XII Congreso de la Asociación Española de Historia Económica
6th to 9th September 2017, Salamanca

¿Es Persistente el Capital Social? Medición Comparativa entre los Siglos XIX y XX

Is Social Capital Persistent? Comparative Measurement in the Nineteenth and Twentieth Centuries[1]

Marta Felis-Rota[2]
(Universidad Autónoma de Madrid)

RESUMEN

Recientemente, ha habido un interés creciente por el capital social, así como por las dificultades asociadas a su medición. Este artículo propone medir el capital social a través de un análisis de componentes principales y presenta los primeros indicadores internacionales de capital social para el siglo XIX. El análisis está basado en una base de datos internacional que data del siglo XIX y contiene un amplio rango de variables socio-económicas. Se han construido indicadores de capital social para los años 1870 y 1890. Hay que destacar que estos indicadores son compatibles con indicadores sociales de mediados del siglo XX, lo que facilita el estudio de la evolución histórica del capital social entre los siglos XIX y XX. A muy largo plazo, podemos encontrar una caída significativa de la posición relativa de los países europeos y los Estados Unidos.

ABSTRACT

Recently, there has been a growing interest in social capital and in the difficulties related to its measurement. This paper proposes to measure social capital by means of

[1] The first version of this paper was published as LSE Economic History working paper No. 103/07. The last version was chapter 3 of the PhD thesis of the author, available in full detail from the London School of Economics and Political Science, London, UK.

[2] Marta Felis Rota, Department of Economic Analysis: Economic Theory and Economic History, Av. Tomás y Valiente, 5, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain. E-mail: marta.felis@uam.es

principal components analysis and presents the first available international social capital estimates for the nineteenth century. The analysis is based on a nineteenth-century international database containing a wide range of socio-economic variables. Social capital indicators are constructed for the years 1870 and 1890. Interestingly enough, these indicators are comparable to mid-twentieth century social indicators. This facilitates the study of the evolution of social capital between the nineteenth and twentieth centuries. In the very long run, one can find a significant decline in the relative position of the European countries and the United States.

Palabras clave: capital social, confianza
Códigos JEL: Z13, N01, O11

Keywords: social capital, trust
JEL classification codes: Z13, N01, O11

I. INTRODUCTION

The term social capital was coined some time ago. In a phrase, social capital is the quality of the civil society, or the civic values (see Putnam, 1993 and 2000). It should interest economists because it is indeed a form of capital. The importance of social capital for economic performance continues to be underlined in economic research (Ostrom, 2011). Felis-Rota (2011) provided a definition of social capital and embedded it in the economics literature in full detail. It is especially important to revisit this topic since trust in public institutions has declined over recent years (Stevenson and Wolfers, 2011).

Quite a lot of research articles deal with lab experimental economics in order to report levels of trust and its economic implications (Bohnet and Huck, 2004; Zak et al., 2005; Hargreaves Heap and Zizzo, 2009; Eckel and Petrie, 2011). Some prove that trust acts as an incentive or motivation to be more productive (Falk and Kosfeld, 2006; Sliwka, 2007). Others are based on rural areas experiments in the field of development economics (Schechter, 2007; Jackson, Rodriguez-Barraquer, and Tan, 2012; Moscona, Nunn, and Robinson, 2017). However, when it comes to history, we need to trust either in natural experiments (Algan and Cahuc, 2010; Nunn and Wantchekon, 2011; Cagé and Rueda, 2016) or make inferences from historical records, which is what I do in this article. As Tabellini (2007) and Grosjean (2011) showed, historical social trust matters. The great difficulty is accounting for it.

Social capital has been shown to have had an effect in economic performance for the recent decades, but there are some indications that it could also have been relevant in the distant past. In his investigation for the European regions, Tabellini (2007) concludes that economic outcomes have their root in historically fundamented attitudes towards trust, respect for others and self-determination. Nevertheless, it is hard to device quantitative tests.

A stream in the economics literature argues that history affects economic performance through institutions. Perhaps the most well-known paper is that by Acemoglu, Jonhson, and Robinson (2001), where they argue that different colonial origins led to different sets of institutions, and the latter gave rise to different economic outcomes. But, what happens when we look at colonisers instead of colonised countries? Is there nothing to say? Guido Tabellini (2007) conducts an empirical investigation on differential economic performance in the European regions, and links this fact to deep cultural roots, based on trust, respect and confidence built up centuries ago. Also along this line, Grosjean (2011) surveys the contribution to current cultural heterogeneity of the historical division between the Ottoman, Habsburg, Russian or Prussian Empires. Both these branches --the colonial origins literature and the historical cultural differences whithin Europe-- survey two separated parts of the world. On the one hand, the colonisers acquiring their culture through centuries of social interaction; on the other hand, the colonised, inheriting the institutional bequest transferred by the colonisers. Now, is it possible to conduct a historical investigation with a sample of countries representing all regions of the world? And, more importantly, can one reach a generalised conclusion of whether culture and institutions are deterministic or not so when it comes to economic performance?

This paper proposes an international index for social capital in the nineteenth century –very close to Tabellini's definition of culture–, which comprises a varied sample of countries around the world, including both colonised and colonisers. Moreover, it compares the new (referring to the past) to the existing (more recent) indicators and assesses the viability and performance of the different measurement alternatives.

How does social capital evolve over time and across countries? Is it path dependent? Adding some time dimension to the study of social capital looks promising. Section II compares three different proposed measures of social capital for the second half of the twentieth century, and looks at the relationships between them. Section III turns the attention into historical data and presents a new social capital index for the late nineteenth century. In Section IV, we are able to look at the inter-temporal evolution of social capital thanks to the newly created index, together with the more recent measurement alternatives. Finally, section V succinctly concludes.

II. COMPARISON OF CONTEMPORARY ALTERNATIVES

Of all the possible ways to try to measure social capital, several empirical proposals are available, but yet none enjoys general acceptance. Therefore, it is convenient to start with a comparison amongst the most popular measuring alternatives. The comparison might turn into an interesting exercise, since it will make us win some perspective on the alternatives plus reveal some of the insights that remain unnoticed until the present moment. This section compares three different twentieth century measurement alternatives, based on the pre-existing measurement attempts. These three are TRUST, CIVIC, and SOCDEV, standing for level of trust in a society, civic engagement, and social development respectively.

Both TRUST and CIVIC have been originally extracted from the World Value Surveys, which periodically runs over a whole range of countries over the world. General trust in people (TRUST) is the percentage of respondents who answered 'yes' to the following question: 'Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?' I amplified TRUST to TRUSTAM by adding extra country data from both the most recent and past rounds of the World Value Surveys. Civic engagement (CIVIC) is the percentage of civic activities in which and average individual participates. The activities included are: social-welfare services for elderly and deprived; education, art, and cultural activities; local community affairs; conservation, environment, ecology; and voluntary associations for health (La Porta et al., 1997). Finally, social development (SOCDEV) was taken from Adelman and Morris (1967). The index is an extraction of factor scores from a principal components analysis including 41 socio-economic variables from 74 countries around the world, for the period 1957-62. Temple and Johnson (1998) used this index before in order to test the economic significance of social arrangements.

With respect to the timing of the measures, the first two correspond to the late twentieth century and have been taken from Knack and Keefer (1997), (hence KK appears in some instances attached to the variable name). Overall, the General Value Survey rounds for 1980, 1990-1 and 1995-6 are included, corresponding to the last decades of the twentieth century, the closest available to the alternative measure[3]. As an average, we can say the two variables are aimed at monitoring social capabilities at the end of the twentieth century. SOCDEV corresponds to the early 1960's, –this is, twenty to thirty-five years earlier–, so we will need to keep this in mind.

[3] The next round of Surveys dating 2002 and beyond are not included because it is beyond the time framework. Comparison of alternatives too far in time would not make sense.

Table 1[4] is a compendium of the data availability for three variables. We can observe that data availability is limited, so the comparisons amongst variables are forced to be restricted to a smaller sample of countries.

These are three different ways to measure social capabilities that have been proposed in the literature. They are conceptually different from each other and may or may not be related. The correlation matrix shows that their relationship, if any, is not always linear (table 2). Later we will find <u>non-linear</u> relationships between them.

In order to investigate deeper the relationships between the three variables, we make use of graphical representations. Only 9 complete cases are available for all three variables. This is due to the fact that the variables come from different sources and were not thought to match and be studied together. A scatter plot can help us to position these cases in the three-dimensional space. Figure 1 is a joint three-dimensional graphical representation of the three proposed indicators. SOCDEV stands for the Adelman and Morris social development index and is the variable positioned in one of the axis. CIVICKK stands for civic engagement as reported by Knack and Keefer (1997) and has been placed in the second axis. Finally, TRUSTAM stands for trust amplified as described above and is the variable positioned in the third axis. Every dot in the three dimensional space reveals the position of the indicated country with respect to the three indicators. For best visualisation on the three-dimensional space, we include two scatter plots representing two different perspectives on the same data; –one with spikes to the floor and a second one with centroid perspective (spikes to the centre of the data) –. The three-dimensional graphical representation offers an overall picture of the data, which reveals an elliptic shape.

Since we are especially interested in historical considerations, the focus of the paper is driven by past-present contrasts. Nonetheless, comparing both contemporary measures for social capital, TRUST and CIVIC, is not of least interest, since they stand for different concepts. This is done by overlapping throughout the paper two plots in one. In the figures 2 and 3 (within twentieth century comparisons), the light coloured dots and lines depict the pair CIVIC *versus* SOCDEV, while the dark dots and lines represent TRUST *versus* SOCDEV. In short, the scatter plots below should be read in the following way: Every graph is composed of two overlapping bi-dimensional scatter

[4] Because of the large amount of calculations involved in this paper, all tables are compiled at the end.

plots, with the historical index in the horizontal axis and the contemporary index in the vertical axis. SOCDEV is common for both overlapping plots and is always positioned in the horizontal axis. It represents the historical measure of social capital. CIVIC and TRUST are always placed in the vertical axis, representing the contemporary measure of social capital. In this way, we can read all the graphs as a historical evolution of social capital, by looking at where countries were positioned in the 1960's (horizontal axis) and where they were positioned in the 1990's (vertical axis).[5]

Figures 2 and 3 represent the historical evolution of social capital in the second half of the twentieth century. SOCDEV corresponds to the years around 1960, while TRUST and CIVIC capture roughly the last two decades of the twentieth century. The two figures are based on the same data, and differ only on visual aids. The first one draws spikes from every country to the mean of the contemporary variable. Both variables CIVIC and TRUST have been standardised and thus vary within the same range. It is particularly interesting to observe where the mean of these two falls. We can observe that the mean of CIVIC is higher than the mean of TRUST. This fact can be due to the formulation of the questionnaire. But we should recall and keep in mind that they do not measure exactly the same concept: One is an index of voluntary participation and the other a percentage of 'yes' or 'no' answers regarding the general level of trust in a country. So there should be room for disparity. Nevertheless, it is still interesting to wonder whether there is a reason beyond formulation of the questionnaire and conceptual disparity behind the differing means. Later in this paper we argue that this is the case.

In figure 3 a line has been fitted to the points using a non-parametric technique called 'lowess' (locally weighted linear regression). This method fits the maximum number of points with the minimum number of iterations. Fifty percent of the points have been fitted with only three iterations. This type of graph is very appealing because it reveals the outliers. For the sake of historical findings, the engagingness of this exercise lays more on unmasking the outliers than on the fitted points that stand on the average. The impossibility of the fit line to match all dots points at the exceptional evolutions (both for the good and for the bad).

[5] STD at the end of the name of the variable means that the variable has been standardised. The variables which do not contain STD at the end of their name were already constructed in a way which allows for comparison.

Striking results are those of India and Japan. They reveal themselves as outstanding performers in social improvement, which is historically consistent with their growth experiences. We can also detect failure stories by looking at the extremely poor contemporary scores compared to the mid-century scores for some Latin American countries like Mexico or Venezuela. Indeed, from figures 1 to 3 we can observe how some countries strikingly detach from the average, defeating the path dependence argument postulated by North. Having said this, the path dependence hypothesis is not refuted but modified. This is saying that socially well-endowed countries do actually leap over the development gap.

Should we have time series information about social evolution, we would be able to determine the timing of the social change: before, during, or after economic growth. As discussed in previous sections, Putnam argues that social change happens up to 70 years ahead of subsequent economic growth. Therefore, we need to go back further into history of social development to be able to contrast this observation. This is done in the next section.

III. CONSTRUCTING A SOCIAL CAPITAL INDEX FOR THE NINETEENTH CENTURY

Constructing a social capital measure for the distant past presents several challenges. The first challenge is the limitations of data availability. Once we turn into the distant past (more than a few decades ago) no surveys can be conducted and one has to rely on data already collected for other purposes. The second challenge, which is going to be addressed now, is to find a quantitative methodology that is flexible enough given the data limitations, but still conveys informative results. For both challenges, one needs to be truly imaginative and make the most out of the resources. Finally, how are the resulting data going to be compared to more recent data? The third challenge is to construct a measure that, at the same time, can be compared –even if imperfectly– to some existing indicator.

III. 1. SOURCES: NEW ADELMAN AND MORRIS DATABASE

Contemporary indicators of social capital based on the World Value Surveys are informative. Yet we need a wider time span in order to bring historical perspective into the analysis of social capital. Having pre-First World War social capital estimations would provide useful historical insights in order to study its evolution and test its persistence.

It is possible to find historical data to fill in the blanks on existing work and give a time dimension to the social capital analysis. At this respect, Adelman and Morris (1988) provide an extensive socio-economic database for the period 1850 to 1914.

The comprehensive nineteenth century series provided by Adelman and Morris are the starting point for our database. The extensive data appendix accompanying their 1988 book is a summary of the work on their data over more than 20 years. It contains cross-sectional data for 23 countries scattered over the globe and referring to 35 summary variables ranging from attitudes to change to political perceptions[6]. The latter depict the socio-economic structure of every country in the sample between 1850 and 1914, being this divided into 3 sub-periods: 1850-1870, 1870-1890, and 1890-1914. Cross-sectional data are supplied for every sub-period. The variables in levels and proportions refer to the initial level of each period, while those capturing change or characteristics refer to the whole of the preceding 20 year period.

The Adelman and Morris database has unique characteristics of which an economist looking for social influences in historical perspective can certainly take advantage of. These are: The database describes the situation of the economy in the late nineteenth century in conjunction with a detailed picture of the institutional framework, and some interesting social attitudes and customs in different countries. This highly valuable database has been explored under its possibilities.

This research reconstructs a similar database to that which Adelman and Morris built for the period 1850 to 1914, and then uses it to construct a Social Development Index (SDI) for the nineteenth century. The variables are extracted from their 1988 book. They are re-codified for convenience but the integrity of the database is preserved. Some variables available from Adelman and Morris (1988), their previous publications, and other posterior sources have been omitted because the alternative variables covering the same concept are preferable in terms of country classification and overall consistency of the database. Re-codification consisted of transferring letter codification (alphabetic

[6] A list of countries can be found in appendix A; a detailed list of variables can be found in appendix B.

order of categories) into numeric codification (categories sorted by ordinal numbers). This turned alphabetically coded variables into numerically coded variables, suitable for the intended statistical analysis. In addition, almost half of the variables have been renovated or updated. [7] Then we performed a principal components analysis.

III. 2. QUANTITATIVE METHODOLOGY

Why do I think that the principal components must convey information about the quality of the society? One could have gathered a bunch of variables which I think describe the quality of a society and calculate a simple average. However, not all variables contribute in the same measure to explain the dispersion in the data; some do capture more variance than others. The principal components analysis (PCA) is a sophisticated weighted average that gives more importance to those variables that deserve it in terms of variance explained. If there is something that makes the countries genuinely different, the principal components analysis will capture it.

In short, the principal components analysis is a data-reduction technique. It aims at giving a description of the relationships between a set of variables in terms of a smaller set of *linear combinations* of these variables. These linear combinations are called Principal Components. The extent to which the relationships between the variables can be adequately described by a *small* set of new variables called Principal Components depends on the correlations between the original variables. The higher the correlation between the original variables, the smaller the number of Principal Components and, thus, the most effective the data reduction is.[8]

The latter characteristic of the Principal Component Analysis will turn out to be very helpful to us. It is well-known that in the social sciences – and especially in economics - many variables are highly correlated. Economic time-series tend to move together, and because of this it is often difficult to separate the effect that one variable has on another. However, here we are concerned with the parsimonious description of a high-dimensional object (many variables) into a small-dimensional one (one index), and as a result the Principal Components Analysis actually *takes advantage* of these high

[7] Please, refer to Chapter 2 of the PhD thesis of Marta Felis-Rota for a detailed discussion of the Adelman and Morris (1988) database, available from the London School of Economics and Political Science, London, UK.
[8] An intuitive explanation of the principal components technique can be found in appendix D.

correlations. It is for this reason that we consider this technique particularly adequate for our purposes.

