

Toomet, Ott; van der Leij, Marco; Rolfe, Meredith

**Working Paper**

## Social Networks and Labor Market Inequality between Ethnicities and Races

Tinbergen Institute Discussion Paper, No. 12-120/II

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* Toomet, Ott; van der Leij, Marco; Rolfe, Meredith (2012) : Social Networks and Labor Market Inequality between Ethnicities and Races, Tinbergen Institute Discussion Paper, No. 12-120/II, Tinbergen Institute, Amsterdam and Rotterdam, <https://nbn-resolving.de/urn:nbn:nl:ui:29-441459>

This Version is available at:

<https://hdl.handle.net/10419/87561>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

TI 2012-120/II

Tinbergen Institute Discussion Paper



# Social Networks and Labor Market Inequality between Ethnicities and Races

*Ott Toomet<sup>1</sup>*

*Marco van der Leij<sup>2</sup>*

*Meredith Rolfe<sup>3</sup>*

<sup>1</sup> *Department of Economics, Tartu University;*

<sup>2</sup> *CeNDEF, Faculty of Economics and Business, University of Amsterdam; De Nederlandsche Bank, and Tinbergen Institute;*

<sup>3</sup> *Department of Management, London School of Economics.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Social Networks and Labor Market Inequality between Ethnicities and Races\*

Ott Toomet<sup>†</sup>    Marco van der Leij<sup>‡</sup>    Meredith Rolfe<sup>§</sup>

November 12, 2012

## Abstract

This paper analyzes the relationship between unexplained racial/ethnic wage differentials on the one hand and social network segregation, as measured by inbreeding homophily, on the other hand. Our analysis is based on both U.S. and Estonian surveys, supplemented with Estonian telephone communication data. In 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 gaps.

JEL codes: J71, J31, Z13

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

---

\*Corresponding Author: Ott Toomet. E-mail: [otoomet@ut.ee](mailto:otoomet@ut.ee). This paper benefitted from comments by Sebastian Buhai, Sanjeev Goyal, Tomasz Mickiewicz, Betsy Sinclair, Tom Snijders, and audiences at Aarhus Business School, Utrecht University, Tartu University, University of Alicante, VATT, and a Workshop on Networks in Oxford in 2008 and 2009. We are grateful to Jane Roberts and the Roper Center for assistance with the Social Capital Benchmark Data. The research was supported by the FAMNET project in EQUALSOC research program. Marco van der Leij would like to thank the Spanish Ministry of Science and Innovation (project SEJ2007-62656) and the NWO Complexity program for financial support. Ott Toomet gratefully acknowledges support from the Estonian Ministry of Science and Education project SF01800037s08. Views expressed are those of the authors and do not necessarily reflect official positions of De Nederlandsche Bank.

<sup>†</sup>Department of Economics, Tartu University

<sup>‡</sup>CeNDEF, University of Amsterdam; De Nederlandsche Bank; and Tinbergen Institute.

<sup>§</sup>Department of Management, London School of Economics.

# 1 Introduction

It has been consistently found that on average, members of ethnic or racial minorities earn less than those of the majority group. For example, Altonji and Blank (1999) report that blacks earn 21% lower hourly wages compared to whites in the United States in 1995. Similar wage gaps also characterize a large number of other groups, for instance whites and Hispanics in the U.S. (Altonji and Blank, 1999), Blacks and Pakistanis in the U.K. (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 Blank (1999) for a review), with employer discrimination assumed to drive the residual wage gap.

A couple of objections against discrimination-based explanations have been raised, however, most famously the paradox that competitive market forces should drive out discriminatory employers and thus lead to the extinction of wage gaps (Arrow, 1972, 1998). Arrow (1998) suggests that social networks, that is, job contacts and other social relations, may provide an alternative explanation for the existence of wage gaps.

The role of social networks in the labor market has been historically the domain of sociologists, see for example Granovetter (1974) for a classic reference. More recently, economists also have devoted attention to this topic. Montgomery (1991) considers a labor market model in which firms employ the social contacts of existing employees to find suitable workers. Job candidates that apply to a vacancy without having a contact at the firm suffer from adverse selection, and hence, firms offer higher wages to referred job candidates. Calvó-Armengol and Jackson (2004) explicitly model the social network as a graph of social links between employed and unemployed workers, in which employed workers forward vacancies to unemployed friends. They show that the probability of finding a job crucially depends on one’s position in the social network; workers in ‘central’ positions have higher employment rates.

There is indeed ample empirical evidence that social networks matter for jobseekers, see Ioannides and Loury (2004) for an overview. On average about half of all workers find their jobs through friends and relatives (Montgomery, 1991). Job search through social contacts is also cheaper and produces more job offers than formal methods (Holzer, 1988). More generally, the sociological and economic literature has shown that one’s individual network position matters for socioeconomic outcomes with regard to manager achievement (Burt, 1992), technological diffusion (Coleman, Katz, and Menzel, 1966), or school performance (Calvó-Armengol, Patacchini, and Zenou, 2009). However, while there is considerable micro-level evidence of the relevance of social network structure on individual economic outcomes, little is known at the macro-level. That is, with the exception of Eagle, Macy, and Claxton (2010), there is no research on whether global structural properties of a community’s social network are related to aggregate economic outcomes.

Existing micro-level evidence, however, is insufficient to establish the existence of a macro-level relationship between network structure and economic outcomes. The insufficiency of existing research follows from the well-documented

failure of macro level economic phenomena to represent mere aggregation of micro-level preferences. For instance, in case of discrimination and segregation, Schelling (1978) shows that small micro-level discriminatory preferences regarding segregation can be amplified to large residential segregation outcomes at the macro-level. On the other hand, Becker (1957) shows that large discriminatory preferences on average do not necessarily result in large discriminatory wage gaps at the macro level. Market forces and labor market sorting drive minorities to the least prejudiced employers, such that, eventually, the wage gap at the macro-level may be rather small.

The aim of this paper is to provide empirical evidence of a macro-level relationship between network structure and economic outcomes. Our economic outcome of interest is the well-documented gap in labor market outcomes and wages between members of racial or ethnic minority and majority groups. We focus on the relationship between the ethnic wage gap and a critical structural network property: inbreeding homophily. Inbreeding homophily is the tendency of individuals to have friendships with similar others, in our case, the tendency of minority (majority) group members to have more friendships with members of their own minority (majority) group, relative to the proportion of same ethnicity friendships one would expect to observe if friends were chosen randomly. Theoretical network models suggest that inbreeding homophily is a crucial factor causing labor market inequality (Montgomery, 1991; Calvó-Armengol and Jackson, 2004; Van der Leij and Buhai, 2008). We examine this empirically.

We employ data from two different countries, Estonia and the United States. In Estonia, we consider wage gaps between Estonians, the native majority group, and Russians, a minority group comprising about 30% of the population in Estonia. We employ a unique data set in order to measure structural properties of the network, inbreeding homophily in particular. This data set consists of a one-day log of fixed-line telephone calls of the main telephone provider in the country. The data set includes information on the location of the sender and receiver of the calls, and most importantly, their ethnic identity. This allows us to construct a measure of inbreeding homophily in each region, which we then compare to the wage gaps between Estonians and Russians in each region in Estonia. For the United States, we consider the wage gaps between blacks and whites at the metropolitan level. In order to obtain a metropolitan measure of inbreeding homophily, we employ data from the Social Capital Benchmark Survey (SCBS), a survey containing information on the social and racial diversity of friendships, income and employment, among others.

Our main result is that unexplained wage gaps at the regional level are larger in regions with a higher degree of inbreeding homophily. In other words, regional wage gaps are larger in regions where residents are more likely to choose same race or ethnicity friends. This result holds for the relationship between network structure and wage gaps in both Estonia and the United States. These results are robust to introducing controls for the relative size of a minority within a region, a measure that has often been employed in the literature on race and economic effects (Card and Rothstein, 2007).

Our paper provides the first empirical evidence at the macro-level relating network segregation to inequality at the labor market. Earlier work by Eagle, Macy, and Claxton (2010) has found that measures of network diversity at the community level are related to the social and economic development of a community. Eagle et al. also derive network structure measures from national

(U.K.) telephone data at the community level. In contrast to Eagle, Macy, and Claxton (2010), however, we focus on the relationship between ethnic and racial inbreeding homophily and ethnic and racial *inequality*. The Estonian telephone call dataset is unique as it also has information on the ethnicity of the telephone holder (as proxied by preferred language). Thus, we can measure the ethnic segregation of social networks directly, while Eagle et al. focus on network diversity in terms of time, space and structural holes.

We do not identify the causal direction of the relation between wage gaps and inbreeding homophily, nor do we intend to. Such causal identification is difficult in general, and in particular with non-experimental non-longitudinal data (Manski, 2000). Instead, we provide a discussion of potential mechanisms leading to the above-mentioned relationship, namely discrimination and prejudice, residential and occupational segregation, and network effects. Future research should provide causal evidence for these mechanisms.

The rest of the paper is divided as follows. Section 2 discusses the network measure of homophily, and explains our empirical strategy. Section 3 presents the results for Estonia and Section 4 those for the United States, Section 5 includes discussion and Section 6 concludes.

## 2 Methodology

In this section we, first, discuss the social network measure that we focus on, inbreeding homophily. Next, we explain the empirical strategy. The main idea of our strategy is the following: in 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 the methodology of Cutler and Glaeser (1997) and Charles and Guryan (2008).

### 2.1 Homophily and inbreeding homophily

People often choose friends who are similar to them in important ways (Blau, 1977; McPherson, Smith-Lovin, and Cook, 2001). This basic principle, captured in the aphorism “birds of a feather flock together”, is termed homophily in the social network literature (Lazarsfeld and Merton, 1954; McPherson, Smith-Lovin, and Cook, 2001). Various types of segregation have been documented in labor markets (Hellerstein and Neumark, 2007), electronic communication (Leskovec and Horvitz, 2007) and friendship relations (Mayer and Puller, 2008).

Following Currarini, Jackson, and Pin (2009), the relative frequency with which individuals socially interact with other people who are similar to themselves is captured by the homophily index.<sup>1</sup> Take a population  $\mathcal{N}$  of size  $N$ , partitioned into two or more groups on the basis of personal characteristics, such as race or ethnicity.  $N_t$  then denotes the size of the group  $\mathcal{N}_t$  containing people of type  $t$ , and  $w_t = \frac{N_t}{N}$  gives the relative proportion of individuals of type  $t$  in the population.

Now, let  $s_i$  equal the number of ties formed by individual  $i \in \mathcal{N}$  with *similar* type individuals, and  $d_i$  the number of friendships ties formed with *different*

---

<sup>1</sup>Homophily is a measure of the isolation dimension of segregation according to the classification by Massey and Denton (1988).

individuals. Given this, the homophily index is computed as:

$$h_i = \frac{s_i}{s_i + d_i} \quad (1)$$

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,  $w_{t_i}$ . A simple and intuitive way to adjust for the group size results in *inbreeding homophily*,<sup>2</sup> defined as (Coleman, 1958):

$$IH_i = \frac{h_i - w_{t_i}}{1 - w_{t_i}} \quad (2)$$

$IH_i$  can range between  $\frac{-w_{t_i}}{1-w_{t_i}}$  (when a person has a strict preference for friendships with out-group members, termed heterophily) and 1 (when a person only forms friendships with in-group members), where  $IH_i = 0$  indicates no racial or ethnic bias (i.e., the ethnic composition of a person’s friendship group mirrors the ethnic composition of the regional population.) The inbreeding homophily index of group  $t$  is simply the average inbreeding homophily of its members

$$IH(t) = \frac{1}{N_t} \sum_{i \in N_t} IH_i.$$

In our analysis, we focus on the inbreeding homophily index of the minority group.

