SFB 649 Discussion Paper 2016-023

A Mortality Model for Multi-populations: A Semi-Parametric Approach

Lei Fang*
Wolfgang K. Härdle*
Juhyun Park*[2]

* Humboldt-Universität zu Berlin, Germany
*[2] Lancaster University, United Kingdom

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

A Mortality Model for Multi-populations: A Semi-Parametric Approach*

Lei Fang†

Wolfgang K. Härdle‡

Juhyun Park§

Humboldt-Universität zu Berlin

Lancaster University

Abstract

Mortality is different across countries, states and regions. Several empirical research works however reveal that mortality trends exhibit a common pattern and show similar structures across populations. The key element in analyzing mortality rate is a time-varying indicator curve. Our main interest lies in validating the existence of the common trends among these curves, the similar gender differences and their variability in location among the curves at the national level. Motivated by the empirical findings, we make the study of estimating and forecasting mortality rates based on a semi-parametric approach, which is applied to multiple curves with the shape-related nonlinear variation. This approach allows us to capture the common features contained in the curve functions and meanwhile provides the possibility to characterize the nonlinear variation via a few deviation parameters. These parameters carry an instructive summary of the time-varying curve functions and can be further used to make a suggestive forecast analysis for countries with barren data sets. In this research the model is illustrated with mortality rates of Japan and China, and extended to incorporate more countries. All numerical procedures are transparent and reproduced on www.quantlet.de.

JEL classification: C14, C32, C38, J11, J13

Keywords: Nonparametric smoothing; Parametric modeling; Common trend; Mortality; Lee-Carter method; Multi-populations

*The authors gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft through the International Research Training Group IRTG 1792 "High Dimensional Non Stationary Time Series" and the Collaborative Research Center CRC 649 "Economic Risk".

†Humboldt-Universität zu Berlin, C.A.S.E. - Center for Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany. E-mail: fanglei@hu-berlin.de

‡Humboldt-Universität zu Berlin, C.A.S.E. - Center for Applied Statistics and Economics, Unter den Linden 6, 10099 Berlin, Germany. Visiting Professor in Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, 6th Level, School of Economics, Singapore 178903.

§Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.

I. Introduction

In recent years, global population trend has received wide-spread attention because we are in a fast-aging society which raises demographic risk in most developed and even some developing countries. Here demographic risk is understood to be an imbalance of the age distribution of a society with the obvious implication of economic growth, social stability, political decisions and resource allocation. The factor demography is in particular important for the Asian region since Asia has been experiencing continuous economic growth, during which time the age structure of the population could affect the employment, the labour force and even social and economic stability.

China, as a large developing Asian country, is experiencing the transformation to an aging society at an even faster rate, and is therefore a good example with which to study demographic risk. However, due to political reasons and delays in construction of a national system collecting statistics dating from the last century, statisticians always face the problem of insufficient and unsatisfying Chinese demographic data sets when they try to apply stochastic demographic models; accordingly it brings about lots of research activities in this field. Japan, China's neighbour, has demonstrated a dramatic demographic change during the last several decades; but fortunately the Japanese government had already set up a complete national statistical system (middle of last century) and thus could provide qualified demographic data sets in longer time horizons to help researchers explore Japan's demographic transition, which can thus be analyzed as a good reference to China as well.

1. Literature Review

Since 1980, one of the challenges in demography has been to analyze and forecast mortality in a purely statistical way without involving the subjective opinions of experts. Lee and Carter (1992) (LC) firstly proposed a stochastic method based on a Singular Value Decomposition technique to explore the unobserved demographic information, which proved insightful and gained a good reputation. Later on, several methods based on stochastic population modelling and forecasting have been developed, see e.g. Cairn et al. (2008), Booth and Tickle (2008) and Booth (2006) for review. Among all the stochastic models, the most popular one is the LC model, which was used to analyze the U.S. mortality rates from 1933 to 1987. Based on the idea of the LC model, comparisons of different methods and some variants or extensions have been developed. Lee and Miller (2001) compared the forecasts of LC model with the U.S. social security system forecasts. Li et al. (2004) proposed another method when there are fewer observations at uneven intervals, and applied it to China and South Korea. Hyndman and Ullah (2007) developed a more general method by treating the underlying demographic process as functional data, employing the functional principal components analysis to extract more than one explaining components and providing robust estimation and forecast.

