public transport usage (Koslowsky et al, 1996; Gottholseder et al, 2009). Lingren et al (2014) found a higher mortality rate among older commuters. Moreover, commuting is associated with a higher level of conflicts within the family (Hammong et al, 2009). Sandow (2013) established that persistent long-distance commuting is associated with a higher divorce rate.

The main focus of this study is on analysing the duration of long-distance commuting and the role of individuals' personal and social characteristics in various exits from commuting. In our analysis we consider the following hazards from commuting: migration further away from the place of work, migration closer to the place of work, separation to non-employment, re-employment further away from the place of residence and re-employment closer to the place of residence. Hence, our main research question is:

Which factors affect the multiple hazards from commuting?

Since commuting behaviour is driven by household characteristics, the sample includes register-based information on individual characteristics. The model is estimated using a discrete time competing risk multinomial logit model in which competing risks are defined as multiple exits from commuting with respect to remaining in the reference state (i.e. commuting).

This paper contributes to the existing literature in several ways. Firstly, we analyse jointly the determinants of duration and multiple hazards from commuting using a competing risks approach. Secondly, we apply discrete time settings that allow us to deal with competing risks in the presence of a large number of ties.[18] Thirdly, we use a comprehensive longitudinal data set based on the administrative registers of Statistics Sweden. The results from this paper can provide insights for policy makers and employers into the duration of stay in the place of work and workers' incentives for changing their place of residence or place of work.

The remainder of the paper is structured as follows. In the next section, the literature review is presented. Sections 3 and 4 report the descriptive statistics and contain the description of the theoretical and empirical models applied in the paper. Section 5 presents the main findings of the research paper. Finally, Section 6 summarizes the conclusions and discussions.

[18] *Ties are the number of commuting spells with exactly the same duration caused by grouping according to yearly intervals (Hess & Persson, 2010).*

3.2 Duration of long-distance commuting and hazards

3.2.1 General description of commuting behaviour

Looking at the macro-level, the availability of infrastructure and long-distance commuting improves the employer–employee matching process. The intensities of the commuting streams have various implications for regional development, such as the mitigation of regional disparities and the reduction of inequality between regions (Hazans, 2004). At the same time, Calthrop (2010) pointed out that infrastructure projects aimed at decreasing the commuting time and expenses speed up the economic growth of regions. At the micro-level, commuting decisions can be viewed as a simultaneous choice of the place of work and place of residence while optimizing the monetary and non-monetary losses associated with commuting (Rouwendal, 1999; Eliason et al, 2003). The empirical findings suggest that individuals who are involved in long-distance commuting experience various positive and negative consequences of this phenomenon. As pointed out by Manning (2003) and later by Rupert et al (2008) and Mulalic et al (2014), long-distance commuters receive partial compensation for the expenses related to long-distance commuting. Fu and Ross (2010) showed that individuals employed in urban agglomerations tend to have higher wages due to the higher productivity levels associated with the agglomeration-related externalities. Moreover, commuters might benefit from the accommodation price differential between the place of work and the place of residence (Potepan, 1994; So et al, 2001; Romani et al, 2003; Gobillon et al, 2007). Additionally, they might enjoy the local amenities in the place of residence that might partially offset the commuting costs (Renkow & Hower, 2000; Clark, 2004). Finally, commuters might obtain positive utility while commuting by actively exploiting the commuting time, for instance by working while travelling or using the Internet (Llyons & Urry, 2005; Gottholmseder, 2009).

Although commuting can be the outcome of an optimizing strategy, it is connected to various monetary and social costs (White, 1977). Empirical researchers have identified the commuting expenses for one hour of commuting time as about 50% of the hourly wage. Calfee and Winston (1998) estimated the willingness to pay (WTP) for one hour of commuting time to be about 20% of the hourly wage. In line with the previous research, De Borger and Forsgerau (2008) found that the WTP constitutes between 10% and 40% of the hourly wage. By contrast, van Ommeren and Forsgerau (2009) identified the total monetary and non-monetary expenses for commuting as around 200% of the hourly earnings.

Commuting leads to various negative social outcomes, such as the negative impact on the household production due to the lack of time available at home (Gimenez & Molina, 2015). Furthermore, a long commuting distance is connected with higher levels of family conflicts (Hamming et al, 2009). Sandow (2013) reported a higher divorce rate in families with one or two commuters. Koslowsky et al (1995) showed that commuting is associated with various health and mental costs.

Regarding the individual perception, commuters experience a higher level of stress associated with long-distance journeys and report lower levels of life satisfaction (Cramer et al, 2008; Gottholmseder, 2009). Sabir et al (2010) proposed that the individual cost of commuting might rise with the risk of traffic accidents or congestion. Hanson et al (2011) and Hoehner et al (2012) suggested that long-distance commuting has an impact on individuals' blood pressure and causes deterioration of their sleep quality, obesity and various other health outcomes. Moreover, it was found that the magnitude of the negative effect on health increases with the commuting distance for users of public transportation systems. Finally, Lingren et al (2014) reported a higher mortality rate for long-distance commuters in the age group above 55 years old.

3.2.2 Various hazards from commuting

In principle, the decision to start or stop commuting can be viewed as a repeated decision about exiting or staying in the commuting state every period of time. If individuals decide to exit a particular earning–distance combination, they face several independent risks, such as: 1) migration closer to the place of work (i.e. a change in the place of residence); 2) migration further away from the place of work; 3) job separation with dropping out of the labour force; 4) re-employment closer to the place of residence; and 5) re-employment further away from the place of residence.

Migration closer to the place of work is a direct consequence of the optimizing behaviour of individuals, whereby they tend to offset their transportation costs by selecting the optimal distance–residence combinations that would generate the highest level of individual utility (Rouwendal, 1999; Shuai, 2012). In a dual-earner household, migration closer to the place of work of an individual can occur if it offsets the losses generated by the change in the commuting distance and labour supply patterns of the individual's partner (Mincer, 1978).

Migration further away from the place of work is driven mainly by two factors: tied moving or moving to choose the level of housing prices and/or amenities composition that can offset the additional commuting expenses associated with increased commuting distances (Potepan, 1994; Renkow & Hower, 2008; Åstrom & Westerlund, 2009). Tied moving is a situation in which an individual experiences losses due to changing the place of residence or work while his or her partner receives gains that offset them. Åstrom and Westerlund (2009) found that tied moving occurs if the household income balances the individual losses associated with moving. Potepan (1994) and Hower and Renkow (2000) reported that individuals often take into consideration the level of specificity of local amenities and prices of accommodation when selecting the area of residence.

Individuals might also tend to select job separation or re-employment in the case of unsatisfactory job–distance combinations. Manning (2003) suggested that higher rates of job separation among long-distance commuters occur if the employees are significantly under-compensated for their commuting time. Generally speaking, job separation in the form of re-employment further from or closer to home or separation to non-employment occurs if individuals are not compensated enough for commuting by their employer. The other cause of the higher separation rate among commuters is a higher level of absenteeism and shirking, which leads to higher involuntary layoff rates among commuters (Ross & Zenoy, 2008; Van Ommeren & Gutiérrez-i-Puigarnau, 2010, 2011).

3.3 Descriptive statistics

The data used in the empirical analysis are drawn from various administrative registers compiled by Statistics Sweden and the Labour Market Board of Sweden and held at the Department of Geography and Economic History at Umea University.[19] The data set represents annual detailed records on various individual personal and demographic characteristics, including income, education, marital status, educational attainments, employment and georeferenced coordinates of the place of work and residence in the United Transverse Mercator system of coordinates.

The sample is constructed for the period 2000 to 2009. The condition of entering the sample is finding a job in the year 2000. Individuals who entered employment later are not included. This condition allows us to ensure that the sample does not suffer from left-truncation and left-censoring issues. Individuals who took a migration decision in the year 2000 are excluded from the analysis, since we are unable to track the competing risk probabilities of the alternative outcomes for those individuals. Essentially we also assume that there are no multiple entries in the sample. The six competing risks that we consider are:

0. Maintaining a stable employment distance status (reference)
1. Migration further away from the place of work (i.e. more commuting)
2. Migration closer to the place of work (i.e. less commuting)
3. Separation to non-employment
4. Re-employment closer (i.e. less commuting)
5. Re-employment further (i.e. more commuting)

Therefore, these 6 competing risk hazards represent our multinomial dependent variable. In these conditions experiencing alternative 0 is defined as surviving in the commuting state at time *t* and represents the baseline category in the model. Basically, our data set is an unbalanced panel with a multinomial discrete dependent variable.

A set of controls, consisting of individuals' gender, income from employment, commuting distance, marital status, previous labour market experience, educational variables, employment status and sector of employment, is explicitly incorporated into the model. The controls are

[19] *Access to the data was kindly offered to us by Prof. Urban Lingren and the Department of Geography and Economic History at Umea University.*

defined as time invariant to avoid reverse causality issues and represent individuals' characteristics in the first period (i.e. 2000).

The main independent variables of interest are *Log of commuting distance* and *Log of earnings*. Distance reflects the logarithm of the Euclidian commuting distance between the individual's place of residence and his or her place of work calculated with the application of the Pythagoras formula. *Log of earnings* represents the logarithm of individual earnings from the main place of employment. An extensive set of other explanatory variables is included to capture the individual heterogeneity within the sample. *Gender* is a dummy variable in which men are coded as 1 and women as 0. This variable is introduced to capture the gender difference in the decision to exit a commuting spell. *Age* represents the age of individuals at the beginning of the spell. Previous studies have documented decreasing residential and labour market mobility of older individuals. *Nationality* is a dummy variable that represents the nationality of individuals, in which individuals of non-Swedish origin are coded as 1. *Children* is a dummy variable that reflects the presence of children under 18 within the household. *Private services*, *Retailing*, *Construction*, *Manufacture* and *Public services* are dummy variables for the sector of employment, representing private services, retailing, construction, manufacturing and public services, respectively. The *Public services* sector of employment is selected as the reference category. *Previous income* is a logarithm of the sum of annual labour market income for the 4 preceding years. It is included to approximate the accumulated financial wealth of individuals. *Unemployment experience* is a dummy variable that indicates the incidence of unemployment of individuals during the preceding 4 years. Hence, the individuals who experienced unemployment at least once during the last 4 years before the beginning of the study period would be encoded as 1, while others would be encoded as 0. The family status of an individual is represented by the following set of dummies: *Married*, *Single parent*, *Living with parents* and *Single*. *Married* is an indicator of being married *or* in "sambo". *Single parent* indicates single parents. *Living with parents* denotes single individuals living with their parents. Finally, *Single* represents a category of single individuals. *Married* is selected as the reference group. The educational variables are represented by the following set of dummies: *Primary education*, *Gymnasium education*, *Post-secondary education<2 years*, *Post-secondary education>2 years and MSc or Ph.D*. *Primary education* is defined as the lowest level of education corresponding to pre-gymnasium education, and *MSc or Ph.D.* is assigned the highest level of education achieved, which corresponds to university education. *Log of time* represents the logarithm of time in the spell.

The results of the descriptive statistics by the type of exit are presented in Table 1. The table provides some useful insights for the description of the data. The descriptive statistics suggest that the earnings of individuals who experience migration are comparatively higher than the earnings of individuals who experience other types of exit. At the same time, individuals who migrate further away from their place of work or change workplace further away from their place of residence have a much smaller commuting distance than individuals who experience the other types of exit. Table 1 indicates a higher number of females who separate to non-employment and migrate further away from their place of work. This may result from the fact that women are more prone to unemployment and tied moving. On average, there are lower numbers of movers among individuals of foreign origin than the other types of exit. Individuals who have experienced migration are younger and less likely to have children than individuals who have selected other types of exit, such as separation to non-employment and re-employment. Among other features, migrants have higher levels of education than those who switch their place of work.