Empirically, several studies suggest that inbreeding homophily is non-monotonically related to the size of the minority group as a proportion of the population (Blau, 1977; Moody, 2001; Currarini, Jackson, and Pin, 2009). 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. Hence, inbreeding homophily levels are highest when ethnic minorities constitute a substantial proportion (e.g., 35-65%) of the local population.

## 2.2 Empirical strategy

Our aim is to relate the minority inbreeding homophily index of a community to the unexplained gap between majority and minority wages within that community. For this purpose, we follow a two-step strategy as in Cutler and Glaeser (1997) and Charles and Guryan (2008).

As a first step, we estimate the wage gap by communities based on the common socio-economic characteristics. Because of the small size of the community samples and the low number of racial minority members, we do not employ the commonly used Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973). Instead, we estimate a common wage regression for the complete data set. We capture the wage gap by including the vector of community dummies  $C$ , vector

---

<sup>2</sup>Analogous measures are also called “eta<sup>2</sup>” (Massey and Denton, 1988), “effective segregation” (Hellerstein and Neumark, 2007) and “isolation index” (Hellerstein, McInerney, and Neumark, 2008).



of ethnic minority dummies  $R$ , and the community and ethnicity cross-effects  $C \cdot R$ . We also add the common socio-economic characteristics  $X$ . We model the individual wage as

$$y_i = \alpha_0 + \alpha'_C C_i + \alpha'_R R_i + \alpha'_{CR} C_i \cdot R_i + \beta' X_i + \varepsilon_i. \quad (3)$$

We also include individual random effects in case of Estonian panel data. The parameter of the community-ethnicity cross-effects,  $\alpha_{CR}$ , captures the unexplained wage gap within the community, and the estimate of this parameter,  $a_{CR}$ , is our measure of interest. The estimation method will depend on the dependent variable, either a continuous or categorical wage variable, and will be discussed in the sections on the results.

We use different sets of individual-specific variables in  $X$ . This is because there is no consensus about the “right” set of explanatory variables; it also allows us to assess the robustness of the results. We estimate a number of models. The first model only includes a constant and gender; each following model nests the previous model and includes additional variables, see Appendix A for details on the sets of explanatory variables for both the Estonian and U.S. data.

The second step involves regressing the community wage gap on the inbreeding homophily variable,  $IH_c$ , and the percentage of the ethnic minority within community  $c$ ,  $w_c$ :

$$a_{CR,c} = \gamma_0 + \gamma_1 IH_c + \gamma_2 w_c + u_c \quad (4)$$

Note that each observation corresponds to a community  $c$ . We estimate this equation by weighted least squares, in which we weight observations by the inverse of the variance of  $a_{CR,c}$ , obtained in the first-stage wage regression<sup>3</sup>.

Our primary interest is the estimate of the parameter  $\gamma_1$ . If this coefficient is significantly different from zero, it suggests that social network structure, in this case inbreeding homophily, are (not necessarily causally) related to wage gaps at the *community* level.

## 3 Estonians and Russians in Estonia

### 3.1 Background

Estonia is a small country in North-Eastern Europe, bordering Russia and Latvia, with a population of about 1.3 million. Before the Second World War, almost 95% of the inhabitants were of ethnic Estonian background speaking Estonian, a language rather different from Russian. During the early years of the war, the country was seized by Soviet troops and later incorporated into the Soviet Union. The years of brutal Stalinist regime destroyed the relations between Estonians and Russians, which had been quite friendly up to that point.

<sup>3</sup>As an alternative to the two-step procedure, it may seem attractive to perform the estimation in a single stage by estimating the individual wage regression in the form:

$$y_i = \alpha_0 + \alpha_{IH} \cdot IH_{c_i} + \alpha_r \cdot R_i + \alpha_{IHR} \cdot IH_{c_i} \cdot R_i + \beta' X_i + \eta_i, \quad (5)$$

where  $IH_{c_i}$  is the estimated inbreeding homophily in the community of individual  $i$ , and  $\alpha_0$ ,  $\alpha_{IH}$ ,  $\alpha_c$ ,  $\alpha_{IHR}$  and  $\beta$  are parameters. In that case,  $\alpha_{IHR}$  would be the main parameter of interest. However, if the true model is specified by equations (3) and (4), then a regression of model (5) would face an endogeneity problem. This can be seen by noting that under (3) and (4), the error term,  $\eta$ , would be a compound of  $\varepsilon$  and  $u'R$ , and hence,  $\eta$  would be correlated with the ethnic dummies,  $R$ . Hence, the two-stage procedure is more appropriate.

After the war, new industrial workplaces attracted many workers from the neighboring Soviet republics. This led to a substantial immigration of mostly Russian-speaking people, with Russian speakers constituting almost 40% of the Estonian population by the final years of the Communist Regime. Estonian language and culture were not directly threatened by the Soviet state (although use of Russian was strongly promoted), but the steady increase of Russian-speaking population gradually eroded the domains where Estonian was commonly spoken. In a number of industries, such as mining and merchant fleet, the army, and most of the large firms, the Estonian language was of little use. As a result, the native population was increasingly concerned about the future of their culture and identity. The authoritarian regime was unwilling to address the concerns of native Estonians, leading to a further deterioration in relations between members of the two language groups. Ethnic Estonians and Russians lived in a largely segregated country, working in largely segregated workplaces and following media outlets in their native language.

Estonia became independent again following the coup of August, 1991. The new country immediately began to implement policies that encouraged the use and teaching of Estonian while the Russian language, no longer compulsory in the school curriculum, began a retreat from public life. The new authorities introduced several measures favoring the native population, including granting citizenship based on knowledge of Estonian, now the sole official language of the country. After two decades of independence, the language groups are still far from being integrated. The relationship between them is commonly considered “normal” though somewhat tense in periods. Most notably, tensions exploded to large-scale riots in Tallinn in spring 2007. In everyday life, the two ethnic groups live in separate social worlds with limited intergroup contact. The separate worlds are also reflected in media which may present quite different viewpoints depending on the language (Korts and Kõuts, 2002). There is little evidence about direct discrimination, but a survey suggests that 37% of Russian speakers perceive discrimination to be common (Pettai, 2002).

In terms of labor market outcomes, ethnic Estonians earn about 15% more than Russians even after controlling for observable characteristics (Leping and Toomet, 2008). The wage gap between the two groups did not exist during the years of Soviet rule, but emerged shortly after Estonia was granted independence. The wage gap has not decreased over time, but is today roughly the same size it was when it first appeared in the early 1990s. It is noteworthy that knowledge of Estonian language does not seem to have a major impact on the size of the gap except at the very bottom of the wage distribution (Toomet, 2011), suggesting that language skills are not sufficient to erase the disadvantages experienced by Russian speakers.

Below, we investigate whether the above mentioned wage differential is related to inbreeding homophily in Estonia at the regional level. The regional units we look at are counties and municipalities. The country is administratively divided into 15 counties, population of which varies from 10,000 to 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 an urban center within commuting distance of less than an hour for most of the inhabitants. Since the number of counties is small, we repeat the analysis at the finer level of municipalities as a robustness check. Estonia is administratively split to 241 municipalities. Due to small number of observations in most of them, we retain only regions which contain at

least 10 observations of both ethnic group and hence our final data includes 55 municipalities. Although the sample size is dramatically larger, municipalities are far less perfect proxies for the local labor market.

### 3.2 Data

The analysis is based on two different data sources: landline telephone communication for the network data and a labor force survey to measure wage gaps.

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 450,000 calls. The dataset includes information that is needed for billing the contract holders, like caller and receiver ID, and duration and time of the call. We also observe the location (district and municipality) of the phones. In addition, the information on the preferred language of the contract holder, either Estonian or Russian, is collected by the telecom company for marketing purposes.

Our data are potentially subject to several types of errors that might bias our measure of inbreeding homophily derived from the telephone data and ethnic information. First, language may not be recorded on all accounts, or may be coded incorrectly. In most cases, the language information is collected only if it is not Estonian (the official language in Estonia). Second, as the land-line phones are household specific, multilingual households are coded as monolingual. Third, there may be systematic difference in use of land-line phones by different ethnic groups.

In order to assess whether any of these potential errors might bias our network measure, we compare the percentage of telephones with Russian-language contract holders to different measures on minority households by counties using year 2000 census data. Statistics Estonia divides households into single- and multi-language households according to the language. In Table 1 we compare the broadest and narrowest measure of households where Russian might be considered as the telephone language. The broadest measure includes 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-language phones ( $w_2$ ) is remarkably similar to the narrow measure of non-Estonian households. The main exceptions are Harju (the capital area) where  $w_2$  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 conclude that the language codes for the phones correspond well to the census household language data, and therefore we do not believe that our derived measure of inbreeding homophily is systematically biased.

We use the telecommunications data to directly analyze inter-ethnic communication. Although we observe just one of the possible communication channels, use of different communication channels is highly correlated (Haythornthwaite, 2005). Moreover, the large data set allows us to analyze network characteristics at the regional level.

We construct an undirected network from the telephone call data. Any phone that was used on the day of analysis constitutes a node. We consider

Table 1: Different measures of non-Estonian households.

Region	NE	NES	$w_2$
Estonia	0.352	0.249	0.270
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

Notes:

NE: percentage of households in which a language, other than Estonian, is spoken

NES: percentage of household in which only a single language but not Estonian is spoken

$w_w$ : percentage of non-Estonian household in the telephone data.

two individuals linked if there is at least one call between them in the data.<sup>4</sup> We exclude all the calls from/to another provider as we have no data on the caller/receiver. We also exclude loops, phones with erroneous location data, and phones that made more than 10 calls.<sup>5</sup>

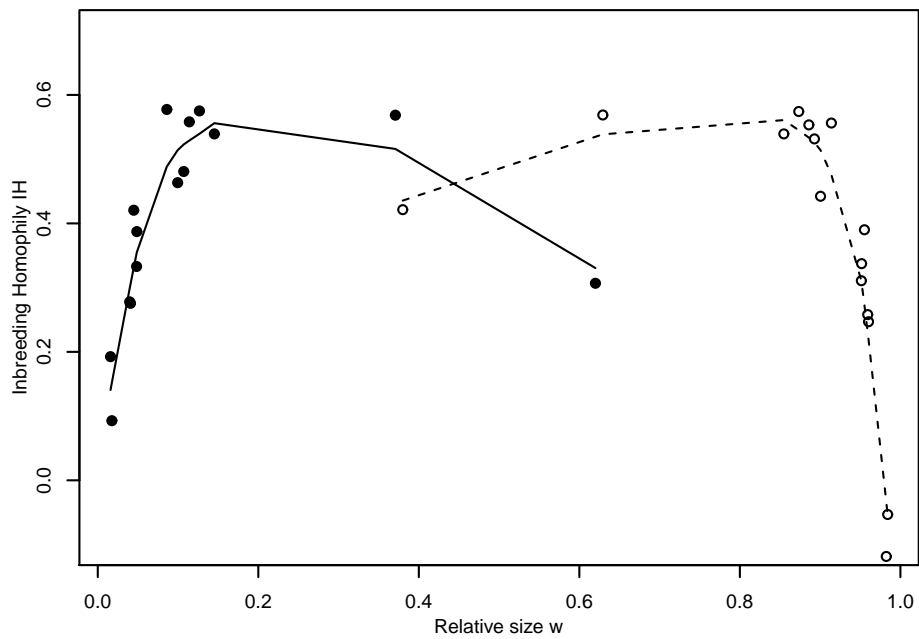
We calculate the regional homophily measures as explained in Section 2.1 above. We take into account all the calls, taken and received by residents of the region, including connections to other regions. The analysis based on intra-regional calls only did not reveal any substantial differences.

