Confirming long waves in time series of German student populations 1830-1990 using filter technics and spectral analysis

Mueller-Benedict, Volker

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Confirming Long Waves in Time Series of German Student Populations 1830 - 1990 Using Filter Techniques and Spectral Analysis

Volker Mueller-Benedict[*]

Abstract: From 1830 till 1990 there were four historical phases of overcrowding at German universities. To investigate these periods of overcrowding members of the research project entitled "Qualifikationskrisen" collected data on student populations in various disciplines at all German universities. Research members then used this data to create 160-year-long time series that demonstrate the impressive growth and cyclical behavior of Student populations at German universities. But to evaluate the number and the length of the cyclical components present in these time series 1 had to use notch filtering and Spectral analysis. Using these tools I detected the same two overlaying cycles in each time series, a longer cycle that is 35 – 40 years in length and a shorter cycle that is 13 – 18 years in length and is only present in the data from the twentieth century. In the following presentation I will present and discuss the meaning of these cycles.

Sections:
1. Data: the stock of the time series
2. The problem and the research question
3. Trend elimination using notch filtering
4. The results of the spectral analysis
5. A discussion of the meaning of the cyclical components

1. Data: the stock of time series

The data, which will be presented in this study, was collected over a period of more than ten years by the members of a research project entitled "Qualifica-

---

[*] Address all communications to Volker Mueller-Benedict, Methodenzentrum Sozialwissenschaften, Platz der Goettinger Sieben, D-37077 Goettingen.
E-Mail: vbenedi@uni-goettingen.de.

tion Crises and Educational Changes in the Higher Education Systems of Germany and Prussia from 1820 to 1945", short "QUAKRI". The first reason to collect data was the hypothesis that the overcrowding crisis in German universities from 1885 – 1890 was manipulated and augmented by the conservative Prussian government to prevent members of the lower social classes from studying and to restore the social exclusivity of the Prussian universities (Herrlitz/Titze 1978). This crisis had a large impact on the public discussion surrounding academic education and culminated in references to a growing "academic proletariat". During a great debate in parliament the Prussian minister of culture, Gossler, asked, with the acclamation of the conservatives, "how many students are necessary to maintain the size of the governing classes?" Then, a secret prognosis of student enrollments, which had been made by the economist Wilhelm LEXIS, was used to successfully push through political measures designed to diminish student populations. But all of these efforts were unsuccessful and alter a short cyclical recession, student populations grew at a much faster rate than they had beforehand (Mueller-Benedict 1991:11 f).

Therefore it became clear that the dynamic of the process that led to overcrowding was independent of political influences. As well, the hypothesis that the overcrowding crisis was manipulated by particular interest groups, Gould not be sustained. Therefore further investigations into the nature of the self - generated dynamics present in the time series were necessary. As a result of these investigations the project team was able to publish a complete stock of data on the changing number of students for all subjects at all universities. This data, together with other interesting information such as the age, sex, and social or class Background of the students, has been published in two books (Titze 1986, 1993).[1]

2. The problem and the research question

The majority of the time series Show impressive growth and cyclical behavior. For example, figure 1 shows four graphs that represent the four primary fields of study – theology, law, medicine and "education science" – at all German universities. The maximum and minimum points of the graphs correlate with the historical dates of political conflicts or public debates about overcrowding and the lack of professionals. But the cycles vary in length and do not maintain the same shape over time. Therefore a statistical time series analysis had to be carried out, with the following research questions

1) What are the main cyclical components?
2) Which changes were produced by the composition and the length of the cyclical components over time?

---

[1] The data will be available in electronic form at die Central Archive for Empirical Social Research (ZA). Internet-Address: www.za.uni-koeln.de

3) Which parallels can be drawn between the length of the detected cycles and the length of specific historical characteristics and or developments in the German and Prussian educational systems?

Figure 1: Student populations of the four primary fields of study in Germany 1830-1945

Medicine

Education science
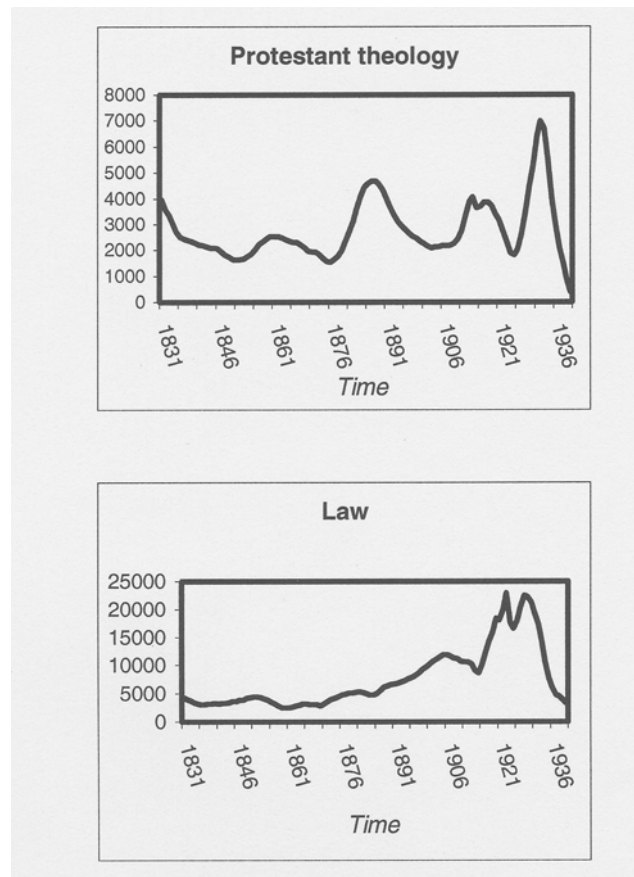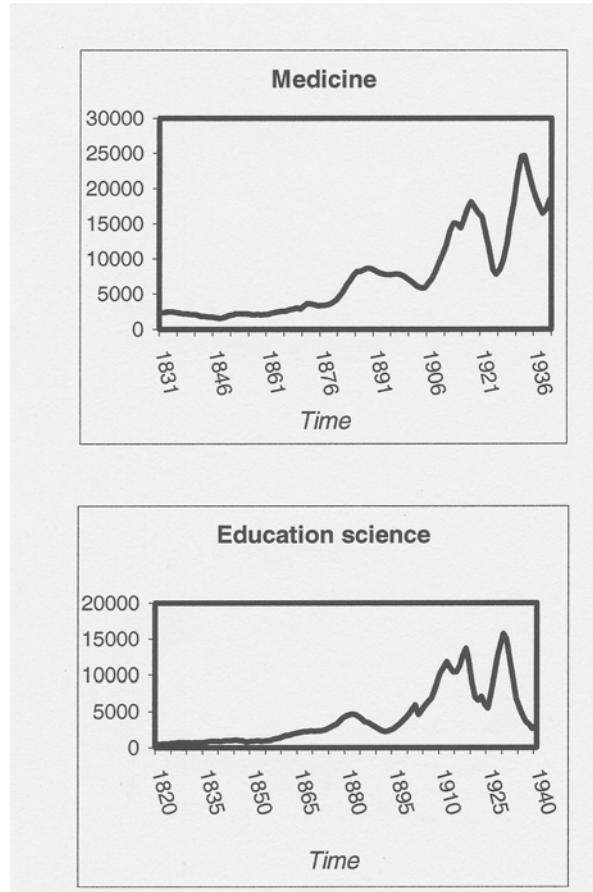
3. Trend elimination and notch filtering

