Social Networks and the Economic Performance of Minorities*

Ott Toomet[†] Meredith Rolfe[‡] Marco van der Leij [§]

May 7, 2009

Abstract

This paper analyses the relationship between unexplained racial/ethnic unemployment and wage differentials and the segregation of social networks, as measured by inbreeding homophily. Our analysis is based on both U.S. and Estonian surveys, supplemented with Estonian telephone communication data. In the case of Estonia we consider the regional variation in economic performance of the Russian minority, and in the U.S. case we consider the regional variation in black-white differentials. Our analysis finds a strong relationship between the size of the differential and network segregation: regions with more segregated social networks exhibit larger unexplained wage and unemployment differential.

JEL codes: J71, J31

Keywords: social networks, wage differential, segregation, race, minorities

1 Introduction

On average, members of ethnic or racial minorities often earn less than those from the majority group. This minority wage gap refers most notably to black and white males in the United States, but similar wage gaps also characterise a large number of other groups, including whites and Hispanics in the US (Altonji and Blank, 1999), Blacks and Pakistanis in UK (Blackaby, Leslie, Murphy, and O'Leary, 2005), Russians and Estonians in Estonia (Leping and Toomet, 2008), Serbians and Albanians in Kosovo (Bhumaik, Gang, and Yun, 2006), and Turks and Bulgarians in Bulgaria (Giddings, 2002). Current treatments of the systematic difference in minority-majority wages focus on the personal characteristics of minority employees (i.e., education, job-related training, etc, see Altonji and

*Corresponding Author: Ott Toomet. E-mail: otoomet@ut.ee. This paper benefitted from comments by Sebastian Buhai, Tomasz Mickiewicz, Betsy Sinclair, Tom Snijders, and audiences at Aarhus Business School, Utrecht University, Tartu University and a Workshop on Networks in Oxford. We are grateful to Indrek Seppo for help with the initial processing of ELFS data, and to Jane Roberts and the Roper Center for assistance with the Social Capital Benchmark Data. Marco van der Leij would like to thank the Spanish Ministry of Science and Innovation for financial support (Project SEJ2007-62656).

†Department of Economics. Tartu University.

‡Saïd Business School, University of Oxford.

§Departamento de Fundamentos del Análisis Económico, Universidad de Alicante.

Blank (1999) for a review), with employer discrimination assumed to drive the residual wage gap.

There are a number of studies, analysing the impact of individual social capital on labor market outcomes. Black, Haviland, Sanders, and Taylor (2006) show that the wage disparity is less an issue for not southern-born with college-educated parents. There is no consensus whether inter-ethnic contacts improve the labor-market outcomes (Kahanec and Mendola, 2007) or not (Danzer and Ulku, 2008). The positive effect of intermarriage (Meng and Meurs, 2006) may partly related to network integration as well.

However, the economic literature pays little systematic attention to how social context affects decisions by both employers and employees. By way of contrast, there is a more substantial literature in both sociology and political science which documents the impact of racial concentration in the local community on various outcomes. The proportion of black people in American communities is known to affect earnings **?**, occupational choices **?**, poverty **?**, schooling outcomes (Card and Rothstein, 2007) and unemployment rates **?**.

There is some evidence that the increasing proportion of racial minority is related to increasing wage disparities up to a maximum, beyond of what the disparity starts falling again (Huffman and Cohen, 2004).

To test these predictions, we introduce two new measures of social context: racial composition of the local community and racial homophily in social networks. We look at the black-white wage gap in the US and the wage gap between ethnic Estonians and Russians in Estonia. We find that both our social context measures are related to the residual gap. However, due to low number of observations, the results are not significant at the conventional levels.

The rest of the paper is divided as follows: In the next section we describe the related literature which focus on the network segregation measures, theoretical explanations of ethnic wage gap and related empirical results. Section **??** describes the datasets we are using and our empirical strategy. Section **??** presents the results, Section **??** includes discussion and Section 7 concludes.

2 Methodology

The main idea of the analysis is the following: as the first step we estimate the unexplained racial/ethnic disparities by regions, and in the second step we treat the estimated disparities as the new dependent variables (this parallels Charles and Guryan (2007) methodology). We focus on two types of labor market outcomes: wage and unemployment rate. Because of low number of unemployed blacks in the sample, we are only able to look at the unemployment differences in Estonian data.

2.1 Homophily

People often choose friends who are similar to them in important ways: females are more likely to choose other females, whites are more likely to choose other whites, and older individuals are more likely to choose friends of a similar age (**??**McPherson, Smith-Lovin, and Cook, 2001). This basic principal, captured in the aphorism "birds of a feather flock together", is termed homophily in the social network literature (**?**McPherson, Smith-Lovin, and Cook, 2001). Various types of segregation are extensively documented in labor markets (Hellerstein and Neumark, 2007), electronic communication (Leskovec and Horvitz, 2007) and friendship relations (Mayer and Puller, 2008).

Mathematically, the relative frequency with which individuals socially interact with other people who are highly similar to themselves may be captured in the homophily index. Take a population of size N, broken into two or more groups on the basis of personal characteristics. N_c then denotes the size of the group containing people with type t, and $w_c = \frac{N_t}{N}$ gives the relative proportion of people of type c in the population.

Now, let s_i equal the number of ties formed by individual i with *similar* type individuals, and d_i the number of friendships ties formed with *different* individuals. Given this, the homophily index is computed as:

$$h_i = \frac{s_i}{s_i + d_i} \tag{1}$$

The homophily index does not take into account the size of the group with members of type t. Indeed, given type-blind forming of ties, the expected value of h_i would be the relative size of the group individual i belongs to, w_i. Therefore, we look at *inbreeding homophily*, defined as (Currarini, Jackson, and Pin, 2008)[1]

$$IH_i = \frac{h_i - w_i}{1 - w_i} \tag{2}$$

Relatively larger groups have more opportunities to interact with similar other than smaller groups, a fact that should lead to greater homophily among larger majority groups **?**. Empirically, numerous studies suggest that racial homophily is related to the size of the minority group as a proportion of the population **?**. For example, homophily is lower among both minority and majority groups in schools where minority students make up only a small fraction of the student population. As the relative proportion of minority students increases, students are increasingly likely to choose friends from their own racial or ethnic group Currarini, Jackson, and Pin (2008).

2.2 Regression

Two-step approach As a first step, we estimate the wage gap by communities based on the common socio-economic characteristics. Because the size of the community samples, and the low number of racial minorities, we do not employ the commonly used Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973). Instead, we estimate a common wage regression for the complete dataset. We capture the wage gap by including the vector of community dummies $\boldsymbol{C}$, vector of racial dummies $\boldsymbol{R}$, and the community and race cross-effects $\boldsymbol{C} \cdot \boldsymbol{R}$. We also add the common socio-economic characteristics $\boldsymbol{X}$. We model the individual wage as

$$\log w_i = \alpha_0 + \boldsymbol{\alpha_C}'\boldsymbol{C}_i + \boldsymbol{\alpha}'_R\boldsymbol{R}_i + \boldsymbol{\alpha}'_{CR}\boldsymbol{C}_i \cdot \boldsymbol{R}_i + \boldsymbol{\beta}'\boldsymbol{X}_i + \varepsilon_i \tag{3}$$

The main variables of interest are the components of $\boldsymbol{\alpha}_{CR}$. We choose national sample as the reference. Accordingly, the components answer the question – what is the expected wage penalty (or gain) being black in the given community.

[1] Analogous measure is also called "effective segregation" (Hellerstein and Neumark, 2007) and "isolation index" (Hellerstein, McInerney, and Neumark, 2008).

	1st model
region	SCBS "community"/Estonian county. National sample is used as the baseline
black	race "non-hispanic black"/ethnicity: Estonian/Non-Estonian
female	gender
	2nd model
age	age, modelled as 4th order polynomial/age group
yedu	years of education/education group
	3rd model
married	living with partner
kids	children in the household, *gender* and *kids* cross-effects
non.citizen	not a US citizen/immigrant status
	4th model
spanish	interview conducted in Spanish/Estonian, Russian, English skills
work.hours	hours worked in average week
	5th model
industry, occupation	

Table 1: Explanatory variables, used for SCBS data

We use different sets of individual-specific variables in $\boldsymbol{X}$. This is because there is no consensus about the "right" set of explanatory variables, it also allows us to check the robustness of the results. We estimate 4 models. First of them only includes constant and gender; the following models include all the variables of previous model plus a few new ones (Table 1).

