DISCUSSION PAPER SERIES

No. 4200

INDIVIDUAL MORTALITY AND MACROECONOMIC CONDITIONS FROM BIRTH TO DEATH

Maarten Lindeboom, France Portrait and
Gerard J van den Berg

LABOUR ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP4200.asp**

ISSN 0265-8003

INDIVIDUAL MORTALITY AND MACROECONOMIC CONDITIONS FROM BIRTH TO DEATH

Maarten Lindeboom, Vrije Universiteit Amsterdam
France Portrait, Vrije Universiteit Amsterdam
Gerard J van den Berg, Vrije Universiteit Amsterdam and CEPR

Discussion Paper No. 4200
January 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4200

January 2004

ABSTRACT

Individual Mortality and Macroeconomic Conditions from Birth to Death*

This Paper analyses the effects of macroeconomic conditions throughout life on the individual mortality rate. We estimate flexible duration models where the individual's mortality rate depends on current conditions, conditions earlier in life (notably during childhood), calendar time, age, individual characteristics, including individual socio-economic indicators, and interaction terms. We use individual data records from Dutch registers of birth, marriage, and death certificates, covering an observation window of unprecedented size (1812-1999). These are merged with historical data on macroeconomic and health indicators. The results indicate a strong effect of macroeconomic conditions during childhood on mortality at all ages. Those who are born in bad times on average have a high mortality rate throughout life, in particular during childhood itself and at ages above 50. Current macroeconomic conditions mostly have an effect on youths and on the elderly.

JEL Classification: C50, I10, J10 and N30
Keywords: business cycle, death, epidemics, health, life expectancy, lifetimes, longevity and recession

Maarten Lindeboom
Department of Economics
Vrije Universiteit Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
THE NETHERLANDS
Tel: (31 20) 444 6033
Email: mlindeboom@econ.vu.nl

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=127149

France Portrait
Department of Economics
Vrije Universiteit Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
THE NETHERLANDS
Tel: (31 20) 444 6155
Email: fportrait@econ.vu.nl

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=156359

Gerard J van den Berg
Department of Economics
Vrije Universiteit Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
THE NETHERLANDS
Tel: (31 20) 444 6132
Fax: (31 20) 444 6020
Email: gjvdberg@xs4all.nl

*This Paper was previously distributed under the name 'Changes in longevity and economic conditions'. The 'Historical Sample of the Netherlands Dataset Utrecht release UT.03' was kindly provided by the International Institute of Social History (IISG) in Amsterdam. We acknowledge comments and suggestions of Paul Gregg, Sören Hojgard and participants of the 2002 European Workshop on Econometrics and Health Economics (Lund), the ESPE 2003 Meeting (New York), and the ESEM 2003 (Stockholm).This Paper is produced as part of a CEPR Research Network on 'New Techniques for the Evaluation of European Labour Market Policies', funded by the European Commission under the Research Training Network Programme (Contract No: HPRN-CT-2000-00071).

Submitted 12 November 2003

1. Introduction

This paper analyzes the effects of macro-economic conditions throughout life on the individual mortality rate. In particular, we aim to distinguish between the effects of contemporaneous macro-economic conditions and the effects of macro-economic conditions earlier in life, controlling for individual-specific socio-economic conditions and other individual mortality determinants.

The effects of contemporaneous conditions and of individual-specific socio-economic conditions on mortality have been much investigated (see e.g. Cutler and Meara, 2001, and references below) and are relatively well understood. A high income enables individuals to spend on goods that improve health and living conditions. At the aggregate level, a high GNP is associated with high public health expenditures, good access to medical care, and a high incidence of medical innovations. All of this leads to low mortality rates. In some cases, increases in GNP are also associated with increases in urbanization leading to epidemics and high mortality (Haines, 2001). A temporary downturn has a positive effect on mortality because of increased insecurity and stress associated with joblessness (Brenner, 1979), and a negative effect because of reduced work-related stress and work injuries and because of reduced opportunity costs of health enhancing activities like sporting (Ruhm, 2000).

In addition to this, macro-economic conditions during pregnancy of the mother and childhood may also have effects on mortality later in life. It is well documented that poor living conditions early in life lead to susceptibility to a wide range of health problems later in life (see e.g. Case, Fertig and Paxson, 2003, and references therein). It is also well documented that exposure to disease or malnutrition early in life often leads to an increase in mortality (see e.g. the recent impressive overview in Doblhammer, 2003). In a recession, the provision of sufficient nutrients and good living conditions for children and pregnant women may be hampered. From a methodological point of view, it is important to take these effects into account in order to understand mortality patterns. For example, if birth cohorts that endured adverse macro-economic conditions early in life respond differently to shocks later in life, and if this is not taken into account, then the effects of contemporaneous conditions may be confounded or misinterpreted. Moreover, knowledge on the magnitude of such effects has important policy implications. If being born in bad times has a positive long-run effect on

mortality then the value of life is reduced for those affected.[1] This would increase the benefits of policies that help children and pregnant women in recessions, for example by way of enhanced provision of food, housing, and health care (see Almond, 2002, for a detailed discussion of policy implications).

In the literature that addresses the effects of macro-economic conditions on mortality, the effects mentioned in the previous paragraph have been difficult to analyze, because of a lack of a sufficiently long time span in the data. For example, Himes (1994), Ruhm (2000), Attanasio and Emmerson (2001), Ferrie (2003), Henderson (2001), and Lichtenberg (2002) use data sets covering 1960-1985, 1972-1991, 1988-1994, 1850-1860, [1900-1903 and 1992-1996], and 1960-1997, respectively. For essentially the same reason, these studies suffer from initial conditions problems, meaning that their samples condition on survival up to the beginning of the observation window. The latter implies that cohorts who were born before that date are only represented by their fittest members. Moreover, often only mortality of adults is examined. These issues are particularly problematic if there is much unobserved heterogeneity in the data.

The data in the existing studies are typically aggregated at the regional or national level. Explaining aggregated mortality rates out of macro-economic conditions also gives rise to endogeneity problems. To some extent these can be handled by using fixed effects panel data methods (Ruhm, 2000). However, both mortality and aggregate production may depend on idiosyncratic shocks. Other studies in the literature focus on the effects of economic conditions on infant mortality (e.g. Lynch and Greenhouse, 1994, and Pritchett and Summers, 1996) but then of course one cannot distinguish between effects of current and past economic conditions.

As mentioned above, some alternative approaches have been applied to investigate the effects of conditions and events early in life on outcomes later in life. A number of studies estimate the effects of individual socio-economic conditions during childhood on health outcomes later in life. However, individual variation in childhood conditions and later health outcomes may be jointly affected by unobserved heterogeneity, leading to simultaneity bias. Contrary to this, using macro-economic conditions during early childhood as determinants of individual mortality does not give rise to such a bias, because these conditions are exogenous

[1] Murphy and Topel (2003) demonstrate that the gains from mortality rate reductions have been enormous.

from the individual point of view. In epidemiology, natural experiments and instrumental variables have recently been applied to estimate the effects of nutrition and disease exposure on mortality later in life, using longitudinal data. For example, Doblhammer (2003) uses month of birth, whereas others use epidemics, wars, famines, or the rate of infant mortality in the cohort, as instruments for conditions in utero and early in life. These studies are primarily focused on medical explanations, and not on economic conditions or policies.

In this paper we advance on the literature by using a unique dataset of individual records of Dutch citizens, called the Historical Sample of The Netherlands (HSN). Our sample covers around 3,000 individuals born in the province of Utrecht in the period 1812-1912. These individuals are followed up to 1999. The variables are from the standardized recordings of vital events (birth, marriage and death) kept by municipalities and provinces. These were introduced along with the imposition of French civil law during the Napoleonic wars. The variables include information on individual socio-economic conditions and demographic circumstances. We merge these with historical time-series data from Statistics Netherlands on macro-economic variables like GNP and on agricultural production, and to external information on the incidence of epidemics. There are no reliable 19th century data on health expenditures and medical innovations, so that we only observe the beginning and end of the causal chain from economic conditions to health to mortality. Also, a sample size of 3,000 precludes a detailed analysis of the effects of specific policy measures like the abolition of child labor.

The empirical analysis consists of the estimation of a duration model for individual survival, or, equivalently, for individual mortality or longevity. For a given individual at a given moment in time, the mortality rate depends on current conditions, conditions earlier in life (notably during childhood), calendar time, age, individual characteristics, including individual socio-economic indicators, and interaction terms. Whereas the natural experiments in the epidemiological literature are targeted towards conditions in utero and during the first months of life, we also consider the effects of conditions during subsequent childhood years. The model specification is designed to be flexible. The results are used to assess the importance of current and past economic conditions on individual mortality. We decompose life expectancies into a factor due to contemporaneous macro-economic conditions and a factor due to macro-economic conditions at the beginning of life. Our estimation results also shed light on the effect of illiteracy of the father, the effect of being born out of wedlock, and

the effect of the parents' social class, on individual mortality. As we will see, despite the methodological differences between studies in the literature, the main outcomes are in strong agreement, and they support the view that economic status in childhood is a key determinant of health and mortality in adulthood.

The present paper is organized as follows. Section 2 presents the data set, discusses some variables that we will use in the analyses, and provides descriptives on mortality rates of different age groups and different birth cohorts. Section 3 presents the duration model. The estimation results are presented in Section 4. Section 5 concludes.

