Working Papers No. 103/07

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

Marta Felis Rota

July 2007

Department of Economic History
London School of Economics
Houghton Street
London, WC2A 2AE

Tel: +44 (0) 20 7955 7860
Fax: +44 (0) 20 7955 7730

Is Social Capital Persistent? Comparative Measurement In The Nineteenth And Twentieth Centuries

Marta Felis Rota

Abstract

Recently, there has been a growing interest in social capital and in the difficulties related to its measurement. In this paper, we propose to measure social capital by means of principal components analysis. Then, we present the first available international social capital estimates for the nineteenth century. Our 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 allows us to study the evolution of social capital between the nineteenth and twentieth centuries. We can make observations about the persistence of the social capital indicator, discovering some exceptional trajectories. In the very long run, we find a significant change in the relative position of the European countries.

1. Introduction To Social Capital

The importance of the structure of society and of the interactions among its agents has always been recognised as relevant factors in the performance of economies. The social links between citizens can be encompassed under the concept of 'social capital' (DiPasquale and Glaeser, 1999: 355). In the words of the sociologist Robert Putnam, social capital 'refers to features of social organisation, such as trust, norms, and networks, that can improve the efficiency of society by facilitating co-ordinated actions' (Putnam, 1993: 167). Margaret Keck and Kathryn Sikkink define networks as '"forms of organization characterized by voluntary, reciprocal and horizontal patterns of communication and exchange" [Keck and Sikkink, 1998:8]. These networks may include not only conventional NGOs, but also local social movements, foundations, the media, churches, trade unions, consumer organizations, intellectuals, parts of regional and international intergovernmental organizations, and parts of the executive and/or parliamentary branches of governments' (Brecher, Costello, and Smith, 2000:83). In Woolcock's (1998) opinion, however, these features are not social capital, but rather the consequences thereof. He adds that the concept can be read in several ways.

Woolcock argues that there are four dimensions of social capital: the size and scope of horizontal associations; social integration, the nature of social ties within communities; the relationship between civil society and the state; and, the quality of governing institutions. However, he agrees that it should be collapsed and understood as a single variable. Fine cites a statement made by Woolcock on the World Bank's email discussion site on social capital:

> 'Several critics, not without justification, have voiced their concern that collapsing an entire discipline into a single variable (especially one with such economic overtones) is a travesty, but there are others who are pleased that mainstream sociological ideas are finally being given their due at the highest levels'(Fine, 2001:139).

As a way of summarising the various proposed definitions, social capital can be understood in the abstract as the quality or health of civil society.

The concept of social capital is nothing new. Putman acknowledges the fact that the concept was described well before his influencing 1993 book. 'The term *social capital* turns out to have been independently invented at least six times over the twentieth century, each time to call attention to the ways in which our lives are made more productive by social ties' (Putnam 2000:19). However, one has to bear in mind that some of the claims on social capital are retrospective claims, which affirm that ‘writers (…) really were talking about social capital when they thought they were writing about something else’ (Harriss, 2001:75).

The ‘six times’ Putnam acknowledges are the following, starting by L. J. Hanifan in 1916, who studied the importance of the community involvement for the well-functioning of schools. A forty years gap separates Hanifan from the book co-authored by John Seely, Alexander Sim, and Elizabeth Loosley in 1956, *Crestwood Heights: A Study of the Culture of Suburban Life*. Keeping on the line of the study of urban societies and neighbourliness, Jane Jacobs presented her work on 1961. It was not until the mid-seventies when Glenn Loury rediscovered social capital once more, writing about the determinants of income differences between members of different ethnic groups. Loury’s discovery attracted the attention of the most influential of the six, James Coleman, who defines social capital according to Loury’s vision as ‘the set of resources that inhere in family relations and in community social organisation and that are useful for the cognitive or social development of a child or young person’ (Coleman, 1990: 300). The sixth according to Putnam was the Marxist theorist Pierre Bordieu who underlined the relevance of social networks in the 1980s.

There are a variety of scattered quantitative studies on the effect of social capital on economic growth and development; see Putnam (1993 and 2000), Temple and Johnson (1998), Knack and Keefer (1997), DiPasquale and Glaeser (1999), and Alesina and La Ferrara (2000) among others. Studies conducted by Easterly and Levine (1997), and Temple and Johnson (1998) aimed at determining the macroeconomic consequences of social arrangements. They indicate that some disappointing policy outcomes have their roots in the nature of societies. La Porta et al. (1997) find some effect of social capability on growth for the period 1970-93; Knack and Keefer (1997) focus on the period 1980-92, and find a stronger link: If the level of trust increases by 10 percent, growth is, on average, 0.8 percent higher per year. This is a very considerable effect. Therefore, it stands clear from the existing research that social capital has some economic consequences, and it makes the difference as far as economic growth is concerned.

Inquiring about the mechanism behind the economic effects of social capital, one important line of research is its role in making institutions work well, through a reduction of transaction costs; '(that is, the costs of monitoring and enforcing agreements), and thus in enabling agents more efficiently to surmount problems of opportunism and shirking' (Putnam, 93:166). Harriss recalls Coleman's example of traders in a Cairo market 'who share information about customers – he shows how the reciprocity and trust which may be an aspect of social relations are of value because they help to reduce many of the costs of transaction, through the communication of information and the kinds of insurance that are created in social networks' (Harriss, 2001:5). This could potentially explain why institutions work better in some places than in others. Institutions, as technology, may be replicated, but yet they achieve different levels of efficiency in different places. This depends among other factors upon the social capability of the society in question. Thus, one of the produces of social capital is that it helps institutions work well.

> According to Douglass North,
> 'It is much easier to describe and be precise about the formal rules that societies devise than describe and be precise about the informal ways by which human beings have structured human interaction. But although they defy (…) neat specification and it is extremely difficult to develop unambiguous tests of their significance, they are important' (North, 1990:36).

In this passage, Douglass North points out the importance of informal organisations and *modus operandi* of human activity. However, he does not take this statement further; he does not explore the phenomenon any further.

The duty of the social scientist is to analyse the social phenomena. No matter how imperfect the measurement of social phenomena is, these should be subject to non-quantitative as well as quantitative analysis. So, even if the quantitative analysis procedure is far from optimal, it might still reveal some interesting stylised facts or links that invite us to look for further detailed evidence and improve the procedure itself. It is the case, indeed of the social arrangements. In other words, taking North more seriously than he did himself could constitute a contribution to the theory of institutions and growth.

Taking North's forceful statement together with the argumentation in the previous paragraphs, we can start by concluding that there is a strong enough theoretical basis which encourages us to pursue further quantitative study in the following direction: All the cited works have a contemporary perspective. To the extent of my knowledge, studying social capital from a historical quantitative perspective has never been done before, and would deserve some effort in order to win a place in the literature as an additional contribution.

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. This paper presents a new social capital indicator for the late nineteenth century; this is done in section II. Section III compares the newly created measure with other social capital measurement alternatives: Section III.a compares three different measurement alternatives for the second half of the twentieth century, while section III.b is an inter-temporal comparison between these and the newly created series for the late twentieth century.

