FLUCTUATIONS IN THE RATE OF BASIC INNOVATIONS. A POISSON REGRESSION APPROACH

by

Bart Verspagen

ECIS, Eindhoven University of Technology
&
MERIT, Maastricht University

First draft

This paper is the result of work that emerged out of many discussions and joint work with my colleague Gerald Silverberg. I am indebted to him for letting me use many of his ideas here. He is not to be blamed for any controversial opinions or remaining errors, however.

1. Introduction

Schumpeter (1939) initiated a debate that echoed into the 1980s and beyond. The main hypothesis from his work that raised controversy was twofold: first, that there exist long (i.e., 50-60 years) waves of economic activity, and, second, that these long waves are driven by the occurrence of so-called basic innovations. Basic innovations are major technological breakthroughs that radically change the nature of economic activity, and provide new opportunities for growth. Examples of such basic innovations are the steam engine, railways, and, more recently, the computer. In his historical analysis, Schumpeter observed three such long waves since the beginning of modern capitalism. Freeman and Soete (1997), in a similar vein, observed five such long waves, i.e., two new ones occurred since the end of Schumpeter's analysis, and identified basic innovations that are associated with each of these.

Schumpeter's specific hypothesis was that basic innovations tend to cluster in the depression periods of the long waves, i.e., when activity rates and profits are low. The basic innovations create a 'bandwagon' of smaller (incremental) innovation, thus replacing existing capital goods ('creative destruction'), and providing new opportunities for expansion. Once the opportunity for further technological improvements runs out, the economy enters the downward part of the cycle again, until the process starts again.

Such a theory obviously raises high hopes on the stimulating effect of innovation on economic growth. Thus, when the world economy entered a major depression in the 1970s and 1980s, 'neo-Schumpeterians' started to promote innovation as a way out of the crisis. This, among other things, raised a debate on the empirical validity of the Schumpeterian idea of a relation between basic innovations and the existence of a long wave.

Schumpeter's ideas had been criticized as early as 1940, in a book review by Kuznets (1940). Kuznets argued that Schumpeter did not convincingly show that basic innovations indeed tend to be clustered in depression periods, and, moreover, that he did not provide a satisfactory theoretical explanation for this phenomenon. The debate in the 1970s and 1980s can be well summarized as a rehearsal of this Schumpeter-Kuznets controversy.

This paper intends to pick up an important suggestion in this debate that has so far remained unexplored. This suggestion was made by Silverberg and Lehnert (1994). They argue that the statistical methods used by the various researchers who tried to answer the question whether or not basic innovations tend to cluster in depression periods are not valid as a result of the specific nature of the underlying data. They correctly point out that innovations are events, rather than stocks or flows, and the distribution that is used to describe the occurrence of innovations must take this into account. As early as 1974, Sahal (1974) had indeed suggested to use a statistical method based on the Poisson distribution to describe time series of incremental innovation.

In the mean time, so-called Poisson regression has become a standard part of the econometric toolbox (e.g., Green, 1995). Thus, the question whether or not the arrival rate of basic innovations fluctuates over time (as the long wave hypothesis would suggest) can well be tackled by a regression model that is in line with the suggestions made by Silverberg and Lehnert. This is exactly what this paper intends to do, by using the same basic innovation databases as were used by authors such as Kleinknecht and Mensch in the 1970s and 1980s.

The rest of this paper is organized as follows. In Section 2, the empirical literature on basic innovation time series and long waves will be briefly reviewed. This section will also introduce the critique to this debate by Silverberg and Lehnert in more detail. Section 3 presents the Poisson regression model. Section 4 implements this model in the context of several time series for basic innovations. Two major research questions are investigated. The first of these asks what is the nature of the relation between the arrival rate of basic innovations and time. Can a process be observed in which there are major fluctuations over time, or do we see a steady monotone increase of the arrival rate? Second, is the arrival rate of basic innovations in any way related to economic variables such as the rate of economic growth? Note that this question touches on a difficult issue with regard to causality, which cannot be solved with the existing data. The estimations that will be provided below therefore aim at testing whether some correlation between basic innovation and economic growth exists, leaving the issue of causality aside.

The last section summarizes the main findings, and draws some conclusions.

2. Innovation time series. Review of the literature and shortcomings

The 1970s and 1980s saw an intense debate on the Schumpeterian hypothesis of clustering innovations. The start of this debate was Mensch (1979), who argued strongly in favour of the Schumpeterian hypothesis that says that basic innovations tend to be clustered in depression periods of the long wave. Mensch also set the tone with regard to methodology in the debate. His approach was to construct a time series of basic innovations, and to use this time series to count the number of innovations occurring each year (using year of innovation rather than year of invention). The Schumpeterian hypothesis was then tested by checking whether or not the average number of innovations was higher in depression periods than in boom periods. The latter periods were determined on the basis of economic data used to test for the existence of long waves. Mensch used a runs test to statistically test the clustering hypothesis.

Haustein and Neuwirth (1982) compiled a different time series on basic innovations, and used spectral analysis to test periodicity in this time series and other time series related to the long wave. They concluded that there are “some doubts when looking at the regular patterns of inventions and innovations by Mensch and Marchetti” (p. 67). In other words, their results suggest a less strictly periodic pattern than was suggested by Mensch.

In the mean time, the interpretation of the innovation time series data by Mensch gained support from van Duijn (1983), Kleinknecht (1981) and Kleinknecht (1987), who provided similar calculations as Mensch with more extensive data sets. However, the original results obtained by Mensch were severely criticized by Freeman, Clark et al. (1982), who argued that the data used and collected by Mensch were not representative and wrongly dated. A further contribution that was critical to the idea that basic innovations tend to cluster in depression periods was made by Solomou (1986). He applied a *z*-test for the null hypothesis that the mean number of basic innovations in two separate periods were drawn from a normal distribution with the same mean.

Kleinknecht (1990) responded to the criticism raised by constructing a new time series of basic innovations that was a compilation of three different time series used earlier (Mensch, Haustein and Neuwirth, Van Duijn). He used a similar null-hypothesis to that applied by Solomou, but relied on a *t*-test rather than a *z*-test. His conclusions were strongly in favour of clustering of basic innovations in the depression periods of the long wave.

Kleinknecht on his turn was strongly criticized by Silverberg and Lehnert (1994). Discussing the contributions by Solomou (1986) and Kleinknecht (1990), they argue that

> "*z* and *t* tests ... are only applicable to a normally distributed random variable. On *a priori* grounds we have argued that the null hypothesis on innovations must be that they are homogeneous Poisson distributed, however, and a histogram of, for example, the Haustein and Neuwirth data (as well as any of the other series we have examined) confirms that they are anything but normally distributed ... both authors apply their tests to sub samples they claim have been selected on *a priori* criteria ... In general, the periodisations employed derive from previous authors such as Mensch, whose runs test did not depend on it, or on the examination of growth rates and the addition of a time lag ... growth rates and a moving average of the innovation data may be highly (cross) correlated, so that the selection of a proper lag against variations in the growth rate series may simply be a method to select sub periods of above- and below-average innovation activity even from a completely random series. This fact would further invalidate any means test (even one appropriate to a Poisson process, such as a binomial statistic)". (p. 98-99)

Thus, Silverberg and Lehnert argue that the statistical methodology used by Solomou and Kleinknecht is flawed, and suggest using a Poisson regression model to model fluctuations in time series of basic innovations (p. 101/103). However, their suggestion has not been taken up so far. This paper intends to fill this gap.

3. Poisson regression

An innovation is an event in time, rather than a flow or stock of some type. Most economic variables (such as production, employment or capital) are either flows or stocks. Contrary to stocks or flows, events are usefully measured only in terms of (non-negative) integer numbers. Hence the point by Silverberg and Lehnert that the number of innovations can not usefully be approximated by a normal distribution, or any distribution that does not take into account the integer nature of the data.

A simplifying assumption is that the probability of occurrence of an innovation within a given interval of time is independent of previous innovations and independent of time. As noted by Silverberg and Lehnert (p. 80), “this is precisely the definition of a time-homogeneous Poisson process”. The probability of a random variable Y generated by a Poisson distribution occurring y times during an interval (choosing units in such a way that the interval’s length is one) is given by

$$Prob(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!},$$

where $y \in \mathbb{N}$, and λ is a parameter, often referred to as the arrival rate of the random variable Y. Note that λ is not necessarily an integer number. The expected number of events per period is equal to λ, which also happens to be the variance of the distribution.

The parameter λ may also be specified endogenously, for example as in $\ln \lambda = \beta'\mathbf{x}$, where $\mathbf{x}$ is a vector of independent variables, and β is a parameter vector. Other specifications for λ are possible, but the example above is often used because it is convenient for estimation purposes. In this case, the parameter vector β can be interpreted as a vector of elasticities of the arrival rate with regard to the independent variables. Such an approach obviously allows one to test the assumption that the Poisson process is time-homogeneous by setting up the null hypothesis that all elements of β referring to other variables than a constant are equal to zero.

Sahal (1974) proposed to use the Poisson model in order to examine the characteristics of various time series of innovations in specific industries. His conclusion was that “invention is properly characterized as a Poisson random process, but [...] its rate is a function of economic forces” (p. 403). Although Sahal did provide some estimates of β-type parameters, these were based on ordinary least square methods. However, when the data contain many zero and small integer values, a maximum likelihood approach based explicitly on the Poisson distribution may be more appropriate. Such a procedure was introduced into the literature on innovation by Hausman, Hall et al. (1984). They estimated a model in which the number of patents of a firm is related to the firm’s R&D expenditures. Elaborations on the approach by Hausman, Hall et al. (1984) were presented by Crepon and Duguet (1997), Crepon and Duguet (1997) and Cincera (1997).