Table 1. Descriptive statistics by the type of exit

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Re-employment closer*	*Re-employment further*
Gender	*0.4883*	*0.5516*	*0.4797*	*0.5013*	*0.5061*
	(-0.4998)	*(-0.4973)*	*(-0.4995)*	*(-0.4999)*	*(-0.4999)*
Age	*32.9409*	*33.2146*	*35.6563*	*34.4414*	*34.3085*
	(-10.4959)	*(-10.7821)*	*(-12.8459)*	*(-11.7194)*	*(-11.707)*
Nationality	*0.1015*	*0.102*	*0.1374*	*0.1175*	*0.1167*
	(-0.302)	*(-0.3027)*	*(-0.3443)*	*(-0.322)*	*(-0.3211)*
Children	*0.158*	*0.1924*	*0.208*	*0.2211*	*0.2217*
	(-0.3647)	*(-0.3942)*	*(-0.4059)*	*(-0.4149)*	*(-0.4154)*
Log of (earnings)	*7.3497702*	*7.5047776*	*6.7186089*	*7.049636*	*7.0358561*
	(0.9616)	*(0.7764)*	*(1.4499)*	*(1.2402)*	*(1.2528)*
Log (distance)	*12.079634*	*57.491843*	*31.176965*	*52.945862*	*14.864245*
	(43.3503)	*(117.6834)*	*(89.4823)*	*(118.5172)*	*(50.0031)*

Table 1: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Re-employment closer*	*Re-employment further*
Private services	*0.0966*	*0.0886*	*0.0572*	*0.0845*	*0.0814*
	(-0.2954)	*(-0.2841)*	*(-0.2322)*	*(-0.2782)*	*(-0.2735)*
Retailing	*0.2343*	*0.2373*	*0.1853*	*0.2328*	*0.2332*
	(-0.4236)	*(-0.4254)*	*(-0.3885)*	*(-0.4226)*	*(-0.4229)*
Construction	*0.0531*	*0.056*	*0.0606*	*0.0511*	*0.0524*
	(-0.2242)	*(-0.2299)*	*(-0.2322)*	*(-0.2203)*	*(-0.2229)*
Manufacture	*0.1724*	*0.1959*	*0.1499*	*0.1332*	*0.1415*
	(-0.3778)	*(-0.3969)*	*(-0.357)*	*(-0.3398)*	*(-0.3485)*
Previous income	*8.131*	*8.0043*	*7.9368*	*8.1028*	*8.0935*
	(-1.2902)	*(-1.3889)*	*(-1.457)*	*(-1.3613)*	*(-1.3744)*
Unemployment experience	*0.2859*	*0.2788*	*0.2973*	*0.2768*	*0.2753*
	(-0.4518)	*(-0.4484)*	*(-0.4571)*	*(-0.4474)*	*(-0.4467)*
Single parent	*0.0679*	*0.0643*	*0.0651*	*0.0615*	*0.0626*
	(-0.2516)	*(-0.2454)*	*(-0.2468)*	*(-0.2402)*	*(-0.2423)*
Living with parents	*0.1415*	*0.1602*	*0.214*	*0.1976*	*0.2084*
	(-0.3485)	*(-0.3668)*	*(-0.4101)*	*(-0.467)*	*(-0.4062)*
Single	*0.436*	*0.4711*	*0.3038*	*0.3214*	*0.3128*
	(-0.4959)	*(-0.4991)*	*(-0.4599)*	*(-0.467)*	*(-0.4636)*
Gymnasium level of education	*0.5187*	*0.5358*	*0.5013*	*0.4844*	*0.4908*
	(-0.4996)	*(-0.4987)*	*(-0.4999)*	*(-0.4997)*	*(-0.4999)*
Post-secondary education <2 years	*0.0779*	*0.0634*	*0.0644*	*0.0702*	*0.0708*
	(-0.2681)	*(-0.2436)*	*(-0.2455)*	*(-0.2555)*	*(-0.2565)*
Post-secondary education >2 years	*0.2686*	*0.2683*	*0.2076*	*0.2591*	*0.2528*
	(-0.4433)	*(-0.4431)*	*(-0.4056)*	*(-0.4381)*	*(-0.4346)*
MSc or PhD	*0.0087*	*0.0092*	*0.007*	*0.0074*	*0.0062*
	(-0.0928)	*(-0.0958)*	*(-0.0835)*	*(-0.0859)*	*(-0.079)*
Number of obs.	*11954*	*12302*	*186361*	*282795*	*283268*

3.4 Econometric method

The causal impact of the income from employment and commuting distance on the probabilities of exiting the distance employment spell is assessed using a discrete competing risk model with the application of a multinomial logit model.

3.4.1 General specification

The main idea behind the choice of the model for estimation lies in the underlying assumptions on the data and hazard distribution. The nature of exits suggests that they occur continuously during the time, although, we are able to observe the hazard only in discrete intervals (years). This fact leads to the significant number of ties, that is, spells of the same length. Ties together with unknown distribution in unobservable heterogeneity lead to significant complications in the estimation of the duration model. Therefore, following Hess and Persson (2010), we use a discrete time hazard setting. In these conditions the choice of the model lies in the selection of the appropriate discrete dependent variable. As pointed out by Beck, Katz and Tucker (1998), the choice of model is rather trivial. Hence, to estimate the competing risk probabilities of exiting the employment commuting spell, we apply the multinomial logit model suggested by Jenkins (2005).

Below we describe the theoretical discrete case duration model used in our analysis. Although the hazard of changing from commuter to non-commuter status occurs continuously, the exits are observed only within the time interval. Therefore, we define the conditional probability of the particular exit occurring at the time $[t_a, t_{a+1})$, a=2000,…,2009 as[20]:

$$\lambda_{im} = \frac{e^{X_i \beta_m}}{\sum_{k=0}^{K} e^{X_i \beta_m}} \tag{1}$$

where λ_{im} denotes the conditional probability of experiencing the event *m* in which the events are: (1) migration closer to the place of work; (2) migration further away from the place of work; (3) separation to non-employment (i.e. unemployment and separation from the labour force); (4) re-employment closer to the place of residence; and (5) re-employment further away from the place of residence. $C_i \beta_m$ is defined as:

$$C_i \beta_m = \beta_{0m} + D\beta_{1m} + Y\beta_{2m} + X\beta_{jm} + \omega_i \tag{2}$$

[20] *In our notation we follow Green (2008) and Cameron and Trivedi (2010).*

where $i=1...N$ indicates cross-sectional units of analysis, D represents individual commuting distance, Y denotes earnings from employment, X encapsulates generic set of variables that captures individual characteristics and life cycle events and ω is an individual error term.

Therefore, the likelihood's contribution for individual *I* can be written as:

$$\zeta_i = \lambda_{i1}^{d_{i1}} * \lambda_{i2}^{d_{i2}} * \dots * \lambda_{iK}^{d_{iK}} \tag{3}$$

where $\lambda_{i1}^{d_{i1}}$ denotes the probability occurring for individual *i* to experience hazard k and d_{i1} is a dichotomous indicator.

The likelihood function for the entire sample of individuals is expressed as:

$$\zeta = \prod_{i=1}^{N} (\lambda_{i1}^{d_{i1}} * \lambda_{i2}^{d_{i2}} * \dots * \lambda_{iK}^{d_{iK}}) \tag{4}$$

Hence, the log transformation of the likelihood function will take the form of:

$$ln\zeta = \sum_{i=1}^{N} \sum_{K=1}^{K} d_{ik} \ln(\lambda_{ik}) = \sum_{i=1}^{N} \sum_{K=1}^{K} d_{ik} \ln\left(\frac{e^{X_i \beta_k}}{\sum_{k=0}^{K} e^{X_i \beta_k}}\right) \tag{5}$$

The marginal effects of regressor X_i on the probabilities are:

$$\frac{\partial P_k}{\partial x_i} = P_k \left[\beta_j - \sum_{k} P_k \, \beta_k\right] \tag{6}$$

Therefore, the parameters provide insights into the probability of experiencing event *m* relative to the probability of survival.

In this setting the issue of endogeneity arises basically due to the fact that past commuting distances and earnings are not randomly assigned to individuals. They may be determined by unobservable variables such as individual ability, which is also correlated with future optimal commuting and migration choices (Manning, 2003; Mulalic et al, 2010). The common way of dealing with such types of problems is to use an instrumental variables approach. In our model the problem is more complicated since we have two variables, *Distance* and *Earnings*, that are likely to suffer from endogeneity issues.

3.4.2 Endogeneity of income

Despite the clear evidence on the positive relation between the length of employment spells and the monthly wage (D'Addio & Rosholm, 2005), there is also evidence that unobserved ability may bias the estimated effects of earnings on the length of the employment spell. In this context high earners might possess a greater ability that is positively correlated with the spell duration. Therefore, the estimates may be upward biased. Since individual ability can potentially simultaneously affect the decision to stay with an employer and the annual earnings, the inclusion of income earned only by luck might be considered to be fairly exogenous in this particular setting. To be more specific, we assume that some firms tend to pay higher wages only by luck (i.e. exogenously to the characteristics of specific workers), possibly due to higher firm-specific productivity or regional externalities. Lewis (1986) and Krueger and Summers (1988) showed a significant variation of wages across sectors and union status. Shea (2000) argued that this earnings premium reflects the "rent" paid by the sector to the workers. We take this evidence further and assume that some firms within sectors also tend to pay rents to their workers or a firm-specific wage premium (Card et al, 2016). To deal with this issue, it is possible to simulate an experiment by transferring money through firms to random workers and then tracing their future employment history. Following Shea (2000), we approximate this experiment by assuming that some firms are more productive than others within a sector for reasons unrelated to their workers' characteristics. Therefore, these firms tend to pay higher wages only by luck. This setting is applicable especially to Sweden, where wage bargaining occurs at the sectoral level between employers' organizations and labour market unions and the minimum wage level is set for all the members of the employer's union. Workers do not generally observe firm productivity or other factors that can drive their wages up; their initial wage level is common to every firm within the same sector owing to a commonly fixed level of wages, collectively established by employers and labour unions. Therefore, the firm fixed effects in an earnings equation of workers after controlling for the sector of employment, individual characteristics and patterns of labour supply reasonably represent a "rent" paid to all the employees.

The model for the estimation of the "firm rent" is specified as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 S_m + \beta_3 W_j + \varepsilon_i \qquad (7)$$

where $i=1...N$ indicates cross-sectional units of analysis, $j=1...K$ indicates firms and $m=1...5$ denotes the individual sector of employment. Y represents the income of individuals, X is a vector of variables that encompasses education, experience and patterns of labour supply, S is the sector in which an individual is employed, W indicates the firm-specific time-invariant effect "firm rent", ε is an individual specific error term and $\beta_{0}, \ldots, \beta_{3}$ are coefficients to be estimated.

We calculate the "firm rent" by using an unbalanced panel conditional on the employment of individual i in firm j and estimating an earnings equation using firm fixed effects among the controls. Subsequently, "firm rent" W is used instead of individual earnings in the main competing risk equation. The results of the earnings equation are reported in Table 2.