2. Constructing A Social Capital Indicator For The Nineteenth Century

The first social capital related quantitative study was undertaken in the 1960's by Adelman and Morris, (see 1965 and 1967 publications). They started from the idea that some societies may be inherently more suited to entrepreneurship and economic development. Societies evolve in a variety of dimensions, such as family relations, the extent of communications, the importance of the middle class, and social mobility. Departing from this idea, they used a series of socio-economic variables in a cross-country historical investigation in order to explain different patterns of development. In fact, moving to more recent studies, Temple and Johnson (1998) proposed the use of the social development index elaborated by Adelman and Morris for the 1960´s as a proxy for social capital. Other empirical work has focused on international differences in the level of trust, and of civic co-operation. The measure of social capital that Putnam uses in his book *Bowling Alone* 'is constructed – like that of "civic community" in *Making Democracy Work* [his previous book]—by combining a number of indices (of "community organizational life", "engagement in public affaires", "community volunteerism", "informal socializing" and of "social trust") which are themselves highly intercorrelated' (Harriss, 2001:55). Papers like that of Alesina and La Ferrara (2000) focus more narrowly on the level of trust for the United States in the period 1974-1994. The level of trust is established by the General Social Surveys (Davis and Smith, 1994), which asked to respondents whether they think that most people can be trusted. Other alternatives are experiments where civic virtues are tested by, for example, "losing" a wallet with fifty dollars in different cities and counting the percentage of wallets being returned (Knack and Keefer, 1997). In their paper 'Are Homeowners Better Citizens?', DiPasquale and Glaeser (1999) point at the importance of being homeowner for the involvement in the social initiatives aiming at the well-being of the local community. Thus, it shows clear that many proposals are available, but yet none enjoys general acceptance.

2.A) 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[1]. 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

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

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.

We reconstructed a similar database to that which Adelman and Morris built for the period 1850 to 1914, and then use it to construct a Social Development Index (SDI) for the nineteenth century. Many of the variables are extracted from their 1988 book and re-codified for convenience. On the other hand, 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.[2] Re-codification consisted of transferring letter codification (alphabetic 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. Then we performed a principal components analysis.

2.B) Results: Principal Components Analysis

The principal components analysis is a data reduction technique that will help us to reduce a large set of variables into a single index. In general, the relationships between a large set of variables can be summarised into a small set of principal components.

The principal components are newly generated variables, obtained from linear combinations of the original variables. The higher the correlation between the original variables, the smaller the number of principal components extracted; and, thus, the more effective the data reduction. This characteristic of the Principal Component Analysis will

[2] Please, refer to the PhD thesis of Marta Felis Rota for a detailed discussion of the Adelman and Morris (1988) database.

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 correlations. It is for this reason that we consider this technique particularly adequate for our purposes[3].

Tables 2.a and 2.b show the total variance explained by the first eight principal components in 1870 and 1890 respectively. Remarkably, the analysis reveals that the first principal component alone explains more than forty 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 (figures 2.a and 2.b 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.

The main principal component of the analysis of the 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 XXXX, standing for social development index in year XXXX. Series for years 1870 and 1890 are presented here; though I expect to expand the list of years to 1850 and 1910. Most of the necessary data for the latter are already available, but the way lagged

[3] An intuitive explanation of the principal components technique can be found in appendix C.

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 2.3 and 2.4. The component scores coefficients shown in table 2.3 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 2.4 shows the first principal component scores obtained for every country and year in the sample. These are available for 22 countries (all countries in the database except China, for which some key variables are lacking). In this way, we have a score for Argentina 1870, another for Argentina 1890, and so on. Scores for China could be obtained by bringing additional information from other sources to complete the missing data. The same is true for other countries shown with 'n/a' (Burma and The Netherlands for some years).

Two additional columns have been added to table 2.4 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 2.4 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. Japan is the country that improved the most in the SDI in absolute terms. The United Kingdom is the only country that appears with a negative sign in the last column. Very close to the United Kingdom but still exhibiting a positive sign are Egypt and France, in this order. These 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 2.4. 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 2.4.

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. Despite the fact that both datasets were constructed by the same authors, samples of countries for the nineteenth and the twentieth century hardly overlap. This results in a small number of countries being in the two samples.

Figures 2.5 and 2.6 depict the historical evolution of social developments over long periods of time. Figure 2.5 represents the change in scores over almost a century, from 1870 to 1960. Figure 2.6 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.

3. What Can We Learn From The Relationship Of The Sdi With Other Measures Of Social Capital?

Several proposals for capturing social capital in the empirics 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.

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?

3.A) Contemporary Comparison Of Alternatives

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 incorporated, using always the most recent observation available[4]. 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.

Table 2.1 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 2.2). Later we will find *non-linear* 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 2.1 is a joint three-dimensional graphical representation of the three proposed indicators. SOCDEV

[4] The last round of Surveys dating 2002 is pending to be incorporated.

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 2.2 and 2.3 (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 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 2.2 and 2.3 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 2.3 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

[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.

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).

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 2.1 to 2.3 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.

3.B) Inter-temporal Comparisons: Nineteenth And Twentieth Centuries

In the previous section, we underlined the basic features of the relationship between social development and other late twentieth century social capital measurement alternatives. In this section, we carry out an inter-temporal comparison of the new social development index for the late nineteenth century and other available measures for the late twentieth century.

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 III.a. In figures 2.7 to 2.12 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 2.7 shows quite different from its twentieth century counterpart (figure 2.2). 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 2.8 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 SDI. This is true both for 1870 and 1890. Thus, there has been a change in the relative position of the level of trust with respect the civic engagement, which appears to be more stable or equalised across countries. The former reaches higher values for the nineteenth century and lower values for the twentieth century, when compared to the more constant index for civic engagement. In a few words, the general level of trust appears to be higher in the late nineteenth century than nowadays. The late-nineteenth century opening of the international economy known as the first wave of globalisation could well be linked to this phenomenon. But throwing such a thrilling hypothesis requires a more careful investigation.

Figures 2.9 and 2.10 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. Here we can observe that trust is more volatile across countries than civic engagement is, and has tended to converge less. 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 is nowadays 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[6].

The tendency of civic engagement levels to converge across countries is confirmed in figures 2.11 and 2.12. 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 convergence tends to be constant, and 2) parabolic layout of trust points.

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. The 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

[6] Economic growth implications are not tested in this paper. For a formal test, see the PhD thesis of Marta Felis Rota.

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 2.12). 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, it is countries in the middle of the spectrum those which have improved the most. 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. However, outliers depart from the trend, doing nothing but confirm that the results are historically consistent. 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 pre-occupating phenomenon, which surprises as counterintuitive, needs a more detailed consideration. 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.

4. 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.

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 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 2.B) Results: Principal Components Analysis

Table 2.A - Total Variance Explained For 1870

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	15,455	42,931	42,931	15,455	42,931	42,931
2	4,402	12,228	55,158	4,402	12,228	55,158
3	3,337	9,270	64,428	3,337	9,270	64,428
4	2,296	6,377	70,805	2,296	6,377	70,805
5	1,773	4,925	75,730	1,773	4,925	75,730
6	1,654	4,593	80,324	1,654	4,593	80,324
7	1,546	4,296	84,619	1,546	4,296	84,619
8	1,171	3,253	87,872	1,171	3,253	87,872

Extraction Method: Principal Component Analysis.