The second step involves regressing the regional wage gap on the regional inbreeding homophily and the percentage of the minority. We weight observations by the inverse of the variance of the wage gap.

Single-Step Approach We estimate the individual wage regression in the form:

$$\log w_i = \alpha_0 + \alpha_{IH} \cdot IH_{r_i} + \alpha_r \cdot R_i + \alpha_{IHR} \cdot IH_{r_i} \cdot R_i + \boldsymbol{\beta}' \boldsymbol{X}_i + \varepsilon_i, \qquad (4)$$

where IH_{r_i} is the estimated inbreeding homophily in the region of individual i, and α_0, α_{IH}, α_r, α_{IHR} and $\boldsymbol{\beta}$ are parameters. The main parameter of interest is α_{IHR}, which captures the effect of belonging to the minority depending on the homophily of the local region.

We cluster the standard errors by region.

3 Estonians and Russians in Estonia

Estonia is a former Soviet republic which houses a large Russian-speaking minority. There has been a substantial unexplained wage gap of 10-15% in favor of Estonian-speaking workers since mid-1990s (Leping and Toomet, 2008). The regional units we look at the current study are counties and municipalities. There are 15 counties in Estonia, population of which varies from 10,000 till

500,000; and the minority percentage from 0.01 to 0.80. The counties are good proxies for the local labor markets as they include a major urban center within a commuting distance of less than an hour for most of the inhabitants.

Estonia is administratively split to 241 districts (municipalities and settlements). Due to small number of observations in most of them, we retain only regions which contain at least 10 observations of both ethnic group, our final data includes 59 municipalities. Although the sample size is dramatically larger, municipalities are far less perfect proxies for the local labor market.

3.1 Data

The analysis is based on two different data sources: landline telephone communication for the network information and labor force survey for the wage gap analysis.

The telecommunication data originates from a landline telephone service provider. We observe all private telephone calls in the providers network during a single day in 2006. The data covers about 200,000 phones and 250,000 calls. The dataset includes the information, needed for billing the contract holders, like caller and receiver ID, and duration and time of the call. We also observe the location (district) of the phones. In addition, the information on the preferred language of the contract holder is collected by the telecom for marketing purposes[2]

The telecommunication data allows us to directly analyze the inter-ethnic communication. Although we observe just one of the possible communication channels, the previous research indicates, that use of different communication channels is highly correlated. Even more, the large dataset easily allows us to analyze the network characteristics at county- and in many cases at district level.

We consider the individuals linked if there is at least one call between them in the data. We exclude all the calls from/to another provider as we have no data on the caller/receiver. We also exclude the loops and phones with erroneous location data.

We calculate the regional homophily measures as explained in Section 2.1 above. We take into account all the calls, taken and received by residents of the region, including connections to other regions. The analysis based on intra-regional calls only did not reveal any substantial difference.

We compute the homophily and inbreeding homophily based on the telecommunication data. The relationship between inbreeding homophily IH and the minority percentage in counties and retained districts is given in Figure 2. As the inbreeding homophily varies dramatically more across the Estonian districts than across SCBS communities, we see two clear hump-shaped curves, one for the majority and another for the minority population (Figure 6). In particular, the regional inbreeding homophily for Russian-speaking persons ranges from about 0 to 0.6, a far larger variability that for the blacks in the SCBS data.

We perform a similar excercise for the clustering as well. We include all types of clustering, including inter-ethnic and inter-region links. Thereafter we compute the average over individual clustering coefficients for all the individuals of both ethnic group in every region. We present the region-specific averages and

[2]See Appendix A for the correspondence between census and telephone household data.

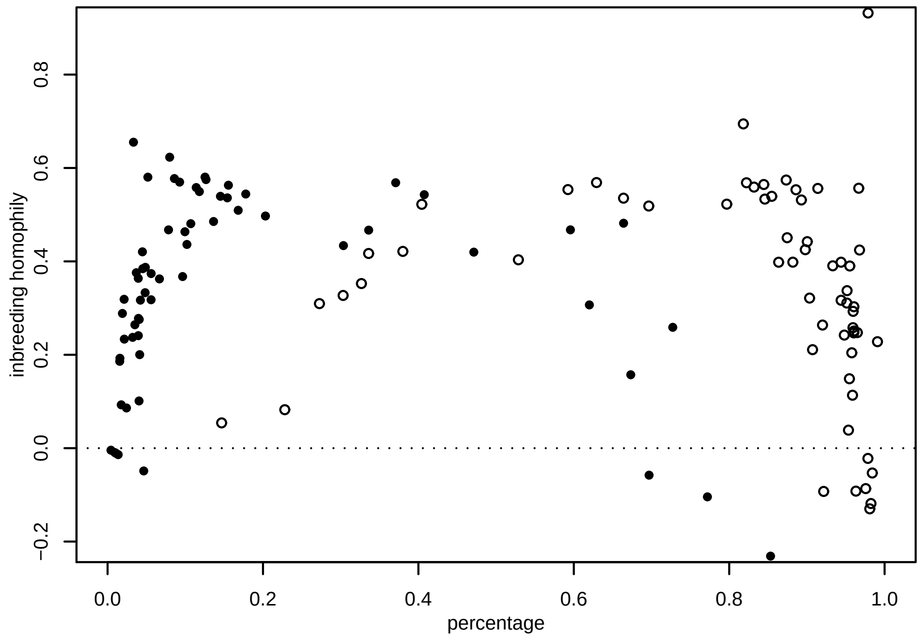

Figure 1: Relationship between the homophily and population percentage by counties. Telecommunication data. Black dots represent Russian-speaking, white dots Estonian-speaking households.

supplement the figure by lowess smoothers for both ethnic groups (Figure **??**). As the figure reveals, the clustering is initially increasing in the population group percentage, and slightly falling above the value about 0.5.

We estimate the income and unemployment models based on the Estonian Labor Force Survey (ELFS). ELFS is conducted quarterly as a semi-rotating panel. We employ information about the monthly salary at the main job, and information on common socio-economic characteristics. We perform the calculations based on both county and municipality of residence.

In order to increase the number of individual observations by smaller regions, we aggregate the individual ELFS observations between 2000-2007 for the counties and 2000-2006 for the administrative districts (we do not have data for districts for 2007). We exclude all the regions with less than 10 observations for each of the ethnic group, in this way we have 15 counties and 59 districts in the final sample. We look only at the individuals between 20 and 60 years of age.

3.2 Results

3.2.1 Unexplained wage differential

We estimate a Mincer-style wage equation, were we include the common socio-economic characteristics (in different combinations), such as gender, age, years of education, immigrant status, family status, and Estonian-, Russian-, and English language skills, industry and occupation. We also include a dummy for non-Estonian ethnicity and dummies for counties. The county-specific unex-

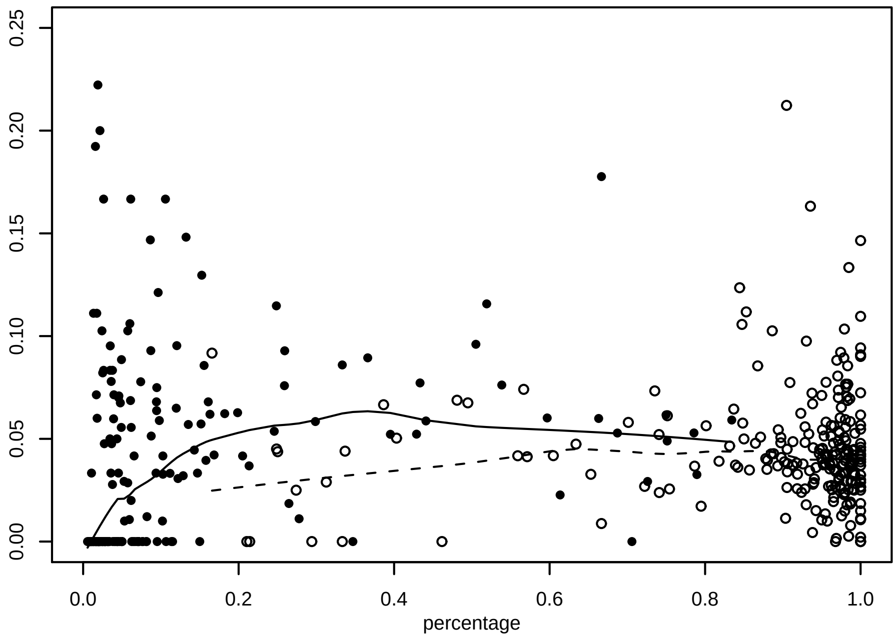

Figure 2: Relationship between clustering and population percentage by districts. Telecommunication data. Black dots represent Russian-speaking, white dots Estonian-speaking households. The lines represent lowess smoothers.

plained wage differentials are captured by cross-effectes between ethnicity and county dummies. We choose to present the gaps with respect to the average over all the regions (Suits, 1984).