The relationship between inbreeding homophily  $IH$  and the minority percentage in counties and municipalities is given in Figure 1. We see two clear hump-shaped curves, one for the majority and another for the minority population, similar to those reported in Currarini, Jackson, and Pin (2009). In particular, the regional inbreeding homophily for Russian-speaking individuals ranges from about 0 to 0.6.

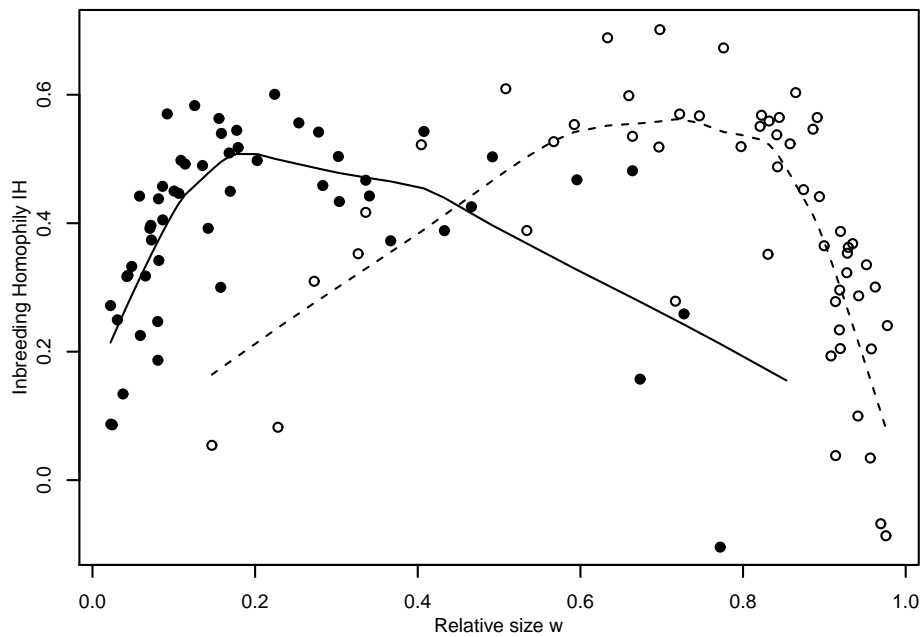
We now turn to the data for the wage regression models. We estimate the income models based on the Estonian Labor Force Survey (ELFS). The ELFS is

<sup>4</sup>We ignore the direction of the link in order to obtain a sufficient amount of observations. Results calculated on bi-directional ties only, were qualitatively the same.

<sup>5</sup>Although the sample only contains private phones, a large number of those are presumably used for business purposes. Below, we excluded all phones with calls to/from more than 10 different phones, removing 8.7% of the phones. The results are robust with respect to this threshold.



(a) Counties



(b) Municipalities

Figure 1: Relationship between inbreeding homophily and population percentage by Estonian counties (upper panel) and municipalities (lower panel). Telecommunication data. Black dots represent average values across the counties for Russian-speaking households, white dots are those for Estonian-speakers. Lines are loess smoothers.

conducted quarterly as a semi-rotating panel where each individual is surveyed 4 times over a  $1\frac{1}{2}$  year time span. In all, around 4 000 individuals are interviewed each quarter. The most important variables in our study are the following. First, we observe the *ethnic nationality*, the ethnic background. Most people in the country identify themselves rather clearly with a distinct ethnic group which we code either as Estonian or non-Estonian. The vast majority of the latter use Russian as their everyday language, and hence below we use non-Estonians and Russians interchangeably. Next, we employ information about the *last net monthly salary at the main job*. Income from other jobs or other sources is not included. The survey also provides information about language skills, in particular the ability to communicate in Estonian. We introduce corresponding dummies for Estonian as well as for English and Russian. Finally, we include common socio-economic controls, such as age, education, and family status.

In order to increase the number of individual observations by smaller regions, we aggregate the individual ELFS observations between 2000-2010 for the counties (2000-2006 for municipalities because of data limitations). We look only at the individuals between 20 and 60 years of age. Table 2 reports the average income (in Estonian kroons) by ethnic group, it's relative difference, sample size, and percentage of Russian-speaking workers for each county. We only present data for workers with positive reported wage. Analogous table for the municipalities is given in the Appendix C.

Table 2: Income and percentage of Russian-speaking workers by counties.

county	wage E	wage R	difference	N	pct R
Harju	6663	4872	-0.27	13364	0.43
Hiiu	4033	3129	-0.22	2033	0.01
Ida-Viru	4185	3551	-0.15	6584	0.84
Järva	4515	3642	-0.19	1717	0.07
Jõgeva	4564	3577	-0.22	2334	0.05
Lääne	4386	3868	-0.12	1307	0.11
Lääne-Viru	4478	4213	-0.06	3941	0.16
Pärnu	4143	4659	0.12	1696	0.04
Põlva	4491	3799	-0.15	2585	0.13
Rapla	4863	4115	-0.15	1958	0.06
Saare	4493	3453	-0.23	2064	0.02
Tartu	5076	3933	-0.23	4574	0.16
Valga	4137	3843	-0.07	2145	0.16
Viljandi	4376	3676	-0.16	3238	0.06
Võru	4257	3712	-0.13	1635	0.05

Notes: Only individuals with positive wage reported.

“wage E” and “wage R”: average wage of Estonian- and Russian speakers respectively.

“difference”: relative difference of these wage measures

“N”: the number of observations in each county.

“pct R”: the relative percentage of Russian-speakers, based on the telephone data.

The table gives a picture of the counties, which vary widely in size and percentage of the minority population. The largest county is Harju (the capital region) with more than 13,000 observations, the smallest is Lääne with about 10 times fewer wage-earners in the sample. The Russian-speaking workers earn less in almost all the regions. Average Russian wages are 10-20% lower in most regions, with the largest difference of 27% observed in Harju county. The only region where Russian-speakers earn more than native Estonians on average is Pärnu county.

### 3.3 Results on wage gaps

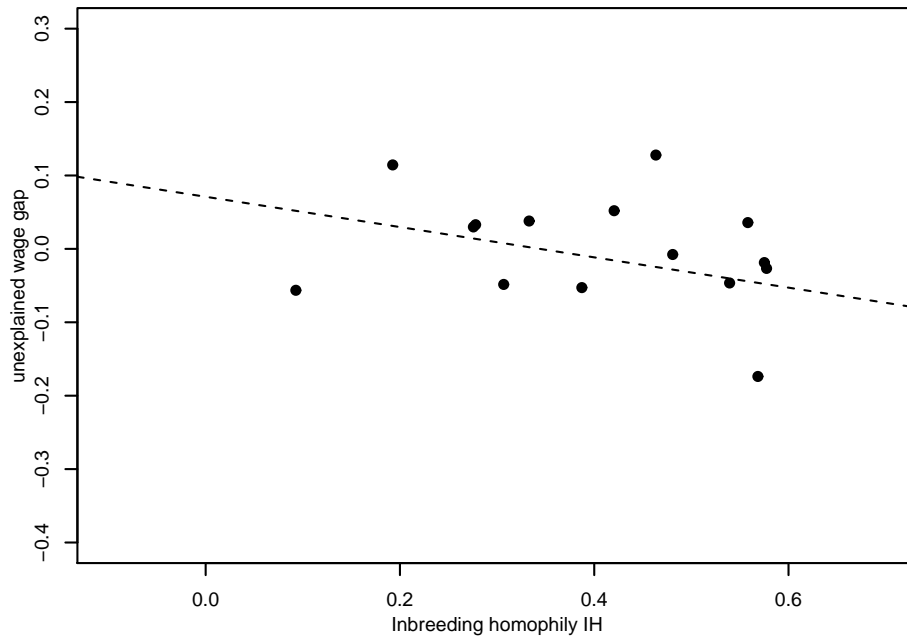
We first investigate the relation between unexplained wage differentials and homophily. We estimate a Mincer-style wage equation, where we include common socioeconomic characteristics (in different combinations), such as gender, age, years of education, immigrant status, family status, and Estonian, Russian and English language skills, industry and occupation (see Appendix A). We also include a dummy for ethnicity and dummies for regions. The region-specific unexplained wage differentials are captured by the ethnicity dummy in the corresponding regions (the cross-effects between ethnicity dummy and regional dummies). We choose to present the gaps relative to the average over all the regions (as in Suits, 1984).

We calculate the regional wage gap by pooled ordinary least squares (OLS) and random effect models. As the pooled OLS model is strongly rejected by the Breusch-Pagan test, we focus on the random effect model below. Results of the first stage wage regression with random effects for model 2 at the county level are given in Table 7 in Appendix A.

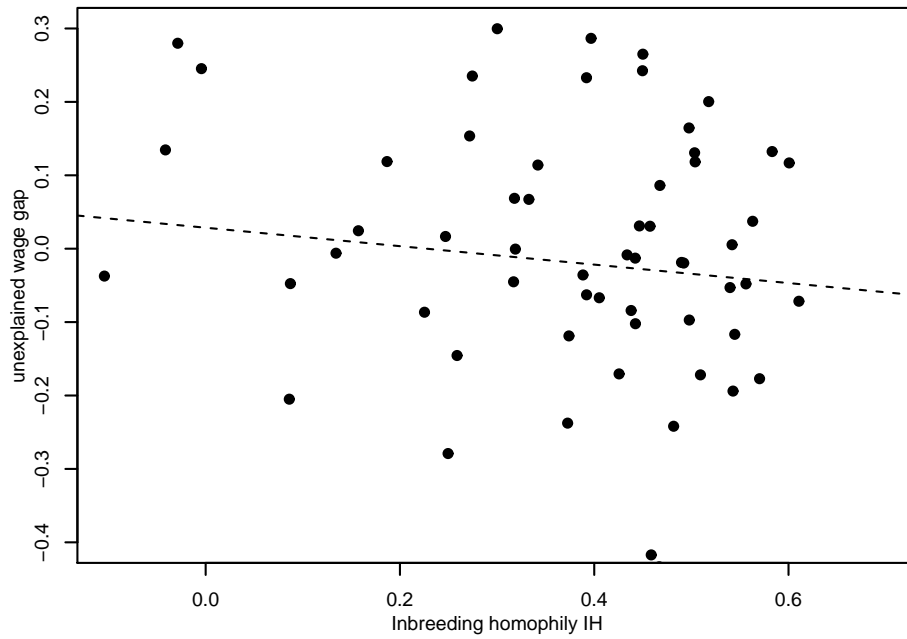
In the second stage, we regress the estimated regional wage gaps, weighted by the corresponding variance estimates, on inbreeding homophily and minority fraction, both measured from the telephone call data set. The results are shown in Table 3. The upper panel shows the results at the county level, and the lower panel at the municipality level. The relationship between the gap and inbreeding homophily is negative and statistically significant for all the models. The relationship is remarkably stable with respect to including different sets of explanatory variables in the first stage wage regression. Moreover, the estimated effect is large. The point estimate of -0.363 (Table 3, Model 2, counties, random effect) suggests that a switch from a county without inbreeding homophily (random matching) to a fully segregated county is associated with an increase of the wage gap by 30 percentage points.

