A Further details on skill measures

Unfortunately, there is no representative workforce survey in Brazil that would provide information about the skills workers require in their occupations. In the US there are two such surveys that ask workers to state the importance of various ability requirements and activities performed in their professional life. The DOT is used in the seminal paper on tasks/skills by Autor et al. (2003). In many subsequent papers, such as Acemoglu and Autor (2011), the O*NET survey, the DOT’s predecessor is used to replicate their skill definitions. Maciente (2013) provides a matching of US occupations to Brazilian occupations at the most disaggregated (6-digit) level using their names and synonyms. As a further validation, each occupation code from both countries is compared to the international ISCO88 classification, for both of which a transition already exists. Despite the efforts, Maciente (2013) admits that the mapping of occupations may introduce unpredictable measurement error in the skill scores. By far and large, those fluctuations will not change that fact that a truck driver requires more manual and less analytical skills than a bank accountant which in turn requires less interpersonal and analytical skills than his director. In other words, the skill ranking of occupations should be largely unaffected by measurement error.

Given the availability of the crosswalk from the O*NET data to Brazilian occupations, we can replicate the definitions for analytical, cognitive and manual skills provided in Acemoglu and Autor (2011). To give an example, the score of analytical skills is composed of the importance of the following elements in the O*NET survey: "analyzing data or information", "thinking creatively" and "interpreting the meaning of information for others". Firpo et al. (2011) is another interesting study based on O*NET data which, however, introduces some proper but related skill definitions. We chose to work with their definition of "face-to-face" skills because it complements the three other skill definitions. Other papers by Michaels et al. (forthcoming), Ehrl and Monasterio (2017) and Bacolod et al. (2009) have also recently stressed the importance of "human interactions" or "soft skills".
face-to-face skills also can be seen as a robustness check in the way the our results are not driven by a specific definition of skills. Some elements in the definitions of cognitive and interpersonal skill overlap. Both definitions include "establishing and maintaining interpersonal relationships" and "coaching and developing others". Additionally, the cognitive skill measure captures "guiding, directing, and motivating subordinates" and "how important being very exact or highly accurate" is. Face-to-face skills in turn include "assisting and caring for others", "working directly with the public", as well as the frequency of "face-to-face discussions".

In the main text, we described the three steps that aggregate the O*NET elements and transform them to AMC-specific skill concentration indices that are ultimately used as our dependent variable. Here we complement the verbal descriptions with the exact formulas.

\[
\text{skill}_{io} = \frac{1}{S} \sum_{s=1}^{S} (\text{O*NET element})_{sio}
\]

\[
\text{\overline{skill}}_{io} = 10 + \frac{\sigma_{\text{skill}}}{\text{\overline{skill}}}
\]

\[
\text{skill}_k = \frac{1}{N_k} \sum_{i \in k} \left( \text{\overline{skill}}_{io} \right)
\]

The first equation shows the aggregation of the \( s \) different O*NET elements for each individual \( i \) with occupation \( o \). The second equation shows the standardization to have a mean of 10 and a standard deviation of 1, where \( \sigma_{\text{skill}} \) is the standard deviation and \( \overline{\text{skill}} \) is the grand mean of the standardized skill variable in the occupation data. The last equation calculates the skill concentration measure \( \text{skill}_k \) in each AMC \( k \). It is a average, i.e., the sum of the workers' standardized skill scores divided by the number of those workers in AMC \( k \).

To avoid repetitive descriptions and results, the study focuses only on the face-to-face skill measure. As noted above, there is some overlap between cognitive and face-to-face skills. Aside, manual skills are largely mirror-image to those other skills. A large intensity of manual skills may as well be interpreted as the absence of analytic and interpersonal skills. The correlation between manual skills and analytic and face-to-face skills is equal to -0.59 and -0.72, respectively. The focus and presentation of one skill concentration is sufficiently informative, the more so as the other skill measures do not add different information. Cognitive and analytic skills basically reproduce the results of face-to-face skills and manual skills yield a reverse image.\(^1\)