But the principal components analysis is not merely a data reduction exercise. The principal components are the underlying factors behind the variables, those factors that make them move together (covariate), and cannot be captured in any other way than in the abstract. The variables are just the reflection or result of those underlying factors that make them move together. And social capital is precisely this glue. This is why I think that a set of variables describing a society can help us capture the social capital behind.

III. 3. PRE-PCA TESTS: OPTIMISING THE VARIABLES

Prior to the principal components analysis (PCA) of the data, there are some statistical preliminary tests that are convenient to run. The first one of these preliminary tests is the Bartlett's test of sphericity. This test checks the validity of the whole exercise. Secondly, one might want to check for the validity of the inclusion of individual variables. There are several ways to test the adequacy variable-by-variable. These are the square multiple correlation coefficient or R-squared of individual variables, a test of simple partial correlations, and, finally, Kaiser´s measure of sampling adequacy[9]. This section goes through all these tests, prior to the principal components analysis.

The first preliminary test is the Bartlett's test of sphericity (Bartlett, 1950). This is a basic test of the significance of the principal components analysis[10]. It checks that the correlations matrix of all variables to be included in the analysis is not an identity matrix. Should that be the case, it would make no sense to run a principal components analysis because this type of analysis is only adequate for variables that are interrelated. The Bartlett's test of sphericity tests the null hypothesis that the correlation matrix is an identity matrix (all zeros except ones in the main diagonal) against the alternative hypothesis that the variables show statistically significant correlations amongst each other. If the null hypothesis is rejected, then the database has passed the first hurdle.

[9] These series of preliminary tests have been suggested by Wuensch (2005).
[10] Actually, the Bartlett's test of sphericity is designed to test the significance of any factor analysis, regardless of the extraction method.

Given the socio-economic character of the variables in the database, one would be very surprised if the database would not pass this test. The principal components analysis has been chosen precisely because of the adequacy of this type of analysis to strongly interrelated variables.

The variables referring to the rate of adoption of new technologies in industry and agriculture respectively are perfectly correlated. The latter has been removed from the analysis in order to avoid multicolinearity problems. Therefore, there is one unique variable referring to the rate of adoption of new technologies, and this corresponds to both industry and agriculture.

As far as the adequacy of variables is concerned, all variables are in the database because there is a theoretical justification to include them. They are all related to or helping to reveal the quality of society in one way or another. However, statistical methods offer the possibility to put all variables through tougher tests of adequacy. First, the partial correlation between two variables is the correlation between the deviations from the mean. The simple partial correlations test consists of calculating partial correlations between all pairs of variables and checking whether these are strong enough for us to think they are contributing to the same underlying factors or components. Second, the squared multiple correlations (SMC) test calculates the R-squared of every variable with all the rest, and, therefore, is a more elaborated measure of adequacy variable-by-variable. For this reason, the first test of simple partial correlations is omitted. Still more sophisticated is the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO). This measure is calculated as 1 minus the ratio between the sum of squares of all pair wise correlations (pair wise variances) and the sum of squares of all partial correlations of that variable with the rest[11]. This ratio is close to 1 when the pair wise correlations with all variables are high and the partial correlations are small, and close to 0 if the pair wise correlations with all other variables are small and the partial correlations high. It can be calculated both for individual variables and overall. For maximum representativeness in a principal components analysis one would want the KMO being close to 1, meaning that direct correlations with other variables are much more important than correlations of deviations from the mean. This is a way to

[11] $KMO = 1 - \frac{\sum_{k \neq i} r_{ik}^2}{\sum_{k \neq i} pr_{ik}^2}$, where r is the correlation coefficient and pr is the partial correlation coefficient.

measure the proportion of total variance explained by a particular variable with respect to the average. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) is used as the main criterion of inclusion/exclusion of variables, although the SMC or R-squared is also calculated for reassurance/illustration purposes.

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy can theoretically take values from minus infinity to 1. In practice, very rarely takes negative values, so it is regarded as running from 0 to 1. Small values mean that overall the variables have too little in common to warrant a good factor analysis. Heuristically, the following labels are given to values of KMO:

Less than 0.49 unacceptable
0.50 to 0.59 miserable
0.60 to 0.69 mediocre
0.70 to 0.79 middling
0.80 to 0.89 meritorious
0.90 to 1.00 marvelous,

as classified in the STATA User’s Manual, based on Kaiser. This measure of sampling adequacy improves as the number of variables increases, the number of effective components decreases, the number of individuals increases, and the general level of correlation increases (Kaiser, 1970:405).

The KMO test is positive overall. The global rating for the adequacy of the sample of variables is 0.72 in a scale of 0 to 1. The KMO test also gives a mark for each one of the variables. A mark lower than 0.5 is suggesting the corresponding variable to be removed from the analysis. From the 32 variables in the first stage principal components analysis 4 are returning a mark lower than 0.5: Export growth group, population growth group, immigration group, and land concentration. Then, a second principal components analysis is performed.

In the second round (removing the 4 irrelevant variables) the overall KMO rate has improved from 0.72 to 0.77, and many more variables seem to fit better in this dimension. Also, 2 of them have become irrelevant: population per farm and population group. This is an invitation to perform a third principal components analysis round removing those two additional variables.

The third round with only 26 of the initial 32 variables is giving acceptable KMO marks for all variables. Therefore, it is the moment to stop. In addition, the overall KMO mark remains at 0.77 (same as last round). In total, we have removed 6 variables from the initial set plus 1 for being redundant. All the rest have been proven to have an empirical on top of theoretical justification to be there. It is interesting to note that the first principal component, characterised by the highest ranked variables, is becoming increasingly more important in terms of total variance captured as successive corrected rounds are performed. In the first round, the first principal component was capturing almost 45 percent of the variation in the data, in the second round it was capturing more than 50 percent, and in the third round it captures 54 percent.[12] The relative importance of the first principal component with respect to the rest is explored in more detail in the next section of this chapter.

As a robustness test for the list of variables eventually included in the PCA let us alter the KMO selection criteria. Before, it has been set at a minimum of 0.5 variable-by-variable, following the classification above. Let us move this barrier up to the next level, 0.6 mark or above variable-by-variable. This is a more strict selection criterion, and it is a standard one. Following this alternative criterion suggests dispensing with some extra variables after the first round: income growth, population per farmland, industrial wage change, agricultural wage change, population group, and form of land tenure, all of these scoring between 0.5 and 0.6, the latter not inclusive, [0.5, 0.6). Following this stricter criterion leaves a satisfying list of variables already in the second round. This has raised the KMO overall mark from 0.77 to 0.87, while leaving the significance level of the Barlett's sphericity test practically untouched. Thus, we shall stick to this criterion; all this keeping in mind that the overall KMO test is returning a global measure above 0.7 in all instances for the whole group of variables considered together. So, the successive rounds of variable-by-variable KMO test should be taken as a refinement of the selection exercise. As a result of the improved suitability, 22 variables are left from the original 32. The total variance explained by the first principal component alone rises from the initial 45 to almost 60 percent thanks to the optimisation of the choice of variables [13].

[12] Exact values are shown in the tables section at the end of the paper.
[13] See tables at the end of the paper.

The new database preserves the spirit of the original Adelman and Morris database but has been renovated and updated with the information available up to date. Therefore, it should be the preferred choice. The initial count of variables available for inclusion in the PCA goes up to 35, thanks to the updating work. A total of 14 variables are new, most of them substituting old ones. The same preliminary tests have now been run for the new database.

As before, the rate of adoption of new technologies in agriculture needs to be removed because of perfect correlation with the rate of adoption of new technologies in industry. Leaving just one of these two, no matter which one, will represent the rate of adoption of new technologies in general. This leaves the initial analysis of the new data with 34 variables to start with.

The old variables for economic role of the government (govt) and representativeness of political institutions (represent) have a clearly superior performance than the new ones (govrel and polity2) in being aligned with the components. This fact is an invitation for the reincorporation of previous alternatives. Let us first do the second round with the suppression of these two as the KMO test suggests. Then, we will move to reincorporation of old adequate variables. In both cases, for the inclusion in the second round, the same restrictive criterion as above has been followed: All variables with a KMO mark below 0.6 have been suppressed.

The final result is the second round with the new database, having replaced govrel and polity2 by their old corresponding variables, govt and represent, due to the fact that they seem to perform better than the new ones. An initial survey has been made with the 34 initial variables, just replacing back these two. Then, all those scoring below 0.6 in the KMO test have been removed. This gives rise to the final selection of variables.

Results are indeed the best in terms of adequacy of the variables that contribute to explain the principal components. This is reflected in the fact that the overall KMO is above 0.9, which corresponds to the highest target for the KMO test, and described as "marvellous". Actually, once the less relevant variables have been suppressed, all 18 eventually included variables have an individual KMO mark around 0.9, as seen in the tables shown at the end of the chapter. This reflects the adequacy of the variables in capturing the underlying principal components. Actually, this selection of variables happens to return also the highest percentage of variance explained by the first principal component alone (above 65 percent), which reinforces the intuitive idea that a summary variable unlocked by the principal components is more than feasible. The principal components selection is explored in more detail in the next section.

Both datasets, an adaptation of the original by Morris and Adelman (1988) and the renewed one used here, have been put to test for best PCA performance potential. Preliminary tests for optimal selection of variables reveal that the new dataset has a higher PCA potential, once cleaned up. However, some of the old variables have been recovered in the final selection, because they seem to capture elements not present in the new data. In all cases there are no problems of insufficient correlation amongst the variables, as shown by the Barlett's sphericity test.

III. 4. RESULTS: PRINCIPAL COMPONENTS ANALYSIS

The principal components are newly generated variables, obtained from linear combinations of the original. The first component extracts the linear combination of the variables that contains most of the variance. Successive orthogonal components explain the variance that is left. Because every variable entering the analysis is startardised to have variance equal to 1, one should be interested in those outcoming principal components that have an eigenvalue greater than 1. This means they can explain more variance than any of the single variables can; the greater the eigenvalue, the larger the amount of variance explained. The sum of all eigenvalues has to be equal to the number of variables. This is why we are only potentially interested in those principal components with an eigenvalue greater than 1, although there are as many principal components (or dimensions) as variables.

The principal components are sorted by the associated eigenvalue: The first principal component is the one that has the highest eigenvalue and thus can explain the greatest amount of variance in the data, the second one is the one that has the second highest eigenvalue, and so on. Out of the 18 principal components extracted, there are only 3 with an eigenvalue greater than one. These three alone capture 80 percent of the total variation in the data. Actually, only one of the has a 2-digit percentage of explained variance. Not only this but, in this case, the first principal component alone explains more than 65 percent of the total variation in the data, making data reduction very effective.

Tables 10 and 11 show the total variance explained by the principal components in 1870 and 1890 respectively, sorted by the amount of variance in the data they are able to capture. Remarkably, the analysis reveals that the first principal component alone explains between 63 and 65 percent of the variation in the data for both years 1870 and 1890. The significance of this first principal component is also very clearly reflected in the scree plots (graphs 2 and 3 corresponding to 1870 and 1980 respectively). The figures for the eigenvalues abruptly decline after the first principal component. So we can take without fear the first principal component as the main underlying unobserved explanatory factor in the data.

How many components to retain? It can be observed both in the explained variance table and the scree plot that the first principal component is overwhelmingly the most representative one and effective in terms of data reduction. This is a quite clear case; nevertheless, it is always beneficial to double check with the theory. In a relatively recent methodological paper, Diana and Tommasi (2002) suggest to stop when the associated eigenvalue is more than one and a half times the following one[14]. In this case, the ratio of the first to the second eigenvalue is in the order of 7, so we are far on the safe side.

The dominant principal component of the analysis of the renovated Adelman and Morris database can be interpreted as the level of socio-economic development. I have extracted a score for each and year I have data for. I named this variable SDI [YEAR], standing for social development index in a given year. Series for years 1870 and 1890 are presented here; though there is potential to expand the list of years to 1850 and 1910, with the necessary amount of time and resources. Most of the necessary data for the latter are already available, but the way lagged variables operated in the computations made the final index non-obtainable for the first and last periods, 1850 and 1914.

The first principal component score coefficients and scores for both years 1870 and 1890 are shown in the tables 12 and 13. The component scores coefficients shown in

[14] In other words, max{j=1, …, k} Aj/Aj+i > 1.5 (Diana and Tommasi, 2002:80).

table 12 are the weights by which variables are multiplied to obtain the country scores. Full name and description of variables can be found in the appendix. Table 13 shows the first principal component scores obtained for every country and year in the sample. These are available for 23 countries (remarkably including China, country for which no data at this respect has ever been produced before, to the extent of my knowledge). In this way, we have a score for Argentina 1870, another for Argentina 1890, and so on, completing the list of 23 countries.[15] Two additional columns have been added to table 13 in order to monitor the evolution of the social development index over time. The third numerical column has been obtained by subtracting SDI 1870 from SDI 1890. The result is the change of the index in these two decades. The last column in table 13 indicates the sign of the change, either positive (increase) or negative (decrease).

Interestingly enough, practically all countries in the sample show an increase in social development for the period under study, 1870 to 1890. New Zealand and Japan are the countries that improved the most in the SDI (more than 1 standard deviation in 20 years), being the average improvement around half a standard deviation. The United Kingdom is the only country that appears to have lost some of its stock of social capital with respect to other countries. After the United Kingdom, but still exhibiting a positive sign of change, are China and France, in this order. There seems to the 2 groups of countries that have improved less than the rest of the sample, these being either Western European countries which departed from a privileged position and therefore do not have so much catching-up potential (UK, France, Belgium), or countries in a very poor stage of development which are not making much progress at that time yet (China, Egypt, India). These 2 groups are the countries that have lost or won less in terms of relative position in the ranking.

The two completely new series are depicted in figure 4. The Social Development Index for 1870 is positioned in the horizontal axis, while the 1890 counterpart lays on the vertical axis. In this way, we can see the change in the positioning of countries during the 20-year period in between. A diagonal 45-degree line has been drawn for ease of interpretation. All countries above the line improved their score in 1890 with respect to 1870. Countries below the line scored lower in 1890 than in 1870. Almost all countries managed to improve their score, as confirmed in table 13.

[15] List of countries in Appendix A.

IV. LONG RUN INTER-TEMPORAL COMPARISONS: NINETEENTH AND TWENTIETH CENTURIES

Both Tabellini (2007) and Grosjean (2011) provide a framework for long run inter-temporal comparison. In their investigation for the European or near-European regions, they conclude that economic outcomes have their root in historically fundamented attitudes towards trust, respect for others and self-determination and, therefore, give a quite deterministic view to economic development, in the direction of what Douglass North pointed at regarding historical determinism (North, 1981).

The newly developed Social Development Index for the late nineteenth century creates some room for a long run inter-temporal comparison of the quality of the society for countries all over the world, including both colonised and colonisers. The new SDI series for 1870 and 1890 can be contrasted to the contemporary measures of social capital. In particular SOCDEV for the early 1960's was constructed with a similar technique. Unfortunately, samples of countries for the nineteenth and the twentieth century overlap thinly. This results in a small number of countries being in the two samples for this specific index.

Figures 5 and 6 depict the historical evolution of social developments over long periods of time. Figure 5 represents the change in scores over almost a century, from 1870 to 1960. Figure 6 depicts the change over a 70 year period, from 1890 to 1960. All countries in the sample have improved notably over these long periods of time.

We have just seen that tracing the change of the Social Development Index over the twentieth century is currently feasible for a small sample of countries. Now, what can we learn from the relationship of the nineteenth century SDI with other twentieth century indicators of social capital? At this point, it turns useful to bring into the analysis the two most popular contemporary alternatives, namely trust and civic engagement. In particular, are there any patterns in which these two contemporary variables proceeding from surveys relate to the nineteenth century newly constructed estimates?

Starting with the most recent first, the relationship between TRUST and CIVIC is illustrated by means of the overlay scatter plot (i. e. overlapping to scatter plots). The scatter plots should be read in the same way described for the twentieth century analysis in section II. In figures 7 to 12 the light coloured dots and lines represent the relation between SDI and TRUST, while the dark dots and lines represent the relation between SDI and CIVIC.

