The influence of socio-demographic factors and region on seasonal mortality in the United States (Work in Progress)

Roland Rau
Max Planck Institute for Demographic Research
Konrad-Zuse-Str. 1
D-18057 Rostock
Germany
Email:`Rau@demogr.mpg.de`
Tel.+49-381-2081109

and
Jutta Gampe
Max Planck Institute for Demographic Research
Konrad-Zuse-Str. 1
D-18057 Rostock
Germany

September 15, 2004

1 Introduction

Seasonal mortality fluctuations are a persistent phenomenon in almost every population. The basic pattern is shaped by the climate as mortality typically peaks in winter and has a trough late in summer. In our analysis, we focussed on explaining the "seasonality paradox" as shown in Figure 1. Countries with relatively warm or moderate climate such as Italy, Portugal, Ireland or the UK experience considerably higher excess winter mortality than cold regions like Sweden, Norway or Finland. This suggests that not climate but social and cultural factors shape the actual amplitude in seasonality. In particular we asked the questions:

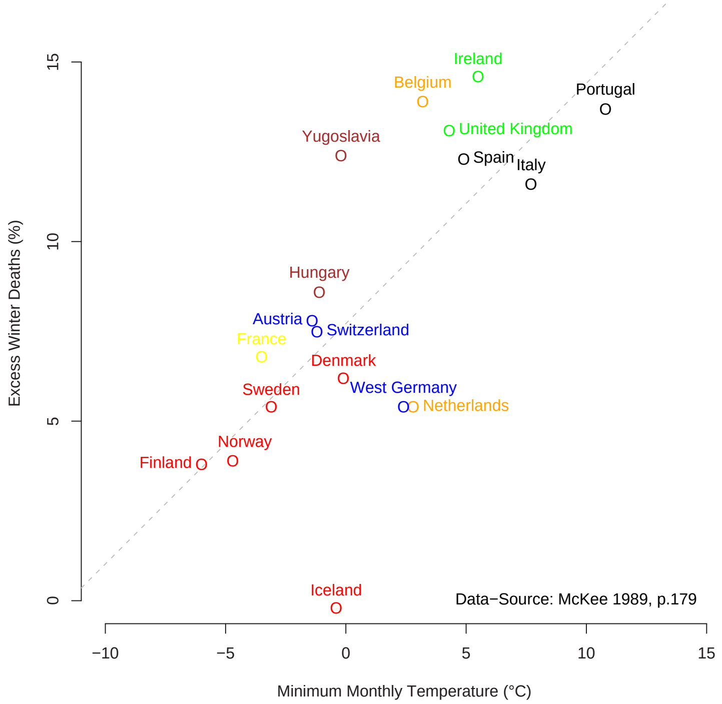
- **Period** Is seasonality over period decreasing with the spread of central heating whose lack is associated in the literature with high levels of winter excess mortality (e.g. Aylin et al., 2001; Donaldson and Keatinge, 1997; Eurowinter Group, 1997; Keatinge et al., 1989; Sakamoto-Momiyama, 1977)?

Age Can we support previous findings (e.g. McDowall, 1981; Näyhä, 1980; Quetelet, 1838; Rau and Doblhammer, 2003) that suggest an increase of mortality with age?

Region How important is the region where you are living for the trajectory of seasonal mortality? From cross-country analyses (Grut, 1987; Healy, 2003; McKee, 1989) we would expect higher seasonality in warm regions of the US than in colder regions.

Education How important are socioeconomic factors which we measured by highest level of education on seasonal mortality? Previous research is ambivalent. Few studies argue that lower social groups are disadvantaged (e.g. Donaldson and Keatinge, 2003), most others found no social gradient (e.g. Lawlor et al., 2002, 2000; Shah and Peacock, 1999).

Figure 1: The Seasonality Paradox: Excess Winter Mortality in Several European Countries



2 Data and Methods

2.1 Data

Our analysis uses the "Multiple Cause of Death" Public Use Files for the years 1959–98 published by the "US Centers for Disease Control and Prevention" (CDC). We included only deaths at ages 50 and higher. The data consists of more than 77. Mio death records. Each record contains information on the sex, month and year of death, and age at death. For our present analysis we also extracted state of residence, state of occurence and for the last 10 years (while it was available) also educational level.