After a first inspection of the time series it became clear that we could expect cycles to be up to 40 years in length. This means that the second goal, detecting changes in cycle lengths during the 120-year-long time series, cannot be reliably achieved. This is because there are only two or three cycles within this 120-year time period. So the first decision that had to be made was to extend the time series up to the present day and as far into the past as possible. The geographical, cultural and political space of the German Empire remained largely

stable from 1830 to 1940, but the Federal Republic of Germany only possessed a part of the German Empire's former territory until 1990, so extending the time series could be a problem. But the German educational system, which builds the basis for the time series, has remained quite stable since 1830 and changes and developments in Germany's educational system took place simultaneously throughout Germany[2]. Therefore, if we ignore the socialist school system that was introduced in the former German Democratic Republic, it is clear that, in comparison to other social institutions, Germany's school system has remained fairly stable over a period of 160 years.

This stability in the German school systems allows us to suppose that the basic mechanisms that generate the time series are the same in the former German Empire and the Federal Republic of Germany. Therefore we are able to connect the time series from the German Empire, which last till the end of World War II, with the time series from the Federal Republic of Germany. As well, most of the universities of the German Empire are still present in the FRG and the average student populations have remained nearly the same. Furthermore, since Prussian data is available since 1820, our time series for the field of education can be extended into the past if we connect the Prussian series with 60% of the data from the Federal Republic of Germany. Thereby we were able to create four extended time series that represent student populations in the four primary fields of study from 1820 till 1990 (figure 2).

Figure 2: Student populations of the four primary fields of study in Germany/Prussia 1830/20-1990

---
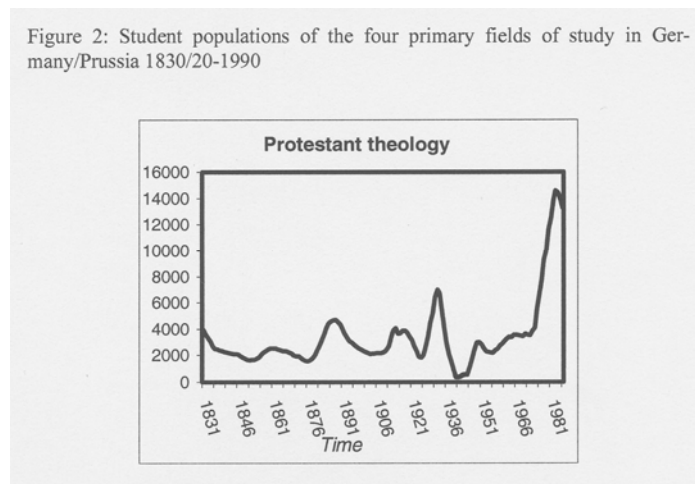
2   The German education system is a parallel system that is composed of schools for lower, middle and higher education as well as a university system established by Wilhelm Humboldt that enjoys academic freedom and scientific autonomy.