Figure 7 shows quite different from its twentieth century counterpart studied in section II (figure 2). In the first place, countries are more widely spread over the social development index range and less over the vertical axis. This indicates convergence from a wide range of social development positions in the nineteenth century to a more equalised level at the end of the twentieth century. Late nineteenth century results are similar (see figure 8 for 1890). Secondly, means for contemporary variables are reversed. Now the mean of TRUST is higher than the mean of CIVIC, both when contrasted to past SDI. This is true both for 1870 and 1890. Also, civic engagement appears to be more stable or equalised across countries than trust. This means that all countries in the display a relatively similar level of contemporary civic engagement, no matter where they were standing in terms of social development in the past. However, this is not true for trust. There seems to be a pattern in the distribution of trust across countries, depending what was their departure point in terms of social development in the past. Let us have a closer look at this phenomenon.

Figures 9 and 10 show the same set of data with fitted lines for 1870-nowadays and 1890 nowadays historical evolutions. In both cases the fitting method was the lowess method, with fifty percent of the points fitted in three iterations (same as in section II graphs). Here we can observe that trust is more volatile across countries than civic engagement is, and has a tendency to be less equalised. This fact stands clear from both 1870-nowadays and 1890-nowadays fit lines. Again, India stands as the most paradigmatic outlier in the sample, showing a spectacular social evolution in the course of the twentieth century. New outliers revealed by the nineteenth century analysis are Norway in the good side, and Brazil, Australia, and France in the down side. We would not have expected this deceiving result from France or Australia, even with the more than one century's perspective. But be aware that we have only very recently realised that India had a big potential for economic growth, which has recently being spectacularly coming out. This was not obvious just twenty years ago. So, we are afraid one could be a catastrophist when auguring growth prospects for France and Australia if one is to judge by the social evolution indications.

The tendency of civic engagement levels to be equalised across countries regardless of their past is confirmed in figures 11 and 12. These present fit lines with ninety-five percent confidence intervals. Quadratic and cubic regression prediction lines were used respectively, according to which method fitted the data best. Again 1870-nowadays and 1890-nowadays analyses show similar results, the main ones being: 1) civic engagement tends to be similar for all countries in the sample, regardless of where they were standing in terms of social development in the past and, 2) parabolic layout of trust observations.

The contemporary civic engagement levels appear to be very similar for almost every country in the sample, regardless of what was the level of social development in the late nineteenth century. Second, open ends mean that social extremes seem to be more unpredictable. This happens by construction of the confidence intervals. Extremes tend to be more unpredictable, since we only have data either from the right or from the left, but not from both sides. Still, this phenomenon is especially acute for the bottom tale of the sample. Countries with very low levels of social development in the late nineteenth century have proved to unfold in all directions: They might evolve into a miracle (India) or turn down to a catastrophe (Brazil).

The 'parabolic layout of trust points' means that the cross-country study of the evolution of trust contrasted with social development in the nineteenth century reveals a parabolic layout. Even allowing for a higher level polynomial would the prediction line turn out to be quasi-parabolic (see figure 12). The relevance of the quadratic term can be tested econometrically. It can be shown that the square of social development in a trust regression is significant at the standard 5 percent level. In other words, countries in the middle of the spectrum have improved the most with respect to their nineteenth century position in the SDI ranking. I interpret this parabola as the combination of two phenomena: the unpredictable direction that the worst scored countries will follow (see previous paragraph), combined with the Abramovitz hypothesis of 'falling behind' for the best historically positioned countries (Abramovitz, 1986). In this way, countries in the middle of the spectrum have the highest predictable prospects for catching up.

A considerable historical perspective is added to the analysis. We can observe a tendency to persistence of the social indicators. So, there is an element of North's hypothesis on path dependency, as suggested by Tabellini (2007) and Grosjean (2011). However, outliers depart from the trend, doing nothing but confirm that the results are historically consistent with future economic growth trajectories. This is the case of India, which shows exceptionally high values in the social development index or Brazil, whose scores are deceivingly poor. A striking characteristic is the finding that some socially well located countries in the nineteenth century show to be loosing their relative position at the end of the twentieth century. This preoccupating phenomenon, which surprises as counterintuitive, needs a more detailed consideration. One needs to keep in mind that the comparison is done with different scales. But, as a first approximation, this dramatic finding is nothing but the proof of what Putnam was pointing at in his 2000 book *Bowling Alone*, detailing the weakening of social values in the North American society. There seems to be a tendency for Western European countries to fall into this group.

V. CONCLUSION

We presented the first international historical estimates for social capital. Two new series of a Social Development Index (SDI) become available: one for 1870 and one for 1890. We showed a new way of looking at social evolution. Together with some other contemporary measurement attempts, the new series allow monitoring the evolution of a social development index over time.

As a conclusion, one can say that it is true that social attitudes are correlated to economic performance as proved in the literature, but it is not true that the former do not change. The inter-temporal comparison of the various social capital proxies shows that cultural revolutions are possible and are indeed observed in the data if one allows for a sufficient time span. The present study looks at the change in the world country ranking of social capital proxies for a period of 70 to 90 years, and the result is that cultural miracles (as growth miracles) exist. When all regions in the world are included instead of only Europe, differentiated patterns of social development trajectories can be pinned down.

Thus, North's hypothesis of path dependency is modified. Practically all countries in the sample show an increase in social development during the intermediate period (1870 to 1890), and all of them reveal a very significant improvement over the twentieth century.

We find some outstanding performers in social improvement, defeating path dependence, and also detect some failure stories. In both cases, the social development trajectories seem to be historically consistent with their subsequent economic growth experiences.

Europe’s relative position with respect to the rest of the World varies. Scandinavian countries are absolute leaders on trust, while they were in the centre of the social development spectrum more that 100 years ago. Meanwhile, some core Western European countries like France or the United Kingdom, who were World leaders once, seem to have lost their privileged positions during the course of the twentieth century.

Finally, different social capital measurement alternatives exhibit different patterns, suggesting that they are simply capturing different aspects. We find statistically significant non-linear relationships between them. In particular, trust describes a parabolic layout with respect to our Social Development Index, and civic engagement stands as surprisingly even across countries.

TABLES AND FIGURES FOR SECTION II - CONTEMPORARY COMPARISON OF ALTERNATIVES

Table 1 – Three alternative measures to monitor social capital

COUNTRY	SOC DEV	TRUST AM	CIVIC KK	COUNTRY	SOC DEV	TRUST AM	CIVIC KK
Afganistan	-1,02	.	.	Lesotho	.	.	.
Algeria	0,18	.	.	Liberia	-1,01	.	.
Angola	.	.	.	Libya	-0,68	.	.
Argentina	1,91	27.0	39.50	Lithuania	.	22	.
Armenia	.	25	.	Luxembourg	.	.	.
Australia	.	47.8	38.27	Madagascar	-1,31	.	.
Austria	.	31.8	41.45	Malawi	-1,57	.	.
Azerbaijan	.	21	.	Malaysia	.	.	.
Bahamas, The	.	.	.	Mali	.	.	.
Bahrain	.	.	.	Malta	.	.	.
Bangladesh	.	21	.	Mauritania	.	.	.
Barbados	.	.	.	Mauritius	.	.	.
Belarus	.	24	.	Mexico	0,75	17.7	34.55
Belgium	.	30.2	38.08	Moldova	.	22	.
Benin	-1,54	.	.	Morocco	-0,57	.	.
Bolivia	-0,35	.	.	Mozambique	.	.	.
Botswana	.	.	.	Myanmar (Burma)	-0,41	.	.
Brazil	0,79	6.7	37.58	Nepal	-1,36	.	.
Bulgaria	.	30.4	.	Netherlands	.	46.2	38.36
Burkina Faso	.	.	.	New Zealand	.	.	.
Burundi	.	.	.	Nicaragua	0,88	.	.
Cambodia	-0,55	.	.	Níger	-1,86	.	.
Cameroon	-1,34	.	.	Nigeria	-0,91	22.9	39.19
Canada	.	49.6	39.74	Norway	.	61.2	40.75
Cape Verde	.	.	.	Oman	.	.	.
Central African Rep.	.	.	.	Pakistan	-0,08	.	.
Chad	-1,70	.	.	Panama	0,84	.	.
Chile	1,39	22.7	36.80	Papua New Guinea	.	.	.
China	.	.	.	Paraguay	0,97	.	.
Colombia	0,66	10	.	Peru	0,68	5	.
	SOC DEV	TRUST AM	CIVIC KK	COUNTRY	SOC DEV	TRUST AM	CIVIC KK

Comoros	.	.	.	Philippines	0,56	6	.
Congo	.	.	.	Poland	.	34.5	.
Costa Rica	0,78	.	.	Portugal	.	21.4	36.89
Cote d'Ivoire	-0,98	.	.	Romania	.	16.1	.
Croatia	.	25	.	Russia	.	24	.
Cyprus	1,08	.	.	Rwanda	.	.	.
Czech Republic	.	30	.	Saudi Arabia	.	.	.
Denmark	.	56.0	40.34	Senegal	-0,52	.	.
Dominica	.	.	.	Seychelles	.	.	.
Dominican Rep.	0,81	26	.	Sierra Leone	-1,39	.	.
Ecuador	0,54	.	.	Singapore	.	.	.
Egypt	0,73	.	.	Slovakia	.	23	.
El Salvador	0,71	.	.	Slovenia	.	16	.
Estonia	.	22	.	Solomon Islands	.	.	.
Ethiopia	-0,99	.	.	Somalia	-1,35	.	.
Fiji	.	.	.	South Africa	0,62	30.5	36.99
Finland	.	57.2	40.64	Spain	.	34.5	38.75
France	.	24.8	36.26	Sri Lanka	0,35	.	.
Gabon	-0,83	.	.	St.Lucia	.	.	.
Gambia	.	.	.	St.Vincent&Grens.	.	.	.
Georgia	.	23	.	Sudan	-0,64	.	.
Germany	.	29.8	39.83	Suriname	0,54	.	.
Ghana	-0,01	23	.	Swaziland	.	.	.
Greece	1,47	.	.	Sweden	.	57.1	41.57
Grenada	.	.	.	Switzerland	.	43.2	40.89
Guatemala	0,35	.	.	Syria	0,57	.	.
Guinea	-1,47	.	.	Taiwan	1,05	42	.
Guinea-Bissau	.	.	.	Tanzania	-1,22	.	.
Guyana	.	.	.	Thailand	0,50	.	.
Haiti	.	.	.	Togo	.	.	.
Honduras	0,26	.	.	Tonga	.	.	.
Hong Kong	.	.	.	Trinidad & Tobago	1,15	.	.
Hungary	.	24.6	.	Tunisia	-0,18	.	.
Iceland	.	41.6	41.07	Turkey	0,88	10.0	42.43
	SOC DEV	**TRUST AM**	**CIVIC KK**	**COUNTRY**	**SOC DEV**	**TRUST AM**	**CIVIC KK**
India	-0,28	34.3	42.65	Uganda	-1,22	.	.

Indonesia	-0,40	.	.	Ukraine	.	31	.
Iran, I.R. of	0,09	.	.	United Arab Emirates	.	.	.
Iraq	-0,03	.	.	United Kingdom	.	44.4	40.07
Ireland	.	40.2	37.51	United Status	.	45.4	40.55
Israel	1,77	.	.	Uruguay	1,59	22	.
Italy	.	26.3	41.23	Vanuatu	.	.	.
Jamaica	1,06	.	.	Venezuela	1,37	14	.
Japan	1,63	40.8	41.79	Vietnam, South	-0,49	.	.
Jordan	0,16	.	.	Western Samoa	.	.	.
Kenya	-0,53	.	.	Yemen, N.Arab	-1,35	.	.
Korea	0,85	38	.	Yugoslavia	.	31	.
Kuwait	.	38.0	39.64	Zaire	.	.	.
Laos	-1,06	.	.	Zambia	-0,89	.	.
Latvia	.	25	.	Zimbabwe	0,14	.	.
Lebanon	1,44	.	.				

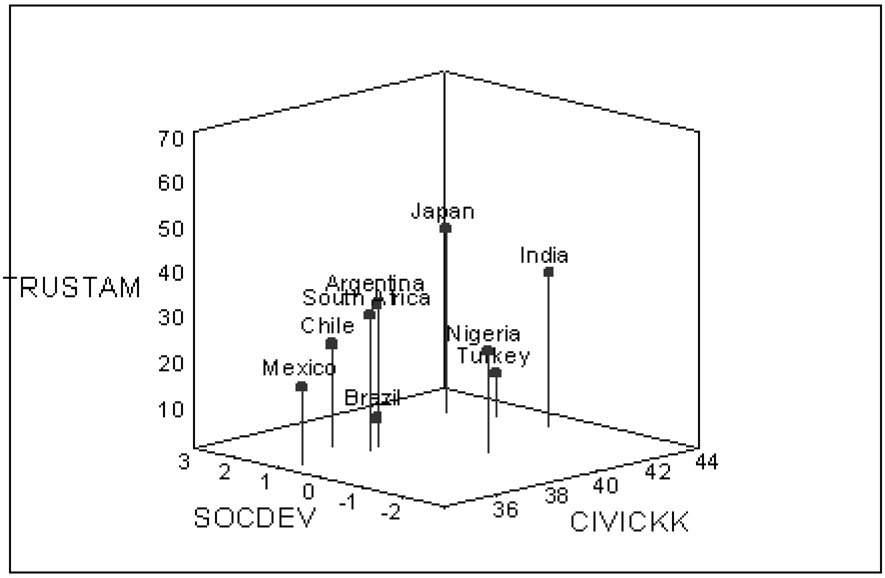
Table 2 – Pearson Correlation Matrix of Social Capital Alternatives

		socdev 1960	**trustam 1990**	**civickk 1990**
socdev 1960	Pearson Correlation	1	,090	-,089
	Tail probability		,723	,820
	Number of cases	75	18	9
trustam 1990	Pearson Correlation	,090	1	,387(*)
	Tail probability	,723		,038
	Number of cases	18	58	29
civickk 1990	Pearson Correlation	-,089	,387(*)	1
	Tail probability	,820	,038	
	Number of cases	9	29	29

* Correlation is significant at the 0.05 level (2-tailed).

The correlation matrix shows that there is no strong linear relationship between these three variables.