The relationship is represented graphically in Figure 2. The negative slope is rather clear. We can also see that the unexplained gaps roughly correspond to the average differences in Table 2 (note that in the Figure, we plot the differences with respect to the country average).

The coefficient is lower for model 5, which includes industry and occupation controls, suggesting that industry and occupation of work are partly related to networks. Estimates of the pooled OLS model are of similar magnitude but slightly lower.



(a) Counties



(b) Municipalities

Figure 2: Relationship between the unexplained wage gap and inbreeding homophily across Estonian counties (upper panel) and municipalities (lower panel). The unexplained wage gap is computed from Model 2 using a random effects estimator



Table 3: Community-wise wage gap as a function of homophily and minority percentage. Estonia

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Counties: pooled OLS					
Constant	0.294*** <i>0.085</i>	0.245*** <i>0.072</i>	0.245*** <i>0.070</i>	0.222*** <i>0.069</i>	0.178*** <i>0.050</i>
<i>IH</i>	-0.599*** <i>0.163</i>	-0.461*** <i>0.137</i>	-0.464*** <i>0.134</i>	-0.414*** <i>0.132</i>	-0.327*** <i>0.095</i>
Minority pct	-0.315*** <i>0.097</i>	-0.351*** <i>0.082</i>	-0.354*** <i>0.080</i>	-0.308*** <i>0.079</i>	-0.275*** <i>0.057</i>
$R^2$	0.637	0.6855	0.699	0.649	0.721
# obs	15	15	15	15	15
Counties: random effect					
Constant	0.237** <i>0.092</i>	0.198** <i>0.083</i>	0.200** <i>0.080</i>	0.178** <i>0.080</i>	0.152** <i>0.063</i>
<i>IH</i>	-0.473*** <i>0.178</i>	-0.363** <i>0.160</i>	-0.369** <i>0.154</i>	-0.323** <i>0.153</i>	-0.271** <i>0.120</i>
Minority pct	-0.279** <i>0.111</i>	-0.298*** <i>0.099</i>	-0.300*** <i>0.095</i>	-0.259*** <i>0.095</i>	-0.239*** <i>0.074</i>
$R^2$	0.512	0.524	0.548	0.480	0.545
# obs	15	15	15	15	15
Municipalities: pooled OLS					
Constant	0.165*** <i>0.063</i>	0.162*** <i>0.059</i>	0.168*** <i>0.058</i>	0.138** <i>0.058</i>	0.108** <i>0.043</i>
<i>IH</i>	-0.353*** <i>0.109</i>	-0.310*** <i>0.101</i>	-0.326*** <i>0.100</i>	-0.272*** <i>0.099</i>	-0.215*** <i>0.075</i>
Minority pct	-0.254*** <i>0.083</i>	-0.288*** <i>0.077</i>	-0.287*** <i>0.076</i>	-0.242*** <i>0.076</i>	-0.189*** <i>0.057</i>
$R^2$	0.2184	0.2481	0.2576	0.2	0.2143
# obs	55	55	55	55	55
Municipalities: random effect					
Constant	0.148** <i>0.067</i>	0.135** <i>0.062</i>	0.140** <i>0.061</i>	0.112* <i>0.060</i>	0.093* <i>0.050</i>
<i>IH</i>	-0.309*** <i>0.117</i>	-0.263** <i>0.109</i>	-0.277*** <i>0.107</i>	-0.227** <i>0.106</i>	-0.191** <i>0.087</i>
Minority pct	-0.255*** <i>0.090</i>	-0.264*** <i>0.084</i>	-0.263*** <i>0.082</i>	-0.218*** <i>0.082</i>	-0.190*** <i>0.067</i>
$R^2$	0.1801	0.19	0.1994	0.1481	0.1614
# obs	55	55	55	55	55
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

## 4 Whites and Blacks in the United States

We now turn to the relationship between wage differentials and inbreeding homophily for whites and blacks in the United States. First, we describe our data sources and the relevant variables.

### 4.1 Data

The U.S. analysis is based on the 2000 Social Capital Benchmark Survey (SCBS), set up by the Saguro Seminar of Harvard University. The Benchmark study provides information on racial attitudes and interracial contact in social networks. SCBS is a telephone survey administered to approximately 30,000 adults living in 42 communities throughout the United States. A community typically corresponds to a metropolitan area or a county. A random sample of between 500 and 1500 respondents is available for each of the 42 communities, along with a 3000 individual national sample and several area specific racial or ethnic oversamples.<sup>6</sup> The response rate varied considerably across the communities, and averaged around 30%. The survey contains various measures of personal social networks, attitudes, socioeconomic background and income.<sup>7</sup> We exclude those areas where less than 10 black respondents were surveyed.

Three communities, St. Paul metropolis, Minneapolis and North Minneapolis, all lie in a single metropolitan area, twin city Minneapolis-St. Paul, and cannot be considered separate labor regions. Minneapolis-St. Paul is the only metropolitan area in the SCBS in which several communities were sampled, and we therefore analyze the data with and without the three communities in Minneapolis-St. Paul. This leaves us with 30 and 27 communities respectively.

The SCBS asks respondents a number of questions related to their personal networks. This provides information both on their network composition, types of people in their networks, and about how much time they spend socializing with those people.

To compute the racial homophily measure, we combine two socialization related questions. The first one assesses the general level of socialization and the second one the level of socialization with friends of a different race. The exact wording of these questions is the following:

**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 assume that socializing is reciprocal and doubled the estimated frequency of having friends over to one's home. Thus, the individual homophily index is:

$$h_i = 1 - \frac{FRDRAC_i}{2 \cdot FRDVISIT_i}. \quad (6)$$

---

<sup>6</sup>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 respondents in Boston, and respondents living within city boundaries in Greensboro city and Delaware.

<sup>7</sup>See <http://www.hks.harvard.edu/saguro/communitysurvey/index.html> for more information on the SCBS and data access.

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

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  $w_i$ . These minority percentages are taken from census data by the designers of the survey. We then take the average over all black respondents within the regions to get the region-specific network measures.

Only information about total household income is available in SCBS. The income is reported in different intervals ( $< \$20,000$ ;  $\$20,000$ - $\$30,000$ ;  $\$30,000$ - $\$50,000$ ;  $\$50,000$ - $\$75,000$ ;  $\$75,000$ - $\$100,000$ ; and  $> \$100,000$ ). As SCBS includes fewer job- and human-capital related measures, we only estimate 3 different models with different sets of explanatory variables, see Appendix A. All these variables were taken from the SCBS.

Table 4 presents summary statistics for each community. We can see that the average income for whites (income W) exceeds that of blacks (income B) in virtually every community. While the number of whites is in the hundreds in every community, the number of blacks is in many cases rather low, even below 30.<sup>8</sup>

We next present the relationship between the inbreeding homophily and the percentage for the blacks in different communities (Figure 3). Most of the communities have quite similar percentage of blacks (between 0 and 0.4) with one clear outlier, North Minneapolis, where the percentage is 0.55. We can distinguish a familiar hump-shaped pattern (see Currarini, Jackson, and Pin, 2009, Figure 4), however, the presence of the ‘‘hump’’ is largely relying on the outlier. The curves for whites and blacks seem to fit well together.

## 4.2 Results

We estimate the association between the regional network segregation and wage gap with an approach similar to the Estonian case, that is, we first perform a wage regression according to (3). As SCBS reports income in intervals we use an interval regression for the first-stage regression (see Wooldridge, 2002, p. 509). From this regression we obtain estimates of the community wage gaps. In the second stage, in order to quantify the relationship, we estimate an OLS model, explaining the estimated community wage gap by inbreeding homophily and minority percentage in the community where the observations are weighted by their estimated variance. As both inbreeding homophily and minority wage gaps may be related to the minority percentage, we regress the regional wage gap on both of these variables. We perform the estimations separately both including and excluding the communities in the Minneapolis metropolitan region.

Results are presented in Table 5. The table shows that the wage gap and inbreeding homophily are indeed negatively related. With Minneapolis region included, the coefficient for  $IH$  is significant, except for model 1, and the magnitude is large. Excluding the outliers, the estimates become statistically insignificant but their values remain essentially the same. Figure 4 plots the inbreeding homophily in the different communities against the residual wage gaps from model 3. The upper panel of the plot explains the role of the outliers; one community, namely North Minneapolis, has substantially lower  $IH$  value.

---

<sup>8</sup>Note that the percentage of blacks (pct B) reported in the SCBS, is based on the 2000 census and not on the racial composition in the survey.

Table 4: Income and percentage of black workers by community in the SCBS data set.

	income W	income B	diff (%)	N W	N B	pct B
national	52072	43636	-0.16	1697	451	0.12
atlanta metro	65157	51278	-0.21	249	150	0.31
baton rouge	60120	40018	-0.33	273	144	0.36
birmingham metro	55188	40643	-0.26	297	115	0.30
charlotte region/14 county	55864	44825	-0.20	973	242	0.20
syracuse/onondaga county	50120	35986	-0.28	399	37	0.09
chicago metro	61148	49791	-0.19	416	111	0.18
cincinnati metro	52908	36959	-0.30	753	98	0.12
east tennessee	42219	35562	-0.16	394	24	0.05
houston/harris county	60638	44432	-0.27	209	77	0.18
kanawha valley (wv)	42846	36474	-0.15	396	23	0.05
kalamazoo co.	53263	49470	-0.07	396	33	0.09
los angeles co.	63789	43344	-0.32	177	48	0.08
st. paul metro	59249	42869	-0.28	397	16	0.02
san diego co.	56907	49422	-0.13	294	22	0.06
san francisco (city)	72508	46900	-0.35	256	40	0.04
detroit metro/7-co.	56263	47709	-0.15	298	96	0.21
winston-salem/forsyth co.	55597	40621	-0.27	484	144	0.23
indiana	48103	38575	-0.20	802	40	0.06
cleveland/cuyahoga co.	53295	38703	-0.27	597	205	0.25
greensboro/guilford co.	56160	42805	-0.24	445	178	0.25
peninsula-silicon valley	74423	64801	-0.13	754	39	0.04
seattle	56392	53916	-0.04	376	30	0.09
grand rapids (city)	47876	33070	-0.31	341	58	0.16
boston (city)	57694	43400	-0.25	256	130	0.23
delaware	54219	42652	-0.21	928	177	0.18
rochester metro (ny)	51087	32550	-0.36	616	134	0.09
minneapolis	53927	36607	-0.32	365	38	0.08
north minneapolis	49504	39298	-0.21	220	143	0.55
denver (city/co.)	54137	43975	-0.19	279	60	0.11
phoenix/maricopa co.	54715	56163	0.03	298	18	0.04

Notes: “income W” and “income B”: mean interval household income of whites and blacks, respectively; “difference”: relative difference of these income measures; “N”: the number of observations in each community; “pct B”: the relative percentage of blacks in the community, based on U.S. census data.

However, a negative correlation of similar size is also present in the data of the other communities (lower panel). The fitted values suggest that the wage gap is 8 percentage points larger in a community with high inbreeding homophily (around .60) compared to a community with low inbreeding homophily (around .45).

