Revisiting the Relationship between Ethnic Diversity and Preferences for Redistribution*

Lena Nekby♣ and Per Pettersson-Lidbom#

This version
September 12, 2015

JEL: D64, I30, Z18, J15, J18

Key words: ethnic diversity, income redistribution

Abstract

In this paper, we revisit the question raised in Dahlberg, Edmark and Lundqvist (2012) concerning a causal relationship between ethnic diversity and preferences for redistribution. We find that their results are based on (i) an unreliable and potentially invalid measure of preferences for redistribution, (ii) an endogenously selected sample and (iii) a mismeasurement of the refugee placement program. Correcting for any of these three problems reveals that there is no evidence of any relationship between ethnic diversity and preferences for redistribution. We also discuss what is currently known about the refugee placement program and to what extent it can be used for estimating causal effects more generally.

* This paper is a heavily revised version of Nekby and Pettersson-Lidbom (2012). We thank Thina Carlsson at the Archives of the Swedish Board of Immigration for generous help with data collection. We also thank the editor and two anonymous referees, Torsten Persson, David Strömberg, Mårten Palme, Peter Skogman Thoursie, Peter Fredriksson, Mikael Lundholm and Mahmood Arai for useful comments.

♣ Lena passed away in July of 2014. Her friendship and her energy will be greatly missed.

Department of Economics, Stockholm University, E-mail: pp@ne.su.se

1. Introduction

In this paper, we revisit the question raised in Dahlberg, Edmark and Lundqvist (2012, henceforth DEL) concerning a causal relationship between ethnic diversity and preferences for redistribution.[1] The question of whether ethnic heterogeneity affects individual behavior, such as preferences, has received considerable attention in the literature (e.g., Alesina and La Ferrara 2005). However, there are few studies that use a credible exogenous source of variation in ethnic diversity. In contrast, DEL exploit a plausible exogenous variation in ethnic heterogeneity arising from a nationwide refugee placement program which placed refugees in municipalities throughout Sweden during 1985–94. DEL also match data on refugee placement to individual survey data on the preferences for redistribution and use the placement policy as an instrument for ethnic diversity as measured by the share of immigrants in the municipality. The instrumental variable results suggest a large and statistically negative relationship between ethnic diversity and preferences for redistribution. However, the empirical design is compromised by three problems, namely (i) that the measurement of attitudes towards redistribution lacks reliability and validity, (ii) endogenous sample selection and (iii) the mismeasurement of the refugee placement program.

Evaluating the validity and reliability of a measurement item in a survey, i.e., whether the response to the survey question measures what it is purported to measure and whether the survey response is consistent, is standard procedure in survey research. However, DEL only report results from one measure of attitudes toward redistribution even though there are at least two other survey items in the Swedish election survey that *a priori* should be equally good or even better measures of preferences for redistribution, at least when it concerns redistribution towards immigrants.[2] Importantly, the results from the two alternative outcome variables do not corroborate the finding in DEL.

The endogenous sample selection is essentially due to the fact that DEL only base their empirical analysis on individuals that answered the survey in two consecutive

[1] In Nekby and Pettersson-Lidbom (2012: Dec. 17) we also discuss other important issues concerning the replication of Dahlberg et al. (2012). We also comment on their reply to our critique, i.e., Dahlberg et al. (2013).

[2] It is noteworthy that the results from these two other outcomes are reported in a previous working paper by Dahlberg and Edmark (2009).

elections in order to create a panel.[3] Thus, the sampling scheme is endogenous since it is based on the outcome, i.e., only individuals which responded to the survey more than once are included in their sample. As a result, there will be a strong sample selection bias since the attrition rate in DEL's sample is 73% relative to the original nationally representative sample of 14,297 observations. In other words, DEL's result is based on a highly selected panel of 1,917 individuals.[4]

The mismeasurement of the refugee placement program is due to the fact that DEL measure the program by the number of grants paid out by the Swedish Immigration Board to local governments.[5] However, it turns out that there is little or no correspondence between this measure and the refugee placement program since the grants (i) cover a large number of individuals that were not part of the program, (ii) were not paid out at the time of the refugee placement but only after a long time lag, given that a local government had submitted a grant application. Consequently, DEL therefore incorrectly label their measure as "the number of refugees placed via the placement program".[6] In fact, DEL's variable is an internal administrative measure used by the Swedish Board of Immigration to register yearly grant payment flows to municipalities which has little to do with the refugee placement program. As a result, DEL's instrumental variable does not accurately measure the placement program and is therefore likely to give biased results due to the mistiming of refugee placements and the endogenous settlements of individuals that were not part of the program.

One of the major contributions of this paper is to properly address the problems with DEL's empirical design. Another important contribution is to discuss what is

---

[3] We pointed out the sample selection problem and discussed how it could be solved with DEL at a seminar at Stockholm University in November 2011 and in several e-mails sent in November and December 2011. To our knowledge, their paper had not been resubmitted to JPE at that time.

[4] DEL also make 8 unreported sample restrictions on the panel data which reduced the number of individuals from 2,702 to 1,917.

[5] The rules governing the payments to municipalities are laid out in Swedish law ("Förordning (1984:683) om statlig ersättning för mottagande av flyktingar och vissa andra utlänningar 1984:683 and Förordning (1990:927) om statlig ersättning för flyktingmottagande m.m.) See links http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Forordning-1984683-om-statl_sfs-1984-683/ and http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Forordning-1990927-om-statl_sfs-1990-927/?bet=1990:927 for more information.

[6] It is noteworthy, that DEL changed the definition of this variable as compared to Edin et al. (2003). In Edin et al. (2003), this variable is correctly expressed as measuring "the number of refugees covered by grants from the Immigration Board."

currently known about the refugee placement program and to what extent it can be used for estimating causal effects more generally. This should be of general interest since the placement program has been used extensively in previous research (e.g., Edin et al. (2003), Edin et al. (2004), Åslund (2005), Åslund and Rooth (2007), Åslund and Fredriksson (2009), Åslund et al. (2010), Åslund et al. (2011) and Grönqvist et al. (2012, 2015)).