One problem with the Poisson model is its characteristic that the mean and variance of the distribution are equal. The empirical data often show a larger variance than mean for the dependent variable, which is termed ‘overdispersion’. For example, Hausman, Hall et al. (1984) observed such a phenomenon in their firm level patent database. A model that can account for overdispersion may be obtained by adding an unobserved random effect to the mean of the Poisson distribution (Hausman, Hall et al., 1984). Greene (1995) shows how this leads to a modified probability function of the type:

$$Prob(Y=y|u) = \frac{e^{-\lambda u}(\lambda u)^y}{y!},$$

where u is a random variable for which some distribution must be assumed. The variable u may, for example, reflect random noise, or cross-sectional heterogeneity (when the model is estimated in the cross-section dimension). Assuming that u is gamma distributed, one obtains the following unconditional distribution (Green, 1995):

$$Prob(Y=y|\mathbf{x}) = \frac{\Gamma(\theta+y)}{\Gamma(y+1)\Gamma(\theta)} r^y (1-r)^\theta, \quad \text{where } r = \frac{\lambda}{\lambda+\theta}.$$

This distribution is known as the negative binomial distribution, and has mean λ and variance $\lambda(1+\lambda/\theta)$, for $\theta>0$. When $\theta=0$, the models reduces to a standard Poisson model, and the variance becomes equal to λ again. The negative binomial model can also be estimated using a maximum likelihood method. Testing the Poisson distribution against the negative binomial distribution can be implemented by a Likelihood Ratio test or a Wald test (Green, 1995) of the null hypothesis $\theta=0$.

4. Empirical results

This paper will follow the suggestion made by Silverberg and Lehnert (1994), and use a Poisson regression approach to model time series of basic innovations. Figure 1 shows histograms for six of the most commonly used time series of basic innovations. Three of these stem from the patent data set prepared by Baker (1976), which was used extensively by Kleinknecht. Kleinknecht (1987) divided the Baker patents into two sets: process innovations and product innovations. The first three histograms in the figure present results for the complete Baker data set (‘Baker Total’), the Baker product innovations (‘Baker - Product’), and the Baker process innovations (‘Baker - process’). The Baker time series run from 1769 until 1970, and thus comprises 202 years. The fourth data set was collected by Haustein and Neuwirth (1982), and runs from 1764 until 1975 (212 years). The fifth data set stems from Kleinknecht (1981). This is the shortest of the six time series, running from 1879 to 1965 (87 years). The last data set is from van Duijn (1983), and runs from 1811 - 1971 (161 years).

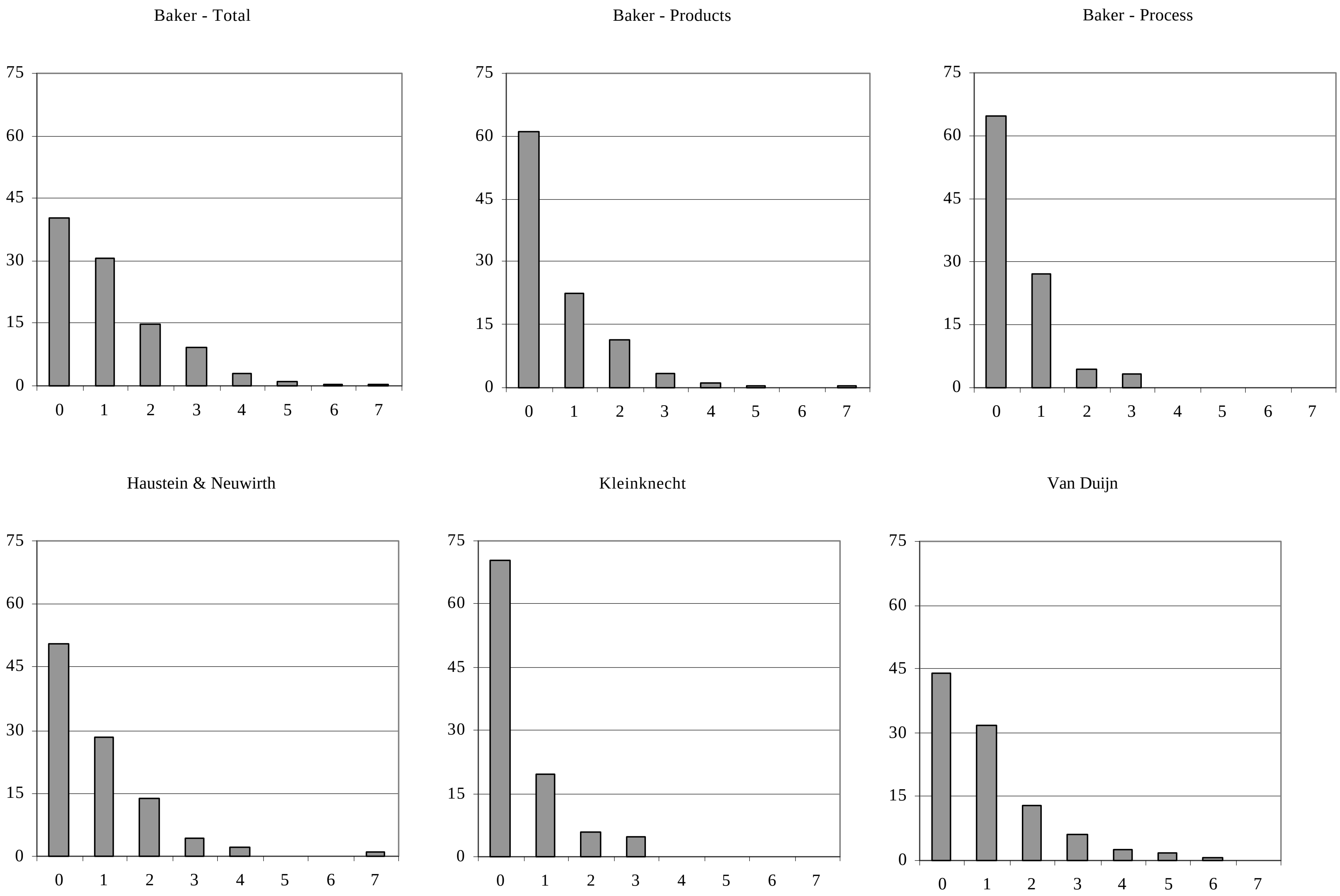

Figure 1. Histograms of basic innovation time series data, on horizontal axis: number of innovations per year; on vertical axis: percentage share of total sample

Table 1. Regression results, Poisson and negative binomial models, arrival rate function of time

Innovation Data Source	N	Start year	*a*	p	*b*	p	θ	p(LR)	Ps. R^2	Best poly nomial	No. max	No. min	Ps. R^2	p(LR)
Poisson model														
Baker - All	202	1769	-0.7488	0.000	0.0073	0.000			0.06	4	1 (1895)	2 (1783, 1942)	0.09	
Baker - Products	202	1769	-1.8707	0.000	0.0115	0.000			0.11	2	1 (1925)	0	0.15	
Baker - Process	202	1769	-0.9807	0.000	0.0020	0.269			0.00	4	1 (1859)	2 (1782, 1932)	0.05	
Haustein & Neuwirth	212	1764	-1.2195	0.000	0.0086	0.000			0.08	5	2 (1878, 1951)	2 (1782, 1902)	0.12	
Kleinknecht	87	1879	-3.8377	0.001	0.0184	0.007			0.05	4	1 (1938)	1 (1896)	0.17	
Van Duijn	161	1811	-0.9835	0.000	0.0072	0.000			0.04	5	2 (1879, 1946)	2 (1825, 1904)	0.07	
US Patents	208	1790	7.1212	0.000	0.0192	0.000			0.86	5	0	0	0.97	
Negative Binomial model														
Baker - All	202	1769	-0.7867	0.000	0.0076	0.000	0.2422	0.011	0.05	2	1 (1926)	0	0.07	0.038
Baker - Products	202	1769	-2.0143	0.000	0.0127	0.000	0.5313	0.001	0.09	3	1 (1924)	0	0.12	0.021
Baker - Process	202	1769	-0.9843	0.000	0.0020	0.302	0.3611	0.126	0.00	4	1 (1859)	2 (1782, 1933)	0.05	0.511
Haustein & Neuwirth	212	1764	-1.2871	0.000	0.0091	0.000	0.3563	0.002	0.07	5	2 (1878, 1951)	2 (1782, 1901)	0.10	0.019
Kleinknecht	87	1879	-4.0982	0.004	0.0200	0.018	0.8532	0.045	0.04	4	1 (1938)	1 (1896)	0.15	0.596
Van Duijn	161	1811	-1.0321	0.001	0.0076	0.000	0.3803	0.003	0.03	5	2 (1880, 1947)	2 (1825, 1904)	0.05	0.030
US Patents	208	1790	4.3454	0.000	0.0371	0.000	0.7806	0.000	0.07	4	0	0	0.16	0.000

Notes: N is number of observations, p gives *p*-value for parameter estimate in previous column, p(LR) is the *p*-value of a Likelihood ratio test of the hypothesis $\theta=0$ (i.e., Poisson vs negative binomial), Ps. R^2 gives pseudo R^2 statistic, 'Best polynomial' gives the order of the highest-order polynomial model in which the last term was significant (orders up till 5 attempted), 'No. max' and 'No. min' gives the number of maxima and minima (respectively) in the 'Best polynomial' model (years between brackets), start- and end-years not counted as minima or maxima. All p-values for parameter estimates (other than p(LR)) are based on a 2-sided *z*-test.

