Finnish Economic Papers – Volume 17 – Number 1 – Spring 2004

INTERETHNIC WAGE VARIATION IN THE HELSINKI AREA*

JAN SAARELA

Department of Social Sciences, Åbo Akademi University,
P.O. Box 311, FI-65101 Vasa, Finland

and

FJALAR FINNÄS

Institutet för Finlandssvensk Samhällsforskning, Åbo Akademi University,
P.O. Box 311, FI-65101 Vasa, Finland

This paper compares wage income of Swedish-speaking and Finnish-speaking employees in the Helsinki metropolitan area. Longitudinal data are analysed with random-effects tobit models. We find that Swedish-speaking males on average have 17 per cent higher wages than Finnish-speaking males. Two thirds of this wage gap can be attributed to characteristics differences, particularly education and age. For females the wage difference is very small. The findings echo previous research in the sense that they point out a favourable labour market performance of the Swedish-speaking minority in Finland and that differences between language groups are larger among males than among females. (JEL: J15, J31)

1. Introduction

Recent studies have shown that people of the Swedish-speaking linguistic minority in Finland have lower unemployment and disability retirement propensity than their Finnish-speaking counterparts (Saarela and Finnäs, 2002a, 2002b, 2003a, 2003b). At least in certain respects, the labour market performance of this minority is thus better than that of the majority.

From the international literature, there are empirical evidence suggesting that population groups who experience relatively low unemployment rates also may have relatively high wages when employment is found (Blackaby et al., 1994, 1998, 2002). The purpose of the present paper is consequently to study whether there also is a wage gap between Swedish-speakers and Finnish-speakers in Finland. No such systematic language-group comparisons of wages, or of any other detailed economic indicators, have previously been performed.[1] Offi-

* *We are grateful for advice from two anonymous referees, Stan Panis, Markus Jäntti, and seminar participants at SOFI and EALE 2003.*

[1] *Smedman (1998) and Kyyrä (1999) are two minor studies that compare wages of the two language groups.*

cial statistical series in Finland neither contain income measures classified by native language.

Direct involvement of Sweden in Finland can be traced already from the twelfth century. Also at even earlier dates, however, Finland seems to have had permanent settlement of Swedish-speakers. Until the beginning of the 20th century, Swedish was the official language of government and the domestic language of the upper social orders, but Finnish was never deliberately suppressed. Swedish-speakers were predominant in the tiny upper class, but both language groups consisted predominantly of rural populations of modest social status. Also after Finland's independence from the Russian Empire in 1917, Swedish-speakers in the country benefited from existing financial resources, higher-level education and cultural life, in spite of them having decreased in both absolute and relative numbers due to emigration, low birth rates and intermarriage.[2] Socio-economic disparities between the two language groups are smaller today but do exist (Finnäs, 2003).

Nowadays, Swedish-speakers amount to barely six per cent of the total population (whereas about two per cent has a mother tongue other than Finnish or Swedish). About 95 per cent of them live concentrated at the southern and western coastlines of the country, in municipalities which are classified as either bilingual or monolingual Swedish.

Helsinki with surrounding municipalities constitutes the country centre, composed of professionals and business interests, in which a major part of the Swedish-speakers live, and in which they traditionally have played a major role in terms of professionals and business interests and thus decision-making and economic life (Allardt and Miemois, 1979, 1982). In the Helsinki area, which has witnessed a tremendous in-migration of foremost Finnish-speakers from other parts of the country, Swedish-speakers are in clear minority, in contrast with the situation in most other parts of the country with Swedish-speaking settlement. Swedish-speakers amount to seven per cent of the population in the Helsinki region, whereas about half of the Swedish-speaking population in Finland live in municipalities where they constitute the local majority. There is also substantial variation in income, education, age, industrial structure, population density, etc. within the settlement area of the Swedish-speakers (Allardt and Starck, 1981; McRae, 1997; Finnäs, 2003). Any attempt to analyse a larger geographical area would therefore be confounded by these preconditions. In order to understand the role of native language in terms of a comparison of wage income between the two language groups, a first study ever on the topic therefore obviously needs to be restricted to the Helsinki metropolitan area.

A common presumption is that Swedish-speakers in Finland are over-represented among those well-to-do. This is also supported by population censuses, saying that Swedish-speakers, particularly those in the Helsinki area, are proportionately better represented in certain higher socio-economic categories in the work force. What is not clear from these data, however, is how much of the socio-economic variation is explainable in terms of factors such as level of education and age structure (McRae, 1997).

It would be fairly natural to expect that Swedish-speakers have higher wages than Finnish-speakers, simply because the socio-economic and demographic composition favours them. Swedish-speakers are, for instance, more likely to have university education than Finnish-speakers (Saarela and Finnäs, 2003c). They are also on average older.

Also some other issues may promote the wage of Swedish-speakers. One such aspect is the ability to speak both Swedish and Finnish, which may be needed in administration as well as in many services professions. Unfortunately, there exist no recent extensive register data, which would provide information about the language skills of each population group. In spite

They come to somewhat different conclusions, probably because they seem to differ in terms of how the analysed data are geographically restricted. Saarela (2002) found that social assistance propensity is lower among Swedish-speaking unemployed than among Finnish-speaking unemployed in a local labour market. Since social assistance is a conditional transfer provided to families with low income, this may be regarded as an indication of language-group differences in the economic situation of the unemployed.

[2] *McRae (1997) provides a further historical background.*

of this lack of adequate data, it is evident that Swedish-speakers to a greater extent than Finnish-speakers are bilingual. The 1950 census indicated that bilingualism is much more frequent among the Swedish-speakers. Other features supporting the argument is the fact that the proportion of people who are the offspring of mixed marriages is higher among the Swedish-speakers than among the Finnish-speakers, and that in-migration of Finnish-speakers from the monolingual Finnish-speaking parts of the country has been very high. Being a fairly small minority in relative numbers, Swedish-speakers in the Helsinki area also obviously need Finnish in much of their everyday life.