The upper plot in figure A.1 shows the average skill scores along the population weighted

\(^1\) Manual skills are mainly concentrated in rural areas and the coefficient of their concentration in wage regressions is negative. This negative coefficient is in line with (Bacolod et al. 2009: 145): "motor skills have a hedonic price that decreases with MSA population". Even though motor skills and manual skills are meant to capture the same, the variables in Bacolod et al. (2009) are defined differently because another, anterior data base (DOT) provides the importance of skills in professions. That is, the more workers that use manual skills intensively are concentrated in one place, the lower the wage. One explanation is that employers gain more bargain power over the workers, since manual skills are largely unskilled activities. Moreover, a large supply of low-skilled workers might depress those wages substantially.
wage distribution in 2010. The lower graph is useful to make sense of the distribution of skills above. It shows the employment shares of 1-digit occupation groups along the same wage distribution aggregated to quintiles. It is noteworthy that analytic and cognitive skill measures are almost parallel. As expected, workers in the bottom fifth of the wage distribution score the highest values in manual skills and exhibit low values for all other skills. These workers are mainly engaged in agriculture and in simple industrial activities. Coherently, at the upper end of the wage distribution, the highest values of analytic, cognitive and face-to-face skills are to be found. Mainly managers, researchers and people with a medium-skilled technical occupation are located in the top fifth of the wage distribution. Workers further to the left have clearly lower and almost monotonically falling values of analytic and cognitive skills.

Between the 2nd and 3rd quintiles, face-to-face skills are predominant, what comes from the fact that many sales jobs are located there. Many occupations in the manufacturing industry obtain higher earnings. The latter demand high manual skills but also require somewhat more analytical and cognitive skills than those professions with lower income in sales and agriculture, for example. Naturally, the workers in the industry rely less on interpersonal / face-to-face skills. All in all, the picture reflects the Brazilian peculiarities, such as a large agricultural sector, but it seems reasonable and in line with both the expectations and findings from, for example, Germany (cf. Spitz-Oener (2006)) or the US (cf. Autor et al. (2008)).

B Robustness

Another concern regarding our identification strategy may be that we are merely capturing a size effect of the region. We have shown that analytic, cognitive and face-to-face skills are concentrated in agglomerations. In general, the relative population size of regions is also quite stable over time. Besides, headquarters, universities and other institutions are traditionally located in large cities. Therefore, it could be that the development of modern industries was not directly affected by the concentration of manufacturing and liberal professions, but both could be correlated with the persistent population size of the region.

To tackle this concern, we include the historic population size of the region to the previous regressions. The estimations in table B.1 are now distinguished by the variables’ year of origin because in this way we avoid a regression with the population in 1872 and 1920, which are obviously highly collinear. Population size is a frequently used proxy for several observably equivalent agglomeration externalities (Duranton and Puga 2004). Thus this exercise also reveals whether the positive effect of the historical skill concentration is independent from other agglomeration economies.

Column (1) in table B.1 indicates that the population size in the past has indeed significantly affected the current concentration of interpersonal skills. This impression changes once the historic skill concentrations are added. As before, their effect clearly dominates
and the coefficients are quite similar to those in table 1. The $R^2$ and the F-value increase sharply once the share of industrial and liberal professions are added to the estimation in column (2). The same holds for the variables from 1872. Thus far we found that the composition of skills in the past does not merely reflect a size effect of the region but an independent and even more relevant channel.

Table B.1: Regressions with historical population size and GDP

| size variable: | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|
| 1920: population | | | | | | | | |
| log(size)     | 0.015*** | 0.001 | -0.079*** | -0.033** | 0.018*** | -0.000 | -0.054*** | -0.033*** |
| (0.005)       | (0.004) | (0.027) | (0.013) | (0.004) | (0.004) | (0.020) | (0.010) |
| log(ind. per 1000) | 0.032*** | -0.161** | 0.032*** | -0.083* | 0.053*** | -0.070* |
| (0.012)       | (0.012) | (0.071) | (0.046) | (0.013) | (0.013) | (0.036) |
| log(lib. per 1000) | 0.052*** | -0.104* | 0.053*** | -0.070* |
| (0.012)       | (0.060) | (0.046) | | |
| log(size) *   | 0.019*** | 0.014*** |
| (0.005)       | (0.005) |
| log(ind. per 1000) | 0.015*** | 0.014*** |
| (0.005)       | (0.004) |
| log(size) *   | | |
| (0.005)       | |
| log(lib. per 1000) | | |
| (0.005)       | |