Law

Medicine

Education science
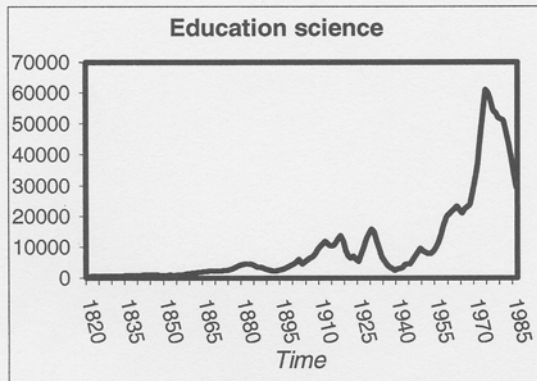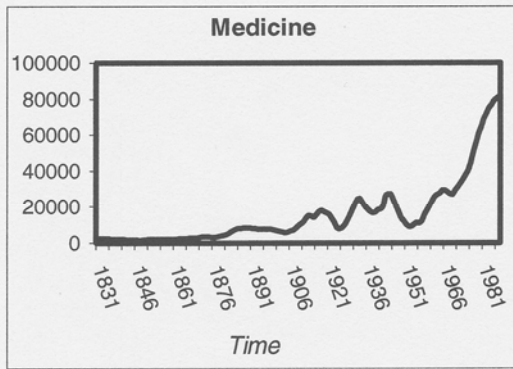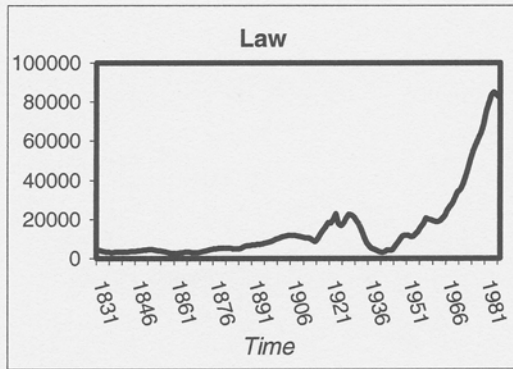
The inherent trend that prevents us from seeing much cyclical behavior is easier to perceive in these graphs with extended data than in previous graphs that only included data up to 1940. This inherent trend, namely, growth in the size of the higher educational system, can be understood by turning to theories of inclusion or growth in the sciences for example. But theories such as these cannot be used to design a quantitative model of increases in student populations at German universities. Therefore the elimination of a trend is more of a technical problem, namely the problem of finding a suitable filter, since none of the specific smooth trend functions have an empirical basis.

To find a suitable filter for trend elimination we first have to decide which model underlies the time series being considered. Two alternatives are an integrated ARIMA-model and a component model consisting of a trend and a cyclical component. In this case the decision is simple: a component model should be used, because we believe that there are social conditions that only generate cycles and other social conditions that only generate growth in student populations. Furthermore, we know that these two social forces can be separated. Therefore we can conclude that it is not necessary to specify the parameters of an approximate ARIMA-model. Instead, we have to choose a suitable linear filter that can extract specific frequencies. This filter has to be a high-pass-filter, which suppresses the long frequencies contained in the time series.

As a result of research and development, especially in the field of econometrics, several high pass filters are available that suppress long frequencies and are very useful for trend elimination, the now widely used Hodrick-Prescott-Filter for example (hetz 1996). But since all of these filters have their specific advantages and disadvantages, it is difficult to determine which filter is the best for a specific problem. The filter I selected is the notch filter, a filter that is less commonly used, but still just as good as these that are used more often (Cadzow 1973, Schulte 1981).

The notch filter transfers all frequencies of the original times series without changing them, but at a single pre-set frequency there is a notch in the transfer - function. Ideally, this notch ensures that a single frequency is not transferred to the output time series. If you set this notch at frequency zero, only non cyclical elements will be removed, meaning, of course, that only the trend will be removed from the time series.

The notch-Filter has three poles and their locations determine the width and shape of the notch. The width should be close or equal to the inverse of the length of the time series, so that all cycles shorter than the complete length of the time series can pass through the filter without being changed. Under these conditions the these poles of a filter with a notch at frequency f are located at

$p_{1,3}$ $= \cos(2\pi f) - d \cos(2\pi f \pm \alpha) + i \sin(2\pi f) - d \sin(2\pi f \pm \alpha)$,

p_2 $= (1-d) \cos(2\pi f) + i (1-d) \sin(2\pi f)$ (1)

d and α = polar coordinates of the filter in the complex plane.

In order to install the notch filter into a computer program, each pole is treated as a single filter so that the time series will pass through a cascade of three filters

$$Y_j(t) = b_j (y_{j-1}(t) - 2 \cos (2\pi f) y_{j-1}(t-1) + y_{j-1}(t-2)) +$$
$$2 \operatorname{Re} p_j y_j(t-1) - |p_j|^2 y_j(t-2) \qquad (2)$$

For j 0 1,2,3, $b_1 = b_2 = 1$, $b_3 =$ constant for normalization. In order to widen the gap of the notch, this filter is used twice. The first time the filter is used, the notch frequency is set at 0 and the second time the filter is used f is an equivalent to half the width of the first notch. This procedure results in the transfer function within the domain of the amplitudes of figure 3.

The small hill at the ground of the notch represents the rest of the left side of the second filters notch. The filter does a good job of suppressing all frequencies from zero to $0.01 = 100$ years and does not change frequencies that are shorter than $0.013 = 77$ years. One reason why this filter has limited application in the social sciences may be that it shifts the phase of the longer cycles. The transfer function in the phase domain results in figure 4.