Another factor of potential importance is culture, which usually includes some notion of shared values, beliefs, expectations, customs, jargon and rituals (Lazear, 1999). Some international studies (Sowell, 1981, 1983; Chiswick, 1983a, 1983b; Simmons, 2003) claim that cultural explanations may lie behind the favourable wage performance of certain minority groups. It is plausible that this is the case also here. O'Leary and Finnäs (2002) argue that intra-group interaction is identifiable along communal lines of much of the social life of Swedish-speakers, including the education system, media and a number of social organisations. On that account, the two groups are somewhat socially separated, which may act to promote linguistic solidarity among the Swedish-speakers. Some scholars also suggest that cultural endowments may explain why morbidity and mortality rates are lower and marital stability higher among Swedish-speakers than among Finnish-speakers (Finnäs, 1997; Hyyppä and Mäki, 2001; Saarela and Finnäs, 2002a). The term culture in this paper refers to such potential differences in lifestyle.

The standard interpretation of unexplained earnings differentials, i.e. differences in wages that remain after having controlled for socio-economic and demographic characteristics, would be that there is discrimination (Blinder, 1973; Oaxaca, 1973). This is not likely to be the case here. Swedish-speakers and Finnish-speakers live intermingled and are guaranteed equal constitutional rights, which among other things imply that there are two parallel education systems up to upper secondary level, with the same curriculum but with instructions in each language. Also in several fields at tertiary level, there are schools providing instruction in Swedish.

Due to non-existing data about language skills and cultural related factors, and thus the impossibility to explicitly disentangle the impact of these issues on the individual wage rate, this study should be seen as a first explorative attempt to compare the wage income of these two population groups. We avoid making direct comparisons with other linguistic or ethnic minorities, because it seems to be difficult finding obvious points of similarity. There is, for instance, no linkage between linguistic and religious dimensions, as is the case in Switzerland, Belgium and Canada.[3]

2. Data

We will use an extract from the longitudinal data file *Työssäkäyntitilasto* (*Employment Statistics*), maintained by Statistics Finland. It consists of annual individual-level information from the years 1987 to 1999, about basic socio-economic and demographic factors, as well as issues related to individuals' labour market situation.

Since a language variable is included in population registers in Finland, official statistics can be used to study language groups. The extract we use has been designed to facilitate comparisons between Swedish-speakers and Finnish-speakers. It is a random sample, consisting of 20 per cent of all Swedish-speakers and 5 per cent of all Finnish-speakers, representing all residents born before 1984, who live in the provinces Uusimaa, Eastern Uusimaa, Varsi-

[3] *O'Leary and Finnäs (2002) propose that the Swedish-speaking minority in the Helsinki area and Protestants in Dublin in the Republic of Ireland are comparable in terms of being ethnic minorities who have been historically dominant, with strong educational and social organisations, and a similar demographic situation. They also suggest that in this context it is not important what people in the minority have in common (language or religion) but rather that they distinguish themselves from the majority, and that they have opportunities for contact with other members of the same group.*

nais-Suomi, Pohjanmaa and the Åland Islands at the end of one or more of the years. The geographical restriction is due to the fact that practically all Swedish-speakers live in this area.

The focus of the present paper is at the Helsinki metropolitan region, which consists of the municipalities Helsinki, Espoo, Vantaa and Kauniainen. The total population in this area is about 900,000.

The variable of central interest is wage. Total annual earnings for each year are used as proxies, since explicit information about wage rates is not available. The number of months a person has been working each year is known. We can therefore restrict the analysis to employees (not self-employed and farmers) who have been working all twelve months of each year. This also reduces sample heterogeneity.

It is not possible to distinguish people in part-time work from those working full time. This should not be a problem, because part-time work is not very common in Finland (Ilmakunnas, 1997). We also exclude people who fall below the "minimum" level for full-time workers, which is about 60,000 FIM in 1999 prices (about 10,000 €), in order to decrease the proportion of potential part-time workers in the data (cf. Kyyrä, 1997).

Statistics Finland have top-coded the earnings data to guarantee anonymity of the persons. This implies that wage rates for sample individuals found in the top decile each year are not known.

We include a variable representing family situation in order to reduce differences in annual earnings caused by variation in working hours (and/or variation in social competence). A variable indicating whether a person was employed for twelve months the previous year is also used, to serve as a crude indicator for labour market experience. Data on earnings for the first observation year, 1987, are therefore not utilised in the analysis.

The data are restricted to people who are 21-59 years old, since participation rates of those in other ages are well below 50 per cent. The total number of observations is 213,177. They represent 33,373 individuals.

The standard practice to obtain real wages would be to adjust with the consumer price index. This would, however, result in different censoring points for each year, and require further adjustments. Therefore, and because the focus is not on individual earnings profiles, we have chosen an alternative method that directly produces adjusted wages with the same censoring point for each year.[4] To achieve this, we for each year compute $Y_t = (c_{1999}/c_t) \cdot y_t$, where c_t is the top decile limit and y_t wage income for year t. As a consequence, we end up with "real wages" Y that are top-coded at 201,000 FIM in 1999 prices.

The original top-coding was made on the basis of all sample individuals, also younger and older people, not full-time workers and those living outside the Helsinki region. Since the overall wage rate of those studied is higher than that of those excluded from analysis, the proportion of right-censored observations is higher than in the raw data. Table 1 provides a description of the real wage distribution by gender and native language. The proportion of people with relatively high wages is substantially larger among Swedish-speaking males than among Finnish-speaking males. For females the difference between language groups is much smaller. The percentage of right-censored observations, i.e. those with a wage of 201,000 FIM or higher, is 53.8 for Swedish-speaking males, 39.2 for Finnish-speaking males, 16.0 for Swedish-speaking females, and 14.1 for Finnish-speaking females.

The statistical estimations, which will be discussed in detail in the following section of the paper, seek to explain the above depicted language-group difference in wages, i.e. if it differs between socio-economic, demographic and labour market related characteristics, and how much can be attributed to these characteristics. Table 2, which provides a description of the distribution of explanatory variables to be used by gender and native language, indicates that there are obvious reasons for doing such an analysis. Swedish-speakers are on average older, higher educated, educated in other fields, work in different industries, have higher levels of employment experience, and live in different

[4] *The results would remain practically the same no matter which adjustment method is used.*

Table 1. Some descriptive statistics of the wage distribution by gender and native language.

		% with wage over (in 1,000 FIM)						
		100	125	150	175	200	Median	n obs.
MALES	Finnish	95.0	84.7	67.9	51.9	39.2	179	79,030
	Swedish	95.6	87.4	75.0	63.3	53.8	201	24,094
FEMALES	Finnish	88.0	61.7	37.4	22.7	14.1	136	86,371
	Swedish	88.4	65.9	42.4	25.4	16.0	142	23,682

The percentage of observations with wages over 200,000 FIM corresponds with the percentage of right-censored observations.