$R^2$ 0.342 0.513 0.501 0.512 0.536 0.503 0.523
F-value 10.89 30.49 16.75 22.78 15.90 30.68 16.33 22.17
RMSE 0.082 0.071 0.071 0.071 0.080 0.071 0.071 0.070

Notes: Regressions in columns (1) to (4) include historic values of log population size whereas columns (5) to (8) include the GDP from 1920 as the size variable. The distribution of industrial or liberal professions is also from the year 1920. All regressions control for the distance to the sea and to the state capital, the railroad network in 1872 as well as for the concentration of slaves and foreigners in each AMC in 1872. Robust standard errors are in brackets. Regressions are weighted by population size in 2010. The number of observations is equal to 446 in all estimations. * denotes significance at ten, ** at five and *** at one percent level.

In columns (3) and (4) we include interaction terms between the population size of the region and the concentration of skills. Regarding the composition of the economy in 1872, we do not observe any significant effect. In 1920, however, the interaction between either of the liberal or industrial skill concentrations shows a positive and highly significant coefficient. At least when the industrial revolution began to gain pace in the 1920, the number of inhabitants in a region seems to have an important impact on the long-run development of industries. Note that despite the negative signs of the share of liberal and industrial professions in columns (3) and (4), their marginal effect (evaluated at the means) on the concentration of face-to-face skills continues to be positive confirming the results from table 2.

Columns (5) to (8) repeat the same estimations with our second historical measure of market size: GDP from 1920. Overall, the results are highly similar to the previous ones. We also experimented with log population density instead of log population size. Both variables are highly correlated and we obtained similar results throughout. Moreover, we prefer population size because the theoretical arguments given previously suggest that market size matters and density is rather a measure of concentration than of size.

A further robustness check in table B.2 includes the concentration of several other professions in 1872 and 1920 to the specification with full controls from the specification in column (6) of table 1. Following the reasoning for the definition of liberal and industrial
professions, we select and aggregate occupations to groups that are more or less comparable in both Censuses. This extension can be seen as a placebo test because we do not expect that unskilled occupations contribute to the regional development. The data show that a large concentration of unskilled and especially agricultural workers in the past hindered the subsequent development of regions, while a concentration in commerce, transport or extraction industries did not have any significant impact.