2.2 Methods

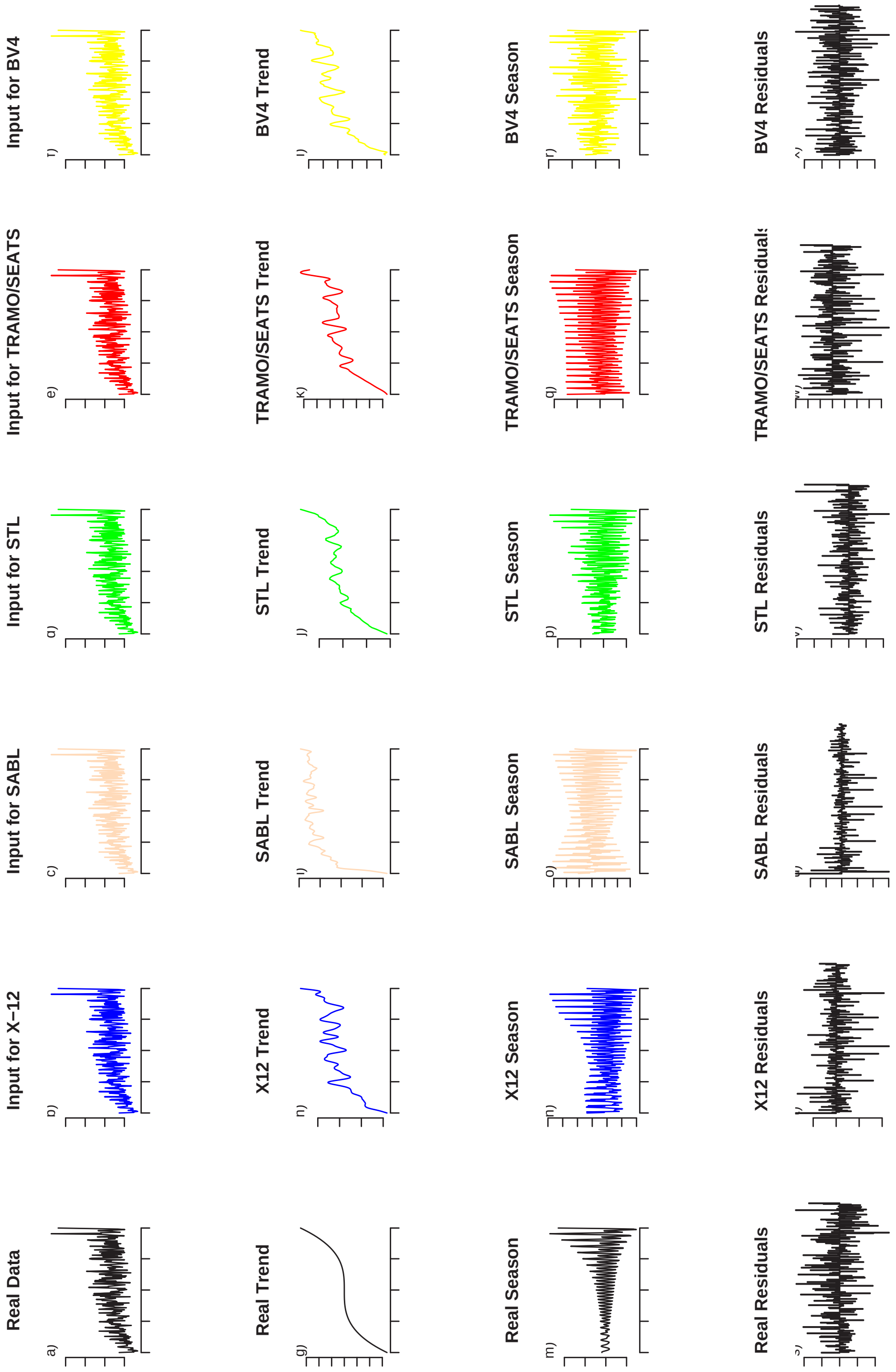
2.2.1 Problems of Standard Methods

For our analysis we first tried via simulation studies whether various standard time-series methods like X-12, STL were able to handle count data with changes in the trend and the seasonal component in the presence of overdispersion. The results were unsatisfactory as Figure 2 on page 4 shows. The left column in black displays the data we input into the various methods. The data which were drawn from a Negative Binomial Distribution to allow overdispersion in the count data contain a third-order polynomial as a trend and an increasing seasonal component. As one can easily visually detect none of the methods we used was able to reflect the input correctly.[1]

---

[1] Please see the following literature for an explanation of the used methods: X-12: Bruce and Jurke (1992); Yaffee (2000); SABL: Cleveland et al. (1981a,b); STL: Cleveland et al. (1990); TRAMO/SEATS: Maravall (2002); BV4: Nourney (1973, 1975, 1983); Speth (2004).

Figure 2: Seasonal Decomposition of Time Series — Unsatisfactory Results from Standard Methods



2.2.2 Model and P-Spline Smoothing

To overcome those methodological problems, we therefore developed the following new model: We denote the monthly number of deaths (for a specific sex and age-category or period) by $Y_t, t = 1, \ldots, T$. For analyses by period, $T = 480$ ($\hat{=}$ Jan '59, …, Dec '98); for analysis by age, $T = 600$ ($\hat{=} 50\frac{0}{12}$ years of age, …, $99\frac{11}{12}$ years of age). We start by assuming that the Y_t are independently Poisson distributed with a log-link and the mean μ_t specified as

$$\ln \mu_t = \alpha_0 + f_0(t) + \sum_{l=1}^{L} \left\{ f_l^{sin}(t) \sin\left(\frac{2\pi l}{12} t\right) + f_l^{cos}(t) \cos\left(\frac{2\pi l}{12} t\right) \right\} \tag{1}$$

Both the additive trend term ($f_0(t)$) and the amplitude modulating functions $f_l^{sin}(t)$ and $f_l^{cos}(t)$ are assumed to be smoothly varying functions over time t. The most simple seasonal model would only fit one sine-cosine term ($L = 1$), by adding more components more complex cyclic patterns could be captured. Model (1) is a varying coefficient model (Hastie and Tibshirani, 1993) which, as demostrated by Eilers and Marx (2002), can be conveniently fit by using P-Splines. Each smooth model component is expanded using a moderately large B-spline basis and smoothness is controlled by penalizing the spline-coefficients by a difference penalty (Eilers and Marx, 1996). The optimal amount of smoothing can be determined by minimizing an information criterion, like AIC, over a grid of values for the smoothing parameter λ. For large models with several functions to be smoothed Eilers and Marx (2002) suggest a multi-dimensional grid-search to determine the optimal compination of smoothing parameters.

