Immigrant-native wage gaps in time series: Complementarities or composition effects?

Joakim Ruist

Department of Economics
University of Gothenburg
Box 640
405 30 Gothenburg, Sweden
joakim.ruist@economics.gu.se
+46 31 786 2915

January 2013

Abstract

This study investigates the role of immigrant composition in creating the observed negative correlation between immigrant supply and immigrant wages in US time series, which has recently been interpreted as evidence of immigrant-native complementarities in production. The main finding is that compositional effects fully explain this empirical pattern. Newly-arrived immigrants, notably Latin Americans, earn less than previous immigrants. Hence their presence decreases average immigrant wages, mechanically. Controlling for this, no negative relation between immigrant supply and immigrant wages remains. More generally, the findings highlight problems in structural estimation of wage effects of immigration, and also the need to distinguish between effects of immigrants of different origins.

Keywords: Immigrant-native complementarities, wage gaps, immigrant composition

JEL codes: J31, J61
1 – Introduction

One of the more contentious issues in the recent academic literature on immigration and the labor market is that of the substitutability between immigrants and natives when these are similar in terms of education and work experience. Whether they are perfect substitutes or not is of first-order importance for how we expect immigration to affect the labor market outcomes of natives and previous immigrants respectively. Recent influential studies from the US (Ottaviano and Peri, 2012), the UK (Manacorda, Manning, and Wadsworth, 2012), and Germany (D’Amuri, Ottaviano, and Peri, 2010) have shown in time series that the immigrant-native wage gap is significantly higher in the parts of the workforce – defined by education and work experience – where immigrants are more numerous. These results have been interpreted as evidence of immigrant-native complementarities in production. However, an alternative interpretation that cannot yet be ruled out is that these results may be mechanical effects of changing immigrant composition. It is known from previous literature that immigrant earnings vary substantially by region of origin and by the number of years since immigration, also when controlling for education and experience (Borjas, 1995; Bratsberg, Barth, and Raaum, 2007). If this variation is not itself due to complementarities but rather to, e.g., differences in human capital transferability to the immigration country, then increasing immigration from certain regions could create the results observed as a purely mechanical compositional effect also in the absence of immigrant-native complementarities.

The study by Ottaviano and Peri (hereafter OP) has been particularly influential. Their estimated degree of immigrant-native complementarity implies that US immigration 1990-2006 decreased previous immigrants’ wages by more than 6% in their simulations, yet increased natives’ wages by about 0.6%. Manacorda, Manning, and Wadsworth estimated even stronger complementarities. Due to the magnitudes and influence of these estimates it is
important to investigate to what extent they are due to complementarities and compositional
effects respectively. The present study investigates this issue on US data. The data underlying
the results from the UK and Germany were not of sufficient quality to enable similar
investigation.

2 – Data and method

Following OP this study analyzes supply and wages of immigrants and natives at four levels
of education and eight levels of work experience. The data used is from the US
Census/American Community Surveys 1960-2006 (Ruggles et al., 2010). See Ottaviano and

2.1 – Immigrants’ earnings heterogeneity

Table 1 documents the variation over time in US immigrant stocks by origin and in
immigrants’ wages by origin. It documents a steady increase in the shares of immigrants from
the low- and middle income regions. The increase in the Latin American (includes the
Caribbean) share is particularly strong, amounting to 37 percentage points of the immigrant
workforce from 1960 to 2006. Looking at wages, high-income region immigrants’ residual
wages are consistently significantly higher than natives’. All other immigrant groups
consistently earn less than natives, almost always with significant differences. Latin American
immigrants have the lowest residual wages in all six years, and their difference to natives is
always significant. The immigrant group that has increased the most over time is thus the one
that consistently has the lowest wages.

Time spent in the US is also an important predictor of immigrant wages, at least for
immigrants from low- and middle income countries. For example, in the year 2000 the
average recently-arrived Mexican immigrant worker with exactly high school education and
21-25 years of potential work experience earned only slightly more than half of what the average native worker with similar education and potential experience did. Yet the average similarly educated and experienced Mexican immigrant that arrived 21-25 years earlier earned about the same as the native worker.