All six histograms show that the highest frequency is found for zero innovations. Also, all histograms show declining frequencies for larger numbers of innovations per year. No time series shows more than seven innovations per year (this occurs twice in the Haustein and Neuwirth series). Silverberg and Lehnert (1994) showed similar histograms and concluded that these show that the assumption of normally distributed data must be rejected at face value.[1] Obviously, this also holds for the six histograms shown here. As noted by Silverberg and Lehnert, this invalidates the statistical methods used by Kleinknecht (1990).

The shape of the histograms suggests that fitting a Poisson or negative binomial model to the time series might be a useful approach. Table 1 presents a number of these models. All models in the table specify the arrival rate in the Poisson model as a function of time only. The first part of the table asks the question whether or not the Poisson process underlying the data is homogenous in time. This part presents regressions in which the following model for the arrival rate was used:

$$\ln \lambda = a + b t,$$

where a and b are parameters to be estimated, and t is time (t=1 in 1764 and increases by one each year). When b is found to be significantly positive, this indicates exponential growth of the arrival rate ($\lambda = Ae^{bt}$, with $A = e^a$). A non-significant value of the estimate for b indicates that a constant arrival rate over time is a better interpretation of the data than an exponential trend.

Besides the six time series already introduced above, the table also contains results for the number of patents issued in the U.S. This is used as an indicator of incremental innovations, and is included for comparison of some of the estimated parameter values with those of the time series for basic innovations. The U.S. patents time series runs from 1790 until 1997 (208 years). It is taken from the U.S. Patent and Trademark Office databases.

The results reject the null-hypothesis of a constant arrival rate over time in all but one cases. The one exception is the Baker process series (both Poisson and negative binomial model). The other five time series show highly significant time trends. This result, which was also found by Silverberg and Lehnert using a different test, seems to confirm that the rate of occurrence of basic innovations has expanded over the history of capitalism.

The results also show that the Poisson model is rejected in favour of the negative binomial model. In other words, the regression results seem to support the hypothesis that the data are overdispersed. All p-values for the LR test of the Poisson model versus the negative binomial model are significant, except the model for the Baker process series (which also did not show a significant trend).

The estimated growth rates for the arrival rate differ substantially between the different time series. Of the five series with significant trends, the Van Duijn and Baker All time series show the

[1] Silverberg and Lehnert also present statistical tests, which by and large confirm their visual impression.

lowest growth rate (both 0.76%, all growth rates in this paragraph for the negative binomial model). The Kleinknecht series shows a growth rate of 2%, which is the highest value for the basic innovations time series. The U.S. patents time series shows a growth rate that is much higher: 3.7% (1.85 times the maximum growth rate for the basic innovations time series). Thus, the occurrence of incremental innovations seems to increase more rapidly than that of basic innovations.

It must be noted, however, that the parameter *b* is only one part of the mean of the arrival rate when the negative binomial model is used (this model is indeed to be preferred over the Poisson model, as argued above). In the negative binomial model, the (conditional) mean of the distribution is the sum of λ and a random effect captured by the gamma distribution. In the present time series context, this random effect must be interpreted as a disturbance factor. Although the expected value of this disturbance is zero, it does have a positive variance, and hence the interpretation must be that although the expected value of the number of basic innovations is equal to the predicted value of λ, this is subject to uncertainty that increases with θ.

The pseudo-R^2 statistics in Table 1 are not very high. However, it must be noted that they cannot be readily compared to the more familiar R^2 statistic used in ordinary least squares regressions. The pseudo-R^2 statistic is also known as the 'likelihood ratio index' (Greene, 1995, p. 936), which is defined as 1 - $\ell(\lambda)/\ell(c)$, where $\ell(\lambda)$ is the likelihood of the model for which the statistic is calculated, and $\ell(c)$ is a similar model with only a constant. Thus, the pseudo-R^2 statistic basically measures by how much the likelihood function increases in the estimated model compared to a model with only a constant.

However, the long wave debate in which much of the discussion on the rate of occurrence of basic innovations took place, obviously goes much further than assuming an exponential increase of this rate. In its extreme form, the long wave hypothesis would assume that the rate of basic innovation shows strict periodicity. A weaker form would suggest that, although periodicity may not be fixed, there would at least be major fluctuations of the rate of occurrence of basic innovations over the history of capitalism.

Probably, the extreme interpretation of strict periodicity is now supported only by a minority of the scholars in the field. Moreover, modern time series models (such as unit root models and long memory models[2]) that separation of trend and cycles is difficult if not impossible. The paper therefore proceeds with a rather simple, if not naive, implementation of the weak long wave hypothesis, i.e., that there are major, but not necessarily strictly periodic, fluctuations in the arrival rate of basic innovations. This is the first research question identified in the introduction.

The first model that will be used to test this hypothesis assumes that the arrival rate of the Poisson or negative binomial model is equal to

$$\ln\lambda = a + bt + ct^2 + dt^3 + et^4 + ft^5,$$

[2] For an interpretation of this type of models in an evolutionary context, see Silverberg and Verspagen (1999).

where *a*, *b*, *c*, *d*, *e* and *f* are parameters that will be estimated. The model thus specifies the natural log of the arrival rate as a polynomial function of time. The highest order of the polynomial function, i.e., 5 was chosen arbitrarily. Slightly higher values did not yield useful results. Lower-order polynomial models were also estimated, i.e., for each time series, five models were estimated, one with highest order polynomial value equal to 1..5.

The estimations for the polynomials of order higher than one are not documented (available from the author on request). The column 'Best polynomial' in Table 1 lists the highest order of the polynomial that was selected from these results. Selection was done on the basis of the significance of the highest-order term in the regression. E.g., when the 5th order term in the degree-5 polynomial estimation was not significant, but the 4th order term in the degree-4 estimation was significant, the column will list "4". Significance of terms with lower order was not used in the selection process. The cut-off value for determining significance was a *p*-value of 0.1.

Obviously, a higher order polynomial provides more opportunities for changes in the sign of the slope of the arrival rate as a function of time. When starting- and end-years are not used as peaks or throughs, a time period spanning two (long) waves would embrace three extreme values (either two peaks and a through, or two throughs and a peak). This would require a polynomial of degree 4. A degree five polynomial, which is the highest possible value in the analysis, could thus span two and half (long) waves at most. Note that with an average length of 50-60 years per long wave, one would expect to find 3-4 long waves over a time span of 200 or slightly more years. Thus, in this light, the maximum degree of the polynomial used (5) would appear somewhat low. Note however, that a value of 6 did not yield any significant results.

The main question that will be looked at using the estimations of the polynomial functions of time is how important non-linearities are in modeling the arrival rate of basic innovations. The 'weak' variant of the long wave hypothesis adopted above would imply that these non-linearities are important. This hypothesis would predict that a monotonic trend, as in the first part of Table 1, would not be an adequate model for the innovation time series. Instead, it would predict that the arrival rate of basic innovations grows rapidly in some period, and less rapidly in others. Possibly, one would even find parts of the arrival rate function that are decreasing.

The last columns of Table 1 already give some indication of the importance of such non-linearities. Concentrating on the negative binomial results (which are very similar to the Poisson results anyway), it turns out that of the six basic innovation time series, four have at least one minimum and maximum value. Note that only 'internal' maxima and minima were counted, i.e., the beginning and end points of the time series were not counted as extreme values. In contrast to the basic innovation time series, the data for incremental innovation (US Patents) do not show any extreme values.