2.2.3 The Impact of Overdispersion

Clearly, there is unobserved heterogeneity in these data. The smooth index is only a proxy for the actual prevailing weather conditions, and individuals, even for narrow age categories, have different susceptibility to death. Both features are well known sources of overdispersion (Barron, 1992; Cameron and Trivedi, 1998). The effect of overdispersion on smoothing methods can be considerable as depicted in the right column in red of Figure 3 (page 7). In this case, we used similar data as in Figure 2 to test our method. Extra variation that is not allowed for by the Poisson model is distributed over the smooth model components leading to serious undersmoothing of the target functions. This phenomenon corresponds to the similar effect that arises when correlated data are smoothed under independence assumptions.

2.2.4 Smoothing Parameter Selection

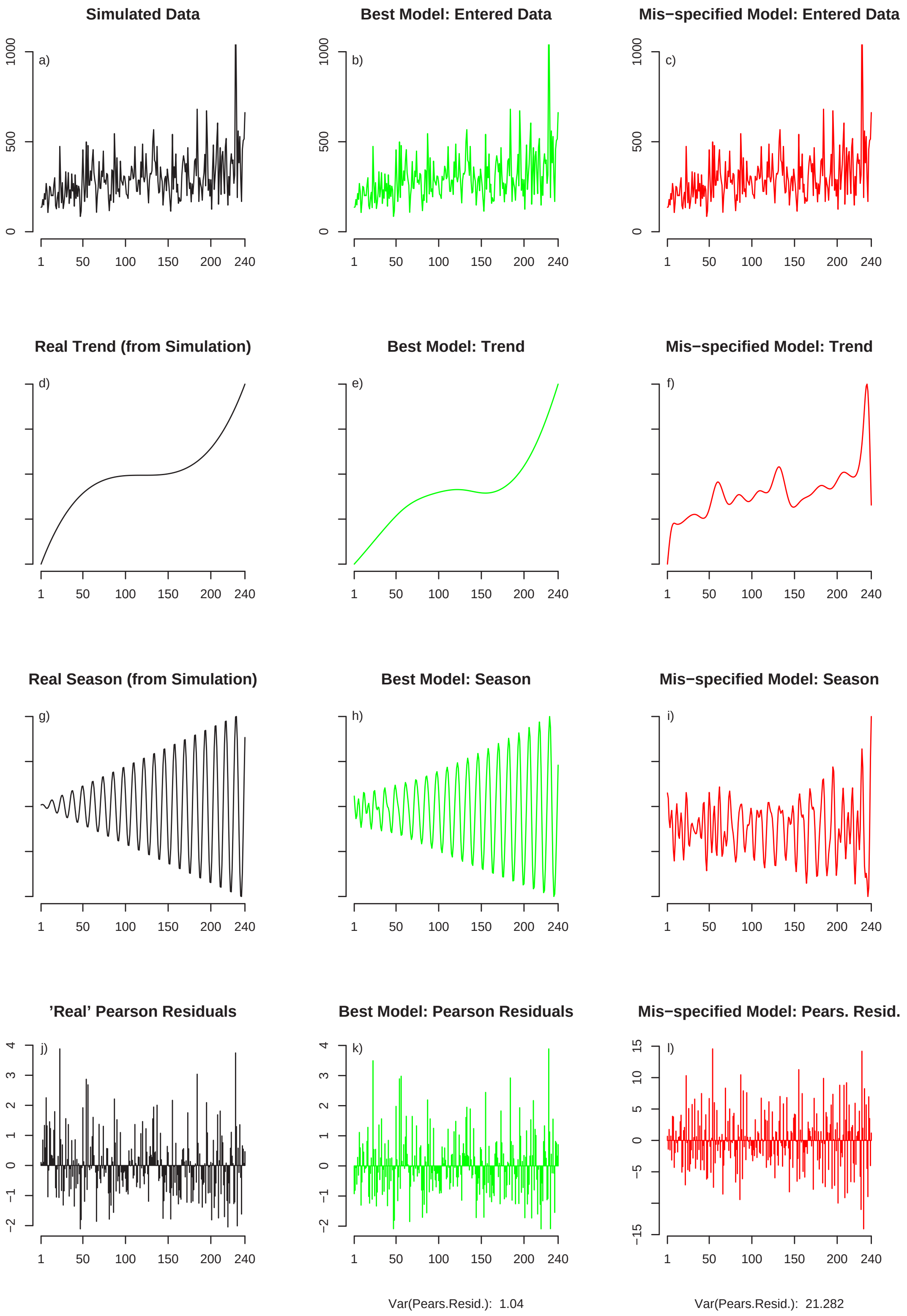
A simple and common extension for overdispersed count data is the Negative Binomial (NB) distribution (Lawless, 1987), arising from a Gamma-Poisson mixture. For a fixed value of the variance τ^2 of the mixing Γ-distribution (with mean 1), the NB is an exponential family and we thus still operate in the GLM framework. Therefore, for a given amount of overdispersion τ^2, we may determine the values of the smoothing parameters as in the Poisson case. An optimal procedure though has to determine which portion of the variation in the data can be attributed to overdispersion and which is due to the structural components in the model. To resolve this question we propose the following two-stage strategy.