Table 2. Variable distributions by gender and native language (%).

	MALES		FEMALES	
	Finn.	Swed.	Finn.	Swed.
Age, 21–24 years	5.2	3.4	4.9	3.6
25–29	14.4	10.9	12.2	10.7
30–34	17.1	13.7	14.1	11.4
35–39	15.7	13.6	14.9	11.9
40–44	15.5	16.3	16.5	15.3
45–49	14.3	17.0	16.3	17.9
50–54	11.3	15.2	13.1	17.5
55–59	6.5	9.9	7.8	11.7
Education level, Basic	24.4	22.4	27.5	27.2
Lower vocational	36.4	25.2	33.2	26.8
Upper vocational	13.1	12.9	20.0	18.4
Undergraduate	7.8	14.7	6.7	13.7
Graduate or postgraduate	18.3	24.8	12.6	13.9
Education field, Teacher, human., aesth.	2.8	3.1	7.7	7.8
Commercial, social science	15.7	27.4	25.8	32.2
Natural science	9.5	7.8	3.2	2.1
Technology	30.8	22.1	5.1	2.5
Medical and health care	1.9	2.4	11.5	12.1
Services	4.8	3.4	7.4	4.4
General or other	34.5	33.9	39.3	38.9
Industry, Manufacturing	19.5	15.9	9.9	8.6
Construction	6.6	2.7	1.2	0.5
Trade	17.4	26.2	14.9	18.7
Hotels and restaurants	1.9	1.6	3.6	1.9
Transport and communications	11.9	11.2	6.5	9.2
Financial intermediation	3.9	6.2	8.7	10.0
Real estate and business services	15.6	14.7	11.2	10.5
Public administration	7.6	5.0	9.2	6.4
Education	4.8	5.6	7.7	7.7
Human health and social services	2.5	2.7	19.6	17.4
Other services	5.5	6.3	6.2	7.7
Other industry	2.7	1.9	1.4	1.5
Not employed for 12 months last year	10.1	8.3	10.7	9.8
Family type, Married, no children	14.3	16.1	14.8	17.0
Married, 1 child	14.7	16.7	13.3	13.2
Married, 2 children	18.9	22.1	15.0	14.7
Married, 2+ children	7.7	9.9	4.7	5.5
Consensual union, no children	9.5	6.9	8.7	6.7
Consensual union, children	5.0	3.6	3.8	3.8
Sole supporter	1.5	1.4	10.2	10.7
Single	23.5	18.5	27.4	25.4
Other	4.9	4.7	2.0	3.0
n observations	79,030	24,094	86,371	23,682
n individuals	12,535	3,669	13,501	3,668

family compositions than Finnish-speaking males. The differences among females are somewhat smaller than among males.

3. Econometric methodology

Since the observed dependent variable wage, denoted by y, is right-censored, the tobit model (Tobin, 1958) that is used is specified as

$$(1) \qquad y_{it} = \begin{cases} y_{it}^* = x_{it}\beta + \varepsilon_{it} & \text{if } y_{it}^* < \tau \\ \tau_y & \text{if } y_{it}^* \geq \tau \end{cases}$$

where y^* is the latent (or index) variable, and τ is the threshold of censoring. Each individual is denoted by i, and each observation by t. A vector of explanatory variables is referred to as x, whereas β is its associated vector of coefficients. The error is denoted by ε_{it}. Since we have longitudinal data, the model used is of the random-effects type.[5] We specify

$$(2) \qquad \varepsilon_{it} = v_{it} + u_i .$$

Unmeasured characteristics are thus in part specific to each observation (v_{it}), and in part individual-specific and constant across time (u_i). Both these components are assumed normally distributed with zero means and independent of one another, so that

$$(3) \qquad Var[\varepsilon_{it}] = \sigma_v^2 + \sigma_u^2 .$$

The parameter σ_u is the standard deviation for the error part related to unobserved individual heterogeneity. The standard deviation of v_{it} is also estimable, as is the case in all tobit models.

If y^* can be assumed normally distributed, the tobit model will provide consistent and efficient estimates of parameters. Maximum likelihood estimation for the model involves dividing the observations into two sets. The first set contains uncensored observations, which maximum likelihood treats in the same way as the linear regression model. The second set contains censored observations, for which we do not know the specific value of y^*. One therefore proceeds by computing the probability of being censored

$$(4) \qquad \Pr(y^* \geq \tau | x) = \Phi\left(\frac{x\beta - \tau}{\sigma}\right),$$

and using this quantity in the likelihood equation. The likelihood function for both sets of observations is then

$$(5) \qquad \ln L = \sum_{Uncens.} \ln \frac{1}{\sigma}\phi\left(\frac{y - x\beta}{\sigma}\right) + \sum_{Cens.} \ln\Phi\left(\frac{x\beta - \tau}{\sigma}\right),$$

where ϕ and Φ are the probability density function and the cumulative density function respectively, for the standard normal distribution, and σ is the standard deviation of ε.

Expected values for the latent outcome, $E(y^*|x) = x\beta$, are our primary focus.

If the wage structure is influenced by factors affecting whether individuals are working all twelve months during the year, a correction for selectivity bias may improve the fit of the model. We have attempted to correct for self-selection in terms of "Heckman's lambda" (Heckman, 1979). The estimate for this parameter turned out to be positive and significant, but did not change the results with respect to language-group differences in wages.[6] Since these types of estimates are increasingly being questioned (Nawata, 1993; Manski, 1995), we have chosen to exclude this specification from the final models, whose results are reported here.

The results are not sensitive for clustering within individuals. There were some indications of heteroskedasticity, which means that the variance of the residual is not constant across language groups. Some tests we performed, however, indicated that this departure from homo-

[5] *The random-effects specification rests on the assumption that the distribution function of errors is independent of explanatory variables, i.e. that unobservable factors are not correlated with explanatory variables (Arellano and Honoré, 1999). Since language group remains constant over time, it is not possible to use fixed effects.*

[6] *For Swedish-speaking males, there was not even an improvement of the log likelihood when Heckman's lambda was included.*

skedasticity was so small in magnitude that the results to be reported here are not severely affected.