Figure 3: Transfer function of the notch filter, parameters as described in the text.
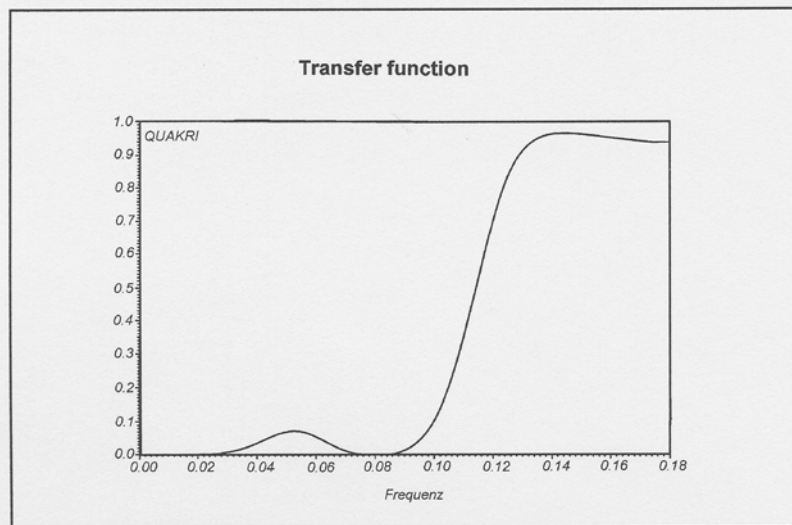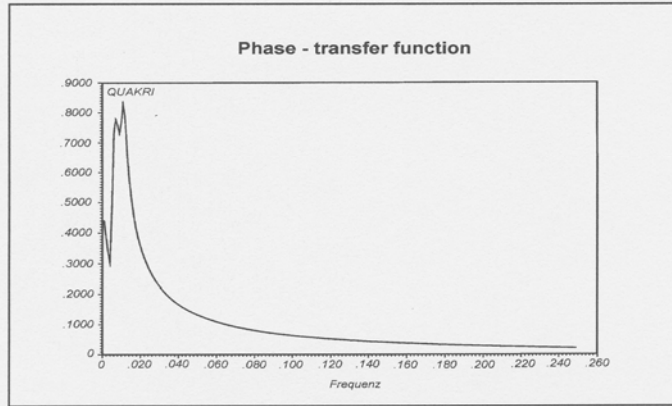
Figure 4: Phase-shift function of the notch filter, parameters as described in the text.
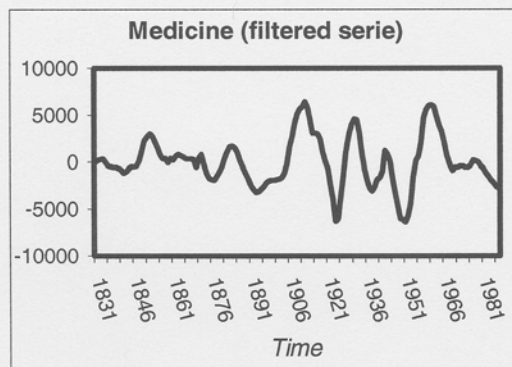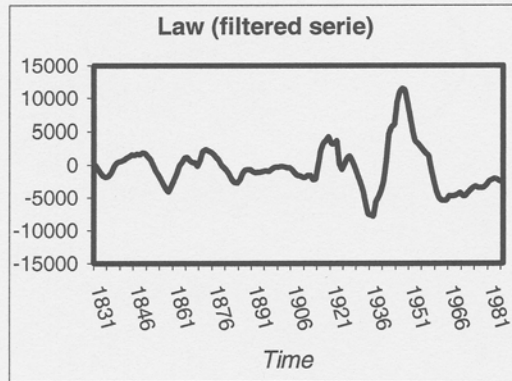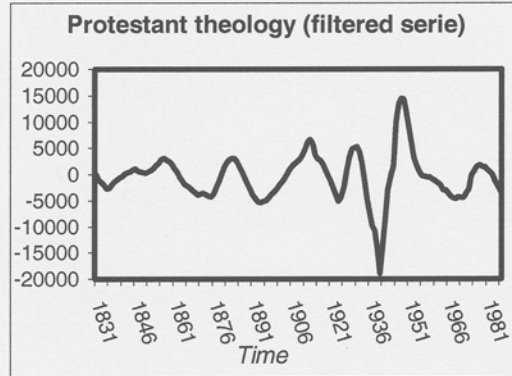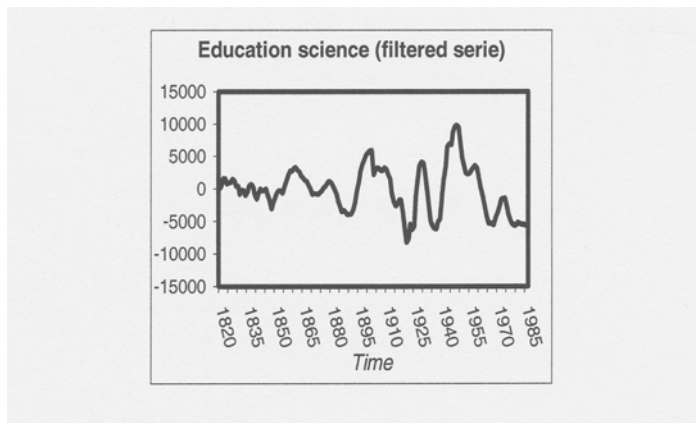
Phase - transfer function

QUAKRI

This image shows that a cycle with a length of 30 years, for example, will experience a phase shift of about 20% after filtering. This means that in the output time series of these cycles the peaks will appear 6 years later.

If we use a component model, the time series are transformed into logarithms before they are filtered. This is necessary because the underlying trend changes the amplitudes of the cycles multiplicatively. Therefore, in this case, overcrowding at German and Prussian universities should not be measured in absolute terms, but in relative terms, since overcrowding is part of a long-term trend. So, to remove the trends from the four primary fields of study the data is transformed into logarithms and filtered using the notch filter just described. This process results in four stationary time series (figure 5).

Here we can see the effectiveness of a notch filter; the trend has been removed and the cycles have been restored. The same is true even at the ends of the time series where the cycles are most effected by trends. If you take a closer look you can even see the phase shifts of some of the longer cycles, during the years between the first and second World Wars for example.

Figure 5: Notch-filtered time series of the student populations of the four primary fields of study in Germany/Prussia 1830/20-1990

Protestant theology (filtered serie)

Law (filtered serie)

Medicine (filtered serie)

Education science (filtered serie)

4. The results of the spectral analysis

The long-term cyclical behavior of the filtered time series seems to be irregular. After closer inspection it becomes clear that the cycles are between 35 and 10 years in length. A mixture of cycles seems to be present, meaning that the single elements of these mixtures have to be isolated. To do this we use spectral analysis in order to separate single frequency components. Because I am interested in changes in the cyclical structure of the series, the time series will be analyzed in parts. But how should the time series be divided? Because the length of the whole series encompasses 168 data points and the length of the longest expected cycle is nearly 40 years, the only possible way to divide the time series is to divide it into two parts, each about 80 years in length. In this case, two of the longest cycles, which are about 40 years long, can be contained in one series so that they can be detected faultlessly.