2.2 – Estimation strategy

This study investigates the role of immigrant earnings heterogeneity and compositional effects of immigration in creating the observed negative correlation in time series between supply of and returns to immigrated labor. This is done by controlling for compositional factors in regressions that are similar to those performed by OP, i.e., by adding control variables and by analyzing subsamples of immigrants separately. Following a long tradition, OP estimated the elasticity of substitution between immigrant and native workers using the equation

$$\ln\left(\frac{W_I}{W_N}\right) = edex + year - \frac{1}{\sigma} \ln\left(\frac{H_I}{H_N}\right).$$

Equation (1) expresses the average wage ($W$) ratio of immigrants ($I$) and natives ($N$) in the education-experience cell as a function of dummies for each education-experience cell ($edex$), and for each year, and of the ratio between total hours worked ($H$) by immigrants and natives in the cell. The regression coefficient on relative supply (hours worked) is interpretable as minus one over the elasticity of substitution between immigrant and native workers, given similar education and experience. The expected sign of the regression coefficient is thus negative. A crucial assumption is that relative supply only affects relative wages in the same cell, with no effects across cells.

In the regressions reported by OP, the coefficient on the relative supply of immigrant workers in Equation (1) has the expected negative sign. This is interpreted as evidence of less than
perfect substitutability between immigrant and native workers. It is clear that the average immigrant wage is lower when immigrants are more numerous. What it yet is to be answered is whether this implies lower expected wages for actual individual immigrants or that when immigrants have been more numerous, larger shares of them have been from lower-earnings categories, driving down the averages. If the regression coefficient of interest stays negative and significant when immigrant composition is controlled for, this further supports the previous conclusion that the coefficient does measure the degree of substitutability. If it does not, compositional effects are found to be the driver of the previous result. It then remains to be analyzed whether the important immigrant earnings heterogeneities are themselves partly due to complementarities.

3 – Results

OP’s estimation results in the pooled sample of men and women are replicated using data and program code from Giovanni Peri’s website. These results are shown in row (1) of Table 2. The remaining rows add control variables for average years since immigration (YSM) among the immigrants in the cell, the fraction of immigrants who are from high income regions (Anglo America, Europe, and Oceania), and the fraction of immigrants who are from Latin America.¹ We see that controlling for years since migration or shares of immigrants from the high-income regions has virtually no effect on the estimated coefficient of interest, which remains negative and significantly different from zero. Yet, when controlling for shares of immigrants being from Latin America, the estimated coefficients on relative immigrant supply are either insignificant or significant with the wrong sign. All indications of a

¹ While the measure of relative immigrant supply was expressed in logs, for the regression coefficient to be interpretable as minus one over the (constant) elasticity of substitution, the fractions of high-income region or Latin American immigrants are not logged as they reflect compositional effects on the averages of immigrant wages.
functional negative correlation between relative supply and relative wages of immigrants disappear. These patterns also look very similar if the sample is restricted to the lower education groups, for which OP identified the most important immigrant-native complementarities.

These results indicate that the share of immigrants in the education-experience cell has served as a proxy for the share of Latin American immigrants among the immigrants in the cell, since the share of Latin Americans among immigrants has been strongly correlated with the share of immigrants in the workforce. The residual (regressed on dummies for all education-experience cells, and all years) correlation between log relative immigrant supply and share of Latin Americans among immigrants is between 0.60 and 0.90 for each of the four education groups.

Another way to arrive at a similar conclusion is to substitute smaller subgroups of immigrants for all immigrants on the left hand side, yet not on the right hand side, of Equation (1). The coefficient of interest is then a measure of the effect of the share of immigrants in the workforce on the average wage of the smaller immigrant subgroup relative to natives. This measure is not interpretable as an elasticity of substitution, yet it serves another purpose. If there are immigrant-native complementarities in production, then the expected wage of the individual immigrant – relative to the average native – decreases with the share of immigrants in the workforce. If the total stock of immigrants is then partitioned into a number of subgroups, this decrease must by construction be visible for at least one of these subgroups. Otherwise the negative effect on the average immigrant wage must be due to changing weights of the subgroups on the immigrant average.

Rows (5) – (8) of Table 2 report the results of these modified regressions for four subgroups of immigrants defined by region of origin. We see that two point estimates are positive and
two are negative, and nothing is significant. With no significant effect in any subsample, the conclusion is that the significant negative effect in the total sample in row (1) is rather due to changing weights of the four subsamples, i.e., a compositional and not a complementarity effect.