Table 2. Estimation of firm fixed effects

Variables	*Model for firm-level fixed effects*	
	Coefficient	*Standard errors*
Gender	*0.2839****	*0.0011*
Age variables		
Age	*0.12909****	*0.00034*
Square age	*-0.00129****	*4.08e-06*
Family status variables		
Married	*-0.05394****	*0.00116*
Single mother/father	*-0.07958****	*0.00202*
Living with parents	*-0.35692****	*0.002*
Children	*-0.03307****	*0.0012*
Education level		
Gymnasium level of education	*0.28223****	*0.00146*
Post- gymnasium level of education <2 years	*0.2091****	*0.00225*
Post-gymnasium level of education >2 years	*0.48259****	*0.00171*
University level of education	*0.89901****	*0.00503*
Nationality	*-0.12439****	*0.00152*
Sector of employment		
Manufacture	*0.15047****	*0.00701*
Construction	*0.47964****	*0.00736*
Retailing	*0.19843****	*0.00651*
Private services	*0.3277****	*0.00639*
Constant	*3.9816****	*0.00746*
Adjusted R^2	*0.2204*	
Number of observations	*3451203*	

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the log of annual earnings.
The standard errors are92eteroscedasticityy robust and clustered at the individual level

In this setting, firm fixed effects represent firm-specific increments to earnings that are unrelated to the observable individual characteristics or ability. Therefore, these rents paid by firms can be used instead of earnings, which are contaminated by the ability bias.

3.4.3 Endogeneity of distance

The commuting distance plays a significant role in the determination of the length of distance employment spells of individuals. Commuting distance, like past earnings, is also likely to be correlated with individual unobservable characteristics, such as the ability to tolerate commuting costs ("commuting tolerance"). The commuting tolerance affects both the accepted commuting distance and the length of the commuting spells.

To address the endogeneity in the commuting distance and evaluate the effect of the commuting distance on the distance employment spell, we implement an estimation strategy based on the inclusion of a significant number of individual variables to capture individual heterogeneity and apply the residual inclusion method to correct for endogeneity bias in the commuting distance (Terza et al, 2008).

Terza et al (2008), Ben-Akiva and Guevara (2009) and Woolbridge (2014) argued that residual inclusion (RI, hereafter) in the main non-linear model along with the main endogenous explanatory variables permits unbiased estimates to be obtained of the main endogenous explanatory variable of interest on the probability of leaving the spell. The main condition for the application of two-stage RI (2SRI, hereafter) is the availability of valid instruments (Terza et al, 2008).

Therefore, in the first stage, the standardized residuals are calculated as follows:

$$D_i = \beta_0 + \beta_1 X_i + \beta_2 I_i + \varepsilon_i \tag{8}$$

where $i=1 \ldots N$ indicates cross-sectional units of analysis (individuals), D specifies the individual commuting distance during commuting spells, I indicates an instrument and X represents a generic set of control variables. I_i are residuals to be applied in the second stage when estimating the MNL to correct for the endogeneity bias in the commuting distance. Geraci et al (2014) focused their attention on the application of the standardized form of the residuals when performing RI to address the endogeneity correction in the multinomial logit models.

To capture the endogeneity of distance, we constructed a synthetic variable that reflects the lagged shortest distance to the nearest place of work of the same type as the individual's first-year sector of employment. In other words, in the conditions of non-random and spatially separated distributions of places of work and residence and individuals' preference-based specialization in particular sectors due to the possession of sector-specific skills that maximize their return to employment, the shortest distance should directly influence the actual commuting patterns, since individuals would choose to minimize the commuting distance subject to a wage offer. The assumption on the existence of a spatial separation between the place of work and the place of residence is confirmed by Wasmer and Zenou (2002, 2006) and Zenou (2009). In addition, Allen and van der Velden (2001) indicate that skill mismatch leads to a significant penalty, suggesting the existence of sector-specific skills and knowledge. The lags of 0, 1, 2, 3 and 4 years used in the past for the instruments ensure that individuals had an opportunity to make a choice in the conditions of specific spatial configurations of the places of work and residence and adjust their commuting distance accordingly. Hence, the orthogonality conditions of the instrument should be met. Table 3 shows the estimates of the distance equation. The lagged spatial mismatch variable emerges as highly significant in the distance equation.

Table 3. Estimation of the correcting term with the application of different lags of instrument.

Variables	*Lag of instrument used in the analysis*			
	Lag 0	*Lag 1*	*Lag 2*	*Lag 3*
Gender	*0.26412****	*0.27396****	*0.28661****	*0.2974****
	(-0.00392)	*(-0.00438)*	*(-0.0045)*	*(-0.00466)*
Age	*0.04213****	*0.03825****	*0.03861****	*0.03403****
	(-0.00145)	*(-0.00164)*	*(-0.00169)*	*(-0.00178)*
Squared Age	*-0.00046****	*-0.00041****	*-0.00042****	*-0.00038****
	(-0.00001)	*(-0.00002)*	*(-0.00002)*	*(-0.000002)*
Private services	*-0.29564****	*-0.15105****	*-0.14700****	*-0.58341****
	(-0.008255)	*(-0.00915)*	*(-0.00937)*	*(-0.00727)*
Retailing	*-0.12819****	*-0.10329****	*-0.08159****	*-0.0691****
	(-0.00651)	*(0.00725)*	*(-0.00751)*	*(-0.00783)*
Construction	*-0.17686****	*-0.10951****	*-0.08455****	*-0.06807****
	(-0.00907)	*(-0.010002)*	*(-0.01029)*	*(-0.01064)*
Manufacture	*-0.62628****	*-0.55072****	*-0.51593****	*-0.49097****
	(-0.00669)	*(-0.00748)*	*(-0.00771)*	*(-0.00799)*

Table 3: Continued

Variables	*Lag of instrument used in the analysis*			
	Lag 0	*Lag 1*	*Lag 2*	*Lag 3*
Single parent	*-0.08910****	*-0.08123****	*-0.10287****	*-0.11629****
	(-11.3)	*(-0.00872)*	*(-0.0088)*	*(-0.00889)*
Living with parents	*0.3046****	*0.3143****	*0.3236****	*0.36382****
	(-0.00812)	*(-0.00896)*	*(-0.00912)*	*(-0.00536)*
Single	*0.08076****	*0.03034****	*-0.01004****	*-0.04054****
	(-0.00485)	*(-0.00523)*	*(-0.00528)*	*(-0.00536)*
Instrument (shortest "sector-distance")	*0.32543****	*0.28492****	*0.25572****	*0.24079****
	(-0.0017)	*(-0.00187)*	*(-0.0019)*	*(-0.00202)*
Gymnasium level of education	*0.19647****	*0.17029****	*0.15267****	*0.11727****
	(-0.00499)	*(-0.00571)*	*(-0.00605)*	*(-0.0066)*
Post-secondary education <2 years	*0.55955****	*0.49467****	*0.4814****	*0.42928****
	(-0.00829)	*(-0.00909)*	*(-0.00928)*	*(-0.00971)*
Post-secondary education >2 years	*0.5213****	*0.46022****	*0.43843****	*0.3945****
	(-0.00616)	*(-0.00684)*	*(-0.00707)*	*(-0.00748)*
MSc or PhD	*0.65066****	*0.56413****	*0.55382****	*0.50855****
	(-0.00883)	*(-0.02406)*	*(-0.0243)*	*(-0.00202)*
Const	*-0.65731****	*-0.36245****	*-0.22684****	*-0.03371*
	(-0.02841)	*(-0.03287)*	*(-0.03451)*	*(-0.00202)*
Number of observations	*732734*	*591620*	*570525*	*539478*
Adjusted R2	*0.09*	*0.0821*	*0.0751*	*0.0712*

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the log of commuting distance.
The standard errors are heteroskedasticity robust and clustered at the individual level

Therefore, equation 2 transforms into:

$$C_i = \beta_{0m} + D\beta_{1m} + \hat{\varepsilon}\beta_{3m} + W\beta_{3m} + X\beta_{jm} + \omega_i \quad (9)$$

where *i=1…N* indicates cross-sectional units of analysis, C denotes competing risks i.e. exits from the commuting spells, D is individual commuting distance corrected from endogeneity with the residuals $\hat{\varepsilon}$ derived from commuting distance equation, W is "firm rent" (firm specific fixed effects that are exogenous from the individual prospective), X denotes a generic set of control variables and ω is individual error term.

3.5 Results

The results from the estimation of the competing risk model are presented in Table 5 without correcting for endogeneity and in Table 6 correcting for endogeneity for the main variables of interest. All the coefficients in Tables 5 and 6 are presented in the form of marginal effects. The first and second columns indicate marginal probabilities for migration further away and closer to the place of work. The third column presents separation to non-employment. The fourth column reports re-employment closer to the place of residence, and the last column represents re-employment further away from the place of residence.

3.5.1 Competing risk model without controlling for endogeneity

The marginal effects of main interest are those of distance and income on the probability of experiencing a hazard from commuting. The results in the first and second columns in Table 4 suggest that a 10% increase in the commuting distance between an individual's place of residence and his or her place of work would lead to a 0.8 percentage points (p.p. hereafter) decrease in the probability of moving away and a 1.8 p.p. increase in the probability of moving closer. At the same time, it would lead to a 4.4 p.p. increase in the probability of separation to non-employment, a 25.4 p.p. increase in the probability of re-employment closer and a 12.5 p.p. decrease in the probability of changing job further away from the place of residence.

Table 4. Competing risk model (without correcting for endogeneity)

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Gender	*-0.00024***	*-0.00028***	*0.00096***	*0.00051*	*0.01363****
	(-0.00008)	*(-0.00008)*	*(-0.0002)*	*(-0.00035)*	*(-0.00035)*
Age	*-0.00015****	*-0.00012****	*0.0005****	*-0.00108****	*-0.0010****
	(-0.000005)	*(-0.000005)*	*(-0.00001)*	*(-0.00001)*	*(-0.00001)*
Nationality	*-0.00026**	*0.00009*	*0.00452****	*0.00711****	*-0.00065*
	(-0.00012)	*(-0.00012)*	*(-0.00041)*	*(-0.0005)*	*(-0.00052)*
Children	*-0.00126****	*0.00027***	*-0.00702****	*-0.0003*	*0.00168****
	(-0.0001)	*(-0.0001)*	*(-0.00033)*	*(-0.00038)*	*(-0.00039)*
Log (distance)	*-0.00077****	*0.00181****	*0.0044****	*0.02542****	*-0.01249****
	(-0.00002)	*(-0.00002)*	*(-0.00008)*	*(-0.00009)*	*(-0.0001)*
Log of (earnings)	*0.00078****	*0.00101****	*-0.01159****	*-0.01047****	*-0.01025****
	(-0.00005)	*(-0.00005)*	*(-0.00013)*	*(-0.00017)*	*(-0.00017)*