2. The data

2.1. Variables

The HSN data have been derived from the registers of birth, marriage, and death certificates (see Mandemakers, 2000, for a general description in English). Currently, we have access to a cleansed sample of 2,975 individuals. This is a random sample of individuals born in the province of Utrecht[2] between 1812 and 1912. The end of the observation window is December 31, 1999.

The data provide information on a limited set of characteristics of the individual, events in his/her life, and his/her socio-economic and demographic environment. Notably, we observe the marital status of the mother at the time of birth, the occupation of the father and the mother, whether or not the father was illiterate, the age of the father and mother at the time of birth of the individual, gender, date of birth, its geographic location, an indicator of whether the individual lived in the countryside at the moments of birth and death, the date of marriage and occupation at the time of marriage, the date of death, and the residence at that date.

The individual lifetime durations are observed in days. If the individual is still alive at the end of 1999 or the individual died in another province or country, then we do not observe

[2] At the time, The Netherlands had 11 provinces. Utrecht is representative of The Netherlands in the sense that it had both rural and urban areas, is located in the center of the country, and had economic characteristics that were similar to the nationwide average.

the date of death, and we can not distinguish between the two explanations for this either. However, if such an individual experienced marriage and/or birth of children while living in the province then the corresponding dates are observed, so the lifetime duration is right-censored at the latest of these dates. Otherwise (i.e., if such events did not occur while living in the province and the individual did not die in the province before 2000), the lifetime duration is right-censored at zero, and the individual is discarded from the data. The latter occurs in 19% of the original sample of 3,669 individuals, leading to our aforementioned sample size of 2,975.

All occupational titles are coded in the 1984 Standard Occupation Classification of Statistics Netherlands. This is a four digit hierarchical code. Subsequently, each occupational codes has been translated into an ISCO occupational code and a social class code. We use the latter in our analyses. It distinguishes between six levels: 1) Lower lower class, 2) Upper lower class, 3) Lower middle class 4) Upper middle class, 5) Lower upper class, 6) Upper upper class. The place of residence at birth has been translated into a urbanization indicator (being 1 iff the individual is born in the cities of Utrecht or Amersfoort).

We merge the individual data records with external information. Most importantly, we use historical time-series data from Statistics Netherlands on annual GNP at each year in the life of the individual (see Smits, Horlings and Luiten van Zanden, 2000). Our choice for GNP instead of obvious alternatives such as GDP is driven by the need for mutually consistent observations for as many years as possible. Figure 1 plots the log annual real per capita GNP in our observation window. It is clear that in addition to the upward trend there are many cyclical fluctuations. Jacobs and Smits (2001) provide a detailed analysis of GDP movements in The Netherlands in the 19th century. Years with low and negative growth are observed more frequently than in the 20th century. GDP fluctuations are strongly correlated to the business cycles in the UK and U.S.. We also considered other macro-economic indicators, like interest and inflation rates (see also Smits, Horlings and Luiten van Zanden, 2000), but these seem less relevant than GNP. As we shall see in Section 3, the analysis is computationally demanding due to the large number of model parameters even with a small set of macro-economic indicators. In most analyses we do however use the share of agricultural production in the annual GNP at the time of birth, as reported by the Dutch Institute of Social History (IISG). This share decreases with industrialization, but in the mid-

19th century it is relatively high due to the effects of trade liberalization between The Netherlands and the UK (Jacobs and Smits, 2001).

In addition to this, we exploit external information on the incidence of epidemics, because these cause pronounced spikes in the mortality rates. At 1830, large epidemics had been absent for around a century, but the period 1830-1875 witnessed a number of dramatic epidemics. The cholera epidemics of 1848-49 and 1866-67 and the smallpox epidemic of 1870-72 each lead to over 20,000 deaths nationwide. This each corresponds to about 0.7% of the population. At the height of such an epidemic, the national annual mortality rate was around 25% higher than otherwise. There were frequent smaller cholera epidemics.[3] After 1875, however, infectious diseases were brought under control, and mortality rates dropped spectacularly. The one notable exception is influenza, which caused an epidemic in 1918 of the same order of magnitude as the worst epidemics of the 19th century.

We also need to take the World War II period (1940-1945) into special account. This has been the only war and occupation on Dutch soil since the Napoleonic era (The Netherlands did not participate in World War I) and included the famine of unprecedented severity of the winter of 1944/45. Mortality rates peaked because of malnutrition and the genocide of Jewish citizens. There are no reliable macro-economic statistics for the World War II period.

Table 1 presents some summary statistics on variables we use below as explanatory variables.[4] Note that the illiteracy indicator is the explanatory variable at the individual level whose average value changes most over the observation window. Missing values of explanatory variables lead to an additional loss of 175 individuals from the sample. In the next subsection we summarize the marginal lifetime duration distributions in the data.[5]

The price to be paid for the fact that the observation window is of unprecedented size concerns the absence of a number of variables that are often used in the mortality literature but that are unobserved in the 19th century records. Notably, we do not observe the

[3] See Ferrie (2001) for the importance of cholera for U.S. mortality in the 1850s.
[4] The level of the "born in urban area" variable is not informative because the sample is stratified by it.
[5] See e.g. Smits, Horlings and Luiten van Zanden (2000) for aggregate time series on birth and mortality in the 19th century. The national population grew from 2.2 million in 1812 to 3.4 million in 1862 to 6.1 million in 1912.

individual's cause of death and aggregate amounts of health expenditures and numbers of medical innovations.

2.2. Descriptive statistics of lifetimes

Table 2 gives the mean, standard deviation and median of lifetimes by birth cohort interval *([1812-1822], [1823-1832], ..., [1903-1912*]), for individuals whose date of death is observed. The last row gives the fraction of individuals whose lifetime is right-censored at a positive value (the over-all fraction in the sample is around 10%). For all but the latest cohorts, these constitute individuals who emigrated out of the province some time after after marriage or birth of children. Note that the fraction for the 1903-1912 birth cohort is relatively low, suggesting that most of these also constitute emigrants instead of individuals who were still alive in 2000. Among the observations that are right-censored at a positive value, 80% has a censored lifetime exceeding 16 years.

It is clear from Table 2 that the mean and median lifetime vary substantially over time and across gender. The dramatic differences between the mean lifetimes of the 1833-1842 and the 1843-1852 birth cohorts are due to the Cholera epidemic of 1849. This epidemic caused widespread death of babies and small children. When reading this table it should however be borne in mind that the estimated standard errors of the mean and median lifetimes by birth cohort interval are non-negligible (typically around 3.5 years for the means). Lifetime distributions have most probability mass close to the boundaries of their support, and because of this the mean and the median are rather unstable sample statistics. It is more informative to examine conditional death probabilities.

Figures 2a-2e present estimates of the conditional probability of death, by birth cohort interval, age category *([0-1), [1-5), [5-20), [20-50)* and *[50-70]),* and gender. For a given gender, the conditional probability of death in age category *[t,t+a)* for a cohort born in year τ-t is defined as the number of individuals in this cohort who die in years $[\tau,\tau+a)$ as a fraction of the number who survived up to τ. These probabilities are aggregated individual mortality rates. The figures report averages over the years within a birth cohort interval. The horizontal axes depict the mid-points of the cohort intervals. The pointwise standard errors of the depicted numbers are typically in the range 0.025-0.05, so that a 95% pointwise confidence interval would have a size in the range 0.1-0.2.

Mortality as a function of age displays the familiar U-shaped curve. Among the uncensored observations, 24% died within a year. For the lowest ages (up to 5), the mortality rate is highest among the cohorts born in the middle of the 19th century. For those aged over 20, mortality is generally declining as a function of calendar time. As usual, the mortality beyond age 50 is higher for men than for women. Note that the mortality of men aged over 50 is remarkably high for those born around the turn of the 20th century and the beginning of the 20th century. This is a well known artifact due to the historically high incidence of smoking within these cohorts of men (Wolleswinkel-Van den Bosch et al., 1998). In addition, it may be that a number of individuals in the latest birth cohort were still alive in 1999 but are missing from the sample because their lifetime durations are censored at zero (see the previous subsection). For this reason, we should not attach great importance to estimation results that are driven by data on old ages for the latest cohort. We return to this in Subsection 4.2.

3. The duration model for individual mortality

3.1. Model specification

The aim is to specify a flexible model for the distribution of the individual lifetime duration T. We take the random variable T to be continuous and nonnegative, so its distribution can be characterized by its hazard rate, which is the individual mortality rate. This mortality rate is a natural starting point of the specification of the model, because of our interest in the dependence of current mortality on current and past macro-economic conditions. Accelerated failure time (AFT) models and regression models are less amenable. For example, AFT models with time-varying explanatory variables (Cox and Oakes, 1984, Van den Berg, 2001) impose a specific structure on the effect of past time-varying explanatory variables on the current mortality rate. This structure entails amongst other things that the effect of values in the recent past has the same sign as the effect of values in the distant past.