Figure 2.a - Scree Plot for 1870

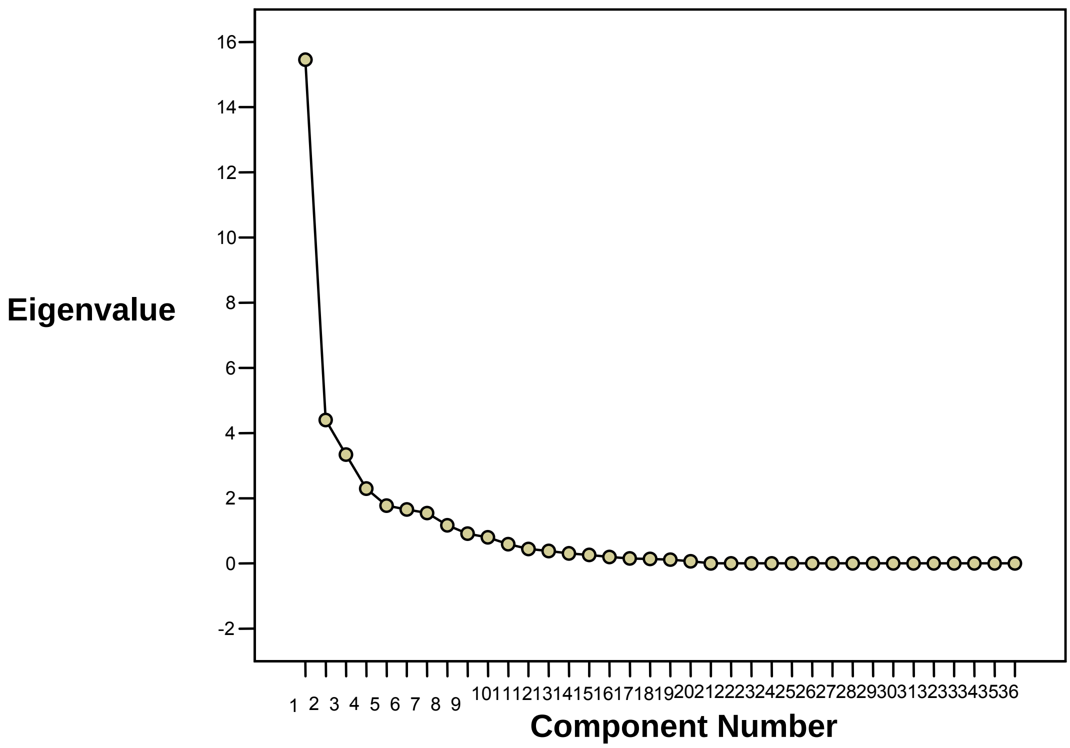


Table 2.B - Total Variance Explained For 1890

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	14,634	40,649	40,649	14,634	40,649	40,649
2	4,368	12,132	52,781	4,368	12,132	52,781
3	3,409	9,468	62,249	3,409	9,468	62,249
4	2,976	8,268	70,517	2,976	8,268	70,517
5	2,088	5,799	76,316	2,088	5,799	76,316
6	1,520	4,223	80,539	1,520	4,223	80,539
7	1,365	3,792	84,331	1,365	3,792	84,331
8	1,170	3,250	87,581	1,170	3,250	87,581

Extraction Method: Principal Component Analysis.

Figure 2.b - Scree Plot for 1890

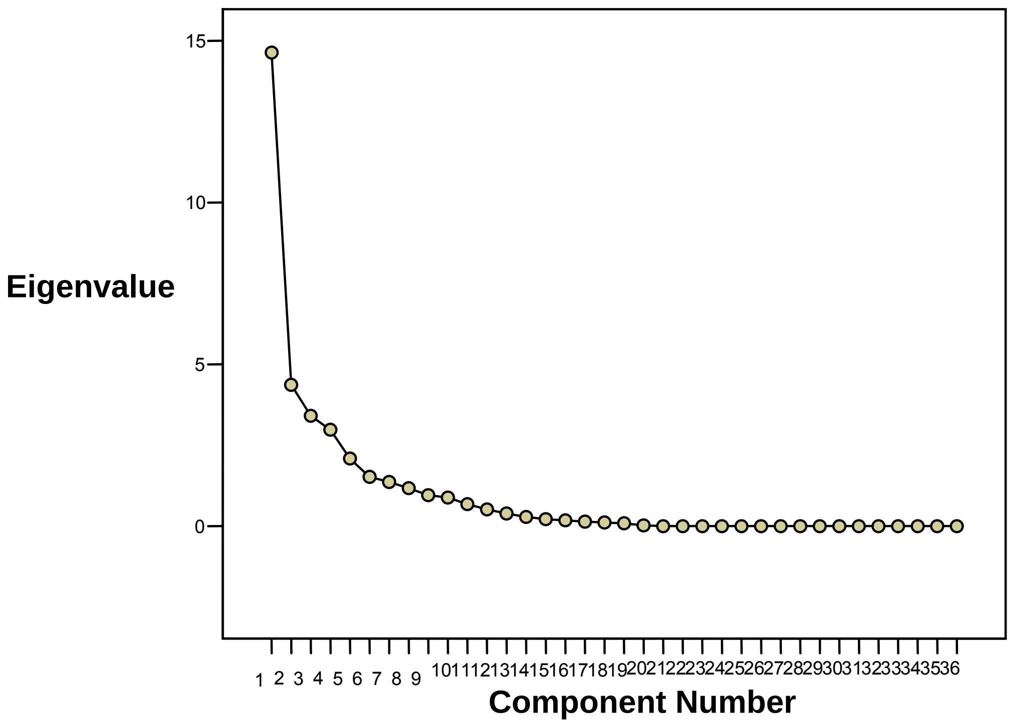


Table 2.3 - First Principal Component Score Coefficients

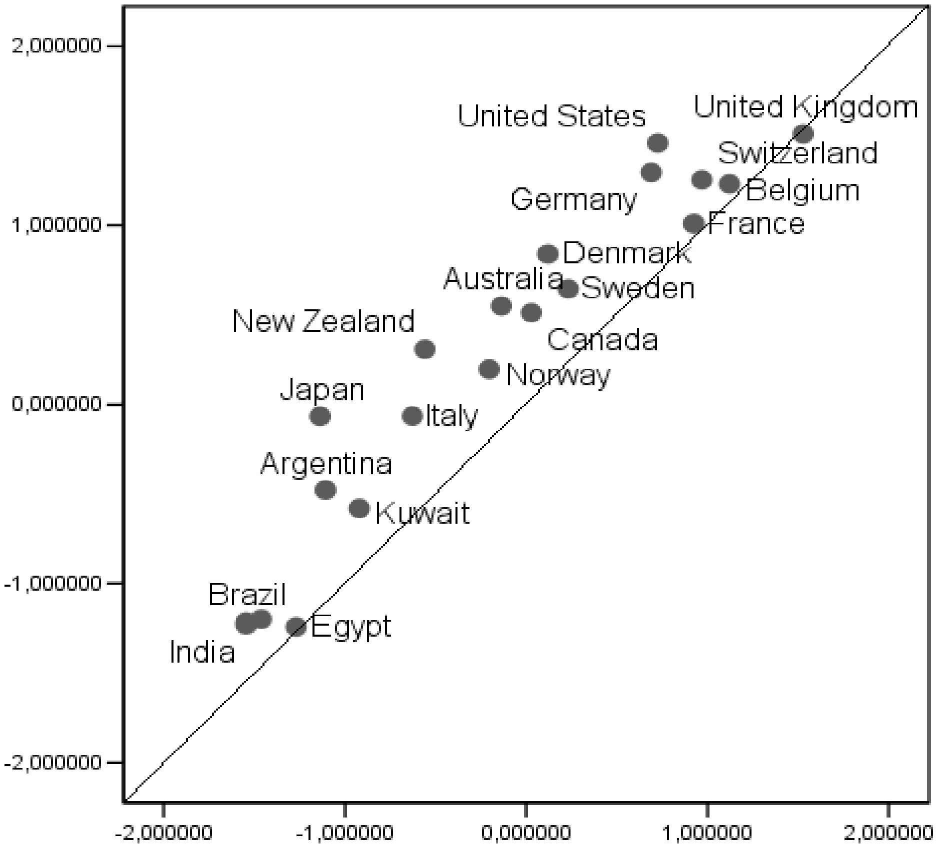
variable	score 1870	score 1890	variable	score 1870	score 1890
AGRICLAB	-,043	-,052	LANDCONC	-,022	-,007
AGRILGRP	-,046	-,049	LANDTENU	,037	,039
AGRITECH	,059	,061	LIT	,052	,062
AGRTECGR	,056	,049	MKTDEV	,060	,063
AGRWCHAN	,006	,028	MKTDEVGR	,058	,050
COLSTAT	,037	,037	MKTDVGRL	,058	,047
ENTREP	,057	,056	POLSTABI	,041	,054
FARMLAND	-,023	-,030	POPGRGRP	-,013	-,005
FOREIGND	,045	,057	POPGROUP	-,009	-,014
GOVT	,024	,004	POPXFARM	,025	,012
IMMIGRP	-,025	-,014	PRIMEDGR	,029	,021
INCGROUP	,051	,060	REPRESEN	,055	,057
INCOMEGR	,033	,044	SHIFTX	,042	,041
INDTECGR	,056	,049	SOCIOPOL	,054	,054
INDUTECH	,055	,058	TOTALAND	-,020	-,026
INDWCHAN	,028	,034	TRANSPGR	,034	,023
INTRANSP	,049	,056	URBANI	,043	,040
LANDADOP	,051	,056	XGRGROUP	,027	,014

Note: The key to the variables can be found in the Appendix B.