Figure 1 - 3D Scatter Plots of Social Capital Alternatives. Spikes to the Floor and Centroid respectively



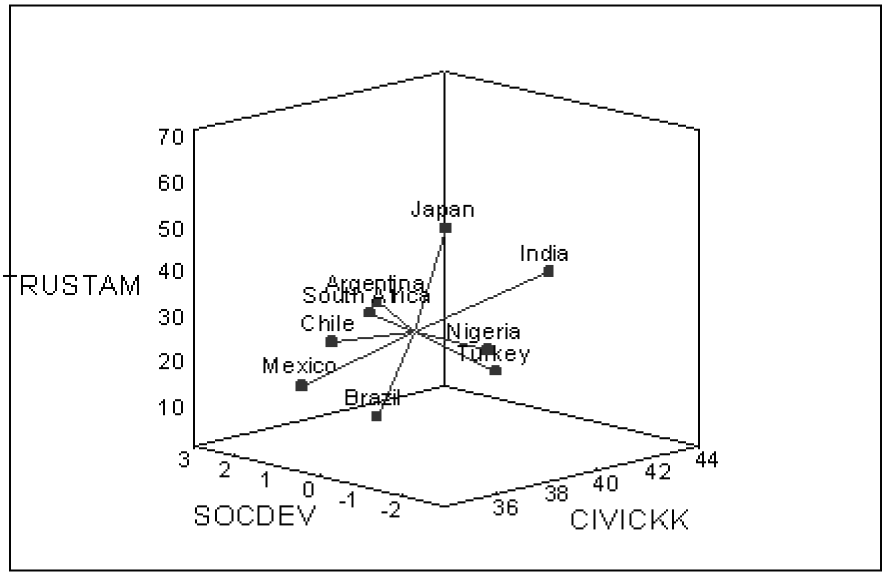
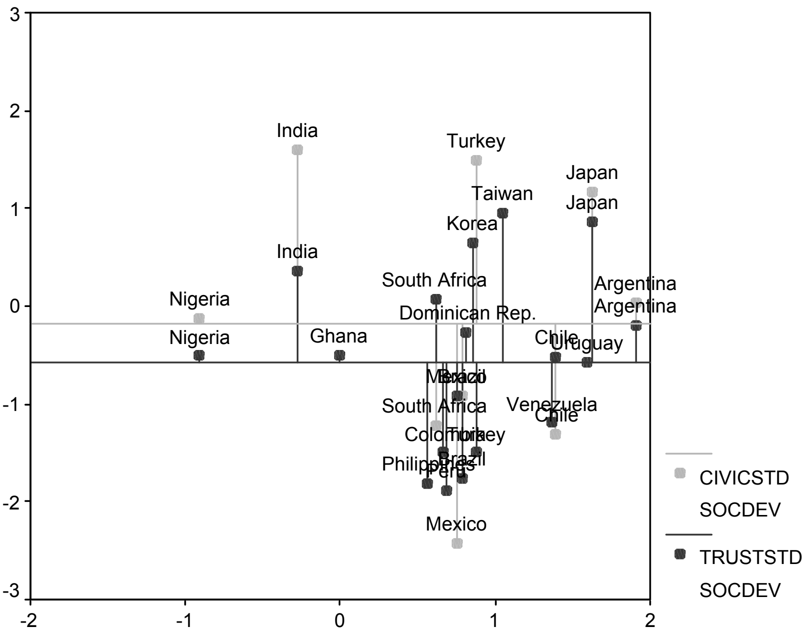
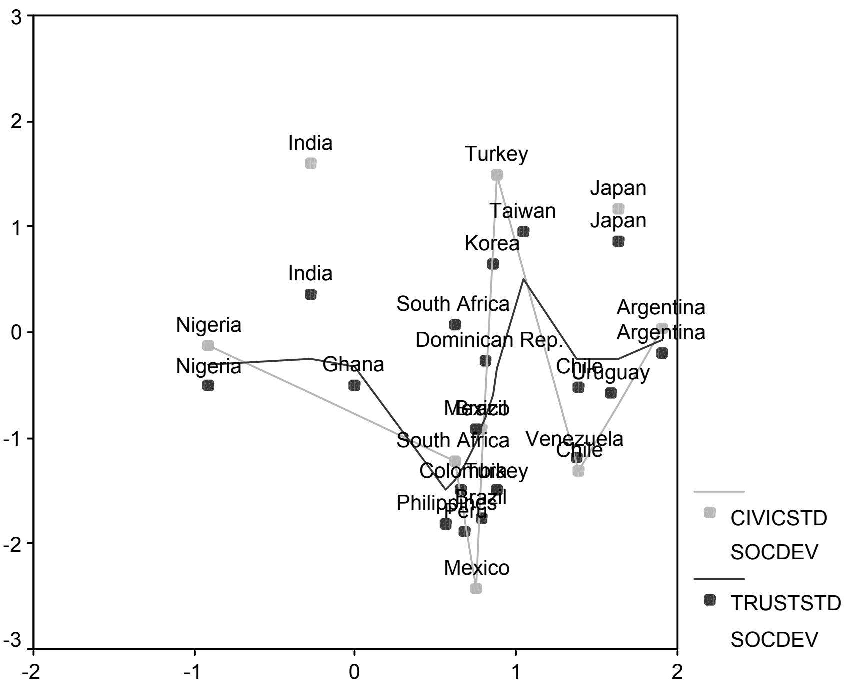


Figure 2 - Historical Evolution, from 1960 (Soc Dev) to Surveys: Overlay Scatter Plot



SOCDEV in the horizontal axis, CIVICSTD (light) and TRUSTSTD (dark) in the vertical axis. Spikes to reference line for each pair. Reference lines are mean of Y

Figure 3 - Overlay Scatter Plot with Fit Line



SOCDEV in the horizontal axis, CIVICSTD (light) and TRUSTSTD (dark) in the vertical axis. Fit Method: Lowess. 50% of points fitted with 3 iterations.

TABLES FOR SECTION III.3 – PRE-PCA TESTS: OPTIMISING THE VARIABLES

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,716
Bartlett's Test of Sphericity	Approx. Chi-Square	1684,231
	df	496
	Sig.	,000

Table 3 – First round variables, after excluding rate of adoption of new technologies in agriculture

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,769
Bartlett's Test of Sphericity	Approx. Chi-Square	1419,284
	df	325
	Sig.	,000

Table 4 – Third (and last) round variables, after excluding rate of adoption of new technologies in agriculture, population per farmland, export growth group, the demographic variables (population group, population growth group and immigration group) and land concentration.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,869
Bartlett's Test of Sphericity	Approx. Chi-Square	1258,497
	df	231
	Sig.	,000

Table 5 – Second (and last) round variables following a stricter criterion of variable-by-variable KMO selection (mark greater or equal than 0.6) for every variable. This has raised the overall KMO measure of sampling adequacy from 0.77 to 0.87, while keeping the significance of the sphericity test practically untouched.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,909
Bartlett's Test of Sphericity	Approx. Chi-Square	1093,367
	df	153
	Sig.	,000

Table 6 – Final selection of variables

Table 7 - Squared Multiple Correlations (SMC) of variables with all other variables

Variables	**Original Data**		**New Data**	
	SMC initial variables	SMC final variables, after KMO test	SMC initial variables	SMC final variables, after KMO test
income	.9841	.9399	.9790	.9270
income growth	.9330	-	.8930	-
technique in industry	.9702	.9290	.9634	.9258
adoption of new techniques	.9349	.8956	.9776	.8501
technique in agriculture	.9714	.9152	.9799	.9029
agricultural labour	.9700	.9043	.9820	.8839
population per farmland	.8578	-	.9160	-
inland transportation	.9581	.9046	.9346	.8739
transportation, growth	.8049	.6294	.9151	.5736
export growth group	.7242	-	-	-
trade	-	-	.9618	-
shift in export structure	.8594	.7928	.8747	.6935
industrial wage change	.8491	-	.8840	-
agricultural wage change	.7164	-	.8710	-
wages	-	-	.8971	-
population	.8951	-	.9654	-
population growth	.8988	-	.9506	-
immigration	.8980	-	.8947	-
migration to pop. growth	-	-	.9050	-
literacy	.9748	.9309	.9879	.8827
primary education growth	.8011	.7330	.9401	-
form of land tenure	.9287	-	.9348	-
land concentration	.7499	-	.8628	-
land adoption	.9302	.7819	.9329	.7689
urbanisation	.9321	.8354	.9361	.7707
entrepreneurship	.9584	.9188	.9841	.8941
role of government	.7112	.6152	.6427	-
socio-politics	.9145	.8743	.9741	.8555
representativeness	.9488	.9042	-	.8479

polity 2	-	-	.9659	-
political stability	.9595	.8665	.9579	-
foreign dependency	.9161	.7965	.9400	-
colonial status	.9414	.8742	.9639	.7587
market development	.9810	.9511	.9922	.9318
market development growth	.9989	.9975	.9988	.9972
mkt develpt growth, lagged	.9989	.9972	.9987	.9969
Number of variables	**32**	**22**	**34**	**18**
KMO sampling adequacy	**.7161**	**.8692**	**.6026**	**.9086**

The SMC shows that all variables are strongly correlated to one another. However the sampling adequacy test reveals that not all of them are so relevant for the components analysis.

Table 8 – Kaiser-Meyer-Olkin (KMO) sample adequacy scores variable-by-variable

Variables	**Original Data**		**New Data**	
	KMO initial variables	KMO final variables	KMO initial variables	KMO final variables
income	.6985	.8589	.8194	.8938
income growth	.5393	-	.5443	-
technique in industry	.7590	.8866	.8517	.8856
adoption of new techniques	.8853	.8781	.6678	.9126
technique in agriculture	.7950	.9330	.6745	.9389
agricultural labour	.6579	.8512	.6578	.8889
population per farmland	.5702	-	.3341	-
inland transportation	.8314	.8674	.8629	.9161
transportation, growth	.7421	.8454	.4820	.8683
export growth group	.4345	-	-	-
trade	-	-	.1977	-
shift in export structure	.7839	.8170	.6892	.9214
industrial wage change	.5994	-	.5535	-
agricultural wage change	.5819	-	.2956	-
wages	-	-	.2857	-
population	.5216	-	.4677	-
population growth	.4177	-	.4521	-
immigration	.4658	-	.2399	-
migration to pop. growth	-	-	.4671	-
literacy	.8090	.8931	.7048	.9332
primary education growth	.6685	.7167	.3733	-
form of land tenure	.5909	-	.5719	-
land concentration	.2078	-	.1454	-
land adoption	.7264	.9337	.6294	.9288
urbanisation	.6957	.8547	.6394	.9138
entrepreneurship	.9029	.9515	.6509	.9444
role of government	.6380	.6545	.3797	-
socio-politics	.8491	.9230	.6929	.9372
representativeness	.8283	.8854	-	.9287
polity 2	-	-	.4022	-

political stability	.6373	.8033	.6323	-
foreign dependency	.7443	.8602	.6904	-
colonial status	.6276	.7379	.5800	.8168
market development	.7861	.9014	.6380	.9322
market development growth	.7401	.8683	.7381	.8713
mkt develpt growth, lagged	.7096	.8651	.7573	.8675
Number of variables	**32**	**22**	**34**	**18**
Overall sampling adequacy	**.7161**	**.8692**	**.6026**	**.9086**

The score differences between the initial and the final selection of variables reflect the improvement in sampling adequacy, in a scale from 0 to 1.

TABLES AND FIGURES FOR SECTION III.4 - RESULTS: PRINCIPAL COMPONENTS ANALYSIS

Table 9 - Total Variance Explained

Principal components/correlation Number of obs = 46
Number of comp. = 18
Trace = 18
Rotation: (unrotated = principal) Rho = 1.0000

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11,723	65,126	65,126	11,723	65,126	65,126
2	1,648	9,158	74,284	1,648	9,158	74,284
3	1,068	5,932	80,216	1,068	5,932	80,216
4	,889	4,936	85,152			
5	,664	3,691	88,843			
6	,453	2,516	91,359			
7	,303	1,683	93,042			
8	,287	1,593	94,635			
9	,209	1,161	95,796			
10	,183	1,017	96,813			
11	,142	,788	97,601			
12	,104	,580	98,181			
13	,087	,482	98,663			
14	,084	,467	99,130			
15	,069	,382	99,512			
16	,051	,284	99,796			
17	,035	,196	99,992			
18	,001	,008	100,000			

Codes of the variables included: income indutech indtecgr agritech agrilgrp intransp transpgr shiftx lit landadop urbani entrep sociopol represen colstat mktdev mktdevgr mktdvgrl Extraction Method: Principal Component Analysis.

This is the second round of the new database, but having replaced govrel and polity2 with the old ones govt and represen to start with, given that they give better results than the new ones. As a result, govt should be removed in any case, but represen stays in the second round. Re-introducing represen makes polstabi and foreignd go; in addition, transpgr cannot be removed.

Graph 1 - Scree Plot

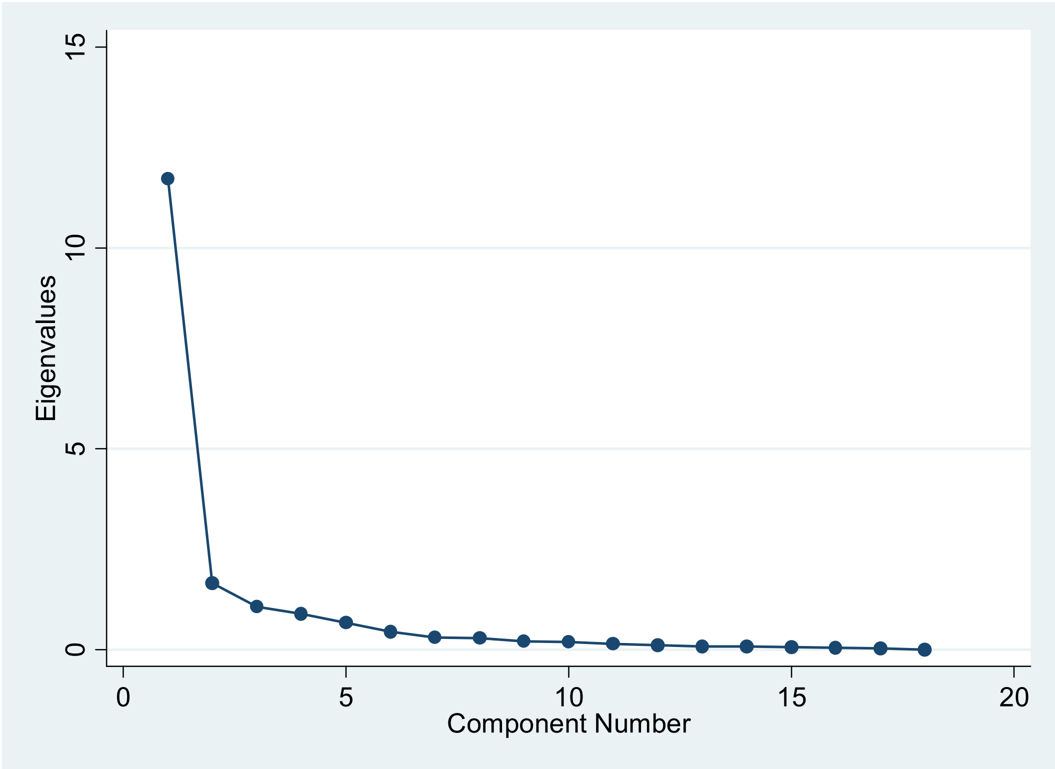


Table 10 - Total Variance Explained for 1870

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11,873	65,961	65,961	11,873	65,961	65,961
2	1,683	9,348	75,308	1,683	9,348	75,308
3	1,259	6,992	82,300	1,259	6,992	82,300
4	,703	3,904	86,204			
5	,693	3,851	90,055			
6	,549	3,053	93,108			
7	,306	1,698	94,806			
8	,207	1,150	95,956			
9	,192	1,064	97,021			
10	,136	,756	97,777			
11	,131	,725	98,502			
12	,084	,468	98,970			
13	,061	,340	99,310			
14	,053	,295	99,605			
15	,033	,185	99,790			
16	,024	,133	99,923			
17	,014	,076	100,000			
18	8,04E-005	,000	100,000			

Codes of the variables included: income indutech indtecgr agritech agrilgrp intransp transpgr shiftx lit landadop urbani entrep sociopol represen colstat mktdev mktdevgr mktdvgrl
Extraction Method: Principal Component Analysis.
Only cases for which year = 1870 are used in the analysis phase.

Graph 2 - Scree Plot for 1870

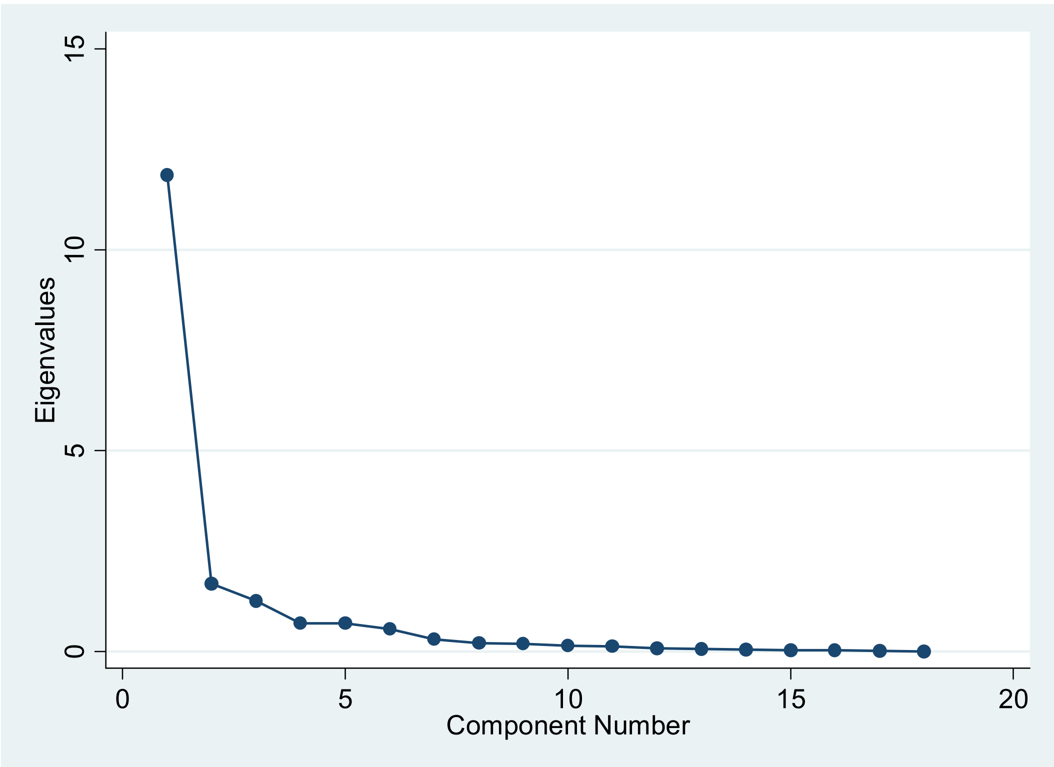


Table 11 - Total Variance Explained for 1890

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11,446	63,587	63,587	11,446	63,587	63,587
2	2,038	11,320	74,907	2,038	11,320	74,907
3	1,279	7,105	82,012	1,279	7,105	82,012
4	,915	5,086	87,098			
5	,587	3,258	90,356			
6	,450	2,500	92,856			
7	,326	1,812	94,668			
8	,246	1,366	96,034			
9	,187	1,041	97,075			
10	,138	,765	97,840			
11	,115	,639	98,478			
12	,082	,453	98,932			
13	,075	,417	99,348			
14	,057	,315	99,664			
15	,047	,259	99,923			
16	,011	,061	99,984			
17	,003	,015	99,999			
18	,000	,001	100,000			

Codes of the variables included: income indutech indtecgr agritech agrilgrp intransp transpgr shiftx lit landadop urbani entrep sociopol represen colstat mktdev mktdevgr mktdvgrl
Extraction Method: Principal Component Analysis.
Only cases for which year = 1890 are used in the analysis phase.