Let the variable τ denote current calendar time. It is assumed that all variation in the mortality rate θ of an individual at a given point of time can be explained by the prevailing age t, individual socio-economic and demographic background characteristics X, current

macro-economic conditions $z(\tau)$, macro-economic conditions earlier in life $z(\tau\text{-}t+i)$ $(i \in [0,t))$, calendar time τ itself, and various interaction terms, so, in obvious notation,

$$\lim_{dt \downarrow 0} \frac{\Pr(t \le T < t+dt \mid T \ge t, \tau, X = x)}{dt} \quad = \quad \theta\big(t,\, x,\, z(\tau),\, z(\tau - t + i) \text{ with } \; i \in [0,t),\, \tau\big) \qquad (1)$$

We model the effects of $z(\tau\text{-}t+i)$ $(i \in [0,t))$ to work by way of the value of $z(\tau\text{-}t+i)$ at birth (i.e., at $i=0$) and by way of the average values of $z(\tau\text{-}t+i)$ within some youth age intervals $(i \in [1,6]$, $[7,14]$, and $[15,20])$. For example, if $t=12$ then $\log \theta$ will be specified to be linear in $z(\tau\text{-}12)$, the average of z in $[\tau\text{-}11, \tau\text{-}6]$, and the average of z in $[\tau\text{-}5, \tau]$. To obtain a compact notation, we define $z_1(\tau\text{-}t,t)$, $z_2(\tau\text{-}t,t)$ and $z_3(\tau\text{-}t,t)$ as the average values of $z(\tau\text{-}t+i)$ within the intervals $[1,\min\{6,t\}]$, $[7,\min\{14,t\}]$, and $[15,\min\{20,t\}]$ for i, respectively, with the restriction that $z_j(\tau\text{-}t,t)=0$ if the corresponding interval is empty (so, in the example with $t=12$, $z_3(\tau\text{-}t,t)=0$).

Our baseline specification for θ is,

$$\log \theta\,\big(t,\, x,\, z(\tau),\, z(\tau - t + i) \text{ with } \; i \in [0,t),\, \tau\big) \;=$$
$$\psi_1(t) + \beta' x + \alpha_1' z(\tau) + \alpha_2' z(\tau - t) + \sum_{j=1}^{3} \alpha_{j+2}' z_j(\tau - t, t) + \psi_2(\tau) \qquad (2)$$

where the α_j $(j=1,\ldots,5)$ are the parameters of interest. In fact, we estimate more general model specifications allowing for various interaction effects between the determinants on the right-hand side. Notably, we allow for interactions between t and $z(\tau)$, between t and $z(\tau\text{-}t)$, and between x and $z(\tau)$.

We first discuss the above baseline specification in more detail. The function ψ_1 represents the age dependence of the individual mortality rate. Note that, contrary to the values of z and z_j, the values of the function ψ_1 (and also the function ψ_2) are unobserved and need to be estimated. We adopt a piecewise constant specification for ψ_1, distinguishing between 10 different age intervals. Concerning x we restrict attention to indicators at birth as opposed to later in life, for the reason that the latter may be endogenous or confounded.[6]

[6] We do not use age of parents at birth because this may also be endogenous. Moreover, it is often missing; e.g. for 80% of the mothers.

Concerning birth in an urban area it should be noted that in the 19th century urban areas had higher mortality rates due to increased risks of infectious diseases and poor public health infrastructure (see De Swaan, 1988, for a description of the situation in The Netherlands, and Haines, 2001, for a quantitative analysis). We also include indicators of social class, illiteracy of the father, and birth out of wedlock. Using data from the Dutch city of The Hague in the 1850s, Kok, Van Poppel and Kruse (1997) find that illegitimate children had very little chance of surviving childhood. Underlying explanations concern the age and social class of the mother as well as the lack of financial support.

For $z(\tau)$ we take annual real per capita GNP at τ, as well as dummy variables for years with epidemics and for World War II. The latter also captures the fact that the GNP variable is missing for that period. Note that current GNP may capture the mortality trend effects associated with increased welfare as well as the instantaneous mortality effects of economic cycles (recall the discussion in Section 1). The vector $z(\tau\text{-}t)$ contains annual real per capita GNP at birth $\tau\text{-}t$, as well as the share of agricultural production in the annual GNP at birth. The latter is supposed to capture the degree of industrialization of society and the associated living conditions and public health infrastructure at birth.[7] If current GNP in $z(\tau)$ captures the trend effects of increased welfare, then GNP at birth in $z(\tau\text{-}t)$ captures the long-run effects of the economic conditions at birth. The variables $z_j(\tau\text{-}t,t)$ are constructed using GNP within youth age intervals *[1,6], [7,14],* and *[15,20]*. The third interval covers the moment of labor market entry of most individuals, so the associated parameter α_5 captures the long-run mortality effect of macro-economic conditions at the moment of labor market entry.

The function ψ_2 captures the mortality effects of all nation-wide changes in contemporaneous conditions that are not included in $z(\tau)$. The observation window covers almost 200 years, and over these years the society has made a dramatic transformation from a primarily agricultural society to a modern industrialized society. The epidemiological literature suggests that mortality changes in The Netherlands in this time span have been affected by changes in climate (directly as well as by way of agricultural production), food availability, the dissemination of nutritional, hygienic, and medical knowledge among the

[7] One could include the epidemics and war dummies into $z(\tau\text{-}t)$. For example, having survived an infectious disease may affect subsequent physical well-being, while relatively healthy individuals will be overrepresented among the survivors of an epidemic.

population, and the increase of infectious diseases and public health provisions like sewage and water supply (De Swaan, 1988, Wolleswinkel-Van den Bosch et al., 1998). Most likely, these effects are not all captured by $z(\tau)$. We therefore include an additive flexible function of calendar time to the log individual mortality rate $\log \theta$. Since $z(\tau)$ is (almost) continuous, it is useful that ψ_2 is continuous as well. In particular, we take $\log \psi_2$ to be the sum of Chebyshev polynomials of the second kind[8] in τ, with unknown coefficients. This specification is flexible and concise in that the number of unknown parameters equals the highest order among the polynomials. Note that the epidemics and war effects are identified from ψ_2 because the former are discrete in τ whereas the latter is continuous in τ. The estimated function ψ_2 may capture a wide range of unknown effects, which makes it difficult to interpret. In this sense its parameters are nuisance parameters.

We now turn to the interaction terms between the determinants at the right-hand side of equation (2) that we add to that right-hand side. Interactions between t and $z(\tau)$ arise if vulnerable age categories (like babies and the elderly) suffer disproportionally from living in bad economic conditions. For other age categories they may just lead to temporary bad health, but for babies and the elderly they may directly lead to higher mortality rates. Conversely, improvements in public health may be particularly beneficial to babies and the elderly (see e.g. Cutler and Meara, 2001). Similarly, it is conceivable that certain types of individuals (characterized by certain values of x) suffer disproportionally from bad economic conditions. This leads to interactions between x and $z(\tau)$.

Interactions between t and $z(\tau-t)$ arise if the mortality effect of bad economic conditions during pregnancy of the mother and childhood varies during lifetime (holding everything else constant). By now there is substantial medical and epidemiological evidence that malnutrition of the mother during pregnancy leads her child to have a particularly high incidence of health problems at ages over 50 (see e.g. Koupilová, 1997).

Some comments are in order concerning our modeling strategy. First, effects of per capita GNP on mortality should not be interpreted as (average) individual income elasticities.

[8] Such polynomials are mutually orthogonal in the observation window, ensuring absence of multicollinearity. See Abramowitz and Stegun (1970), for details, and Abbring, Van den Berg and Van Ours (2002) for an application in duration analysis.

GNP is correlated with the levels of technology and public health expenditure. To some extent, the estimated GNP effects will reflect the effects of changes in those conditions.[9]

Secondly, the model, including the interaction effects, is econometrically identified. As is well known, models with additive age, cohort, and time effects are unidentified because year of birth plus age equals calendar time (see Abbring, Van den Berg and Van Ours, 2002 for an analysis in duration models). Suppose that both ψ_1 and ψ_2 are piecewise constant with step size equal to one year (i.e., suppose they can be represented by yearly age and time dummies). If the effect of macro-economic conditions at birth $\alpha_2' z(\tau-t)$ would also be replaced by annual cohort dummies then the model would not be identified. Thus, from an econometric point of view, identification is established by replacing cohort dummies by an unknown linear function of an observable. Of course, in our setting this is not problematic, as we are precisely interested in the causal effect of macro-economic conditions on mortality, and not in a mechanic decomposition of mortality variation into age, time, and cohort effects.[10,11]

3.2. Unobserved heterogeneity

As is well known, ignoring unobserved heterogeneity of mortality determinants across individuals may result in biased estimates of the duration model parameters (see Van den Berg, 2001, for an overview) although the problem is less severe in the virtual absence of right-censoring, as in our case. Proportional Hazard models (specifying $\log\theta$ to be additive in the elapsed duration t and the explanatory variables) can be extended to allow for an additive unobserved heterogeneity term in $\log\theta$, leading to Mixed Proportional Hazard (MPH) models that are still identified. However, in practice the estimates of MPH model parameters are sensitive to arbitrary functional form assumptions. Moreover, as shown by Van den Berg

[9] See also Pritchett and Summers (1996) for a discussion on this.

[10] This is analogous to the analysis of cohort size effects on wages; see e.g. Macunovich (1999).