Table 2.4 – First Principal Component Scores

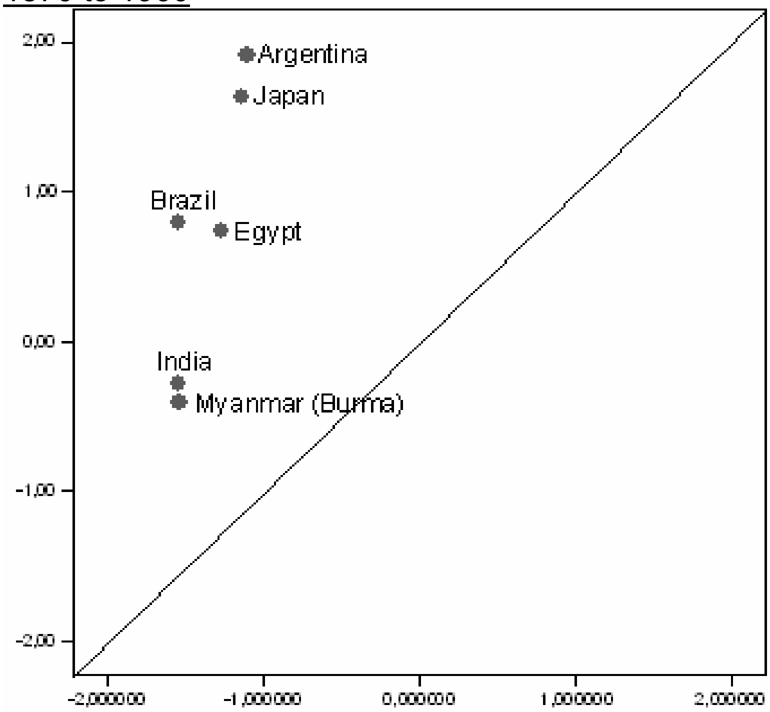
Country	SDI 1870	SDI 1890	Change SDI 1890 –SDI 1870	Sign of Change
Argentina	-1,108	-0,477	0,631	+
Australia	-0,137	0,550	0,687	+
Belgium	1,122	1,231	0,109	+
Brazil	-1,545	-1,214	0,330	+
Burma	-1,538	n/a	n/a	n/a
Canada	0,029	0,513	0,484	+
Denmark	0,119	0,841	0,722	+
Egypt	-1,269	-1,242	0,028	+
France	0,925	1,010	0,085	+
Germany	0,690	1,296	0,606	+
India	-1,545	-1,230	0,315	+
Italy	-0,629	-0,065	0,564	+
Japan	-1,137	-0,066	1,071	+
Netherlands	n/a	0,974	n/a	n/a
New Zealand	-0,560	0,309	0,868	+
Norway	-0,205	0,197	0,402	+
Russia	-1,463	-1,198	0,265	+
Spain	-0,921	-0,580	0,342	+
Sweden	0,232	0,646	0,414	+
Switzerland	0,968	1,254	0,286	+
United Kingdom	1,529	1,511	-0,018	-
United States	0,726	1,459	0,734	+

Figure 2.4 - Scatter Plot For The New Social Development Index: 1870 Against 1890



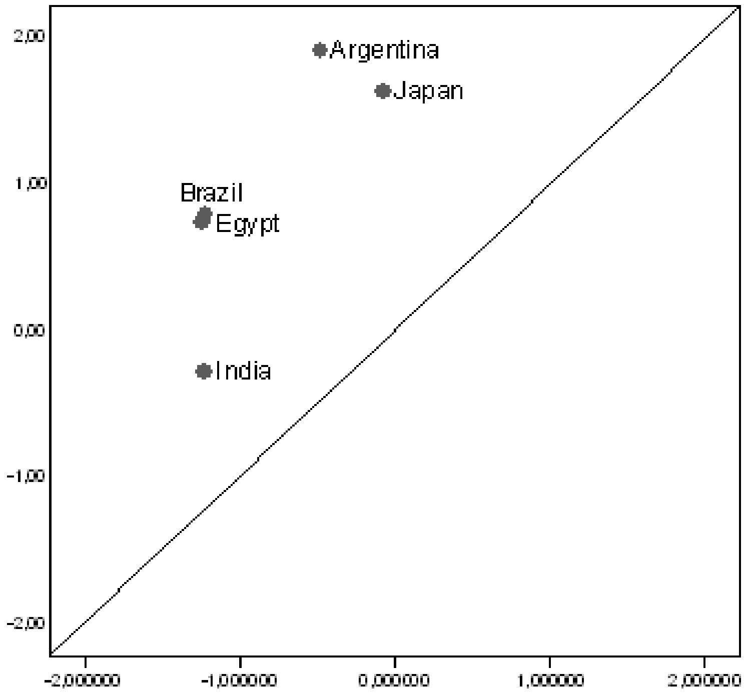
SDI 1870 in the horizontal axis. SDI 1890 in the vertical axis.
Almost all countries improved over this period.

Figure 2.5 - 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 2.6 - Scatter plot for Social Development Index. Historical Evolution from 1870 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.

Tables And Figures For 3.A) Contemporary Comparison Of Alternatives

Table 2.1 – 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	.	.	íger	-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	.
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	.	30	.	Saudi Arabia	.	.	.

Republic							
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
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 2.2 – Pearson Correlation Matrix

		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 2.1 - 3D Scatter Plots: Spikes to the Floor and Centroid respectively