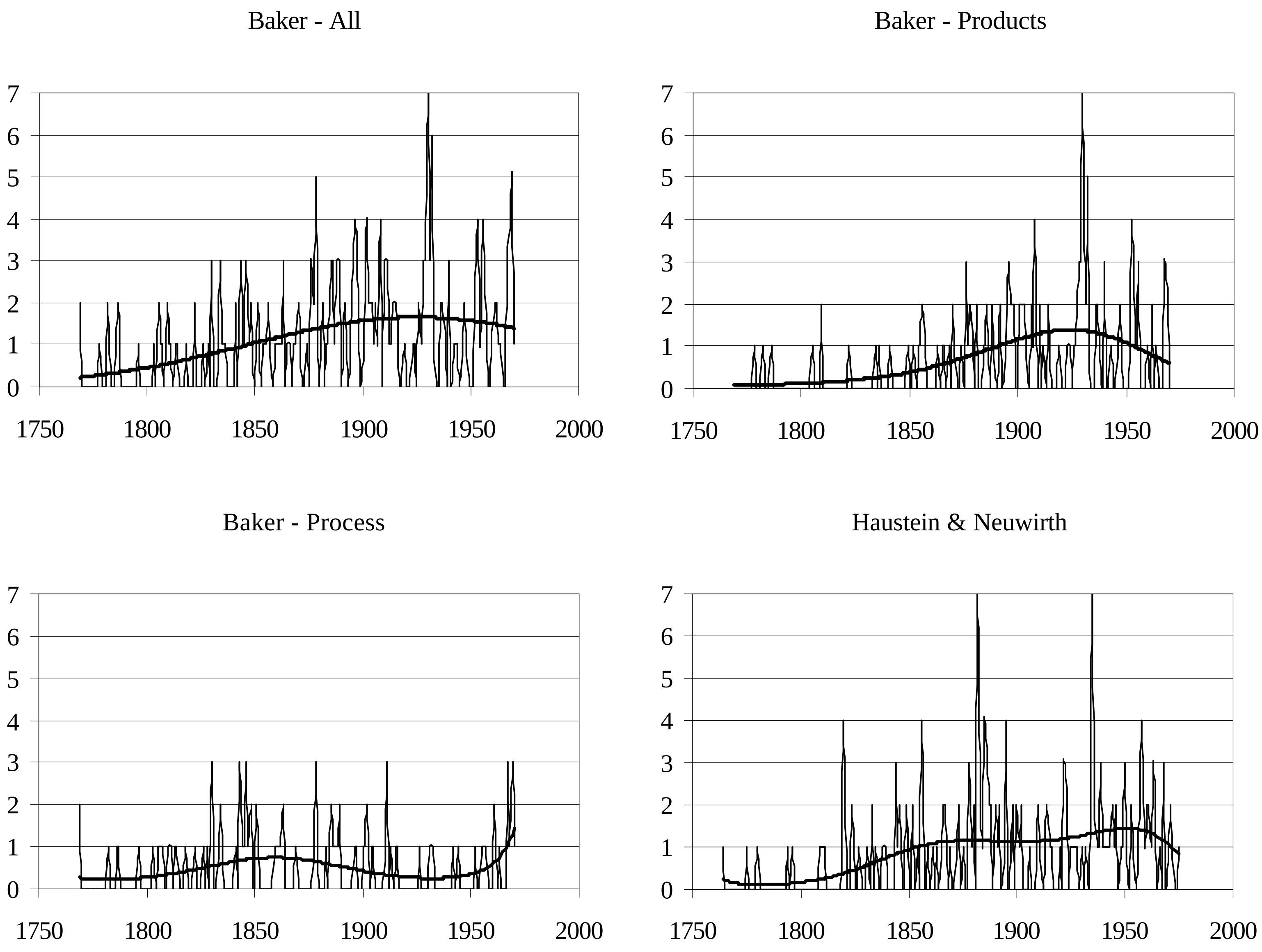

Figure 2. Innovation time series (thin line) and estimated deterministic part of Poisson arrival rate (thick line, estimations in Table 1)

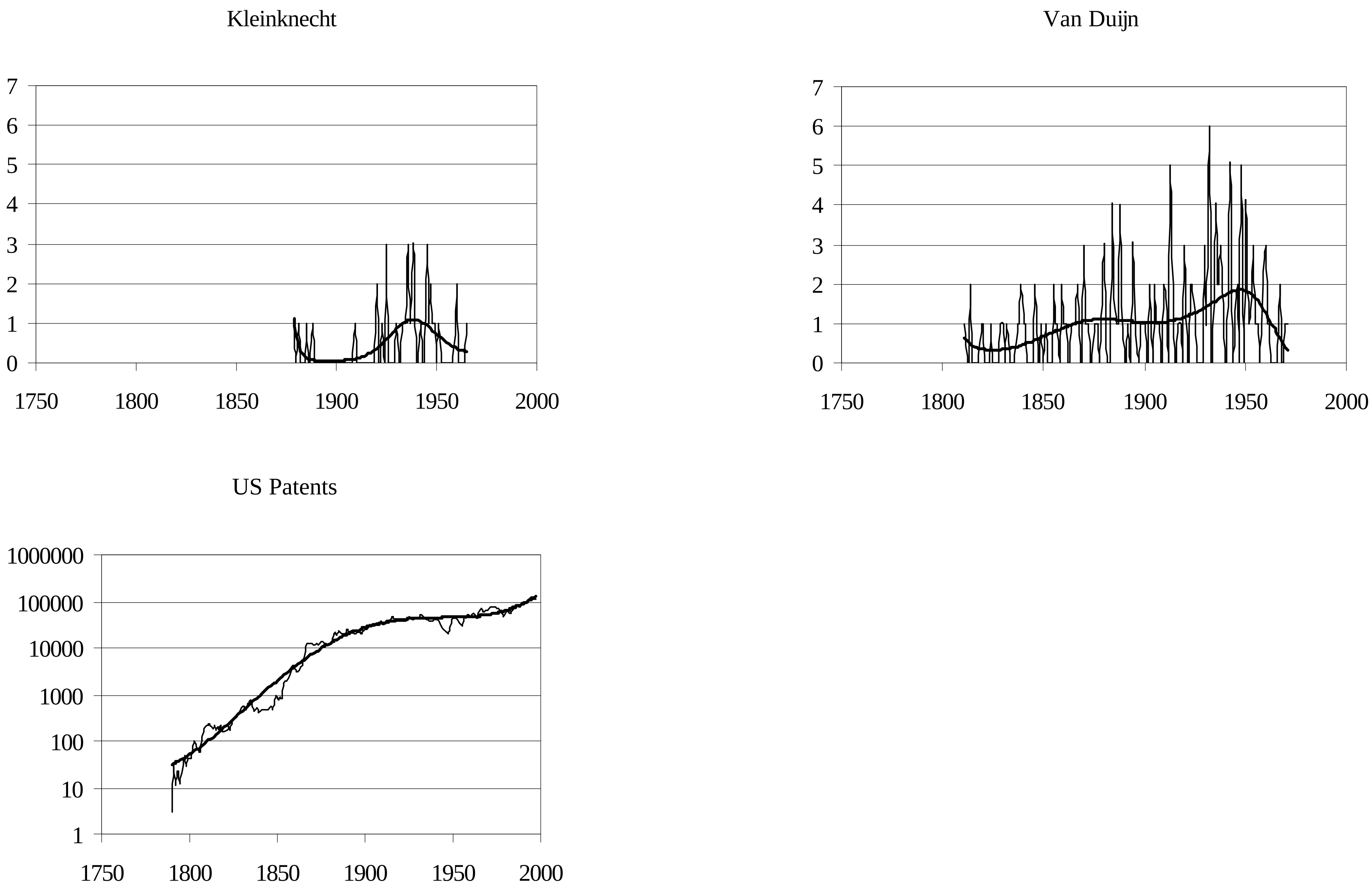

Figure 2. Innovation time series (thin line) and estimated deterministic part of Poisson arrival rate (thick line, estimations in Table 1)

Figure 2 gives a more detailed picture of the shape of the polynomial functions for the arrival rates. The Baker - process, Kleinknecht and Van Duijn time series show the most pronounced non-linear pattern. For the Baker - process and Van Duijn time series, a steady increase of the arrival rate until the second half of the 19th century is observed. After that, there is a fall (Baker - process), or approximately flat trend (Van Duijn). The first half of the 20th century shows a marked increase in the arrival rate for both time series. This upward trend starts later (1933) for the Baker process series than for the Van Duijn series (1904). The Van Duijn series shows a peak in 1947, and a strong decline afterwards. The Baker - process series keeps rising sharply until the end of the sample period. The Kleinknecht series, which is rather short as compared to the other series, shows a similar sharp increase in the first half of the 20th century as the Van Duijn series does. The period before this upswing is only marginally covered by the Kleinknecht series.

The other three time series for basic innovations also show an increase in the arrival rate in the early period. For the Baker total and Baker products time series, this upward trend lasts until the 1920s, for the Haustein and Neuwirth data, the upward trend ends in 1878 (this is more in line with the Van Duijn and Baker process data). The Haustein and Neuwirth data do not show marked fluctuations in the period after 1878, however, as was the case for the Van Duijn and Baker process series. Although there are some fluctuations in the Haustein Neuwirth data in the 20th century, these fluctuations are not very substantial. They are generally in line, however, with the Van Duijn time series in terms of the dates for local maxima and minima.

The conclusion with regard to a (weak) long wave pattern in the innovation time series data remains somewhat undetermined. Some of the time series (Van Duijn, Baker process) support an interpretation of multiple periodic declines and upswings in basic innovation activity. Other time series (Baker products, Haustein & Neuwirth) rather point to a once and for all increase of basic innovation activity during the last part of the 19th century and early 20th century.

In order to investigate this issue further, a new innovation time series was created by merging the two main time series used in Table 1, i.e., Haustein and Neuwirth's data and Van Duijn's data. Kleinknecht's's time series is too short as compared to the two other basic innovation time series. The Baker series were not used because they are based on patent data rather than innovation data, and hence are less compatible with the other two time series.

Kleinknecht (1990) already coined the idea of a merged sample on the basis of the two series used here. He also added innovation data collected by Mensch.[3] The construction of the merged innovation time series used here differs substantially from the Kleinknecht approach, however. The main difference refers to the overlap, i.e., those innovations which are covered in both sources. Kleinknecht constructs a time series in which the innovations that occur more than once in the three time series he considers are double counted. In other words, he simply adds up the numbers of innovations per year in the three innovation time series. Kleinknecht justifies this

[3] As noted above, the sample by Mensch was heavily disputed by Freeman *et al*., which is why these data are not used here.

procedure by arguing that it provides some implicit weighting scheme, in which the important innovations (i.e., those on which all sources agree) are weighted more heavily.