[11] Our data can be aggregated over individuals and used for nonparametric estimation of age×cohort specific mortality rates. These can subsequently be regressed on age, cohort, and time dependent aggregate indicators, allowing for age and cohort fixed effects, similar to the analyses of Lichtenberg, (2001), Cutler and Meara (2001) and Deaton and Paxson (2001). This would entail the loss of the information on variation across individuals. Moreover, as

(2001), the unobserved heterogeneity distribution is identified from interaction effects between t and x in the observed log hazard, so identification rests on the assumed absence of interaction terms in the individual log hazard. Estimation of models with interaction terms at the individual level *and* unobserved heterogeneity should therefore be performed with caution. On the other hand, it should be noted that most identification results assume absence of time-varying explanatory variables (like our $z(\tau)$) although these may convey useful additional information.

Unobserved heterogeneity poses an additional problem in duration analysis if the current individual hazard rate is allowed to depend on the value of an explanatory variable at a point of time in the past but after the beginning of the spell (like, in our case, the value of $z_1(\tau\text{-}t,t)$, $z_2(\tau\text{-}t,t)$ and $z_3(\tau\text{-}t,t)$). This is most easily explained by way of an example (see e.g. Vaupel and Yashin, 1985). Suppose that GNP has a strong instantaneous effect on mortality between ages 1 and 6 but that GNP between these ages has no effect on mortality later in life, and suppose that individuals differ in terms of whether a "high mortality gene" is present. Now consider the birth cohort of, say, 1847. If GNP is low in 1848-1853 then the survivors after 1853 contain relatively many individuals with the good gene and low mortality throughout their subsequent life. If unobserved heterogeneity is not taken into account, but GNP between ages 1 and 6 is included as an explanatory variable in the mortality rate at all ages, then this variable will pick up the dynamic selection effect driven by the unobserved heterogeneity. This leads to an incorrect conclusion concerning the importance of macro-economic conditions during childhood for mortality later in life. However, as is clear from the example, the dynamic selection effect can be expected to generate a *positive* relation between GNP during childhood and observed mortality later in life, whereas in fact the economic, epidemiological and medical reasoning leads one to expect a *negative* relation. So, when unobserved heterogeneity is not taken into account, it may create an upward bias in the estimated effect of GNP during childhood on the mortality rate later in life. In other words, if we find a negative effect without taking account of unobserved heterogeneity, then the true effect is likely to be at least as negative.

In the empirical analysis below we estimate models without unobserved heterogeneity as well as a model in which an unobserved heterogeneity term is added to the right-hand side

---

indicated in Section 1, it cannot handle joint dependence of mortality and economic conditions on idiosyncratic shocks.

of specification (2) for the log individual mortality rate. We assume that this term has a flexible discrete mass-point distribution[12] in the population of newborns. In Section 5 we outline how additional data could be of help to estimate more general models with unobserved heterogeneity.

For all model specifications, the likelihood function is readily derived (see e.g. Lancaster, 1990). The lifetime durations that are right-censored (recall Subsection 2.1) are treated as independently right-censored observations.[13] We estimate the models with GAUSS maximum likelihood routines.

4. Estimation results

4.1. Parameter estimates

Table 3 presents the estimation results for the model specification (2) including interactions between t and $z(\tau)$. The estimates concern the mortality rate, so a positive value is associated with a short lifetime. For the age dependence function ψ_1 we report the log value. The lifetime time unit is 1 year (but recall that lifetimes are recorded in 1/365 years).

Females, individuals from a higher social class, and individuals whose father was literate have lower mortality rates than their counterparts. The finding with respect to social class is consistent with the large literature on the health-income gradient. The effect of being born out of wedlock is extremely large. The latter is in line with the literature on infant mortality (see Subsection 2.1).

We now turn to the effects of macro-economic conditions. Here we focus on the raw estimates, whereas in Subsection 4.3 we assess the quantitative importance of these determinants in detail, and in Section 5 we discuss the policy implications. The most striking result is that the GNP value during early childhood has a significantly negative effect on the

[12] The discrete distribution is computationally convenient and (with an unspecified number of points of support) does not impose strong functional form restrictions.

[13] The independence assumption can of course be criticized, but the data preclude a further analysis. This also applies to the assumption that individuals with right-censored durations at zero (see Subsection 2.1) are a random subset. The latter assumption is particularly problematic for individuals born close to the end of the observation window.

mortality rate throughout life. In other words, macro-economic conditions during early childhood have long lasting effects on mortality later in life. The early childhood period covers important physical and mental development stages of the child. The availability of health facilities, good living conditions, and sufficient nutrition can have large effects on the success of this development, with long run implications for mortality. The effect of GNP at birth is insignificant. This should not be taken as evidence that economic conditions around the date of birth are irrelevant in the long run. Long run effects of individual socio-economic conditions at birth may simply dominate the long run effects of macro-economic conditions at birth because around birth the nutrients are taken from the mother whereas between ages 1-7 they are taken directly from the food brought into the household. Note that unobserved heterogeneity cannot explain the estimated long run effects, because, from Subsection 3.2, it would give a positive effect of GNP during early childhood on the mortality rate later in life.

We subsequently estimate a model in which the effects of GNP early in life are allowed to vary with the age of the individual. Specifically, we interact the average of GNP at ages 0-7 with an indicator of whether age exceeds 50. The corresponding estimate is equal to -0.20 (standard error 0.09, t-value 2.1). Evidently, bad macro-economic conditions during childhood have a particularly strong effect on mortality at higher ages. This is in line with the recent medical literature mentioned in Subsection 3.1.

The GNP value during the period in which individuals typically enter the labor market (age 15-20) is not of influence for later mortality. The effect of the share of the agricultural sector in GNP has a positive coefficient, implying that industrialization has on average reduced mortality risks.

The estimated instantaneous effect of GNP is also strong. The corresponding coefficients are all negative and mostly significant, implying that mortality rates are lower in periods of high production. The size of the instantaneous effect varies with age. A likelihood ratio test of the restriction that the effect of contemporaneous GNP is age-independent results in rejection.[14] The estimated effect is very large and significant for children below 14. It is also large and significant for individuals aged between 15 and 34. For those aged above 34 the estimated effect is much smaller. Even for those aged above 90, the estimated effect is

[14] Estimates of the restricted model are in Table 4. There are 9 restrictions. The chi-square statistic equals 37.6, which by far exceeds the 95% critical value of 19.0. In the restricted

smaller than for those aged between 15 and 34, although it is highly significant. Across all ages, the effect is strongest for the ages 1-7. Recall that we also found that the GNP in this age interval has the strongest long run effect on mortality later in life. So, macro-economic conditions between age 1 and 7 have the strongest instantaneous effect as well as the strongest long run effect on mortality. Also, like in the case of the long run effects, the fact that the instantaneous macro-economic effect is smaller for babies is probably because the health of babies primarily depends on the health of the mother.

The findings with respect to the instantaneous effects of the GNP are in line with the literature. For example, Pritchett and Summers (1996) find positive effects of GNP on infant mortality in developing countries. The findings are not necessarily at odds with Ruhm (2000), who reports higher mortality rates in periods of economic expansion and attributes this to work-related stress, work injuries, and opportunity costs of health enhancing activities. His data are from the late 20th century, and the underlying explanations may be less relevant in the 19th century, where being out of work could lead to starvation. However, when we interact contemporaneous GNP with a 19th century dummy then the estimates show absence of differential effects of GNP. Note also that Ruhm (2000) considers adult individuals. We find that the effects are strongest for children, and in fact our estimate for individuals aged between 35 and 60 is insignificantly different from zero. Of course, mortality due to individual labor market decisions does not apply to children.

Concerning epidemics, the model specification only allows for the cholera epidemic in 1849, the smallpox epidemic in 1870, and the influenza epidemic of 1918. The effects of other epidemics are insignificant and are omitted (see Subsection 4.2 for more details). The included epidemics and World War II give, as expected, rise to increased mortality. The effect of World War II is not significant, but one has to realize that the World War II dummy also measures the effect of missing data during that period.

The calendar time function ψ_2 captures all contemporaneous nation-wide effects that are not included in GNP and the share of agricultural production. It turns out that a fourth order polynomial suffices for an adequate fit (additional terms do not have explanatory power). The estimated polynomial is increasing throughout the observation window. This means that during the 19th and 20th century there were mortality reducing factors that can not

---

model, GNP at birth has a significant long run effect, but this merely captures the omission of the strong instantaneous effect of GNP at birth on infant mortality.

captured by GNP or the share of agricultural production. It is not difficult to speculate about what these factors are, but of course the estimation results do not give a compelling interpretation. Finally, the age dependence function (or baseline hazard) ψ_1 has the expected U-shaped form, meaning that the hazard rates are the highest for infants and for the elderly.

To illustrate the estimates, we plot the mortality rate as a function of age for three cohorts (1812, 1870 and 1912), for males and for females (see Figure 3). In our calculations we used sample averages of the individual characteristics and the actual values of the macro-economic conditions. For the 1912 cohort we stopped the calculations at age 70. A first look at the hazard rates reveals the usual U-shaped pattern and the fact that mortality rates for women are lower than for men. The 1870 cohort experienced a smallpox epidemic at birth, which explains why its mortality rate at birth is so much higher than the corresponding rate of the 1812 cohort. This can be seen even more clearly from the lower panel of Figure 3. This lower panel also shows that, for a given cohort, there are many fluctuation in the mortality rates. These reflect the effects of events like wars, epidemics and changes in macro-economic conditions. For example, the large dip around the ages 30-35 in the most upper line (1812 vs 1912) reflects the effect of World War II on the mortality rate of the 1912 cohort. For the 1870 cohort the same effect is found around ages 70-75.