In light of limited data access combined with fragile quality, less technically refined methods

are available for Asian countries compared to, for example, developed western countries. An exception is the stochastic population approach on Asian data by Li et al. (2004), who implemented the LC model to sparse data. In their work, they generated central forecast with just the first and last observations along the time horizon, improved the estimates by additional observations and evaluated its performance with other existing methods. Raftery et al. (2012) proposed a Bayesian method for probabilistic population projections for all countries, where the Bayesian hierarchical models, estimated via Markov chain Monte Carlo, are applied to the United Nations population data. In cases of limited data and similar demographic trends between two populations (regional or national level), the Bayesian stochastic modelling for two populations is proposed by Cairns, Blake et al (2011). This motivates us to analyze Chinese demography via taking Japan into reference.

An interesting finding from Fang and Härdle (2015) is that China has a demographic trend closer to Japan than to Taiwan, particularly visible in the mortality trend. On the economic level, China and Japan have been both important economies in last several decades and the development pattern is also quite similar. Hanewald (2011) found that the LC mortality index k_t correlates significantly with macroeconomic fluctuations in some periods, which provides a good reference with which to connect the mortality trends between China and Japan. The research from Härdle and Marron (1990) on semiparametric comparison of regression curves and also the one from Kneip and Engel (1995) suggest one potential way to analyse demographic common trend between two countries and even among multi-countries.

2. Goal and Outline

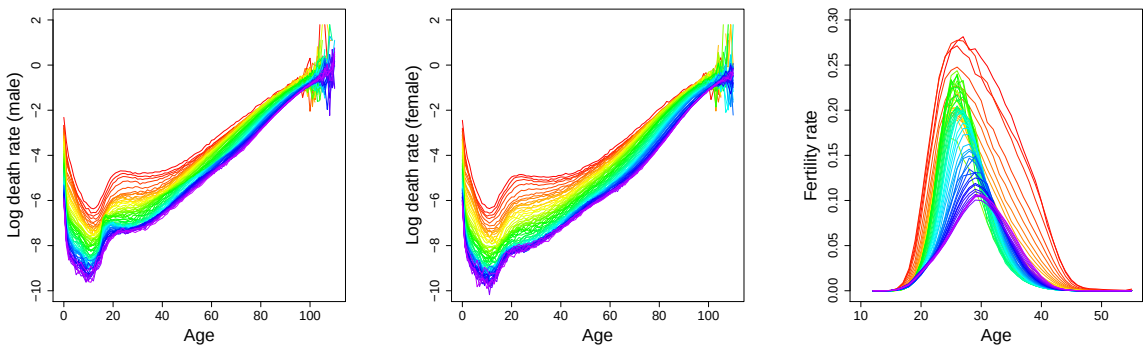


Figure 1: *Japan's descriptive demography - male (left) and female (middle) mortality dynamics from 1947 to 2012, fertility (right) dynamics from 1947 to 2009, rates in different years are plotted in rainbow palette order.*

MuPoMo

Due to the sensitivity of fertility to social policies and induced unpredictability, our research is restricted to mortality analysis. In this research, we apply the LC method to mortality data sets,

make use of the investigated mortality time-varying indicator k_t to compare the similarities of different countries via a semi-parametric comparison approach and accordingly propose potential mortality forecasting improvements. In the following section, we will discuss the methodology in more details. Section III focuses on an empirical research of mortality data sets, including China, Japan and the other 34 countries obtained from the Human Mortality Database. We will analyse mortality similarities between China and Japan, furthermore extend it to a global common mortality trend and sub-group pattern. More discussions on economic insights, global aging trend influences and suggestions will be organised in the last section.