Two-Stage Estimation We calculate the regional wage gap by pooled OLS and random effect models. As the OLS is strongly rejected by Breush-Pagan test, we focus on the random effect model below. We estimate analogous second-stage regional OLS as for the US (equation 6). The relationship between the gap and inbreeding homophily is negative for all the models, and statistically significant in all but one of the models (see also Figure 3). The relationship is remarkably stable with respect to including different sets of explanatory variables. The settlement-based model is also robust with respect to the time period under study (see Appendix B.1). The point estimates (Table 4) suggests that increasing the inbreeding homophily by 0.1 is associated to increase of wage gap by $0.1 \times 0.28 = 2.8$ percent.

Single-Stage Estimation The results for single-stage estimation are given in the lower panel of Table 4.

3.2.2 Unemployment rate differential

We estimate analogous models for unemployment, but not including industry and occupation. We employ probit and linear probability models with- and

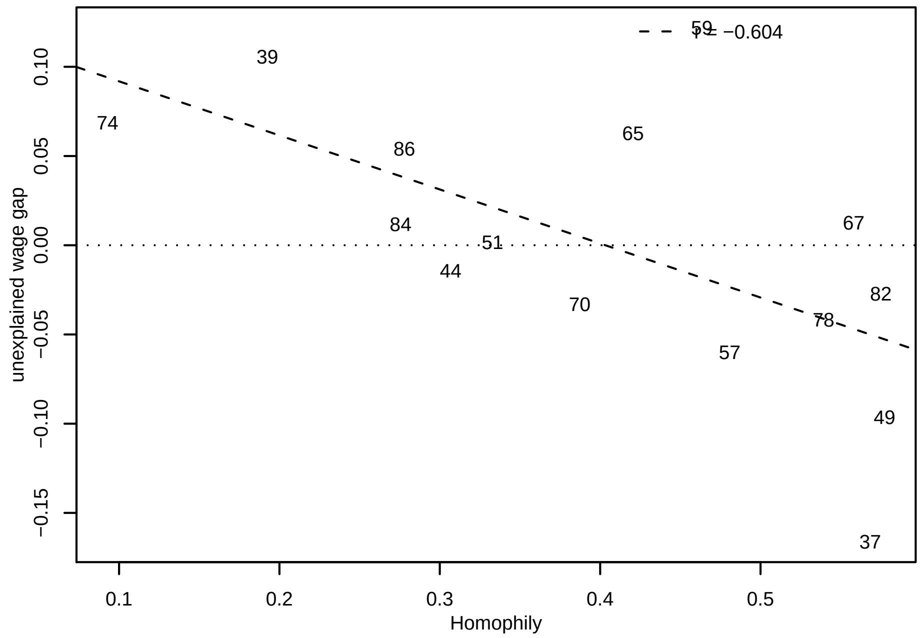

Figure 3: Relationship between the unexplained wage gap and inbreeding homophily. Estonian data.

without random effect (Figure 4 and Table 3). The robustness of the results with respect to the time period is assessed in the Appendix B.1.

The data reveals a positive correlation between network segregation and minority unemployment.

3.2.3 Clustering and wage gap

We also analyse the relationship between regional clustering and unexplained wage gap. Figure 5 shows that the seemingly strong correlation is related to a single outlier, the county 39 (which appears to be the smallest Estonian county). Removing the outlier, the relationship becomes insignificant when controlling for the minority percentage (Table **??**). Even more, clustering for counties and for municipalities show an opposite (although statistically insiginificant) sign.

4 Whites and Blacks in the United States

[Something else here.]Next, we describe the main data sources and relevant variables.

4.1 Data

The US analysis is based on two data sets: the 2000 Social Capital Benchmark Survey (SCBS) **?**, and the integrated public use microdata 5% sample from the 2000 Census (IPUMs) **?**. The Benchmark study provides information on racial attitudes and interracial contact in social networks. This information is

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
			Counties – pooled OLS		
Constant	0.126*	0.120*	0.118*	0.071*	0.033
	0.039	*0.016*	*0.020*	*0.025*	*0.041*
IH	-0.290*	-0.253*	-0.253*	-0.178*	-0.083
	0.120	*0.050*	*0.063*	*0.078*	*0.127*
Minority pct	-0.105	-0.163	-0.148	-0.047	-0.057
	0.261	*0.108*	*0.136*	*0.166*	*0.271*
R^2	0.413	0.782	0.688	0.377	0.057
# obs	15	15	15	15	15
			Counties – random effect		
Constant	0.131*	0.130*	0.130*	0.131*	0.078*
	0.031	*0.031*	*0.027*	*0.027*	*0.025*
IH	-0.272*	-0.270*	-0.282*	-0.284*	-0.159*
	0.088	*0.089*	*0.077*	*0.077*	*0.072*
Minority pct	-0.164	-0.166	-0.141	-0.141	-0.138
	0.179	*0.179*	*0.156*	*0.155*	*0.147*
R^2	0.552	0.547	0.623	0.627	0.424
# obs	15	15	15	15	15
			Districts – random effect		
Constant	0.167*	0.168*	0.167*	0.168*	0.178*
	0.052	*0.052*	*0.052*	*0.051*	*0.039*
IH	-0.289*	-0.290*	-0.290*	-0.294*	-0.337*
	0.124	*0.124*	*0.123*	*0.122*	*0.093*
Minority pct	-0.131	-0.131	-0.125	-0.125	-0.148
	0.140	*0.140*	*0.138*	*0.137*	*0.105*
R^2	0.112	0.113	0.114	0.118	0.230
# obs	59	59	59	59	59
			Districts, single-stage, OLS, clustered on region		
$IH \cdot R$	-0.423*	-0.423*	-0.405*	-0.405*	-0.221*
	0.173	*0.173*	*0.168*	*0.168*	*0.109*
R^2	0.290	0.290	0.298	0.298	0.461
			Explanatory variables		
constant, cubic time, gender	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
age, education		$\surd$	$\surd$	$\surd$	$\surd$
marriage, kids, immigrant status			$\surd$	$\surd$	$\surd$
language skills				$\surd$	$\surd$
industry, occupation					$\surd$

Note: standard errors in italics.
*: significant at 5% level

Table 2: Community-wise wage gap as a function of homophily and minority percentage. Estonia

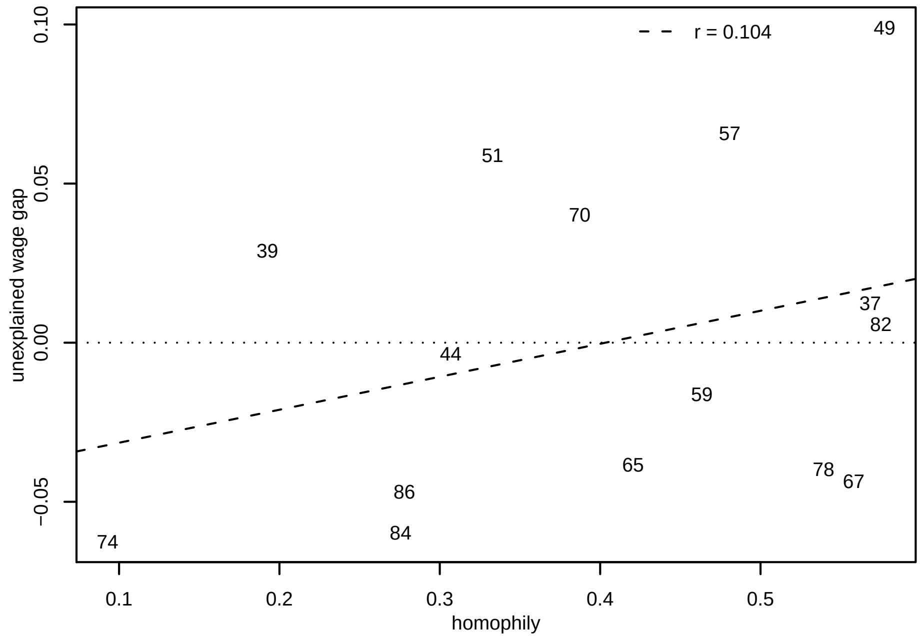

Figure 4: Relationship between the unexplained unemployment differential and inbreeding homophily. Estonian data.