Graph 3 - Scree Plot for 1890

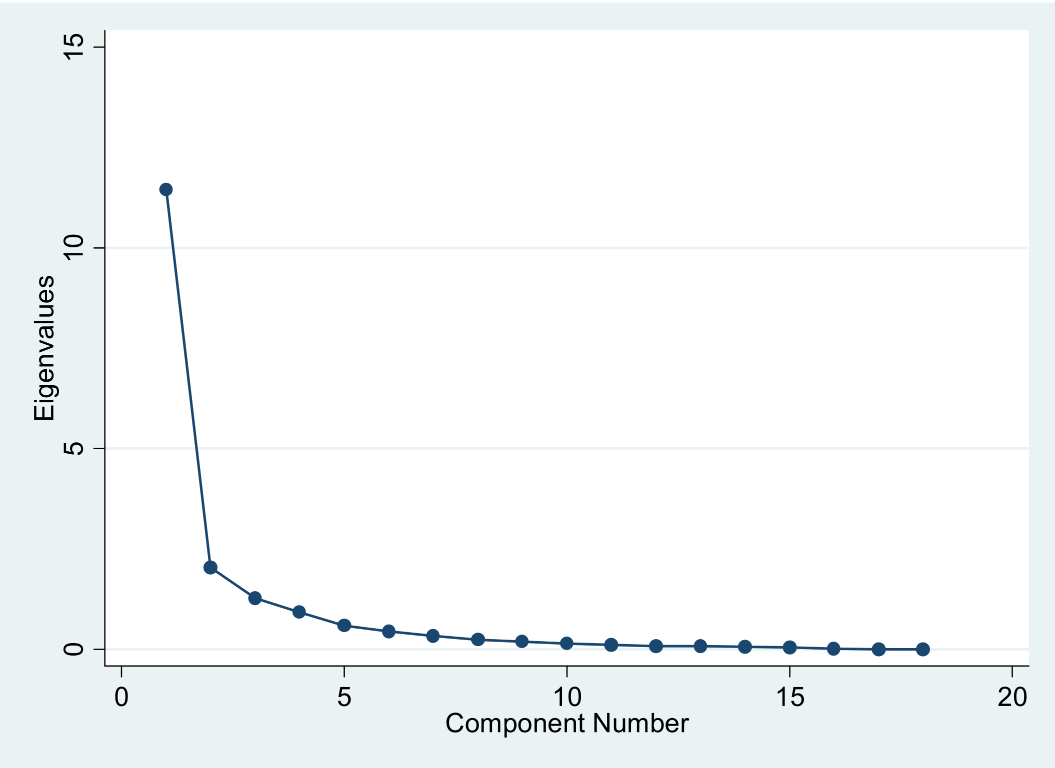


Table 12 - First Principal Component Score Coefficients (Loadings)

Variable	(1) Joint score	(2) Score 1870	(3) Score 1890
Income	.2463	.2363	.2539
Technique in industry	.2603	.2605	.2583
Adoption of new techniques	.2368	.2492	.2230
Technique in agriculture	.2650	.2630	.2666
Agricultural labour	-.2229	-.2181	-.2282
Inland transportation	.2445	.2319	.2577
Transportation, growth	.1744	.1833	.1496
Shift in export structure	.1897	.1907	.1864
Literacy	.2467	.2290	.2667
Land adoption	.2217	.2208	.2362
Urbanisation	.2036	.2041	.1953
Entrepreneurship	.2537	.2574	.2531
Socio-politics	.2509	.2500	.2515
Representativeness	.2529	.2468	.2565
Colonial status	.1731	.1818	.1770
Market development	.2750	.2721	.2787
Market development growth	.2486	.2601	.2338
Mkt develpt growth, lagged	.2435	.2578	.2247

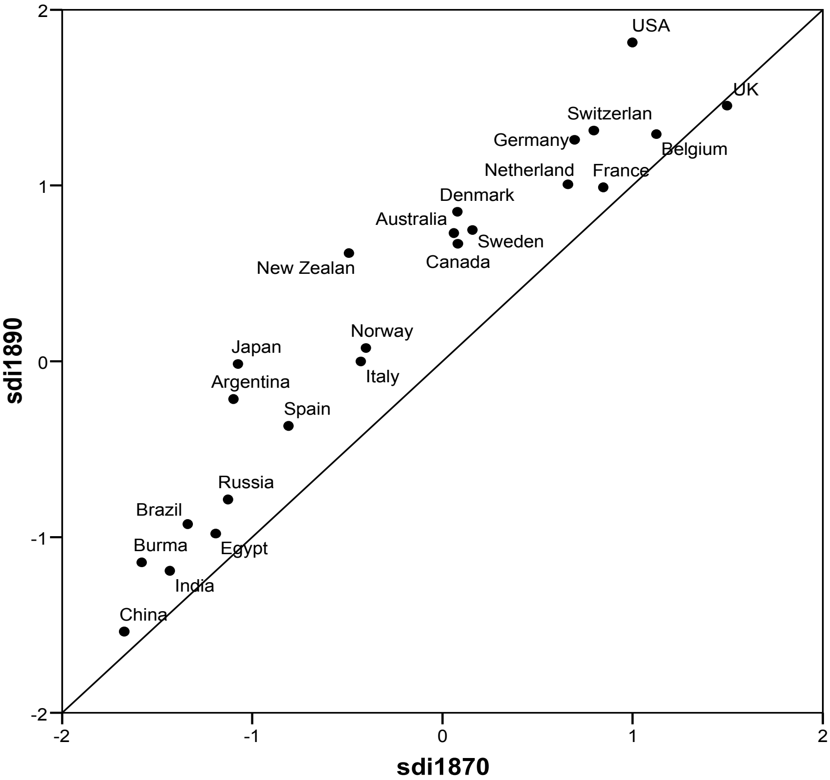
Notes: The weight given to each variable is determined by the eigenvectors of the correlation matrix of all variables. Columns in the table are the eigenvectors associated to the first principal component. (1) Joint score coefficients; common weights across periods. This is the preferred option. (2) Weights corresponding to 1870 only; (3) weights corresponding to 1890 only.
Normalisation: Sum of Squares column =1

Table 13 – First Principal Component Scores

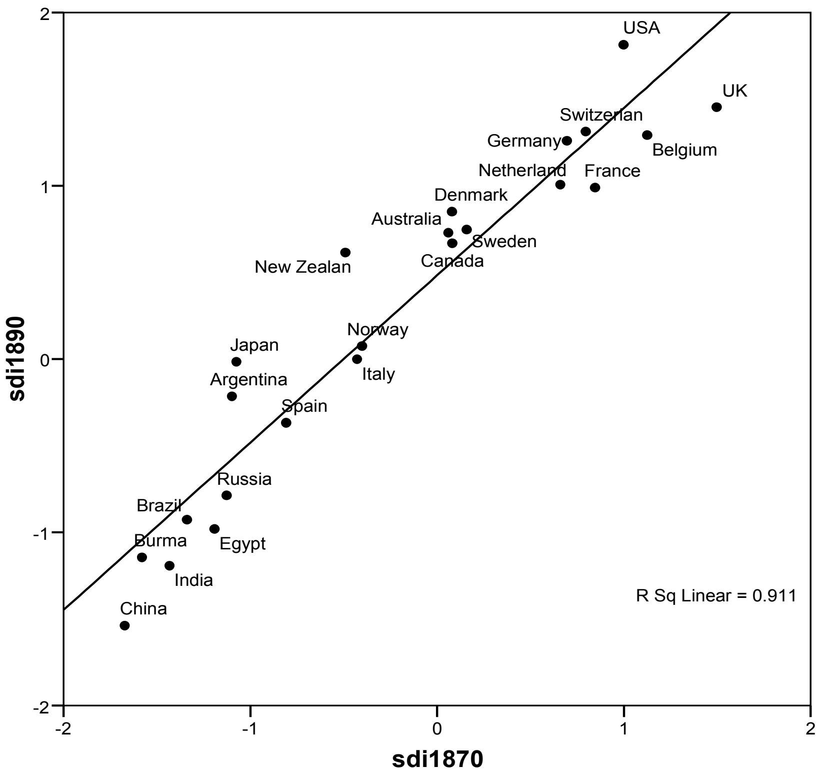
Country	SDI 1870	SDI 1890	Change SDI 1890 –SDI 1870	Sign of Change
Argentina	-1.099	-0.215	0.885	+
Australia	0.060	0.730	0.669	+
Belgium	1.125	1.292	0.168	+
Brazil	-1.339	-0.926	0.413	+
Burma	-1.581	-1.144	0.437	+
Canada	0.081	0.669	0.588	+
China	-1.673	-1.538	0.135	+
Denmark	0.079	0.851	0.772	+
Egypt	-1.192	-0.980	0.212	+
France	0.846	0.990	0.145	+
Germany	0.695	1.261	0.566	+
India	-1.433	-1.192	0.242	+
Italy	-0.429	-0.000	0.429	+
Japan	-1.075	-0.015	1.060	+
Netherlands	0.660	1.007	0.348	+
New Zealand	-0.492	0.615	1.107	+
Norway	-0.402	0.076	0.478	+
Russia	-1.128	-0.786	0.341	+
Spain	-0.809	-0.367	0.442	+
Sweden	0.158	0.747	0.589	+
Switzerland	0.796	1.314	0.518	+
United Kingdom	1.496	1.455	-0.042	-
United States	0.999	1.814	0.815	+

Note: The scores presented here have been standardised to have mean 0 and standard deviation 1. This results in a maximum observed score of 1.814 and a minimum of -1.673. To recover the unstandardised scores, multiply by 3.424. Italy score -0.000 is marked with the negative sign because for more than 4 significative ciphers, the number is negative.

Figure 4 - Scatter Plot for the New Social Development Index: 1870 against 1890

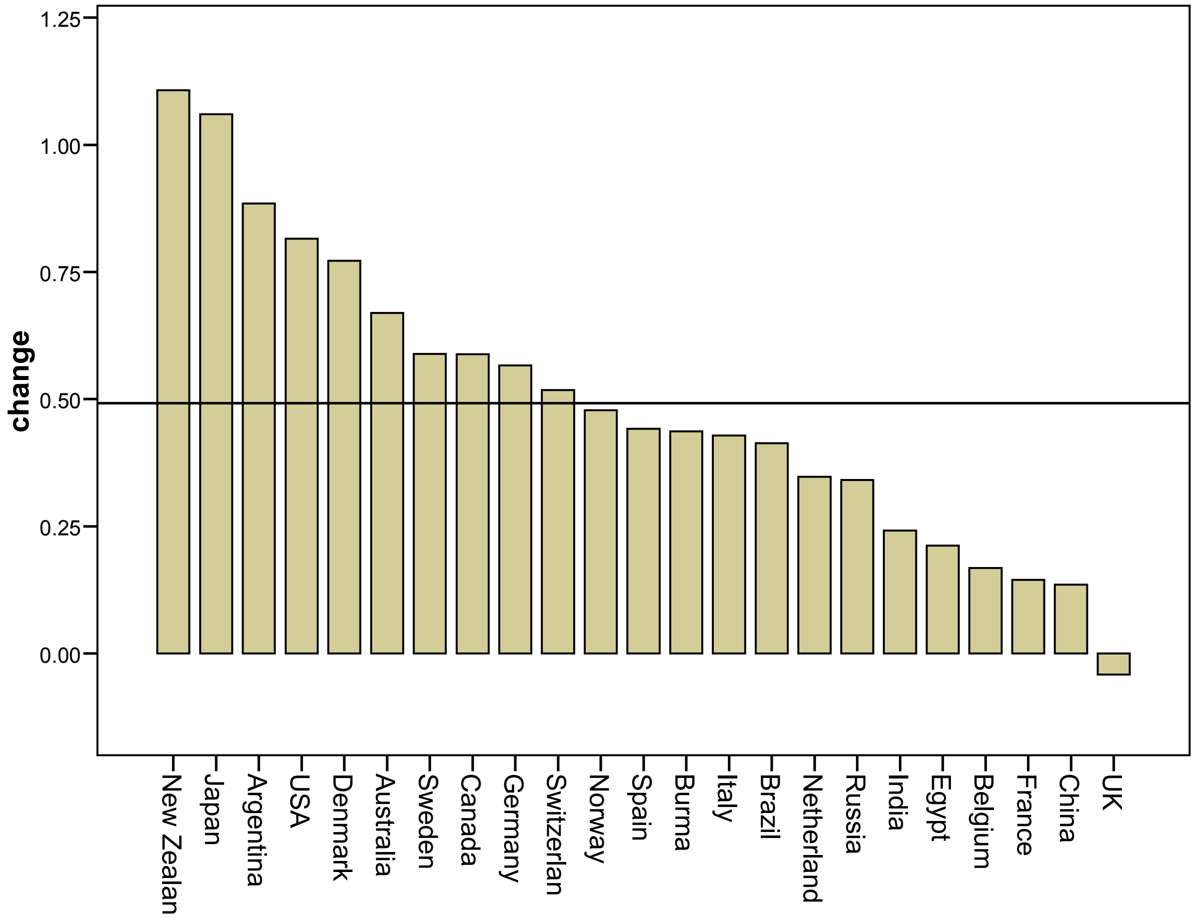


SDI 1870 in the horizontal axis. SDI 1890 in the vertical axis. Almost all countries lie above the 45-degree line. This means all countries except UK improved over this period.



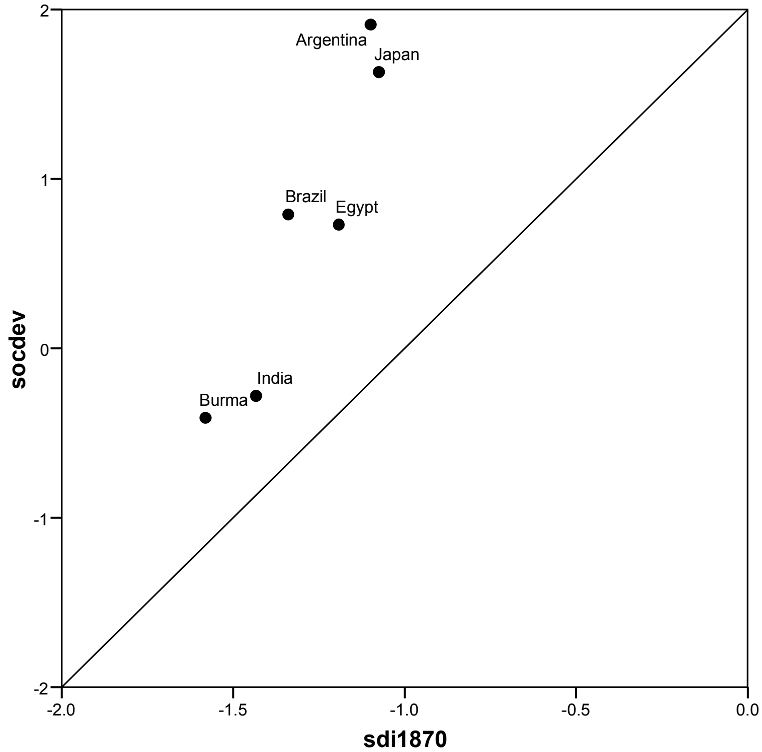
Scores with fit line at average standard deviation. Scores have improved an average of around half standard deviation in 2 decades.