Because of the technical limitations of spectral analysis, only a very crude separation of the frequencies is possible under these conditions. The base frequencies of the analysis are 1/84, 1/42, 3/84, 1/21, 5/84, etc...

This means for example that with a spectral window of length 7 we were not able to separate cycles that are 84 years long from cycles that are 13 years long. Therefore, we have to use a spectral window of length 3, if we want to separate cycles in our empirical data that are 42 years in length from those that are 21 years in length.

To demonstrate the best possible results under these conditions, I first conducted a spectral analysis with simulated data. This data was generated using an AR(2) and an AR(4)-process

$$y(t) = \quad 1.85y(t-1) - 0.95\ y(t-2) \tag{3}$$
$$y(t) = \quad 3.5775y(t-1) - 4.9987y(t-2) + 3.2287y(t-3) - 0.8145y(t-4) \tag{4}$$

The theoretical spectrum of such processes can be obtained using calculus. The first series has peaks at a frequency of 0.0426 or 23.5 years and the second series has peaks at a frequency of 0.033 or 33 years and at a frequency of 0.0713 or 14 years. The spectra can be seen in figure 6.

Figure 6: Theoretical spectra of an AR(2)- and an AR(4)-process with generating formula (3) and (4)
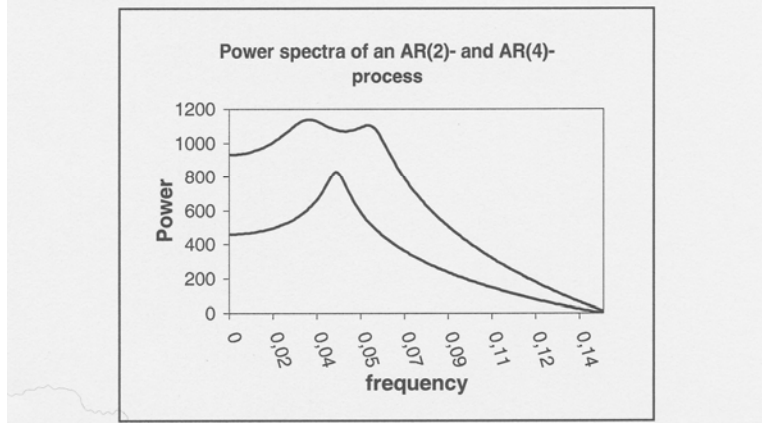
Then I generated 390 and 80 data points by using the formula of the AR-processes just mentioned and calculated their spectra by using the spectral analysis parameters already mentioned (figure 7).

By comparing the theoretical spectra with the spectra of the simulated data, you can see that the peaks of the spectra of the simulated data Set have been shifted to the basic frequencies of the spectral analysis. For example, the theoretical peak with a frequency of 0.033 is moved to 0.024 and spreads out to the next basic frequency of 0.0375. These shifts Show us the limitations of long wave separation in short historical time series. But this simulation also Shows us that we can expect to be able to separate shorter cycles from cycles that are 25 years or longer.

Figure 7: Spectra of two realizations (390 and 80 data points) of the AR(4)-process with generating formula (4)
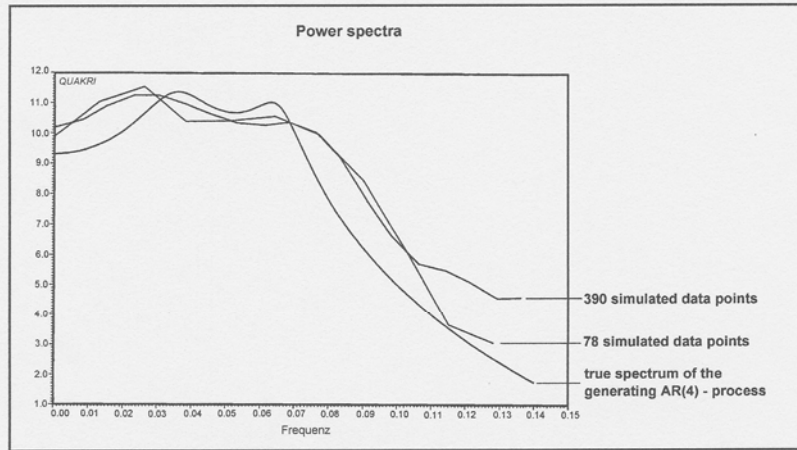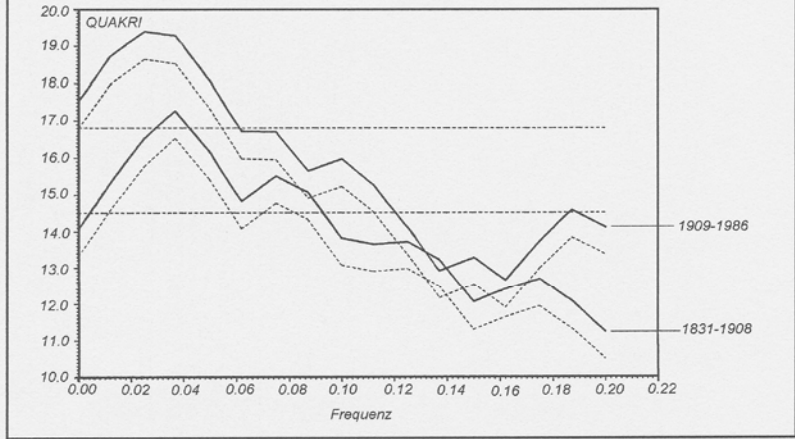