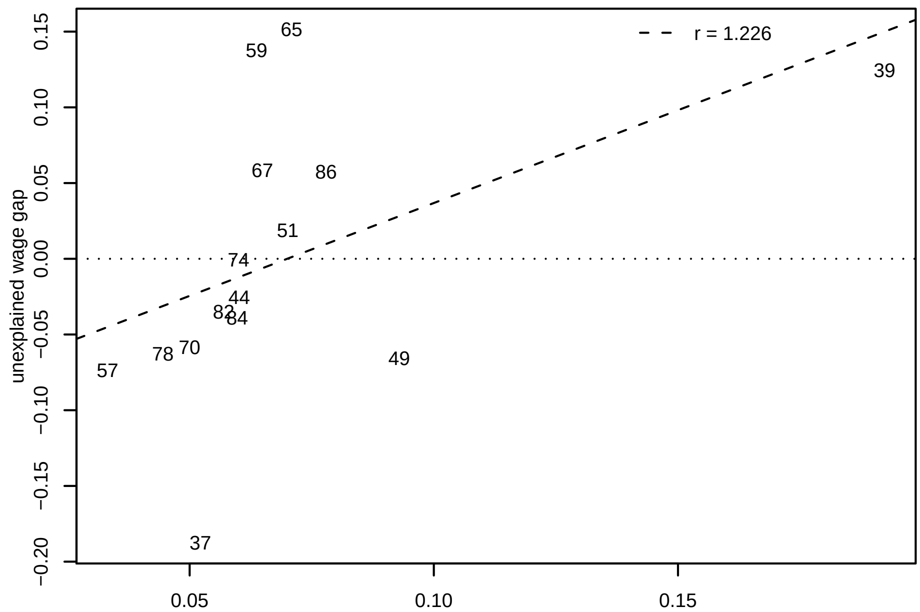

Figure 5: Relationship between the unexplained wage gap and clustering. Estonian data.

matched onto wage gap estimates obtained from the IPUMs data, and matched

Variable	Model 1	Model 2	Model 3	Model 4
Counties – LPM with random effect				
Constant	-0.060.	-0.049	-0.057.	-0.046
	0.030	*0.028*	*0.029*	*0.030*
IH	0.171.	0.152	0.175.	0.157
	0.095	*0.090*	*0.093*	*0.095*
Minority pct	-0.022	-0.037	-0.045	-0.077
	0.193	*0.182*	*0.188*	*0.190*
R^2	0.234	0.208	0.247	0.627
# obs	15	15	15	15
Districts – pooled probit				
Constant	0.075	0.066	0.060	0.054
	0.056	*0.059*	*0.060*	*0.060*
IH	0.057	0.063	0.076	0.060
	0.114	*0.118*	*0.115*	*0.116*
Minority pct	-0.116	-0.098	-0.095	-0.120
	0.097	*0.100*	*0.097*	*0.097*
R^2	0.034	0.029	0.037	0.043
# obs	76	76	76	76
Districts – LPM with random effect				
Constant	-0.076*	-0.078*	-0.082*	-0.073*
	0.036	*0.036*	*0.036*	*0.036*
IH	0.175.	0.179*	0.191*	0.174.
	0.089	*0.089*	*0.089*	*0.089*
Minority pct	0.035	0.041	0.044	0.022
	0.085	*0.086*	*0.085*	*0.086*
R^2	0.053	0.055	0.063	0.050
# obs	76	76	76	76
Explanatory variables				
constant, cubic time, gender	$\surd$	$\surd$	$\surd$	$\surd$
age, education		$\surd$	$\surd$	$\surd$
marriage, kids, immigrant status			$\surd$	$\surd$
language skills				$\surd$

Note: standard errors in italics.
*: significant at 5% level; .: at 10% level
LPM = linear probability model

Table 3: region-wise gap in unemployment rate as a function of homophily and minority percentage. Estonia

at the geographic level of the public use microdata area (PUMA).

SCBS is a telephone survey administered to approximately 30,000 adults living in 42 communities throughout the United States. A random sample of between 500 and 1500 respondents is available for each of the 42 communities, along with a 3000 person national sample and several area specific racial or

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Counties – random effect, including outlier					
Constant	-0.027	-0.026	-0.025	-0.025	0.007
	0.044	*0.044*	*0.040*	*0.040*	*0.039*
cl	0.836*	0.825*	0.810*	0.807*	0.284
	0.360	*0.361*	*0.327*	*0.326*	*0.309*
Minority pct	-0.210	-0.211	-0.216	-0.216	-0.186
	0.250	*0.250*	*0.227*	*0.227*	*0.224*
R^2	0.402	0.547	0.439	0.439	0.157
# obs	15	15	15	15	15
Counties – random effect, excluding outlier					
Constant	-0.078	-0.077	-0.075	-0.074	-0.040
	0.104	*0.104*	*0.094*	*0.094*	*0.093*
cl	1.628	1.627	1.592	1.586	1.029
	1.517	*1.519*	*1.377*	*1.374*	*1.361*
Minority pct	-0.201	-0.202	-0.208	-0.208	-0.178
	0.258	*0.259*	*0.234*	*0.234*	*0.231*
R^2	0.155	0.155	0.182	0.101	0.109
# obs	14	14	14	14	14
Municipalities – random effect					
Constant	0.089*	0.089*	0.088*	0.088*	0.081*
	0.038	*0.038*	*0.037*	*0.037*	*0.030*
cl	-0.580	-0.581	-0.583	-0.580	-0.489
	0.524	*0.524*	*0.513*	*0.513*	*0.415*
Minority pct	-0.129	-0.128	-0.127	-0.126	-0.179
	0.142	*0.142*	*0.139*	*0.139*	*0.113*
R^2	0.042	0.042	0.043	0.043	0.077
# obs	60	60	60	60	60
Explanatory variables					
constant, cubic time, gender	√	√	√	√	√
age, education		√	√	√	√
marriage, kids, immigrant status			√	√	√
language skills				√	√
industry, occupation					√

Note: standard errors in italics.
*: significant at 5% level

Table 4: Region-wise wage gap as a function of clustering and minority percentage. Estonia

ethnic oversamples.[3] The response rate varied considerably across the communities, and averaged around 30%. The survey contains various measures of

[3]Screens for African-American and hispanic samples were used in Rochester, Cuyohoga and National. Screens were also used to identify an additional 200 lower-income Rs in Boston, and respondents living within city boundaries in Greensboro city and Delaware.

personal social networks, attitudes, socio-economic background and income.

IPUMS is the public microuse data available from the U.S. Census bureau, and represents a sample of 5% of all Census respondents. [More on IPUMS] (For more on census geography, see).) A public-use microdata area, or PUMA, is is a contiguous geographic area of approximately 100,000 residents, and is the smallest area in the Census geography for which a sample of the public-use census data is available.

We computed wage-gap and network measures at two levels of aggregation. First, we worked with the boundaries defined by the creators of the SCBS sample. This approach results in XX communities, after excluding those where less than 10 minorities were surveyed in the sample. To increase both the number of communities and the quality of the income and wage-gap estimates, we matched the SCBS network data with wage-gap estimates based on the 2000 IPUMS 5% sample. To do this, we obtained the geographically sensitive codes directly from Roper, and used area matches generated by the the geographic translation database **?** to identify the PUMAs of all SCBS respondents. XXXX respondents had geographic information that could not be identified, and these respondents were excluded. Respondents from up to XX PUMAs were represented in any given SCBS sample community. The resulting dataset contained information on XX PUMAs, after excluding XX PUMAs with less than 10 minority respondents.