- Fix a grid of values for the overdispersion parameter, i.e. the variance of the Γ-distribution: $\tau_1^2, \ldots, \tau_M^2$.
- For each of these (fixed) values $\tau_m^2 (m = 1, \ldots, M)$ minimize the AIC to obtain the optimal smoothing parameters $(\lambda_1^m, \ldots, \lambda_C^m)$, where C is the number of components to be smoothed in (1).
- For these smoothing parameters calculate the Pearson residuals according to the NB model currently under consideration (i.e. the fixed value τ_m^2).

$$p_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\hat{\omega}_t}} \qquad \hat{\omega}_t = \hat{\mu}_t + \tau_m^2 \hat{\mu}_t^2$$

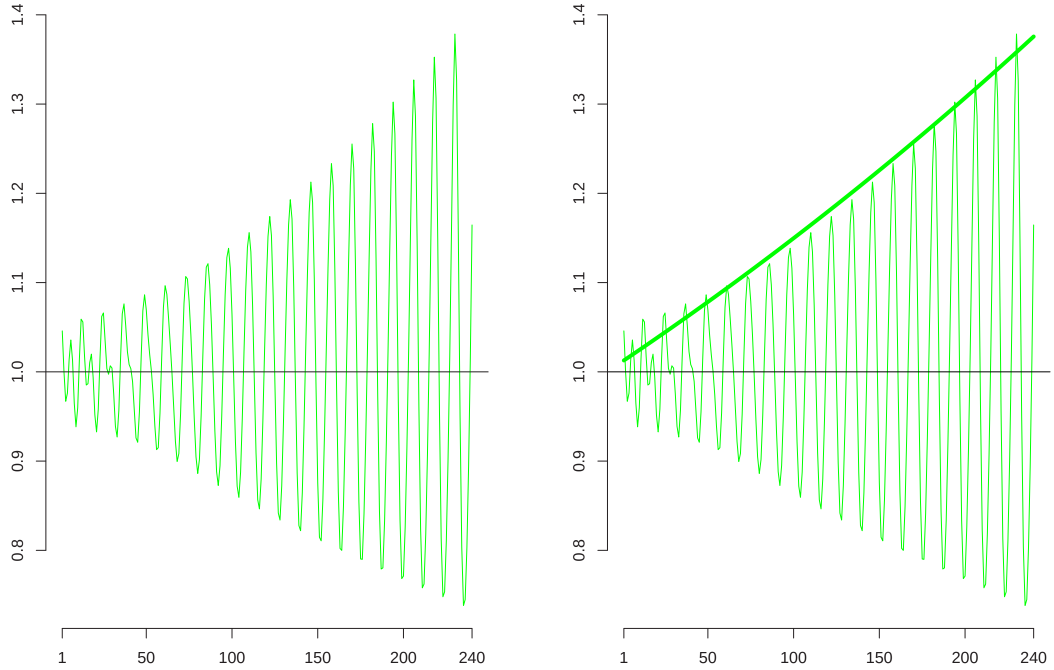
Using this approach yields results which are displayed in the middle column in Figure 3 on page 7 in green: The trend as well as the seasonal component which is returned corresponds almost perfectly to our data we have entered which are shown in black in the left column.

Figure 3: Simulated Data, “Optimal” Model and a Mis-specified Model



As an indicator for seasonality we used the amplitude of the seasonal component as indicated by the bold green line in the right panel of Figure 4 on page 8.