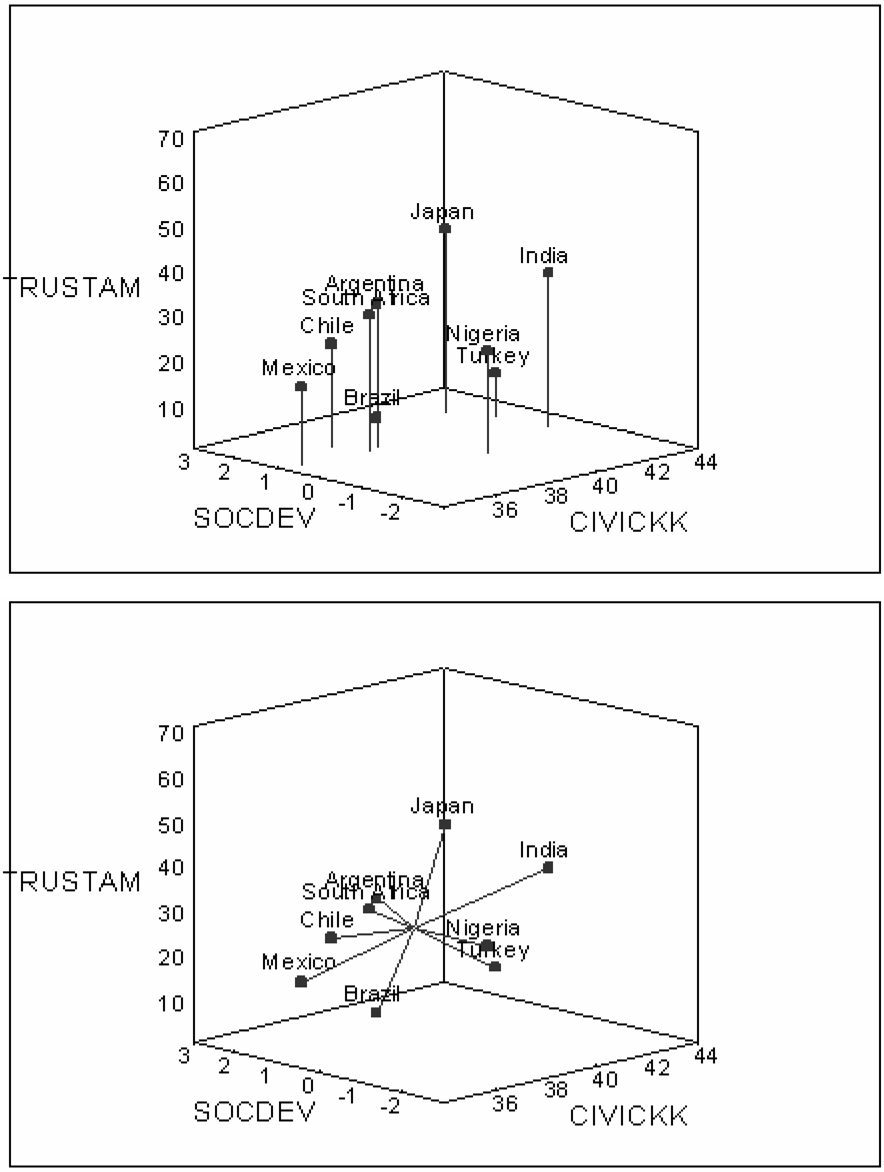
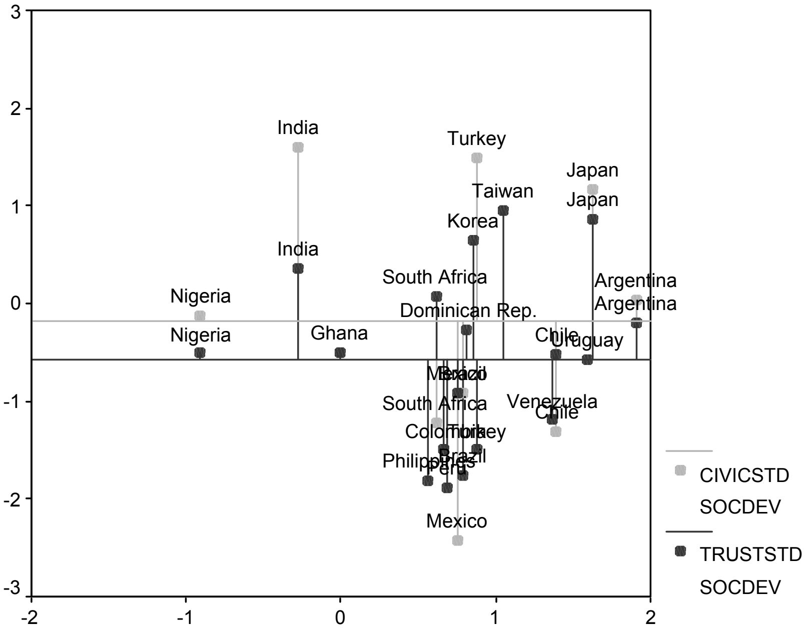
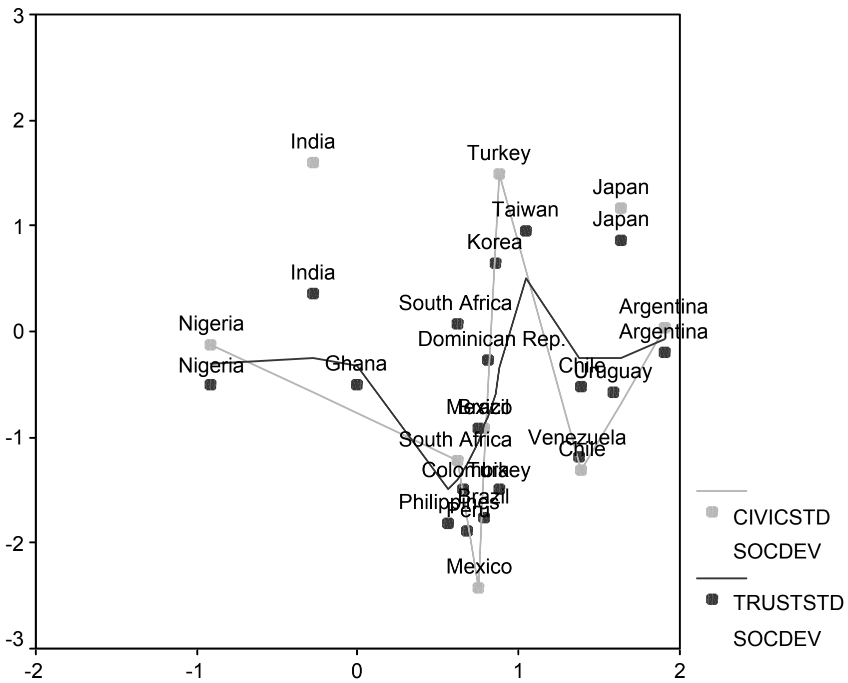


Figure 2.2 - Historical Evolution: Overlay Scatter Plot



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.

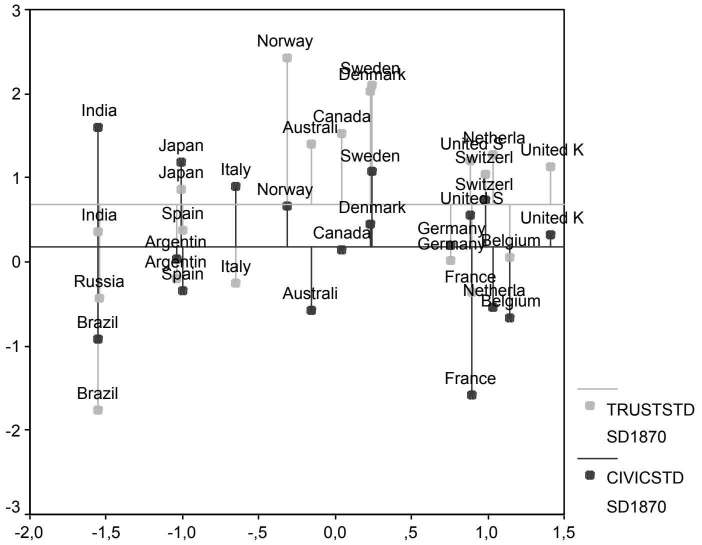
Figure 2.3 - Overlay Scatter Plot with Fit Line



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

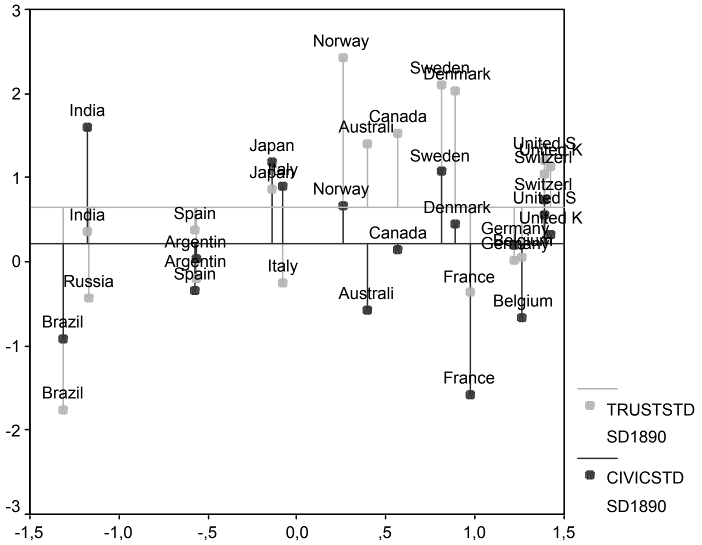
Figures For 3.B) Intertemporal Comparisons: Nineteenth And Twentieth Centuries

Figure 2.7 - Historical Evolution 1870-Nowadays: Overlay Scatter Plot with Spikes