The source of non-network variables depends on the level of aggregation. For the wage analysis, we used variables on the monthly salary at the main job, and information on common socio-economic characteristics. These variables were taken from the SCBS for the community sample level, and from the IPUMS for the PUMA level. The fraction of a minority in a community in the SCBS communities is taken from census data by the designers of the survey, and calculated directly from the IPUMS data at the PUMA level.

The SCBS asks respondents a number of questions designed to find out more about their personal networks. Respondents are asked both about network composition, or the types of people in their networks, and about how much time they spend socialising with those people. There are unfortunately only two questions which measure inter-racial friendships. The first asks respondents asks respondents to report the time spent socialising with friends of a different race; the second asks whether they know at least one person from the indicated racial group (white, black, asian, hispanic).

To compute the racial homophily measure, we used the first question regarding time spent socialising with friends of a different race in combination with a general question about time spent socialising. The text of the two questions is as follows:

FRDVISIT How many times in the past twelve months have you had friends over to your home?

FRDRAC How many times in the past twelve months have you been in the home of a friend of a different race or had them in your home?

To make the two questions comparable, we assumed that socialising is reciprocal and doubled the estimated frequency of having friends over to one's home.

Thus, the individual homophily index is written as follows:

$$h_i = 1 - \frac{FRDRAC_i}{2 \cdot FRDVISIT_i}. \tag{5}$$

We truncate h to be in the interval $[0, 1]$.

To measure both the racial and general diversity of the respondent's personal network, we computed additive indices from 10 questions with the same general text:

DO YOU KNOW Thinking about friends...XX

The racial diversity index was based on responses to questions about friends in the four racial categories: white, black, asian, and hispanic. To measure general network diversity, we also included in the additive index whether or not the respondent had a friend in six additional categories: welfare recipient, business owner, gay, XX, XX, XX.

Previous studies have found that people with more diverse networks in general are likely to have more friends of a different racial background **?**. As expected, the calculated racial homophily index was moderately correlated with both racial and general network diversity at the individual level, and more highly correlated at the aggregate level. At the individual level, we found that respondents with reported more time spent socialising at home with friends of a different race also reported higher levels of both racial($r = 0.XX$) and general ($r = 0.XX$) diversity in their personal networks. Furthermore, racial homophily at the community level was highly correlated with the proportion of respondents who stated having at least one black friend ($r = 0.XX$), as well as the average network diversity within the community.

Inbreeding homophily – interaction relative to opportunity for interaction. Inbreeding homophily is computed by taking the homophily measure and subtracting the percentage of own race in city (adjusted to respondent's race.) Divide this term by 1 - percentage own race to yeild inbreeding homophily.

The individual inbreeding homophily IH_i is calculated in the same way as defined in (2) where we use the minority percentages in corresponding communities for w_i. We later aggregate both types of homophilies across the regions to get the region-specific network measures. We calculate the standard errors using the intra-community variation of homophily. Note that the percentage of whites and blacks does not sum to unity as there are more racial groups represented in the communities, omitted in this study.

We first present the relationship between the inbreeding homophily and the percentage for the blacks in different communities (Figure 6).

Most of the communities have quite a similar percentage of blacks (between 0 and 0.4), less so of whites (between 0.4 and 0.9). We can distinguish a familiar hump-shaped pattern (see Currarini, Jackson, and Pin, 2008, Figure 4). The curves for whites and blacks seem to fit well together. North Minneapolis forms a single outlier for blacks, the low end of whites have many more points of observation.

Currarini, Jackson, and Pin (2008) have shown that the likelihood of forming friendships with other individuals of the same race depends on the proportional population size of the racial group. This relationship follows a non-linear or hump-shaped pattern, with people most likely to choose friends of the same race

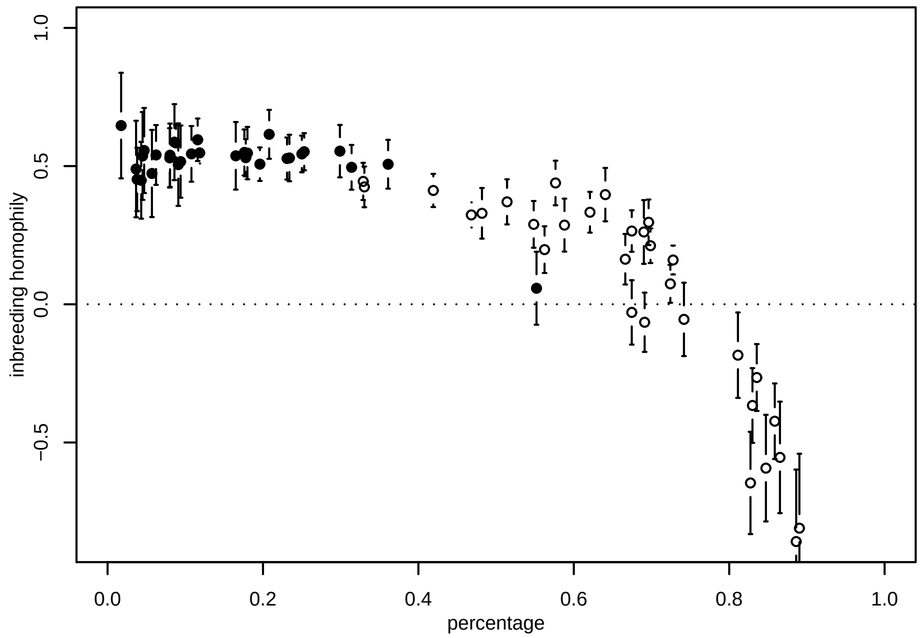

Figure 6: Relationship between the black homophily and percentage by communities. Black dots represent blacks and white dots whites. US data.

when their racial group makes up half of the total population. As can be seen in Figure **??**, . We do not replicate **?** finding that minority proportion affects network size, however. This is not entirely unexpected, given that they look friendships among teenagers who can choose friends from relatively small total populations of a few hundred, while the population from which our respondents choose friends is orders of magnitude larger.

4.2 Results

We approach the problem of estimating the associationship between the regional network segregation and wage gap in two different ways.

As SCBS reports income in intervals we use interval regression.

Below, we focus on the blacks. As the the outlier, North Minneapolis, has disproportionately large impact on the results, we exclude it from the results below. We Present a cross-plot of inbreeding homophily and the unexplained wage gap (based on Model 2) in Figure 7. We see a negative relationship (correlation = -0.2).

In order to quantify the relationship, we estimate an OLS model, explaining the wage gap by the inbreeding homophily and minority percentage:

$$\widehat{\Delta w_r} = \alpha_0 + \alpha_1 IH_r + \alpha_2 w_r + u_r, \qquad (6)$$

where $\widehat{\Delta w}$ is the estimated wage gap, w is the minority percentage and r denotes regions. We weight the observations by inverse of the estimated variance of $\widehat{\Delta w}$. As both inbreeding homophily and minority wage gap are related to the minority percentage, we regress the regional wage gap on both of these variables (Table 5,

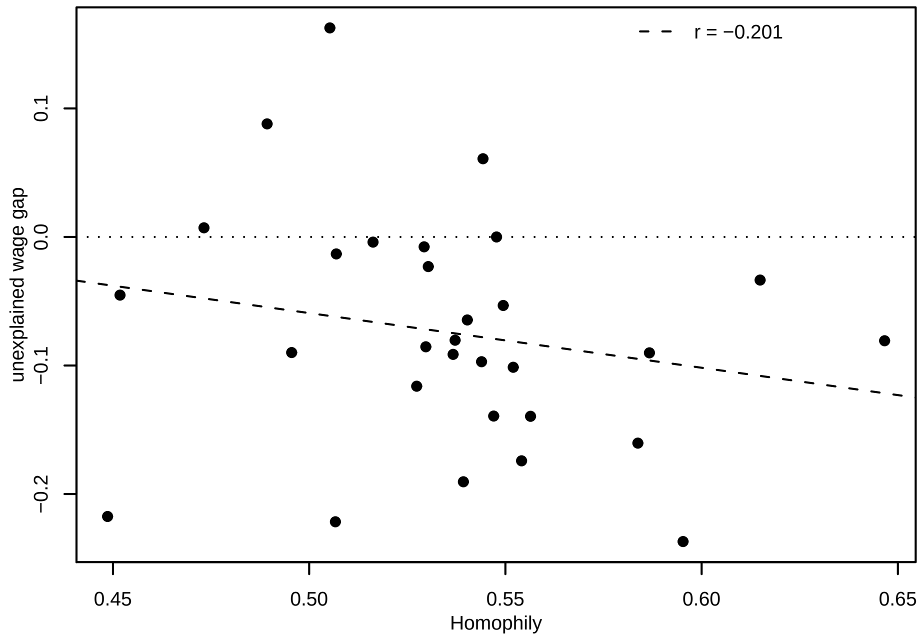

Figure 7: Relationship between the black-white wage gap and inbreeding homophily by communities. Wage gap based on model 2. US data.

upper panel). The table confirms that the wage gap and inbreeding are indeed negatively related. Even more, the relationship is negative and statistically significant even if we control for the minority percentage.