Graph 4 – SDI Change from 1870 to 1890



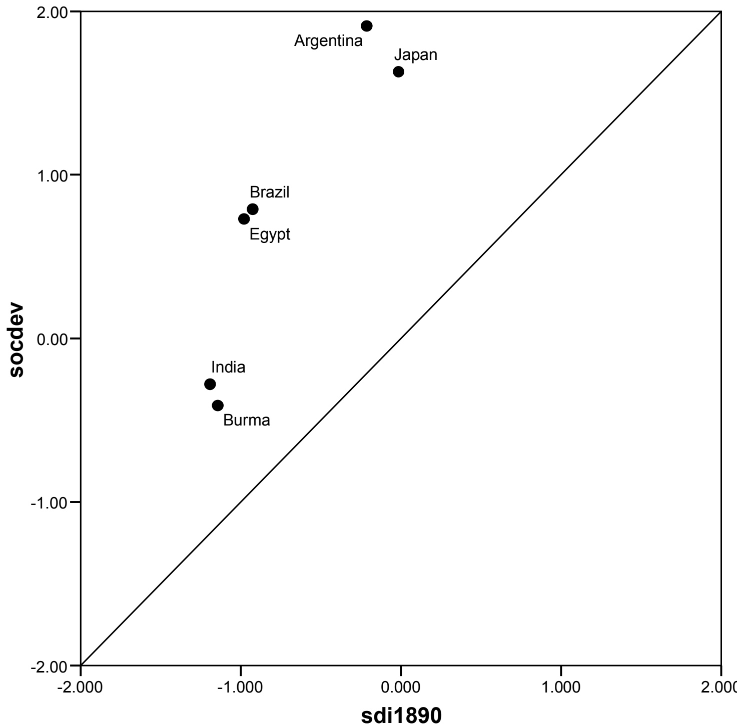
Reference line set equal to the mean =.492

Figure 5 - Scatter plot for Social Development Index. Historical Evolution from 1870 to 1960



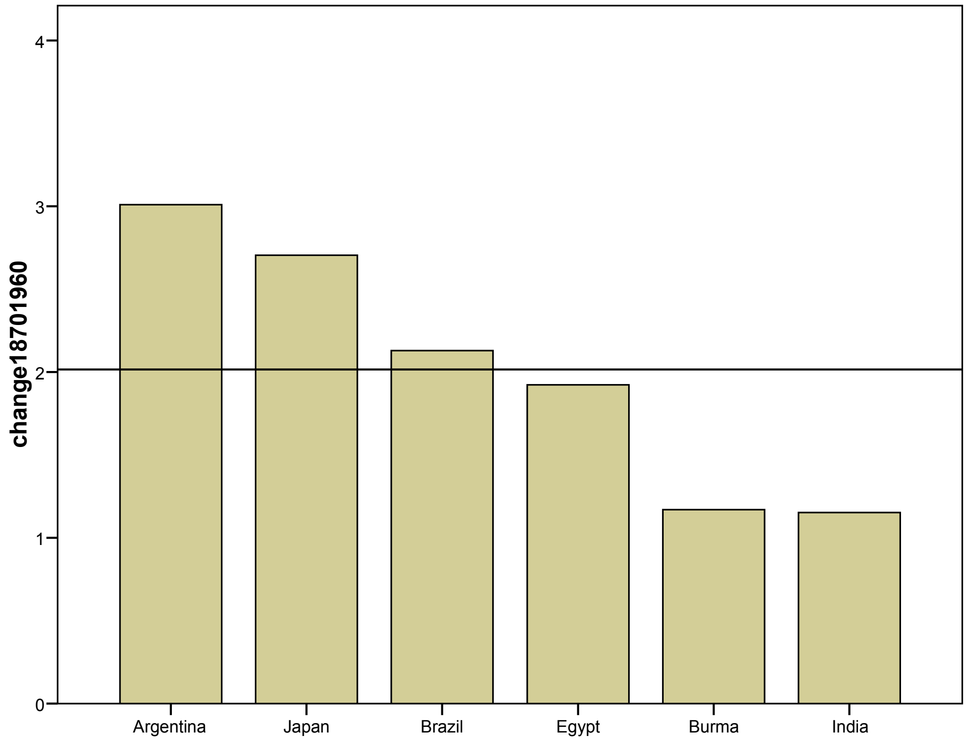
SDI 1870 in the horizontal axis. SOCDEV in the vertical axis.
All countries in the sample improved notably during the period 1870-1960.

Figure 6 - Scatter plot for Social Development Index. Historical Evolution from 1890 to 1960



SDI 1890 in the horizontal axis. SOCDEV in the vertical axis.
All countries in the sample improved notably during the period 1890-1960.

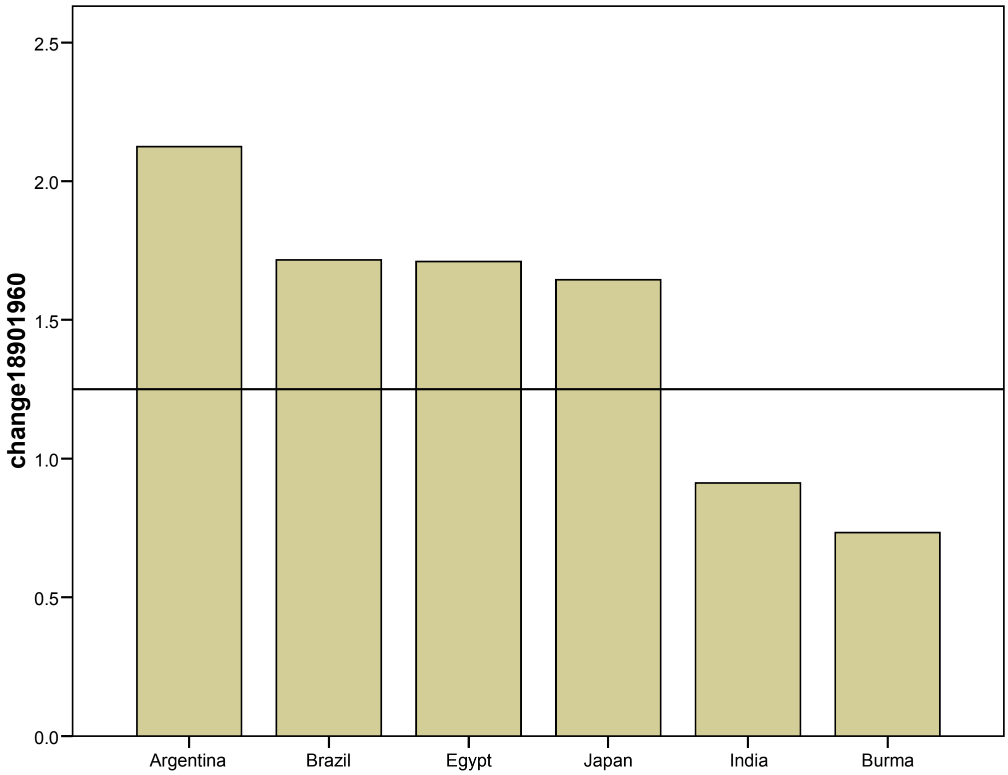
Graph 5 – SDI Change from 1870 to 1960



Reference line set at mean value = 2.02

Countries have improved an average of 2 standard deviations between 1870 and 1960 (90 years).

Graph 6 – SDI Change from 1890 to 1960

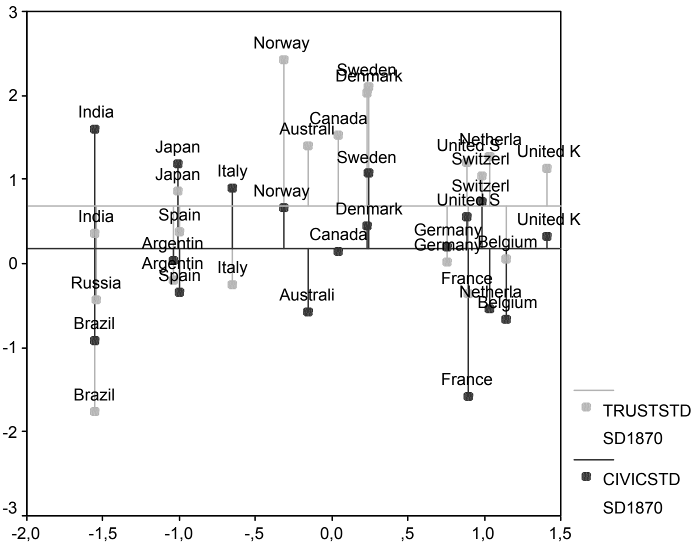


Reference line set at mean value = 1.47

Countries have improved an average of around one and a half standard deviations between 1890 and 1960 (70 years).

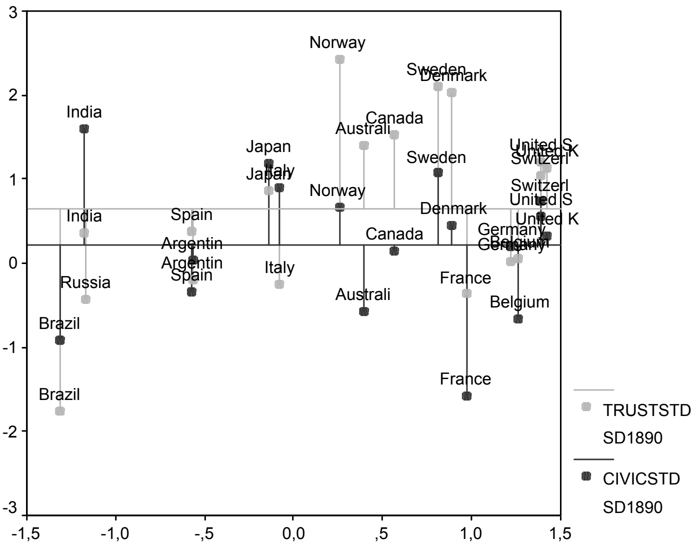
FIGURES FOR SECTION IV - INTERTEMPORAL COMPARISONS: NINETEENTH AND TWENTIETH CENTURIES

Figure 7 - Historical Evolution 1870-Nowadays: Overlay Scatter Plot with Spikes



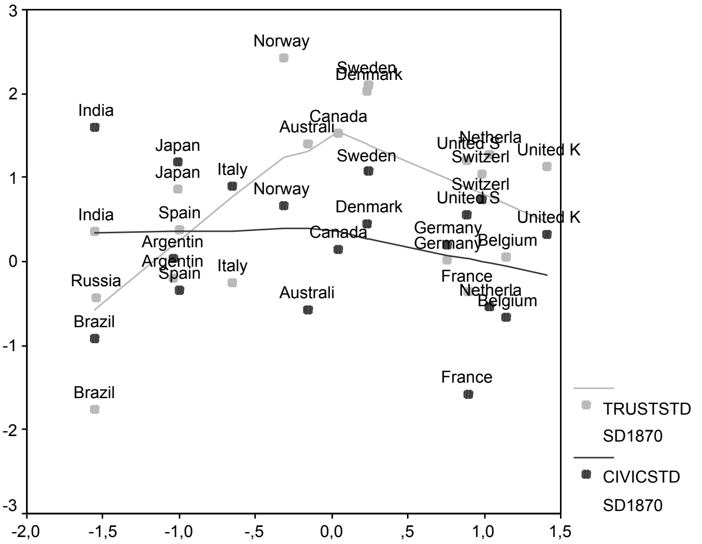
SDI 1870 in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.

Figure 8- Historical Evolution 1890-Nowadays: Overlay Scatter Plot with Spikes



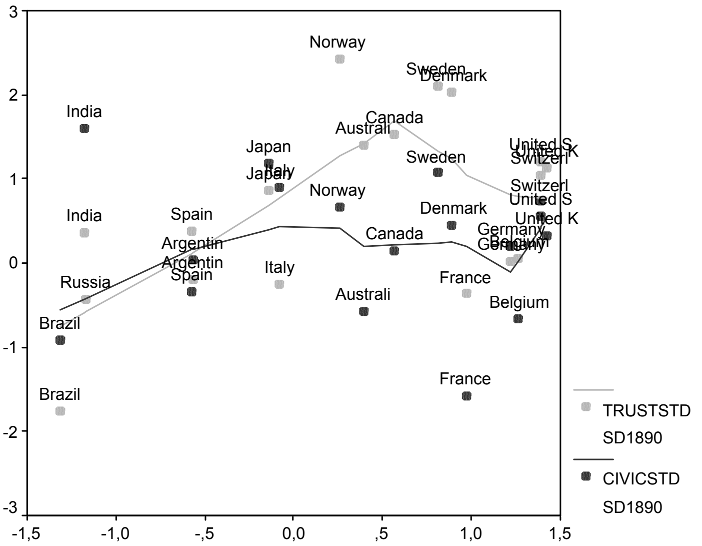
SDI 1890 in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.

Figure 9 - Historical Evolution 1870-Nowadays: Overlay Scatter Plot with Fitted Line



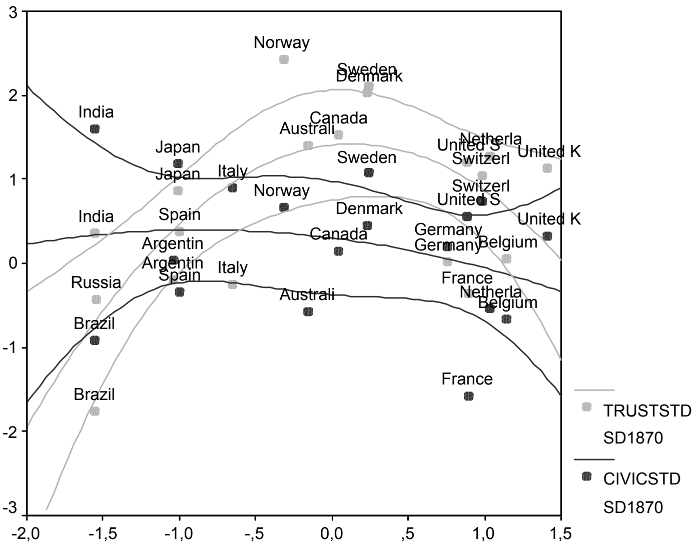
SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Lowess. 50% of points fitted with 3 iterations.

Figure 10 - Historical Evolution 1890-Nowadays: Overlay Scatter Plot with Fitted Line



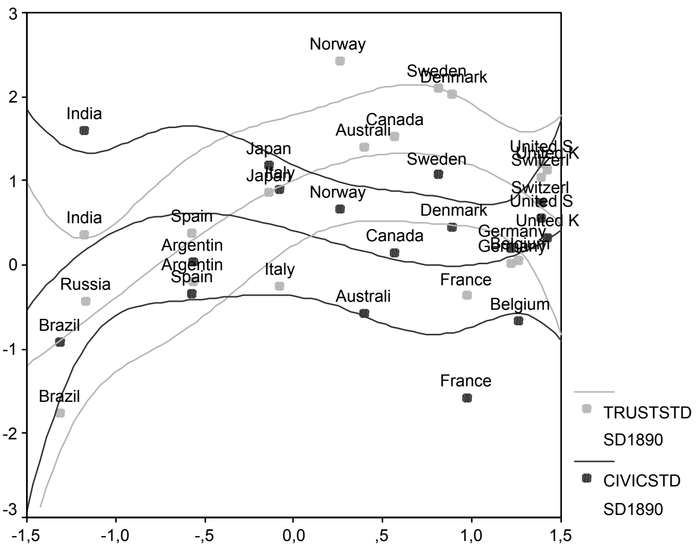
SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Lowess. 50% of points fitted with 3 iterations.

Figure 11 - Historical Evolution 1870-Nowadays. Overlay Scatter Plot with Quadratic Regression Lines



SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Quadratic regression prediction lines.