Table 4: Continued

Variables	Type of hazard				
	Migration further	Migration closer	Separation to non-employment	Reemployment closer	Reemployment further
Private services	*0.00062****	*-0.00013*	*-0.01687****	*-0.00205*	*0.00543****
	(-0.00012)	*(-0.00018)*	*(-0.00056)*	*(-0.00056)*	*(-0.00058)*
Retailing	*0.00067****	*-0.0001*	*-0.01585****	*-0.00549****	*0.00547****
	(-0.00009)	*(-0.0001)*	*(-0.00038)*	*(-0.00043)*	*(-0.00044)*
Construction	*0.00065****	*-0.00072****	*0.00143**	*-0.01691****	*-0.00014*
	(-0.00017)	*(-0.00017)*	*(-0.00058)*	*(-0.00075)*	*(-0.00076)*
Manufacture	*0.00016*	*0.00046****	*-0.00952****	*-0.01904****	*-0.01034****
	(-0.00011)	*(-0.0001)*	*(-0.0004)*	*(-0.00049)*	*(-0.0005)*
Previous income	*-0.00005*	*-0.00044****	*-0.0042****	*0.00128****	*0.000166****
	(-0.00003)	*(-0.00003)*	*(-0.00011)*	*(-0.00015)*	*(-0.00015)*
Unemployment experience	*0.0009*	*-0.00005*	*0.0033****	*0.00048*	*0.00042*
	(-0.00008)	*(-0.00008)*	*(-0.00028)*	*(-0.00034)*	*(-0.00034)*
Single parent	*0.00158****	*0.00213****	*0.00413****	*0.00902****	*0.00573****
	(-0.00017)	*(-0.00017)*	*(-0.00051)*	*(-0.00066)*	*(-0.00064)*
Living with parents	*-0.00017*	*0.00138****	*0.01447****	*-0.00374****	*0.01314****
	(-0.00013)	*(-0.00014)*	*(-0.00064)*	*(-0.00064)*	*(-0.00073)*
Single	*0.00121****	*0.00295****	*0.00952****	*0.00955****	*0.00715****
	(-0.00009)	*(-0.0001)*	*(-0.00033)*	*(-0.00039)*	*(-0.00039)*
Log of time	*-0.00078****	*-0.0011****	*-0.02632****	*-0.06688****	*-0.0619****
	(-0.00003)	*(-0.00004)*	*(-0.00016)*	*(-0.00023)*	*(-0.00022)*
Gymnasium level of education	*0.00017*	*0.00029**	*-0.00507****	*-0.00103**	*0.00276****
	(-0.00003)	*(-0.00004)*	*(-0.00039)*	*(-0.00051)*	*(-0.00052)*
Post-secondary education <2 years	*0.0007****	*-0.00054***	*-0.00042****	*0.00146**	*0.01251****
	(-0.00017)	*(-0.00018)*	*(-0.0006)*	*(-0.00073)*	*(-0.00074)*
Post-secondary education >2 years	*0.00068****	*0.00032**	*-0.01279****	*0.00457****	*0.01352****
	(-0.00014)	*(-0.00014)*	*(-0.00045)*	*(-0.00056)*	*(-0.00057)*
MSc or PhD	*0.00142****	*0.00101***	*-0.0011*	*0.00411**	*0.00413**
	(-0.00039)	*(-0.00038)*	*(-0.00148)*	*(-0.00177)*	*(-0.00189)*
Number of observations				*1801188*	
Overall p-value				*0.00000*	
Pseudo R^2				*0.0858*	

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the multinomial variable denoting the competing risks.

The standard errors are bootstrapped

The results on commuting distance on the probability of reemployment closer to the place of residence can potentially encapsulate the mechanical effects of distance i.e. any accepted workplace can be located closer to the place of residence. This reason could lead to the potential inflation in magnitude of coefficients on distance on the probability of reemployment closer with evident increasing effect. To investigate this issue, we split our samples by the quartiles of commuting distance distribution and conduct a separate analysis for all subsamples. Our results demonstrated in Table 5 indicate that there is no significant increasing effect of distance on the probability of reemployment closer for the individuals who commute less than 15 kilometers and constitute 3 quartiles of commuting distribution. Hence, we can conclude that, the coefficients on commuting distance reported in the Table 4 are not distorted by the pure mechanization in the job search process.

Table 5. Investigation of the mechanical effect of commuting distance

Commuting distance	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment closer*
> 4 km	*-0.00040*	*-0.00014*	*-0.00053*	*0.02542****	*-0.01445****
	(0.00052)	*(0.00020)*	*(0.00159)*	*(0.00178)*	*(0.00211)*
between 4 km and 7.5 km	*-0.00140***	*0.00553****	*0.00933****	*0.02221****	*-0.00558**
	(0.00054)	*(0.00043)*	*(0.00177)*	*(0.00214)*	*(0.00223)*
between 7.5 km and 15 km	*-0.00127****	*0.00396****	*-0.00225*	*0.02338****	*-0.01266****
	(0.00036)	*(0.00051)*	*(0.00147)*	*(0.00187)*	*(0.00176)*
> 15 km	*-0.00054****	*0.00123****	*0.01150****	*0.03910****	*-0.01008****
	(0.00013)	*(0.00022)*	*(0.00057)*	*(0.00077)*	*(0.00064)*

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the multinomial variable denoting the competing risks.
The standard errors are bootstrapped

Furthermore, income has a positive, significant impact on the probability of changing residence both closer to and further away from the place of work. The results in columns 1 and 2 of Table 4 suggest that a 10% increase in income is associated with a 0.78 p.p. increase in the probability of migration further away from the place of work and a 1 p.p. increase in the probability of migrating closer to the place of work. The probabilities of separation to non-employment and re-employment are negatively affected by the level of earnings. In this case a 10% increase in the earnings leads to an 11.6 p.p. decrease in the probability of separation to non-employment and to about a 10 p.p. decrease in the probability of re-employment closer and the probability of re-employment further away.

The results suggest that men have a lower probability of migration either closer or further away than females. By contrast, men have a higher probability of separation either to non-employment or to another place of work. That supports the previous findings that indicate higher levels of workplace attachment among females. Age has a significantly negative impact on all the types of labour force mobility except for separation to non-employment. The results show that individuals of foreign origin have a lower probability of migrating further away and re-employment further away than natives. Individuals with an origin other than Swedish instead tend to have a higher probability of moving closer to their place of work, changing job closer to their place of residence and separating to non-employment. These results suggest that foreigners have lower commuting tolerance than natives.

The presence of children below 18 years somewhat reduces the labour market mobility and stimulates migration closer to the place of work. The marginal effects of employment in the different sectors indicate the spatial configuration of the sectors and possible possession of sector-specific skills. Employment in the private services sector and retailing has a positive impact on the probability of migration and re-employment further away, whereas the impact on migration closer, separation to non-employment and re-employment closer is negative. The reason for that may lie in the spatial configuration of the sectors. Employment in manufacturing has a positive impact on the probability of migration further away and closer to the place of work. Furthermore, employment in manufacturing has a negative impact on the probability of separation to non-employment and re-employment. Previous income has a negative impact on the probability of migration and separation to non-employment and a positive impact on re-employment. The previous experience of unemployment has a positive (although insignificant) impact on the probabilities of experiencing all types of hazards. The marginal effects on the family-related variables suggest that single individuals are more prone to change the place of work and residence. Education has a positive impact on migration further away and migration closer to the place of work and re-employment. Quite expectedly, education has a negative impact on the probability of separation to non-employment.

3.5.2 Competing risk model corrected for endogeneity

As discussed before, the "firm rent" is introduced into the main model to overcome the problem of endogeneity bias in the earnings. The results with the application of the one-year-lagged RI term are presented in Table 6. At the same time, competing risk models with the application of correcting terms for 0-period, 2-period and 3-period lags are demonstrated in Appendix 1A, Appendix 1B and Appendix 1C, respectively.

Table 6. Competing risk model with correction for endogeneity of the earnings and commuting distance

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Gender	*0.00029***	*-0.00043****	*0.00133***	*0.00578****	*0.01035****
	(-0.0007)	*(-0.0001)*	*(-0.00039)*	*(-0.00046)*	*(-0.00048)*
Age	*0.00133****	*-0.00008****	*0.00059****	*-0.00091****	*-0.00083****
	(-0.0000049)	*(-0.0000053)*	*(-0.00002)*	*(-0.00002)*	*(-0.00002)*
Nationality	*0.00006*	*0.00036**	*0.0055****	*0.00604****	*0.00072*
	(-0.00013)	*(-0.00014)*	*(-0.00051)*	*(-0.00065)*	*(-0.00067)*
Children	*-0.00077****	*0.00037****	*-0.00578****	*0.00034*	*0.00091**
	(-0.00009)	*(-0.00009)*	*(-0.00034)*	*(-0.00041)*	*(-0.00043)*
Log of commuting distance	*-0.00157****	*0.00020****	*0.00457****	*0.01079****	*-0.000298*
	(-0.00012)	*(-0.00011)*	*(-0.00002)*	*(-0.00053)*	*(-0.00055)*
Correcting term	*0.00158****	*-0.00081****	*-0.00095*	*0.01997****	*-0.01545****
	(-0.00019)	*(-0.00017)*	*(-0.00065)*	*(-0.00081)*	*(-0.00087)*
Log Firm rent	*0.00091****	*0.0003***	*-0.01789****	*-0.01630****	*-0.00542****
	(-0.00019)	*(-0.0001)*	*(-0.00037)*	*(-0.00046)*	*(-0.00052)*
Private services	*0.00046***	*-0.00022*	*-0.01197****	*0.00655****	*0.00308****
	(-0.00015)	*(-0.00014)*	*(-0.0007)*	*(-0.00071)*	*(-0.00074)*
Retailing	*0.00075****	*-0.00044****	*-0.00867****	*0.00079*	*0.00387****
	(-0.00012)	*(-0.00015)*	*(-0.00052)*	*(-0.00059)*	*(-0.00061)*
Construction	*0.00069****	*-0.00105****	*0.00296****	*-0.01021****	*-0.00562****
	(-0.00018)	*(-0.00019)*	*(-0.00065)*	*(-0.00088)*	*(-0.00087)*
Manufacture	*-0.00029****	*-0.00012*	*-0.00309****	*-0.00922****	*-0.00885****
	(-0.00012)	*(-0.00012)*	*(-0.0005)*	*(-0.00059)*	*(-0.00063)*
Previous income	*0.000001*	*-0.00011**	*-0.00642****	*-0.00155****	*-0.00104****
	(-0.00005)	*(-0.00005)*	*(-0.00016)*	*(-0.00023)*	*(-0.00025)*
Unemployment experience	*0.00028***	*-0.00015*	*0.00548****	*0.00322****	*0.00556****
	(-0.00009)	*(-0.0001)*	*(-0.00035)*	*(-0.00045)*	*(-0.00043)*

Table 6: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Single parent	*0.00165****	*0.0025****	*0.00437****	*0.00573****	*0.00849****
	(-0.00019)	*(-0.00021)*	*(-0.00062)*	*(-0.00077)*	*(-0.00076)*
Living with parents	*0.00013*	*0.00198****	*0.00932****	*-0.00357***	*0.00412***
	(-0.00023)	*(-0.00028)*	*(-0.00119)*	*(-0.00117)*	*(-0.00129)*
Single	*0.0016****	*0.00285****	*0.00798****	*0.0066****	*0.00513****
	(-0.00011)	*(-0.00012)*	*(-0.00041)*	*(-0.00047)*	*(-0.00047)*
Log of time	*-0.00052****	*-0.00074****	*-0.02459****	*-0.05769****	*-0.0512****
	(-0.00004)	*(-0.00004)*	*(-0.0002)*	*(-0.00027)*	*(-0.00027)*
Gymnasium level of education	*0.00018*	*0.000012*	*-0.00546****	*-0.00028*	*0.00127**
	(-0.00013)	*(-0.00014)*	*(-0.00046)*	*(-0.0006)*	*(-0.00061)*
Post-secondary education <2 years	*0.000720****	*-0.00046**	*-0.00472****	*0.00608****	*0.00772****
	(-0.0002)	*(-0.00021)*	*(-0.00077)*	*(-0.00092)*	*(-0.00094)*
Post-secondary education >2 years	*0.00088****	*0.00037**	*-0.01359****	*0.00593****	*0.00774****
	(-0.00015)	*(-0.00015)*	*(-0.00055)*	*(-0.00068)*	*(-0.00069)*
MSc or PhD	*0.00182****	*0.00088**	*-0.00868****	*0.00317**	*-0.00275*
	(-0.00037)	*(-0.0004)*	*(-0.00163)*	*(-0.00189)*	*(-0.00203)*
	Number of observations		*1801188*		
	Overall p-value		*0.00000*		
	Pseudo R^2		*0.0858*		

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the multinomial variable denoting the competing risks.