A spectral analysis with these settings produces the following spectra for the four detrended time series:
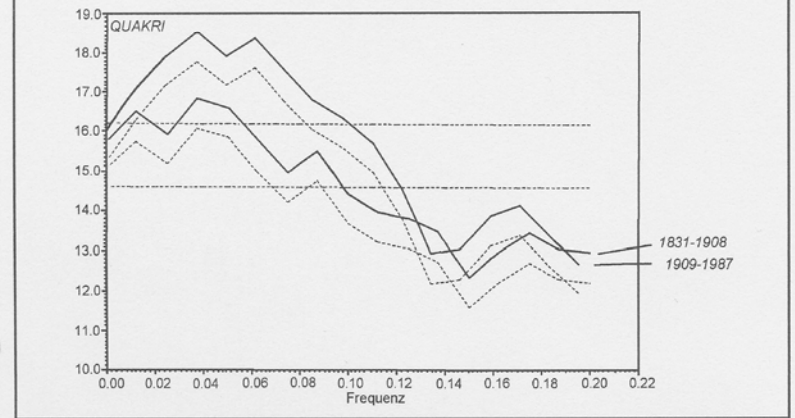
Figure 8: Spectra of the four detrended time series of the student populations of the four primary fields of study in Germany/Prussia 1830/20-1990
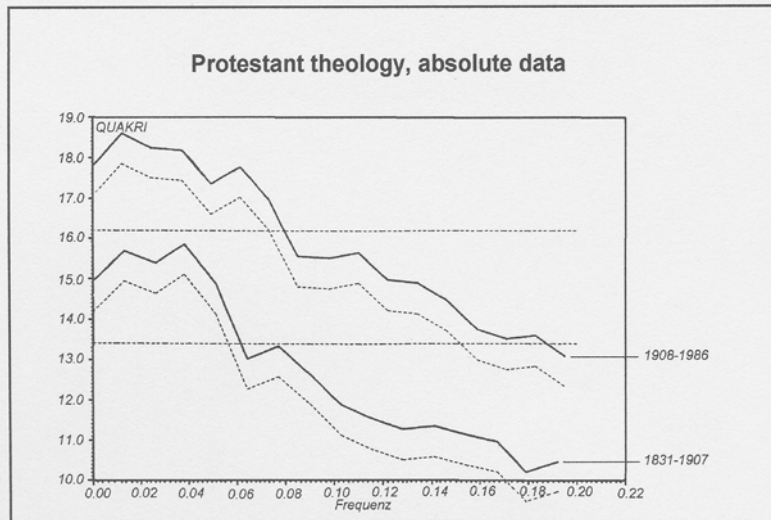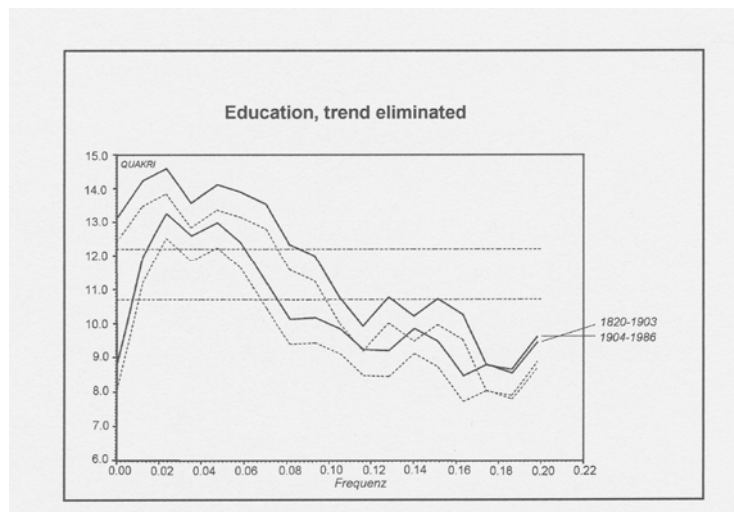
Law, trend eliminated

Medicine, trend eliminated

Education, trend eliminated

In table 1 these results are summarized. In the first row of this table the spectra of the unfiltered series for the field of evangelical theology are included. The inclusion of these spectra allows us to analyze the impact of the detrending procedure. An inspection of the table leads us to the following conclusions:

1. One main cyclical component can be isolated in all series. This main cyclical component is between 26 and 41 years in length. In other words, this cyclical component has a length that is close to the second or third basic frequency, a length that cannot be isolated. Therefore this main cyclical component is about 33 years long on average.

2. The second cyclical component, which lies significantly above the white-noise-border, only occurs in some of the series. If this cyclical component exists in the first half of the series, it has less spectral power than the longer cyclical component. Therefore we can conclude that this cycle emerges during the first half of the series. The second cycle is between thirteen and 19 years in length, has an average frequency of 15 years, and can be well separated from the longer cycle.

3. A comparison of the spectra of the raw and filtered theology data Shows that the filtereng process does not harm the theology data. The peak of the longer cycle has been shifted, but this shift lies within the width of the spectral window and stems from the spectral mass with a frequency of near zero, a frequency that was completely removed by the filter, but exists still in the raw data.