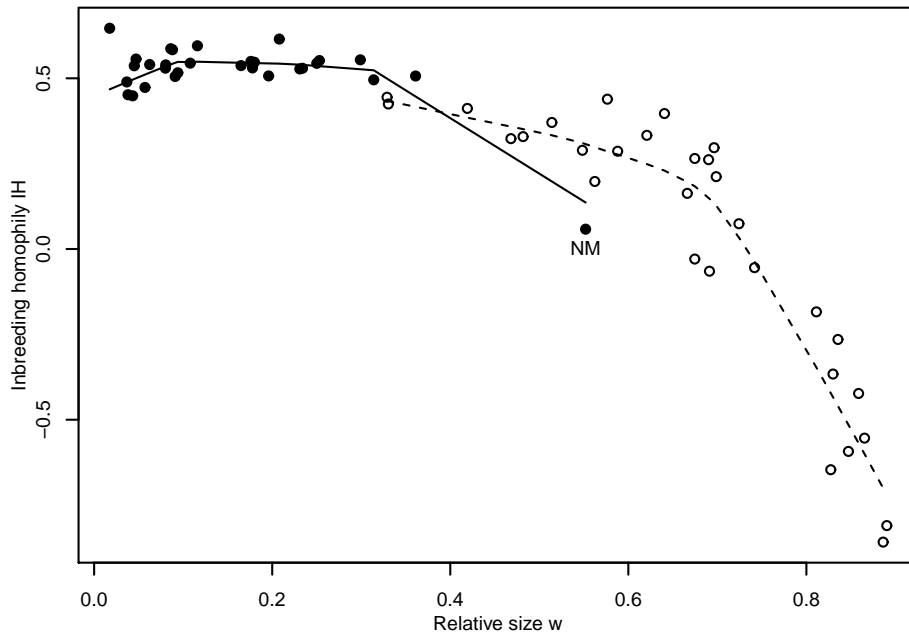


Figure 3: Relationship between the black homophily and percentage by communities in the United States. Black dots represent blacks and white dots whites. The lines represent respective smoothed averages. The observation labeled NM represents North-Minneapolis.

## 5 Potential Explanations

Our results indicate that a relationship exists between inbreeding homophily and wage gaps at the regional level, both in Estonia and in the United States. What drives these results? We discuss three possible explanations: discrimination, residential or occupational segregation, and social network effects. We do not intend to test or empirically dissect these explanations, as our data does not allow for a definite identification strategy. However, the discussion should give us guidelines for future research on this topic.

### 5.1 Prejudice and Discrimination

One oft-discussed explanation for differences in labor market outcomes is discrimination by the majority group. Becker (1957) provided the first economic analysis of discrimination and labor markets in cases where white agents have explicit preferences for market interaction with whites only. Since then, a large literature has studied the relationship between discrimination, both preference-based and statistical, and labor market outcomes, see Charles and Guryan (2011) for a recent overview. In particular, Charles and Guryan (2008) find direct empirical support for Becker’s theory of prejudice and sorting; namely, that black workers are more likely to work for the least prejudiced employers.

If individuals have explicit preferences for market interaction with a particular race, then one should expect these racial preferences to be even stronger in the case of social interactions. The very same discriminatory preferences should

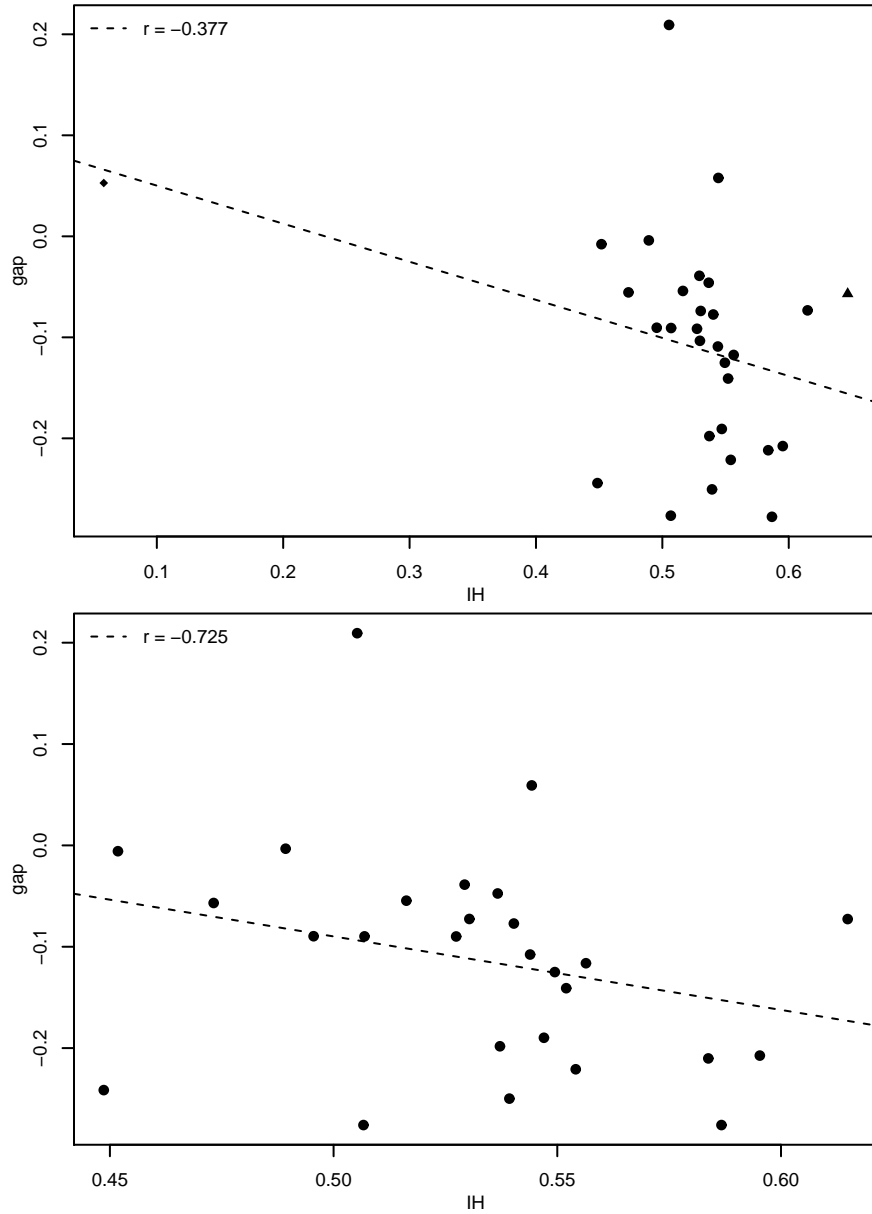


Figure 4: Relationship between the black-white wage gap and inbreeding homophily by communities in the United States. Wage gap based on model 2. The lines represent weighted regression fits and corresponding coefficients. The upper panel includes the outliers (in particular, North Minneapolis), the lower panel is based on the estimations where the outliers are excluded.

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

Dependent variable: inbreeding homophily $IH$			
Variable	Model 1	Model 2	Model 3
All regions			
(Intercept)	0.083	0.243*	0.225*
	<i>0.160</i>	<i>0.133</i>	<i>0.133</i>
$IH$	-0.365	-0.566***	-0.540***
	<i>0.250</i>	<i>0.208</i>	<i>0.209</i>
Minority pct	-0.217	-0.277	-0.292
	<i>0.220</i>	<i>0.183</i>	<i>0.183</i>
nObs	30	30	30
rSquared	0.07371	0.215	0.199
Outliers excluded			
(Intercept)	-0.011	0.236	0.250
	<i>0.327</i>	<i>0.272</i>	<i>0.271</i>
$IH$	-0.176	-0.552	-0.587
	<i>0.605</i>	<i>0.504</i>	<i>0.503</i>
Minority pct	-0.259	-0.282	-0.286
	<i>0.246</i>	<i>0.206</i>	<i>0.206</i>
nObs	27	27	27
rSquared	0.04745	0.116	0.1231
Explanatory variables			
constant, region	✓	✓	✓
age, education, citizenship		✓	✓
family			✓

Notes: standard errors in italics.  
Significance level: \* – 10%, \*\*\* – 1%

therefore imply inbreeding homophily. Applying a revealed preference approach based on a model of racial preferences and meeting opportunities to high school friendship data, Currarini, Jackson, and Pin (2009) indeed find that high school children have discriminatory preferences; they value a friendship with a child of a different race at only 80% of the value of a friendship with a child of the same race. Wimmer and Lewis (2010) estimate an ERG model on Facebook data of college students. They, too, find that ethnicity-based preferences are a main driver of homophily, although not the only one. These two studies provide evidence that inbreeding homophily is at least partially driven by discriminatory preferences.

It is therefore plausible that these discriminatory preferences at least partially drive both inbreeding homophily and wage gaps, hence the relation between them. This hypothesis may be examined by repeating the analysis for the United States, but including a variable that measures the amount of prejudice within a particular region, similar to the approach of Charles and Guryan (2008). If the relation between homophily and wage gaps is driven by discriminatory

preferences, then including a proxy for the mediating factor (i.e., prejudice) should reduce or eliminate the estimated relationship between homophily and wage gaps.

## 5.2 Residential Segregation

Another branch of the literature suggests that ethnic wage gaps are related to residential segregation with social networks playing only a minor role. This explanation suggests that residential segregation is the primary cause or consequence of ethnic wage gaps, and inbreeding homophily is just a residual effect of residential segregation, without necessarily having a feedback effect on the labor market.

Naturally, residential segregation leads to inbreeding homophily. Residential proximity is still a primary force behind friendship formation. Hence, residential segregation limits the opportunity to meet and befriend members from other races or ethnicities, leading to inbreeding homophily (Wimmer and Lewis, 2010).

Theories on why segregation may be related to wage gaps, often refer to local human capital externalities (Benabou, 1993; Lundberg and Startz, 1998; Sethi and Somanathan, 2004). Indeed, there is a substantial heterogeneity in school quality, many of the lower quality schools being located in ghettos.

In the United States, residential segregation according to race is all too present and a primary policy concern. It has been an object of intense study, both in theory, in measurement and in its relation to discrimination and economic inequality (Cutler and Glaeser, 1997; Echenique and Fryer, 2007; Cutler, Glaeser, and Vigdor, 2008). Although earlier studies suggested that more residential segregation leads to more racial inequality (Cutler and Glaeser, 1997; Bayard, Hellerstein, Neumark, and Troske, 1999), later studies suggest that the relationship might be negative instead, and that it depends on circumstances such as income level (Charles and Guryan, 2008; Cutler, Glaeser, and Vigdor, 2008).

The Estonian case, however, strongly suggest that residential segregation is not a likely explanation of the observed relationship - at least not in Estonia. This is in part because neighborhoods have never been as segregated in Estonia as they are in the United States. For example, the capital Tallinn has a Russian minority of around 40%, but none of the neighborhoods has a Russian population of more than 75% (Toomet, Silm, Saluveer, Tammaru, and Ahas, 2012). More striking evidence against the residential segregation thesis, however, comes from the swift emergence of a wage gap shortly after Estonian independence in 1991 (Leping and Toomet, 2008). No wage gap existed prior to 1991, and it is unlikely that segregation increased rapidly enough after 1991 to have played a major role in the initial emergence of a wage gap during the early 1990s. This suggests that residential segregation did not play a major role in the emergence of a wage gap in Estonia, although it may play a smaller role in sustaining the gap once established via other mechanisms.

Notwithstanding the compelling evidence provided by the Estonian natural experiment, one might still wish to use statistical controls to assess the potential contribution of residential segregation to both inbreeding homophily and labor market outcomes. Such an exercise would require that measures of both segregation and network homophily be included in a single regression. As these measures are almost certainly highly correlated, research along such lines would



require the use of exceptionally high-quality data. To our knowledge, such data are not currently available.