Table B.2: Regressions with alternative historical skills

|                  | (1)    | (2)    | (3)    | (4)    | (5)    | (6)    |
|------------------|--------|--------|--------|--------|--------|--------|
| 1872: log(ind. per 1000) | 0.046*** | 0.025* | 0.028** | 0.028** |        |        |
|                  | (0.014) | (0.012) | (0.013) | (0.011) |        |        |
| 1872: log(liberal per 1000) | 0.019 | -0.008 | 0.003 | -0.003 |        |        |
|                  | (0.016) | (0.008) | (0.011) | (0.007) |        |        |
| 1872: log(agric. per 1000) | -0.023 | -0.006 | -0.020 | -0.009 |        |        |
|                  | (0.015) | (0.011) | (0.013) | (0.010) |        |        |
| 1872: log(fishermen per 1000) | 0.008 | 0.005 | 0.010 | 0.001 |        |        |
|                  | (0.010) | (0.008) | (0.009) | (0.008) |        |        |
| 1872: log(unskilled per 1000) | -0.003 | -0.011** | -0.009* | -0.011** |        |        |
|                  | (0.007) | (0.004) | (0.005) | (0.004) |        |        |
| 1872: log(maritime per 1000) | -0.012 | 0.002 | -0.009 | 0.002 |        |        |
|                  | (0.018) | (0.007) | (0.010) | (0.006) |        |        |
| 1920: log(ind. per 1000) | 0.012 | 0.011 | 0.008 | 0.007 |        |        |
|                  | (0.016) | (0.017) | (0.013) | (0.013) |        |        |
| 1920: log(liberal per 1000) | 0.029* | 0.029* | 0.033** | 0.034** |        |        |
|                  | (0.016) | (0.015) | (0.014) | (0.015) |        |        |
| 1920: log(agric. per 1000) | -0.022* | -0.026** | -0.036*** | -0.036*** |        |        |
|                  | (0.013) | (0.012) | (0.017) | (0.016) |        |        |
| 1920: log(extraction per 1000) | -0.009 | -0.002 | -0.010* | -0.005 |        |        |
|                  | (0.006) | (0.007) | (0.005) | (0.006) |        |        |
| 1920: log(commerce per 1000) | 0.033*** | 0.030*** | 0.014 | 0.016 |        |        |
|                  | (0.012) | (0.011) | (0.012) | (0.012) |        |        |
| 1920: log(transp. occ. per 1000) | 0.016 | 0.007 | 0.018* | 0.012 |        |        |
|                  | (0.010) | (0.010) | (0.009) | (0.009) |        |        |
| R²               | 0.240  | 0.506  | 0.538  | 0.414  | 0.544  | 0.569  |
| F-value          | 6.928  | 35.48  | 34.91  | 17.82  | 21.64  | 37.51  |
| RMSE             | 0.088  | 0.071  | 0.069  | 0.078  | 0.069  | 0.067  |

Notes: Regressions also control for the share of slaves and foreigners in each AMC, as well as for the existence of a railroad network, GDP in 1920, the distance to the Sea and the state capital. Robust standard errors are in brackets. The number of observations is equal to 446 in all estimations. Regressions are weighted by population size in 2010. * denotes significance at ten, ** at five and *** at one percent level.

As an additional robustness check, we repeat the entire analysis but use the concentration of the three highest paid (i.e. most skilled) occupations instead of the concentration of skills. The three occupation groups we use are: managers, skilled technicians and scientists. The data show in all three cases that the concentration of these occupations is also positively related to the concentration of liberal and industrial occupations in the beginning of the 20th century. Positive interactions effects between occupation concentration and population size in the past are also observed analog to table B.1. This extension shows that neither our main analysis is severely biased by a measurement error in the skill variables, nor are the results driven by a specific occupation group such as managers, for example due to the concentration of headquarters in certain cities.

We test for further heterogeneity between regions and for spatial dependence in a final extension. As argued in section 2, there is a huge gap in the economic development between
the South/Southeast and North/Northeast of Brazil. Running our main regression separately for these two parts of the country generates a few additional insights. Columns (1) and (2) of table B.3 shows that both the concentration of industrial and liberal professions in the past are significantly related to the current distribution of interpersonal skills. However, the regression has a much better fit in the more developed region, i.e., the South. The circumstance that the industrial concentration is significant in 1872 in column (2) but only significant in 1920 for the Northern states suggests that the industrial take-off occurred earlier in the Southern part of the country. The results in columns (3) and (4) adjust our main regression for spatial dependence between regions using the two most common models: the spatial autoregressive (SAR) and the spatial error model (SEM), see Royuela and García (2015). Both models indicate that the spatial dependence is statistically important. The autoregressive specification has the better adjustment according to the AIC. Consequentially, the coefficients of liberal and industrial profession in the past are slightly lower than in the previous OLS regressions but despite controlling for the development of the neighbor regions, the region’s own endowment with those professions does not cease to be important.