As is standard in the empirical literature on between-group comparisons of earnings, we will determine how much of any wage differential is due to characteristics differences and how much is due to coefficient differences (i.e., due to differences in the returns to the characteristics). The latter part is often attributed to discrimination, but it can in principle be related to anything that is not associated with observable characteristics, such as for instance culture.

The approach adopted here is the one proposed by Neumark (1988) and Oaxaca and Ransom (1994). It attempts to estimate the competitive wage structure that would exist in the absence of "discrimination" and use these estimates as weights in the decomposition of the wage gap. An obvious advantage of this approach is that it leads to a unique solution and so avoids the index number problem associated with the initial methodology of Blinder (1973) and Oaxaca (1973).[7]

Formally, the wage gap can be decomposed as

$$(6)\quad \overline{\ln Y^s} - \overline{\ln Y^f} = \left[\hat{\beta}^*\left(\bar{x}^s - \bar{x}^f\right)\right] + \\ + \left[\bar{x}^s\left(\hat{\beta}^s - \hat{\beta}^*\right) - \bar{x}^f\left(\hat{\beta}^f - \hat{\beta}^*\right)\right]$$

where Y is real wage, s and f refer to Swedish-speakers and Finnish-speakers respectively, x is a row vector of characteristics, and $\hat{\beta}$ is a vector of estimated coefficients. $\hat{\beta}^*$ is an estimate of the "non-discriminatory" wage structure and is derived by using the cross product matrices as weights from the wage equation such that

$$(7)\quad \hat{\beta}^* = \Omega\hat{\beta}^s + (1-\Omega)\hat{\beta}^f,$$

where $\Omega = (x^{s'}x^s + x^{f'}x^f)^{-1}x^{s'}x^s$ is the Oaxaca-Ransom weighting matrix. The wage structure given by (7) is equivalent to running a regression on the pooled data.

The first term on the right-hand side of (6) represents the difference in wage that is attributed to wage-related characteristics, which proxy productivity. The second term represents the wage differential that is due to differences in returns to these characteristics.

Since the model is linear and the regression line goes through the sample means of the data, the characteristic component can be further decomposed into its individual components, for instance the proportion of the wage gap explained by differences in education level. The coefficient component cannot be further divided in this way, since individual decompositions are arbitrary, being influenced by transformations of the data and the choice of omitted variable categories (Jones, 1983).

Estimations have been performed with *aML* version 1.04.

4. Results

It is essential to allow the intercept as well as the returns to each of the explanatory variables to vary between language groups within genders. We have therefore estimated separate tobit models for Finnish-speaking and Swedish-speaking males and females, respectively (the results can be seen in Appendix 1). In order to get an overview of how each model reflects the latent wage rate, the distribution of the predicted wage for the sample individuals is provided in Figure 1 for males and in Figure 2 for females. The wage distribution of Swedish-speaking males is more skewed to the right than that of Finnish-speaking males, as already suggested by Table 1. The spread is also higher for Swedish-speakers. For females, the wage distribution of each language group is close to identical, with a very small advantage for the Swedish-speakers.

The estimation results are summarised in Table 3, which reflects the percentage difference between Swedish-speakers and Finnish-speakers in the predicted wage for each characteris-

[7] *Blackaby et al. (1998, 2002) and Simmons (2003) use the Neumark-Oaxaca-Ransom approach, whereas Drinkwater and O'Leary (1997), Trejo (1997) and Saarela and Finnäs (2002a) perform decompositions according to the Blinder-Oaxaca setting. Alternative approaches to the index number problem have been proposed by Reimers (1983) and Cotton (1988).*

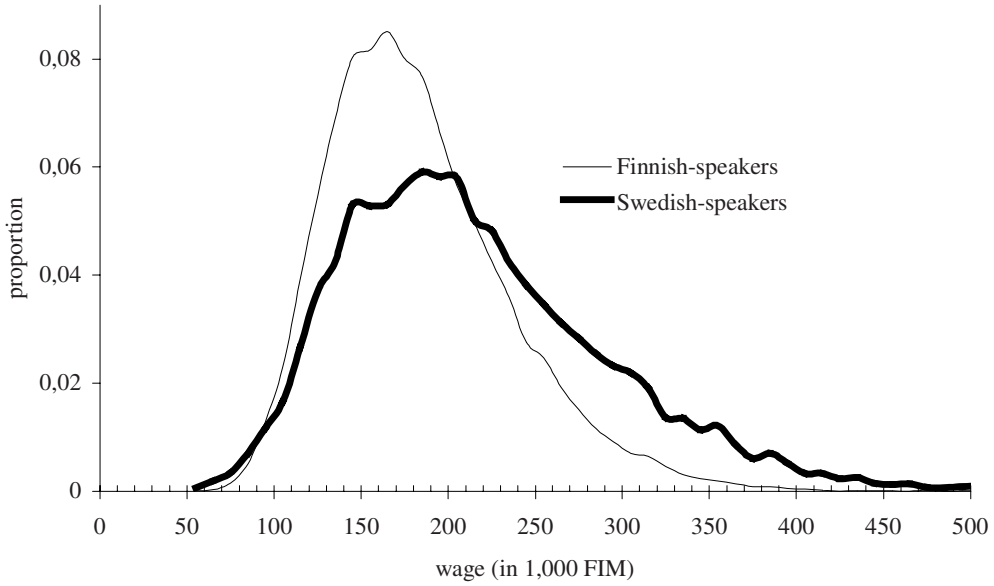

Figure 1. Distribution of predicted wage for males.