### 5.3 Structural Network Theories

A third possibility is that homophily in social networks causes wage differentials between ethnicities and races. Members of minority groups might find it more difficult to find jobs if they lack the social ties to members of the majority group, who often possess more and better quality information about available jobs. Minority groups who prefer to form friendships with those from a similar racial or ethnic background might inadvertently limit their access to job information.

There is a large body of empirical literature about job search channels and job quality at the individual level, see Granovetter (1974) for an overview. Fewer studies analyze the relationship between ethnic or racial wage gaps and network properties. Ioannides and Loury (2004) 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.

An in-depth case study by Royster (2007) of 25 white and 25 black graduates from the same vocational school identifies several mechanisms that put black graduates at a disadvantaged situation when entering the labor market. In particular, black students have fewer and less powerful contacts than those of those of their white peers. White students were more likely to have established and maintained social contacts in “all male and all-white spaces”, such as bars and taverns.

Although homophily may sustain temporary wage gaps, it is not obvious that homophily should lead to *persistent* wage gaps in models where job contacts matter, unless one assumes that the disadvantaged group has fewer job contacts for exogenous reasons. Calvó-Armengol and Jackson (2004) argue that small gaps in labor outcomes are magnified by drop-out decisions. If only employed or high-wage workers provide job references, then ethnic minority members with many unemployed job contacts are more likely to drop out of the labor market, which negatively affects the drop-out decisions of their friends, leading to a substantial difference in drop-out decisions and unemployment rates between the ethnic minority and the majority.

Van der Leij and Buhai (2008) have an explicit model in which inbreeding homophily creates persistent unemployment and wage gaps through occupational segregation. A key factor in this model is the education decision that individuals make before entering the labor market. Individuals prefer to choose the same education as the majority of their ethnic group, as inbreeding homophily leads them to believe that they obtain more job contacts, and hence better job opportunities, by doing so. These benefits in terms of job opportunities outweigh the potential disadvantage of a lower wage.

## 6 Conclusions

In this paper we analyze the relationship between network segregation, that is, inbreeding homophily, and unexplained wage gaps. We consider two very dif-

ferent societies and labor markets; ethnic Estonians and Russians in the former Soviet republic of Estonia, and black and white workers in the United States. For this purpose, we employ three data sources: the Social Capital Benchmark Survey 2000 (SCBS) for the U.S. network and income measures, telecommunication data for the Estonian network, and the Estonian Labor Force Survey (ELFS) for the income measures.

For both the U.S. and Estonia, we establish a negative relation between wage gaps and network segregation. Less contact between the racial or ethnic groups is related to larger unexplained gaps disfavoring the minority group. The negative correlation persists even when controlling for the relative size of the minority population in the region.

We then discuss three possible reasons for this relationship: 1) discrimination and prejudice in both the friendship selection and employer hiring processes; 2) differential investment in human capital due to residential segregation producing both network segregation and the wage gap; and 3) minority difficulty in finding jobs due to network segregation, particularly in light of labor force drop-out and peer-influenced education decisions. None of the three mechanisms discussed is well understood. Moreover, it is likely that all three mechanisms are shaping observed outcomes simultaneously, making it more difficult to disentangle their relative contributions to observed wage gaps. Distinguishing between the proposed mechanism would require better longitudinal data on both labor markets and network segregation, in particular the segregation in job contact networks, and these data are not currently available.

However, for policy purposes, understanding and empirically disentangling these mechanisms is important. With the advent of job contact social networking sites, such as LinkedIn, we hope that a better understanding of the relationship between social, segregation and inequality will be within our reach.

## A First stage regressions

### A.1 Explanatory variables

Table 6 explains the explanatory variables in the first stage regressions for all the Estonian and U.S. models. The variables are analogous, but not exactly the same due to institutional differences and data availability.

Table 6: Explanatory variables.

<i>variable</i>	Estonia	U.S.
		1st model
<i>region</i>	county	SCBS “community”
<i>black</i>		National sample is used as the baseline race, non-hispanic black (reference: non-hispanic white)
<i>non-est</i>	ethnicity: Non-Estonian (reference: Estonian)	
<i>female</i>	gender	region and minority cross-effects
<i>date</i>	cubic polynomial of observation date	
		2nd model
<i>age</i>	age groups (< 25, 25 – 35, 35 – 50, 50–)	5th-order polynomial in age
<i>education</i>	education groups (less than high school, high school, college)	years of education
		3rd model
<i>married</i>	living with partner	
<i>kids</i>	number of children in the household (0-3, 4-6, 7-17 year olds separately) <i>married</i> and <i>female</i> cross-effects <i>gender</i> and <i>kids</i> cross-effects	0-5 and 6-17 year olds separately
<i>immigrant</i>	born outside Estonia	
<i>Non citizen</i>		not a U.S. citizen
<i>Spanish</i>		Interview conducted in Spanish
		4th model
<i>Russian, Estonian, English</i>	relevant language skills (coded as able/not able to speak)	
		5th model
<i>industry, occupation</i>	industry, occupation of work	

## B Full results

### B.1 Estonia (counties)

Table 7: First stage regression. Full results for Estonia at county level, random effect models.

Dependent variable: log household wage					
Explanatory variable	Model 1	Model 2	Model 3	Model 4	Model 5
constant	8.835*** <i>0.008</i>	8.806*** <i>0.008</i>	8.738*** <i>0.010</i>	8.628*** <i>0.015</i>	8.632*** <i>0.032</i>
county 39	-0.426*** <i>0.018</i>	-0.339*** <i>0.017</i>	-0.338*** <i>0.017</i>	-0.324*** <i>0.016</i>	-0.305*** <i>0.015</i>
county 44	-0.342*** <i>0.022</i>	-0.286*** <i>0.021</i>	-0.286*** <i>0.021</i>	-0.253*** <i>0.021</i>	-0.242*** <i>0.019</i>
county 49	-0.369*** <i>0.018</i>	-0.313*** <i>0.017</i>	-0.318*** <i>0.017</i>	-0.288*** <i>0.017</i>	-0.247*** <i>0.016</i>
county 51	-0.325*** <i>0.016</i>	-0.252*** <i>0.015</i>	-0.257*** <i>0.015</i>	-0.234*** <i>0.015</i>	-0.202*** <i>0.014</i>
county 57	-0.323*** <i>0.020</i>	-0.265*** <i>0.019</i>	-0.267*** <i>0.019</i>	-0.244*** <i>0.019</i>	-0.227*** <i>0.017</i>
county 59	-0.347*** <i>0.014</i>	-0.273*** <i>0.013</i>	-0.274*** <i>0.013</i>	-0.253*** <i>0.013</i>	-0.218*** <i>0.012</i>
county 65	-0.431*** <i>0.018</i>	-0.370*** <i>0.017</i>	-0.374*** <i>0.017</i>	-0.342*** <i>0.017</i>	-0.312*** <i>0.016</i>
county 67	-0.314*** <i>0.015</i>	-0.256*** <i>0.014</i>	-0.258*** <i>0.014</i>	-0.238*** <i>0.014</i>	-0.213*** <i>0.013</i>
county 70	-0.283*** <i>0.017</i>	-0.223*** <i>0.016</i>	-0.224*** <i>0.015</i>	-0.203*** <i>0.015</i>	-0.181*** <i>0.014</i>
county 74	-0.334*** <i>0.017</i>	-0.270*** <i>0.016</i>	-0.274*** <i>0.016</i>	-0.253*** <i>0.016</i>	-0.244*** <i>0.015</i>
county 78	-0.227*** <i>0.013</i>	-0.213*** <i>0.012</i>	-0.213*** <i>0.012</i>	-0.207*** <i>0.012</i>	-0.173*** <i>0.011</i>
county 82	-0.417*** <i>0.018</i>	-0.358*** <i>0.017</i>	-0.364*** <i>0.016</i>	-0.337*** <i>0.016</i>	-0.318*** <i>0.015</i>
county 84	-0.384*** <i>0.014</i>	-0.324*** <i>0.014</i>	-0.326*** <i>0.014</i>	-0.301*** <i>0.014</i>	-0.272*** <i>0.013</i>
county 86	-0.385*** <i>0.018</i>	-0.326*** <i>0.017</i>	-0.328*** <i>0.017</i>	-0.301*** <i>0.017</i>	-0.274*** <i>0.015</i>
non-est	-0.290*** <i>0.011</i>	-0.264*** <i>0.010</i>	-0.264*** <i>0.010</i>	-0.228*** <i>0.011</i>	-0.195*** <i>0.010</i>
date <sup>1</sup>	97.141*** <i>0.664</i>	95.210*** <i>0.629</i>	95.342*** <i>0.627</i>	93.946*** <i>0.631</i>	94.078*** <i>0.591</i>
date <sup>2</sup>	11.933*** <i>0.527</i>	11.770*** <i>0.509</i>	11.803*** <i>0.508</i>	11.574*** <i>0.506</i>	11.849*** <i>0.486</i>
date <sup>3</sup>	-13.479*** <i>0.450</i>	-13.223*** <i>0.439</i>	-13.145*** <i>0.438</i>	-13.045*** <i>0.437</i>	-12.766*** <i>0.425</i>
woman	-0.285***	-0.332***	-0.292***	-0.291***	-0.232***

Table 7 – continues. . .