II. Methodology

In this section, we firstly introduce the parameters of interests and then outline the LC method, semi-parametric comparison of nonlinear curves and common trend modeling in details.

1. Notations and Parameters of Interest

In this research, parameters of interests are age-specific mortality rates from multiple countries. We use the symbols m to denote mortality rate. All the parameters are indexed by a one-year age group, denoted by x, and in addition indexed by time, denoted by t. For instance, $m(x,t)$ is the mortality rate for age x in year t.

Based on the LC model, $m(x,t)$ is decomposed into average age pattern and time-varying index k_t (for single country, state or region), which plays an important role in the following research of semi-parametric comparison of nonlinear curves and common trend modeling. When it comes to multiple countries, we use $k_i(t)$ to denote the derived time-varying mortality indicator for country i, with $i \in \{1, ..., N\}$.

2. Lee-Carter (LC) Method

The benchmark LC model employs the Singular Value Decomposition (SVD) to analyse the time series on the log of the age-specific mortality. The method relies on the standard statistical analysis of the time series. Nonetheless, the LC model does not fit well in some cases where missing data is common or the horizon of time series is not sufficient, the reason being the assumption of long-term stationarity.

The basic idea for demography dynamics analysis is to regress mortality $m(x,t)$ on non-observable regressors for prediction. The regressors are obtained via SVD of the demographic indicators. It separates the age pattern from the time-dependent components, takes time series analysis on the time-dependent components only and hence forecasts the future trend.

The mortality rate $m(x,t)$ is hence calibrated via the following model:

$$\log\{m(x,t)\} \quad = \quad a_x + b_x k_t + \varepsilon_{x,t}, \tag{1}$$

or

$$m(x,t) \quad = \quad \exp(a_x + b_x k_t + \varepsilon_{x,t}),$$

where $m(x,t)$ denotes the historical mortality rate at age x in year t, a_x is the derived age pattern averaged across years, b_x stands for the sensitivity of the mortality rates to the change of k_t, reflecting how fast the mortality rate changes over ages, k_t represents the only time-varying index of mortality level, and $\varepsilon_{x,t}$ is the residual term at age x in year t with $\mathsf{E}(\varepsilon_{x,t}) = 0$ and $\mathsf{Var}(\varepsilon_{x,t}) = \sigma_\varepsilon^2$.

Three unobserved parameters a_x, b_x and k_t in the single equation (1) mean that the LC model is over-parameterized and therefore two normalisation constraints are imposed:

$$\sum k_t = 0, \ \sum b_x = 1\,.$$

By SVD, one obtains k_t and b_x. The evolution of k_t can be fitted by ARIMA techniques, like the Box-Jenkins procedure. After model identification, Lee and Carter (1992) found that a random walk with drift describes k_t quite well:

$$k_t = k_{t-1} + d + e_t$$

where d is the drift parameter reflecting the average annual change and e_t is an uncorrelated error. Other research taken by Chan et al. (2008) pointed out that the mortality index k_t may be better fitted with a trend-stationary model for Canada, England and United States when accounting for a break in mortality rate decline during the 1970s.

Given an h-step ahead forecasting k_{t+h}, we could, in return, forecast the mortality rates in future period $t + h$ via the following formula:

$$m(x,t+h) \quad = \quad \exp(a_x + b_x k_{t+h}).$$

3. Semi-Parametric Comparison of Nonlinear Curves

When the observable curves are noisy versions of similar regression curves, comparison of regression curves from related samples is not trivial. As demonstrated in Fig. 2, when the differences among similar curves mainly rely on shifted time axis and vertical re-scaling, comparison of similar curves could be simplified by quantifying differences through parameters describing horizontal and vertical shifts. In addition, nonparametric smoothing techniques could be used to estimate the underlying curves when the solid theory is unavailable in modeling them. Hence, the general semi-parametric models where a comparison of nonparametric smoothed curves is framed in a parametric way are desirable in our research.