The crucial question that remains to be answered is then whether the low earnings of the Latin Americans are themselves due to complementarities. Possibly, it could be the case that although the average immigrant is a perfect substitute for the average native in production, the average Latin American immigrant is not. This question is easily answered by substituting Latin American immigrants for all immigrants, and non-Latin American immigrants and natives for natives in Equation (1). The resulting coefficient of interest is reported in row (9) of Table 2. It is very small, and far from providing any significant support for the hypothesis of complementarities between Latin American immigrants and other workers.

4 – Discussion

The results presented in this study do not imply that immigrant and native workers are indeed perfect substitutes. The opposite still seems to be a plausible hypothesis, as immigrants are substantially more concentrated in some occupations than in others (Peri and Sparber, 2009). Yet the existence or magnitude of immigrant-native complementarities should not be seen as established by the time series evidence presented in previous studies. A plausible reason for why this empirical strategy fails to identify complementarities is likely to be found in the construction of the aggregates. By the empirical specification, each of the 32 education-experience cells is in fact treated as an isolated labor market. This has strong implications: for example wages of high school dropouts with 21-25 years of experience are completely insulated from supply of high school dropouts with 15-20, or 26-30, years of experience. A more plausible empirical setup would have to allow for a richer variety of cross-effects.
The method of estimating/simulating wage effects of immigration in structural frameworks, with assumed national-level production functions and parameters estimated from time series data, was originally used by Borjas (2003). It was an alternative to earlier more direct estimation methods, which suffered from potentially serious endogeneity issues. OP later highlighted the lack of robustness of Borjas’ results. In turn, Borjas, Grogger, and Hanson (2012) highlighted the lack of robustness of OP’s results. Dustmann and Preston (2012) have also questioned the construction of the education-experience cells. The structural method is also based on quite strong assumptions regarding stability in parameter trends in the production functions, and crucially that parameters do not respond to immigrant inflows (evidence that production technology does respond to immigration is presented by Lewis, 2003). Adding the findings of this study, i.e., that compositional effects can importantly bias the results, it does seem questionable that this new method has indeed produced any more trustworthy results.

The findings in this study also more generally highlight a weakness in previous literature on immigration and labor markets: too often ignoring that since immigrants of different origin perform very differently on the labor market their presence should be expected to affect other workers differently. This issue ought to be better accounted for in future research on the subject.
References


Table 1. Immigrant shares and wages by origin

<table>
<thead>
<tr>
<th>Origin</th>
<th>Share of immigrant workforce %</th>
<th>Residual wage relative to natives %</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA, Eu, Oc</td>
<td>81</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Latin America</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Asia</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Africa</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.20)</td>
</tr>
</tbody>
</table>

Notes: “AA, Eu, Oc” refers to Anglo America, Europe, and Oceania. Latin America includes the Caribbean.

Residual wages relative to native wages are given by one minus the coefficients on dummies for each region in a log weekly wage regression – for full time workers only – that includes a dummy for each education-by-experience cell. P values in parentheses refer to the significance of the estimated dummy parameter. Standard errors are clustered at the education-experience cell.
Table 2. Effects of relative immigrant supply on relative immigrant wages

<table>
<thead>
<tr>
<th>Specification</th>
<th>(1) All workers</th>
<th>(2) Full time workers only</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Original</td>
<td>-0.026*</td>
<td>-0.039***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>(2) + YSM</td>
<td>-0.041**</td>
<td>-0.054***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>(3) + AA, Eu, Oc</td>
<td>-0.033***</td>
<td>-0.045***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>(4) + Latin</td>
<td>0.047***</td>
<td>0.037**</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>(5) AA, Eu, Oc subsample</td>
<td>0.011</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>(6) Latin subsample</td>
<td>-0.004</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>(7) Asia subsample</td>
<td>0.021</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>(8) Africa subsample</td>
<td>-0.069</td>
<td>-0.086</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>(9) Latin complementarity</td>
<td>-0.004</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>

Notes: Each cell reports the estimate of the parameter of interest from a separate regression. Robust standard errors in parentheses. n=160 when YSM is included (no data from 1960); n=192 otherwise. *, **, and *** refer to significance at the 10%, 5%, and 1% levels respectively. Observations are education-experience cells. Fixed effects for education-experience cells and for years are included. AA, Eu, Oc refer to Anglo America, Europe, and Oceania.