It is clear that such an implicit weighting procedure is not adequate in the context of a Poisson regression approach. This is why a different approach was chosen here. This approach consists of identifying the innovations that are covered by both samples, and counting those only once. A complication in this procedure is that, as noted by Kleinknecht, the innovation dates of the same innovation often differ between the two sources. The majority of overlap cases are dated in a range of 10 year, but differences up to 50 years happen. In all cases, the earliest date was used to assign the innovation to the merged sample. The merged sample contains 88 innovations that occur in both samples[4], 90 innovations that only occur in the Haustein and Neuwirth sample, and 70 cases that are only listed in the Van Duijn sample. The merged series thus has 248 innovations, ranging from 1764 to 1976. The complete listing of all innovations in the merged sample, as well as the original Haustein and Neuwirth and Van Duijn sources, together with their assignment to the merged sample, are given in the annex.

The merged basic innovation sample was used to perform the same regressions as in Table 1 and Figure 2. The hypothesis of a time invariant Poisson process was rejected like in the cases of most other innovation time series. The following equation resulted from the negative binomial model (p-values in brackets):

$\ln \lambda =$	-1.1421	+	0.0104 t,	$\theta =$	0.2663	
	(0.000)		(0.000)		(0.002),	Ps. $R^2 = 0.09$.

Like in most other cases in Table 1, the Poisson model is rejected in favour of the negative binomial model. The 'best polynomial' model that was estimated was a 5-degree negative binomial model (Ps. $R^2 = 0.13$, p(LR) = 0.068). The predicted arrival rate for this model, along with the raw data is depicted in Figure 3.

The results for the merged sample completely support the previous results for the Van Duijn series. There are two periods with a steep increase in the arrival rate of basic innovations: the period 1800- 1860, and the period 1900- 1950. The arrival rate function has an inflection point with an approximately flat level around 1880.

This pattern can be compared in more detail with Kleinknecht's periodization. Note that in the regressions underlying Figure 3, polynomials up to degree 7 were tried, but no values beyond 5 were significant. Thus, the data do not seem to provide any support for more local maxima/minima in the time series for the arrival rate than are present in Figure 3. Kleinknecht (1990, p. 87) finds periods of relatively low innovation activity during 1861-1881, 1901-1927 and

[4] This is a considerably larger overlap than reported by Kleinknecht. He reports 72 overlapping innovations. Part of the difference is explained by the fact that Kleinknecht's limits his merged sample to the period of 1800 - 1968. The merged sample used here contains two cases outside this time span (the blast furnace, 1796, and the microprocessor, 1971). This still leaves 14 of the present overlap cases (or 20% of the Kleinknecht overlap) that Kleinknecht does not consider as overlap. Unfortunately, Kleinknecht does not provide details of what he considers overlap (this is not important for his analysis anyway).

1962-1968. Periods with high innovation activity in his analysis are 1881-1901 and 1927-1962.[5] From the point of view of the predicted Poisson arrival rate in Figure 3, this periodization is not very enlightening. The first three periods together (1861-1927, low-high-low in Kleinknecht's analysis) comprise a period with a rather flat trend in Figure 3, and cannot be readily distinguished from each other in terms of the arrival rate. The period before Kleinknecht's start (pre-1860s), on the other hand, is an interesting period in terms of the arrival rate, because it shows a marked increase.

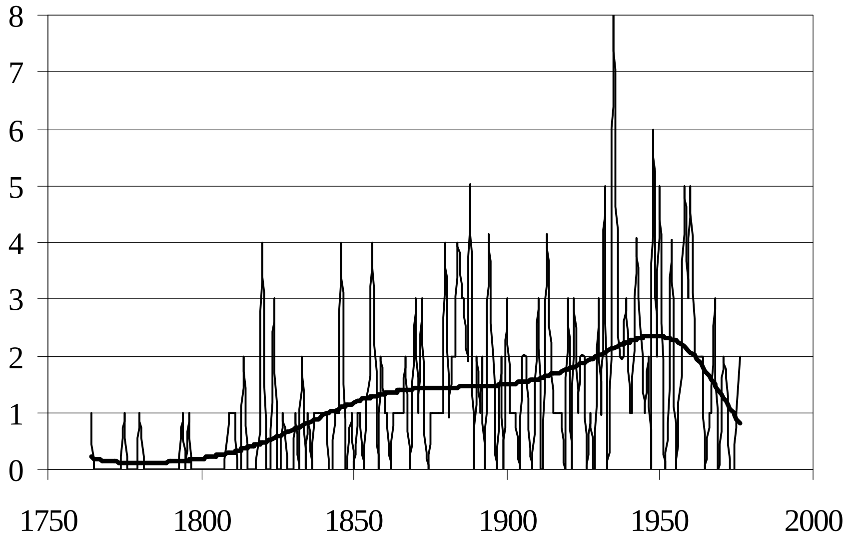

Figure 3. Data and estimated arrival rate for merged sample

The conclusions from Figure 3 match more closely to Kleinknecht's interpretation for the period after 1927. He finds high activity for 1927-1962, which is largely consistent with Figure 3 (although Figure 3 would place the peak somewhat earlier than 1962). Also the falling trend after the 1950s confirms Kleinknecht's interpretation.

Thus, the conclusion seems to be that there are major fluctuations over time in the arrival rate of basic innovations. However, these fluctuations do not seem to be of a shape as suggested by a long wave hypothesis. More specifically, the observed fluctuations in the arrival rate, among other things, have a lower frequency than would be suggested by a long waves interpretation. Instead of three or more complete cycles since 1750, the data tend to support an interpretation in which the arrival rate of basic innovations sharply increased during two distinct periods in

[5] These periods are from the part of his table labeled '12 years lead of innovation wave on economic long wave'. Conclusions for periods with 15 years lead are similar. Kleinknecht does not provide tests for years before 1860.

history: the first half of the 19th century and the period before the second world war.

This raises the question as to whether these fluctuations can be related to economic variables such as the growth rate of GDP, investment, profits, or capacity utilization. This is the second research question identified in the introduction. Unfortunately, time series of 200 years or more are hard to come by in economics statistics. Therefore, it is only possible to obtain a handful of indicators to which the time series of basic innovations can be related. The primary goal of the analysis is to test for a relation between the long-run cyclical behaviour of the world economy and the time series of basic innovations.

The primary source of the data used is Dumenil and Levy (1993), who present a long-run database for the U.S. economy. Three indicators are taken from this database. The first is the profit rate (profits as a percentage of the capital stock) in the business part of the U.S. economy. The second is the rate of capacity utilization, which is estimated by Dumenil and Levy on the basis of their data on capital stocks and GDP. It must be noted that this variable is derived from business cycles in those variables, rather than directly measured (for details see Dumenil and Levy). The third indicator from Dumenil and Levy is the logarithmic deviation of GDP from an exponential trend ('GDP cycle'). In addition to these indicators, an additional indicator for investment in the United Kingdom was taken from Maddison (1993). The variable used is constructed in a similar way as the GDP cycle variable, and is termed Investment cycle.

All of these variables are admittedly selected primarily on the basis of the length of the time series, rather than on the basis of their theoretical relevance or methodological quality in preparing the data. For any time series that goes back into the 19th or 18th century, the quality of the data will be poor compared to present day methodological standards. All four variables are used as broad indicators of the state of affairs in the world economy, although they refer specifically to a single country only. The basic innovations in the various time series obviously stem from a larger set of countries than just the U.S. and the U.K. Unfortunately, however, an indicator of sufficient time span for the world economy as a whole is not available.

Moreover, the issue of causality is a difficult one. The long waves discussion suggest that causality runs two ways, i.e., from innovations to the economic variables used here, and backwards. For example, Mensch (1979) and Kleinknecht (1987) put forward the so-called 'depression trigger' hypothesis as an explanation of why basic innovations tend to cluster in depression periods. This hypothesis states that entrepreneurs show satisfycing behaviour, and that in times of booms, the normal profit rates that can be realized are so high as to discourage investment in risky major innovations. Only in depression times, when 'normal' profits soar, will entrepreneurs be willing to take the risks associated with investment in basic innovations. This 'theory' would suggest that (low) profits 'cause' basic innovations. Schumpeter's original ideas obviously state that basic innovations lead to higher profit rates, after a time lag during which the basic innovation begins to diffuse. Thus, the causal relationship between profits and basic innovations would be a complex and multi-directional one.

The statistical analysis here cannot give any indication of causality. The aim is therefore relatively modest: to investigate whether or not there is some statistically significant relationship between the economic variables used and the basic innovations time series data. When and if such a relationship is indeed found, this must be interpreted as a possible direction for further theoretical work, rather than a confirmation of rejection of a certain theoretical perspective. Note, however, that none of the economic variables were lagged or leading the innovation time series.

Table 2 gives the results of the analysis. The table presents negative binomial regressions including a constant, a linear time trend (as before), and one of the four economic variables introduced above. As in Table 1, the U.S. patent variable is included to compare the results for basic innovations to results for incremental innovations. The latter variable is labeled ‘Variable X’. As for Table 1, the equations were also estimated in a pure Poisson form, but because the majority of cases pointed to a significant improvement when the negative binomial model was used (see the ‘p(LR)’ column), only these results are presented. In cases where the p(LR) column shows a value above 0.1, i.e., where the Poisson model was not rejected, the Poisson results were similar to the ones documented.

The table points out that about one third of all equations (11 of 32 cases) show a significant ‘variable X’. There are three time series for which none of the economic variables is significant: Baker process, Haustein and Neuwirth and Van Duijn. For the Baker - all series, only one variable is significant (investment cycle). For the (short) Kleinknecht series, two variables are significant: capacity utilization and GDP cycle. This is similar for the merged basic innovation time series of Figure 3. For Baker products, three of the four economic variables are significant. Only the GDP cycle variable is not significant in this case. The U.S. patent time series (incremental innovation) also shows three significant variables. Here the investment cycle variable is not significant.

