

The Immigrant-Native Test Score Gap: A Quantile Regression Perspective

Einreichung für die Frühjahrstagung der DGS-Sektion „Soziale Ungleichheit und Sozialstrukturanalyse“, GESIS MANNHEIM, 22.-24.03.2023

Sebastian E. Wenz, Leibniz Institute for the Social Sciences, Germany

Usually, researchers in social stratification turn to traditional regression models that focus on conditional means—e.g., linear regression or hierarchical linear models (HLMs)—when assessing differences between immigrants and natives in metric outcomes, such as test scores in educational assessments. These models estimate the conditional mean function, $E(y|x) = x\beta$, and, thus, estimate differences between the mean test score of immigrants and the mean test score of natives. However, such models ignore that differences in means between immigrants and natives—or any other two or more sociodemographic groups, for that matter—are not necessarily the same as differences in other points (e.g., the 10th percentile, the median, or the 90th percentile) of the outcome’s distribution.

In my contribution I show how quantile regression (Koenker & Bassett 1978, Koenker 2005) provides a convenient way of assessing differences between immigrants and natives beyond the mean by estimating the conditional quantile function, $Q_\tau(y|x) = x\beta_\tau$.

I illustrate the typical use of quantile regression for studying immigrants with an analysis of all 79 countries and regions in OECD’s PISA 2018 study. I show that, in many countries, the immigrant-native educational achievement gap in quantiles differs remarkably from the mean difference. I discuss potential statistical and substantive explanations—the consequences of processes of (self-)selection, immigration, integration, and discrimination—for such differences between linear regression models and quantile regression models and show that these differences also vary across countries.

My analyses show that there are countries in which the *immigrant disadvantage* does *not* differ across quantiles (e.g., USA, United Kingdom, Denmark), in which it is *larger* for lower quantiles than for higher quantiles (e.g., Germany, Finland, Uruguay), and there are countries in which it is *smaller* for lower quantiles than for higher quantiles (e.g., Albania, Brazil, Indonesia).

In some countries that show an *immigrant advantage*, I observe *no* differences across quantiles (e.g., Jordan, Saudi Arabia), in some countries or regions the advantage is *larger* for lower quantiles than for higher quantiles (e.g., Macao, Singapore), and in some it is *smaller* for lower quantiles than for higher quantiles (e.g., United Arab Emirates, Panama). Last but not least, there are countries in which *both* an immigrant disadvantage and an immigrant advantage can be observed when looking at multiple points of the test-score distributions: I find an *immigrant disadvantage for lower* quantiles but *immigrant advantage for higher* quantiles in Latvia, Malta, and Malaysia. The reversed pattern of an immigrant *disadvantage for higher* quantiles but *immigrant advantage for lower* quantiles I find in Argentina.

I conclude by a brief discussion of the most important limitations and misconceptions around quantile regression (for recent discussions see, e.g., Borgen et al 2021, Wenz 2019).