

Åslund, Olof; Nordström Skans, Oskar

**Working Paper**

## How to Measure Segregation Conditional on the Distribution of Covariates

Working Paper, No. 2007:27

**Provided in Cooperation with:**

Department of Economics, Uppsala University

*Suggested Citation:* Åslund, Olof; Nordström Skans, Oskar (2007) : How to Measure Segregation Conditional on the Distribution of Covariates, Working Paper, No. 2007:27, Uppsala University, Department of Economics, Uppsala,  
<https://nbn-resolving.de/urn:nbn:se:uu:diva-12431>

This Version is available at:

<https://hdl.handle.net/10419/82726>

**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.*



UPPSALA  
UNIVERSITET

Working Paper 2007:27

*Department of Economics*

# How to Measure Segregation Conditional on the Distribution of Covariates

Olof Åslund and Oskar Nordström Skans

Department of Economics  
Uppsala University  
P.O. Box 513  
SE-751 20 Uppsala  
Sweden  
Fax: +46 18 471 14 78

Working paper 2007:27  
December 2007  
ISSN 1653-6975

## HOW TO MEASURE SEGREGATION CONDITIONAL ON THE DISTRIBUTION OF COVARIATES

OLOF ÅSLUND AND OSKAR NORDSTRÖM SKANS

Papers in the Working Paper Series are published  
on internet in PDF formats.  
Download from <http://www.nek.uu.se>  
or from S-WoPEC <http://swopec.hhs.se/uunewp/>

# How to Measure Segregation Conditional on the Distribution of Covariates<sup>♦</sup>

by

Olof Åslund<sup>♦</sup> and Oskar Nordström Skans<sup>♦</sup>

December 7, 2007

## Abstract

This short paper proposes a non-parametric method of accounting for the distribution of background characteristics when testing for segregation in empirical studies. It is shown and exemplified—using data on workplace segregation between immigrants and natives in Sweden—how the method can be applied to correct any measure of segregation for differences between groups in the distribution of covariates by means of simulation, and how analytical results can be used when studying segregation by means of peer group exposure.

**Keywords:** Segregation, Covariates, Workplaces, Immigrants

**JEL-codes:** C1, J1, J2

---

<sup>♦</sup> We are grateful for comments from Per Johansson, Eva Mörk, Roope Uusitalo, seminar participants at IFAU and Växjö University. The work is partly financed by a grant from the Swedish council for working life and social research (FAS). The order of authors is according to the English alphabet and is not related to contribution.

<sup>♦</sup> IFAU (Institute for labour market policy evaluation), Uppsala University and SNS (Center for Business and Policy Studies). Postal address: IFAU, P.O. Box 513, SE-751 20 Uppsala, Sweden, Phone: +46-18-471 70 89, E-mail: olof.aslund@ifau.uu.se

<sup>♦</sup> IFAU and Uppsala University Phone: +46-18-471 70 79, E-mail: oskar.nordstrom\_skans@ifau.uu.se.

# 1 Introduction

The concept of segregation aims to capture systematic sorting over units (neighborhoods, schools or workplaces) by individuals belonging to different groups. In the vast literature, empirical studies have been accompanied by methodological work investigating properties of various measures of segregation.<sup>1</sup> One aspect that has received little attention, however, is how to account for the fact that part of the segregation along the dimension of primary interest (e.g. gender or race) may be caused by sorting according to other dimensions, such as educational attainment. A straightforward parallel is wage regressions where one wishes to isolate differences across groups that cannot be explained by differences in other observed characteristics.<sup>2</sup> This paper presents a non-parametric method accounting for the distribution of such underlying characteristics. The method can be used to adjust any measure of segregation for differences between groups in the distribution of underlying covariates. It can also serve as the basis of analytical tests for segregation conditional on the distribution of covariates when using exposure as the indicator of segregation. We first present and discuss the methods, and then illustrate them using data on ethnic workplace segregation in Sweden.

## 2 The general case

One peculiarity in the literature on segregation is that most traditionally used measures such as the Duncan index and the Gini coefficient (see Massey and Denton (1988) for definitions) refer to a completely *even* distribution as the baseline—although segregation is arguably defined as the *systematic* sorting of individuals over units. As pointed out by Winship (1977) and more re-

---

<sup>1</sup> Massey and Denton (1988) review twenty segregation indices. Recent work describing the characteristics of existing measures of segregation includes Flückiger and Silber (1999) and Hutchens (2004). Papers discussing multigroup segregation measures include Boisso et al. (1994) and Reardon and Firebaugh (2002). Some further discussion is provided in our working paper version Åslund and Skans (2005a).