Table 1: Frequency and power of the main cyclical components of time series of students of the four primary fields for two time areas (Mueller-Benedict 1991:219)

Career/time	Cycle length		Power		Varian.
	1. peak	2. peak	1. peak	2.peak	
Prot.theology not filtered 1831-1908	263	-	15.8	-	13.4
Prot.theology not filtered 1909-1986	78.0	15.6	18.6	17.7	16.2
Prot.theology 1831-1908	26.3	-	18.4	-	15.5
Prot.theology 1909-1986	38.4	15.6	19.5	19.7	17.3
Law 1831-1909	26.3	13.0	17.2	15.5	14.5
Law 1909-1986	26.3	13.0	19.4	16.7	16.9
Medicine 1831-1908	26.3	11.2	16.8	15.5	14.6
Medicine 1909-1986	26.3	15.6	18.6	18.5	16.2
Education 1820-1903	41.6	20.8	17.9	17.6	15.5
Education 1904-1986	41.6	15.6	19.0	18.8	16.8

In summarizing these results, we can conclude that the shape of each of the four time series, which represent the four primary fields of academic study, are primarily dominated by two cycles: a longer cycle covering a period of about 33 years, and a shorter cycle covering a period of about 15 years. Additionally we can conclude that the shorter cycle came into being during the second half of the 19th century.

These results lead us to pose the following question regarding the meaning of these cycles: what kind of structural conditions can lead to these two special cycle lengths? To find an answer to this question we first analyzed the typical career paths as well as the institutional contexts and regulations governing the careers of individuals who received a degree in one of the four primary fields of study.

During the first half of the time period considered, all the professions we considered in this study gained autonomy. The prerequisites and exams required for entry into a given career were prescribed and the full turn of a career cycle was regulated. In fact, the careers of teachers and vicars were organized

by central institutions. In this sense their careers were as strictly organized and regulated as the careers of civil servants such as judges and public prosecutors. Physicians, on the other hand, organized themselves by means of a their own central institution, which regulated the number of medical offices, thereby guaranteeing physicians lifelong employment. One of the consequences of these regulations was, and this has not changed till recently, that very few people dropped out of one of these professions before they reached the prescribed retirement age. This meant that all people who got a job in one these professions retained their position during their entire working life, and people who finished their studies during times when there were few chances for employment, ended up waiting a long time to get one of the few available posts.

Based on these fact we would expect that the length of time which a typical professional spent working in one of these careers would be a structural constant and would be present in the time series. Since the typical career of a Professional began between the ages of 25 and 30, following studies at a university and internships, and lasted until retirement between the ages of 60 and 65, this "structural constant" would be between 30 and 40 years long. But how would career length lead to peaks in the time series? To answer this question we collected data on some of the age structures present in the four primary professions (Titze/Nath/Mueller-Benedict 1985). This data provided us with the following figure 9.

Figure 9: Age-structure of teachers in higher education in Prussia 1886- 1937.
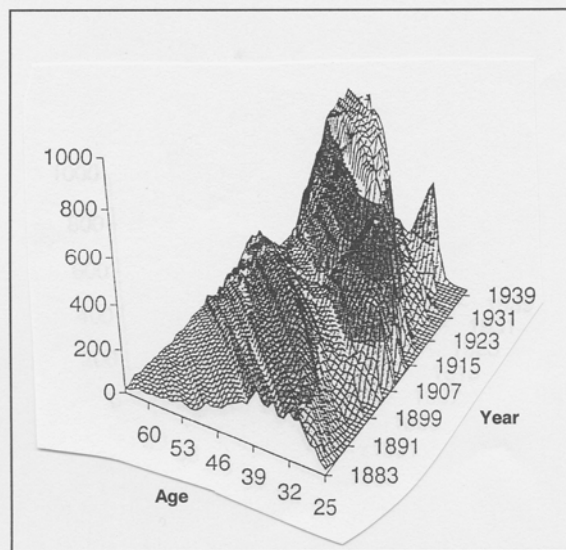
Here you can see that there are waves in the age structures that pass through the entire time series. When one of these waves ends with retirement, new careers are started and a new wave begins. So the longer of the two cycles present in the four time series is a result of the changes that are caused by generational shifts. These generational shifts have such a large effect on the time series because the age structures in all four professions are very asymmetric and generate cycles of demand in each of the professions. For example, when the peak of this age structure reaches retirement age, the number of vacancies is very high, and when the peak approaches 45 years, the demand for new professionals is very low. Since a single typical career lasts between 30 and 40 years, cycles of demand are generated that last just as long. As well, changing cycles of demand cause cycles in student populations during the same time periods. Therefore the longer cycle, which is about 33 years in length, represents changes in demand as a result of generational cycles in all four professional careers.

This mechanism that generates cycles obviously works best in a stable career in which the number of jobs remains stable over time and demand for workers is only a result of the need for replacements following retirements. The only career that fulfills these conditions is protestant theology. Additionally, if there is demand for an increase in the number of professionals in a given career, the long waves of this career's age structure will be disturbed. As a result, these long waves will not be as easy to discern as the long waves in the age structures of stable careers.

But how we can explain the presence of shorter cycles, especially their emergence after the foundation of the second German Empire? Economic analyses of market supply and demand tell us that we can expect cycles in supply and demand for goods that take a long time to produce. This means that some goods are only available a while after the demand for these goods has been established. As a result, there are alternating periods of surplus and shortage because of the time lag between supply and demand. These cycles of supply and demand are more extreme if there is not a pricing mechanism that allows the price of a good to change in order to dampen lags in production.

As for academic careers, waves are very stable in comparison to the prices of "normal" economic goods. As well, it often takes 8 years or Tonger to complete the academic education necessary to become an academic professional. Therefore, in the market for academic professionals the conditions are such that cycles are likely to exist. Economic analyses also Show that cycle lengths in the academic market are twice as long as the production time, namely the time it takes to become an academic professional. That means that in academic labor markets we can expect cycles to be between 12 and 18 years in length. This length is exactly the same as the length of the second cycle we found in the time series. Therefore the second cycle is a typical supply cycle that is

generated by a time lag between the decision to pursue an academic career and entry into an academic career, a time lag of about 15 years.