Table B.3: Regressions with spatial heterogeneity and spatial dependence

| estimation: | OLS | OLS | SAR | SEM |
|-------------|-----|-----|-----|-----|
| sample:     | North | South | full | full |
| 1872: log(ind. per 1000) | -0.000 | 0.049*** | 0.01 | 0.016* |
|             | (0.014) | (0.013) | (0.008) | (0.008) |
| 1872: log(liberal per 1000) | 0.004 | -0.001 | 0.001 | -0.005 |
|             | (0.009) | (0.008) | (0.007) | (0.007) |
| 1920: log(ind. per 1000) | 0.037** | 0.016 | 0.014* | 0.019** |
|             | (0.016) | (0.013) | (0.008) | (0.009) |
| 1920: log(liberal per 1000) | 0.046** | 0.060*** | 0.023** | 0.030*** |
|             | (0.019) | (0.018) | (0.009) | (0.010) |
| Observations | 240 | 206 | 450 | 450 |
| R²          | 0.381 | 0.684 |
| F-value     | 8.583 | 55.05 |
| AIC         | -827 | -838 |
| Wald Test   | 94.2*** | 125.2*** |
| ρ           | 0.480*** |
| λ           | 0.545*** |

Notes: Regressions also control for the share of slaves and foreigners in each AMC, as well as for the existence of a railroad network, GDP in 1920, the distance to the Sea and the state capital. Robust standard errors are in brackets. Regressions are weighted by population size in 2010. * denotes significance at ten, ** at five and *** at one percent level.

2 The "South" sample includes the official regions South and Southeast while the "North" sample includes the Central-West, North and North-East of the Brazilian states. Note that whether the Central-Western states are included in the "North" or "South" sample makes no qualitative difference for the results.

3 The spatial weight matrix is built from the distances between states considering the neighbors according to the queen first-order type specification.
C The concentration of high-skilled occupations

As a final robustness check, we substitute the concentration of skills for the concentration of certain occupations. To be accurate, our measure of occupational concentration is defined as the log of the number of people in a certain occupation per 1000 workers within each region. As explained previously on the basis of figure A.1 in section 3.3, those occupation groups with the highest remuneration are appropriate substitutes because they also exhibit the highest scores in face-to-face and analytical skills. These three occupational groups are: managers & directors, scientists and medium-skilled technicians.

Figure C.1 illustrates the spatial distribution of these jobs. Unequivocally, managers, scientists and medium-skilled technicians are also most frequently encountered in heavily populated regions. In this representation, the concentrations of these professions also vary quite substantially – between 1 and 267 per 1,000 inhabitants in an AMC. Continuing the examples in section 4.1, the municipality Areias registers 17 managers and 3 skilled technicians per 1,000 inhabitants whereas these statistics are equal to 116 and 86 in Florianópolis.

The information on occupations does not emerge from a worker survey, as skill measures do. In fact, the concentration of occupations is a simple and objective count that is directly derived from official data. Our skill measures may possibly be exposed to criticism because we import them from a US workforce survey. The occupation concentration only contains measurement error to the extent that workers’ jobs are misclassified despite the utilization of the most disaggregated occupation classification. Only if such erroneous entries do not occur at random across regions – and we have no reason to presume that – does this pose a threat to our identification. Still, in our view, skills are a preferable measure because skills provide a more intuitive understanding as to why positive externalities emerge.

For the sake of space, we move directly to the estimation of the main regressions, in analogy to those in the main text. Table C.1 shows that for the concentration of managers, science & teaching related occupations, and skilled technicians, there is also a positive and highly significant relation to the skill concentration in the past. In particular, the concentration of liberal dominates over the over three skill variables and is the only coefficient with statistical significance in columns (1) to (3). In univariate regressions, at least the concentration of both liberal and industrial skills in 1920 are significant in all three cases. For the science related occupations – which is most directly related in the generation of knowledge – the $R^2$, F-value and the estimated coefficient itself indicate that the strongest relation to the supply of localized knowledge related professions in 1920. Columns (4) to (7) show that the current distribution of managers and scientists exhibits a positive interaction between the population size and the concentration of liberals and industrials in 1920. For the technical occupations (not shown in table C.1), no significant interaction terms is obtained. In sum, the fact that results similar to those in the main text are obtained throughout this robustness check clearly indicates that observed development is not tied to one specific occupation group or the provision of certain institutions like headquarters or research institutions but rather to skills that are common across the three occupation groups.
Table C.1: Regressions with occupation concentrations