<sup>2</sup> A small number of previous studies have attempted to correct for covariates using more restrictive methods, e.g. Spriggs and Williams (1996), Bayer et al. (2004), Kalter (2000) and Reardon et al. (2000). In applied work parallel to this paper, Hellerstein and Neumark (2005) uses methods resting on logic similar to the one presented here. However, their presentation is more limited and, in contrast to this paper, relies entirely on simulations.

cently by Carrington and Troske (1997), evenness will be a very misleading baseline if the units are small, so that there is a lot of *random* variation in the composition of different units. As a consequence researchers should instead compare actual measures of segregation,  $Z$ , to a comparable “expected” baseline,  $E(Z)$ —a baseline which can be derived by randomly allocating individuals over the actual units (e.g. workplaces) and calculating  $Z$  on the randomized data. As Carrington and Troske (1997) show, this may play a substantial role in actual data.

We propose to generalize this idea to account not only for the random variation, but also for variation which is driven by systematic differences in the distribution of underlying covariates. As an illustration, assume that a population of  $N$  individuals is divided into two groups consisting of  $M$  immigrants and  $N-M$  natives, and that we are interested in whether these groups are segregated over workplaces.<sup>3</sup> To the extent that immigrants and natives differ in e.g. skills or residential allocation, this may affect their allocation over workplaces. Even though the origins of such differences are not necessarily independent of workplace segregation, studying segregation *conditional* on place of residence or observed skills will provide additional insights into the process at hand.

To fix ideas, think of a data-generating process where each *individual* has a set of characteristics, captured by a specific realization ( $x$ ) of a vector  $X$ , and each *job* requires a certain realization of  $X$ . Thus,  $x$ -type individuals are only suitable for  $x$ -type jobs. The type is determined by characteristics such as where the workers live and their education. Define *segregation conditional on  $X$*  to occur *if* workers are non-randomly sorted given the predetermined allocation of different types as captured by different values of  $X$  among immigrants and natives.

---

<sup>3</sup> Throughout we discuss ethnic workplace segregation, which was also the example provided by Carrington and Troske (1997)

In order to find segregation conditional on the distribution of  $X$  we estimate, for each realization  $(x)$  of the vector  $X$ , the propensity  $p(x)$  of individuals holding jobs with characteristics  $x$  to be immigrants. As long as  $X$  contains discrete characteristics,  $p(x)$  may be estimated non-parametrically as the fraction of minority individuals:

$$p(x) = \frac{M(X = x)}{N(X = x)}. \quad (1)$$

Note that  $p(x)$  can be interpreted as the propensity of individuals holding jobs with characteristics  $x$  to be immigrants. As a consequence,  $p(x)$  could also be estimated as a “propensity score” by postulating immigration status as some function of  $X$  along the lines of Rosenbaum and Rubin (1983).<sup>4</sup> This alternative allows the researcher to include continuous covariates and may be useful if sample sizes are small. However, since this would require the use of a parametric function in order to balance the  $X$ :s, it would be somewhat more restrictive<sup>5</sup> and would also require diagnostics for whether the  $X$ :s really are balanced. For this reason, and in the interest of brevity, we choose to only illustrate the non-parametric version in this paper.

The propensity equation (1) can be used to achieve a counterfactual distribution by randomly allocating minority status to individuals within each “cell”  $x$ , using a probability of being an immigrant equal to the fraction of immigrants in the cell as given by the equation. To be precise: for each individual we allocate a continuous random variable  $v$ , which is uniformly distributed between 0 and 1. Then we infer a counterfactual minority status  $q$  based on the relationship between  $v$  and  $p(x)$  using (1):

---

<sup>4</sup> In practice, this typically amounts to estimating a logit or probit for minority status and then using the estimates for assigning each person a predicted minority status.