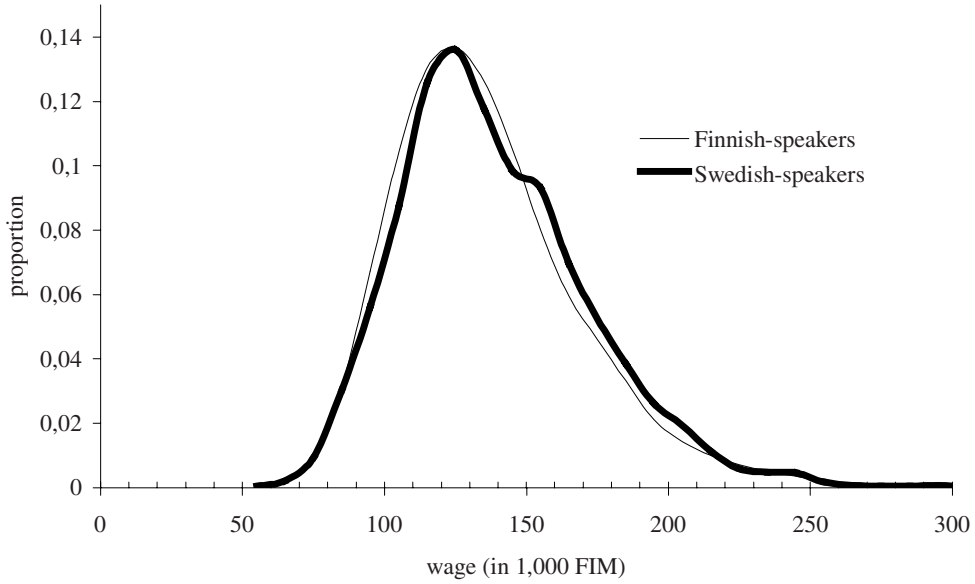

Figure 2. Distribution of predicted wage for females.

tic of the explanatory variables. Each of the other variables have been set to its reference value (i.e., each estimate is simply added to the constant).

For males, there is consistently a wage advantage for Swedish-speakers. There are also some indications that this wage gap is larger in higher ages than in lower ages. In order to see if this can be attributed to a cohort effect, and not due to a pure age effect, we have tried to

Table 3. Language-group difference in predicted wage.

	Males	Females
Age, 21–24 years	+	0
25–29	+	0
30–34	++	0
35–39	+++	0
40–44	+++	0
45–49	+++	0
50–54	+++	0
55–59	+++	0
Education level, Basic	+	0
Lower vocational	+++	0
Upper vocational	++	0
Undergraduate	+++	0
Graduate or postgraduate	+++	0
Education field, Teacher, hum., aest.	+	0
Commercial, social science	+++	0
Natural science	+++	+
Technology	++	0
Medical and health care	+++	–
Services	++	+
General or other	+++	0
Industry, Manufacturing	++	0
Construction	+++	–
Trade	+++	0
Hotels and restaurants	+	0
Transport and communications	+++	0
Financial intermediation	+++	0
Real estate and business services	++	0
Public administration	+	0
Education	+	0
Human health and social services	+	0
Other services	+++	0
Other industry	++	0
Employed 12 months last year, Yes	+++	0
No	+++	0
Family type, Married, no children	+++	0
Married, 1 child	+++	0
Married, 2 children	+++	0
Married, 2+ children	+++	0
Consensual union, no children	++	0
Consensual union, children	+++	0
Sole supporter	+++	0
Single	++	0
Other	++	0

Percentage difference (Swedish-speakers versus Finnish-speakers) in predicted wage for each characteristic, with each other variable set to its reference value: –10 to –5 (–); –5 to +5 (0); +5 to +10 (+); +10 to +15 (++); +15 to +20 (+++).

include also a variable representing the decade in which the person was born. These results indicated that a substantial part of the above depicted language-group difference may be due to differences across birth cohorts. However, since the observation period is too short to facilitate any reliable conclusions on this point, we have chosen not to outline the results of these estimations here.

Table 4. Decomposition of the language-group wage difference, Swedish-speakers versus Finnish-speakers.

	Males	Females
Mean wage differential, FIM	31,900	3,300
…as approximate geometric mean	0.174	0.024
Difference due to coefficients (%)	35.4	112.7
Difference due to characteristics (%)	64.6	–12.7
Components of characteristics effect (%)		
Age	37.9	26.3
Education level	81.7	83.6
Education field	4.1	22.6
Industry	7.4	3.5
Employed 12 months last year	0.9	0.3
Family type	5.9	5.4
Unobserved individual heterogeneity	–37.9	–41.7

It may also be noted that the language-group wage gap is smallest among people with basic education only. This could be the case if the returns to underlying characteristics, such as language skills, is not equally rewarding for people with low education as it is for people with higher levels of education. For the other characteristics there is no evident pattern, which would have suggested that factors such as bilingualism are particularly important in certain socio-demographic groups.

For females, differences between language groups are very small. In some of the very few cases there is any difference, they even favour the Finnish-speakers.

Results of the decomposition are summarised in Table 4. It shows that Swedish-speaking males on average have 17 per cent higher wages than Finnish-speaking males. Two thirds of this wage gap can be attributed to characteristics differences, whereas one third should be explained by different returns to the characteristics. Decomposing the characteristics effect tells us that most of this is due to differences in education level and age. Unobserved individual heterogeneity contributes negatively to the entity, saying that adding random effects changes the estimated coefficients so that the predicted earnings gap is higher. The fit of each model im-

proves substantially when the component for unobserved individual heterogeneity is included (see Appendix 2).

From Table 4 we can also see that there is a Swedish-speaking advantage in terms of a higher average wage also for females, but that it is only two per cent. In contrast to what is generally assumed, Swedish-speaking females are worse equipped than Finnish-speaking females, but the return to their characteristics is better. This result resembles that of Saarela and Finnäs (2002a). They compare the two language groups with regard to disability retirement and find that also in terms of avoiding disability retirement, socio-economic and demographic characteristics are less favourable for Swedish-speaking females than for Finnish-speaking females. The overall advantage of Swedish-speaking females can thus be attributed totally to differences in the returns to these characteristics.

Education level and age are, as for males, important components of the characteristics effect, as well as the field of education. Adding random effects has the similar impact as for males.

5. *Discussion and conclusions*

In this paper we find that Swedish-speaking males in the Helsinki metropolitan area on average have 17 per cent higher wage income than their Finnish-speaking counterparts. Two thirds of this wage gap can be explained by characteristics differences, foremost a more favourable education and age distribution for the Swedish-speakers. For females the difference between language groups is very small.

The reasons to why the wage gap varies between genders may be several. One could be that there is a difference in language skills between Finnish-speaking females and Finnish-speaking males that drives the results, because it is reasonable to assume that practically every Swedish-speaker in the Helsinki region is bilingual. There are no indications from the data, however, that the language-group wage gap is highest in jobs where language skills should be highly ranked, such as among people working at hotels and restaurants. Another possibility is that social contacts promoting wages differ between genders and that they are specifically prominent among Swedish-speaking males.