5 Potential Explanations

6 Unobserved Productivity Differences

Another possible explanation for this finding is related to unobserved geographical variables which are correlated to the wage gap. A previous analysis indicates that the black-white wage gap is related to regional industrial composition, the percentage of black and immigrant minority, union coverage and percentage of casual employment. These variables together explain around 10-25% of the variation of the wage gap across the U.S. metropolitan areas (McCall, 2001). Unfortunately, these measures are not available at the community level.

For both US and Estonian pooled OLS results, the relationship is most clear (has the highest sgnificance level) if we control for age and education, and also for family characteristics. Adding further controls for language, working hours, and industry/occupation will further dilute the relationship. Although the differences are not statistically significant, it hints that language skills and industry and occupation are tightly related to the network structure. Adding controls to these variables into the regression will implicitly also control for the network segregation. The random effect models, however, show surprisingly similar outcomes if industry and occupation are not included. Stronger network

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
		SCBS data – cross section OLS			
Constant	0.206	0.202	0.127	0.118	
	0.190	*0.150*	*0.175*	*0.168*	
IH	-0.530	-0.506	-0.422	-0.397	
	0.327	*0.258*	*0.300*	*0.288*	
Minority pct	-0.392	-0.249	-0.180	-0.181	
	0.264	*0.208*	*0.240*	*0.231*	
R^2	0.106	0.126	0.068	0.066	
# obs	29	29	29	29	
		Explanatory variables			
constant, cubic time, gender	√	√	√	√	√
age, education		√	√	√	√
marriage, kids, immigrant status			√	√	√
language skills				√	√
industry, occupation					√

Note: standard errors in italics.
*: significant at 5% level

Table 5: Community-wise wage gap as a function of homophily and minority percentage

segregation is related to minorities working in less wealthy industries.

6.1 Negative Stereotypes, Discrimination and Trust

The Becker (1957) taste-based discrimination theory has resulted in extensive literature on the mechanisms behind the racial wage gap. Here we review the literature related to social networks and labor market disparities. Trivially, if the Beckerian models suggest the income disparity to fall in network integration, given we assume the discriminatory taste is monotonically related to network integration. A number of models assume the presence of segregation (as Sattinger (1996)). Naturally, if higher network integration would eliminate the segregation, those models will break apart and income disparities will vanish.

A branch of statistical discrimination literature, initiated by Phelps (1972), explains discrimination by noisy signals about unobserved productivity and "cultural distance". The members of the majority group can more easily read the "signals" of their groupmates (see also Cornell and Welch, 1996; Lundberg and Startz, 2007). In these models network integration decreases racial disparities, if the integration can be integrated as the ability to read the minority signals.

Geographical variation has been used establishing a negative relationship between the minority wage and prejudices against minorities (Charles and Guryan, 2007; Waisman and Larsen, 2009).

Cobb Clark and Antecol (2006) show that self-reported discriminatory incidences are positively related to the percentage of other racial groups in the community for all the groups, except for blacks where the relationship is the opposite.

Social context is also known to mediate the link between negative attitudes towards minorities and community level indicators of social capital. While racial resentment) and negative implicit attitudes towards blacks are widespread throughout the American citizenry **??**, the social and political salience of race in local American communities varies depending on the racial composition of the community. In communities with large and politically active minorities, negative attitudes towards blacks are most likely to be translated into decreased levels of generalized and cross-racial social trust and decreased provision of education and other social services **??**. In communities with small racial minorities, negative racial attitudes have little impact on social trust or public policy. Racial composition of local communities also affects the provision of public goods in Africa **?**, and .... **?**.

Given the link between local social context and the political relevance of race (and ethnic bacground), it is likely that similar processes are at work in the local labor markets as well. Employers and employees may be more likely to actively discriminate against minorities when negative attitudes towards minorities are socially and politically salient. Furthermore, if increased political salience of minority status decreases educational expenditures in a community, social context may further affect the wage gap by decreasing the education, training and skill level of minority workers. Reductions in public welfare and educational spending would make it difficult for poorer members of a community to obtain the training and social support needed to get and keep good jobs.

Previous studies have also identified a relationship between interracial trust and minority proportion of the population. Again, the relationship is curvilinear: the lowest levels of interracial trust are typically found in cities where the minority and majority populations are of approximately equal size. There is some question as to whether the changes in trust are more generalized or apply only to interracial context. There are several potential measures of trust. [[edit below]] [[Need to discuss theoretical direction – is trust about implicit attitudes, or simply reflect homophily – which comes first?]]

6.2 Segregation Theories

However, the reverse causality is also feasible. The pre-existing income differential may lead to residential segregation and in this way to segregation of the networks as well.

The previous studies on the role of racial segregation and wage gap are inconclusive. According to Bayard, Hellerstein, Neumark, and Troske (1999), the black-white segregation explains a substantial part of the corresponding wage gap. However, Charles and Guryan (2007) finds the racial wage gap across the US states to be *negatively* related to the workplace segregation: as more segregated are the racial groups as smaller is the unexplained wage gap.

Coordination failure type of models (such as Mailath, Samuelson, and Shaked, 2000; Moro and Norman, 2004) do not include a segregation or cultural distance measure. However, certain generalisations do. Chaudhuri and Sethi (2008) shows that integrating the racial groups in presence of peer effects in human capital acquisition leads to less inter-group inequality.

There are several arguments for a non-monotonic relationship between network segregation and income disparities.

Kahanec (2006) develops a model where ethnic groups invest in different types of qualification, depending on the network size and social distance. In his model, integration helps minorities to gain access to large majority network. However, it also evades the gain from ethnic specialisation. If the elasticity of substitution of different types of labor is large, the latter effect dominates and integration leads to falling relative income for the minority. In two-dimensional segregation model (the dimensions are income and race), Sethi and Somanathan (2004) show that least segregated are middle-income-gap communities. Given high income disparities, blacks cannot afford to move to the (rich) white neigbourhood. If the disparities are low, it does not pay off in terms of better (richer) neighbourhood.

6.3 Structural Network Theories

We establish a negative relationship between network segregation and unexplained wage gap. Although this analysis is not able to determine the causality of the relationship, it still gives a few suggestions. A natural explanation for the outcome is the flow of information on job openings in the networks. Less connections between the majority and minority social networks makes it less likely that the members of the (less well off) minority community will be able to apply (either formally or informally) to the good jobs.

Social context might affect opportunities or preferences for social interaction between members of the minority and majority groups. Members of minority groups might find it more difficult to find jobs if they lack the weak social ties to provide information about available jobs (Granovetter, 1973). Minority groups who prefer to form friendships with those from a similar racial or ethnic background (racial homophily) might inadvertently limit their access to job information. Even more, lack of social ties between the minority and majority groups may be related to screening discrimination (Cornell and Welch, 1996; Lundberg and Startz, 2007) and in this way limit the minority access to high-skilled jobs.

There is a large body of empirical literature about the job search channels and job quality at individual level. The results are either inconclusive (see e.g. Mahuteau and Junankar, 2008; Loury, 2006) or support the importance of networks (Cingano and Rosolia, 2006). However, several theoretical considerations point to a possible link between the groups' labor market outcomes and how tightly their networks are linked.

If the employed individuals pass job information to their unemployed friends, different initial unemployment may lead to persistent wage gap across non-connected networks (Calvo-Armengol and Jackson, 2007; Fontaine, 2008).