<sup>5</sup> The fully interacted model given by the non-parametric approach is arguably particularly important in studying dispersion over units. Consider a simple example where majority and minority are equally distributed over workplaces in two regions, and also over two skill groups, but where the fraction of high-skilled minority working in one region is larger than the corresponding fraction of the majority. A regression controlling only for region and skill will then explain nothing of the observed workplace segregation, whereas an interacted model will account for the unequal distribution across industry/region “cells”.

$$\begin{aligned} q_i &= 0 \text{ if } v > p(x_i) \\ q_i &= 1 \text{ if } v \leq p(x_i) \end{aligned} \quad (2)$$

Then, we can calculate counterfactual measures of segregation  $E(Z|X)$  based on the distribution of  $q$  using *any* measure of segregation, and contrast it to the actual measure  $Z$ . Since replicated simulations provide measures of uncertainty, we also get tools for inference just as in the unconditional case discussed by Carrington and Troske (1997). The comparison between actual and expected segregation can be quantified in many ways depending on the measure used. Here we choose to use a conditional version of the “index of systematic segregation” originally proposed for the unconditional case by Carrington and Troske (1997):

$$\hat{Z}^X = \frac{Z - E(Z | X)}{1 - E(Z | X)} \quad (3)$$

The interpretation is that  $\hat{Z}^X$  measures actual segregation as a fraction of the largest possible segregation conditional on the distribution of  $X$ . Empirical results using this index are presented in Table 1 in section 4.

The methodology presented above is applicable to any measure of segregation. Moreover, the methodology can be used to condition on any number of characteristics and all types of (discrete) characteristics, either unit based and/or individual based (i.e. some or all characteristics may vary within units).

Although the exposition above is limited to a two-group case (minority vs. not minority), it is straightforward to generalize the logic to multi-group cases by making both the propensity equation (1) and the assignment equation (2) group specific, i.e. let  $p^g(x) = \frac{M^g(X=x)}{N(X=x)}$  be the



fraction migrants from (say) country of origin  $g$  among all individuals with characteristics  $x$ , and use  $p^g(x_i)$  together with  $v$  determine the allocation of group  $g$  status. Instead of choosing a single cut-off for  $v$ , we then use intervals to determine group status. If there are among those with characteristics  $x$  are e.g. 5 percent Finnish immigrants, 5 percent Danish immigrants, and 90 percent natives, then a person with characteristics  $x$  is counted as a Finn if his/her random number is below 0.05, as a Dane if it is between 0.05 and 0.10 and as a native if the number is 0.10 or greater.<sup>6</sup>

### 3 A special case with analytical results

Even though the general method presented in the previous section is powerful and generally applicable, a drawback is that the analysis requires simulations. It is however possible to derive tests for segregation analytically by using a measure of *peer exposure*. This measure is a slightly modified version of the index of isolation, and has been used in previous studies. As a statistic of segregation, peer exposure carries the virtue of intuitive economic meaning. It also highlights the link between segregation and social interactions by explicitly capturing an individual's contacts with others of certain characteristics.<sup>7</sup> However, as concluded by Massey and Denton (1988), different indices capture different aspects of segregation, suggesting that the choice of index may depend on the empirical question. Also, peer exposure does not have all the desired properties of a segregation measure given by Hutchens (2004).<sup>8</sup> We do thus not argue that peer exposure is generally ideal or superior to other measures, but that it has nice analytical and interpretational properties useful in many economic settings.

---

<sup>6</sup> Note that the random number is continuous so equality occurs with probability zero.

<sup>7</sup> Echenique and Fryer (2005) argue along similar lines and present a high degree of correlation between their proposed "spectral segregation index" and indices of exposure or isolation.

<sup>8</sup> As a matter of fact, neither does any other available measure of segregation.

To outline a test of segregation based on exposure, assume as before that a population of  $N$  individuals is divided into two groups consisting of  $M$  immigrants and  $N-M$  natives, and that we are interested in whether these groups are sorted randomly over workplaces. If individuals were sorted randomly, each coworker of each worker would be of immigrant origin with a probability equal to  $M/N$  (if oneself is excluded).<sup>9</sup> This holds in expectation for all coworkers of each worker. Thus, immigrants and natives should (in a world without systematic sorting) on average have a share of immigrant coworkers equal to the fraction of immigrants in the population. This prediction can be tested directly, without simulations.

To make this definition precise, define individual  $i$ 's exposure ( $e$ ) to immigrants as the fraction of immigrants among the coworkers:

$$e_i^u \equiv \frac{1}{n^u - 1} \sum_{\substack{j \in u \\ j \neq i}} D_j^m \quad (4)$$

where  $D$  is an immigrant dummy and  $n$  is the number of workers at workplace  $u$ . What (4) does is simply to sum the number of immigrants at workplace  $u$  among the *coworkers* of worker  $i$  (thereby  $j \in u$  and  $j \neq i$ ), and then divide this by the number of coworkers. The average immigrant exposure among immigrant workers is retrieved by summing the observed individual exposure for all immigrants and dividing it by the total number of migrants:

$$\overline{e^m} \equiv \frac{1}{M} \sum_{D_i^m=1} e_i^u \quad (5)$$

---

<sup>9</sup> Formally this is an approximation since it excludes oneself from both denominator and nominator for immigrants. Thus, an immigrant has expected exposure  $(M-1/N-1)$  whereas it is  $(M/N-1)$  for a native. It is trivial in all but tiny sample sizes.