We find that there is no evidence for any relationship between ethnic diversity and preferences for redistribution when correcting for the three problems with DEL's empirical approach. We also argue that there currently exists too little statistical information about the refugee placement program for a compelling assessment of to what extent the program can be used to more generally estimate causal effects. Specifically, there is currently no available information about the actual placement of the refugees at the individual level. Thus, we caution people to draw any strong conclusion from this line of work until this data has been collected from the Archives of the Swedish Board of Immigration. Moreover, we also argue that a firm understanding of the local placement policy is necessary for a credible analysis since previous research has wrongly assumed that that placement policy was characterized by a "top-down" rather than a "bottom-up" approach.

The rest of the paper is structured as follows. Section 2 tests the reliability and validity of the measure of preferences. Section 3 discusses the endogenous sample selection problem while Section 4 discusses the problem with the measurement of the refugee placement program. Section 5 reports results from additional specification checks while Section 6 discusses and concludes the paper.

2. Reliability and validity of the measure of preferences

In this section, we investigate whether the result in DEL is robust to alternative survey measures of preferences for redistribution since it is well-known that survey responses can have problems with both reliability and validity which, in turn, may lead to biased findings (e.g., Bound et al. (2001) and Hyslop and Imbens (2001)).[7] In fact, Hyslop and

[7] Alesina and Giuliano (2011) also note another problem with electing preferences for redistribution from survey questions, namely that it is hard to distinguish whether a respondent is in favor of social insurance or redistribution since some aspects of the welfare state are primarily redistributive while others mostly

Imbens (2001) show that the bias of a regression coefficient can be away from zero, i.e., overestimation, even in a regression model with a classical measurement error. This type of bias occurs when both the regressor and the outcome variable are measured with correlated errors. Consequently, DEL's result might be plagued by such bias since both the dependent and the explanatory variables are measured with errors as further discussed below. Alternatively, DEL's finding may also be a result of "p-hacking" (e.g., Brooder et al. (2015), Simmons et al (2011)),[8] since DEL only report results from one measure of preferences for redistribution even though there are other survey items measuring the same construct.[9] Anyway, to rule out these types of problems in DEL's analysis, it is important to test whether these alternative measures of preferences confirm DEL's result or not.

The survey item used by DEL is the answer to the question "what is your opinion about the following proposal", namely "reduce social benefits" and where the following five-point scale is used: (1) very good proposal, (2) fairly good proposal, (3) neither good nor bad proposal, (4) fairly bad proposal and (5) very bad proposal.[10] In the same section of the survey, respondents are also asked about the questions "reduce the public sector" and "increase economic support to immigrants so they can maintain their own culture". *A priori* these questions also seem to be valid measures of attitudes toward redistributions. Indeed, the question about "economic support to immigrants" seems to be an ideal measure given DEL's research question, i.e., whether there exists a causal relationship between ethnic diversity and preferences for redistribution.

Table 1 shows the results. In comparison, Column 1 presents the results using DEL's measure "reduce social benefits" while Column 2 displays the results from "reduce the public sector" and Column 3 from "increase economic support to immigrant."

---

provide social insurance. The question analyzed by DEL, "reduce social benefits," seems to be more about social insurance than redistribution.

[8] P-hacking refers to the practice of reanalyzing data in many different ways to yield a target result, namely a p-value below .05

[9] DEL analyze 9 survey questions according to their computer code, i.e., (1) "accept fewer refugees into Sweden," (2) "reduce income differences in society," (3) "reduce the public sector," (4) "increase economic support to immigrant so they can maintain their own culture," (5) "reduce third world aid," (6) "retain nuclear power," (7) "increase the proportion of health care run by private interest," (8) "reduce social benefits" and (9) "reduce defense spending." Two of the outcomes (6 and 7) are reported in DEL as "placebo analyses" i.e., results where the p-values is not statistically significant. However, we find that all survey questions expect for "reduce social benefits" are not statistically significant at the 5% level.

[10] Individuals reporting "don't know/don't want" are excluded from the analysis.

Importantly, the results in Table 1 do not corroborate the finding in DEL since the estimated effects in Columns 2 and 3 are not statistically significant.[11] Moreover, the estimated effect in Column 2 is even of the opposite sign as compared to DEL's estimate, even though both questions are about *reducing* redistribution. Moreover, while the sign of the estimated effect in Column 3 is consistent with DEL's estimate effect since the question is about an increase rather than a decrease, the size of the effect is considerably smaller than DEL's estimate, i.e., 0.35 versus 0.18.

To conclude, the analysis in this section suggests that the results are based on an unreliable and potentially invalid measure of preferences for redistribution, at least when it concerns redistribution towards immigrants.

3. Endogenous sample selection

Starting with the problem of endogenous sample selection, DEL uses individual data from the Swedish Election Surveys which is a representative sample of the Swedish voting-eligible population. However, DEL only use the response from 1,917 individuals while there are 9,620 available observations from the total sample survey size of 14,297 for all the included elections surveys from 1985, 1988, 1991 and 1994.[12] One reason for the large reduction in the number of observations is that DEL select individuals that are present in two consecutive election surveys in order to create a panel. However, by doing this, they destroy the annual (i.e., cross-sectional) representativeness of the original sample. As a result, their findings may be biased due to endogenous sample selection. It is also crucial to note that neither individual or panel data is required for econometric identification in DEL's empirical design since all their right-hand side variables, including the regressor of interest (the share of immigrants in a municipality) and the

[11] We could not exactly replicate the result in DEL. Our point estimate is –0.367 (s.e.=0.163) using 1,911 observations which should be compared to DEL's estimate –0.347 (s.e.=0.156) using 1,917 observations. One reason for the failed replication is that there are discrepancies between their data and the data from the citied sources. For further discussion about the replication, see our working paper Nekby and Pettersson-Lidbom (2012)

[12] The Swedish election survey consists of about 3,500 individuals each election year. The survey has a rotating panel design in which half of the sample has been interviewed in connection with the previous election and the other half in connection with the following election. DEL use four waves of the election surveys but they only base their estimation on data from respondents in the rotating panel. The attrition rate in their analysis can therefore be calculated as (14297 –1917*2)/14297=73%.

instrumental variable (the refugee placement policy), only vary at the group-time level, which is the municipal-time level in this case.