Figure 12 - Historical Evolution 1890-Nowadays. Overlay Scatter Plot with Cubic Regression Lines



SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Cubic regression prediction lines.

APPENDIX A

LIST OF COUNTRIES

Recoding of variable 'country' into 'countryn' in order to assign a value to every country:

COUNTRY Old Value	COUNTRYN New Value	Value Label
Argentina	1	Argentina
Australia	2	Australia
Belgium	3	Belgium
Brazil	4	Brazil
Burma	5	Burma
Canada	6	Canada
China	7	China
Denmark	8	Denmark
Egypt	9	Egypt
France	10	France
Germany	11	Germany
India	12	India
Italy	13	Italy
Japan	14	Japan
Netherlands	15	Netherlands
New Zealand	16	New Zealand
Norway	17	Norway
Russia	18	Russia
Spain	19	Spain
Sweden	20	Sweden
Switzerland	21	Switzerland
UK	22	UK
US	23	US

APPENDIX B

KEY TO THE VARIABLES IN THIS CHAPTER

(O) means variable in the adaptation of the original Morris and Adelman (1988) databse. (N) means variable in the new database. (F) means definitive final list of variables that go into SDI[16].

O	N	F	Variable Code	Description
x			incgroup	Classification for level of per capita income
	x	x	income	GDP per capita
x			incomgram	Classification for rate of change in per capita income in the past 20 years
	x		incomegr	GDP per capita growth
x	x	x	indutech	Level of development of techniques in industry
x	x	x	indtecgr	Classification for rate of improvement of techniques in industry (lagged, referred to the last 20 years)
x	x	x	agritech	Classification for level of development of techniques in agriculture
x	x		agrtecgr	Classification for rate of improvement of techniques in agriculture
x	x	x	agrilgrp	Classification for percentage of labour force in agriculture
x	x		popxfarm	Population per square kilometres of farmland
x	x	x	intransp	Level of development of inland transportation
x	x	x	transpgr	Classification scheme for rate of improvement of inland transportation (lagged)
x			xgrgroup	Classification scheme for rate of growth of total real exports
	x		trade	Volume of exports relative to GDP
x	x	x	shiftx	Classification for degree of shift in structure of export sector
x	x		indwchan	Classification for direction of change in average real wages in industry
x	x		agrwchan	Classification for direction of change in average real wages or income of the employed agricultural poor

[16] A complete list of variables, including working (and finally non-included) variables,with explanations, sources and construction notes is provided in chapter 2 of the PhD thesis of Marta Felis-Rota.

	x		wages	Real wages
x			popgroup	Classification for total population
	x		pop	Total population in thousands
x			popgrgrp	Classification for rate of population growth in the last 20 years
	x		popgr	Cumulative population growth in the last 20 years (%)
x			immigrp	Classification for net immigration
	x		immi	Net migration, in thousands (immigration with positive sign and emigration with negative sign)
	x		mitopopgr	Rate of net migration over total population growth (%)
x	x	x	lit	Classification of extent of adult literacy
x	x		primedgr	Classification for rate of spread of primary education in the past 20 years
x	x		landtenu	Classification for predominant form of land tenure and holding
x	x		landconc	Classification for concentration of landholdings
x	x	x	landadop	Classification for favourableness of land system to adoption of improvements
x	x	x	urbani	Classification for extent of urbanisation
x	x	x	entrep	Classification for favourableness of attitudes towards entrepreneurship
x			govt	Classification for extent of domestic economic role of government in the past 20 years
	x		govrel	Public expenditure as a percentage of GDP
x	x	x	sociopol	Classification for socioeconomic character of national political leadership in the past 20 years
x		x	represen	Classification for strength of national political institutions in the past 20 years
	x		polity2	Revised Polity variable score, ranging from -10 to 10 (from Polity IV, variable under the same name)
x	x		polstabi	Classification for extent of political stability in the past 20 years
x	x		foreignd	Classification for degree of foreign economic dependence in the past 20 years

x	x	x	colstat	Classification for colonial status
x	x	x	mktdev	Component scores for composite indicator of level of development of market institutions up to the given date
x	x	x	mktdevgr	Component scores for composite indicator of rate of spread of market institutions in the last 20 years
x	x	x	mktdvgrl	Component scores for composite indicator of rate of spread of market institutions in the last 20 years (lagged)

APPENDIX C

THE BASICS OF PRINCIPAL COMPONENTS ANALYSIS

The Calculation of the Principal Components

Suppose we have a dataset with n observations on k variables. We denote these variables by $x_1, x_2, \ldots, x_k$. These data can be arranged in a matrix X with n rows and k columns: $X=(x_1\ x_2\ \ldots\ x_k)$.[17] Each column in this matrix contains the data on a particular variable. For instance, the n observations on the first variable, x_1, are in the first column of X. As we have k variables, we can compute all the sample variances ($Var(x_1)$, $Var(x_2)$, … , $Var(x_k)$) of the variables and all the sample covariances between these variables ($Cov(x_1,x_2)$, $Cov(x_1,x_3)$, etcetera) and arrange them in a variance-covariance-matrix S. This matrix has k rows and k columns. The diagonal of this matrix contains the variances and the other elements of the matrix contain the covariances. For example, the element in the second row and the third column of S contains $Cov(x_2,x_3)$.

The calculation of the Principal components proceeds via the eigenvalues and eigenvectors of the sample covariance matrix S.[18] This matrix can be described by k positive eigenvalues ($\lambda_1, \lambda_2, \ldots, \lambda_k$) and k corresponding eigenvectors ($e_1, e_2, \ldots, e_k$) that have unit length and are orthogonal to each other. Therefore, after calculating the variance-covariance matrix S we can compute (λ_1, e_1), (λ_2, e_2),…, (λ_k, e_k), where we have arranged this sequence in such a way that λ_1 is that largest eigenvalue, λ_2 the one but largest eigenvalue, etcetera. There is a simple relationship between these eigenvalues and eigenvectors of S and the Principal Components.

The relationship is as follows: in the introduction we described the Principal Components as linear combinations of the original variables. A linear combination of the variables $x_1, x_2, \ldots, x_k$ is a weighted sum like

$$P = a_1x_1 + a_2x_2 + \ldots + a_kx_k.$$

[17] In what follows, we assume that the reader is familiar with matrices, and basic matrix multiplication. For more information see Johnson and Wichern (2002), chapter 2.

[18] Eigenvalues and eigenvectors of a matrix are explained briefly in the appendix to this chapter. A more technical discussion can be found in Johnson and Wichern (2002), chapter 2:98-100.

Here, the weights are given by the vector $a=(a_1, a_2, \ldots, a_k)$. Now, it can be shown[19] that the Principal Component that is responsible for the highest variance in the data is the linear combination P_1 with weights equal to the values in the eigenvector of S that corresponds to the largest eigenvalue, i.e. the values in e_1. In short, we obtain the first Principal component by taking the vector a in the formula above equal to the vector e_1. The second Principal Component is the linear combination P_2 with weights equal to the values in the eigenvector of S that corresponds to the one-but-largest eigenvalue, i.e. the values in e_2, and so on and so on. As a result, we will be able to describe the n *observations* we have on the k *variables* of the original data in X as n *factor-scores* on k *Principal Components*. Hence, we effectively transformed an n by k matrix X of data into an n by k matrix $P=(P_1, P_2, \ldots, P_k)$ of factor scores.

At first sight it seems that we have not reduced the data at all. We started with an n by k matrix X and arrive at an n by k matrix P. The difference between X and P, however, is that the columns of P are now independent (orthogonal to each other), each of the columns pointing at a separate independent dimension of variation in the data. Moreover, P is constructed in such a way that its first column (P_1) is the linear combination of the original variables in the data that has the largest variance of all possible linear combinations of the original variables. The second column is the linear combination of the variables in the orginal data that has the second-largest variance, and so on. This fact has important consequences for data-reduction. Indeed, if we remove the last column of P, we know that we remove the direction in the data that has the lowest variance. The resulting columns may well still contain most of the variance that was available in the original data matrix X. Hence, by dropping the last column of P we made sure that at least as possible variation was lost. The same holds for dropping the second-last column and so on. In some cases we can still capture most of the variation in the data by concentrating on only the first or the first two Principal Components.

Geometrical Explanation of Principal Components

Principal Component Analysis is most easily understood graphically in the case where there are only two variables, X1 and X2. Suppose we have a dataset X=(X1 X2) with some observations on these two variables. We can display the variation in these two variables over different observations in a scatterplot, as in Figure 13 below. From this

[19] See Johnson and Wichern (2002), chapter 8.

graph we can see that X1 and X2 are positively correlated, and that X1 has a larger variance than X2. If we were interested in reducing the number of variables (data-reduction) while preserving most of the variation in the original data, we would choose to keep X1 and to drop X2. However, we can do better than that.

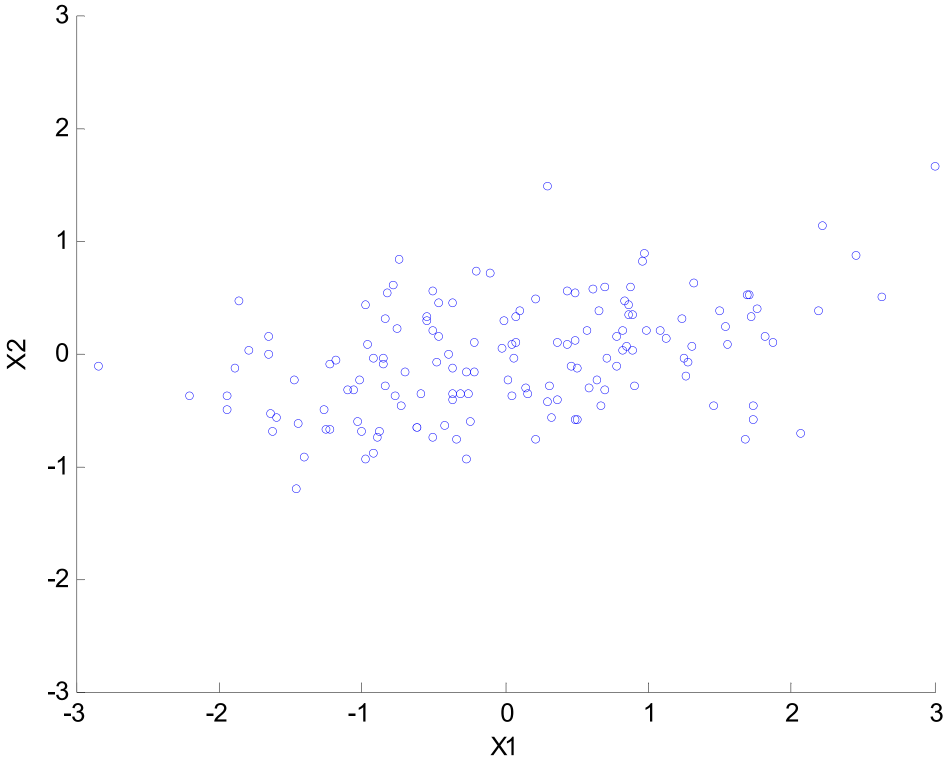


Figure 13 - Scatter plot of X1 and X2.

There is positive covariance between X1 and X2, and the variance of X1 is larger than the variance of X2.

In Figure 14, two dashed lines are added to the figure. The most horizontal line of the two is the axis along which most of the variation in the data is concentrated. The vector in the graph that follows the direction of this axis turns out to be proportional to the eigenvector e_1 (the eigenvector corresponding to the largest eigenvalue λ_1 of S, the variance-covariance matrix of X). The length of the vector drawn in the graph is exactly λ_1.[20] The second axis that is drawn in the graph runs in the direction of e_2. The length of this vector in the graph is λ_2.

The first axis mentioned above is called the first Principal Component. Note that the points that lie on this axis are simply linear combinations of points on the original axes. The second new axis is called the second Principal Component. Note that the two new

[20] A technical note: the associated contour is the contour that describes the set of points that have statistical distance from the center-point equal to 1.

axes are just a rotated version of the original axes. With respect to these two new axes, all of the points in the graph have new coordinates. Each point now has a

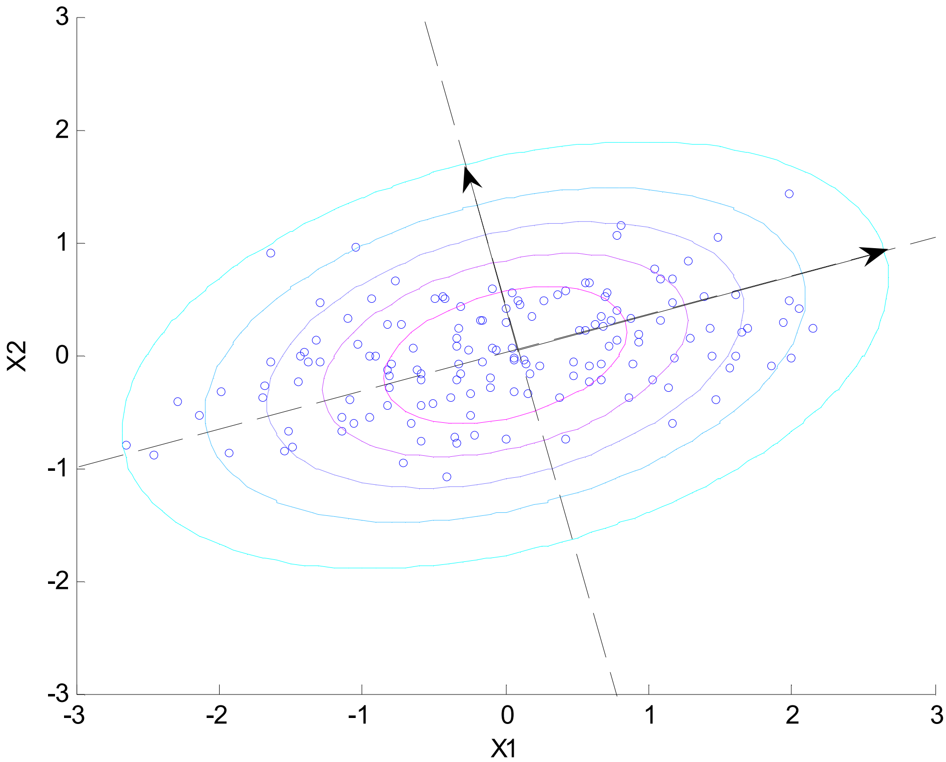


Figure 14 - Scatter Plot with Principal Components.
Let S denote the sample variance Covariance matrix of the variables X1 and X2. The eigenvector corresponding to the largest eigenvalue of S is the more horizontal vector in the graph. Its length is the corresponding eigenvalue. Likewise, the more vertical vector is the other eigenvector of S and its length is the smallest eigenvalue of S. Here, the 'most horizontal' axis is the axis with respect to which the data have the widest range of coordinates (factor scores). Therefore, this axis is the first Principal Component. The more vertical axis is the second Principal Component.

coordinate with respect to the first Principal Component and a coordinate with respect to the second Principal Component. These new coordinates are called factor-scores. The idea now is, that if the correlation between X1 and X2 would have been very high, most of the variation in the data would be due to variation in the factor scores on the first Principal Component. Hence, data-reduction would ideally proceed by keeping the first Principal Component and dropping the second Principal Component. This procedure preserves much more variation than the data-reduction strategy mentioned above, where simply X1 was kept and X2 was dropped.

Conceptual Definition of Eigenvalues and Eigenvectors

There are different types of multiplication possible, depending on what is multiplied by what. The simplest case is premultiplication of a variable x by a scalar λ. It is easy to understand what happens in this case: we obtain λ x, λ times the original variable x. The second type of multiplication we can consider is pre-multiplication of a vector x by a scalar λ. Again, it is easy to understand what will happen to the vector x: we obtain a vector λ x that has the same direction as x, but a length that is λ times the length of the original vector x. The problem occurs in the third type of multiplication that we will consider. It is much harder to understand what happens if we pre-multiply a vector variable x by a matrix A. We obtain the vector Ax, which we will call y.[21] It is typically not clear how the original vector x relates to the resulting vector y. In the abstract, we understand scalar multiplication much better than matrix multiplication.