Now, to calculate the “expected” exposure in the absence of segregation conditional on the distribution of covariates, let  $\varepsilon_i^X$  denote the average immigrant propensity among the colleagues of individual  $i$ , given the  $X$ -characteristics of the colleagues:

$$\varepsilon_i^X = \frac{1}{n^u - 1} \sum_{\substack{j \in u \\ j \neq i}} p(x_j) \quad (6)$$

(6) differs from (4) in that it sums the expected rather than the actual number of immigrants. Analogous to (5), we can compute *average expected exposure* among immigrants by:<sup>10</sup>

$$\overline{\varepsilon^{m,X}} = \frac{1}{M} \sum_{D_i^m=1}^i \varepsilon_i^X. \quad (7)$$

We may write the null hypothesis of *no segregation conditional on the distribution of  $X$*  as:

$$H_o(X) : \overline{e^m} = \overline{\varepsilon^{m,X}} \quad (8)$$

If (8) holds, there is no segregation above what is driven by differences in background characteristics. It is worth noting that we have retained the non-parametric feature of the analysis and have thus not imposed any distributional assumptions on the relationship between the covariates and our segregation measure. No simulations are required since we simply compare the actual mean to the expected value, both of which are easily computed from data.

As already mentioned, there is an ongoing debate on which properties a measure of segregation should have, and which measure to use in different economic settings. Two recent examples in this strain are Hutchens (2004) and Echenique and Fryer (2005). It is important to note that equation (8) is a prediction which should hold in a world without segregation conditional on the

---

<sup>10</sup> Note that with only one possible realization of  $X$  (i.e. with no covariates), (7) collapses to the fraction of immigrants in the population as suggested in the introductory paragraph in this section.

distribution of  $X$ . Rejection of  $H_o$  is therefore in itself proof of segregation, since it means that there is a systematic sorting of individuals over units. The issue is then how to quantify segregation, which—as indicated by the vivid debate—can be done in a countless number of ways.

One way to make the distance between actual and expected exposure meaningful is to use the index of systematic segregation defined in equation (3) above. Another is to compare the two proportions (the actual and expected propensity for a coworker of an immigrant to be an immigrant) as we would any other two proportions using e.g. odds ratios. The *overexposure odds ratio* may thus be defined as the ratio between the odds of actual and expected exposure:

$$R^X \equiv \frac{\bar{e}/(1-\bar{e})}{\bar{\varepsilon}^X/(1-\bar{\varepsilon}^X)} \quad (9)$$

The straightforward interpretation is that the odds of an actual immigrant having a certain colleague being an immigrant is  $R_X$  times what it would have been if the distribution was random, given the distribution of  $X$ . Table 2 in section 4 presents some empirical illustrations using this ratio.

Note that the advantages of these analytical results lie in the flexibility and convenience stemming from the fact that the measures are linear functions of individual based statistics. Since the proportions underlying both equations (8) and (9) are based on analytically calculated values for each individual (from eq:s 4 and 6), the test and the overexposure odds ratio can easily be computed for any subpopulation (e.g. what is the tendency for segregation among low educated males?).

The computational simplicity is even more useful in a multigroup setting, since the number of simulations required increases by the number of subgroups. In a multigroup setting, one can of course study exposure in any dimension (see Åslund and Nordström Skans 2005b for empirical

examples). The closest parallel to the two-group example above is to compute average and expected *own-group exposure* among immigrants, i.e. exposure to Finns among Finnish immigrants and to Danes among Danish immigrants. To see how this can be done, define individual own-group exposure  $e_i^{gu}$  analogous to (4), but replace  $D_j^m$  by a dummy for group  $g$  status,  $D_j^g$ . Average own-group exposure is then given by

$$\overline{e^G} = \frac{1}{M} \sum_{g=1}^G \sum_{D_i^g=1} e_i^{gu} \quad (10)$$

where the observed exposure to group  $g$  is first summed for all members of group  $g$ , then summed over all  $G$  immigrant groups, and then divided by  $M$ —the total number of individuals belonging to any of the  $G$  immigrant groups. Define also the expected exposure to group  $g$  ( $\mathcal{E}_i^{gX}$ ) in the same way as in (5), but use instead the group-specific propensity  $p^g(x_j)$ . The expected average own-group exposure is computed as

$$\overline{\mathcal{E}^{G,X}} = \frac{1}{M} \sum_{g=1}^G \sum_{D_i^g=1} \mathcal{E}_i^{gX} \quad (11)$$