To formally illustrate this point, we first start by noting that DEL estimate the following population regression model

(1) $$Y_{igt} = \alpha_i + \lambda_t + \beta X_{gt} + \delta W_{gt} + v_{igt},$$

where Y_{igt} is a measure of the preferences for redistribution for individual i in municipality g in time period t, X_{gt} measures the municipal immigrant share, W_{gt} is a set of control variables, α_i is an individual fixed effect and λ_t is a time-fixed effect. The parameter of interest is β, which measures the causal effect of ethnic diversity on the preferences for redistribution in the population, i.e., the Swedish voting-eligible population. Most importantly, this causal effect can also be estimated by aggregating the individual data relationship, i.e., equation (1), into cells based on municipality and time, that is,

(2) $$\overline{Y}_{gt} = \alpha_g + \lambda_t + \pi X_{gt} + \theta W_{gt} + \overline{v}_{gt},$$

where α_g is a grouped fixed effect. Note that the right-hand side variables, X and W, are *completely* unaffected by the aggregation due to the fact that these variables are constant within municipality-time cells. It is now easily shown that the parameter π in equation (2) is *identical* to the parameter β in equation (1), if equation (2) is estimated by weighted least squares (WLS) and where the weights are the number of eligible voters in each cell (Angrist and Pischke 2009, p 39). In other words, the micro data regression (1) is *identical* to a grouped data regression estimated by WLS where the weights are the population shares. As a result, individual data is not required for identification, only data at the group level. Thus, both repeated cross-sectional and panel data can be used for estimating the parameter β in equation (1), as long as the data is representative of the underlying population, since the sample shares would then be identical to the population shares, absent sampling variability. In other words, it is possible to consistently estimate β using all available observations in the Swedish election survey since this data constitutes a random sample from the population. This is also true for DEL's instrumental variable approach since their instrument only varies at the same level as all other right-hand side variables, X and W.[13]

However, DEL do not use a representative sample. Indeed, DEL's approach of creating a panel means that their sampling of observations is endogenous, i.e., sampling in which the probability of selection varies with the dependent variable, even after conditioning on the explanatory variables. Thus, their estimator will *not* be consistent for any population parameter of interest, including the local average treatment effect (LATE), since DEL's estimator uses the wrong (endogenous) weights, i.e., sample shares. Nonetheless, the estimator is consistent if it is weighted by the inverse probabilities of selection, i.e., population shares (e.g., Wooldridge (2002), Cameron and Trivedi (2005, ch. 24)).[14] It is important to stress that the rationale for weighing in this case is not related to correcting for heteroscedasticity or identifying the population average partial effect (on this point see e.g., Solon et al. (2015)). Instead, weighting is needed on consistency grounds due to endogenous sampling.[15]

To illustrate the problem with endogenous sample selection, Table 2 displays the sample used by DEL with 1,917 individuals and two other additional samples from the Swedish Election survey: the total sample with 9,620 observations and the extended panel sample with 2,702 individuals. It is noteworthy that DEL make 8 additional, undiscussed, sample restrictions which reduced the extended panel sample from 2,702 to 1,917 individuals. In the DEL sample, the average municipality-time cell size is 2.2, 3.2 in the extended panel while it is 8.4 for the total sample, i.e., the average cell size differs by a factor of four from the smallest to the largest sample. Moreover, the share of cells with zero observations is 25 percent in DEL's sample, while it is only 3 percent in the total sample. Thus, there will be a large number of cells in DEL's analysis that get zero weights in the estimation, which directly lead to a bias. A second source of bias is that the

[13] The way in which one gets rid of the fixed effects is not relevant for the argument made here. For example, the individual data regression will yield exactly the same estimate as the grouped data regression if individual fixed effects are included in equation (1) or if grouped fixed effects are included in equation (2). Alternatively, we can use a difference transformation, which is the approach taken by DEL.
[14] This method is known as inverse probability weighting and was first discussed by Horvitz and Thompson (1952).
[15] Arellano (2014, p. 22) also discusses the rationale for weighing under stratified sampling. He argues that if there is parameter heterogeneity at the individual level, weighting by population shares is required for estimating population parameters such as LATE.

non-empty cells are weighted incorrectly, i.e., by the endogenous sample weights rather than by the exogenous population weights. The problem of using sample weights in the estimation is exacerbated when there is a large variation in the population shares and when the cell sizes are very small. For example, the largest municipality, Stockholm, had 551,791 eligible voters in 1994 while the smallest municipality had only 2,346 eligible voters (the mean is 23,329 and the standard deviation is 42,757). Thus, one would expect that cell sizes would range from about 1 to 289 since the typical survey size in the Election Studies was about 3,500.[16] Consequently, in the total sample, the sample shares should correspond rather closely to the population shares since there is relatively little non-response (about 30 %). Thus, the problem of endogenous sampling should be of little importance in the total sample. In sharp contrast, when the survey sample size is very small, as in DEL's analysis, we would expect endogenous sampling to be much more of a problem.

To investigate to what extent DEL's analysis is affected by endogenous sample selection, we estimate the effect of ethnic diversity on the preferences for redistribution in all these three samples and check whether the results are robust when weighting the regressions by the population shares instead of the sample shares. Table 3 shows the results from this analysis. Column 1 reproduces the instrumental variable estimate from DEL of –0.347, based on their micro data specification with a sample of 1,917 observations.[17] Column 2 instead shows the results from the group-level specification with 641 cells and weighted by the sample shares. As expected, it gives exactly the same estimate as the individual data specification. Thus, the group-level specification clearly illustrates that individual data is not required for identification; only aggregate data at the group level. Note also that there is *no* gain in statistical efficiency from using micro data rather than grouped data, since the cluster-robust standard errors are exactly the same

[16] The probability of selection is (2,346/6,672,157) in the smallest municipality and (551,791/6,672,157) in the largest municipality where 6,672,157 is the total number of eligible voters in Sweden in 1994.