We now turn to estimation results allowing for unobserved heterogeneity. Table 4 reports estimates of a model in which the unobserved heterogeneity term v has a discrete distribution with two points of support (below we consider more general specifications). For computational reasons we omit the interaction terms between t and $z(\tau)$. This complicates a comparison to Table 3. We therefore also report the estimates of a model without unobserved heterogeneity and without these interaction terms. The results with unobserved heterogeneity are similar to those without. Sign and significance of the main coefficients are unaffected. As usual, allowing for unobserved heterogeneity causes the estimated age dependence to be less negative and covariate effects to be further away from zero.

4.2. Validation of the results with external data and sensitivity analyses

The data source we use constitutes the only source of mortality statistics available, so a cross validation with other data is not possible. However, we may compare the estimated average lifetimes to those published by Statistics Netherlands for the whole country since 1868

(keeping in mind the caveat in Subsection 2.2 on the use of averages as location measures of lifetime distributions). Figure 4 depicts average lifetimes at birth for different five-year cohorts, for men and women. The dark bars are based on our model estimates whereas the blank bars are from Statistics Netherlands. The results are strikingly similar. This confirms our statement in Subsection 2.1 that the province of Utrecht is quite representative of the rest of the Netherlands. Also, Figure 4 can be taken to suggest that the effect of gender on the mean lifetime by cohort is correctly modeled. We made a similar comparison for the average residual lifetime at age 12.5 (nationwide available since 1868). The results are again in close agreement. This suggests that the modeling of mortality during childhood by cohort is correct.

We now report estimation results for models with additional explanatory variables. For sake of brevity we do not present the full sets of results. Those that are not mentioned are virtually identical to those reported in the previous subsection. First, we estimate models with interaction terms between gender and all other explanatory variables listed in Table 3. This amounts to separate estimation for men and women. The estimates of the parameters of interest for men and women are very similar to each other and to those reported in Table 3 (as should be expected from Figure 2, the estimated age dependence differs). This is of course in line with the fact mentioned in the previous paragraph that the model without these interaction terms is well able to explain aggregate life expectancies by cohort for both genders. Not surprisingly, the number of significant parameters by gender is smaller than in Table 3. For women, the effect of the GNP that prevailed between age 1 and 7 is larger than for men (–0.5 versus –0.3, with standard errors of around 0.2). This suggests that women suffer relatively heavily from adverse macro-economic conditions during childhood.

We also considered the inclusion of other indicators of contemporaneous macro-economic conditions in $z(\tau)$, notably the interest rate, the inflation rate, and the share of exports in GNP. These all turn out to be insignificant and quantitatively unimportant.

We also estimate models with additional dummy variables for 1866 and 1871 in $z(\tau)$, because of the nationwide cholera and smallpox epidemics in those years, but the corresponding coefficients are insignificantly positive and very small, so that it can be concluded that these epidemics did not have a sizeable effect on mortality in the province of Utrecht. We subsequently investigated whether the epidemics primarily affected certain age groups. It is well known that cholera and smallpox often disproportionally inflict small

children and elderly persons. The estimated interaction effects of epidemics and age indicators are sometimes sizeable though always insignificant (note the small number of individuals in a certain age category in a certain epidemic). For example, the interaction of the 1870 smallpox epidemic and the "age below 1" indicator has coefficient 0.52 (standard error 0.33), and the death rate for babies is 3 times higher than in adjacent years. However, the other estimates are virtually the same as in Subsection 4.1.

Other potentially relevant events are the abolition of child labor in 1874 and the discovery of penicillin in 1929. However, indicators in $z(\tau)$ of $\tau<1874$ and $\tau<1929$ are insignificant for all age categories. Presumably, the effects of these events on mortality trickled down rather slowly, in which case they are captured by the estimate of the ψ_2 function. An indicator for World War I is also insignificant.

We also estimate models with the season of birth as additional x variables. These turn out to be insignificant. For spring, summer, and fall, the estimates are –0.10, –0.01, and 0.02 (standard errors 0.06). We also estimate models with additional macro-economic indicators at birth as individual explanatory variables, notably the real price of wheat and meat and the aggregate rate of childbirth. These are all strongly insignificant.[15]

In Subsection 3.1 it was argued that individuals with certain x values may suffer disproportionally from bad contemporaneous macro-economic conditions. We estimate models with interactions between contemporaneous GNP on the one hand, and social class and literacy status of the father on the other. The interaction effects are insignificant. This implies that the mortality of individuals from different social classes is equally responsive to bad current macro-economic conditions.

As argued in Section 2, the results may be biased by the exclusion of individuals from the sample when they are still alive in 2000 and their lifetime is right-censored at zero. We investigate this by re-estimating the model with a sample in which the latest cohort is fully omitted. The results are very similar to those above. The only noticeable difference concerns the interaction between average GNP at ages 0-7 and the indicator of whether age exceeds 50. The corresponding estimate is now insignificantly different from zero.

Concerning the modeling of unobserved heterogeneity one may argue that a discrete distribution with two points of support is restrictive. We therefore also estimate models

[15] The result on food prices is in line with Bengtsson and Lindström (2003) who report an insignificant effect of the price of rye at birth, in Sweden in 1766-1894.

allowing for additional points of support. In case of more than three mass points the estimates of mass points coincide. In case of three mass points the main estimation results are very similar to those in Table 4.

To conclude this subsection, the main results are qualitatively and quantitatively robust with respect to a very wide range of assumptions.

4.3. The quantitative importance for mortality of macro-economic conditions at various stages of life

In this subsection we assess the quantitative importance of contemporaneous and past macro-economic conditions on age-specific mortality rates and life expectancies, using the numbers reported in Table 3.

The elasticities of the age-specific mortality rate with respect to contemporaneous GNP and with respect to GNP during childhood are straightforward to calculate. In the basic model they equal the corresponding (age dependent) coefficient times the corresponding GNP level, so they are independent of individual characteristics except age. The fact that they are linear in the GNP levels is an artifact of the specification of the log individual mortality rate as a linear function of GNP. This leads to very large elasticities for the higher ages in the latest cohort, and these should obviously not be taken too seriously, so we focus on averages across cohorts. Moreover, note that an elasticity corresponds to a change of the GNP level in a certain age (or time) interval, but the length of this interval differs across different elasticities. The elasticity relating GNP during childhood to the mortality rate later in life concerns a simultaneous change in all years in the age interval 1-7. Formally, the elasticity for a contemporaneous change concerns a change in the current level only, but in reality, because of the continuity of GNP, and because the GNP effect is identified from non-experimental data, it makes more sense to interpret it as a change in a certain interval around the point of interest. In case of children it does not make sense to distinguish between a contemporaneous effect and a long run effect. Basically, the GNP effect for children is the sum of the corresponding elasticities.

Keeping this in mind, we find that the elasticity of the child mortality rate with respect to GNP is typically very large (around –1.4). Among adults, the elasticity with respect to

GNP in early childhood is also rather large (around –0.7), whereas the elasticity with respect to contemporaneous GNP is small (around –0.2).

Now let us turn to effects on life expectancies at birth, i.e. on the expectations of the lifetime distributions described by the estimated model. As we have seen in Subsection 2.2, life expectancies at birth are not very robust as location measures of lifetime distribution. This means that levels and fluctuations in life expectancies are not very informative. However, the effects of various changes in GNP on life expectancies turn out to be rather robust over time. Note also that the values of the above-mentioned elasticities cannot be translated easily into values of elasticities of life expectancies, because of the age dependence of the mortality rate.[16] This is even true for the elasticity of the residual life expectancy conditional upon survival into adulthood with respect to GNP during childhood.

In the remainder, we focus on "average" individuals, defined by the sample average of the individual characteristics in a specific cohort. We also average over cohorts. We merely report rounded-off numbers.

Suppose that at each point of time during a lifetime, the GNP would be 10% larger than it was in reality. On average, this increases the life expectancy by around 12.5%. In years this is around 5 years, given a baseline average life expectancy of 40 years. Roughly speaking, 6.5% from this 12.5% gain is due to a reduction in childhood mortality. This corresponds to around 2.5 extra years of life. Another 3.5% from this 12.5% (1.5 years extra years of life) is due to the long run effects of GNP during childhood. Finally, 2.5% from this 12.5% (1 year) is due to the effect of contemporaneous GNP for adults. The ranking by importance of these three contributions is in agreement to the ranking of the corresponding elasticities. Increases in GNP lead primarily to a reduction in childhood mortality. Secondly, improved macro-economic conditions during childhood have a long run effect on adult mortality. Thirdly, and quantitatively least importantly, increases in GNP lead to an instantaneous reduction of adult mortality.

The decomposition results depend somewhat on the sequence of decomposition, because the components are not additive. Moreover, if one considers heterogeneous individuals then the aggregate long run effect of GNP during childhood is smaller than above,

[16] Most probability mass of the lifetime distributions is concentrated in two specific areas: around a peak at zero, representing child mortality, and at a hump at high ages. The precise location of the latter hump varies over calendar time.

because it is mitigated by the selection effect of GNP during childhood on the composition of survivors.