The standard errors are bootstrapped

The results indicate that (log) "firm rent" positively affects the migration decision both further away from the place of work and closer to the place of work. At the same time, it has a negative impact on the decision to separate from the job place. The results suggest that a 10% increase in the firm earnings leads to a 0.9 p.p. increase in the probability of moving further away from the place of work and a 0.3 p.p. increase in the probability of moving closer to the place of work. At the same time, a 10% increase in the "firm premium" leads to a 17.9 p.p. decrease in the probability of separation to non-employment, a 1.3 p.p. decrease in the probability of re-employment closer and a 5.4 p.p. decrease in the probability of moving further away.

The results for commuting distance free of endogeneity bias suggest that individuals' tendencies to move further away and find re-employment further away are negatively associated with unobservable factors affecting the commuting distance. The results indicate that a 10% increase in the commuting distance decreases the probability of moving further away from the place of work by 1.6 p.p. and the probability of re-employment further away by 0.3 p.p. Meanwhile, the above-mentioned increase in the commuting distance leads to a 0.2 p.p. increase in the probability of migration closer to the place of work, a 4.6 p.p. increase in the probability of separation to non-employment and a 10.8 p.p. increase in the probability of re-employment closer to the place of residence.

Table 7. Comparison of estimates with and without correcting for the endogeneity of the individual earnings and commuting distances

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Estimates without correction for endogeneity					
Log (distance)	*-0.00077****	*0.00181****	*0.0044****	*0.02542****	*-0.01249****
	(-0.00002)	*(-0.00002)*	*(-0.00008)*	*(-0.00009)*	*(-0.0001)*
Log of (earnings)	*0.00078****	*0.00101****	*-0.01159****	*-0.01047****	*-0.01025****
	(-0.00005)	*(-0.00005)*	*(-0.00013)*	*(-0.00017)*	*(-0.00017)*
Estimates with correction for endogeneity					
Log (distance)	*-0.00157****	*0.00020****	*0.00457****	*0.01079****	*-0.000298*
	(-0.00012)	*(-0.00011)*	*(-0.00002)*	*(-0.00053)*	*(-0.00055)*
Firm rent	*0.00091****	*0.0003***	*-0.01789****	*-0.01630****	*-0.00542****
	(-0.00019)	*(-0.0001)*	*(-0.00037)*	*(-0.00046)*	*(-0.00052)*

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the multinomial variable denoting the competing risks.
The standard errors are bootstrapped

The other results indicate that gender has a positive impact on all the types of hazard except migrating closer. Age affects positively migration further away and separation to non-employment, while migration closer, re-employment closer and re-employment further away are negatively affected by age. Individuals of foreign origin have a higher probability of experiencing all the types of hazard. The results show that employment in private services and retailing increases the probability of migration further away and re-employment closer to and further away from the place of residence. By contrast, it decreases the probability of separation to non-employment and migration closer to the place of work. The probabilities of separation to non-employment and migration further away are positively affected by employment in the construction industry. On the contrary, the probabilities of migration closer, re-employment closer and re-employment further away are negatively influenced by employment in construction. Manufacturing exerts a negative impact on all the types of hazards. Previous income decreases the probability of experiencing all the exits. At the same time, unemployment experiences have a positive effect on the probabilities of experiencing all the types of exits. The dummy variables on the family status suggest that individuals with a family status other than marriage are more mobile and consequently have a higher probability of exiting the employment commuting spell. Education has a positive effect on the residential and labour market mobility, while the impact on unemployment is negative.

Table 8. Comparison of estimates for the single and coupled subsamples of population

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Single individuals					
Log (distance)	*-0.00179****	*0.00288****	*0.00352****	*0.01238****	*-0.00210**
	(0.00027)	*(0.00022)*	*(0.00076)*	*(0.00087)*	*(0.00096)*
Firm rent	*0.00132****	*0.00080****	*-0.02082****	*-0.01551****	*-0.00719****
	(0.00026)	*(0.0002)*	*(0.00061)*	*(0.00074)*	*(0.00088)*
Married individuals					
Log (distance)	*-0.00164****	*0.00138****	*0.00591****	*0.01252****	*-0.00096*
	(0.00016)	*(0.00014)*	*(0.0006)*	*(0.00077)*	*(0.0008)*
Firm rent	*0.00066****	*-0.00005*	*-0.01623****	*-0.01691****	*-0.00384****
	(0.00016)	*(0.00014)*	*(0.00052)*	*(0.00069)*	*(0.00079)*

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the multinomial variable denoting the competing risks.
The standard errors are bootstrapped

A comparison of the results with and without endogeneity correction indicates the significant role of unobservable factors in determining the influence of distance and earnings on the length of the employment commuting spells. Indeed, the results in Table 6 suggest the presence of unobserved traits that significantly inflate the probability of separation from the workplace and deflate the probabilities of migration with respect to distance. The estimated coefficients in Table 7 also indicate that the firm premium paid to workers plays a significantly larger role in the decision to migrate closer or further away or separate from the place of work. The application of the firm premium, free from the endogeneity bias, increases the impact of income on the probabilities of migration further away, separation to non-employment and re-employment closer to the place of work while reducing the impact of income on the probabilities of re-employment closer and migration further away. The correction for endogeneity in distance leads to an increase in the impact of distance on the probability of migration further away from the place of work while reducing the impact on the probabilities of selecting the other exits.

Since previous studies (Nakosteen et al, 2008; Åstrom & Westerlund, 2009) have demonstrated very different migration and commuting behaviour for married and unmarried individuals, we conducted similar analyses for the married and single subsamples. In the case of married individuals, we took into account the fact that the spouse's characteristics might play a significant role in determining the length of the commuting and employment behaviour of individuals by including the labour market characteristics of the spouses in the model. The estimates of the main regressors of interest for the individuals with and without partners are presented in Table 8. The output tables from the estimation of the competing risk model for the two subsamples of individuals are presented in Appendices B.1 and B.2.

The results clearly suggest that the earnings premium paid by employers and the commuting distance have a much smaller influence on the migration and employment decisions of individuals with partners than those of the single cohort of individuals.

3.6 Conclusions

The main aim of this paper was to analyse the role of earnings and commuting distance in the determination of the length of employment commuting spells in Sweden. We modelled five alternative ways to terminate these spells: migration closer to the place of work, migration further away, separation to non-employment, re-employment closer to the place of residence and re-employment further away from the place of residence. The data used in the analysis represent a combination of administrative registers carried by Statistics Sweden and the Swedish Tax Board. They contain precise administrative records about individual earnings and personal characteristics. The information is available on an annual basis. The period of study was selected to be from 2000 to 2009. Applying precise information on individuals, we estimated discrete time competing risk models using a multinomial logit setting. The endogeneity of earnings was accounted for by substituting earnings with the presumably exogenous "firm rent" paid to workers. At the same time, the endogeneity of distance was accounted for using two-stage residual inclusion and exploiting the distance to the closest sector-specific employment alternative as an instrument.

Our results indicate that, when addressing endogeneity, past earnings increase the likelihood of migrating closer to and further away from an individual's residence with respect to maintaining the same place of work. The potential explanations could include the possibility to afford better housing or dwellings in areas with a richer level of amenities. Besides that, earnings quite intuitively have a significant negative impact on the probability of separation and re-employment. Another important result is the evidence of a tendency of individuals to reduce their commuting; that is, the commuting distance has a significantly positive impact on the probabilities of re-employment and migration closer to the place of residence, while separation and migration further away are negatively affected by the commuting distance.

Our results provide interesting insights for employers and policy makers. On the one side, our paper establishes a link between the firm premium paid to workers and their attachment to the place of work. Furthermore, it allows policy makers to predict regional migration flows based on observable workers' and firms' characteristics.

References

Beck, N., Katz, J. N., and Tucker, R., (1998). Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science*, pp. 1260-1288.

Bratti,, M. (2003). Labour force participation and marital fertility of Italian women: The role of education. *Journal of Population economics*, 16(3), pp. 525-554.

Calfee, J. and Winston, C. (1998). The value of automobile travel time: implications for congestion policy. *Journal of Public Economics*, 69(1), pp 83-102.

Calthrop, E., De Borger, B. and Proost, S. (2010). Cost-benefit analysis of transport investments in distorted economies. *Transportation Research Part B: Methodological*, 44(7), pp. 850-869.

Cameron, A. C. and Trivedi, P. K. (2010). Microeconometrics using stata.

Card, D., Cardoso, A.R., Heining, J. and Kline, P. (2016). Firms and Labor Market Inequality: Evidence and Some Theory.

Clark, T. N. (2004). Urban amenities: lakes, opera, and juice bars: do they drive development?. *The city as an entertainment machine. research and urban policy*,9, pp. 103-140.

Cramer, H., Evers, V., Kemper, N., and Wielinga, B. (2008). Effects of autonomy, traffic conditions and driver personality traits on attitudes and trust towards in-vehicle agents. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM* (Vol. 3, pp. 477-482). IEEE.

D'Addio, A. C., and Rosholm, M. (2005). Exits from temporary jobs in Europe: a competing risks analysis. *Labour Economics*, *12*(4), pp. 449-468.

De Borger, B., and Fosgerau, M. (2008). The trade-off between money and travel time: A test of the theory of reference-dependent preferences. *Journal of Urban Economics*, 64(1), pp. 101-115.

Geraci, A., Fabbri, D., and Monfardini, C. (2014). Testing exogeneity of multinomial regressors in count data models: does two stage residual inclusion work?

Giménez, J. I., and Molina, J. A. (2015). Commuting Time and Household Responsibilities: Evidence Using Propensity Score Matching.

Gobillon, L., Selod, H., and Zenou, Y., (2007). The mechanisms of spatial mismatch. *Urban studies*, 44(12), pp. 2401-2427.

Gottholmseder, G., Nowotny, K., Pruckner, G. J., and Theurl E. (2009). Stress perception and commuting. *Health Economics, 18*, pp. 559–576.

Greene, W. H.(2008). *Econometric analysis*. Granite Hill Publishers.

Gutiérrez-i-Puigarnau,, E., and van Ommeren, J. N. (2010). Labour supply and commuting. *Journal of Urban Economics*, *68*(1), pp. 82-89.

Guevara, C. A., and Ben-Akiva, M. (2008) A Lagrange multiplier test for the validity of instruments in MNL models: An application to residential choice. In *EUROPEAN TRANSPORT CONFERENCE 2008; PROCEEDINGS*.