The wage gap among males may potentially also be due to cultural differences. Swedish-speakers in the Helsinki area have been remarkably successful in maintaining their relatively high socio-economic position. It has therefore been proposed that their behaviour acts to fix social distances between the two groups, and to maintain social prerogatives and support the existing social structure (O'Leary and Finnäs, 2002; Saarela and Finnäs, 2003c). In some sense this argument is supported by present data, which show that the language-group wage gap is smallest among those with basic education only. It should also be borne in mind that most of the Swedish-speakers constitute the native population in this area, which may help to maintain social status over generations. On the other hand, our data indicate that part of the wage gap could be due to a cohort effect, which implies that such behaviour may have become less important for younger generations.

It cannot be ruled out, however, that if there is an association between social status and social integration, it enhances wage discrimination against the Finnish-speakers. If Swedish-speakers to a higher extent than Finnish-speakers work in high status occupations, which we cannot fully observe from the data, it may promote linguistic solidarity and wage setting (cf. Hechter, 1978). The wage gap would then be caused by intra-group interaction in the Swedish-speakers' social and cultural life.

Our findings clearly indicate that there is much left to explore about these issues. One future approach could be to study the potential wage gap in regions with different language structure, considering that previous research suggests that the widest language-group difference in social structure is in the Helsinki region (Tunkelo, 1933; McRae, 1997; Allardt and Starck, 1981; Finnäs, 2003). Another possibility could be to link intergenerational information to the present data, in an attempt to increase our understanding of the role of social and cultural heritage, or to construct and develop new data including questions about language proficiency and lifestyles.

References

Allardt, E., and K.J. Miemois (1979). Roots Both in the Centre and the Periphery: The Swedish-speaking Population in Finland. Research Reports No. 24. Research Group for Comparative Sociology, Helsinki University.

– **(1982).** "A Minority in Both Centre and Periphery: An Account of the Swedish-speaking Finns." *European Journal of Political Research* 10, 265–292.

Allardt, E., and C. Starck (1981). *Språkgränser och samhällsstruktur: Finlandssvenskarna i ett jämförande perspektiv*. Stockholm: Almqvist and Wiksell.

Arellano, M., and B. Honoré (1999). "Panel Data Models: Some Recent Developments." In *Handbook of Econometrics*, Vol. 5, 3229–3296. Eds. J.J. Heckman and E. Leamer. Amsterdam: North-Holland.

Blackaby, D.H., K. Clark, D.G. Leslie, and P.D. Murphy (1994). "Black-white Male Earnings and Employment Prospects in the 1970s and 1980s: Evidence for Britain." *Economics Letters* 46, 273–279.

Blackaby, D.H., D.G. Leslie, P.D. Murphy, and N.C. O'Leary (1998). "The Ethnic Wage Gap and Employment Differentials in the 1990s: Evidence for Britain." *Economics Letters* 58, 97–103.

– **(2002).** "White/Ethnic Minority Earnings and Employment Differentials in Britain: Evidence from the LFS." *Oxford Economic Papers* 54, 270–297.

Blinder, A.S. (1973). "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8, 436–455.

Chiswick, B.R. (1983a). "An Analysis of the Earnings and Employment of Asian-American Men." *Journal of Labor Economics* 1, 197–214.

– **(1983b).** "The Earnings and Human Capital of American Jews." *Journal of Human Resources* 18, 313–336.

Cotton, J. (1988). "On the Decomposition of Wage Differentials." *Review of Economics and Statistics* 70, 236–243.

Drinkwater, S.J., and N.C. O'Leary (1997). "Unemployment in Wales: Does Language Matter?" *Regional Studies* 31, 581–591.

Finnäs, F. (1997). "Social Integration, Heterogeneity, and Divorce: The Case of the Swedish-speaking Population in Finland." *Acta Sociologica* 40, 263–277.

– **(2003).** "The Swedish-speaking Population on the Finnish Labour Market." *Yearbook of Population Research in Finland* 39, 91–101.

Hechter, M. (1978). "Group Formation and the Cultural Division of Labour." *American Journal of Sociology* 84, 293–318.

Heckman, J.J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47, 153–162.

Hyyppä, M.T., and J. Mäki (2001). "Individual-level Relationships between Social Capital and Self-rated Health in a Bilingual Community." *Preventive Medicine* 32, 148–155.

Ilmakunnas, S. (1997). Female Labour Supply and Work Incentives. Doctoral Dissertation. Studies No. 68. Labour Institute for Economic Research, Helsinki.

Jones, F.L. (1983). "On Decomposing the Wage Gap: A Critical Comment on Blinder's Method." *Journal of Human Resources* 18, 126–130.

Kyyrä, T. (1997). Työllistyneiden alkupalkkojen määräytyminen. VATT-Research Reports No. 39. Government Institute for Economic Research, Helsinki.

– **(1999).** Post-unemployment Wages and Economic Incentives to Exit from Unemployment. VATT-Research Reports No. 56. Government Institute for Economic Research, Helsinki.

Lazear, E.P. (1999). "Culture and Language." *Journal of Political Economy* 107, S95–S126.

Manski, C. (1995). *Identification Problems in the Social Sciences*. Cambridge: Harvard University Press.

McRae, K.D. (1997). *Conflict and Compromise in Multilingual Societies: Finland*. Waterloo, Ontario: Wilfred Laurier University Press.

Nawata, K. (1993). "A Note on the Estimation of Models with Sample-selection Biases." *Economics Letters* 42, 15–24.

Neumark, D. (1988). "Employers' Discriminatory Behaviour and the Estimation of Wage Discrimination." *Journal of Human Resources* 23, 279–295.

Oaxaca, R. (1973). "Male-female Wage Differentials in Urban Labor Markets." *International Economic Review* 14, 693–709.

Oaxaca, R., and M.R. Ransom (1994). "On Discrimination and the Decomposition of Wage Differentials." *Journal of Econometrics* 61, 5–21.

O'Leary, R., and F. Finnäs (2002). "Education, Social Integration and Minority-Majority Group Intermarriage." *Sociology* 36, 235–254.

Reimers, C.W. (1983). "Labour Market Discrimination against Hispanic and Black Men." *Review of Economics and Statistics* 65, 570–579.