Note that the prediction by the model of the mean elasticity of the life expectancy with respect to GNP during life (1.25) is larger than expected from the observed changes of GNP and life expectancy over the observation window. This reflects a deficiency of the model specification, in particular the linear specification of the log individual mortality rate as a function of GNP. From this point of view, a more flexible specification is warranted (recall however the trade-off with computational requirements).

5. Conclusions

One of our key findings is that the effect of macro-economic conditions during childhood on mortality later in life is stronger at ages above 50. Such results can only be obtained from data with observation windows of unprecedented size, like ours, covering lifetime histories of individuals born between 1812 and 1912. To study the effects of contemporaneous and past macro-economic conditions on mortality, we specified a flexible class of duration models for individual mortality. These allow for contemporaneous and lagged individual and macro-economic explanatory variables as well as for a range of interaction terms between mortality determinants, a calendar time polynomial capturing developments that are not fully synchronous with GNP, and indicators of events like epidemics.

Controlling for individual socio-economic background characteristics, we find strong effects of macro-economic conditions on mortality. The most striking result is that (macro-) economic conditions during early childhood (age 1-7) have a large significantly negative effect on the mortality rate throughout life. The effects of GNP at birth and GNP around the time of labor market entry are insignificant. Bad macro-economic conditions during childhood have a particularly strong effect (relatively and absolutely) on mortality at higher ages (above 50). Also, women suffer more than men from adverse economic conditions during childhood.

The estimated instantaneous effect of GNP is also strong. Mortality rates are lower in periods of high production. The size of the instantaneous effect varies with age: again children are most affected. Elderly individuals are affected to a lesser degree. So conditions

between age 1 and 7 have the strongest instantaneous effect as well as the strongest long run effect on mortality. Also, like in the case of the long run effects, the instantaneous macro-economic effect is smaller for babies, presumably because the health of babies primarily depends on the health of the mother. Bad current macro-economic conditions affect the mortality across different social classes in equal amounts. The main results are qualitatively and quantitatively robust with respect to a wide range of assumptions.

We used the estimation results to decompose the effect of a simultaneous increase of GNP at each point of time during a lifetime on the life expectancy. On average, a 10% increase adds 5 years of life. Around half of this is due to a reduction in childhood mortality. Almost a third is due to the long run effects of improved macro-economic conditions during childhood. Finally, a fifth is due to the effects of contemporaneous GNP for adults.

It could be argued that long run effects of childhood conditions in Western societies were more important in the 19th century than they are now, because of the shift in the mortality spectrum from infectiuous diseases to chronic diseases. At present our data do not enable us to address this, because all individuals are born before 1913, and most of those born later are still alive. However, the very recent epidemiological literature using natural experiments demonstrates that cohort effects on health and mortality later in life are also significant in the 20th century (see e.g. Almond, 2002, and the survey in Doblhammer, 2003). Moreover, our results are confirmed by recent studies of the effect of individual socio-economic conditions on health outcomes like illness indicators later in life. These studies invariably point towards childhood conditions as crucial determinants of health later in life.

Our results therefore indicate that from a policy point of view it is particularly useful to focus on children aged between 1 and 7 in bad economic conditions. The contemporaneous mortality of these children, as well as mortality later in their life, can be greatly reduced if their conditions are improved upon. Note that nowadays, in Western societies, childhood mortality is much lower than in our data. The current increase of life expectancies in Western societies is propelled by improvements in survival at high ages. However, according to our results and those in studies of health outcomes, these improvements may be driven to a large extent by improved conditions during childhood.

Note that our study does not reveal the mechanisms behind the effects of macro-economic conditions on mortality rates. For example, we do not touch upon the question how GNP affects health investments and improvements in the health care system, or how these act

upon mortality. The reason for this is that, for most of the years covered by our data, indicators of this are not observed. Alternatively, with a much larger sample size, one could focus on specific medical innovations and policy measures in more detail, like the abolition of child labor.

This issue is related to the role of unobserved heterogeneity at the individual level in the analysis. We have seen that the estimation results are robust with respect to the inclusion of parametrically distributed unobserved heterogeneity that is independent of other explanatory variables. However, in reality, a recession usually does not act with equal force on all individuals. The decomposition results for the effect of a simultaneous GNP increase should therefore be interpreted as averages across individuals with different responses. Those who are strongly affected by the business cycle during childhood gain more years than reported, whereas those who are not gain nothing. To proceed, one may want to estimate models where the macro-economic indicators interact with unobserved heterogeneity in the individual mortality rate. It is doubtful whether such a model is non-parametrically identified with single spell lifetime duration data. One way to proceed is by recording family trees in the data and including family-specific effects in the analysis, assuming that family members are affected by recessions in the same way. To some extent, family-specific effects may also capture genetic differences. However, to obtain sufficient numbers of family members in the data we require substantially larger sample sizes. These are not available yet, so this is a topic for further research.

References

Abbring, J.H., G.J. van den Berg and J.C. van Ours (2002), The anatomy of unemployment dynamics, *European Economic Review* 46, 1785-1824.

Abramowitz, M. and I.A. Stegun (1970), *Handbook of Mathematical Functions; With Formulas, Graphs and Mathematical Tables* (Dover Publications, New York).

Almond, D.V. (2002), Cohort differences in health: a duration analysis using the National Longitudinal Mortality Study, Working paper, University of Chicago.

Attanasio, O.P. and C. Emmerson (2001), Differential mortality in the UK, NBER Working paper 8241, Cambridge.

Bengtsson, T. and M. Lindström (2003), Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766-1894, *International Journal of Epidemiology* 32, 286-294.

Brenner M. (1979), Mortality and the national economy: a review and the experience of England and Wales, 1936-1976, *The Lancet* 8142, 568-573.

Case, A., A. Fertig and C. Paxson (2003), From cradle to grave? The lasting impact of childhood health and circumstance, Working paper, Princeton University.

Cox, D. and D. Oakes (1984), *Analysis of Survival Data* (Chapman and Hall, London).

Cutler, D.M. and E. Meara (2001), Changes in the age distribution of mortality over the 20th century, NBER Working paper 8556, Cambridge.

Deaton, A. and C. Paxson (2001), Mortality, income, and income inequality over time in Britain and the United States, NBER Working paper 8534, Cambridge.

De Swaan, A. (1988), *In Care of the State; State Formation and Collectivization of Health Care, Education and Welfare in Europe and America in the Modern Era* (Oxford University Press, New York).

Doblhammer, G. (2003), The late life legacy of very early life, Working paper, Max Planck Institute for Demographic Research.

Ferrie, J.P. (2003), The poor and the dead: socioeconomic status and mortality in the U.S., 1850-1860, in: D.L. Costa (ed.), *Health and Labor Force Participation over the Life Cycle: Evidence from the Past* (University of Chicago Press, Chicago).

Haines, M.R. (2001), The urban mortality transition in the United States, 1800-1940, *Annales de Demographie Historique*, 33-64.

Himes, C.L. (1994), Age patterns of mortality and cause-of-death structures in Sweden, Japan, and the United States, *Demography* 31, 633-650.

Jacobs, J. and J.P. Smits (2001), Business cycles in the Netherlands, 1815-1913, Working paper, Groningen University.

Kok, J., F. van Poppel and E. Kruse (1997), Mortality among illegitimate children in mid-nineteenth-century The Hague, in: C.A. Corisini and P.P. Viazzo (eds.), *The Decline of Infant and Child Mortality; The European Experience: 1750-1990* (Martinus Nijhoff, Dordrecht).

Koupilová, I. (1997), *Fetal Growth, Social Factors and Circulatory Diseases* (Uppsala University).

Lancaster, T. (1990), *The Econometric Analysis of Transition Data* (Cambridge University Press, Cambridge).

Lichtenberg, F.R. (2002), Sources of U.S. longevity increase, NBER Working paper 8755, Cambridge.

Lynch, K.A. and J.B. Greenhouse (1994), Risk factors for infant mortality in nineteenth-century Sweden, *Population Studies* 48, 117-133.

Macunovich, D.J. (1999), The fortunes of one's births: relative cohort size and the youth labor market in the U.S., *Journal of Population Economics* 12, 215-272.

Mandemakers, K. (2000), Historical sample of the Netherlands, in: P. Hall et al. (eds.), *Handbook of International Historical Microdata for Population Research* (Minnesota Population Center, Minneapolis).

Murphy, K.M. and R. Topel (2003), The value of health and longevity, Working paper, University of Chicago.

Pritchett, L. and L.H. Summers (1996), Wealthier is healthier, *Journal of Human Resources* 31, 841-868.

Ruhm C.J. (2000), Are recessions good for your health?, *Quarterly Journal of Economics* 115, 617-650.