Figure 4: Seasonal Component from Figure 3 and its Amplitude

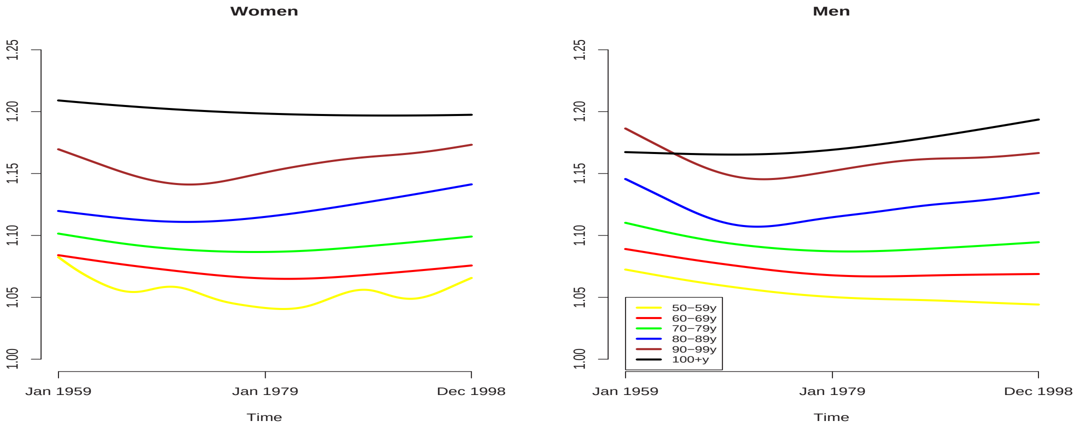


3 Results

3.1 Results by Period and Sex

Figure 5 (page 9) shows the change in the amplitude for seasonality in deaths by 10-year-age-groups for the whole observation period from Janury 1959 until December 1998. The left panel illustrates results for women, whereas the right panel deals with men. For both sexes we see the same general trends: the older the people (=the darker the lines), the higher is the seasonal amplitude. Changes over age will be examined in subsequent parts of this paper. Right now the focus is on changes over time. What we discover is some support for Feinstein's (2002) finding: younger age-groups seem to have a constant or slightly decreasing trend as indicated by the yellow and red lines — especially for men. People who died at an age above 80 (blue, brown, and black lines), however, have to suffer from higher fluctuations in seasonality towards the end of the observation period.

Figure 5: Seasonality of Deaths over Time by Sex and Age-Group

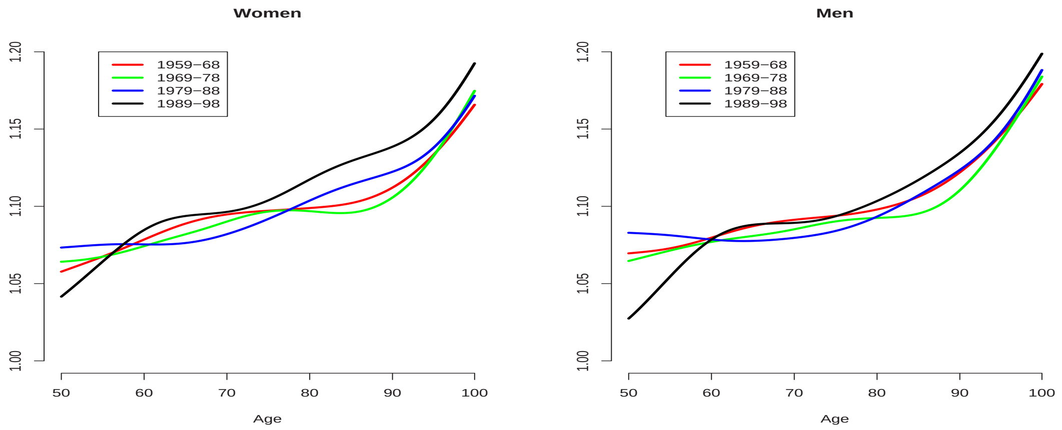


3.2 Results by Age and Sex

Of prime interest for demographers are not only death patterns for women and men over time but — maybe even more important — with age. Figure 6 (page 10) gives a first impression how seasonality in deaths changes with age. The left panel shows seasonality of deaths for women with four colored solid lines indicating 10-year-calendar periods. The right panel shows results from the same analysis for men.

The general trend for both sexes shows — as expected — higher seasonality with age. The increase is far from linear. We could make a distinction for women as well as for men by grouping the first three decades together (red: 1959–68; green: 1969–78; blue: 1979-88)

Figure 6: Seasonality of Mortality by Sex and 10-year-calendar period

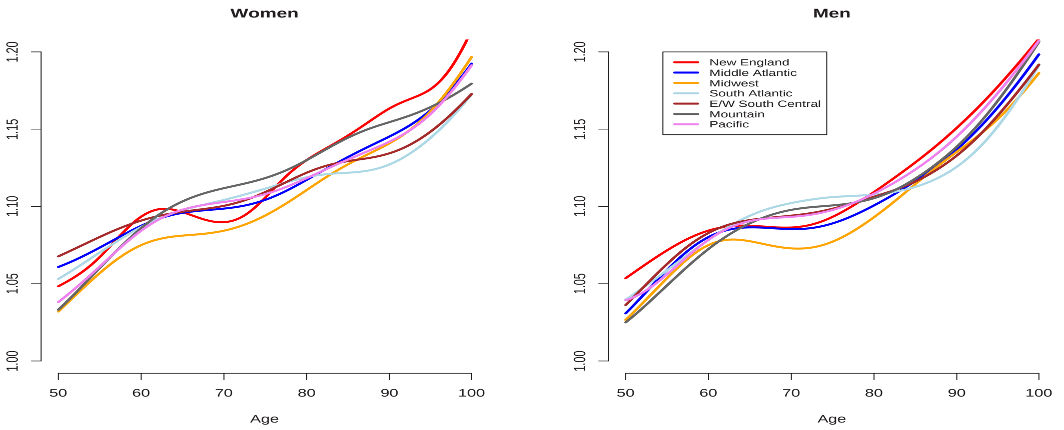


and contrast them with the last 10 years (1989–98 in black): until age 80 the increase is relatively moderate. Then, at the highest ages, seasonality bends sharply upwards. The black solid line in both panels represents changes with age for the most recent decade in the analysis (1989–98). One can differentiate three stages: Compared to previous decades, seasonality is relatively low at age 50 and increases until age 60 where it is roughly on level terms. Between 60 and 75/80 years seasonality remains relatively constant. After age 80 seasonality in deaths from all causes is increasing and shows higher values than in the past for the same ages.