Naturally, by altering the group to which individual exposure is measured, and by modifying over which groups the summations are made, the method can be applied to study “cross-exposure” in various dimensions (e.g. between different ethnic groups with linguistic ties). This is illustrated in Table 2 in the following section where we separately analyze immigrants’ exposure to “ethnic” peers and to other immigrants.

## 4 Illustrations

We here illustrate the methods using the example of ethnic workplace segregation in Sweden.<sup>11</sup>

---

<sup>11</sup> For a more detailed description of ethnic segregation in the Swedish labor market, see Åslund and Skans (2005b).

Our data cover all Swedish establishments with at least 2 employees in year 2000, containing 3,457,951 individuals (whereof 9.83 percent are foreign-born) at 219,235 establishments. Each worker is characterized by location (289 municipalities), human capital (3 age groups (16–29, 30–49 and 50–65), 7 education groups and gender) and industry (38 dummies).<sup>12</sup> Industry dummies can be seen as proxies for unobserved factors making some workers more suitable for jobs in certain industries.

We show two sets of results: (i) Table 1 provides simulation results of the index of systematic segregation based on four segregation measures: the commonly used Duncan and Gini coefficients, the by Hutchens (2004) proposed square root index, and peer exposure as defined by equation (4) above; (ii) Table 2 shows analytical results for the overexposure odds ratio.

Row (1) of Table 1 shows the actual levels of the different measures of workplace segregation in Sweden. As a first counterfactual, row (2) shows what their expected values would be in a world with unconditional random allocation. There is “evidence” of segregation also under random assignment, which highlights the argument for using randomness rather than evenness as the baseline.<sup>13</sup> The tiny dispersion between replications as captured by the standard deviation suggests that statistical uncertainty in the simulated distributions is not a concern in this application. The index of systematic segregation provided by the numbers in brackets varies substantially across the columns, but is very similar for the square root index and the measure of exposure.

Our primary interest here is in how the level of calculated ethnic segregation changes when we condition on different sets of background characteristics. Moving across rows in the respective

---

<sup>12</sup> See Åslund and Skans (2005a) for details on the data.

<sup>13</sup> In their application to US workplace data, Carrington and Troske (1997) report levels of segregation that are relatively similar to ours. Their actual level of the Duncan (Gini) index is 0.504 (0.605), the unconditional expected level is 0.337 (0.488), which gives system segregation of 0.251 (0.344).

columns of Table 1 illustrates this. Conditioning on human capital as measured by the interaction of gender, age and education (row (3)) has very little impact on the expected values (or on systematic segregation). While differences in human capital do not explain why immigrants and natives do not work together, more can be attributed to the geographic distribution (row (4)). In row (5) we combine human capital with municipality and again see very little extra effect from including human capital. Row (6) combines human capital, municipality and industry, which means that we compute the probability of being an immigrant given that you are e.g. a middle aged woman with upper secondary education, working in retail in a particular town.<sup>14</sup> The extent of systematic segregation is much lower in row (6) than in row (2), showing that controlling for covariates matters a great deal. However, actual and expected segregation are statistically and economically different from each other for all the four measures even in the most tightly specified model.

Table 2 presents results using the analytical method described in section 3. One advantage of this approach is that it is easy to get results on exposure in various dimensions for subgroups of a population. We illustrate this by studying whether recently arrived (within 10 years) immigrants (panel A) differ from other immigrants (panel B) in their exposure to immigrants. We separately study exposure to “ethnic” peers as measured by region of origin<sup>15</sup> and to immigrants of other descent. In this case we choose to only contrast the actual distribution to the unconditional expectation and to the expectation conditional on the interaction of human capital, municipality and industry, i.e. the most extensive set of covariates.

Exposure to immigrants in general is displayed in column (i). As we see in the first rows of the two panels, recent immigrants exhibit higher actual immigrant exposure (27 percent) than other immigrants do (21 percent). As was the case in Table (1), statistical uncertainty is negligible.