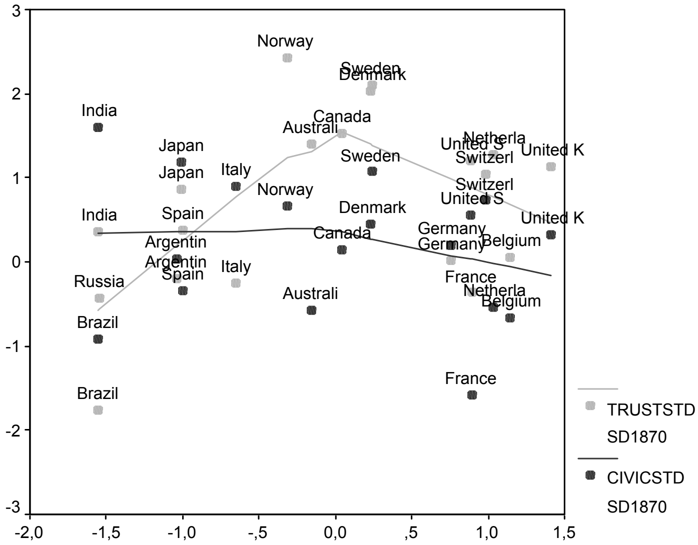
SDI 1870 in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.

Figure 2.8- Historical Evolution 1890-Nowadays: Overlay Scatter Plot with Spikes



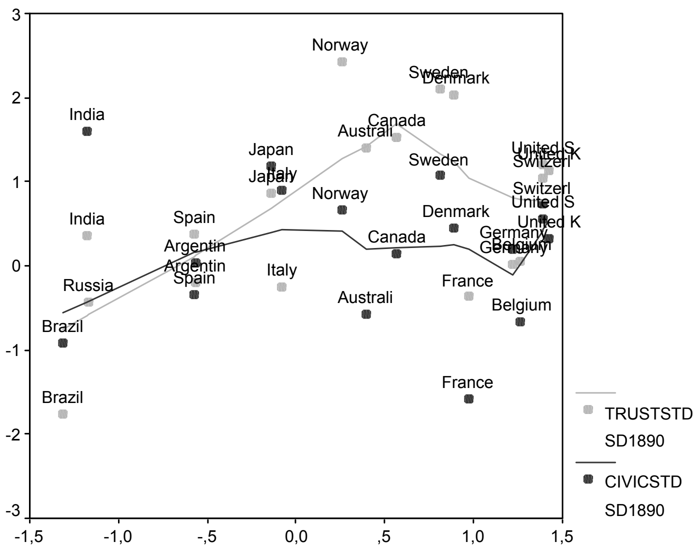
SDI 1890 in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.

Figure 2.9 - Historical Evolution 1870-Nowadays: Overlay Scatter Plot with Fitted Line



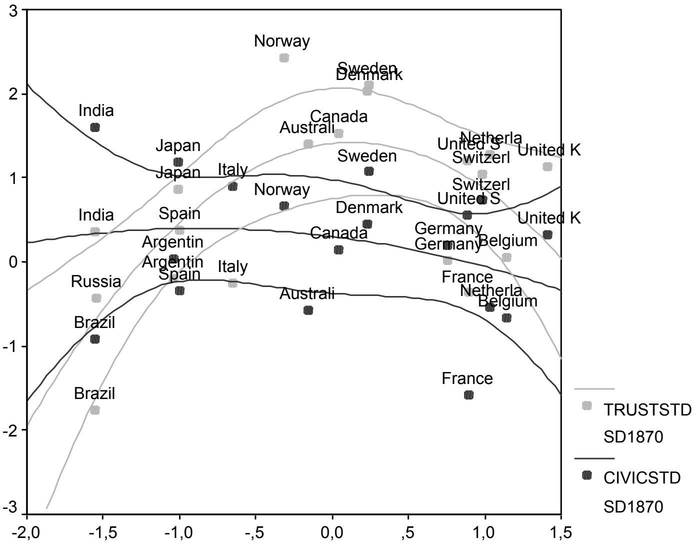
SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Lowess. 50% of points fitted with 3 iterations.

Figure 2.10 - Historical Evolution 1890-Nowadays: Overlay Scatter Plot with Fitted Line



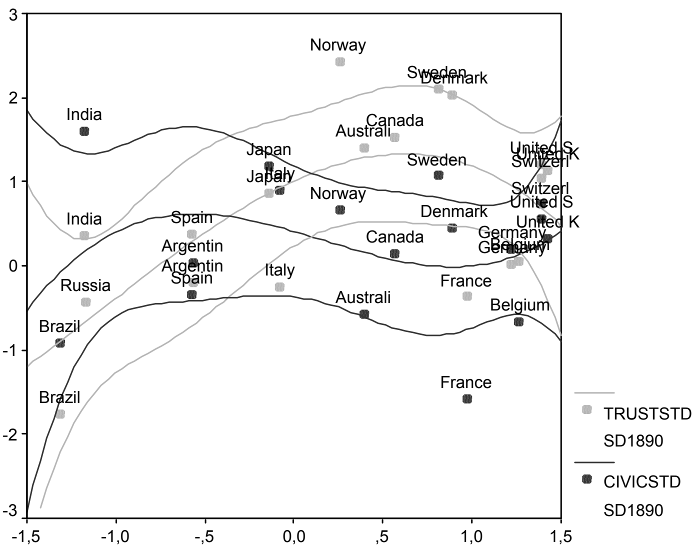
SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Lowess. 50% of points fitted with 3 iterations.

Figure 2.11 - 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 2.12 - 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

<u>List Of Countries</u>

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

COUNTRY	COUNTRYN	
Old Value	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

List Of Variables And Codification

This is the read file attached to our digitalisation and re-codification of the original Morris and Adelman (1988) database.

41 variables (+ COUNTRY and YEAR):

36 included in the factor analysis. Non-included ones are signed with a star *.

INCGROUP = Classification for level of per capita income. Re-codification: 5 to 0 corresponding to range A to F in Morris and Adelman (1988).

*INCOMEAM = Index of level of per capita income (UK 1890=100). Information summarised by Morris and Adelman in the variable INCGROUP. Thus, INCGROUP used instead in the factor analysis.

INCOMGRAM = Classification for rate of change in per capita income in the past 20 years. Re-codification: 4 to 0 corresponding to A to D- in the original.

INDUTECH = Level of development of techniques in industry. Re-codification: 5 to 0 corresponding to A to F in the original.

INDTECGR = Classification for rate of improvement of techniques in industry (lagged, referred to the last 20 years). Re-codification: 6 to 0 corresponding to A to G in the original database.

AGRITECH = Classification for level of development of techniques in agriculture. Re-codification: 6 to 0 corresponding to A to G in the original.

AGRTECGR = Classification for rate of improvement of techniques in agriculture. Re-codification: 5 to 0 corresponding to A to F in the original.

AGRILGRP = Classification for percentage of labour force in agriculture.
Re-codification: 6 to 0 corresponding to A to G in the original.

AGRICLAB = Percentage of labour force in agriculture.

Notes: - Percentage of total population used when percentage of active population not available.
- Most recent study used when discrepancies between sources.
- Oscillations in actually reported years in many cases. Special cases: India and Italy 1890, with estimation actually corresponding to 1901. Russia and Switzerland 1890 is for 1900. China 1910 corresponds to 1930 figure. Russia 1910 is actually the figure corresponding to 1926.

Within the category of 'Relative abundance of agricultural resources' we find several descriptive variables:

TOTALAND = Total land area in thousands of squared kilometres.

FARMLAND = Farmland in thousands of squared kilometres. This is assumed to be constant over the period of study, which is not necessarily true. So, further improvements on the database can be achieved by undertaking further research on the evolution of the amount of farmland in the second half of the nineteenth century. Starting point is Clark (1957) for being the source of the data presented herein.

POPXFARM = Population per square kilometres of farmland.