Saarela, J. (2002). "Språkgruppsskillnader i utkomststödstagande." *Ekonomiska Samfundets Tidskrift* 55, 89–97.

Saarela, J., and F. Finnäs (2002a). "Language-group Differences in Very Early Retirement in Finland." *Demographic Research* 7, 49–66.

– **(2002b).** "Ethnicity and Unemployment in Finland." *Ethnic Studies Review* 25, 26–37.

– **(2003a).** "Unemployment and Native Language: The Finnish Case." *Journal of Socio-Economics* 32, 59–80.

– **(2003b).** "Can the Low Unemployment Rate of Swedish-speakers in Finland be Attributed to Language-group and Industrial Structure?" In *Essays on Labour Market Outcomes in the Bilingual Area of Finland*. J. Saarela. Doctoral Dissertation. Department of Economics and Statistics, Åbo Akademi University.

– **(2003c).** "Social Background and Education of Swedish-speakers and Finnish-speakers in Finland." *European Journal of Education* 38, 445–456.

Simmons, W.O. (2002). "The Black Earnings Gap: Discrimination or Culture." *Journal of Socio-Economics* 31, 647–655.

Smedman, M. (1998). Inkomstens bestämningsfaktorer – Teoretisk bakgrund samt en empirisk undersökning av tjänstemäns inkomster på en tvåspråkig arbetsmarknad. Master's Thesis. Department of Social Sciences, Åbo Akademi University.

Sowell, T. (1981). *Markets and Minorities*. Oxford: Basil Blackwell.

– **(1983).** *The Economics and Politics of Race: An International Perspective*. New York: William Morrow.

Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26, 24–36.

Trejo, S.J. (1997). "Why Do Mexican Americans Earn Low Wages?" *Journal of Political Economy* 105, 1235–1268.

Tunkelo, A. (1933). "Yrke och språk." *Statistiska Översikter* 10, 33–45.

Appendix 1: Estimation results of random-effects tobit models

Table A1. Results of tobit models for ln(wage/1000), males.

	Finnish-speakers		Swedish-speakers	
Age, 21–24 years	–0.2637	(0.0032)	–0.3531	(0.0081)
25–29	–0.1519	(0.0024)	–0.2306	(0.0060)
30–34	–0.0794	(0.0021)	–0.1065	(0.0050)
35–39	–0.0311	(0.0022)	–0.0419	(0.0055)
40–44	–	–	–	–
45–49	–0.0005	(0.0025)	0.0138	(0.0056)
50–54	–0.0158	(0.0026)	–0.0017	(0.0054)
55–59	–0.0404	(0.0035)	–0.0102	(0.0071)
Education level, Basic	–0.0264	(0.0026)	–0.1108	(0.0056)
Lower vocational	–	–	–	–
Upper vocational	0.0737	(0.0024)	0.0434	(0.0060)
Undergraduate	0.2139	(0.0034)	0.2188	(0.0070)
Graduate or postgraduate	0.3338	(0.0031)	0.3423	(0.0067)
Education field, Teacher, hum., aest.	–0.0599	(0.0054)	–0.1658	(0.0105)
Commercial, social science	–	–	–	–
Natural science	0.0665	(0.0033)	0.0712	(0.0077)
Technology	–0.0059	(0.0027)	–0.0316	(0.0058)
Medical and health care	0.0245	(0.0062)	0.0338	(0.0153)
Services	–0.0053	(0.0041)	–0.0610	(0.0103)
General or other	–0.0404	(0.0033)	–0.0512	(0.0071)
Industry, Manufacturing	0.0223	(0.0021)	–0.0089	(0.0048)
Construction	0.0164	(0.0030)	0.0227	(0.0106)
Trade	–	–	–	–
Hotels and restaurants	–0.0560	(0.0052)	–0.1582	(0.0104)
Transport and communications	–0.0204	(0.0023)	–0.0145	(0.0052)
Financial intermediation	0.0824	(0.0041)	0.1112	(0.0082)
Real estate and business services	–0.0236	(0.0022)	–0.0604	(0.0051)
Public administration	–0.0372	(0.0030)	–0.1224	(0.0076)
Education	–0.0768	(0.0035)	–0.1504	(0.0074)
Human health and social services	–0.1034	(0.0047)	–0.1669	(0.0110)
Other services	–0.0317	(0.0031)	–0.0431	(0.0065)
Other industry	–0.0255	(0.0044)	–0.0863	(0.0102)
Employed 12 months last year, Yes	–	–	–	–
No	–0.1063	(0.0021)	–0.1222	(0.0051)
Family type, Married, no children	–0.0328	(0.0024)	–0.0468	(0.0054)
Married, 1 child	–0.0232	(0.0024)	–0.0234	(0.0051)
Married, 2 children	–	–	–	–
Married, 2+ children	–0.0011	(0.0029)	0.0029	(0.0066)
Consensual union, no children	–0.0746	(0.0026)	–0.1164	(0.0063)
Consensual union, children	–0.0675	(0.0031)	–0.0786	(0.0080)
Sole supporter	–0.0574	(0.0059)	–0.0609	(0.0127)
Other	–0.1188	(0.0032)	–0.1486	(0.0067)
σ_u	0.2087	(0.0007)	0.2508	(0.0017)
σ_v	0.1719	(0.0003)	0.1954	(0.0006)
Constant	5.2221	(0.0036)	5.3755	(0.0080)
Log likelihood	–5,758.2975		–3,929.8789	
n observations	79,030		24,094	
n individuals	12,535		3,669	

Standard errors are in parentheses.

Table A2. Results of tobit models for ln(wage/1000), females.