---

<sup>14</sup> Indeed, basing the analysis on such a large number of covariates gives a conservative measure of segregation that

The unconditional expectation (rows 2) is by definition the same for the two groups. By including the covariates we can explain some, but not much, of the actual difference. As a consequence, the odds ratios of overexposure displayed in rows (3) are higher for recent immigrants (2.6) than for other immigrants (2.0).

Recent immigrants are on average from smaller immigrant groups which suggests that exposure to ethnic peers (column (ii)) would be smaller for this group in a world without segregation. This is reflected by the difference between the unconditional expectations in the two panels. When covariates are included, part of this difference in expected exposure disappears, suggesting that the recently arrived more often work in “typical” immigrant locations and industries. Similarly, in both panels the model explains more of the “ethnic” exposure than of the exposure to other groups, suggesting that the distribution of the underlying covariates explain systematic sorting *between* different immigrant groups.

A general message of Table (2) is that there is more overexposure to countrymen than to other immigrants among both recent and non-recent immigrants. However, statistically significant overexposure clearly exists in both dimensions (the expected exposure is in all cases well outside the confidence interval of the actual exposure). The results also suggest that recent immigrants are more overexposed to other immigrants (especially to their ethnic peers) than are non-recent immigrants, even after controlling for human capital, place of residence, industrial affiliation and all interactions between these variables. This raises interesting questions regarding the role of ethnic contacts in the assimilation process; questions that have to be left for future research.

---

probably understates the true level of ethnic segregation, but makes the remaining amount more convincing.

<sup>15</sup> See Åslund and Skans (2005b) for a list of the 26 regions available in the data.



## 5 Summary

When studying segregation it is often desirable to purge the analysis of confounding factors in order to, e.g., identify racial segregation that cannot be attributed to differences in background characteristics such as educational attainment. This short paper describes a method of conditioning on covariates when testing for segregation. We show that this method can be applied to any measure of segregation. We further show that with an index of peer exposure we can test analytically for segregation conditional on the distribution of covariates. This index is well-defined at the individual level, which makes it possible to study how the tendency for segregation varies between individuals, and how various groups are sorted relative to each other.

The empirical illustrations show that accounting for covariates has a strong impact on the estimated level of segregation. Nevertheless, substantial workplace segregation between immigrants and natives remains after controlling for a very rich set of background characteristics. Using the analytical methods we also show that the tendency for segregation is higher among recent immigrants and that there is more excess segregation—also conditional on covariates—towards ethnic peers than towards other immigrants.

## References

- Åslund O and O N Skans (2005a) "Measuring Conditional Segregation: Methods and Empirical Examples", Working Paper 2005:12, IFAU, Uppsala.
- Åslund O and O N Skans (2005b) "Will I See You at Work? Ethnic Workplace Segregation in Sweden 1985–2002", Working Paper 2005:24, IFAU, Uppsala.
- Bayer P, R McMillan and KS Rueben (2004), "What Drives Residential Segregation? New Evidence Using Census Microdata", *Journal of Urban Economics* 56, 514–535.
- Boisso D, K Hayes, J Hirschberg and J Silber (1994), "Occupational Segregation in the Multi-dimensional Case. Decomposition and Tests of Significance", *Journal of Econometrics* 61, 161–171.
- Carrington WJ and KR Troske (1997), "On Measuring Segregation in Samples with Small Units", *Journal of Business and Economics Statistics* 15 (4), 402–409.
- Echenique F and RG Fryer (2005), "On the Measurement of Segregation", NBER working paper 11258.
- Flückiger Y and J Silber (1999), *The Measurement of Segregation in the Labor Force*, Physica-Verlag, Heidelberg.
- Hellerstein J and D Neumark (2005), "Workplace Segregation in the United States: Race, Ethnicity, and Skill.", NBER working paper 11599.
- Hutchens R (2004), "One Measure of Segregation", *International Economic Review* 45 (2), 555–578.
- Kalter F (2000), "Measuring Segregation and Controlling for Independent Variables" Arbeitspapiere Nr 19, 2000, Mannheimer Zentrum für Europäische Sozialforschung.
- Massey DS and NA Denton (1988), "The Dimensions of Residential Segregation", *Social Forces* 67 (2), 281–315.
- Reardon SF, JT Yun and T McNulty Eitle (2000), "The changing structure of School Segregation: Measurement and Evidence of Multiracial Metropolitan-Area School Segregation, 1989–1995", *Demography* 37(3), 351–364.
- Reardon SF and G Firebaugh (2002), "Measures of Multigroup Segregation", in Stolzenberg RM (ed.), *Sociological Methodology* 32, 2002, 33–67, Blackwell Publishing, Boston, MA.
- Rosenbaum P and D Rubin (1983) "The central role of the propensity score in observational studies for causal effects" *Biometrika* 70, 41–55.
- Spriggs WE and RM Williams (1996) "A Logit Decomposition Analysis of Occupational Segregation: Results for the 1970s and 1980s", *Review of Economics and Statistics* 78(2), 348–355.
- Winship C. (1977) "A Revaluation of Indexes of Residential Segregation" *Social Forces* vol 55, No4, 1058–1066.