INTRANSP = Level of development of inland transportation. Re-codification: 4 to 0 corresponding to A to E in the original.

TRANSPGR = Classification scheme for rate of improvement of inland transportation (lagged). Re-codification: 5 to 0 corresponding to A to E in the original. Note: Categories C1 and C2.

XGRGROUP = Classification scheme for rate of growth of total real exports. Re-codification: 3 to 0 corresponding to A to D in the original.

* AVGXGR = Average annual rate of growth of real exports (%). XGRGROUP used instead in the factor analysis.

* AVGXGRP = Average annual rate of growth of exports in current prices (%). Real exports classification preferred in the original. This allows comparisons.

SHIFTX = Classification for degree of shift in structure of export sector. Re-codification: Range from 3 to 0 replacing the original A to D.

INDWCHAN = Classification for direction of change in average real wages in industry. Re-codification: Range from 4 to 0 corresponding to A to E in the original.

AGRWCHAN = Classification for direction of change in average real wages or income of the employed agricultural poor. Re-codification: From 4 to 0 corresponding to A to E in the original.

POPGROUP = Classification for total population. Numerical codification: From 100 to 0 corresponding to A to G. Updated with the previously unused table for 1914. Also available for 1830 (table A25 in Morris and Adelman, 1988:386). Update not included into the analysis.

* POP = Total population. POPGROUP used instead in the original.

POPGRGRP = Classification for rate of population growth in the last 20 years. Numerical codification: From 100 to 0 corresponding to A to F.

IMMIGRP = Classification for net immigration. Numerical codification: From 100 to 0 corresponding to A to F.

* IMMI = Net migration (immigration with positive sign and emigration with negative sign). IMMIGRP used instead in the original.

LIT = Classification of extent of adult literacy. Re-codification: Range from 0 to 100 corresponding to A to J in the original. Notice that the order of the numerical classification has been reversed in order to transform the illiteracy variable into literacy. This is intuitively convenient since we assign higher values to better outcomes in this way.

PRIMEDGR = Classification for rate of spread of primary education in the past 20 years. Numerical codification: From 100 to 0 corresponding to A to E.

LANDTENU = Classification for predominant form of land tenure and holding. Numerical codification: From 100 to 0 corresponding to A to G.

LANDCONC = Classification for concentration of landholdings. Numerical codification: From 100 to 0 corresponding to A to G.

LANDADOP = Classification for favourableness of land system to adoption of improvements. Numerical codification: From 100 to 0 corresponding to A to I.

URBANI = Classification for extent of urbanisation. Numerical codification: From 100 to 0 corresponding to A to D.

ENTREP = Classification for favourableness of attitudes towards entrepreneurship. Numerical codification: From 100 to 0 corresponding to A to F.

GOVT = Classification for extent of domestic economic role of government in the past 20 years. Numerical codification: From 100 to 0 corresponding to A to E.

SOCIOPOL = Classification for socioeconomic character of national political leadership in the past 20 years. From 100 to 0 corresponding to A to D.

REPRESEN = Classification for strength of national political institutions in the past 20 years. From 100 to 0 corresponding to A to D.

POLSTABI = Classification for extent of political stability in the past 20 years. Codification: From 100 to 0 corresponding to A to D.

FOREIGND = Classification for degree of foreign economic dependence in the past 20 years. Inverted numerical re-codification: From 0 to

100 corresponding to A to G. We assign a higher punctuation to economic independence, unlike the original. This eases the conceptualisation of the variable, since economic dependence is taken to be negative.

COLSTAT = Classification for colonial status. Numerical re-codification: Inverted from 0 to 100 corresponding to A to F.

Finally, the market institutional development variables come from an earlier study by the same authors. They selected the following three composite indicators, which summarise the level of institutional development of markets. For a detailed explanation, visit Adelman and Morris (1978).

MKTDEV = Component scores for composite indicator of level of development of market institutions up to the given date.

MKTDEVGR = Component scores for composite indicator of rate of spread of market institutions in the last 20 years.

MKTDVGRL = Component scores for composite indicator of rate of spread of market institutions in the last 20 years (lagged); i. e., the observation corresponding to 1850 is capturing the spread of market institutions from 1830 to 1850.

Appendix C

Quantitative Methodology: The Basics Of Principal Components Analysis

The principal components analysis is a data-reduction technique. In short, it aims to give 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 Principal Components depends on the correlations between the variables. The higher the correlation between the original variables, the smaller the number of Principal Components.

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

The Calculation of the Principal Components

In order to make the above description more precise, we need to consider the technique in more detail. Before we can do this, it is necessary to introduce some notation.

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: $X=(x_1\ x_2\ \ldots\ x_k)$.[7] 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.[8] 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

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

---

[7] 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.
[8] 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[9] 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

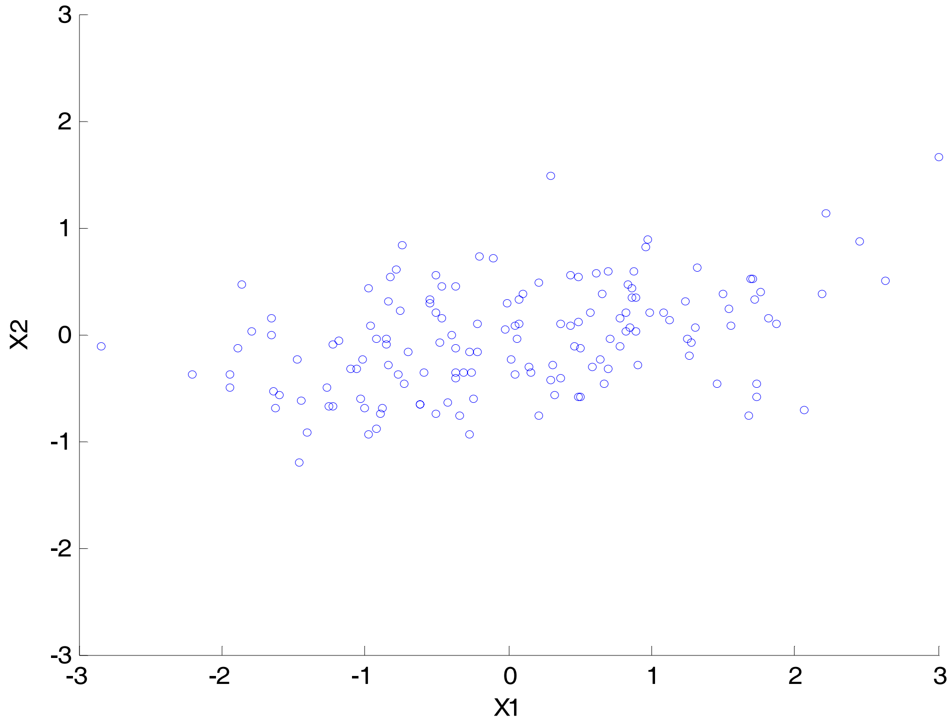
[9] See Johnson and Wichern (2002), chapter 8.

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 Graph 2.1 below. From this 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.

Graph 2.1 - 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 Graph 2.2, 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.[10] 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.