Thus, the evidence of a systematic relationship between economic variables and the arrival rate of basic innovations is at most rather weak. It thus seems as if any theory that suggests a general economic mechanism, operating over the whole history of capitalism, to explain the rate of basic innovation, does not find much empirical support from the regressions presented here. The long run fluctuations in the rate of basic innovation that were identified before (Table 1 and Figures 2 and 3) thus seem to be related to economic developments only in a partial way. Other, not primarily economic factors might be an important part of the story. One may think, for example, of the changing role of (basic) science, institutional changes (such as the formation of industrial R&D labs), or government policy.

Table 2. Negative binomial regressions using ‘economic’ explanatory variables

Innovation Data Source and ‘Var X’	N	Start date	Constant	p	Time	p	Var X	p	θ	p(LR)	Ps. R^2
Baker-All, Profits	102	1869	1.0397	0.062	0.0006	0.855	-2.3640	0.036	0.1906	0.059	0.01
Baker-All, Capacity Utilization	102	1869	0.5250	0.293	-0.0006	0.842	-0.9460	0.257	0.2224	0.031	0.00
Baker-All, Investment cycle	141	1830	-0.0928	0.745	0.0032	0.103	0.5034	0.041	0.1576	0.078	0.01
Baker-All, GDP cycle	102	1869	0.4956	0.322	0.0003	0.929	-0.3216	0.679	0.2325	0.025	0.00
Baker-Products, Profits	102	1869	0.9597	0.167	0.0005	0.902	-3.3425	0.015	0.3022	0.037	0.02
Baker-Products, Capacity Utilization	102	1869	0.2560	0.679	-0.0015	0.702	-2.0052	0.043	0.3372	0.022	0.01
Baker-Products, Investment cycle	141	1830	-1.3084	0.002	0.0083	0.003	0.6629	0.054	0.4002	0.007	0.03
Baker-Products, GDP cycle	102	1869	0.1537	0.804	-0.0004	0.909	-1.0374	0.266	0.3766	0.013	0.00
Baker-Process, Profits	102	1869	-0.7535	0.417	0.0004	0.936	-0.1545	0.936	0.3547	0.267	0.00
Baker-Process, Capacity Utilization	102	1869	-0.7769	0.338	0.0004	0.939	1.8866	0.231	0.3215	0.304	0.01
Baker-Process, Investment cycle	141	1830	-0.1232	0.772	-0.0037	0.228	0.3688	0.345	0.4024	0.114	0.01
Baker-Process, GDP cycle	102	1869	-0.7062	0.397	-0.0003	0.950	1.4751	0.286	0.3273	0.298	0.01
Haust & Neuw, Profits	107	1869	0.2661	0.648	-0.0002	0.954	-0.0418	0.972	0.1985	0.064	0.00
Haust & Neuw, Capacity Utilization	107	1869	0.2819	0.577	-0.0005	0.862	-1.1830	0.159	0.1791	0.090	0.01
Haust & Neuw, Investment cycle	146	1830	-0.4452	0.166	0.0039	0.069	0.1025	0.693	0.2432	0.023	0.01
Haust & Neuw, GDP cycle	107	1869	0.2069	0.678	0.0001	0.972	-1.2060	0.126	0.1777	0.091	0.01
Kleinknecht, Profits	87	1879	-4.0956	0.004	0.0201	0.031	-0.0523	0.985	0.8519	0.047	0.04
Kleinknecht, Capacity Utilization	87	1879	-3.9853	0.004	0.0186	0.025	-2.5918	0.050	0.5600	0.161	0.06
Kleinknecht, Investment cycle	87	1879	-4.0128	0.005	0.0192	0.024	-0.3168	0.522	0.8100	0.056	0.04
Kleinknecht, GDP cycle	87	1879	-3.5551	0.010	0.0165	0.042	-2.9957	0.025	0.5191	0.187	0.07
Van Duijn, Profits	103	1869	-0.0250	0.971	0.0022	0.569	-0.3512	0.791	0.4707	0.001	0.00
Van Duijn, Capacity Utilization	103	1869	-0.0814	0.894	0.0017	0.647	-0.9312	0.326	0.4536	0.002	0.00
Van Duijn, Investment cycle	142	1830	-0.7794	0.033	0.0060	0.015	-0.1246	0.657	0.3736	0.003	0.02
Van Duijn, GDP cycle	103	1869	-0.1448	0.811	0.0023	0.543	-1.0719	0.229	0.4461	0.002	0.01
Merged, Profits	108	1869	0.4739	0.339	0.0014	0.611	-0.3825	0.710	0.1696	0.044	0.00
Merged, Capacity Utilization	108	1869	0.4060	0.343	0.0008	0.748	-1.4481	0.038	0.1351	0.102	0.01
Merged, Investment cycle	147	1830	-0.3238	0.239	0.0053	0.003	-0.0969	0.645	0.1617	0.039	0.02
Merged, GDP cycle	108	1869	0.3142	0.453	0.0016	0.525	-1.5813	0.015	0.1260	0.123	0.02
US Patents, Profits	121	1869	8.7538	0.000	0.0137	0.000	-1.8846	0.000	0.0335	0.000	0.09
US Patents, Capacity Utilization	121	1869	8.2891	0.000	0.0132	0.000	-0.4393	0.034	0.0510	0.000	0.07
US Patents, Investment cycle	150	1830	5.7549	0.000	0.0286	0.000	0.1702	0.301	0.4684	0.000	0.05
US Patents, GDP cycle	121	1869	8.2804	0.000	0.0133	0.000	0.3627	0.043	0.0512	0.000	0.07

Note: all cases where LR rejects negative binomial yield similar results for Poisson model. Notes: N is number of observations, p gives *p*-value for parameter estimate in previous column, p(LR) is the *p*-value of a Likelihood ratio test of the hypothesis θ=0 (i.e., Poisson vs negative binomial), Ps. R^2 gives pseudo R^2 statistic, ‘Best polynomial’ gives the order of the highest-order polynomial model in which the last term was significant (orders up till 5 attempted), ‘No. max’ and ‘No. min’ gives the number of maxima and minima (respectively) in the ‘Best polynomial’ model (years between brackets), start- and end-years not counted as minima or maxima. All p-values for parameter estimates (other than p(LR)) are based on a 2-sided *z*-test.

The few variables that are significant seem to point to a consistent sign of the relationship, however. The rate of capacity utilization is significant in three cases, and all three of these show negative signs. In other words, when capacity utilization is low, basic innovations are more frequent. The GDP cycle variable is significant twice, and again the signs are both negative. When GDP is below its trend level, basic innovations are more frequent. Although the analysis here has not tried to identify a long wave pattern in the GDP data, both of these findings, if anything, support the ideas by Schumpeter, Mensch and Kleinknecht that basic innovations tend to be more frequent in the downswings of the long wave.

The investment cycle variable is also significant in two cases. Both of these cases show a positive sign, indicating that basic innovations tend to occur in periods of high investment. The relationship between the investment cycle and the GDP cycle is not very clear. Long wave theory does not seem to have a unique answer to the question whether investment leads or lags GDP cycles. The data used here does not point to a significant correlation between GDP and investment cycles. Thus, the positive signs on the investment cycles cannot easily be related to the long wave hypothesis, as was done for GDP and capacity utilization. In any case, the finding that basic innovations are facilitated by a positive attitude towards investment does not seem to be unreasonable. Finally, the one case where the profit rate variable is significant points to a positive relation.

The U.S. patent variable, compared to most series of basic innovations, shows rather strong results, with three of the four economic variables turning up significantly. There is only one time series for basic innovations that shows three significant variables. It is interesting to note that this time series is also constructed by using patent data. Thus, it seems as if patents, or incremental innovations in general, are much more strongly related to economic variables than basic innovations. This is indeed an intuitive finding, which could be expected on the basis of the results of, for example, Schmookler (1966).

5. Summary and conclusions

This paper took up a suggestion made by Silverberg and Lehnert (1994), when they discussed the empirical evidence on the 'Schumpeterian hypothesis' that basic innovations tend to cluster in depression periods. Silverberg and Lehnert argued that the statistical methods used in most of the papers that investigate this hypothesis empirically are flawed because they rely on a normal distribution to describe the innovation data. A Poisson distribution is more appropriate because of the fact that innovations are events, rather than stocks or flows.

This paper thus used the technique of Poisson regression to model the occurrence of basic innovations in the period since the mid-1700s. It uses the same basic data as were used by authors such as Mensch, Kleinknecht and Solomou before. Two sets of regressions were undertaken, each aimed at a different research question. The first set of regressions investigated whether or not there are any major fluctuations over time in the rate at which basic innovations occur. To this, end, the arrival rate of basic innovations was modeled as a polynomial function of time. The results indicate that there are indeed such fluctuations. A number of time series for basic innovations, although not all of them, suggest a major increase in the arrival rate in the first part of the 19th century. This increasing trend seems to level off after (roughly) 1860. Another sharp increase takes place from the 1920s to 1950. These long-run movements suggest that basic innovation is not a 'steady' and regular phenomenon. Major changes in the speed of basic innovation seem to have taken place in the history of capitalism. These changes do not, however, seem to be so frequent and regular as to support for a true long wave pattern in the Schumpeterian sense.