**Table 1** Immigrant-native workplace segregation in Sweden in year 2000 – simulation results.

	(i) Duncan	(ii) Gini	(iii) Square root index	(iv) Exposure
(1) Actual mean	0.4686	0.6482	0.2907	0.2248
<i>Expected values, (standard deviations) and [index of systematic segregation]</i>				
<i>X</i> -variables	(i) Duncan	(ii) Gini	(iii) Square root index	(iv) Exposure
(2) Unconditional	0.2593 (0.0005) [0.2826]	0.4018 (0.0006) [0.4119]	0.1660 (0.0008) [0.1495]	0.0983 (0.0003) [0.1402]
(3) Conditional on human capital	0.2651 (0.0005) [0.2769]	0.4075 (0.0006) [0.4062]	0.1681 (0.0008) [0.1473]	0.1047 (0.0003) [0.1341]
(4) Conditional on municipality	0.3511 (0.0007) [0.1811]	0.4948 (0.0007) [0.3037]	0.1993 (0.0008) [0.1141]	0.1191 (0.0003) [0.1199]
(5) Conditional on human capital and municipality	0.3542 (0.0007) [0.1772]	0.4999 (0.0007) [0.2965]	0.2020 (0.0008) [0.1111]	0.1269 (0.0003) [0.1121]
(6) Conditional on human capital, municipality and industry	0.4072 (0.0007) [0.1035]	0.5679 (0.0007) [0.1859]	0.2371 (0.0007) [0.0703]	0.1598 (0.0003) [0.0773]

*Note:* Calculations are based on all (3,457,951) Swedish employees in 2000, employed in one of 219,235 plants which have more than one employee. There are in total 340,041 employed immigrants in the data. Expected values are means from 500 random replications based on equation (2). Standard deviations describe the dispersion between these replications. Index of systematic segregation is defined in equation (3). Definitions of the Gini and Duncan index can be found in Massey and Denton (1988), the square root index is according to Hutchens (2004), exposure is defined in equation (5). Human capital is captured by gender, age (3 dummies) and education (7 dummies). Data include 289 municipalities and 38 industry dummies. All variables are completely interacted in all models.

**Table 2** Ethnic workplace segregation in Sweden in year 2000 –analytical results for exposure.

(A) Recent immigrants (less than 10 years since immigration)			
	(i) To all immigrants	(ii) To “ethnic” peers	(iii) To other immigrants
(1) Actual exposure (s.e.)	0.2733 (0.0043)	0.0735 (0.0014)	0.1997 (0.0039)
(2) Unconditional expectation Odds ratio	0.0983 4.2406	0.0053 15.1136	0.0931 2.9572
(3) Conditional expectation* Odds ratio	0.1771 2.5803	0.0251 3.2443	0.1520 1.9364
(B) Non-recent immigrants (10 years or more since immigration)			
	(i) To all immigrants	(ii) To “ethnic” peers	(iii) To other immigrants
(1) Actual exposure (s.e.)	0.2058 (0.0025)	0.0624 (0.0009)	0.1435 (0.0023)
(2) Unconditional expectation Odds ratio	0.0983 2.9233	0.0109 6.1805	0.0875 2.0991
(3) Conditional expectation* Odds ratio	0.1531 1.9992	0.0302 2.2743	0.1229 1.5537

*Note:* Calculations are based on all (3,457,951) Swedish employees in 2000, employed in one of 219,235 plants which have more than one employee. There are in total 340,041 employed immigrants in the data, whereof 95,555 are “recent”. Standard errors in parentheses are cluster-corrected for dependencies within workplaces. “Ethnic peer exposure” refers to immigrants born in the same (out of 26) region – “exposure to other immigrants” refers to exposure to immigrants born in any other part of the world. \*Conditional expectation is calculated for the interaction between gender, age (3 dummies), education (7), municipality (289) and industry (38).