3.3 Results by Region, Age, and Sex

Figure 7 shows the development of seasonal mortality by age and region in the US for women and men for the last observed decade, 1989–1998. The results are rather unexpected: previous studies usually indicated that regions with a warm or moderate climate (e.g. the UK, Ireland, Portugal, Spain, Greece) tend to have higher seasonal fluctuations in mortality and deaths than colder regions such as Russia, Canada or Scandinavian countries (Eurowinter Group, 2000; Eurowinter Group, 1997; Grut, 1987; Healy, 2003; McKee, 1989). This has usually been attributed to the fact that people in colder regions have higher indoor temperatures and avoid exposure to outdoor cold. If those findings could have been converted to the United States, one would assume that the regions "South Atlantic" and "East/West South Central", colored in light blue and brown, respectively, should show higher seasonality than other regions. According to the "Köppen Climate Classification", all states covered in these two regions belong to the "Humid Subtropical Climate". Surprisingly, they do not deviate in any way from the other regions in the United States which are less humid and cooler. This underlines that social and cultural factors are important forces in shaping the seasonal pattern of deaths as climate appears to be negligible. It has to be mentioned, though, that “region” in the United States is not only correlated with climate but also with socio-economic status and life expectancy. Residents in New England spent on average more time in school than women and men in the regions “South Atlantic” or “South Central”.[2] At the same time, life expectancy is also lower in those regions (Pickle et al., 1996). This could suggest also an alternative explanation: there are two opposing forces which cancel each other out. On the one hand, the regional differences do exist as in Europe between warm and cold regions. That would imply that the southern states show higher seasonality than the states in the north-east. On the other hand, this differential is counteracted by a selection effect. Mortality is higher in the south of the United States. Due to these higher death rates, frail people tend to die at younger ages than in the North which should have a rather depressing effect on seasonality. We consider the first explanation (no regional differences) to be more likely than the balanced outcome of two opposing forces. If the latter was true, it would require a social gradient by education: due to a selection effect, people with low education should also show lower amplitudes in their mortality fluctuations. As will be shown later, a social gradient is observable — with the opposite direction, though: people with an academic degree have generally lower seasonality than people with only a few years spent in formal education.

Figure 7: Seasonality of Deaths by Age, Sex and Region, 1989–1998



3.4 Seasonality by Region & Period

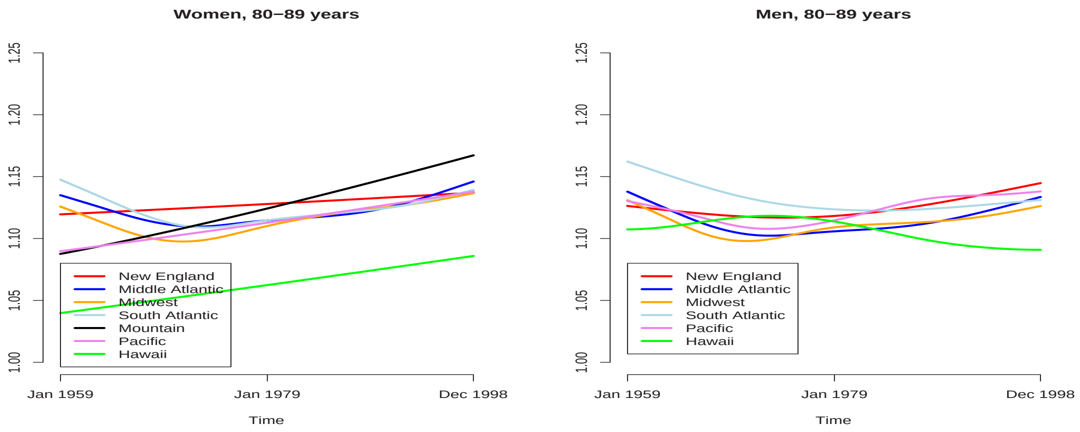
Figure 8 portrays how seasonal death fluctuations have changed over time in various regions of the United States. For reasons of clarity, only results for the age-group 80-89 years have

---

[2] Based on our own calculations using the number of years spent in school of deceased women and men. The results were similar for all ages above 50 as well as for people being 80 years old.