| occupation:          | (1)       | (2)       | (3)       | (4)       | (5)       | (6)       | (7)       |
|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| occupation:          | managers  | science   | technical | managers  | science   |           |           |
| 1872: log(ind. per 1000) | -0.008    | 0.039     | 0.004     |           |           |           |           |
|                      | (0.049)   | (0.078)   | (0.062)   |           |           |           |           |
| 1872: log(liberal per 1000) | -0.025    | -0.017    | -0.047    |           |           |           |           |
|                      | (0.041)   | (0.064)   | (0.050)   |           |           |           |           |
| 1920: log(ind. per 1000) | -0.091**  | -0.020    | 0.006     | -0.182    | -0.650**  |           |           |
|                      | (0.045)   | (0.068)   | (0.059)   | (0.217)   | (0.319)   |           |           |
| 1920: log(liberal per 1000) | 0.397***  | 0.591***  | 0.364***  | 0.405**   | 0.066     |           |           |
|                      | (0.049)   | (0.067)   | (0.061)   | (0.204)   | (0.266)   |           |           |
| 1920: log(population) | -0.115    | -0.004    | -0.345*** | -0.112*   |           |           |           |
|                      | (0.089)   | (0.048)   | (0.131)   | (0.060)   |           |           |           |
| 1920: log(population)* | 0.028     | 0.085***  |           |           |           |           |           |
|                      | (0.018)   |           |           |           |           |           |           |
| 1920: log(population)* |          | -0.006    |           | 0.043**   |           |           |           |
|                      |           |           |           | (0.017)   |           |           |           |

\( R^2 \) \hspace{1cm} F-value \hspace{1cm} RMSE

\( 0.534 \hspace{1cm} 37.93 \hspace{1cm} 0.253 \)

\( 0.684 \hspace{1cm} 92.51 \hspace{1cm} 0.366 \)

\( 0.586 \hspace{1cm} 34.82 \hspace{1cm} 0.309 \)

\( 0.361 \hspace{1cm} 24.30 \hspace{1cm} 0.297 \)

\( 0.522 \hspace{1cm} 37.85 \hspace{1cm} 0.256 \)

\( 0.570 \hspace{1cm} 65.37 \hspace{1cm} 0.426 \)

\( 0.687 \hspace{1cm} 114.6 \hspace{1cm} 0.363 \)

Notes: Regressions also control for the share of slaves and foreigners in each AMC, as well as for the existence of a railroad network, the distance to the Sea and the state capital. Robust standard errors are in brackets. The number of observations is equal to 446 in all estimations. Regressions are weighted by population size in 2010. * denotes significance at ten, ** at five and *** at one percent level.
Figure A.1: Skills and occupations along the wage distribution

Notes: The upper graph shows the standardized skill scores based on the O*NET along the wage distribution for the entire working population interviewed for the Brazilian Census in 2010 using a kernel weighted regression. The lower graph aggregates the wage percentiles to quintiles and shows the share of workers in each occupation group (1-digit classification).
Figure C.1: Spatial concentration of occupations – AMC means

Notes: The circles in each graph represent the AMCs’ concentration in each of the three highest skilled occupation (1-digit) groups. The results from the corresponding linear regression are reported below each graph.
D Additional tables and figures

Figure D.1: Spatial delineation of AMCs

Notes: The map shows how the municipalities in their current definition from 2010 are aggregated to the consistent regions (AMCs) that remain stable over the period 1872-2010. The borders of municipalities are represented by the thin lines whereas the bold lines indicate the borders of AMCs.