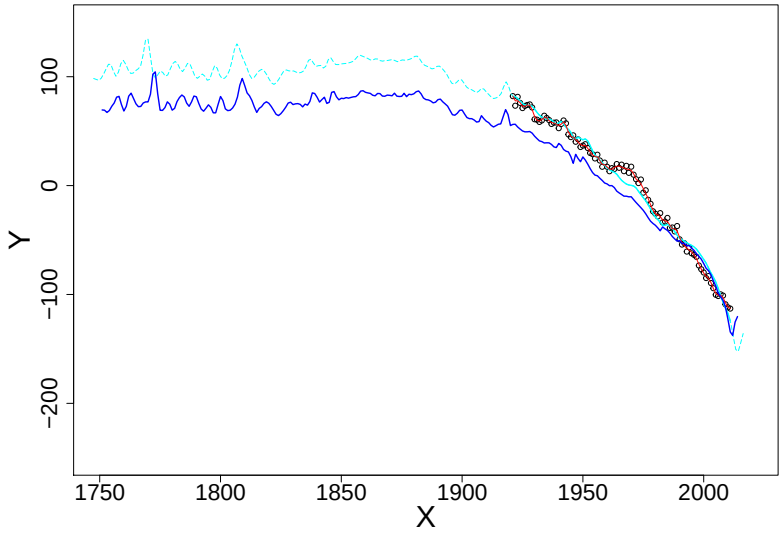


Figure 2: *Semi-parametric Comparison of Nonlinear Curves: the dark blue curve and the red curve have similar pattern, while the light blue curve is semi-parametrically shifted to represent the red curve via the dark blue curve.*

MuPoMo

Simply denoting the underlying curves by f_1 and f_2, the semi-parametric comparison of the nonlinear curves is framed as below in our research:

$$f_2(t) = \theta_1 f_1\left(\frac{t-\theta_2}{\theta_3}\right) + \theta_4, \tag{2}$$

where we assume that f_2 has a similar pattern to f_1 and $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\top$ are shape deviation parameters. Since normally only noisy measurements of the curves are available, f_1 itself can be further estimated by nonparametric smoothing methods, which in turn gives an estimate of f_2. More detailed discussions on this method can be referred to Härdle and Marron (1990) on semiparametric comparison of regression curves.

4. Common Trend Modeling

When more than two regression curves share similar pattern or trend, a common trend model can be built based on the technique of semi-parametric comparison of nonlinear curves we discussed previously. Suppose we are given N noisy curves $\mathrm{Y}_i, i = 1, \ldots, N$ that exhibit some similar patterns. A general regression model can be expressed as,

$$\mathrm{Y}_i = f_i(t) + \varepsilon_i\,,$$

where f_i denote unknown smoothing regression functions while ε_i represent independent errors with mean 0 and variance σ_i^2.

The relationship among these similar curves can be described as

$$f_i(t) = \theta_{i1} g\left(\frac{t-\theta_{i2}}{\theta_{i3}}\right) + \theta_{i4}. \tag{3}$$

Here $\theta_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}, \theta_{i4})$ are unknown parameters describing shape deviations, and g is a unknown function specifying the common shape of these curves, which can be interpreted as a reference curve. The model in Equ. (3) is commonly known as shape invariant model (SIM), firstly proposed by Lawton et al. (1972) and further studied by Kneip and Engel (1995) and provides an extension of the model in Equ. (2) to multiple curves. A detailed investigation of this model to estimate the mortality trend will be given in the following section.

III. Empirical Research

In this section we will focus on empirical research of mortality data sets, including China, Japan and the other 36 countries. We will analyse mortality similarities between China and Japan, furthermore extend it to a global common mortality trend and sub-group pattern.