been plotted. Due to numerical optimization problems, only six regions were possible to be displayed for men (missing: “Alaska”, “Mountain” and “East/West South Central”) and seven for women (missing: “Alaska” and “East/West South Central”). Despite this unfortunate loss of information, several interesting features can be observed: The decrease in seasonality discovered in Figure 5 (page 9) did not occur in the US as a whole. Rather three regions were responsible for this development for women and for men likewise: Middle Atlantic, South Atlantic and the Midwest. They showed decreasing seasonality for the first decade observed. All other regions already showed an increase during that period. With the exception of Hawaii, as indicated by the green line, trends are converging for the remaining regions since the late 1960s. This suggests that the existing climatic differences have become less and less relevant over time as social circumstances and living conditions have become more alike in all regions. Hawaii represents an outlier — especially for women. One could either argue that seasonality in Hawaii is smaller than in other regions because of the predominant tropical climate. There, less precautions are required to avoid cold-related mortality during certain seasons as the temperature varies there less than in other (climatic) regions of the United States. It could be, however, also a statistical artifact due to the small number of deaths on Hawaii compared to the other analyzed regions. This latter hypothesis receives support from the study by Seto et al. (1998). They found differences between winter and summer mortality from coronary artery disease mortality of 22%. This shows that seasonal mortality in Hawaii does not differ from the United States as a whole since we found roughly the same results in a descriptive analysis of winter/summer differences for cardiovascular diseases (Winter/Summer Ratio 1.206).

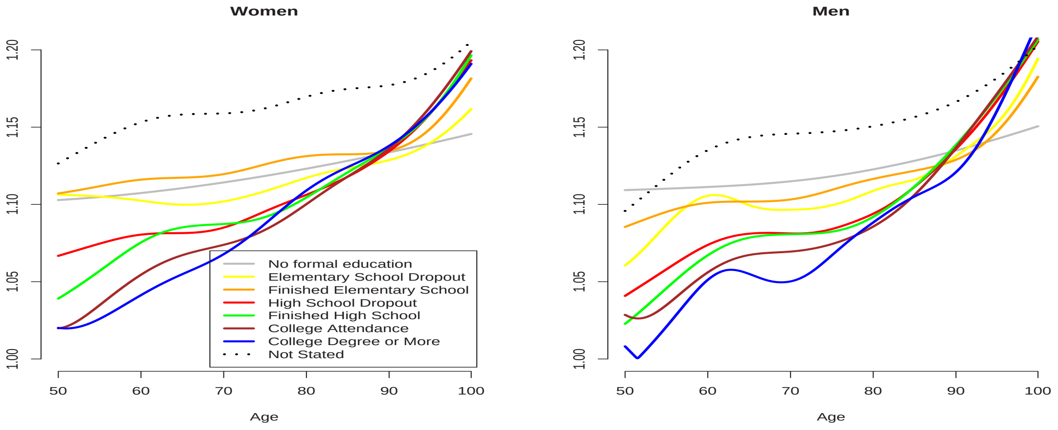
Figure 8: Seasonality of Deaths by Region and Sex, 1959–1998



3.5 Seasonality by Education & Age

Educational level serves as an indicator for socioeconomic status. How this variable affects seasonal fluctuations in deaths over age for women and men during the period 1989–98 is portrayed in Figure 9. For women and men alike, seasonal fluctuations are the highest for the category “not stated” given by the dashed black line. Apart from that residual category, a clear social gradient in seasonal mortality is observable until age 90. The biggest difference is to be seen between people who have earned a college degree or more (blue solid line) and who have received no formal education at all (grey solid line). Persons who belong to the highest educational group have the lowest seasonal amplitude and vice versa. Again, it is remarkable how little women and men differ from each other in terms of seasonal fluctuations. The social gradient diminishes with age and vanishes completely for both sexes at about age 90. The path to convergence is interesting: people with highest completed education show a relatively steep slope whereas the pattern of people without any formal education is rather constant over time. One could therefore argue that people with relatively poor education face seasonal fluctuations in deaths throughout large parts of their adult lives which highly educated people only have to face at very advanced ages. Our estimates show that education does not matter for seasonal mortality when people are 90 years old. It is hard to make any inferences about the last years in our age span until the 100th birthday. It seems as if people without any formal education (grey line) do not become more susceptible to stressful environmental living conditions. Whether a direct effect or an indirect (compositional) effect or both cause this stationary pattern is hard to answer. A direct effect would assume that people with low education are so weak in general that they die regardless of the current season. Contrastingly, a selection effect is also imaginable: as people with lower education tend to die at younger ages (Valkonen, 1989), only a highly selected subgroup is still alive at ages above 90. It is possible, that those people are especially strong in withstanding environmental stress during winter. This latter hypothesis receives further support when the development after age 90 is investigated for the other educational groups. A social gradient is still observable but the other way round. However, the ones facing higher seasonal fluctuations are highly educated people whereas people with less education display smaller seasonal amplitudes. This pattern is possibly a reflection of a compositional effect as people with higher education are less selected than people with lower education.

Figure 9: Seasonality Mortality by Age, Sex and Educational Status, 1989–1998



References

Aylin, P., S. Morris, J. Wakefield, A. Grossinho, L. Jarup, and P. Elliott (2001). Temperature, housing, deprivation and their relationship to excess winter mortality in Great Britain, 1986–1996. *International Journal of Epidemiology 30*, 1100–1108.

Barron, D. N. (1992). The Analysis of Count Data: Overdispersion and Autocorrelation. *Sociological Methodology 22*, 179–220.

Bruce, A. G. and S. R. Jurke (1992). Non-Gaussian Seasonal Adjustment: X-12 ARIMA Versus Robust Structural Models. SRD Report Series 92/14, Bureau of the Census, Statistical Research Division, Washington, D.C.