Fu, S., and Ross S. L. (2010). Wage Premium in Employment Clusters: Agglomeration or Worker Heterogeneity?. *US Census Bureau Center for Economic Studies Paper No. CES-WP-10-04*.

Eliasson, K., Lindgren, U. and Westerlund, O. (2003). Geographical labour mobility: migration or commuting?. *Regional Studies*, *37*(8), pp.827-837.

Jenkins, S. P. (2005). Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.*

Hansson, E., Mattisson, K., Björk, J., Östergren, P. O. and Jakobsson, K. (2011). Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden. *BMC Public Health*, *11*(1), pp. 834.

Hazans, M. (2004). Does commuting reduce wage disparities?. *Growth and Change*, *35*(3), pp. 360-390.

Hess, W. and Persson, M. (2012). The duration of trade revisited. *Empirical Economics*, *43*(3), pp. 1083-1107.

Hoehner, C. M., Carolyn E. B, Barlow, M. S., Allen P. and Schootman, M. (2012). Commuting distance, cardiorespiratory fitness, and metabolic risk. *American Journal of Preventive Medicine, 42*, pp. 571–578

Hämming, O., Gutzwiller. F. and Baurer G. (2009). "Work–life conflict and associations with work- and nonwork-related factors and with physical and mental health outcomes: a nationally representative cross-sectional study in Switzerland". *BMC Public Health*, *9,* p.435.

Koslowsky, M., Aizer, A. and Krausz, M. (1996). Stressor and personal variables in the commuting experience. *International Journal of Manpower,17*(3), pp.4-14.

Krueger, A. and Summers L. (1988). Efficiency wages and the inter-industry wage structure. *Econometrica,* 56, pp. 259–293.

Lewis, H.G. (1986) . *Union Relative Wage Effects: A Survey University of Chicago Press*, Chicago

Lyons ,G. and Urry J. (2005). Travel time use in the information age. *Transportation Research Part A*, 39 , pp. 257–276.

Lundholm, E. (2010). Interregional migration propensity and labour market size in Sweden, 1970–2001. *Regional studies*, *44*(4), pp. 455-464.

Manning, A. (2003). The real thin theory: monopsony in modern labour markets. *Labour Economics*, *10*(2),pp. 105-131.

Mincer, J. (1978). Family migration decisions. *Journal of Political Economy*, 86, pp.749–773.

Mulalic, I., Van Ommeren, J. N.and Pilegaard, N. (2014). Wages and Commuting: Quasi-natural Experiments' Evidence from Firms that Relocate. *The Economic Journal*, *124*(579),pp. 1086-1105.

van Ommeren, J. N. and Gutiérrez-i-Puigarnau, E. (2011). Are workers with a long commute less productive? An empirical analysis of absenteeism. *Regional Science and Urban Economics*, *41*(1), pp. 1-8.

Van Ommeren, J. and Rietveld P. (2002). *Commuting, spatial search and labour market bargaining*. No. 02-039/3. Tinbergen Institute Discussion Paper, Van Ommeren, J. and Fosgerau M. (2009). Workers' marginal costs of commuting. *Journal of Urban Economics* 65 (1), pp. 38-47.

Potepan, M. J.(1994). Intermetropolitan migration and housing prices: simultaneously determined?. *Journal of Housing Economics*, *3*(2), pp.77-91.

Plaut P. (2006). The intra-household choices regarding commuting and housing. *Transportation Research Part A* 40, pp. 561–571.

Renkov, M..and Hoover, D. (2000). Commuting, Migration and Rural-Urban Population Dynamics. *Journal of Regional Science, 40(2):*pp. 261-287.

Romaní, J., Suriñach, J. and Artiís, M. (2003). Are commuting and residential mobility decisions simultaneous?: The case of Catalonia, Spain. *Regional Studies*, *37*(8), pp. 813-826.

Ross, S. L., and Zenou, Y. (2008). Are shirking and leisure substitutable? An empirical test of efficiency wages based on urban economic theory. *Regional Science and Urban Economics*, *38*(5), pp. 498-517.

Rouwendahl, J.(1998). Search theory, Spatial Labour Markets and commuting. *Journal of urban Economics*, (43), pp.1–22.

Rouwendahl, J. (1999). Spatial job search and commuting distances. *Regional sciense and urban economics*, (29), pp. 491-517.

Rouwendal, J. (2004). Search Theory and Commuting Behavior. *Growth and Change*, 35(3), pp.391–418.

Ruppert, P., Stancanelli, E., and Wasmer, E. (2009). Commuting, wages and bargaining power. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, pp. 201-220.

Sabir, M., Ommeren, J., Koetse, M. J., and Rietveld, P. (2010). Impact of weather on daily travel demand. *Proceedings of the Tinbergen Institute discussion paper. VU University, Amsterdam*.

Sandow, E.(2013). Till work do us part: the social fallacy of long-distance commuting. *Urban Studies*

Sandow, E., Westerlund, O. and Lindgren, U.,(2014). Is your commute killing you? On the mortality risks of long-distance commuting. *Environment and Planning A*, *46*(6), pp.1496-1516.

Shea, J. (2000). Does parents' money matter?. *Journal of public Economics,77*(2), pp. 155-184.

So, K. S., Orazem, P. F. and Otto, D. M. (2001). The effects of housing prices, wages, and commuting time on joint residential and job location choices. *American Journal of Agricultural Economics*, *83*(4), pp. 1036-1048.

Shuai, X., (2012). Does Commuting Lead to Migration?. *The Journal of Regional Analysis & Policy* 42 (3).

Terza, J. V., Basu, A. and Rathouz, P. J. (2008). Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *Journal of Health Economics, 27*(3), pp. 531–543.

White, M. J. (1977). A model of residential location choice and commuting by men and women workers. *Journal of Regional Science*, *17*(1), pp. 41-52.

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, *182*(1), pp. 226-234.

Åström, J. and Westerlund, O. (2009). *Sex and migration: who is the tied mover?* (No. 787). Umeå University, Department of Economics.

Appendix A

Table A.1. Competing risk model with correction for endogeneity of the earnings and commuting distance with using the lag 0 correcting term.

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Gender	*0.0002**	*-0.00046****	*0.00152****	*0.00565****	*0.01118****
	(-0.00011)	*(-0.00010)*	*(-0.00040)*	*(-0.00046)*	*(-0.00048)*
Age	*-0.00009****	*-0.00008****	*0.0006096****	*-0.00087****	*-0.00079****
	(-0.000005)	*(-0.000005)*	*(-0.00002)*	*(-0.00002)*	*(-0.00002)*
Nationality	*0.00009*	*0.00037**	*0.00559****	*0.00531****	*0.00035*
	(-0.00014)	*(-0.00015)*	*(-0.00054)*	*(-0.00066)*	*(-0.00068)*
Children	*-0.00077****	*0.0004****	*-0.00569****	*0.00071**	*0.00107**
	(-0.00009)	*(-0.00009)*	*(-0.00035)*	*-0.0004*	*(-0.00042)*
Log of commuting distance	*-0.00129****	*0.00207****	*0.00381****	*0.0104****	*-0.0024****
	(-0.00011)	*(-0.00010)*	*(-0.0003)*	*(-0.00047)*	*(-0.00049)*
Correcting term	*0.00119****	*-0.0010****	*0.00053*	*0.02022****	*-0.01174****
	(-0.00018)	*(-0.00016)*	*(-0.00062)*	*(-0.00073)*	*(-0.0007)*
Firm rent	*0.00066****	*0.00027**	*-0.01746****	*-0.01629****	*-0.00541****
	(-0.00012)	*(-0.00010)*	*(-0.00038)*	*(-0.00047)*	*(-0.00055)*
Private services	*0.00045***	*-0.00016*	*-0.01219****	*0.00611****	*0.00448****
	(-0.00015)	*(-0.00015)*	*(-0.00074)*	*(-0.00073)*	*(-0.00075)*
Retailing	*0.00061****	*-0.00043***	*-0.00829****	*0.0011**	*0.00526****
	(-0.00012)	*(-0.00012)*	*(-0.0005)*	*(-0.00061)*	*(-0.00064)*
Construction	*0.00062***	*-0.00102****	*0.00318****	*-0.01012****	*-0.00508****
	(-0.00017)	*(-0.00019)*	*(-0.00066)*	*(-0.00089)*	*(-0.00090)*
Manufacture	*-0.00015*	*-0.00013*	*-0.00330****	*-0.00885****	*-0.00914****
	(-0.00012)	*(-0.00012)*	*(-0.00051)*	*(-0.00062)*	*(-0.00066)*
Previous income	*0.0000002*	*-0.00011**	*-0.00651****	*-0.00159****	*-0.00086***
	(-0.00005)	*(-0.00005)*	*(-0.00017)*	*(-0.00025)*	*(-0.00027)*
Unemployment experience	*0.00032****	*-0.000099*	*0.00552****	*0.00353****	*0.00567****
	(-0.00009)	*(-0.00010)*	*(-0.00038)*	*(-0.00046)*	*(-0.00045)*
Single parent	*0.00165****	*0.00264****	*0.0038****	*0.00486****	*0.00827****
	(-0.00018)	*(-0.0002)*	*(-0.00062)*	*(-0.00077)*	*(-0.00080)*
Living with parents	*0.00022*	*0.00209****	*0.00931****	*-0.00394***	*0.00518****
	(-0.00025)	*(-0.00031)*	*(-0.00122)*	*(-0.00123)*	*(-0.00140)*
Single	*0.00168****	*0.00292****	*0.00784****	*0.00645****	*0.00451****
	(-0.00011)	*(-0.00014)*	*(-0.00044)*	*(-0.00051)*	*(-0.00050)*
Log of time	*-0.00051****	*-0.00073****	*-0.02477****	*-0.05647****	*-0.04988****
	(-0.00004)	*(-0.00004)*	*(-0.00020)*	*(-0.00026)*	*(-0.00027)*

Table A.2: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Gymnasium level of education	*0.00013*	*0.00009*	*-0.00549****	*0.00032*	*0.00128**
	(-0.00013)	*(-0.00014)*	*(-0.00047)*	*(-0.00063)*	*(-0.0006)*
Post-secondary education <2 years	*0.00060***	*-0.00055**	*-0.00444****	*0.00736****	*0.00801****
	(-0.00019)	*(-0.00021)*	*(-0.00078)*	*(-0.00094)*	*(-0.00097)*
Post-secondary education >2 years	*0.00078****	*0.00031**	*-0.01348****	*0.00677****	*0.00817****
	(-0.00015)	*(-0.00016)*	*(-0.00057)*	*(-0.0006)*	*(-0.0007)*
MSc or PhD	*0.00146***	*0.00076**	*-0.00806****	*0.00462**	*-0.0023*
	(-0.00042)	*(-0.00041)*	*(-0.00173)*	*(-0.00202)*	*(-0.00216)*
Number of observations			*1801188.00*		
Overall p-value			*0.0000*		
Pseudo R2			*0.08*		

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the multinomial variable denoting the competing risks.

The standard errors are bootstrapped

Table A.2. Competing risk model with correction for endogeneity of the earnings and commuting distance with using the lag 2 correcting term.