The second set of regressions asked the question whether there is any relationship between basic innovation and a number of economic variables. Although the quality of the applied data is weak, and causality cannot be established, the results seem to point out that these relationships are weak. Economic factors such as the investment cycle, GDP cycle, capacity utilization rate or the profit rate are only weakly related to basic innovations. The results for the capacity utilization rate and the GDP cycle do confirm, however, that innovations tend to be more frequent in depression periods.

Is there a final verdict on the debate between the supporters and critics of the Schumpeterian hypothesis of basic innovations clustering in depressions? Probably not. The results point to a weakly significant relationship of the type that Schumpeter, Mensch and Kleinknecht have advanced (i.e., basic innovations are more frequent in depressions). But at the same time, long-run changes in the speed of basic innovations seem to be less frequent than a long wave theory would suggest. All of this points to one direction: long-run changes in the rate of basic innovation are a real phenomenon, but the idea of a strictly periodic long weave is too simple. Moreover, the causes of such long-run changes in the frequency of basic innovations are likely to be rather complex, and probably beyond the reach of a simple economic theory.

References

Baker, R. (1976). *New and improved: inventors and inventions that have changed the modern world*. London, British Museum Publications.

Cincera, M. (1997). "Patents, R&D and technological spillovers at the firm level: some evidence from econometric count models for panel data." *Journal of Applied Econometrics* **12**: 265-280.

Crepon, B. and E. Duguet (1997). "Estimating the innovation function from patent numbers: GMM on count panel data." *Journal of Applied Econometrics* **12**: 243-263.

Crepon, B. and E. Duguet (1997). "Research and development, competition and innovation pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity." *Journal of Econometrics* **79**: 355-378.

Dumenil, G. and D. Levy (1993). *The Economics of the Profit Rate. Competition, Crises and*

Historical Tendencies in Capitalism. Aldershot, Edward Elgar.

Freeman, C., J. Clark, et al. (1982). *Unemployment and Technical Innovation*. London, Pinter.

Freeman, C. and L. Soete (1997). *The Economics of Industrial Innovation. 3rd Edition*. London and Washington, Pinter.

Greene, W. (1995). *Econometric Analysis, 3rd ed.* New Jersey, Prentice Hall.

Hausman, J., B. H. Hall, et al. (1984). “Econometric Models for Count Data with an Application to the Patents-R&D Relationship.” *Econometrica* **52**: 909-938.

Haustein, H.-D. and E. Neuwirth (1982). “Long Waves in World Industrial Production, Energy Consumption, Innovations, Inventions, and Patents and Their Identification by Spectral Analysis.” *Technological Forecasting and Social Change* **22**: 53-89.

Kleinknecht, A. (1981). “Observations on the Schumpeterian Swarming of Innovations.” *Futures* **13**.

Kleinknecht, A. (1987). *Innovation Patterns in Crisis and Prosperity. Schumpeter's Long Cycle Reconsidered*. London, Macmillan.

Kleinknecht, A. (1990). “Are there Schumpeterian Waves of Innovations.” *Cambridge Journal of Economics* **14**: 81-92.

Kuznets, S. (1940). “Schumpeter's Business Cycles.” *American Economic Review* **30**.

Maddison, A. (1993). “Standardised Estimates of Fixed Capital Stock: A Six Country Comparison.” *Innovazione e Materie Prime*.

Mensch, G. (1979). *Stalemate in Technology. Innovations Overcome Depression*. Cambridge, Ballinger.

Sahal, D. (1974). “Generalized Poisson and Related Models of Technological Innovation.” *Technological Forecasting and Social Change* **6**: 403-436.

Schmookler, J. (1966). *Invention and Economic Growth*. Cambridge MA, Harvard University Press.

Schumpeter, J. A. (1939). *Business Cycles: A theoretical, historical and statistical analysis of the capitalist process*. New York, McGraw-Hill.

Silverberg, G. and D. Lehnert (1994). Growth fluctuations in an evolutionary model of creative destruction. *The Economics of Growth and Technical Change. technologies, Nations, Agents*. G. Silverberg and L. Soete. Aldershot, Edward Elgar**:** 74-108.

Silverberg, G. and B. Verspagen (1999). Long Memory in Time Series of Economic Growth and Convergence. Maastricht, MERIT.

Solomou, S. (1986). “Innovation clusters and Kondratieff long waves in economic growth.” *Cambridge Journal of Economics* **10**.

Van Duijn, J. J. (1983). *The Long Wave in Economic Life*. London, Allen & Unwin.

Annex. Data on basic innovations

Table A1. Merged sample of basic innovations (used in Figure 3)

Item number	Innovation	Year
89	Spinning machine	1764
90	Steam engine	1775
91	Automatic band loom	1780
92	Sliding carriage	1794
81	Blast furnace	1796
48	Steam ship	1809
95	Whitney's method	1810
80	Crucible steel	1811
183	Street lighting (gas)	1814
184	Mechanical printing press	1814
78	Lead chamber process	1819
77	Quinine	1820
98	Isolated conduction	1820
99	Rolled wire	1820
100	Cartwright's loom	1820
3	Steam locomotive	1824
61	Cement	1824
66	Puddling furnace	1824
101	Pharma fabrication	1827
102	Calciumchlorate	1831
79	Telegraphy	1833
103	Urban gas	1833
104	Rolled rails	1835
87	Electric motor	1837
67	Photography	1838
9	Bicycle	1839
88	Vulcanized rubber	1840
7	Arc lamp	1841
105	Jacquard loom	1844
106	Lathe	1845
107	Inductor	1846
108	Electrodynamic measuring	1846
185	Rotary press	1846
186	Anaestetics	1846
187	Steel (puddling process)	1849
188	Sewing machine	1851
109	Plaster of paris	1852
14	Aluminium	1854
40	Safety match	1855
189	Bunsen burner	1855
36	Refined steel/Bessemer steel	1856
84	Steel pen / Fountain pen	1856

110	Tare colours industry	1856
111	Baking powder	1856
190	Elevator	1857
76	Lead battery	1859
191	Drilling for oil	1859
54	Internal combustion engine	1860
39	Soda works	1861
19	Anilin dyes	1863
192	Siemens-Martin steel	1864
112	Paper from wood	1865
4	Deep sea cable	1866
50	Dynamite	1867
75	Dynamo	1867
113	Commutator	1869
60	Typewriter	1870
193	Celluloid	1870
194	Combine harvester	1870
85	Margarine	1871
6	Thomas steel	1872
46	Reinforced concrete	1872
114	Drum rotor	1872
115	Preservatives	1873
51	Sulphuric acid	1875
195	Four-stroke engine	1876
5	Telephone	1877
116	Nickel	1878
53	Electric Railway	1879
12	Incandescent lamp	1880
47	Water turbine	1880
117	Jodoforme	1880
196	Half-tone process	1880
197	Electric power station	1881
118	Veronal	1882
119	Cable	1882
120	Antipyrin	1883
121	Coals whisks	1883
10	Steam turbine	1884
122	Chloroforme	1884
198	Punched card	1884
199	Cash register	1884
38	Syntethic fertilizers	1885
52	Transformer	1885
123	Synthetic alcaloids	1885
124	Magnesium	1886
125	Electric welding	1886
200	Linotype	1886
72	Phonographe	1887

126	Electrolyse	1887
28	Motor car	1888
59	Pneumatic tyre	1888
127	Electric counter	1888
201	Portable camera	1888
202	Alternating-current generator	1888
86	Man-made fibres	1890
128	Chemical fibres	1890
129	Melting by induction	1891
83	Acetylene welding	1892
130	Accounting machine	1892
41	Cinematography	1894
131	Antitoxines	1894
203	Motor cycle	1894
204	Monotype	1894
8	Diesel engine	1895
132	Drilling machine for mining	1895
205	Electric automobile	1895
206	X-rays	1896
37	Aspirin	1898
133	Arc welding	1898
134	Air ship	1900
135	Synthesis of indigo	1900
207	Submarine	1900
136	Holing machine	1901
137	Electric steel making	1902
208	Safety razor	1903
209	Viscose rayon	1905
210	Vacuum cleaner	1905
138	Acetylen	1906
211	Chemical accelerator for rubber vulcanization	1906
212	Electric washing machine	1907
15	Gyro compass	1909
2	Airplane	1910
69	Bakelite (Phenol plastics)	1910
139	High voltage isolation	1910
65	Vacuum tube	1913
71	Assembly line	1913
213	Thermal cracking	1913
214	Domestic refrigerator	1913
140	Ammonia synthesis	1914
141	Tractor	1914
215	Stainless steel	1914
142	Tank	1915
32	Synthetic rubber	1916
42	Cellophane	1917

1	Zip fastener	1918
29	AM Radio	1920
216	Acetate rayon	1920
217	Continuous thermal cracking	1920
26	Synthetic detergents	1922
57	Insuline	1922
143	Synthesis of methanol	1922
35	Continuous rolling	1923
218	Dynamic loudspeaker	1924
219	Leica camera	1924
144	Deep frozen food	1925
220	Electric record player	1925
145	Coal hydrogenation	1927
17	Power steering	1930
221	Polystyrene	1930
222	Rapid freezing	1930
223	Freon refrigerants	1931
34	Crease-resistent fabrics	1932
224	Gas turbine	1932
225	Polyvinylchloride	1932
226	Antimalaria drugs	1932
227	Sulfa drugs	1932
56	Fluor lamp	1934
146	Diesel locomotive	1934
147	Fischer-Tropsch procedure	1934
11	Radar	1935
13	Ballpoint pen	1935
30	Rockets/guided missiles	1935
31	Plexiglas	1935
62	Magnetophone	1935
70	Catalytic cracking	1935
82	Colour photo	1935
148	Gasoline	1935
16	Television	1936
149	Photoelectric cell	1936
228	FM radio	1936
150	Vitamins	1937
229	Electron microscope	1937
20	Helicopter	1938
230	Nylon	1938
21	Polethylene	1939
55	Automatic gears	1939
151	Hydraulic gear	1939
27	Antibiotics (penicilline)	1940
152	Cotton picker	1941
43	Jet engine/plane	1942
45	DDT	1942