1. Data Sets

The demographic data sets are collected from different sources: China data sets are extracted from the China Statistical Year Book, the other 35 countries are obtained from the Human Mortality Database (HMD). For data sets on China the mortality rates are gender-age specific starting from 0 to 90+ years old, while the mortality rates of other countries are gender-age specific from 0 to 110+ since the Human Mortality Database makes estimates and adjustments on the raw data to extend into wider age group.

As for the sample size, they are all different as well: China mortality data spans from 1994 to 2010 with missing data of years 1996, 1997, 2001 and 2006, while the sample size from other countries range from 14 years (Chile) to 261 years (Sweden). Referring to the missing values, we use moving average of neighboring five years to compute these.

Recall that the definition of the mortality rate is the number of deaths per 1000 living individuals per single calender year. To fit the mortality trend more precisely and for visual convenience, we present the log mortality.

2. Two-Country Case

We start with comparing the mortality trends between China and Japan. In the left of Fig. 3, the mortality trends of China and Japan are displayed. It reads that mortality trends from both gender groups of China correlate with those of Japan respectively, and due to limited sample they all seemingly reflect linear mortality trend over time even though in theory it can not be this case.

To provide intuitive comparison between these two countries, the horizontal shift of Japan's mortality curve over time axis is plotted in the right part of Fig. 3. The dotted lines from left to

right are shifted Japan's smoothed trends of 20-, 23- and 25- years forward respectively. Graphically we see that Japan's mortality trend is 23 years earlier than China.

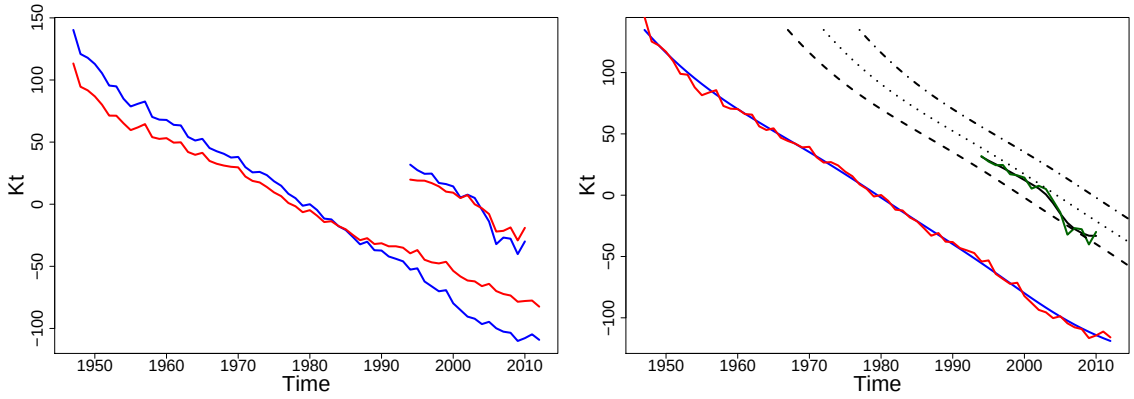


Figure 3: *China mortality trend vs. Japan mortality trend: female, male (left) and Japan trend, Japan smoothed trend, China trend and China smoothed trends (right).The dotted lines (right) from left to right are shifted Japan's smoothed trends of 20-, 23- and 25- years forward respectively.*

MuPoMo

2.1 Model

To parameterize the potential relationship between China and Japan mortality trend, we specify the model as following, and use k_t derived from LC model in Equ. (1).

$$\log\{m(x,t)\} \quad = \quad a_x + b_x k_t + \varepsilon_{x,t}\,,$$

Then we infer China's mortality trend via Japan's trend through the technique of semi-parametric comparison of regression curves in Equ. (2)

$$k_c(t) = \theta_1 k_j\left(\frac{t-\theta_2}{\theta_3}\right) + \theta_4, \tag{4}$$

where $k_c(t)$ is the time-varying indicator for China, $k_j(t)$ is the time-varying indicator for Japan, and $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\top$ are shape deviation parameters.