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Gender	*0.00027***	*-0.00039****	*0.00174****	*0.0055****	*0.00967****
	(-0.00010)	*(-0.00011)*	*(-0.00042)*	*(-0.00049)*	*(-0.0005)*
Age	*-0.00009****	*-0.00008****	*0.00065****	*-0.00087****	*-0.0008****
	(-0.00000)	*(-0.000005)*	*(-0.00002)*	*(-0.0000)*	*(-0.00002)*
Nationality	*0.00004*	*0.00045***	*0.00554****	*0.00568****	*0.00104*
	(-0.00014)	*(-0.00015)*	*(-0.00058)*	*(-0.00068)*	*(-0.00071)*
Children	*-0.0007****	*0.00032***	*-0.0055****	*0.00044*	*0.00112**
	(-0.00009)	*(-0.00010)*	*(-0.00036)*	*(-0.00043)*	*(-0.00044)*
Log of commuting distance	*-0.00125****	*0.00190****	*0.00392****	*0.0117****	*0.00002**
	(-0.00013)	*(-0.0001)*	*(-0.00049)*	*(-0.00058)*	*(-0.00063)*
Correcting term	*0.00117****	*-0.00066***	*0.00022*	*0.01784****	*-0.01514****
	(-0.00021)	*(-0.00019)*	*(-0.00075)*	*(-0.00091)*	*(-0.00099)*

Table A.2: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Firm rent	*0.00075****	*0.00022**	*-0.01710****	*-0.01623****	*-0.00547****
	(-0.00013)	*(-0.00011)*	*(-0.00039)*	*(-0.00051)*	*(-0.00056)*
Private services	*0.00029**	*-0.00017*	*-0.01170****	*0.0061****	*0.00352****
	(-0.00015)	*(-0.00015)*	*(-0.00076)*	*(-0.00074)*	*(-0.0007)*
Retailing	*0.0006***	*-0.00048****	*-0.00776****	*0.00065*	*0.00385****
	(-0.00013)	*(-0.00012)*	*(-0.00052)*	*(-0.00064)*	*(-0.00065)*
Construction	*0.00058***	*-0.00100****	*0.00343****	*-0.00991****	*-0.00582****
	(-0.00018)	*(-0.00020)*	*(-0.00070)*	*(-0.00086)*	*(-0.00091)*
Manufacture	*-0.00032**	*-0.00012*	*-0.00307****	*-0.00842****	*-0.00843****
	(-0.00013)	*(-0.00012)*	*(-0.00052)*	*(-0.00062)*	*(-0.00065)*
Previous income	*-0.000006*	*-0.00016**	*-0.00741****	*-0.00187****	*-0.00046*
	(-0.00007)	*(-0.00006)*	*(-0.00024)*	*(-0.00031)*	*(-0.00033)*
Unemployment experience	*0.00023**	*-0.00005*	*0.00489****	*0.00302****	*0.00626****
	(-0.00009)	*(-0.00011)*	*(-0.00039)*	*(-0.0004)*	*(-0.00046)*
Single parent	*0.00159****	*0.00247****	*0.00425****	*0.00484****	*0.00817****
	(-0.0001)	*(-0.00022)*	*(-0.00061)*	*(-0.00081)*	*(-0.00083)*
Living with parents	*0.0001*	*0.00213****	*0.00899****	*-0.00302**	*0.00262**
	(-0.00024)	*(-0.00032)*	*(-0.00124)*	*(-0.00131)*	*(-0.00133)*
Single	*0.00164****	*0.0028****	*0.00766****	*0.0058****	*0.00492****
	(-0.00011)	*(-0.00013)*	*(-0.0004)*	*(-0.00053)*	*(-0.0005)*
Log of time	*-0.00049****	*-0.0007****	*-0.02428****	*-0.05542****	*-0.04885****
	(-0.00004)	*(-0.00004)*	*(-0.00021)*	*(-0.0002)*	*(-0.00027)*
Gymnasium level of education	*0.00013*	*0.00015*	*-0.00489****	*-0.00064*	*0.00163**
	(-0.00013)	*(-0.00014)*	*(-0.00047)*	*(-0.00062)*	*(-0.00064)*
Post-secondary education <2 years	*0.00058***	*-0.00044**	*-0.00434****	*0.0052****	*0.00840****
	(-0.00019)	*(-0.00022)*	*(-0.00082)*	*(-0.00095)*	*(-0.00098)*
Post-secondary education >2 years	*0.00075****	*0.00038**	*-0.01226****	*0.00521****	*0.00789****
	(-0.00015)	*(-0.00016)*	*(-0.00058)*	*(-0.00070)*	*(-0.00074)*
MSc or PhD	*0.0014***	*0.00095**	*-0.0077****	*0.00387**	*-0.0025*
	(-0.00041)	*(-0.00042)*	*(-0.00175)*	*(-0.00210)*	*(-0.00221)*
Number of observations			*1801188*		
Overall p-value			*0.00000*		
Pseudo R2			*0.0819*		

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the multinomial variable denoting the competing risks.

The standard errors are bootstrapped

Table A.3. Competing risk model with correction for endogeneity of the earnings and commuting distance with using the lag 3 correcting term.

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Gender	*0.00019**	*-0.00042****	*0.00199****	*0.005497****	*0.00922****
	(-0.00010)	*(-0.00012)*	*(-0.00043)*	*(-0.0004)*	*(-0.00056)*
Age	*-0.00009****	*-0.00008****	*0.00067****	*-0.00085****	*-0.0008****
	(-0.000005)	*(-0.000005)*	*(-0.00002)*	*(-0.00002)*	*(-0.00002)*
Nationality	*0.000006*	*0.00049***	*0.0054****	*0.00609****	*0.0004*
	(-0.00015)	*(-0.00016)*	*(-0.00057)*	*(-0.00073)*	*(-0.00077)*
Children	*-0.00066****	*0.00033***	*-0.0057****	*0.0005*	*0.00097**
	(-0.00010)	*(-0.00010)*	*(-0.00036)*	*(-0.00044)*	*(-0.00044)*
Log of commuting distance	*-0.0011****	*0.00186****	*0.00394****	*0.01224****	*0.00114*
	(-0.00015)	*(-0.00014)*	*(-0.00054)*	*(-0.00066)*	*(-0.00071)*
Correcting term	*0.0009****	*-0.00061***	*0.00033*	*0.0166****	*-0.01679****
	(-0.00023)	*(-0.00021)*	*(-0.00085)*	*(-0.00103)*	*(-0.00113)*
Firm rent	*0.00076****	*0.00011*	*-0.01706****	*-0.01639****	*-0.00508****
	(-0.00014)	*(-0.00011)*	*(-0.00040)*	*(-0.0005)*	*(-0.00058)*
Private services	*0.0002*	*-0.00007*	*-0.01194****	*0.00546****	*0.00304****
	(-0.00016)	*(-0.00016)*	*(-0.00075)*	*(-0.00075)*	*(-0.00082)*
Retailing	*0.00051****	*-0.00036***	*-0.00767****	*0.00098*	*0.0036****
	(-0.00012)	*(-0.00013)*	*(-0.00055)*	*(-0.0006)*	*(-0.00068)*
Construction	*0.00041**	*-0.00097****	*0.00336****	*-0.00954****	*-0.0070****
	(-0.00018)	*(-0.00020)*	*(-0.00068)*	*(-0.00094)*	*(-0.00093)*
Manufacture	*-0.00032***	*-0.0001558*	*-0.00306****	*-0.00794****	*-0.00855****
	(-0.00012)	*(-0.00012)*	*(-0.00054)*	*(-0.00065)*	*(-0.00068)*
Previous income	*0.00004*	*-0.00012**	*-0.00739****	*-0.00204****	*-0.0005*
	(-0.00007)	*(-0.00007)*	*(-0.00025)*	*(-0.00035)*	*(-0.00036)*
Unemployment experience	*0.00027***	*-0.00007*	*0.00494****	*0.00332****	*0.0062****
	(-0.00010)	*(-0.00011)*	*(-0.00039)*	*(-0.00048)*	*(-0.00048)*
Single parent	*0.00186****	*0.0026****	*0.00401****	*0.00507****	*0.00839****
	(-0.00021)	*(-0.00022)*	*(-0.00067)*	*(-0.00084)*	*(-0.0008)*
Living with parents	*0.00021*	*0.0022****	*0.00834****	*-0.00219*	*0.0014*
	(-0.00026)	*(-0.00033)*	*(-0.00133)*	*(-0.00133)*	*(-0.00134)*
Single	*0.00161****	*0.00278****	*0.00766****	*0.00545****	*0.00452****
	(-0.00011)	*(-0.00014)*	*(-0.00046)*	*(-0.00052)*	*(-0.00053)*
Log of time	*-0.00049****	*-0.00068****	*-0.02391****	*-0.0547****	*-0.048235****
	(-0.00004)	*(-0.00005)*	*(-0.00020)*	*(-0.00029)*	*(-0.00027)*

Table A.3: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Reemployment closer*	*Reemployment further*
Gymnasium level of education	*0.00011*	*0.00016*	*-0.00436****	*-0.00117**	*0.00191***
	(-0.00013)	*(-0.00014)*	*(-0.00048)*	*(-0.00062)*	*(-0.00064)*
Post-secondary education <2 years	*0.0004**	*-0.00042**	*-0.00393****	*0.00449****	*0.0083****
	(-0.00020)	*(-0.00023)*	*(-0.00083)*	*(-0.00095)*	*(-0.00101)*
Post-secondary education >2 years	*0.00066****	*0.00043**	*-0.01151****	*0.00479****	*0.00811****
	(-0.00015)	*(-0.00016)*	*(-0.00058)*	*(-0.00072)*	*(-0.00076)*
MSc or PhD	*0.00143***	*0.00108***	*-0.006967****	*0.00456**	*-0.00252*
	(-0.0004)	*(-0.00041)*	*(-0.00170)*	*(-0.00208)*	*(-0.00222)*
Number of observations			*1801188*		
Overall p-value			*0.00000*		
Pseudo R2			*0.0807*		

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the multinomial variable denoting the competing risks.

The standard errors are bootstrapped

Appendix B

Table B.1. Competing risk model with correction for endogeneity of the earnings and commuting distance estimated on the single subsample of population.