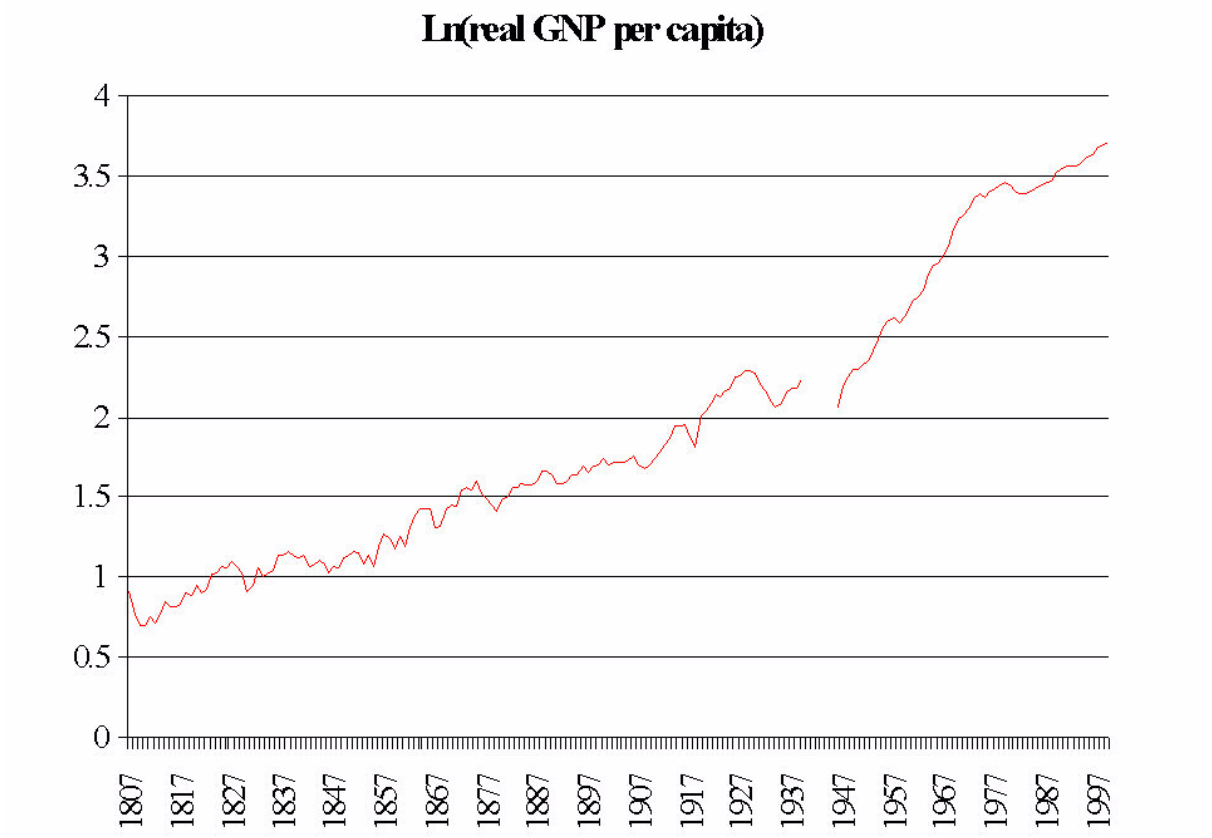
Smits, J.P., E. Horlings and J. Luiten van Zanden (2000), Dutch GNP and its components, 1800-1913, Working paper, Groningen University.

Van den Berg, G.J. (2001), Duration models: specification, identification, and multiple durations, in: J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics, Volume V* (North-Holland, Amsterdam).

Vaupel, J.W. and A.I. Yashin (1985), The deviant dynamics of death in heterogeneous populations, in: N.B. Tuma (ed.), *Sociological Methodology 1985* (Jossey-Bass, San Francisco).

Wolleswinkel-Van den Bosch, J., F. van Poppel, E. Tabeau and J. Mackenbach (1998), Mortality decline in the Netherlands during the period 1850-1992: a turning point analysis, *Social Science and Medicine* 47, 429-444.

Figure 1. Logarithm of annual real per capita GNP in The Netherlands between 1807 and 1997.



Note: real GNP is measured in 1,000 Euros with 1995 as base year.

Table 1. Descriptive statistics of some explanatory variables by birth cohort interval.

birth cohort interval midpoint	1817	1827	1837	1847	1857	1867	1877	1887	1897	1907
Female (%)	46	52	54	50	45	47	50	49	52	47
Social class (range 1-6):										
average	2.7	2.4	2.4	2.5	2.4	2.4	2.6	2.5	2.5	2.5
standard deviation	1.3	1.2	1.2	1.2	1.3	1.3	1.3	1.2	1.2	1.1
Father not illiterate (%)	70	72	74	76	80	83	90	92	93	98
Mother unmarried at birth (%)	5.3	4.4	2.4	2.8	2.8	4.1	0.9	2.3	1.2	0.7
Born in urban area (%)	36	40	38	34	39	34	39	40	42	52
Share agriculture in GNP (%)	26	23	22	25	28	28	25	22	19	19
Real per capita GNP at birth	2.3	2.8	3.0	3.0	3.2	4.0	4.5	5.0	5.3	6.2

Note: real GNP is measured in 1,000 Euros with 1995 as base year.

Table 2. Descriptive statistics of lifetimes by birth cohort interval and gender.

Birth cohort interval midpoint		1817	1827	1837	1847	1857	1867	1877	1887	1897	1907
Male	average	39.0	38.6	39.6	32.1	32.0	36.2	41.1	48.7	50.2	49.3
	standard deviation	33	31	32	33	34	35	36	36	34	32
	median	38.7	39.0	44.3	20.0	17.0	24.6	46.6	47.8	66.0	65.8
Female	average	40.8	37.9	37.3	34.5	33.4	34.8	41.6	48.7	54.2	47.2
	standard deviation	32	33	33	31	36	35	35	36	35	34
	median	34.4	39.1	34.7	31.0	10.8	19.7	54.7	63.8	71.5	65.7
	% right-censored	8	8	6	13	10	12	10	8	15	6

Note: reported moments and medians are based on uncensored lifetimes only.

Figure 2. Estimated conditional probability of death by birth cohort interval, age category, and gender.

Figure 2a: Hazard rates for ages [0,1) (in %)

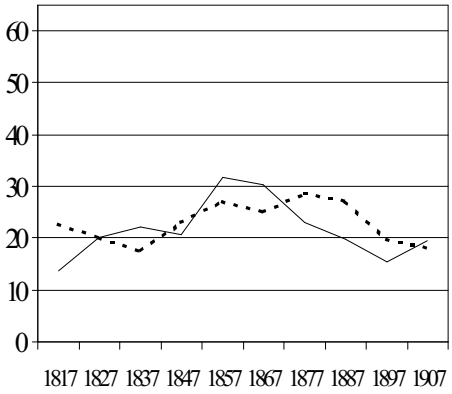


Figure 2b: Hazard rates for ages [1,5) (in %)

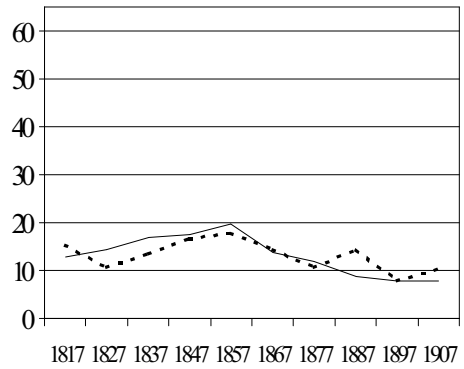


Figure 2c: Hazard rates for the category [5,20) (in %)

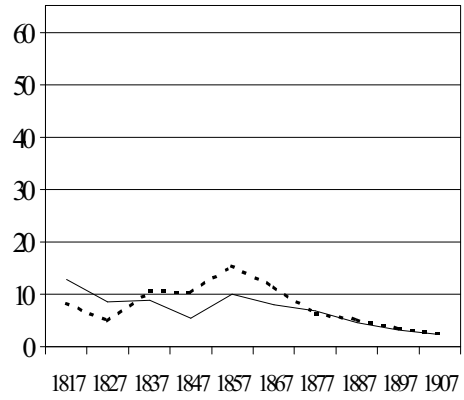


Figure 2d: Hazard rates for ages [20,50) (in %)

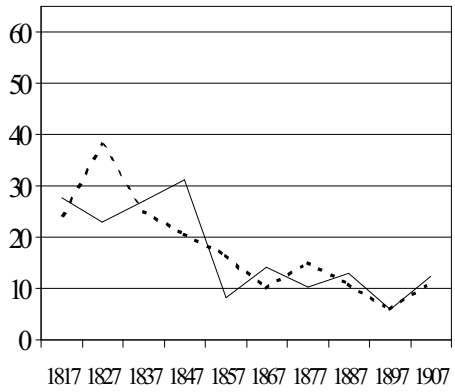
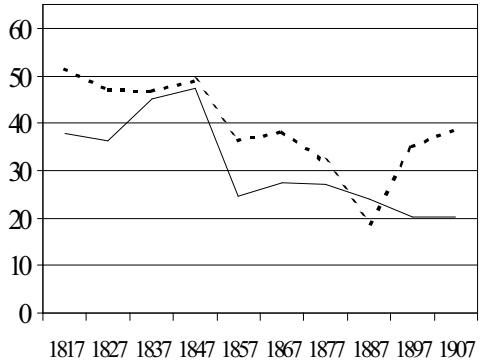
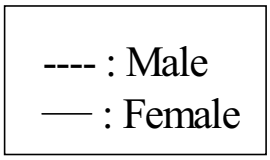


Figure 2e: Hazard rates for ages [50,70) (in %)





Note: "hazard rates" are conditional probabilities of death as defined in Subsection 2.2.