[17] The results in Columns 1-3 of Tables 3 and 6 are based on DEL's dataset, which is available from the Swedish National Data Service (SND 0906). The results in Columns 4-6 are based on the raw data that we collected from original data sources (e.g., Valundersökningarna, Statistics Sweden, the Archives of the Swedish Board of Immigration).

from the individual- and the group-level specifications.[18] Thus, the individual data specification in DEL is identical to a group-level specification in all respects.

Turning to the test of sample selection bias, we first start by testing whether weighting (sample versus population shares) matters. Column 3 displays the group-level regression weighted by the population shares. The estimate is now reduced from –0.347 to –0.177 and it is no longer statistically different from zero. A Hausman *t*-test equal to 5 also strongly rejects that the two estimates are the same. Thus, this result strongly indicates that DEL results are due to sample selection.

We next estimate the grouped-data specification in the extended panel sample. Column 4 shows that the estimate from the group-level specification weighted by the population shares is –0.125. In the extended panel sample, the number of groups has increased from 641 to 728 and the average cell size has increased by a factor of 1.5 relative to DEL's sample. This extended panel sample is therefore more representative than the limited panel sample used by DEL. The estimated effect has decreased even more than in DEL's analysis, which again clearly illustrates that the estimated effect is biased due to sample selection.

Finally, we estimate the grouped data specification with the total sample, which is the most representative sample. Column 5 shows that the estimated effect is –0.109 when the grouped regression is weighted by population shares, which is again considerably smaller than in DEL but almost the same as the estimated effect of –0.125 from the extended panel sample in Column (4). Moreover, Column 6 shows that the estimated effect is –0.099 when the regression is weighted by the sample shares. Thus, weighting does *not* matter in the total sample since the two weighted estimates are almost identical. As a result, this result strongly suggests that there is *no* sample selection problem in the total sample.

Turning to the precision of the estimated effects, Table 3 reveals that the standard errors are significantly smaller for the group specification in the total sample than for DEL's sample, i.e., 0.081 vs. 0.156. Moreover, the standard errors are even smaller (0.069) if the estimates are based on the micro data for the total sample. The increase in

[18] Typically, the standard errors would be different in group and micro data specifications. However, in DEL's regressions, they happen to be identical since the panel includes exactly two observations per individual.

statistical precision is due to the fact that the total sample has many more cells and much larger cell sizes. Importantly, however, it is still impossible to reject the null hypothesis of no effect in the total sample despite the fact that the standard errors are much smaller than in DEL's analysis.

To conclude, the analysis in this section unambiguously suggests that the significant result of DEL is a consequence of endogenous sample selection.

4. Mismeasurement of the refugee placement program

Turning to the problem of correctly measuring the refugee placement program, DEL describes the refugee placement program as "Under the program, refugees arriving to Sweden were consequently not allowed to decide themselves where to settle but were assigned to a municipality through municipality-wise contracts, coordinated by The Immigration Board". However, DEL do not use the number of contracted refugees to which any given municipality committed annually. Instead, DEL measure the refugee placement program by the yearly payments of grants from the Swedish Board of Immigration (SIV) to local governments for *all* newly arrived refugees as discussed previously. However, these payments cover refugees that were targeted by the refugee placement program as well as other refugees that were not targeted by the placement policy.[19] Moreover, the grants were paid out with a considerable time lag which implies that the yearly payments do not necessarily correspond closely to the number of currently placed refugees.

Unfortunately, we cannot directly test the magnitude of the measurement problem since there does not exist any published data on refugees placed within the placement program as previously noted. However, we can still get a sense of the size of the measurement problem by comparing the number of residence permits with DEL's variable since "whether individuals were subjected to the placement policy or not depended solely on when they received their residence permits" (Edin et al. (2003)). Moreover, Edin et al. argue that "During 1987–1991 the placement rate, i.e., the fraction of refugee immigrants assigned an initial municipality of residence by the Immigration Board, was close to 90 percent". As a result, we would expect a close correspondence

[19] These state grants include payments for tied-movers, refugees and other asylum seekers, none of which are placed in the municipality via the placement program.

between the yearly number of residence permits and DEL's variable if the measurement problem is small. However, Column 3 in Table 4 reveals that there are very large percentage differences of the order of 10-70% (with both negative and positive differences) between these two measures, i.e., Column 1 shows the yearly number of residence permits, Column 2 shows the number of yearly grant payments and Column 3 displays the percentage difference between these two variables. Thus, we conclude that DEL's variable shows little correspondence with their labeling "the number of refugees placed via the placement program."

To try to solve the problem of correctly measuring the refugee placement program, we decided to collect data on the yearly number of contracted refugees at the municipal level. Although the contracted number of refugees may deviate from the actual number of placed refugees, the contracts are still arguably the best way of measuring the refugee placement program since they are determined before the actual placement and therefore not affected by timing problems or endogenous refugee settlement issues. Naturally, the contracts may be endogenous due to local political forces in the past. However, using the actual number of placed refugees would not avoid this problem, since current political factors would now come into play in addition to all other factors affecting refugee settlements. In other words, in this paper, we take DEL's approach of using the placement program as a credible exogenous source of variation in ethnic heterogeneity at face value, i.e., both the actual number of placed refugees and the number of contracted refugees can be valid instruments. Instead, in this section, we analyze to what extent their instrument—grant payments—can credibly capture any behavioral responses related to the Swedish refugee placement policy.