Based on these results we can conclude that the overall shape of the time series is the product of a demand and a supply cycle. The demand cycle triggers the supply cycle, but because these cycles have different lengths, the effects of the demand cycle vary, resulting in both larger and smaller distances between the supply and demand cycles.

The question left to answer is how the shorter supply cycles emerged. By using the time series of individual universities for which data is also available prior to 1800, we can confirm that the shorter cycles are not present before the first half of the 19th century. Why is this so? Well, economic modeling tells us that if supply cycles don't damp each other down and converge, the reaction must be of a quantity that exceeds demand. This principle allows us to conclude that shorter cycles can only emerge if social mobility is possible. In this case social mobility means that access to the education necessary to become a professional is not limited to members of social classes occupied by professionals. Historical events around this time provide us with another hint. During the first half of the 19th century, industrialization and modernization broke up the hierarchical order of society and the local isolation of economies, allowing social and regional mobility to grow. Then, increased social and regional mobility allowed nationwide labor markets to come into being. Before these processes had begun, the populations available for recruitment into a professional career were not large enough to cause noticeably large supply cycles. Therefore, from a historical perspective, shorter cycles are "modern" cycles that resulted from the increased importance of markets and individual freedom. Longer cycles, on the other hand, resulted from the social closure, or social immobility, that was typical in most societies prior to the mid 19`h century. In other words, until 1850 the demand for professionals was governed by the generational cycles of a profession's existing members.

Thus far historical processes were described, and the question that arises is whether the processes will continue to generate cycles in the near future as well. This is a difficult question to answer for two reasons: The first reason is that in these research studies we let aside the nature of trend figure growth, which has dominated student populations, especially in the last 40 years. In fact, in Germany, and in other European countries as well, many people are debating the future of European educational systems, and the overall growth of the higher educational system will depend on the results of this debate. The second reason why it is difficult to answer this question is that the modernization processes in our societies tends to dissolve social structures, especially the images of the professions. This is why a professional today is expected to continue his studies during his entire life and can no longer expect to remain in the same professional field after graduating from a university. This means that the

generation cycle that dominated the shape of the time series in the past has now been weakened in many professional careers.

Nonetheless, for some typical careers the generational cycle is still at play. For example, if we take a look at the age structure of German high school teachers during the last 40 years, we can see that in the next few years there will be a large increase in the demand for teachers (figure 10). This increased demand will be a continuation of the processes already described processes that have been at play for about 200 years; at least as far as the careers of high school teachers are concerned.

Figure 10: Age-structure of teachers in higher education in the Federal Republic of Germany 1955- 1990 (data source: unpublished combination of several statistical publications)
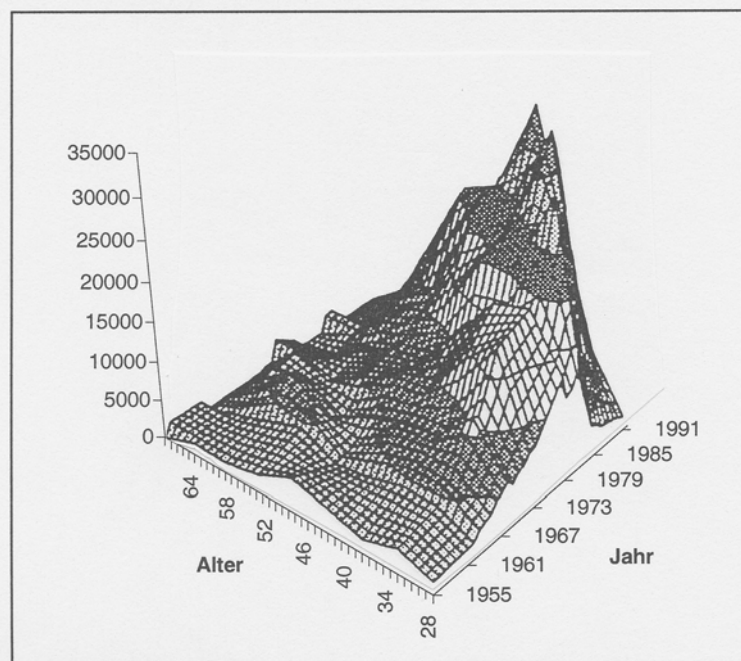
Literature:

Cadzow, J. 1973: Discrete Time Systems. Englewood Cliffs.

Herrlitz, H., Titze, H. 1978: Überfüllung als bildungspolitische Strategie, p.348-370 in: Die Deutsche Schule 1978.

Metz, R. 1996: Der Einsatz des Hodrick-Prescott Filters zur Trendbestimmung in ökonomischen Zeitreihen, p. 48-80, in: Historical Social Research / Historische Sozialforschung 21(1996), H.2.

Mueller-Benedict, V. 1991: Akademikerprognosen und die Dynamik des Hochschulsystems. Frankfurt: Campus.

Schulte, H. 1981: Statistisch-methodische Untersuchungen zum Problem langer Wellen. Königstein.

Titze, H. unter Mitarbeit von H. Herrlitz, V. Mueller-Benedict u. A. Nath 1986: Datenhandbuch zur deutschen Bildungsgeschichte. Band I, Teil 1: Das Hochschulstudium in Preußen und Deutschland 1820-1941. Göttingen: Vandenhoek.

Titze, H. unter Mitarbeit von H. Herrlitz, V. Mueller-Benedict u. A. Nath 1993: Datenhandbuch zur deutschen Bildungsgeschichte. Band I, Teil 2: Wachstum und Differenzierung der deutschen Universitäten 1830-1945. Göttingen: Vandenhoek.

Titze, H., Nath, A., Mueller-Benedict, V. 1985: Der Lehrerzyklus. Zur Wiederkehr von Überfüllung und Mangel im Höheren Lehramt in Preußen, p. 97-126 in: Zeitschrift für Pädagogik 31(1985).