Table D.1: Summary statistics

| variable                  | mean  | std. dev. | min.  | max.  | source   |
|---------------------------|-------|-----------|-------|-------|----------|
| 2010: analytical skill conc. | 9.407 | 0.105     | 9.127 | 9.897 | O*NET    |
| 2010: face-to-face skill conc. | 9.779 | 0.111     | 9.458 | 10.097| O*NET    |
| 2010: cognitive skill conc.  | 9.554 | 0.095     | 9.294 | 9.814 | O*NET    |
| 2010: manual skill conc.    | 10.098| 0.139     | 9.575 | 10.574| O*NET    |
| 1872: log(ind. per 1000)    | 2.325 | 0.698     | 0.532 | 4.482 | census   |
| 1872: log(liberal per 1000) | 1.609 | 0.806     | -1.478| 3.976 | census   |
| 1920: log(ind. per 1000)    | 3.732 | 0.682     | 1.467 | 5.679 | census   |
| 1920: log(liberal per 1000) | 1.696 | 0.642     | -0.196| 3.521 | census   |
| 1872: log(population)       | 9.568 | 0.89      | 7.356 | 12.524| census   |
| 1920: log(population)       | 10.552| 0.98      | 8.349 | 13.953| census   |
| 2010: log(population)       | 11.931| 1.278     | 8.215 | 16.642| census   |
| distance to sea            | 1.763 | 2.365     | 0     | 19.937| own calc.|
| 1920: railroad dummy       | 0.578 | 0.494     | 0     | 1     | ipeadata |
| distance to state capital   | 1.847 | 1.5       | 0     | 8.324 | ipeadata |
| 1872: log(slaves per 1000)  | 4.73  | 0.752     | 1.789 | 6.353 | census   |
| 1872: log(foreigners per 1000) | 1.592 | 1.568   | -2.79 | 6.214 | census   |
| 1920: log(GDP)             | 9.112 | 1.275     | 5.675 | 14.14 | ipeadata |

Notes: The table shows the mean, standard deviation, minimum and maximum in the sample for the most important variables in this paper. The last column indicates the variables' data source. The number of observations is equal to 446 for all variables.
Figure D.2: Correlation between analytical skill mean and historical trades

Notes: The circles in each graph represent the AMCs’ analytical skill average and each one of the four instrumental variables in the sample. The results from a weighted linear regression of each of the IVs on the analytical skill concentration are indicated below each graph.

References

ACEMOGLU, D. and AUTOR, D. H. (2011). Skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics, vol. 4, Elsevier – North Holland, pp. 1043–1171.

AUTOR, D. H., KATZ, L. F. and KEARNEY, M. S. (2008). Trends in US wage inequality: Revising the revisionists. Review of Economics and Statistics, 90 (2), 300–323.

—, LEVY, F. and MURNANE, R. J. (2003). The skill content of recent technological change: An empirical investigation. Quarterly Journal of Economics, 118 (4), 1279–1333.

BACOLOD, M., BLUM, B. S. and STRANGE, W. C. (2009). Skills in the city. Journal of Urban Economics, 65 (2), 136–153.

DURANTON, G. and PUGA, D. (2004). Micro-foundations of urban agglomeration economies. In V. Henderson and J. F. Thisse (eds.), Handbook of Regional and Urban Economics, vol. 4, Elsevier – North Holland, pp. 2063–2117.

EHRL, P. and MONASTERIO, L. M. (2017). Skills, agglomeration economies and historical settlement conditions. mimeo.

FIRPO, S., FORTIN, N. M. and LEMIEUX, T. (2011). Occupational tasks and changes in the wage structure. IZA Discussion Paper No. 5542.
Figure D.3: Correlation between cognitive skill mean and historical trades

Notes: The circles in each graph represent the AMCs’ cognitive skill average and each one of the four instrumental variables in the sample. The results from a weighted linear regression of each of the IVs on the analytical skill concentration are indicated below each graph.

Maciente, A. N. (2013). The determinants of agglomeration in Brazil. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Michaels, G., Rauch, F. and Redding, S. J. (forthcoming). Task specialization in US cities from 1880–2000. Journal of the European Economic Association, 18715.

Royuela, V. and García, G. A. (2015). Economic and social convergence in Colombia. Regional Studies, 49 (2), 219–239.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. Journal of Labor Economics, 24 (2), 235–270.