We continue by analyzing the statistical relationship between the municipality-wise contracts and DEL's variable to assess to what extent the grant payments reflect the contracted number of refugees. Thus, we will estimate a regression of the following form

$$Y_{it}=\alpha_i+ \lambda_t + \varphi X_{it} + \psi W_{it} + u_{it} \quad (3)$$

where Y_{it} is the number of payments to municipality i at time t, X_{it} is the yearly number of contracted refugees, and W_{it} is the same set of control variables as used by DEL (e.g., population size, welfare spending, income tax base, housing vacancies, and political variables). The equation is estimated by WLS for it to be consistent with the micro data relationship. We would expect the parameter φ to be close to one if there is a close correspondence between grant payments and contracts. Table 5 shows the results. To check the sensitivity of the relationship across time, we add one year at a time, implying that Row 1 shows the results for 1986-87, Row 2 for 1986-88, Row 3 for 1986-89 and so on. Surprisingly, Column (1) reveals that there is little or no statistical relationship between the two measures when both are expressed in levels since the estimate is both negative and positive and never statistically significant except for one specification, namely the one that includes 1994.[20] However, DEL normalized their variable with the size of the municipality population. Only then is it possible to find a statistically significant positive correlation between two measures (Column 2), but the estimate is still far from one since it is mostly between 0.5 and 0.6. Once again, this analysis confirms that there is a large measurement problem of the refugee placement program in DEL's analysis, since their measure does not correspond closely with the contracted number of refugees or the number of placed refugees. Thus, the conclusions we draw from the above analyses is that their measure—grant payments—cannot accurately capture any behavioral responses related to the Swedish placement policy.[21] Indeed, grant payments are simply an internal administrative measure used by the Board of Immigration to register payment flows.

It is also noteworthy that DEL redefines the refugee placement program in terms of population shares rather than what is stated in the contracts, namely the number of refugees. Whether the refugee placement program should a priori be expressed in levels

[20] The observations from 1994 should not be included since the refugee placement program ended on July 1st 1994.

[21] A referee correctly pointed out that if there are heterogeneous effects in an instrumental variable setting, then each valid instrument captures a different LATE. Consequently, it is theoretically possible to define a number of LATEs, depending on the instrument. Our point here is that there are potentially three candidates for instrumental variables for ethnic diversity: (i) the contracted number of refugees, (ii) the actual number of placed refugees and (iii) DEL's variable. However, our argument is that only (i) or (ii) might be a valid instruments since DEL's instrument is *a priori* invalid due to that it does not accurately measure the placement program. Particularly, the fact that the transfers of grants occur with a lag makes DEL measure a poor instrument. This has to do with the fact that the independent and dependent variables are all measured contemporaneously (t) in DEL's analysis while the instrument is measured at some future date (e.g., t+1). An instrument based on future events does not make any sense. Moreover, DEL's instrument also violates the strict exogeneity assumption, which is a necessary requirement for the fixed effects model used in this paper.

or as a share is worthy of further considerations. Nonetheless, we still stick with DEL's definition when analyzing how DEL's results are affected by using the contracts rather than the payments as the instrumental variable.

Table 6 displays the results from comparing DEL's instrument with the new instrument based on the municipality-wise contracts. Starting with DEL's sample, the IV estimate is reduced from –0.347 to –0.221 and it is no longer statistically significant from zero. The cluster robust first-stage *F* statistic is 32 for this new instrument while it is 66 for DEL's instrument. Thus, there will be no problems due to a weak instrument, since both instruments are strong according to a statistical test developed by Olea and Pflueger (2013) that is robust to heteroscedasticity, autocorrelation, and clustering.[22] Consequently, any difference between the two IV estimates cannot be due to violations of the assumption of instrument relevance, but must be due to a violation of instrument exogeneity. In other words, it seems that the DEL instrument is biased due to the fact that their instrument does not correctly reflect the placement policy.

It is also possible to investigate how DEL's result would change if the two problems were simultaneously addressed. The new instrument deals with the problem of correctly measuring the placement program and the sample selection bias can be dealt with using the same type of tests that we used in the previous section. Thus, we start by comparing the group-level specification weighted by sample share and the population share in the sample with 1,917 individuals. Columns 2 and 3 in Table 6 show that the estimate is reduced from –0.221 to –0.109 in these group-level specifications. Column 4 shows that in the extended panel, the estimated effect is 0.046 when the group-level specification is weighted by the population shares. Thus, the estimated effect is slightly positive if both problems are dealt with simultaneously. In the total sample, the estimated effect from the population weighted regression is 0.056 (Column 5) which is almost the same as in the extended panel in Column 4. Thus, the estimated effect is again slightly positive when the two problems are properly dealt with. If the grouped regression in the full sample is weighted by the sample share, the estimated effect is 0.065 (Column 6). Thus, once more, weighting does not matter in the full sample which suggests that there

[22] According to Olea and Pflueger (2013), if the cluster robust F-statistics is larger than 23.1, the instrument is not weak.

is no problem of sample selections in the full sample despite some issues with nonresponse.

5. Specification checks

In this section, we report three specification checks. The first specification check is to test whether the effect of ethnic heterogeneity on preferences for redistribution is heterogeneous. In other words, can the significant results in DEL be rationalized by heterogeneous effects? One way of testing for the presence of unmodeled heterogeneity of effects is to compare weighted and unweighted estimates (e.g., DuMouchel and Duncan (1983), Solon et al. (2015)). We perform this specification check using the most credible specifications in Table 6, i.e., those in Columns 5 and 6, to avoid the problem of endogenous sampling and mismeasurement of the refugee placement program. The unweighted estimate from this regression is –0.036 with a standard error (s.e.) of 0.142. Thus, this suggests that there is little or no heterogeneity in the estimated effect since this estimate does not differ much from the weighted estimates in Table 6.

The second specification check tests for whether our results are sensitive to related but alternative measures of preferences for redistribution from the Swedish election survey as discussed in section 2. We report results from two additional survey questions, namely (i) “increase economic support to immigrants so they can maintain their own culture” (ii) and “reduce the public sector”. For these two outcomes, we find no effect since the weighted estimates are –0.005 (s.e.=0.085) and –0.046 (s.e.=0.104), respectively.