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Re-employment closer*	*Re-employment further*
Gender	*-0.00015*	*-0.00059***	*0.00258****	*0.00356****	*0.00914****
	(0.00022)	*(0.00021)*	*(0.00076)*	*(0.00084)*	*(0.00087)*
Age	*-0.00018****	*-0.00017****	*0.00053****	*-0.00106****	*-0.00095****
	(0.00001)	*(0.00001)*	*(0.00003)*	*(0.00004)*	*(0.00004)*
Nationality	*-0.0001*	*0.00012*	*0.00497****	*0.00620****	*0.00234**
	(0.00029)	*(0.00032)*	*(0.00097)*	*(0.0011)*	*(0.00116)*
Children	*-0.00008*	*0.00107****	*-0.00311****	*0.00076*	*0.00449****
	(0.00027)	*(0.00026)*	*(0.00086)*	*(0.00096)*	*(0.00097)*
Log of commuting distance	*-0.00179****	*0.00288****	*0.00352****	*0.01238****	*-0.00210**
	(0.00027)	*(0.00022)*	*(0.00076)*	*(0.00087)*	*(0.00096)*
Correcting term	*0.00152****	*-0.00112****	*0.00189*	*0.01969****	*-0.01487****
	(0.00041)	*(0.00034)*	*(0.0012)*	*(0.00136)*	*(0.00152)*
Firm rent	*0.00132****	*0.00080****	*-0.02082****	*-0.01551****	*-0.00719****
	(0.00026)	*(0.0002)*	*(0.00061)*	*(0.00074)*	*(0.00088)*
Private services	*0.00003*	*-0.00039*	*-0.01120****	*0.00896****	*0.00107*
	(0.00032)	*(0.0003)*	*(0.00122)*	*(0.00118)*	*(0.00122)*
Retailing	*0.00111****	*-0.00050**	*-0.01194****	*0.00197**	*0.00385****
	(0.00024)	*(0.00025)*	*(0.00094)*	*(0.00095)*	*(0.00105)*
Construction	*0.00144****	*-0.00153****	*0.0003*	*-0.01219****	*-0.00349**
	(0.00038)	*(0.0004)*	*(0.00133)*	*(0.00154)*	*(0.0015)*
Manufacture	*-0.00042*	*0.00014*	*-0.00544****	*-0.01000****	*-0.01138****
	(0.00025)	*(0.00024)*	*(0.0009)*	*(0.00102)*	*(0.00105)*
Previous income	*0.0001*	*-0.00021**	*-0.00803****	*-0.00214****	*-0.00296****
	(0.00011)	*(0.0001)*	*(0.00028)*	*(0.00036)*	*(0.00037)*
Unemployment experience	*-0.00002*	*-0.00034*	*0.00472****	*0.00387****	*0.00632****
	(0.0002)	*(0.0002)*	*(0.00063)*	*(0.00074)*	*(0.00078)*
Single parent	*-0.00034*	*0.00064**	*-0.01055****	*0.00282**	*0.00498****
	(0.00034)	*(0.00032)*	*(0.00103)*	*(0.00124)*	*(0.00131)*
Living with parents	*-0.00164****	*0.00032*	*-0.00322**	*-0.00715****	*0.00439**
	(0.00038)	*(0.00037)*	*(0.00146)*	*(0.00154)*	*(0.00173)*
Single	*0.00013*	*0.00112****	*-0.00484****	*0.00418****	*0.00489****
	(0.00029)	*(0.00026)*	*(0.00089)*	*(0.00099)*	*(0.00108)*

Table B.1: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Re-employment closer*	*Re-employment further*
Log of time	*-0.00084****	*-0.00153****	*-0.02525****	*-0.06489****	*-0.06002****
	(0.00009)	*(0.00009)*	*(0.00037)*	*(0.0005)*	*(0.00049)*
Gymnasium level of education	*0.00023*	*0.00019*	*-0.00627****	*-0.00057*	*0.00259**
	(0.00027)	*(0.00028)*	*(0.0008)*	*(0.00101)*	*(0.00104)*
Post-secondary education <2 years	*0.00090**	*-0.00059*	*-0.00502****	*0.00390**	*0.00844****
	(0.00039)	*(0.00043)*	*(0.00139)*	*(0.00158)*	*(0.00167)*
Post-secondary education >2 years	*0.00134****	*0.00092***	*-0.01654****	*0.00576****	*0.01134****
	(0.00031)	*(0.00032)*	*(0.00097)*	*(0.00112)*	*(0.00119)*
MSc or PhD	*0.00011*	*0.00182**	*-0.01414****	*0.00352*	*-0.00064*
	(0.0013)	*(0.00091)*	*(0.00358)*	*(0.00403)*	*(0.00453)*
Number of observations			*626933*		
Overall p-value			*0.00000*		
Pseudo R^2			*0.0978*		

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*
Note. The dependent variable is the multinomial variable denoting the competing risks.
The standard errors are bootstrapped

Table B.2. Competing risk model with correction for endogeneity of the earnings and commuting distance estimated on the married subsample of population with application of additional covariates for spouse employment status.

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Re-employment closer*	*Re-employment further*
Gender	*0.00036**	*-0.00026*	*-0.00105*	*0.00527****	*0.01038****
	(0.00015)	*(0.00016)*	*(0.00063)*	*(0.00078)*	*(0.00081)*
Age	*-0.00002*	*-0.00003**	*0.00075****	*-0.00063****	*-0.00082****
	(0.00001)	*(0.00001)*	*(0.00006)*	*(0.00007)*	*(0.00007)*
Nationality	*0.00005*	*0.00040**	*0.00384****	*0.00388****	*-0.00051*
	(0.00019)	*(0.00019)*	*(0.00078)*	*(0.001)*	*(0.00106)*
Children	*-0.00065****	*0.00001*	*-0.00498****	*0.0002*	*-0.00043*
	(0.0001)	*(0.0001)*	*(0.00041)*	*(0.00051)*	*(0.00052)*
Log of commuting distance	*-0.00164****	*0.00138****	*0.00591****	*0.01252****	*-0.00096*
	(0.00016)	*(0.00014)*	*(0.0006)*	*(0.00077)*	*(0.0008)*
Correcting term	*0.00181****	*-0.00045**	*-0.00329****	*0.01701****	*-0.01417****
	(0.00025)	*(0.00022)*	*(0.00093)*	*(0.00116)*	*(0.00123)*
Firm rent	*0.00066****	*-0.00005*	*-0.01623****	*-0.01691****	*-0.00384****
	(0.00016)	*(0.00014)*	*(0.00052)*	*(0.00069)*	*(0.00079)*

Table B.2: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Re-employment closer*	*Re-employment further*
Private services	*0.00066****	*0.00002*	*-0.01262****	*0.00336****	*0.00293***
	(0.00018)	*(0.00018)*	*(0.00093)*	*(0.00098)*	*(0.00101)*
Retailing	*0.00063****	*-0.00031**	*-0.00701****	*-0.001*	*0.00385****
	(0.00015)	*(0.00015)*	*(0.0007)*	*(0.00084)*	*(0.00086)*
Construction	*0.00034*	*-0.00087****	*0.00450****	*-0.00958****	*-0.00588****
	(0.00023)	*(0.00024)*	*(0.00086)*	*(0.00119)*	*(0.00121)*
Manufacture	*-0.00024*	*-0.00018*	*-0.00161**	*-0.00924****	*-0.00648****
	(0.00016)	*(0.00015)*	*(0.00067)*	*(0.00085)*	*(0.00088)*
Previous income	*-0.00006*	*-0.00001*	*-0.00504****	*-0.00114***	*0.00063*
	(0.00007)	*(0.00008)*	*(0.00025)*	*(0.00037)*	*(0.00041)*
Unemployment experience	*0.00040****	*0.00002*	*0.00520****	*0.00311****	*0.00480****
	(0.00012)	*(0.00013)*	*(0.00049)*	*(0.00064)*	*(0.00065)*
Single parent	*0.00250****	*0.00681****	*0.00719****	*0.00980****	*0.00763***
	(0.0006)	*(0.00091)*	*(0.00197)*	*(0.00238)*	*(0.0024)*
Living with parents	*0.00607*	*0.0184*	*-0.00513*	*0.02803*	*0.02331*
	(0.00563)	*(0.01175)*	*(0.01363)*	*(0.02012)*	*(0.01965)*
Single	*0.00456****	*0.01213****	*0.01581****	*0.01303****	*0.01380****
	(0.0008)	*(0.00121)*	*(0.0023)*	*(0.00254)*	*(0.00263)*
Log of time	*-0.00036****	*-0.00030****	*-0.02301****	*-0.05368****	*-0.04617****
	(0.00005)	*(0.00006)*	*(0.00027)*	*(0.00037)*	*(0.00036)*
Gymnasium level of education	*0.00025*	*0.00013*	*-0.00416****	*-0.00145*	*0.00116*
	(0.00017)	*(0.00017)*	*(0.00063)*	*(0.00087)*	*(0.00088)*
Post-secondary education <2 years	*0.00081****	*-0.0003*	*-0.00405****	*0.00413***	*0.00646****
	(0.00024)	*(0.00026)*	*(0.00103)*	*(0.00129)*	*(0.00131)*
Post-secondary education >2 years	*0.00069****	*0.00002*	*-0.01061****	*0.00348****	*0.00483****
	(0.0002)	*(0.0002)*	*(0.00078)*	*(0.00101)*	*(0.00104)*
MSc or PhD	*0.00178****	*0.00026*	*-0.00440**	*-0.00043*	*-0.0052*
	(0.0004)	*(0.00045)*	*(0.00203)*	*(0.00253)*	*(0.00269)*
Public services (Spouse)	*-0.00033**	*0.0002*	*0.00102*	*-0.00529****	*-0.00335****
	(0.00017)	*(0.00018)*	*(0.00076)*	*(0.00093)*	*(0.00094)*
Private services (Spouse)	*-0.00032*	*-0.00023*	*-0.00095*	*0.00004*	*0.00012*
	(0.00022)	*(0.00024)*	*(0.00102)*	*(0.00121)*	*(0.00122)*
Retail (Spouse)	*-0.00007*	*-0.00007*	*0.00135*	*-0.00273***	*-0.00235**
	(0.00018)	*(0.0002)*	*(0.00081)*	*(0.001)*	*(0.00102)*
Construction (Spouse)	*-0.00037*	*-0.00019*	*-0.00002*	*-0.00441****	*-0.00372***
	(0.00024)	*(0.00027)*	*(0.00103)*	*(0.00133)*	*(0.00134)*

Table B.2: Continued

Variables	*Type of hazard*				
	Migration further	*Migration closer*	*Separation to non-employment*	*Re-employment closer*	*Re-employment further*
Manufacture (Spouse)	*-0.00040**	*0.00011*	*0.00016*	*-0.00472****	*-0.00696****
	(0.00018)	*(0.0002)*	*(0.00081)*	*(0.00101)*	*(0.00103)*
MSc or PhD (Spouse)	*0.00078*	*0.00013*	*0.00042*	*0.00041*	*0.00225*
	(0.0004)	*(0.00047)*	*(0.00207)*	*(0.00255)*	*(0.00262)*
Post-secondary education >2 years (Spouse)	*0.00030**	*-0.00004*	*-0.00174***	*0.00167**	*0.00155**
	(0.00012)	*(0.00013)*	*(0.00055)*	*(0.00066)*	*(0.00067)*
Post- secondary education <2 years (Spouse)	*-0.0001*	*-0.00001*	*0.00059*	*-0.00075*	*0.00212**
	(0.00019)	*(0.00021)*	*(0.00084)*	*(0.00104)*	*(0.00104)*
Log of earnings (Spouse)	*-0.00007***	*-0.00005*	*-0.00024*	*0.00089****	*0.00013*
	(0.00003)	*(0.00003)*	*(0.00012)*	*(0.00017)*	*(0.00017)*
Age (Spouse)	*-0.00003**	*0.00001*	*-0.00018***	*-0.00025****	*-0.00001*
	(0.00001)	*(0.00001)*	*(0.00006)*	*(0.00007)*	*(0.00007)*
Nationality (Spouse)	*0.00002*	*0.00027*	*0.00374****	*0.00469****	*0.00119*
	(0.00019)	*(0.00019)*	*(0.00079)*	*(0.00101)*	*(0.00105)*
Commuting distance (Spouse)	*0.00011***	*0.00031****	*-0.00019*	*-0.00186****	*0.00223****
	(0.00003)	*(0.00004)*	*(0.00015)*	*(0.00019)*	*(0.00019)*
	Number of observations		*926892*		
	Overall p-value		*0.00000*		
	Pseudo R^2		*0.0773*		

Significance level: ""p<0.05, "**"p<0.01, "***" p<0.001*

Note. The dependent variable is the multinomial variable denoting the competing risks.

The standard errors are bootstrapped