Table 3. Parameter estimates of the baseline model for individual lifetime duration.

variable	estimate	t-statistic
Individual characteristics		
Female	-0.112	-2.7
Social class	-0.036	-2.0
Father is not illiterate	-0.096	-1.5
Mother was unmarried at the time of birth	2.100	3.5
Born in urban area	0.063	1.5
Long run effects of macro-economic conditions		
Share of agriculture in GNP at birth	0.016	1.9
Real per capita GNP at birth	-0.099	-0.8
Average real per capita GNP at age 1-7	-0.345	-2.6
Average real per capita GNP at age 8-14	-0.083	-1.2
Average real per capita GNP at age 15-20	0.014	0.4
Instantaneous effects of macro-economic conditions		
Contemporaneous GNP at age 0-1	-0.342	-2.5
Contemporaneous GNP at age 2-7	-0.438	-2.8
Contemporaneous GNP at age 8-14	-0.270	-2.0
Contemporaneous GNP at age 15-34	-0.184	-2.8
Contemporaneous GNP at age 35-50	-0.027	-0.8
Contemporaneous GNP at age 51-60	-0.025	-1.1
Contemporaneous GNP at age 61-70	-0.034	-2.5
Contemporaneous GNP at age 71-80	-0.050	-3.8
Contemporaneous GNP at age 81-90	-0.078	-5.6
Contemporaneous GNP at age 90+	-0.105	-5.5
Instantaneous effects of epidemics and wars		
1848 Cholera epidemic	0.623	2.9
1870 Smallpox epidemic	0.719	4.0
1918 Influenza epidemic	0.288	1.0
World War II	0.070	0.5
Miscellaneous contemporaneous effects		
first-order Chebyshev polynomial in time	3.083	7.0
second-order Chebyshev polynomial in time	1.403	6.2
third-order Chebyshev polynomial in time	0.824	8.2
fourth-order Chebyshev polynomial in time	0.491	5.6
Log age dependence		
Age 0-1	1.297	1.9
Age 2-7	-0.660	-0.9
Age 8-14	-1.125	-1.4
Age 15-34	-1.193	-1.8
Age 35-50	-1.973	-3.2
Age 51-60	-1.480	-2.6
Age 61-70	-0.841	-1.6
Age 71-80	0.089	0.2
Age 81-90	1.177	2.4
Age 90+	2.223	3.9
Log likelihood function	9078.663	
# of individuals	2628	

Note: effects on mortality rate reported.

Figure 3. Mortality rates and differences between mortality rates, for 3 cohorts, by gender.

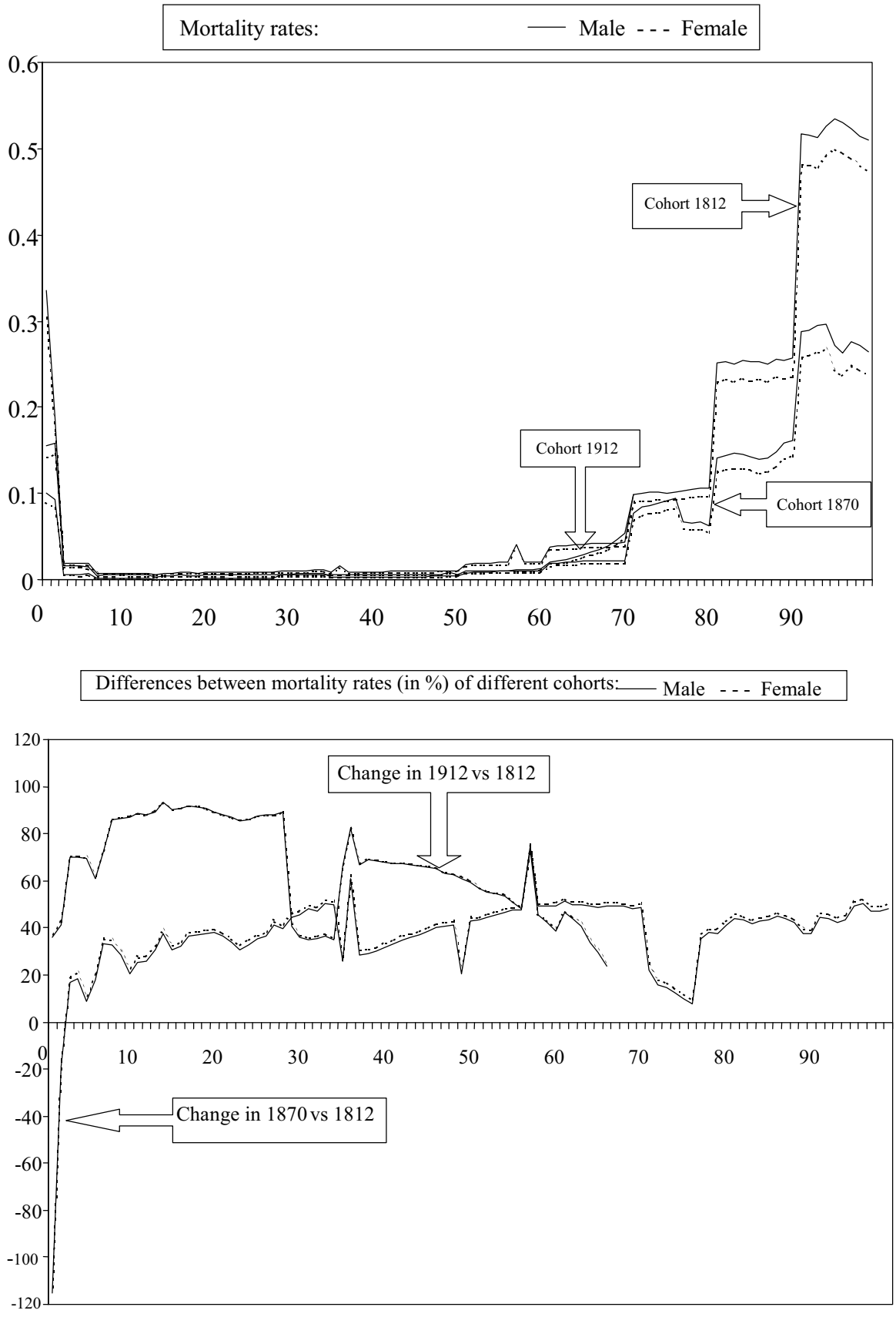


Table 4. Parameter estimates of models with and without unobserved heterogeneity.

variable	estimate	t-statistic	estimate	t-statistic
Individual characteristics				
Female	-0.112	-2.7	-0.157	-3.2
Social class	-0.038	-2.1	-0.028	-1.3
Father is not illiterate	-0.091	-1.4	-0.145	-2.0
Mother was unmarried at the time of birth	2.137	3.6	3.303	5.2
Born in urban area	0.068	1.6	0.066	1.3
Long run effects of macro-economic conditions				
Share of agriculture in GNP at birth	0.017	12.0	0.014	1.4
Real per capita GNP at birth	-0.220	-2.5	-0.325	-3.2
Average real per capita GNP at age 1-7	-0.287	-3.5	-0.227	-2.5
Average real per capita GNP at age 8-14	-0.010	-0.2	-0.007	-0.1
Average real per capita GNP at age 15-20	0.062	1.9	0.060	1.8
Instantaneous effects of macro-economic conditions				
Contemporaneous GNP	-0.051	-4.3	-0.051	-4.3
Instantaneous effects of epidemics and wars				
1848 Cholera epidemic	0.627	3.0	0.917	3.4
1870 Smallpox epidemic	0.694	3.9	0.768	3.2
1918 Influenza epidemic	0.327	1.2	0.344	1.2
World War II	0.005	0.0	0.009	0.1
Miscellaneous contemporaneous effects				
first-order Chebyshev polynomial in time	2.444	5.9	2.647	5.7
second-order Chebyshev polynomial in time	1.108	5.4	1.059	4.3
third-order Chebyshev polynomial in time	0.743	8.1	0.758	6.7
fourth-order Chebyshev polynomial in time	0.281	3.7	0.257	2.8
Log age dependence				
Age 0-1	0.362	2.4	0.028	2.0
Age 2-7	0.048	2.4	0.034	2.1
Age 8-14	0.001	2.4	0.008	2.2
Age 15-34	0.008	2.6	0.009	2.3
Age 35-50	0.010	2.8	0.012	2.5
Age 51-60	0.019	3.0	0.023	2.7
Age 61-70	0.039	3.1	0.048	2.8
Age 71-80	0.099	3.2	0.120	2.9
Age 81-90	0.219	3.2	0.266	2.9
Age 90+	0.426	2.7	0.519	2.5
Unobserved heterogeneity distribution				
$\log(\Pr(v=1)/(1-\Pr(v=1)))$ $(\Pr(v=1)\equiv 1-\Pr(v=c))$			1.432	15.5
			(0.807)	
$\exp(c)$			3.905	34.9
Log likelihood function	9097.470		8908.799	
# of individuals	2628		2628	

Figure 4. Average lifetime at birth by gender and five-year cohort interval.