The third specification check is to test whether the definition of immigrants is of consequence for the result. DEL measure the municipal share of immigrants as the proportion with non-OECD citizenship (according to OECD membership prior to 1994). Given the question of interest, i.e. the impact of ethnic heterogeneity on preferences for redistribution, it seems more natural to define immigrants according to country of birth. A definition based on citizenship is likely to mask part of the heterogeneity that DEL aim at measuring since a proportion of individuals with non-OECD origin are likely to have changed citizenship and will not be counted as immigrants. Indeed, in a European perspective, Sweden is characterized by a relatively high naturalization rate; 65 percent of the foreign born are Swedish citizens (OECD, 2006). As citizenship is not an observable characteristic, the correlation between native preferences for redistribution and ethnic diversity is more likely to be based on the number and the composition of the foreign born in any given municipality. Our results are robust to defining immigrants according to country of birth since the weighted estimate is 0.179 (s.e=0.286).

6. Discussion and conclusion

In this section, we discuss what is currently known about the Swedish refugee placement program and to what extent it can be used for estimating causal effects more generally.

One of the major obstacles to research directed at estimating causal effects by exploiting the refugee placement program as a source of exogenous variation is that there does not exist any published data on the actual placement of the refugees. DEL measure the number of refugees placed via the placement program using data on grant payments from the Board of Immigration to the municipalities. However, DEL's measure is inaccurate since the payments show little correspondence with the number of placed refugees as previously discussed.

There is also a number of other papers besides DEL that exploit the placement of refugees as a source of exogenous variation, e.g., Edin et al. (2003), Edin et al. (2004), Åslund (2005), Åslund and Rooth (2007), Åslund and Fredriksson (2009), Åslund et al. (2010), Åslund et al. (2011) and Grönqvist et al. (2012, 2015). These studies share the common approach of using administrative data at the individual level to try to indirectly measure who was part of the placement program and in which municipally they were initially placed.

Edin et al. (2003) was the first study using this "indirect approach" of identifying refugee immigrants. They use data on country of origin and municipality of residence at the end of the year to identify refugee immigrants. Thus, the indirect approach assumes that the (i) placement rate, i.e., the fraction of refugee immigrants assigned an initial municipality of residence by the Immigration Board, is 100%, (ii) that the initially assigned municipality is the same as the municipality of residence at the end of the year and (iii) that the country of birth unambiguously defines those eligible for the refugee placement policy.

It is hard to assess the plausibility of these assumptions since there does not exist any data to test them. For example, Edin et al. (2003) argue that the placement rate was close to 90% for the period 1987-1991. But this claim is impossible to verify given the lack of data. However, it is still possible to get a rough sense of the plausibility of some of these other assumptions. For example, we know that the internal migration rate is very high for immigrants during the period of study. Dahlberg and Edmark (2008) find that 40% of the immigrants have moved to another municipality within a four-year period.[23] Thus, such a high internal migration could potentially compromise the indirect approach. A comparison of the total number of residence permits to the number of immigrants identified by the indirect approach also provides a check of the reliability of the indirect method. A total of 55,064 residence permits were granted for the period 1987-1991 (see Column 1 in Table 4). However, the indirect approach can only identify 9,883 individuals in the Edin et al. study. Thus, an attrition rate of 82% from the original experimental population necessarily leads to a strong sample selection bias. All the above problems associated with the indirect approach could, however, be solved if individual data on the actual placement of the refugees is collected.[24]

Another important issue in this literature concerns the role of municipalities in the placement of refugees. Previously, we have described that the Swedish Board of Immigration and local municipal governments negotiated and signed contracts concerning the number of refugees to which any given municipality committed annually for a given period of time. Thus, this description suggests that municipal governments play a key role in shaping the Swedish refugee placement program. However, the current literature describes the Swedish Board of Immigration as the key player, while the municipalities and the municipality-wise contracts essentially play no role. Thus, the assumption implicitly made in this literature is that the placement policy was implemented via a "top-down" approach. For example, Åslund et al. (2009) write that "the Swedish Board of Immigration was given the task of assigning newly arrived refugee immigrants to an initial municipality of residence" and "Assigning a refugee to a municipality was conditional to having found a vacant apartment within that particular municipality. (Since individuals were assigned to an apartment, they were in practice

[23] This estimate does not include those immigrants who moved to another municipality within the year of their initial placement.

[24] Edin et al. (2003) use a weighting procedure to deal with the problems associated with the indirect method. However, since they are using the same variables as DEL, i.e., grant payments, rather than the actual placed refugees, as weights, this approach cannot solve the problem.

assigned to a neighborhood)" and "Available public housing essentially determined the placement".

However, there is a number of compelling reasons to question the top-down approach. First, the Swedish Board of Immigration has no jurisdiction over municipalities. Thus, refuges cannot be placed in a municipality by the Board of Immigration (SIV) without an explicit consent from the municipality (Soinen 1992).[25] Second, the municipalities had their own local agencies (e.g., "den kommunala invandrarbyrån") that were responsible for the placement of refugees within the municipality (Soinen 1992). For example, the municipality of Bollnäs did not place refugees in the first available apartment but choose to place them in neighborhoods with few social problems and where the apartments were of high quality.[26] Moreover, the local refugee policy differed sharply between the municipalities (Soinen 1992).[27] Third, the claim that the placement of refugees was essentially determined by available public housing receives little empirical support. As part of our re-analyzes of DEL, we collected data on annual public housing vacancies, available at two different dates, March 1st and September 1st. When we analyze this data, we find little difference in the number of contracted refugees between municipalities with zero housing vacancies and those with available vacancies. On average, slightly more than 50 refugees were contracted on a yearly basis to municipalities with no housing vacancies while, on average, about 70 refugees were contracted to municipalities with available housing vacancies. In addition, we find that there is a negative relationship between the contracted number of refugees and housing vacancies. Thus, taken together, the above discussion suggests that the refugee placement program is characterized by a "bottom-up" rather than a "top-down" approach.