153	Heavy water	1942
231	Continuous catalytic cracking	1942
24	Silcones	1943
232	Aerosol spray	1943
233	High-energy accelerators	1943
44	Streptomycine	1944
154	Titanreduction	1944
22	Sulzer loom	1945
68	Oxygen steelmaking	1946
234	Phototype	1946
49	Numerically controlled machine tools	1948
58	Continuous steel making	1948
235	Orlon	1948
236	Cortisone	1948
237	Long-playing record	1948
238	Polaroid land camera	1948
155	Thonet furniture	1949
156	Polyester	1949
18	Computer	1950
23	Transistor	1950
25	Xerography	1950
239	Terylene	1950
240	Radial tyre	1950
157	Double-floor railway	1951
158	Cinerama	1953
241	Colour television	1953
33	Nuclear energy	1954
242	Gas chromatograph	1954
243	Remote control	1954
244	Silicon transistor	1954
159	Air compressed building	1956
160	Atomic ice breaker	1957
161	Space travel	1957
162	Stitching bond	1958
163	Holography	1958
164	Transistor radio	1958
165	Diffusion process	1958
245	Fuel cell	1958
166	Quartz clocks	1959
246	Polyacetates	1959
247	Float glas	1959
167	Maser	1960
168	Micro modules	1960
248	Polycarbonates	1960
249	Contraceptive pill	1960
250	Hovercraft	1960
64	Integrated circuit	1961

169	Planar process	1961
73	Laser	1962
251	Communication satellite	1962
170	Implementation of ions	1963
171	Epitaxy	1963
172	Synthetic leather	1964
173	Transistor laser	1964
174	Optoelectronic diodes	1966
252	Wankel-motor	1967
74	Video	1968
175	Light emitting fluor display	1968
176	Minicomputers	1968
177	Quartz watches	1970
63	Microprocessor	1971
178	Electronic calculator	1971
179	Light-tunnel technology	1972
180	16-bit microprocessor	1975
181	16384 bit RAM	1976
182	Microcomputer	1976

Note: items 1-88 occur in both databases, items 89-182 occur only in Haustein and Neuwirth data, items 183-252 occur only in Van Duijn data.

Table A2. Innovations in the Haustein and Neuwirth time series that were matched to innovations in the Van Duijn time series

Innovation	Year	To which item in merged series?
Blast furnace	1796	to 81
Steamer	1809	to 48
Crucible cast steel	1811	to 80
Lead-chamber process	1819	to 78
Chinin fabrication	1820	to 77
Locomotive	1824	to 3
Puddling furnace	1824	to 66
Telegraphy	1833	to 79
Photography	1838	to 67
Bicycle (pedal)	1839	to 9
Cement	1844	to 61
Arc lamp	1844	to 7
Generator of current	1849	to 87
Hard rubber	1852	to 88
Aluminium	1854	to 14
Refined steel	1856	to 36
Steel pen	1856	to 84
Lead accumulator	1859	to 76
Soda works	1861	to 39
Production of analin	1863	to 19
Deep sea cable	1866	to 4

Safety matches	1866	to 40
Dynamite	1867	to 50
Dynamo	1867	to 75
Thomas steel	1872	to 6
Typewriter	1873	to 60
Sulphuric acid production	1875	to 51
Telephone	1878	to 5
Electric locomotive	1879	to 53
Incandescent lamp	1880	to 12
Cooking fat	1882	to 85
Electricity	1882	to 87
Electric heating	1882	to 87
Long distance conduction	1882	to 87
Synthetic fertilizers	1885	to 38
Transformers	1885	to 52
Combustion engine	1886	to 54
Phonograph	1887	to 72
Tyres with air compression	1888	to 59
Water turbine	1890	to 47
Welding by acetylene	1892	to 83
Steam turbine	1895	to 10
Automobile	1895	to 28
Cinematography	1895	to 41
Electric railway	1895	to 53
Diesel engine	1897	to 8
Aspirin	1898	to 37
Steel concrete	1902	to 46
Gyro compass	1909	to 15
Pheno plastics	1910	to 69
Airplane	1911	to 2
Conveyor belt production	1913	to 71
Synthetic rubber	1916	to 32
Electronic tubes	1920	to 65
Detergents/synthetic	1922	to 26
Radio	1922	to 29
Insuline	1922	to 57
Zip fastener	1923	to 1
Continuous rolling	1923	to 35
Cellophane	1926	to 42
Power steering	1930	to 17
Crease-resistant fabrics	1932	to 34
Fluorescent lamp	1934	to 56
Ball-point pen	1935	to 13
Rockets	1935	to 30
Plexiglass	1935	to 31
Magnetophone	1935	to 62
Catalytic cracking	1935	to 70

Colour film	1935	to 82
TV	1936	to 16
Radar	1939	to 11
Helicopter	1939	to 20
Automatic gears	1939	to 55
Synthetic fibres	1939	to 86
Antibiotics	1940	to 27
DDT	1942	to 45
Jet engine	1943	to 43
Streptomycine	1944	to 44
Sulzer loom	1945	to 22
Silicons	1946	to 24
Oxygen-process	1946	to 68
NC machines	1948	to 49
Continuous steelmaking	1948	to 58
Computer	1950	to 18
Transistor	1950	to 23
Xerographie	1950	to 50
Polyethylene	1953	to 21
Nuclear power station	1954	to 33
Integrated circuits	1961	to 64
Laser	1962	to 73
Video-tape recorder	1968	to 74
Microprocessor	1971	to 63

Table A3. Innovations in the Van Duijn time series that were matched to innovations in the Haustein and Neuwirth time series

Innovation	Year	To which item in merged series?
Crucible steel	1811	to 80
Sulphuric acid (lead chamber process)	1819	to 78
Quinine	1820	to 77
Portland cement	1824	to 61
Coke blast furnace	1829	to 81
Steam locomotive	1830	to 3
Puddling furnace	1832	to 66
Electric motor	1837	to 87
Steamship (Atlantic crossing)	1838	to 48
Photography	1839	to 67
Electric telegraph	1839	to 79
Vulcanized rubber	1840	to 88
Arc lamp	1841	to 7
Safety match	1855	to 40
Bessemer steel	1856	to 36
Lead battery	1859	to 76
Internal combustion engine	1860	to 54
Sodium carbonate	1861	to 39
Aniline dyes	1865	to 19

Atlantic telegraph cable	1866	to 4
Dynamite	1867	to 50
Dynamo	1867	to 75
Typewriter	1870	to 60
Margarine	1871	to 85
Reinforced concrete	1872	to 46
Sulphuric acid	1875	to 51
Telephone	1877	to 5
Electric railway	1879	to 53
Thomas oven	1879	to 6
Incandescent lamp	1880	to 12
Water turbine	1880	to 47
Steam turbine	1884	to 10
Fountain pen	1884	to 84
Transformer	1885	to 52
Bicycle	1885	to 9
Aluminium	1887	to 14
Motor car	1888	to 28
Cylindrical record player	1888	to 72
Pneumatic tyre	1889	to 59
Mechanical record player	1889	to 72
Rayon (nitro-cellulose pr.)	1892	to 86
Motion picture film	1894	to 41
Diesel engine	1895	to 8
Rayon (cuprammonium pr.)	1898	to 86
Aspririn	1899	to 37
Oxy-acetylene welding	1903	to 83
Airplane	1910	to 2
Bakelite	1910	to 69
Gyro compass	1911	to 15
Synthetic fertilizer (nitrogen)	1913	to 38
Vacuum tube	1913	to 65
Assembly line	1913	to 71
Cellophane	1917	to 42
Zip fastener	1918	to 1
AM radio	1920	to 29
Continuous hot strip rolling	1923	to 35
Insulin	1923	to 57
Synthetic detergents	1930	to 26
Synthetic rubber	1932	to 32
Crease-resisting fabrics	1932	to 34
Radar	1935	to 11
Plexiglas	1935	to 31
Magnetic tape recorder	1935	to 62
Colour photography	1935	to 82
Television	1936	to 16
Catalytic cracking	1937	to 70

Helicopter	1938	to 20
Fluorescent lamp	1938	to 56
Polyethylene	1939	to 21
Penicillin	1942	to 27
Guided missiles	1942	to 30
Jet airplane	1942	to 43
DDT	1942	to 45
Silicones	1943	to 24
Ball-point pen	1945	to 13
Streptomycin	1946	to 44
Automatic transmission (passenger cars)	1948	to 55
Sulzer loom	1950	to 22
Xerography	1950	to 25
Power steering (passenger cars)	1951	to 17
Electronic computer	1951	to 18
Transistor	1951	to 23
Continuous casting of steel	1952	to 58
Oxygen steel making	1953	to 68
Numerically controlled machine tools	1955	to 49
Nuclear energy	1956	to 33
Integrated circuit	1961	to 64
Laser	1967	to 73
Video cassette recorder	1970	to 74
Micro-processor	1971	to 63