## WORKING PAPERS\*

Editor: Nils Gottfries

- 2006:28 Per Engström, Ann-Sofie Kolm and Che-Yuan Liang, Maternal Addiction to Parental Leave. 18pp.
- 2006:29 Jonas Björnerstedt and Andreas Westermark, Delay in Bargaining with Externalities. 27pp.
- 2006:30 Pär Österholm, Incorporating Judgement in Fan Charts. 36pp.
- 2006:31 Mikael Carlsson and Andreas Westermark, Monetary Policy and Staggered Wage Bargaining when Prices are Sticky. 26pp.
- 2007:1 Mikael Elinder, Local Economies and General Elections. 26pp.
- 2007:2 Ouarda Merrouche, The Long Term Impact of French Settlement on Education in Algeria. 19pp.
- 2007:3 Ouarda Merrouche, The Long Term Effect of Education Spending Decentralization on Human Capital in Spain. 15pp.
- 2007:4 Erik Post, Macroeconomic imbalances and exchange rate regime shifts. 38pp.
- 2007:5 Christian Andersson, Teacher density and student achievement in Swedish compulsory schools. 31pp.
- 2007:6 Thomas Aronsson, Sören Blomquist and Luca Micheletto, Where Should the Elderly Live and Who Should Pay for their Care? A Study in Demographics and Geographical Economics. 22pp.
- 2007:7 Sören Blomquist and Vidar Christiansen, Public Provision of Private Goods and Nondistortionary Marginal Tax Rates. 17pp.
- 2007:8 Marcus Eliason and Henry Ohlsson, Living to Save Taxes. 13pp.
- 2007:9 Åsa Ahlin and Eva Mörk, Effects of decentralization on school resources: Sweden 1989-2002. 31pp.
- 2007:10 Henry Ohlsson, The equal division puzzle – empirical evidence on intergenerational transfers in Sweden. 20pp.
- 2007:11 Daniel Hallberg and Mårten Lagergren, Moving in and out of public geriatric care in Sweden. 26pp.
- 2007:12 Per Engström, Wage Formation and Redistribution. 22pp.
- 2007:13 Henry Ohlsson, Tax avoidance – a natural experiment. 21pp.

---

\* A list of papers in this series from earlier years will be sent on request by the department.

- 2007:14 David Kjellberg and Erik Post, A Critical Look at Measures of Macroeconomic Uncertainty. 27pp.
- 2007:15 Mikael Carlsson and Andreas Westermarck, Optimal Monetary Policy under Downward Nominal Wage Rigidity. 52pp.
- 2007:16 Robin Douhan and Anders Nordberg, Is the elephant stepping on its trunk? The problem of India's unbalanced growth. 33pp.
- 2007:17 Annika Alexius and Bertil Holmlund, Monetary Policy and Swedish Unemployment Fluctuations. 27pp.
- 2007:18 Meredith Beechey and Pär Österholm, The Rise and Fall of U.S. Inflation Persistence. 23pp.
- 2007:19 Henry Ohlsson and Donald Storrie, Long term effects of public policy for displaced workers in Sweden – shipyard workers in the West and miners in the North. 26pp.
- 2007:20 Niklas Bengtsson, How responsive is body weight to transitory income changes? Evidence from rural Tanzania. 38pp.
- 2007:21 Karin Edmark, Strategic Competition in Swedish Local Spending on Childcare, Schooling and Care for the Elderly. 38pp.
- 2007:22 Fredrik Johansson, How to Adjust for Nonignorable Nonresponse: Calibration, Heckit or FIML? 25pp.
- 2007:23 Henry Ohlsson, The legacy of the Swedish gift and inheritance tax, 1884–2004. 25pp.
- 2007:24 Ranjula Bali Swain and Fan Yang Wallentin, DOES MICROFINANCE EMPOWER WOMEN? Evidence from Self Help Groups in India. 26pp.
- 2007:25 Bertil Holmlund and Martin Söderström, Estimating Income Responses to Tax Changes: A Dynamic Panel Data Approach. 34pp.
- 2007:26 N. Anders Klevmarken, Simulating the future of the Swedish baby-boom generations. 60pp.
- 2007:27 Olof Åslund and Oskar Nordström Skans, How to Measure Segregation Conditional on the Distribution of Covariates. 17pp.

See also working papers published by the Office of Labour Market Policy Evaluation  
<http://www.ifau.se/>