---

[25] Soinen (1993, p 173) writes " SIV hade inget mandat för att direkt styra kommuner eller kommunala myndigheter, vare sig det handlade om att sluta avtal om mottagande med dem, eller om innehållet i mottagandet. Någon organisationsstyrning i meningen att överordnade (statliga) organ styr underordnade organ på lägre förvaltningsnivåer kunde det sålunda inte vara tal om".

[26] Soinen (1992, p. 96) writes "Flyktingarna skulle inte tilldelas första bästa lägenheter som råkade stå tomma. Inte heller skulle flyktingfamiljema koncentreras till ett visst bostadsområde. Sett ur flyktingarnas perspektiv betydde denna policy att de fick bra bostäder i bostadsområden som inte belastades av sociala problem."

[27] Soinen (1992, p. 155) writes "Valmöjligheterna för organiseringen av mottagandet liksom hur arbetet bedrevs på den lokala nivån var stora. Resultatet blev tydliga skillnader i kommunernas flyktingarbete såväl organisatoriskt som i policyhänseende."

The bottom line of this discussion is that the Swedish refugee placement policy did not work as previously described. Specifically, the municipalities played an essential role in shaping the refugee placement policy. Thus, the empirical design currently used in this literature is not compelling since the role of municipalities has not been taken into account at all. Consequently, an empirical design based on a firm understanding of the working of the placement policy at the local level is necessary for a credible analysis.

To conclude, a re-analysis of the results reported in Dahlberg, Edmark and Lundqvist (2012) shows that the results are compromised by an unreliable and potentially invalid measure of preferences for redistribution, sample selection bias and mismeasurement of the refuge placement policy. Correcting for either of these problems reveals that there is no evidence of any relationship between ethnic diversity and preferences for redistribution.

References

Alesina, A., and E. La Ferrara (2005). “Ethnic Diversity and Economic Performance.” *Journal of Economic Literature*, 43:762–800.

Alesina, A. and P. Giuliano, (2011), Preferences for redistribution, in J. Benhabib, M. Jackson and A. Bisin, eds, `Handbook of Social Economics, Vol. 1A', North Holland, Amsterdam, pp. 93{131.

Angrist, Joshua and Jörn Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton University Press.

Arellano, Manuel (2014), “Econometrics of Survey Data,” Lecture notes, CEMFI. (http://www.cemfi.es/~arellano/survey-metrics-sep2014-tr.pdf)

Åslund, Olof (2005), “Now and forever? Initial and subsequent location choices of immigrants,” *Regional Science and Urban Economics*, 35(2), 141–165.

Åslund, Olof and Dan-Olof Rooth (2007), “Do when and where matter? Initial labour market conditions and immigrant earnings,” *The Economic Journal*, 117 (518), 422–448.

Åslund, Olof and Peter, Fredriksson (2009), “Peer Effects in Welfare Dependence – Quasi-experimental Evidence,” *Journal of Human Resources*, 44(3), 799–825.

Åslund, Olof, Per Anders Edin, Peter Fredriksson, and Hans Grönqvist (2011), “Peers, Neighborhoods, and Immigrant Student Achievement – Evidence from a Placement Policy” *American Economic Journal: Applied Economics*, 3(2), 67-95.

Åslund Olof, John Östh and Yves Zenou (2010), “How important is access to jobs? Old question—improved answer,” *Journal of Economic Geography*, 10 (3): 389-422.

Bound, John, Charles Brown, and Nancy Mathiowetz (2001), “Measurement error in survey data.” Handbook of Econometrics 5: 3705-3843.

Brooder, Abel, Lé, Mathias; Sangnier, Marc and Zylberberg, Yanos, (2015), “Star Wars: the Empirics Strike Back” forthcoming in AEJ: applied

Cameron, Colin and Pravin Trivedi (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, Cambridge

Dahlberg, Matz and Karin Edmark (2009), “Ethnic Diversity and Preferences for Redistribution,” mimeo (https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIPF65&paper_id=404)

Dahlberg, Matz, Karin Edmark and Heléne Lundqvist (2012), “Ethnic Diversity and Preferences for Redistribution,” *Journal of Political Economy*, 120 (1), 41-76

Dahlberg, Matz, Karin Edmark and Heléne Lundqvist (2013), “Ethnic Diversity and Preferences for Redistribution: Reply” Working Paper Series 2013:4, Uppsala University, Department of Economics.

Dahlberg, Matz and Karin Edmark (2008), “Is There a ‘Race-to-the-Bottom’ in the Setting of Welfare Benefit Levels? Evidence from a Policy Intervention.” *Journal of Public Economics*, 92:1193-1209.

DuMouchel, W. K., and G. J. Duncan (1983), “Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples,” *Journal of the American Statistical Association*, 78, 535–543.

Edin, Per Anders, Peter Fredriksson and Olof Åslund (2003), “Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment,” *Quarterly Journal of Economics* 118, 329–357.

Edin, Per Anders, Peter Fredriksson and Olof Åslund (2004), “Settlement Policies and the Economic Success of Immigrants,” *Journal of Population Economics* 17, 133–155.

Grönqvist, Hans, Per Johansson and Susan Niknami (2012) “Income Inequality and Health: Lessons from a Refugee Residential Assignment Program,” *Journal of Health Economics*. 31(4), 617–629.

Grönqvist, Hans, Susan Niknami and Per-Olof Robling (2015) “Childhood Exposure to Segregation and Long-Run Criminal Involvement: Evidence from the ‘Whole of Sweden’ Strategy”, SOFI Working Paper No. 1, Stockholm University.

Horvitz, D. G.; Thompson, D. J. (1952). “A Generalization of Sampling without Replacement from a Finite Universe”. *Journal of the American Statistical Association*, 47: 663–685.

Hyslop, Dean and Guido Imbens (2001), “Bias From Classical and Other Forms of Measurement Error,” *Journal of Business and Economic Statistics*, October 2001, Vol. 19, No. 4, 475-81.