	Finnish-speakers		Swedish-speakers	
Age, 21–24 years	–0.2641	(0.0032)	–0.2917	(0.0069)
25–29	–0.1812	(0.0021)	–0.1799	(0.0045)
30–34	–0.1070	(0.0018)	–0.1228	(0.0035)
35–39	–0.0434	(0.0018)	–0.0490	(0.0035)
40–44	–	–	–	–
45–49	–0.0012	(0.0019)	0.0069	(0.0035)
50–54	–0.0282	(0.0020)	–0.0146	(0.0036)
55–59	–0.0694	(0.0027)	–0.0455	(0.0048)
Education level, Basic	–0.0548	(0.0022)	–0.0330	(0.0041)
Lower vocational	–	–	–	–
Upper vocational	0.0616	(0.0018)	0.0464	(0.0037)
Undergraduate	0.1831	(0.0030)	0.1813	(0.0048)
Graduate or postgraduate	0.3300	(0.0027)	0.3051	(0.0053)
Education field, Teacher, hum., aest.	–0.0658	(0.0030)	–0.0549	(0.0052)
Commercial, social science	–	–	–	–
Natural science	0.0123	(0.0036)	0.0765	(0.0107)
Technology	–0.0145	(0.0031)	0.0141	(0.0083)
Medical and health care	0.0277	(0.0025)	–0.0225	(0.0047)
Services	–0.0559	(0.0029)	–0.0112	(0.0069)
General or other	–0.0117	(0.0024)	–0.0320	(0.0046)
Industry, Manufacturing	0.0491	(0.0023)	0.0410	(0.0048)
Construction	0.0074	(0.0047)	–0.0569	(0.0121)
Trade	–	–	–	–
Hotels and restaurants	–0.0145	(0.0037)	–0.0571	(0.0099)
Transport and communications	0.0402	(0.0028)	0.0183	(0.0043)
Financial intermediation	0.0806	(0.0025)	0.0679	(0.0043)
Real estate and business services	0.0201	(0.0021)	0.0324	(0.0040)
Public administration	–0.0081	(0.0024)	0.0132	(0.0053)
Education	–0.0124	(0.0027)	–0.0464	(0.0048)
Human health and social services	–0.0331	(0.0022)	–0.0123	(0.0042)
Other services	–0.0055	(0.0026)	–0.0247	(0.0046)
Other industry	–0.0063	(0.0042)	–0.0175	(0.0091)
Employed 12 months last year, Yes	–	–	–	–
No	–0.0586	(0.0019)	–0.0708	(0.0037)
Family type, Married, no children	0.0623	(0.0020)	0.0672	(0.0041)
Married, 1 child	0.0054	(0.0019)	0.0172	(0.0039)
Married, 2 children	–	–	–	–
Married, 2+ children	–0.0415	(0.0028)	–0.0601	(0.0046)
Consensual union, no children	0.0654	(0.0024)	0.0824	(0.0051)
Consensual union, children	–0.0310	(0.0029)	–0.0350	(0.0058)
Sole supporter	0.0361	(0.0022)	0.0393	(0.0043)
Single	0.0779	(0.0018)	0.0823	(0.0037)
Other	0.0289	(0.0039)	0.0264	(0.0066)
σ_u	0.1777	(0.0006)	0.1916	(0.0010)
σ_v	0.1633	(0.0002)	0.1696	(0.0004)
Constant	4.9091	(0.0027)	4.9102	(0.0053)
Log likelihood	9,292.5848		1,275.0031	
n observations	86,371		23,682	
n individuals	13,501		3,668	

Standard errors are in parentheses.

Appendix 2: Unobserved individual heterogeneity and log likelihood improvement

The fit of each of the four models improves substantially when the component for unobserved individual heterogeneity is included. This is illustrated in Figure A1, where we have plotted the value of the log likelihood for different values of σ_u (this is an exercise facilitated by the grid search option available in *aML* version 2.00 beta).

For Finnish-speaking females and Swedish-speaking females, the final log likelihood is even positive. This is a result of the small range in which the dependent variable varies. The standard errors, σ_u and σ_v, are small, which implies that the probability density is very large, and the log likelihood thus positive. To put it formally, consider just uncensored (continuous) outcomes. Their likelihood is the probability density,

$$\text{(A1)} \quad \phi(y) = \frac{1}{\sqrt{2\pi s^2}} \exp\left[-0.5 \left(\frac{y}{s} \right)^2 \right],$$

where the mean is assumed to be zero and the standard deviation to be *s*. The log likelihood is then

$$\text{(A2)} \quad \ln[\phi(y)] = -0.5 \left[\ln(2\pi s^2) + \left(\frac{y}{s} \right)^2 \right].$$

If *s* is sufficiently small, this quantity is positive.

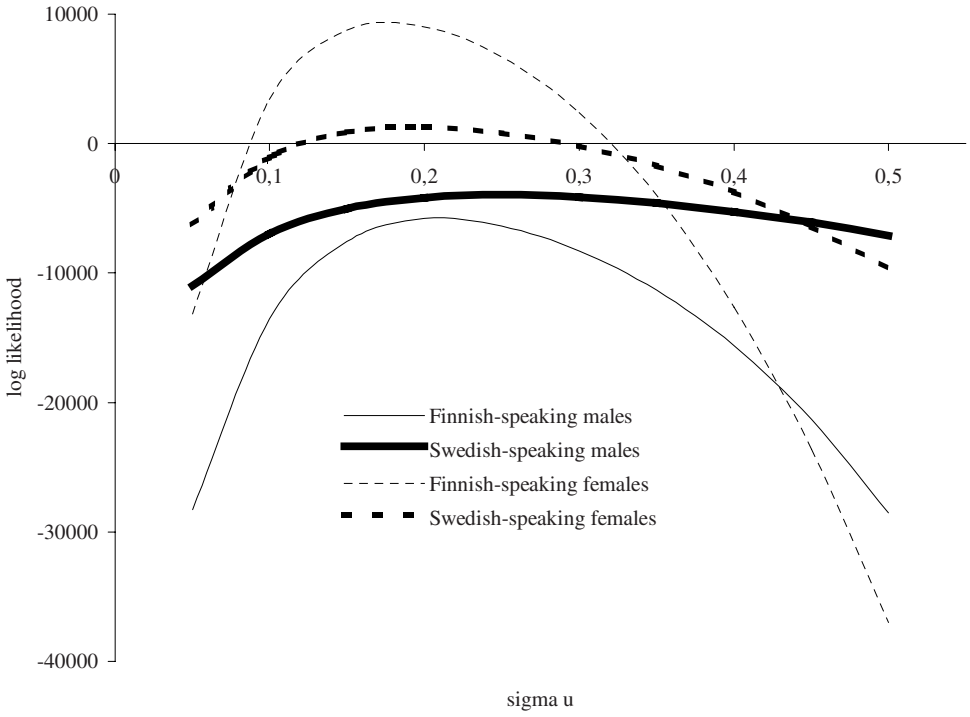

Figure A1. Log likelihood plotted against unobserved individual heterogeneity.