In order to make the relationship between x and Ax clear, we would really have to calculate the result of multiplying A with x. A priori, we do not have any intuition as to what the result will be. In order to obtain a better intuition of what pre-multiplying with A does to x, it is natural to ask if A maybe acts as scalar multiplication for *some* vectors x. Then, at least for those vectors, we would understand what A does. In other words, we ask ourselves if for a particular matrix A there exists a scalar λ and a corresponding vector x such that $Ax= \lambda x$. If there exists such a scalar and such a vector, we call the scalar λ an eigenvalue of A, and the corresponding vector x the eigenvector of A corresponding to the eigenvalue λ. In particular, to emphasize that this vector x is rather special, we will denote it by e, instead of x. It turns out that for any k by k symmetric matrix A,[22] there actually exist k eigenvalues, (λ_1, λ_2, … , λ_k) and k corresponding eigenvectors (e_1, e_2, … , e_k). These eigenvectors are orthogonal to each other, and can be chosen to have unit length. Details about the calculation of these eigenvalues and eigenvectors can be found in Johnson and Wichern (2002). Most statistical software packages have pre-programmed routines that calculate eigenvalues and eigenvectors for any k by k matrix. The important thing to know about them is that eigenvectors identify the areas (sets of vectors) for which the matrix A works as scalar multiplication. The eigenvalues are the corresponding scalars.

[21] If A is a k by k matrix and x a k by 1 vector, the result y is a k by 1 vector.
[22] Remember that for Principal Component Analysis we are interested in the eigenvalues and eigenvectors of the sample covariance matrix S. This matrix is symmetric.

APPENDIX D

PRE-PCA PRELIMINARY TESTS:

TABLES FOR SECOND BEST ALTERNATIVES

Table 14 – Kaiser-Meyer-Olkin (KMO) sample adequacy scores variable-by-variable, Second Best alternatives

Variables	Original Data Criterion: KMO≥.5		All New Data* Criterion: KMO≥.6
	2nd Round	3rd (and final) round	2nd (and final) round
Income	.7140	.6883	.8544
income growth	.6397	.6341	-
Technique in industry	.7663	.7928	.9074
adoption of new techniques	.8745	.8907	.8931
Technique in agriculture	.8128	.7966	.9359
agricultural labour	.7574	.7511	.8292
population per farmland	.4896	-	-
inland transportation	.8317	.8909	.9378
transportation, growth	.7403	.8078	-
export growth group	-	-	-
Trade	-	-	-
shift in export structure	.7833	.8362	.9079
industrial wage change	.6503	.7454	-
agricultural wage change	.6588	.6674	-
Wages	-	-	-
population	.4850	-	-
population growth	-	-	-
immigration	-	-	-
Migration to pop. growth	-	-	-
Literacy	.8056	.7354	.8690
primary education growth	.6565	.6059	-
form of land tenure	.6260	.5573	-
land concentration	-	-	-
land adoption	.7227	.7141	.9240
urbanisation	.8963	.8715	.9240

entrepreneurship	.8856	.8808	.9471
role of government	.5757	.6140	-
socio-politics	.8768	.8778	.9370
representativeness	.8176	.7891	-
polity 2	-	-	-
political stability	.6482	.6354	.7800
foreign dependency	.8160	.8164	.8479
colonial status	.6930	.6820	.7741
market development	.8347	.8448	.9060
market development growth	.8037	.7882	.8588
mkt develpt growth, lagged	.7728	.7570	.8551
Number of variables	**28**	**26**	**18**
Overall sampling adequacy	**.7661**	**.7694**	**.8859**

KMO measures sampling adequacy, in a scale from 0 to 1.
* All new data refers to the new database keeping goverel and Polity2, which have been replaced by the old variables due to better performance in the first best option.

Table 15 - Squared Multiple Correlations (SMC) of variables with all other variables Second Best alternatives

Variables	**Original Data Criterion: KMO≥.5**		**All New Data* Criterion: KMO≥.6**
	2nd Round	3rd (and final) round	2nd (and final) round
Income	.9796	.9793	.9341
income growth	.8749	.8503	-
Technique in industry	.9650	.9478	.9180
adoption of new techniques	.9189	.9055	.8592
Technique in agriculture	.9586	.9548	.9048
agricultural labour	.9432	.9411	.9038
population per farmland	.8279	-	-
inland transportation	.9395	.9072	.8610
transportation, growth	.7539	.7135	-
export growth group	-	-	-
Trade	-	-	-
shift in export structure	.8415	.8038	.6853
industrial wage change	.7833	.7163	-
agricultural wage change	.6009	.5864	-
Wages	-	-	-
population	.8841	-	-
population growth	-	-	-
immigration	-	-	-
Migration to pop. growth	-	-	-
Literacy	.9736	.9684	.9165
primary education growth	.7783	.7756	-
form of land tenure	.9000	.8735	-
land concentration	-	-	-
land adoption	.9235	.9122	.7664
urbanisation	.8596	.8583	.7749
entrepreneurship	.9546	.9469	.8979
role of government	.6916	.6552	-
socio-politics	.9021	.8912	.8536
representativeness	.9390	.9368	-

polity 2	-	-	-
political stability	.9407	.9264	.8285
foreign dependency	.8850	.8590	.7502
colonial status	.9274	.9208	.8265
market development	.9692	.9657	.9394
market development growth	.9982	.9981	.9973
mkt develpt growth, lagged	.9982	.9981	.9970
Number of variables	**28**	**26**	**18**
Overall sampling adequacy	**.7661**	**.7694**	**.8859**

* All new data refers to the new database keeping goverel and Polity2, which have been replaced by the old variables due to better performance in the first best option.

REFERENCES

Abramovitz, Moses, 1986, 'Catching Up, Forgoing Ahead, and Falling Behind', *The Journal of Economic History*, 46 (2):385-406.

Acemoglu, Daron, Simon Johnson, and James A. Robinson, 2001, 'The Colonial Origins of Comparative Development: an Empirical Investigation', *American Economic Review*, 91 (5): 1369-1401.

Adelman, Irma, and Cynthia T. Morris, 1965, 'A Factor Analysis of the Interrelationship Between Social and Political Variables and Per Capita Gross National Product', *The Quarterly Journal of Economics*, 79 (4): 555-578.

——— , 1967, *Society, Politics, and Economic Development: A Quantitative Approach*, John Hopkins Press (Baltimore).

——— , 1978, 'Patterns of Market Expansion in the Nineteenth Century: A Quantitative Study', in George Dalton, ed., *Research in Economic Anthropology*, JAI Press (Greenwich, Conn.): 231-324.

Alesina, Alberto and Eliana La Ferrara, 2000, 'Who Trusts Others?', CEPR Discussion Paper Series, 2646.

Algan, Yann, and Pierre Cahuc, 2010, 'Inherited Trust and Growth.' *American Economic Review,* 100(5): 2060-92.

Allen, Robert, 2004, 'Progres and Poverty in Early Modern Europe', Seminar series at the IHR (London).

Anderson, Eugene Newton, and Pauline Safford R. Anderson, 1967, *Political Institutions and Social Change in Continental Europe in the Nineteenth Century*, University of California Press (Berkeley and Los Angeles).

Anderson, Perry, 1974, *Lineages of the Absolutist State*, New Left Books (London).

Barro, Robert J. and Jong-Wha Lee, 'International Measures of Schooling Years and Schooling Quality', *American Economic Review*, Papers and Proceedings 86 (2), 1996, pp. 218-23. (Updated dataset at: http://www.worldbank.org/research/growth/ddbarle2.htm)

Bartlett, 1950, 'Test of Significance in Factor Analysis', *British Journal of Psychology*, 3:77-85.

Blyth, Mark, 2002, *Great Transformations: Economic Ideas and Institutional Change in the Twentieth Century*, Cambridge University Press (Cambridge,UK).

Bohnet, Iris, and Steffen Huck, 2004, ‘Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change.’ *American Economic Review,* 94(2): 362-366.

Cagé, Julia, and Valeria Rueda, 2016, ‘The Long-Term Effects of the Printing Press in Sub-Saharan Africa.’ *American Economic Journal: Applied Economics,* 8(3): 69-99.

Cairncross, Alexander Kirkland, 1962, *Factors in Economic Development*, Allen & Unwin (London).

Cohen, Daniel and Marcelo Soto, ‘Growth and Human Capital: Good Data, Good Results’, CEPR discussion paper 3025, (2001).

Davis, J. A. and T. W. Smith, 1994, *General Social Surverys, 1972-1994*, produced by the National Opinion Research Center (Chicago), distributed by The Roper Center for Public Opinion Research (Storrs, Connecticut).

De la Fuente, Angel and Rafael Doménech, ‘Human Capital in Growth Regression: How Much Difference Does Quality Data Make’, CEPR discussion paper 2466, (2000).

Deutch, Karl W., 1953, *Nationalism and Social Communication: An Inquiry Into the Foundations of Nationality*, MIT Technology Press and John Wiley & Sons (Cambridge, MA).

Diana, Giancarlo, and Chiara Tommasi, 2002, ‘Cross-validation Methods in Principal Components Analysis: A Comparison”, *Statistical Methods & Applications*, 11:71-82.

DiPasquale, Denise, and Edward L. Glaeser, 1999, ’Incentives and Social Capital: Are Homeowners Better Citizens?’ *Journal of Urban Economics*, 45: 354-384.

Eckel, Catherine C., and Ragan Petrie, 2011, ‘Face Value.’ *American Economic Review,* 101(4): 1497-1513.

Falk, Armin, and Michael Kosfeld, 2006, ‘The Hidden Costs of Control.’ *American Economic Review,* 96(5): 1611-1630.

Felis-Rota, Marta, 2008, ‘Social Capital in Historical Perspective: A Principal Components Approach to International Measurement and its Economic Implications, 1870-2000’, PhD thesis, Department of Economic History, London School of Economics and Political Science, London, UK.

——— , 2011, ‘Un paseo por la literatura sobre capital social desde una perspectiva económica’, *Revista de Historia Económica - Journal of Iberian and Latin American Economic History*, Cambridge University Press, vol. 29 (3): 461-475.

Furnival, John S., 1931, *An Introduction to the Political Economy of Burma*, Burma Book Club (Rangoon).

Gerschenkron, Alexander, 1962, *Economic Backwardness in Historical Perspective: A Book of Essays*, Harvard University Press (Cambridge, MA).

Groningen Growth and Development Centre (GGDC) and The Conference Board, *Total Economy Database*, January 2005, http://www.ggdc.net

Grosjean, Pauline, 2011, 'The Weight of History on European Cultural Integration: A Gravity Approach.' *American Economic Review,* 101(3): 504-08.

Hanley, Susan Bell, and Yamamura, Kozo, 1977, *Economic and Demographic Change in Preindustrial Japan, 1600-1868*, Princeton University Press (Princeton).

Hargreaves Heap, Shaun P., and Daniel John Zizzo, 2009, 'The Value of Groups.' *American Economic Review,* 99(1): 295-323.

Harriss, John, 2002, *Depoliticizing Development: The World Bank and Social Capital*, Anthem Press (London).

Helleiner, Eric, 1994, *States and the Reemergence of Global Finance*, Cornell University Press (Ithaca).

Heston, Alan, Robert Summers and Bettina Aten, *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

Jackson, Matthew O., Tomas Rodriguez-Barraquer, and Xu Tan, 2012, Social Capital and Social Quilts: Network Patterns of Favor Exchange.' *American Economic Review,* 102(5): 1857-97.

Johnson, Richard A. and Dean W. Wichern, 2002, *Applied Multivariate Statistical Analysis*, Prentice-Hall (London).

Kaiser, Henry F., 1970, 'A Second Generation Little Jiffy', *Psychometrika*, 35 (4, December):401-415.

Kirshner, Jonathan, 1999, 'Keynes, Capital Mobility and the Crisis of Embedded Liberalism', *Review of International Political Economy* 6 (3, Autumn):313-337.

Knack, Stephen, and Philip Keefer, 1997, 'Does Social Capital Have an Economic Payoff? A Cross-Country Investigation', *The Quarterly Journal of Economics*, CXII:1251-1288.

Krueger, Alan B. and Mikael Lindahl, 'Education for Growth: Why and For Whom', NBER working paper 7591 (2000).

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny, 1997, 'Trust in Large Organisations', *American Economic Review Papers and Proceedings*, 87: 333-338.

Lee, William Robert, 1977, *Population Growth, Economic Development, and Social Change in Bavaria, 1750-1859*, Arno Press (New York).

Lindert, Peter, 2003, "Voice and Growth: Was Churchill Right?" *Journal of Economic History* 63, 2 (June): 315-350.

MacNamara, Kathleen R., 1998, *The Currency of Ideas: Monetary Politics and the European Union*, Cornell University Press (Ithaca): esp. 54-5, 82-7.

Maddison, Angus, 2003, *The World Economy: Historical Statistics*, OECD Development Centre (Paris).

Manli, Bryan F. J., 1991, *Multivariate Statistical Methods: A Primer*, Chapman and Hall (London).

Milward, Alan S., and S. B. Saul, 1979, The *Economic Development of Continental Europe, 1780-1870*, 2nd ed, George Allen & Unwin, (London).

Moore, Barrington, Jr., 1966, *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*, Beacon Press (Boston).

Morris, Cynthia Taft and Irma Adelman, 1988, *Comparative Patterns of Economic Development, 180-1914*, The John Hopkins University Press (Baltimore and London).

Moscona, Jacob, Nathan Nunn, and James A. Robinson, 2017, 'Keeping It in the Family: Lineage Organization and the Scope of Trust in Sub-Saharan Africa.' *American Economic Review,* 107(5): 565-71.

Mukerjee, Tapan, 1972, 'Theory of Economic Drain: Impact of British Rule on the Indian Economy, 1840-1900' in *Economic Imperialism: A Book of Readings.*

North, Douglass, 1981, *Structure and Change in Economic History*, Norton, New York.

North, Douglass Cecil, 1990, *Institutions, Institutional Change and Economic Performance*, Cambridge University Press (Cambridge, UK).

Nunn, Nathan, and Leonard Wantchekon, 2011, 'The Slave Trade and the Origins of Mistrust in Africa.' *American Economic Review,* 101(7): 3221-52.

Ostrom, Elinor, 2011, 'Reflections on "Some Unsettled Problems of Irrigation".' *American Economic Review*, 101(1): 49-63.

Polanyi, Karl, 1944, *The Great Transformation: The Policial and Economic Origins of Our Time*, Farrar & Rinehart (New York), reprinted in 1957 by Beacon (Boston).

Pollard, Sidney, 1988, Review of Cynthia Taft Morris and Irma Adelman, 1988, *Comparative Patterns of Economic Development, 180-1914*, The John Hopkins University Press (Baltimore and London), *The Economic History Review* 41 (4):668-670.

Putnam, Robert D., *Making Democracy Work: Civic Traditions in Modern Italy*, (Princeton University Press, 1993).

—— , *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster, (New York, 2000).

Read, Ronald C., 1972, *A Mathematical Background for Economists and Social Scientists*, Prentice-Hall (Englewood Cliffs, NJ).

Ruggie, John Gerald, 1982, 'International Regimes, Transactions, and Change: Embedded Liberalism in the Postwar Economic Order', *International Organization* 36 (2, Spring):379-415.

Schechter, Laura, 2007, 'Theft, Gift-Giving, and Trustworthiness: Honesty Is Its Own Reward in Rural Paraguay.' *American Economic Review,* 97(5): 1560-1582.

Simon, Carl P. and Blume, Lawrence, 1994, *Mathematics for Economists*, W. W. Norton & Company (New York and London).

Sliwka, Dirk, 2007, 'Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes.' *American Economic Review,* 97(3): 999-1012.

STATA User's Manual.

Stevenson, Betsey, and Justin Wolfers, 2011, 'Trust in Public Institutions over the Business Cycle', *American Economic Review,* 101(3): 281-87.

Tabellini, Guido, 2007, 'Culture and Institutions: Economic Development in the regions of Europe', http://www.igier.uni-bocconi.it/whos.php?vedi=327&tbn=albero&id_doc=177 .

Temple, Jonathan, and Paul A. Johnson, 'Social Capability and Economic Growth', *The Quarterly Journal of Economics* 113 (3), 1998, pp. 965-990.

Warren, Bill, 1980, *Imperialism: Pioneer of Capitalism*, New Left Books (London).

Wuensch, Karl L., 2005, Principal Components Analysis – SPSS, East Carolina University, Department of Psychology, http://core.ecu.edu/psyc/wuenschk/MV/FA/PCA-SPSS.doc

Zak, Paul, J., Karla Borja, William T. Matzner, and Robert Kurzban, 2005, 'The Neuroeconomics of Distrust: Sex Differences in Behavior and Physiology.' *American Economic Review,* 95(2): 360-363.