A number of studies analyzes the relationship between the unexplained wage gap and network properies. Ioannides and Loury (2004) in their review note that the use of friends and relatives for job search differs across racial and ethnic groups. However, they stress that it is difficult to interpret that variation. Hellerstein, McInerney, and Neumark (2008) suggest that race matters – low-skilled blacks get jobs only when employer hire other blacks. Royster (2007) identifies several mechanisms which put black vocational school graduates at a disadvantaged situation when entering the labor market, including lack of access to as many and powerful contacts as their white peers. Those contacts were party established in “male and all-white spaces”, such as bars and taverns.

Our analysis is related to the literature on social isolation and economic outcomes (Postlewaite and Silverman, 2005). Although the members of the minority group are not necessarily deprived of social contact, the lack of weak ties to the majority group may lead to analogous weak labor market outcomes as the case of social isolation. However, Danzer and Ulku (2008) do not find any evidence that strong ties to the members of the majority group improve the (individual) income of the minority households.

7 Conclusions

We analyze the relationship between network segregation (inbreeding homophily) and unexplained wage gap. We use data for two very different societies and labor markets – we look at racial differences in the U.S. and ethnic differences in Estonia.

We employ three data sources: Social Capital Benchmark Survey 2000 (SCBS) for the US network and income measures, telecommunication data for the Estonian network- and Estonian Labor Force Suvey (ELFS) for the income measures.

We establish that unexplained wag gap is negatively related to network segregation. Less contacts between the racial or ethnic groups is related to larger unexplained gap in favor of the majority group. The negative correlation persists even when controlling for the minority percentage.

A Language codes in the telephone and census data

The coding of language is subject of several types of errors. First, language may not be coded, or coded in a wrong way. In most cases, the language information is collected only if it is not Estonian (the official language in Estonia). Second, as the landline phones are household specific, multilingual household are coded as monolingual. Third, there may be systematic difference in use of landline phones by different ethnic groups. However, we have no evidence on this.

Here we compare the percentage of telephones, coded as used by Russians, with different measures on minority households by counties using year 2000 census data. The Statistics Estonia divides households into single- and multi-language household according to the language. In the table 6 we compare the broadest and narrowest measure of households where Russian might be considered as the telephone language. The broadest measure include all the households where at least one language other than Estonian is spoken (column NE), the narrowest are the single-language households where Estonian is not spoken (column NES).

In most cases, the percentage of Russian-lanugage phones (w_2) is remarkably similar to the narrow measure of non-Estonian households. The main exceptions are Harju (the capital area) where w_w is between the broad and narrow measure, and Russian-dominated Ida-Viru, where the percentage of Russian-language phones falls short of the narrow measure by 4 percentage point. We may conclude that the language codes for the phones correspond well to the census household language data.

B Robustness analysis

B.1 Time period for Estonia

We analyse the dependence of the estimated parameters on the selected time period. We choose all yearly intervals between 1997 and 2006 for counties and 1997-2006 for settlements. This results in 66 different intervals (55 for settlements), the longest one being 1997-2007 (1997-2006), the shortest ones being the 11 (10) individual years. We depict the 95% confidence bounds on the IH term in the regression (6) (Figure 8). Upper confidence bound is given in left and lower in right panel. The negative values are depicted in red, positive in white. Note the that there is no data in the lower-right part of the figure as no interval ends before it starts. The area around the diagonal is most noisy as the intervals are shortest there and hence the number of observations is low. We focus on the model 2, the figures for the other models were qualitatively similar.

The figure indicates that the statistically significant negative relationship between Δw and IH is present only for a number of relatively long observation periods. For most of the intervals, the relationship is not statistically significant (the upper bound is positive, the lower bound negative). However, the relationship based on settlements is negative at the 95% confidence level for all the periods, except the shortest ones.

Analogous figure for the unemployment regression (Figure **??**) is somewhat less robust.

Region	NE	NES	w_2
Estonia	0,352	0,249	
Harju	0,440	0,307	0,371
Hiiu	0,041	0,016	0,016
Ida-Viru	0,828	0,663	0,620
Jõgeva	0,134	0,080	0,086
Järva	0,101	0,045	0,048
Lääne	0,163	0,093	0,107
Lääne-Viru	0,204	0,114	0,099
Põlva	0,082	0,043	0,045
Pärnu	0,164	0,096	0,114
Rapla	0,109	0,050	0,049
Saaremaa	0,032	0,011	0,018
Tartu	0,213	0,139	0,145
Valga	0,217	0,133	0,127
Viljandi	0,100	0,045	0,041
Võru	0,088	0,046	0,040

Table 6: Different measures of non-Estonian households.

Notes:
NE: percentage of households in which a language, other than Estonian, is spoken
NES: pecentage of household in which only a single language but not Estonian is spoken
w_w: percentage of non-Estonian household in the telephone data.

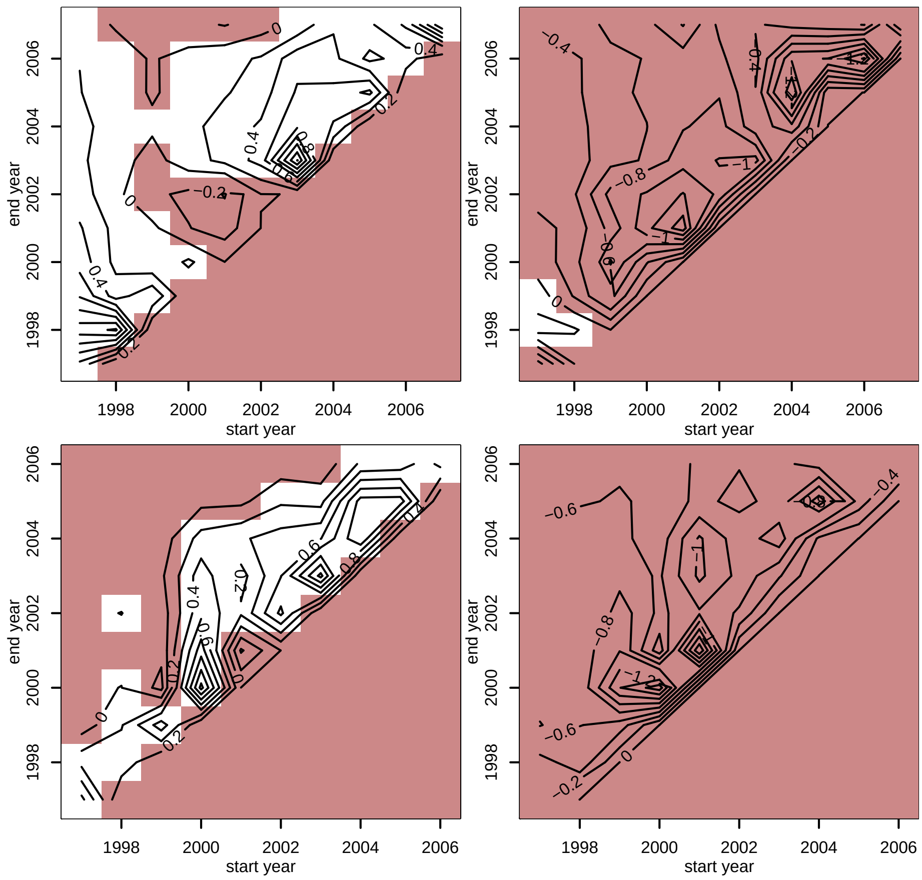

Figure 8: Wage regression, random effect model. 95% confidence bounds of the parameter estimates. Upper bound on left-, and lower bound on the right panel. Negative values in red, positive in white. County-based model in the upper, and settlement-based model in the lower panel.

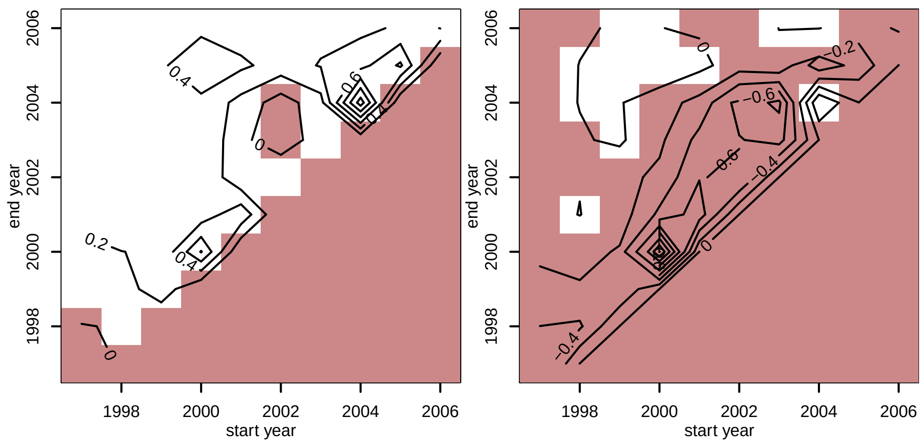

Figure 9: Unemployment regression, linear probability model with random effects. 95% confidence bounds of the parameter estimates. Upper bound on left-, and lower bound on the right panel. Negative values in red, positive in white. Settlement-based model.