Nekby, Lena and Per Pettersson-Lidbom (2012: Dec 17), “Revisiting the Relationship between Ethnic Diversity and Preferences for Redistribution”, mimeo, Stockholm University.
(http://www.ne.su.se/polopoly_fs/1.214878.1418656425!/menu/standard/file/RevisitingEthnicPreferencesDec17.pdf)

Olea, José Luis Montiel, and Carolin Pflueger. (2013) “A Robust Test for Weak Instruments.” *Journal of Business & Economic Statistics* 31(3), 358-369.

Simmons, Joseph P., Leif D. Nelson and Uri Simonsohn (2011), “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant,” *Psychological Science*, 22(11) 1359–1366.

Soininen, Maritta (1992), Det kommunala flyktingmottagandet. Genomförande och organisation (Local refugee-care – implementation and organization) Stockholm, CEIFO, Centre for Research in International Migration and Ethnic Relations, Stockholm University

Soininen, Maritta (1993), 1985 års flyktingomhändertagande: från reformbeslut till genomförande under ändrade villkor. In Tomas Hammar (ed), *Invandring, forskning, politik: en vänbok till Tomas Hammar*, Stockholm, (1993), CEIFO, Centre for Research in International Migration and Ethnic Relations, Stockholm University.

Solon, Gary, Steven J. Haider and Jeffrey Wooldridge (2015), “What Are We Weighting For?” *Journal of Human Resources*, 50(2), 301-316.

Woldridge, Jeffrey (2002), “Inverse probability weighted M-estimators for sample selection, attrition, and stratification,” *Portuguese Economic Journal*, 1, 117–139

Table 1. Test of the reliability and validity of measuring preferences for redistribution

Survey question	Reduce social benefits (1)	Reduce the public sector (2)	Increase economic support to immigrants (3)
IV estimate	-0.347 (0.156)	0.093 (0.159)	0.183 (0.130)
Observations	1,917	1,917	1,917

Note: The result in Column 1 is based on DEL's data set (i.e., SND 0906) while the data in Column 2 and 3 are extracted from the original data source, i.e., Svenska Valundersökningarna, and merged to DEL's data.

Table 2. Sample information

	DEL's sample (1)	Extended panel sample (2)	Total sample (3)
Number of observations	1,917	2,702	9,620
Average cell size (min, max)	2,2 (0, 51)	3,2 (0, 68)	8,4 (0, 203)
Share of empty cells (%)	25	15	3

Table 3. Tests of sample selection bias

	DEL's sample			Extended panel sample	Total sample	
Level of analysis	Individual	Group	Group	Group	Group	Group/Individual
	(1)	(2)	(3)	(4)	(5)	(6)
IV estimate	–0.347	–0.347	–0.177	–0.125	–0.109	–0.099
	(0.156)	(0.156)	(0.154)	(0.146)	(0.081)	(0.079)
						[0.069]
Sampling weights		Sample shares	Population shares	Population shares	Population shares	Sample Shares
Observations	1,917	641	641	728	1,106	1,106

Notes: Standard errors are clustered at the municipal level are within parentheses. The clustered standard errors for the micro data specification with 9620 observations is within brackets. The results in Columns 1-3 are based on DEL's data set (i.e., SND 0906). The results in Columns 4-6 are based on the raw data that we collected from the original data sources (e.g., Valundersökningarna, Statistics Sweden, and the Archives of the Swedish Board of Immigration).

Table 4. Number of residence permits and grant payments

Year	Residence permits (1)	Grant payments (DELs variable) (2)	Percentage difference between Columns (2) and (1) (3)
1985	7,314	12,235	67
1986	11,486	14,839	29
1987	14,042	18,665	33
1988	16,125	17,935	11
1989	24,879	21,173	-14
1990	12,839	22,251	73
1991	18,663	18,842	1
1992	12,791	18,546	45
1993	36,482	25,218	-31
1994	44,875	62,853	40

Note: The information about residence permits comes from the Swedish immigration board (see link http://www.migrationsverket.se/download/18.39a9cd9514a34607721127f/1421326239958/Beviljade+uppeh%C3%A5llstillst%C3%A5nd+1980-2014.pdf)

Table 5. The relationship between grant payments and contracted refugees

	Levels (1)	Share (2)
Estimate for period: 86-87	0.131	0.556
	(0.407)	(0.085)
Estimate for period: 86-88	0.325	0.586
	(0.358)	(0.062)
Estimate for period: 86-89	0.422	0.647
	(0.422)	(0.047)
Estimate for period: 86-90	0.049	0.650
	(0.141)	(0.042)
Estimate for period: 86-91	0.022	0.614
	(0.107)	(0.039)
Estimate for period: 86-92	-0.132	0.495
	(0.075)	(0.086)
Estimate for period: 86-93	-0.211	0.501
	(0.226)	(0.086)
Estimate for period: 86-94	0.612	0.640
	(0.160)	(0.086)

Note: A full set of municipality and time fixed effects are included in the specifications as well as a full set of control variables: population size, welfare spending, income tax base, housing vacancies, and three political variables. Standard errors are reported in the parentheses and clustered at the municipality level. The results are based on the raw data that we collected from the original data sources (e.g., Statistics Sweden, and the Archives of the Swedish Board of Immigration).

Table 6. The instrumental variable specification based on contracted number of refugees

	DEL's sample			Extended panel sample	Total sample	
Level of analysis	Individual	Group	Group	Group	Group	Group/Individual
	(1)	(2)	(3)	(4)	(5)	(6)
IV estimate	–0.221	–0.221	–0.109	0.046	0.056	0.065
	(0.210)	(0.210)	(0.197)	(0.214)	(0.114)	(0.094)
						[0.081]
Sampling weights		Sample shares	Population shares	Population shares	Population shares	Sample Shares
Observations	1,917	641	641	728	1,106	1,106

Notes: Standard errors are clustered at the municipal level are within parentheses. The clustered standard errors for the micro data specification with 9620 observations is within brackets. The results in Columns 1-3 are based on DEL's data set (i.e., SND 0906) but where we have used our instrument in their specification. The results in Columns 4-6 are based on the raw data that we collected from the original data sources (e.g., Valundersökningarna, Statistics Sweden, and the Archives of the Swedish Board of Immigration).