Cameron, A. C. and P. K. Trivedi (1998). *Regression Analysis of Count Data.* Cambridge, UK: Cambridge University Press.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics 6*, 3–73.

Cleveland, W. S., S. J. Devlin, and I. J. Terpenning (1981a). The details of the SABL transformation, decomposition and calendar methods. Technical report, Computing Information Library, Bell Labs, Murray Hill, NJ.

Cleveland, W. S., S. J. Devlin, and I. J. Terpenning (1981b). The SABL statistical and graphical methods. Technical report, Computing Information Library, Bell Labs, Murray Hill, NJ.

Donaldson, G. and W. Keatinge (1997). Mortality related to cold weather in elderly people in southeast England, 1979-94. *British Medical Journal 315*, 1055–1056.

Donaldson, G. and W. Keatinge (2003). Cold related mortality in England and Wales; influence of social class in working and retired are groups. *Journal of Epidemiology and Community Health 57*, 790–791.

Eilers, P. H. C. and B. D. Marx (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science 11*, 89–102.

Eilers, P. H. C. and B. D. Marx (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics 11*, 758–783.

Eurowinter Group (2000). Winter mortality in relation to climate. *International Journal of Circumpolar Health 59*, 154–159.

Eurowinter Group (1997). Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *Lancet 349*, 1341–1346.

Feinstein, C. A. (2002). Seasonality of deaths in the U.S. by age and cause. *Demographic Research 6*, 469–486.

Grut, M. (1987). Cold-related death in some developed countries. *The Lancet* (8526), 212. 24 January 1987.

Hastie, T. and R. Tibshirani (1993). Varying-Coefficient Models. *Journal of the Royal Statistical Society. Series B (Methodological) 55*, 757–796.

Healy, J. (2003). Excess winter mortality in Europe: a cross country analysis identifying key risk factors. *Journal of Epidemiology and Community Health 57*, 784–789.

Keatinge, W., S. Coleshaw, and J. Holmes (1989). Changes in seasonal mortalities with improvement in home heating in England and Wales from 1964 to 1984. *International Journal of Biometeorology 33*, 71–76.

Lawless, J. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics 15*, 209–225.

Lawlor, D., R. Maxwell, and B. Wheeler (2002). Rurality, deprivation, and excess winter mortality: an ecological study. *Journal of Epidemiology and Community Health 56*, 373–374.

Lawlor, D. A., D. Harvey, and H. G. Dews (2000). Investigation of the association between excess winter mortality and socio-economic deprivation. *Journal of Public Health Medicine 22*, 176–181.

Maravall, A. (2002). Brief Description of the Programs. Downloadable at: `http://www.bde.es/servicio/software/tramo/summprogs.pdf`.

McDowall, M. (1981). Long term trends in seasonal mortality. *Population Trends 26*, 16–19.

McKee, C. M. (1989). Deaths in winter: Can Britain learn from Europe? *European Journal of Epidemiology 5*(2), 178–82.

Näyhä, S. (1980). *Short and medium term variations in mortality in Finland.* Ph. D. thesis, Department of Public Health Service, University of Oulu, Finland.

Nourney, M. (1973). Methode der Zeitreihenanalyse. *Wirtschaft und Statistik*, 11–17.

Nourney, M. (1975). Weiterentwicklung des Verfahrens der Zeitreihenanalyse. *Wirtschaft und Statistik*, 96–101.

Nourney, M. (1983). Umstellung der Zeitreihenanalyse. *Wirtschaft und Statistik*, downloadable at: `http://www.destatis.de`.

Pickle, L. W., M. Mungiole, G. K. Jones, and A. A. White (1996). Atlas of United States Mortality. Technical report, National Center for Health Statistics, Hyattsville, MD.

Quetelet, A. (1838). *De l'influence des saisons sur la mortalité aux différens ages dans la Belgique*. Bruxelles, B: M. Hayez.

Rau, R. and G. Doblhammer (2003). Seasonal mortality in Denmark. The role of sex and age. *Demographic Research 9*, 197–222.

Sakamoto-Momiyama, M. (1977). *Seasonality in Human Mortality. A Medico-Geographical Study*. Tokyo, JP: University of Tokyo Press.

Seto, T. B., M. A. Mittleman, R. B. Davis, D. A. Taira, and I. Kawachi (1998). Seasonal variation in coronary artery disease mortality in Hawaii: observational study. *British Medical Journal 316*, 1946–1947.

Shah, S. and J. Peacock (1999). Deprivation and excess winter mortality. *Journal of Epidemiology and Community Health 53*, 499–502.

Speth, H.-T. (2004). Komponentenzerlegung und Saisonbereinigung ökonomischer Zeitreihen mit dem Verfahren BV4.1. Methodenberichte, Heft 3, 2004, Statistisches Bundesamt, Wiesbaden, D.

Valkonen, T. (1989). Adult mortality and level of education: a comparison of six countries. In J. Fox (Ed.), *Health inequalities in European countries*. missing.

Yaffee, R. (2000). *Time Series Analysis and Forecasting with Applications of SAS and SPSS*. San Diego, CA: Academic Press.