Table 7 – continued

Explanatory variable	Model 1	Model 2	Model 3	Model 4	Model 5
	<i>0.006</i>	<i>0.005</i>	<i>0.007</i>	<i>0.007</i>	<i>0.007</i>
county 39×non-est	0.312***	0.288***	0.287***	0.247**	0.189*
	<i>0.109</i>	<i>0.106</i>	<i>0.106</i>	<i>0.105</i>	<i>0.102</i>
county 44×non-est	0.164***	0.125***	0.124***	0.120***	0.082***
	<i>0.025</i>	<i>0.024</i>	<i>0.024</i>	<i>0.024</i>	<i>0.022</i>
county 49×non-est	0.139**	0.147**	0.154**	0.126**	0.094*
	<i>0.065</i>	<i>0.061</i>	<i>0.060</i>	<i>0.060</i>	<i>0.055</i>
county 51×non-est	0.232***	0.212***	0.216***	0.189***	0.172***
	<i>0.057</i>	<i>0.054</i>	<i>0.054</i>	<i>0.054</i>	<i>0.050</i>
county 57×non-est	0.148***	0.166***	0.162***	0.137**	0.118**
	<i>0.057</i>	<i>0.054</i>	<i>0.053</i>	<i>0.053</i>	<i>0.050</i>
county 59×non-est	0.326***	0.302***	0.295***	0.282***	0.233***
	<i>0.031</i>	<i>0.029</i>	<i>0.029</i>	<i>0.029</i>	<i>0.027</i>
county 65×non-est	0.314***	0.226***	0.212***	0.190**	0.197***
	<i>0.080</i>	<i>0.075</i>	<i>0.075</i>	<i>0.075</i>	<i>0.069</i>
county 67×non-est	0.217***	0.210***	0.211***	0.194***	0.151***
	<i>0.038</i>	<i>0.035</i>	<i>0.035</i>	<i>0.035</i>	<i>0.033</i>
county 70×non-est	0.113*	0.121**	0.126**	0.107*	0.104**
	<i>0.059</i>	<i>0.056</i>	<i>0.055</i>	<i>0.055</i>	<i>0.051</i>
county 74×non-est	0.194**	0.117	0.137	0.107	0.093
	<i>0.087</i>	<i>0.084</i>	<i>0.083</i>	<i>0.083</i>	<i>0.080</i>
county 78×non-est	0.122***	0.127***	0.127***	0.114***	0.084***
	<i>0.029</i>	<i>0.027</i>	<i>0.027</i>	<i>0.027</i>	<i>0.025</i>
county 82×non-est	0.159***	0.155***	0.154***	0.142***	0.108***
	<i>0.044</i>	<i>0.041</i>	<i>0.041</i>	<i>0.041</i>	<i>0.037</i>
county 84×non-est	0.191***	0.204***	0.204***	0.181***	0.163***
	<i>0.049</i>	<i>0.046</i>	<i>0.046</i>	<i>0.046</i>	<i>0.043</i>
county 86×non-est	0.269***	0.207***	0.209***	0.171***	0.120*
	<i>0.070</i>	<i>0.066</i>	<i>0.066</i>	<i>0.066</i>	<i>0.062</i>
age -24		-0.085***	-0.057***	-0.090***	-0.078***
		<i>0.008</i>	<i>0.009</i>	<i>0.009</i>	<i>0.009</i>
age 25-34		0.007	0.016**	-0.001	-0.001
		<i>0.006</i>	<i>0.006</i>	<i>0.006</i>	<i>0.006</i>
age 50-		-0.083***	-0.079***	-0.071***	-0.064***
		<i>0.006</i>	<i>0.006</i>	<i>0.006</i>	<i>0.006</i>
education < HS		-0.165***	-0.162***	-0.142***	-0.096***
		<i>0.007</i>	<i>0.007</i>	<i>0.007</i>	<i>0.007</i>
education college		0.344***	0.344***	0.307***	0.224***
		<i>0.007</i>	<i>0.007</i>	<i>0.007</i>	<i>0.007</i>
immigrant			0.007	0.004	0.002
			<i>0.013</i>	<i>0.013</i>	<i>0.012</i>
married			0.051***	0.052***	0.036***
			<i>0.005</i>	<i>0.005</i>	<i>0.005</i>
kids 0-3			0.037***	0.038***	0.041***
			<i>0.009</i>	<i>0.009</i>	<i>0.009</i>
kids 4-6			0.051***	0.051***	0.043***
			<i>0.009</i>	<i>0.009</i>	<i>0.009</i>

Table 7 – continues...

Table 7 – continued

Explanatory variable	Model 1	Model 2	Model 3	Model 4	Model 5
kids 7–17			0.037*** <i>0.007</i>	0.038*** <i>0.007</i>	0.034*** <i>0.007</i>
woman×kids 0–3			-0.087*** <i>0.014</i>	-0.087*** <i>0.014</i>	-0.096*** <i>0.013</i>
woman×kids 4–6			-0.065*** <i>0.013</i>	-0.065*** <i>0.013</i>	-0.060*** <i>0.012</i>
woman×kids 7–17			-0.046*** <i>0.009</i>	-0.046*** <i>0.009</i>	-0.042*** <i>0.009</i>
Russian				0.035*** <i>0.007</i>	0.022*** <i>0.007</i>
Estonian				0.032*** <i>0.010</i>	0.024** <i>0.009</i>
English				0.119*** <i>0.007</i>	0.090*** <i>0.006</i>
industry B					0.173*** <i>0.036</i>
industry C					0.363*** <i>0.024</i>
industry D					0.171*** <i>0.011</i>
industry E					0.257*** <i>0.018</i>
industry F					0.234*** <i>0.013</i>
industry G					0.109*** <i>0.012</i>
industry H					0.055*** <i>0.017</i>
industry I					0.217*** <i>0.013</i>
industry J					0.321*** <i>0.025</i>
industry K					0.043*** <i>0.014</i>
industry L					0.145*** <i>0.014</i>
industry M					-0.001 <i>0.013</i>
industry N					0.063*** <i>0.014</i>
industry O					-0.010 <i>0.015</i>
industry P					-0.496*** <i>0.084</i>
industry Q					0.074 <i>0.168</i>
occupation 1					0.080***

Table 7 – continues...

Table 7 – continued

Explanatory variable	Model 1	Model 2	Model 3	Model 4	Model 5
					<i>0.028</i>
occupation 2					0.017
					<i>0.028</i>
occupation 3					-0.074***
					<i>0.028</i>
occupation 4					-0.160***
					<i>0.028</i>
occupation 5					-0.256***
					<i>0.028</i>
occupation 6					-0.166***
					<i>0.032</i>
occupation 7					-0.166***
					<i>0.028</i>
occupation 8					-0.172***
					<i>0.028</i>
occupation 9					-0.346***
					<i>0.028</i>
# obs	65751	65751	65751	65751	65738

Notes: Standard errors in italics

Significance levels: \* – 10%; \*\* – 5%, \*\*\* – 1%

## B.2 United States

Table 8: Full results. U.S. data. Models without the outlier regions.

Dependent variable: log household wage			
Explanatory variable	Model 1	Model 2	Model 3
atlanta metro	0.333*** <i>0.056</i>	0.244*** <i>0.052</i>	0.254*** <i>0.053</i>
baton rouge	0.133*** <i>0.047</i>	0.077* <i>0.044</i>	0.082* <i>0.044</i>
birmingham metro	0.099** <i>0.049</i>	0.064 <i>0.048</i>	0.065 <i>0.048</i>
charlotte region/14 county	0.099*** <i>0.032</i>	0.102*** <i>0.029</i>	0.111*** <i>0.029</i>
syracuse/onondaga county	-0.036 <i>0.043</i>	-0.078** <i>0.040</i>	-0.076* <i>0.040</i>
chicago metro	0.265*** <i>0.047</i>	0.155*** <i>0.042</i>	0.163*** <i>0.042</i>
cincinnati metro	0.049 <i>0.035</i>	0.063* <i>0.033</i>	0.060* <i>0.033</i>
east tennessee	-0.199*** <i>0.044</i>	-0.114*** <i>0.042</i>	-0.105** <i>0.042</i>
houston/harris county	0.210*** <i>0.052</i>	0.148*** <i>0.049</i>	0.157*** <i>0.050</i>
kanawha valley (wv)	-0.167*** <i>0.048</i>	-0.140*** <i>0.045</i>	-0.132*** <i>0.045</i>
kalamazoo co.	0.034 <i>0.043</i>	0.001 <i>0.040</i>	0.003 <i>0.041</i>
los angeles co.	0.232*** <i>0.058</i>	0.124** <i>0.055</i>	0.133** <i>0.055</i>
san diego co.	0.096** <i>0.049</i>	0.019 <i>0.044</i>	0.029 <i>0.043</i>
san francisco (city)	0.452*** <i>0.049</i>	0.251*** <i>0.043</i>	0.280*** <i>0.043</i>
detroit metro/7-co.	0.169*** <i>0.054</i>	0.138*** <i>0.048</i>	0.147*** <i>0.048</i>
winston-salem/forsyth co.	0.083* <i>0.043</i>	0.032 <i>0.039</i>	0.042 <i>0.039</i>
indiana	-0.038 <i>0.035</i>	0.006 <i>0.033</i>	0.006 <i>0.033</i>
cleveland/cuyahoga co.	0.053 <i>0.039</i>	0.015 <i>0.034</i>	0.024 <i>0.035</i>
greensboro/guilford co.	0.112*** <i>0.042</i>	0.065* <i>0.038</i>	0.070* <i>0.038</i>
peninsula-silicon valley	0.552*** <i>0.034</i>	0.417*** <i>0.031</i>	0.424*** <i>0.031</i>
seattle	0.092** <i>0.046</i>	-0.019 <i>0.042</i>	-0.003 <i>0.042</i>
grand rapids (city)	-0.077	-0.116***	-0.115***

Table 8 – continues. . .



Table 8 – continued

Explanatory variable	Model 1	Model 2	Model 3
	<i>0.049</i>	<i>0.044</i>	<i>0.044</i>
boston (city)	0.098*	-0.014	0.010
	<i>0.050</i>	<i>0.047</i>	<i>0.047</i>
delaware	0.095***	0.071**	0.072**
	<i>0.032</i>	<i>0.030</i>	<i>0.030</i>
rochester metro (ny)	-0.008	-0.027	-0.026
	<i>0.038</i>	<i>0.035</i>	<i>0.035</i>
denver (city/co.)	0.067	-0.031	-0.006
	<i>0.051</i>	<i>0.046</i>	<i>0.046</i>
phoenix/maricopa co.	0.113**	0.047	0.042
	<i>0.055</i>	<i>0.048</i>	<i>0.048</i>
Black	-0.221***	-0.148***	-0.155***
	<i>0.041</i>	<i>0.038</i>	<i>0.038</i>
female	-0.126***	-0.124***	-0.132***
	<i>0.012</i>	<i>0.012</i>	<i>0.012</i>
atlanta metro×black	-0.096	-0.083	-0.096
	<i>0.091</i>	<i>0.087</i>	<i>0.086</i>
baton rouge×black	-0.330***	-0.275***	-0.283***
	<i>0.081</i>	<i>0.077</i>	<i>0.076</i>
birmingham metro×black	-0.232***	-0.210**	-0.211**
	<i>0.089</i>	<i>0.085</i>	<i>0.085</i>
charlotte region/14 county×black	-0.126*	-0.081	-0.087
	<i>0.067</i>	<i>0.064</i>	<i>0.063</i>
syracuse/onondaga county×black	-0.346***	-0.263**	-0.263**
	<i>0.128</i>	<i>0.123</i>	<i>0.123</i>
chicago metro×black	-0.125	-0.121	-0.120
	<i>0.094</i>	<i>0.084</i>	<i>0.083</i>
cincinnati metro×black	-0.175*	-0.202**	-0.216**
	<i>0.091</i>	<i>0.091</i>	<i>0.090</i>
east tennessee×black	0.029	-0.049	-0.058
	<i>0.169</i>	<i>0.143</i>	<i>0.146</i>
houston/harris county×black	-0.258**	-0.195*	-0.205**
	<i>0.106</i>	<i>0.105</i>	<i>0.104</i>
kanawha valley (wv)×black	-0.044	-0.108	-0.111
	<i>0.161</i>	<i>0.153</i>	<i>0.157</i>
kalamazoo co.×black	-0.042	-0.054	-0.059
	<i>0.114</i>	<i>0.101</i>	<i>0.100</i>
los angeles co.×black	-0.266**	-0.223*	-0.217
	<i>0.133</i>	<i>0.135</i>	<i>0.134</i>
san diego co.×black	-0.027	-0.051	-0.060
	<i>0.164</i>	<i>0.151</i>	<i>0.151</i>
san francisco (city)×black	-0.422***	-0.232**	-0.260**
	<i>0.119</i>	<i>0.114</i>	<i>0.113</i>
detroit metro/7-co.×black	-0.057	-0.049	-0.072
	<i>0.097</i>	<i>0.086</i>	<i>0.086</i>
winston-salem/forsyth co.×black	-0.145*	-0.089	-0.099
	<i>0.085</i>	<i>0.083</i>	<i>0.082</i>

Table 8 – continues. . .