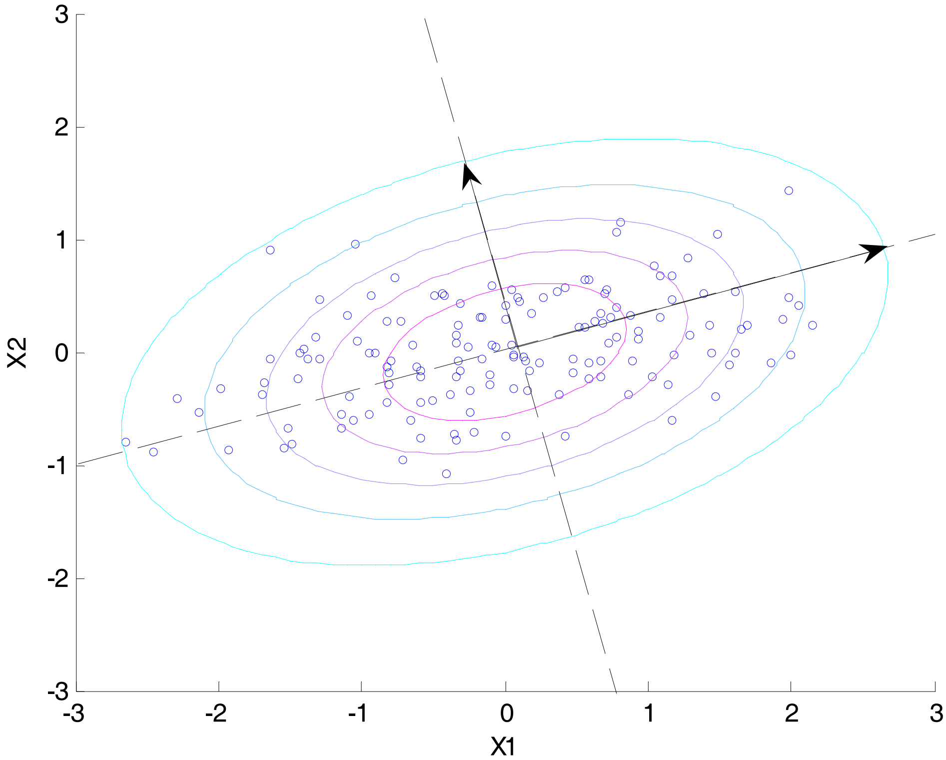
The first axis mentioned above is called the first Principal Component. Note that the points that lie on this axis are simply linear combinations of points on the original axes. The second new axis is

---

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

called the second Principal Component. Note that the two new 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 coordinate with respect to the first Principal Component and a

<u>Graph 2.2 - Scatter Plot with Principal Components.</u>



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 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.[11] 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= λx. If there exists such a scalar and

[11] If A is a k by k matrix and x a k by 1 vector, the result y is a k by 1 vector.

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,[12] there actually exist k eigenvalues, $(\lambda_1, \lambda_2, \dots, \lambda_k)$ and k corresponding eigenvectors $(e_1, e_2, \dots, e_k)$. 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.

[12] Remember that for Principal Component Analysis we are interested in the eigenvalues and eigenvectors of the sample covariance matrix S. This matrix is symmetric.

References[13]

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.

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*

[13] Ask Marta Felis Rota (m.felis-rota@lse.ac.uk) for the referenced html files in case the web page has changed.

and Proceedings 86 (2), 1996, pp. 218-23. (Updated dataset at: http://www.worldbank.org/research/growth/ddbarle2.htm)

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

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).

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

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

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

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.

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

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).

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

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.

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

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).

LONDON SCHOOL OF ECONOMICS
ECONOMIC HISTORY DEPARTMENT WORKING PAPERS
(from 2003 onwards) For a full list of titles visit our webpage at
http://www.lse.ac.uk/

2003

WP70 The Decline and Fall of the European Film Industry: Sunk Costs, Market Size and Market Structure, 1890-1927
Gerben Bakker

WP71 The globalisation of codfish and wool: Spanish-English-North American triangular trade in the early modern period
Regina Grafe

WP72 Piece rates and learning: understanding work and production in the New England textile industry a century ago
Timothy Leunig

WP73 Workers and 'Subalterns'. A comparative study of labour in Africa, Asia and Latin America
Colin M. Lewis (editor)

WP74 Was the Bundesbank's credibility undermined during the process of German reunification?
Matthias Morys

WP75 Steam as a General Purpose Technology: A Growth Accounting Perspective
Nicholas F. R. Crafts

WP76 Fact or Fiction? Re-examination of Chinese Premodern Population Statistics
Kent G. Deng

WP77 Autarkic Policy and Efficiency in Spanish Industrial Sector. An Estimation of the Domestic Resource Cost in 1958.
Elena Martínez Ruiz

WP78 The Post-War Rise of World Trade: Does the Bretton Woods System Deserve Credit?
Andrew G. Terborgh

WP79 Quantifying the Contribution of Technological Change to Economic Growth in Different Eras: A Review of the Evidence
Nicholas F. R. Crafts

WP80 Bureau Competition and Economic Policies in Nazi Germany, 1933-39
Oliver Volckart

2004

WP81 At the origins of increased productivity growth in services. Productivity, social savings and the consumer surplus of the film industry, 1900-1938
Gerben Bakker

WP82 The Effects of the 1925 Portuguese Bank Note Crisis
Henry Wigan

WP83 Trade, Convergence and Globalisation: the dynamics of change in the international income distribution, 1950-1998
Philip Epstein, Peter Howlett & Max-Stephan Schulze

WP84 Reconstructing the Industrial Revolution: Analyses, Perceptions and Conceptions of Britain's Precocious Transition to Europe's First Industrial Society
Giorgio Riello & Patrick K. O'Brien

WP85 The Canton of Berne as an Investor on the London Capital Market in the 18^{th} Century
Stefan Altorfer

WP86 News from London: Greek Government Bonds on the London Stock Exchange, 1914-1929
Olga Christodoulaki & Jeremy Penzer

WP87 The World Economy in the 1990s: A Long Run Perspective
Nicholas F.R. Crafts

2005

WP88 Labour Market Adjustment to Economic Downturns in the Catalan textile industry, 1880-1910. Did Employers Breach Implicit Contracts?
Jordi Domenech

WP89 Business Culture and Entrepreneurship in the Ionian Islands under British Rule, 1815-1864
Sakis Gekas

WP90 Ottoman State Finance: A Study of Fiscal Deficits and Internal Debt in 1859-63
Keiko Kiyotaki

WP91 Fiscal and Financial Preconditions for the Rise of British Naval Hegemony 1485-1815
Patrick Karl O'Brien

WP92 An Estimate of Imperial Austria's Gross Domestic Fixed Capital Stock, 1870-1913: Methods, Sources and Results
Max-Stephan Schulze

2006

WP93 Harbingers of Dissolution? Grain Prices, Borders and Nationalism in the Hapsburg Economy before the First World War
Max-Stephan Schulze and Nikolaus Wolf

WP94 Rodney Hilton, Marxism and the Transition from Feudalism to Capitalism
S. R. Epstein
Forthcoming in C. Dyer, P. Cross, C. Wickham (eds.)
***Rodney Hilton's Middle Ages, 400-1600* Cambridge UP 2007**

WP95 Mercantilist Institutions for the Pursuit of Power with Profit. The Management of Britain's National Debt, 1756-1815
Patrick Karl O'Brien

WP96 Gresham on Horseback: The Monetary Roots of Spanish American Political Fragmentation in the Nineteenth Century
Maria Alejandra Irigoin

2007

WP97 An Historical Analysis of the Expansion of Compulsory Schooling in Europe after the Second World War
Martina Viarengo

WP98 Universal Banking Failure? An Analysis of the Contrasting Responses of the Amsterdamsche Bank and the Rotterdamsche Bankvereeniging to the Dutch Financial Crisis of the 1920s
Christopher Louis Colvin

WP99 The Triumph and Denouement of the British Fiscal State: Taxation for the Wars against Revolutionary and Napoleonic France, 1793-1815.
Patrick Karl O'Brien

WP100 Origins of Catch-up Failure: Comparative Productivity Growth in the Hapsburg Empire, 1870-1910
Max-Stephan Schulze

WP101 Was Dick Whittington Taller Than Those He Left Behind? Anthropometric Measures, Migration and the Quality of life in Early Nineteenth Century London
Jane Humphries and Tim Leunig

WP102 The Evolution of Entertainment Consumption and the Emergence of Cinema, 1890-1940
Gerben Bakker

WP103 Is Social Capital Persistent? Comparative Measurement in the Nineteenth and Twentieth Centuries
Marta Felis Rota