References

Altonji, J. G., and R. M. Blank (1999): "Race and gender in the labor market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3C, chap. 48, pp. 3143–3259. Elsevier Science Publishers B.V.

Bayard, K., J. Hellerstein, D. Neumark, and K. Troske (1999): "Why are Racial and Ethnic Wage Gaps Larger for Men than for Women? Exploring the Role of Segregation," Working Paper 6997, National Bureau of Economic Research.

Becker, G. S. (1957): *Economics of Discrimination.* Chicago University Press, Chicago, IL.

Bhumaik, S. K., I. N. Gang, and M.-S. Yun (2006): "Ethnic conflict and economic disparity: Serbians and Albanians in Kosovo," *Journal of Comparative Economics*, 34, 754–773.

Black, D., A. Haviland, S. Sanders, and L. Taylor (2006): "Why Do Minority Men Earn Less? A Study of Wage Differentials among the Highly Educated," *Review of Economic Studies*, 88(2), 300–313.

Blackaby, D., D. Leslie, P. Murphy, and N. O'Leary (2005): "Born in Britain: How are native ethnic minorities faring in the British labour market?," *Economics Letters*, 88, 370–375.

Blinder, A. S. (1973): "Wage discrimination: Reduced form and structural estimates," *Journal of Human Resources*, 8(4), 436–455.

Calvo-Armengol, A., and M. O. Jackson (2007): "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of Economic Theory*, 132, 27–46.

Card, D., and J. Rothstein (2007): "Racial segregation and the black-white test score gap," *Journal of Public Economics*, 91(11-12), 2158–2184.

Charles, K. K., and J. Guryan (2007): "Prejudice and the economics of discrimination," Working Paper 13661, NBER, 1050 Massachusetts Avenue, Cambridge, MA 02138, USA.

Chaudhuri, S., and R. Sethi (2008): "Statistical Discrimination with Peer Effects: Can Integration Eliminate Negative Stereotypes?," *Review of Economic Studies*, 75, 579–596.

Cingano, F., and A. Rosolia (2006): "People I know: Workplace networks and job search outcomes," Temi di discussione 600, Banca d'Italia.

Cobb Clark, D. A., and H. Antecol (2006): "Racial and Ethnic Discrimination in Local Consumer Markets: Exploiting the Army's Procedures formatching Personnel to Duty Locations," Working Paper 544, The Australian National University, Centre for Economic Policy Research.

Cornell, B., and I. Welch (1996): "Culture, Information and Screening Discrimination," *Journal of Political Economy*, 104(3), 542–571.

Currarini, S., M. O. Jackson, and P. Pin (2008): "An Economic Model of Friendship: Homophily, Minorities and Segregation," *Econometrica*, forthcoming.

Danzer, A. M., and H. Ulku (2008): "Determinants of Integration and its mpact on the Economic Success of Immigrants: A Case Study of the Turkish Community in Berlin," Discussion Paper 3762, IZA, P.O.Box 7240, 53072 Bonn, Germany.

Fontaine, F. (2008): "Why are similar workers paid differently? The role of social networks," *Journal of Economic Dynamics and Control*, 32, 3960–3977.

Giddings, L. (2002): "Has the shift toward markets hurt ethnic minorities? Changes in ethnic earnings differentials in Bulgaria's early transition," *International Journal of Manpower*, 23(1), 9–31.

Granovetter, M. S. (1973): "The strength of weak ties," *American Journal of Sociology*, 78(6), 1360–1380.

Hellerstein, J. K., M. McInerney, and D. Neumark (2008): "Measuring the Importance of Labor Market Networks," Working Paper 14201, National Bureau of Economic Research.

Hellerstein, J. K., and D. Neumark (2007): "Workplace Segregation in the United States: Race, Ethnicity, and Skill," *SSRN eLibrary*.

Huffman, M. L., and P. N. Cohen (2004): "Racial Wage Inequality: Jobs Segregation and Devaluation across U.S. Labor Markets," *American Journal of Sociology*, 109(4), 902–936.

Ioannides, Y. M., and L. D. Loury (2004): "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, 42, 1056–1093.

Kahanec, M. (2006): "Ethnic Specialization and Earnings Inequality: Why Being a Minority Hurts but Being a Bog Minority Hurts More," Discussion Paper 2050, IZA, P.O.Box 7240, 53072 Bonn, Germany.

Kahanec, M., and M. Mendola (2007): "Social Determinants of Labor Market Status of Ethnic Minorities in Britain," Discussion Paper 3146, IZA, P.O.Box 7240, 53072 Bonn, Germany.

Leping, K.-O., and O. Toomet (2008): "Emerging ethnic wage gap: Estonia during political and economic transition," *Journal of Comparative Economics*, forthcoming.

Leskovec, J., and E. Horvitz (2007): "Worldwide Buzz: Planetary-Scale Views on an Instant-Messaging Network," Technical Report MSR-TR-2006-186, Microsoft Research, Redmont, WA.

Loury, L. D. (2006): "Some Contacts Are More Equal than Others: Informal Networks, Job Tenure, and Wages," *Journal of Labor Economics*, 24(2), 299–318.

LUNDBERG, S., AND R. STARTZ (2007): "Information and Racial Exclusion," *Journal of Population Economics*, 20(3), 1432–1475.

MAHUTEAU, S., AND P. N. JUNANKAR (2008): "Do Migrants get Good Jobs in Australia? The Role of Ethnic Networks in Job Search," *Economic Record*, 84, S115–S130.

MAILATH, G. J., L. SAMUELSON, AND A. SHAKED (2000): "Endogenous Inequality in Integrated Labor Markets with Two-Sided Search," *American Economic Review*, 90(1), 46–72.

MAYER, A., AND S. L. PULLER (2008): "The Old Boy (and Girl) Network: Social Network Formation on University Campuses," *Journal of Public Economics*, 92(1-2), 329–347.

MCCALL, L. (2001): "Sources of Racial Wage Inequality in Metropolitan Labor Markets: Racial, Ethnic, and Gender Differences," *American Sociological Review*, 66(4), 520–541.

MCPHERSON, M., L. SMITH-LOVIN, AND J. M. COOK (2001): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Socioloy*, 27, 415–444.

MENG, X., AND D. MEURS (2006): "Intermarriage, Language, and Economic Assimilation Process: A Case Study of France," dp 2461, IZA, P.O.Box 7240, 53072 Bonn, Germany.

MORO, A., AND P. NORMAN (2004): "A general equilibrium model of statistical discrimination," *Journal of Economic Theory*, 114, 1–30.

OAXACA, R. L. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14(3), 693–709.

PHELPS, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62(4), 659–661.

POSTLEWAITE, A., AND D. SILVERMAN (2005): "Social isolation and inequality," *Journal of Economic Inequality*, 3, 243–262.

ROYSTER, D. A. (2007): "What Happens to Potential Discouraged? Masculinity Norms and the Contrasting Institutional and Labor Market Experiences of Less Affluent Black and White Men," *The ANNALS of the American Academy of Political and Social Science*, 609(1), 153–180.

SATTINGER, M. (1996): "Search and discrimination," *Labour Economics*, 3, 143–167.

SETHI, R., AND R. SOMANATHAN (2004): "Inequality and Segregation," *Journal of Political Economy*, 112(6), 1296–1321.

SUITS, D. B. (1984): "Dummy Variables: Mechanics V. Interpretation," *Review of Economic Studies*, 66(1), 177–180.

WAISMAN, G., AND B. LARSEN (2009): "Do attitudes toward immigrants matter?," Stocholm University, Copenhagen School of Business.