Table 8 – continued

Explanatory variable	Model 1	Model 2	Model 3
indiana×black	0.023 <i>0.147</i>	-0.071 <i>0.128</i>	-0.051 <i>0.129</i>
cleveland/cuyahoga co.×black	-0.140* <i>0.076</i>	-0.123* <i>0.072</i>	-0.123* <i>0.072</i>
greensboro/guilford co.×black	-0.152* <i>0.079</i>	-0.110 <i>0.073</i>	-0.111 <i>0.072</i>
peninsula-silicon valley×black	-0.089 <i>0.124</i>	-0.012 <i>0.112</i>	0.001 <i>0.111</i>
seattle×black	0.127 <i>0.115</i>	0.214** <i>0.109</i>	0.214** <i>0.109</i>
grand rapids (city)×black	-0.230* <i>0.130</i>	-0.193 <i>0.123</i>	-0.197 <i>0.121</i>
boston (city)×black	-0.120 <i>0.091</i>	-0.032 <i>0.086</i>	-0.046 <i>0.085</i>
delaware×black	-0.114 <i>0.075</i>	-0.059 <i>0.070</i>	-0.058 <i>0.069</i>
rochester metro (ny)×black	-0.292*** <i>0.086</i>	-0.196** <i>0.078</i>	-0.204*** <i>0.077</i>
denver (city/co.)×black	0.002 <i>0.129</i>	0.067 <i>0.113</i>	0.038 <i>0.113</i>
phoenix/maricopa co.×black	0.011 <i>0.190</i>	-0.009 <i>0.158</i>	0.003 <i>0.157</i>
Education		0.093*** <i>0.002</i>	0.095*** <i>0.002</i>
age <sup>1</sup>		8.640*** <i>0.634</i>	8.949*** <i>0.641</i>
age <sup>2</sup>		-6.613*** <i>0.629</i>	-4.621*** <i>0.658</i>
age <sup>3</sup>		-2.113*** <i>0.639</i>	-2.150*** <i>0.638</i>
age <sup>4</sup>		1.296** <i>0.646</i>	0.382 <i>0.649</i>
age <sup>5</sup>		-2.902*** <i>0.641</i>	-2.713*** <i>0.640</i>
# kids 0–5			0.016** <i>0.007</i>
# kids 6–17			0.060*** <i>0.007</i>
Not citizen			-0.058 <i>0.042</i>
spanish			-0.039 <i>0.053</i>
Constant	10.867*** <i>0.021</i>	9.582*** <i>0.036</i>	9.522*** <i>0.037</i>
$\sigma$	0.660*** <i>0.005</i>	0.605*** <i>0.005</i>	0.602*** <i>0.005</i>
# obs	12 964	12 964	12 963

Notes: Standard errors in italics  
Significance level: \* – 10%; \*\* – 5%; \*\*\* – 1%  
*age* is specified as 5th degree orthogonal polynomials.

## C Estonian Municipalities, Descriptive Statistics

This table reports the average income (in Estonian kroons) by ethnic group, it's relative difference, sample size, and percentage of Russian-speaking workers for each county.

Table 9: Income and percentage of Russian-speaking workers by counties.

Municipality code	wage E	wage R	difference	N	pct R
140	5200	4377	-0.16	108	0.22
154	3940	3728	-0.05	102	0.61
183	4697	3483	-0.26	712	0.19
198	5051	4183	-0.17	209	0.08
229	3283	3267	-0.00	47	0.47
245	5779	5415	-0.06	258	0.12
248	4211	3676	-0.13	937	0.05
251	4016	2800	-0.30	511	0.65
260	3314	2730	-0.18	68	0.16
292	3722	2392	-0.36	379	0.05
295	6981	4451	-0.36	465	0.17
309	3364	3546	0.05	332	0.61
317	4825	4340	-0.10	391	0.11
320	4467	2866	-0.36	63	0.63
322	3775	3410	-0.10	2069	0.86
345	4554	3890	-0.15	221	0.41
349	4316	3157	-0.27	1041	0.02
424	5553	4624	-0.17	159	0.62
446	5122	4027	-0.21	649	0.86
498	2814	3511	0.25	58	0.34
511	4219	3138	-0.26	3365	0.97
518	5351	2632	-0.51	105	0.13
566	4333	3161	-0.27	733	0.10
568	3912	3081	-0.21	150	0.09
580	4281	3861	-0.10	252	0.62
605	4032	3843	-0.05	123	0.13
625	4011	3410	-0.15	1846	0.19
645	3694	3609	-0.02	74	0.61
657	3381	3490	0.03	46	0.61
662	3789	2615	-0.31	181	0.08
663	4789	3805	-0.21	1058	0.15
669	4436	3799	-0.14	625	0.03
674	4097	2781	-0.32	248	0.15
707	3723	2495	-0.33	314	0.07
718	5159	3111	-0.40	359	0.12
735	3808	3302	-0.13	659	0.96
741	4492	4434	-0.01	151	0.42
751	3739	3498	-0.06	38	0.45
758	4009	4827	0.20	354	0.04
770	3931	4482	0.14	306	0.20

Table 9 – continues...

Table 9 – continued

Municipality code	wage E	wage R	difference	N	pct R
784	6013	4236	-0.30	9956	0.50
786	3706	2858	-0.23	342	0.21
790	4206	4418	0.05	584	0.47
795	4729	3623	-0.23	3844	0.20
831	4027	4817	0.20	134	0.19
835	3663	3482	-0.05	704	0.07
854	3664	3654	-0.00	1175	0.29
861	4164	3586	-0.14	99	0.16
868	4928	2739	-0.44	131	0.47
890	6800	5889	-0.13	297	0.06
892	3646	3018	-0.17	266	0.08
897	4265	3197	-0.25	1285	0.09
900	3620	3378	-0.07	464	0.11
918	3438	3662	0.07	325	0.06
919	3740	3250	-0.13	1041	0.09

Notes: Only individuals with positive wage reported.

“wage E” and “wage R”: average wage of Estonian- and Russian speakers respectively.

“difference”: relative difference of these wage measures

“N”: the number of observations in each county.

“pct R”: the relative percentage of Russian-speakers, based on the telephone data.

## 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.
- ARROW, K. J. (1972): "Some mathematical models of race in the labor market," in *Racial Discrimination in Economic Life*, ed. by A. Pascal. Lexington Books.
- (1998): "What has economics to say about racial discrimination," *Journal of Economic Perspectives*, 12(2), 91–100.
- 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.
- BENABOU, R. (1993): "Workings of a city: Location, education, and production," *The Quarterly Journal of Economics*, 108(3), 619–652.
- 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.
- 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.
- BLAU, P. (1977): *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. Free Press New York.
- BLINDER, A. S. (1973): "Wage discrimination: Reduced form and structural estimates," *Journal of Human Resources*, 8(4), 436–455.
- BURT, R. S. (1992): *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.
- CALVÓ-ARMENGOL, A., AND M. JACKSON (2004): "The effects of social networks on employment and inequality," *American Economic Review*, 94(3), 426–454.
- CALVÓ-ARMENGOL, A., E. PATAcCHINI, AND Y. ZENOU (2009): "Peer effects and social networks in education," *Review of Economic Studies*, 76(4), 1239–1267.
- 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 (2008): "Prejudice and Wages: An Empirical Assessment of Becker's 'The Economics of Discrimination'," *Journal of Political Economy*, 116(5), 773–809.

- (2011): “Studying discrimination: Fundamental challenges and recent progress,” *Annual Review of Economics*, 3, 479–511.
- COLEMAN, J. S. (1958): “Relational analysis: The study of social organizations with survey methods,” *Human Organization*, 17(4), 28–36.
- COLEMAN, J. S., E. KATZ, AND H. MENZEL (1966): *Medical Innovation: A Diffusion Study*. Bobbs-Merrill, Indianapolis.
- CURRARINI, S., M. O. JACKSON, AND P. PIN (2009): “An Economic Model of Friendship: Homophily, Minorities and Segregation,” *Econometrica*, 77(4), 1003–1045.
- CUTLER, D. M., AND E. L. GLAESER (1997): “Are ghettos good or bad,” *Quarterly Journal of Economics*, 112(3), 827–872.
- CUTLER, D. M., E. L. GLAESER, AND J. L. VIGDOR (2008): “When are ghettos bad? Lessons from immigrant segregation in the United States,” *Journal of Urban Economics*, 63, 759–774.
- EAGLE, N., M. MACY, AND R. CLAXTON (2010): “Network diversity and economic development,” *Science*, 328, 1029–1031.
- ECHENIQUE, F., AND R. G. FRYER (2007): “A measure of segregation based on social interactions,” *Quarterly Journal of Economics*, 122(2), 441–485.
- 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. (1974): *Getting a job: A study of contacts and careers*. Harvard University Press.
- HAYTHORNTHWAITE, C. (2005): “Social networks and internet connectivity effects,” *Information, Communication and Society*, 8(2), 125–147.
- 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*.
- HOLZER, H. (1988): “Search Method Use by Unemployed Youth,” *Journal of Labor Economics*, 6(1), 1–20.
- IOANNIDES, Y. M., AND L. D. LOURY (2004): “Job Information Networks, Neighborhood Effects, and Inequality,” *Journal of Economic Literature*, 42, 1056–1093.
- KORTS, K., AND R. KÕUTS (2002): “Inter-ethnic communication in Estonian media (in Estonian),” in Valk (2002), pp. 234–245.
- LAZARFELD, P., AND R. MERTON (1954): “Friendship as a social process: A substantive and methodological analysis,” *Freedom and Control in Modern Society*, 18, 18–66.

- 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.
- LUNDBERG, S., AND R. STARTZ (1998): “On the Persistence of Racial Inequality,” *Journal of Labor Economics*, 16(2), 292–323.
- MANSKI, C. (2000): “Economic analysis of social interactions,” *Journal of Economic Perspectives*, 14(3), 115–136.
- MASSEY, D. S., AND N. A. DENTON (1988): “The Dimensions of Residential Segregation,” *Social Forces*, 67(2), pp. 281–315.
- 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.
- MCPHERSON, M., L. SMITH-LOVIN, AND J. M. COOK (2001): “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–444.
- MONTGOMERY, J. (1991): “Social networks and labor-market outcomes: Toward an economic analysis,” *American Economic Review*, 81(5), 1408–1418.
- MOODY, J. (2001): “Race, school integration, and friendship segregation in America,” *American Journal of Sociology*, 107(3), 679–716.
- OAXACA, R. L. (1973): “Male-Female Wage Differentials in Urban Labor Markets,” *International Economic Review*, 14(3), 693–709.
- PETTAI, I. (2002): “Mutual tolerance of Estonians and non-Estonians (in Estonian),” in Valk (2002), pp. 213–233.
- 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.
- SCHELLING, T. (1978): *Micromotives and Macrobehavior*. WW Norton & Company.
- 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.
- TOOMET, O. (2011): “Learn English, not the Local Language! Ethnic Minorities in the Baltic States,” *American Economic Review*, 101(3), 526–531.



- TOOMET, O., S. SILM, E. SALUVEER, T. TAMMARU, AND R. AHAS (2012): “Ethnic segregation in residence, work, and free-time: Evidence from mobile communication,” University of Tartu.
- VALK, A. (ed.) (2002): *Estonia and Estonians in a comparative perspective (in Estonian)*. Tartu University Press, Tiigi 78, Tartu 50410, Estonia.
- VAN DER LEIJ, M. J., AND I. S. BUHAI (2008): “A social network analysis of occupational segregation,” Working Papers 192, Fondazione Eni Enrico Mattei.
- WIMMER, A., AND K. LEWIS (2010): “Beyond and below racial homophily: ERG models of a friendship network documented on Facebook,” *American Journal of Sociology*, 116(2), 583–642.
- WOOLDRIDGE, J. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT press.