Understanding θ It is probably easiest to understand Equ. (4) by starting with $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\top = (1, \theta_2, 1, \theta_4)^\top$.

- θ_1 is the general trend adjustment, here selected as 1.
- θ_2 is the time-delay parameter
- θ_3 is the time acceleration parameter, here selected as 1.
- θ_4 is the vertical shift parameter

In Fig. 4, it demonstrates how θ influences shift of these two curves. In the left of Fig. 4, China's female k_t can reach a similar behavioral area by shifting horizontally $\theta_2 = -23$, while in

the right it shows that another acceptable area could be obtained via approximately vertical shift of $\theta_4 = 85$.

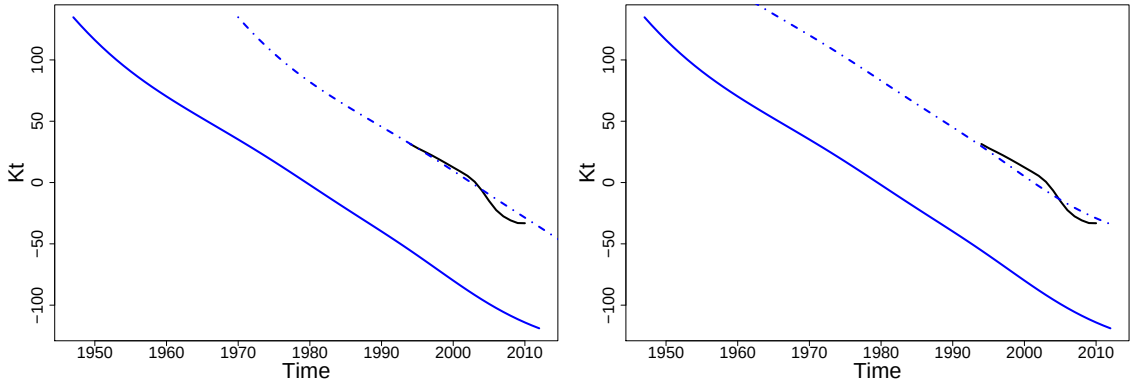


Figure 4: *Time delay $\theta_2 = -23$ (left) and vertical shift $\theta_4 = 85$ (right).*

MuPoMo

2.2 Estimation

In order to find the optimal solution for shape deviation parameters, we minimize the following loss function.

$$\min_{\theta} \int_{t_c} \left\{ \hat{k}_c(u) - \theta_1 \hat{k}_j \left(\frac{u - \theta_2}{\theta_3} \right) - \theta_4 \right\}^2 w(u) du, \tag{5}$$

where $\hat{k}_c(t)$ and $\hat{k}_j(t)$ are the nonparametric estimates of the original time-varying indicators $k_c(t)$ and $k_j(t)$, and t_c is the time interval of China mortality data.

The comparison region needs to satisfy the following condition, in order to make sure the parameter estimation is compared only in the common region defined by $w(u)$.

$$w(u) = \prod_{t_j} 1_{[a,b]}\{(u - \theta_2)/\theta_3\},$$

where t_j is the time interval of Japan's mortality data, $a \geq inf(t_j)$ and $b \leq sup(t_j)$. To consider the importance of more recent data's impact on future trend, we could even impose different weights on different support intervals.

Algorithm To estimate the parameters by the nonlinear least squares estimation criterion in Equ. (5), we first obtain the estimates of k_c and k_t by nonparametric local linear smoothing, denoted by $\hat{k}_c$ and $\hat{k}_t$ respectively. Then we set up the initial estimates $\theta^0 = (\theta_1^0, \theta_2^0, \theta_3^0, \theta_4^0)$ and solve the nonlinear least squares estimation problem by iteratively updating the estimates until convergence.

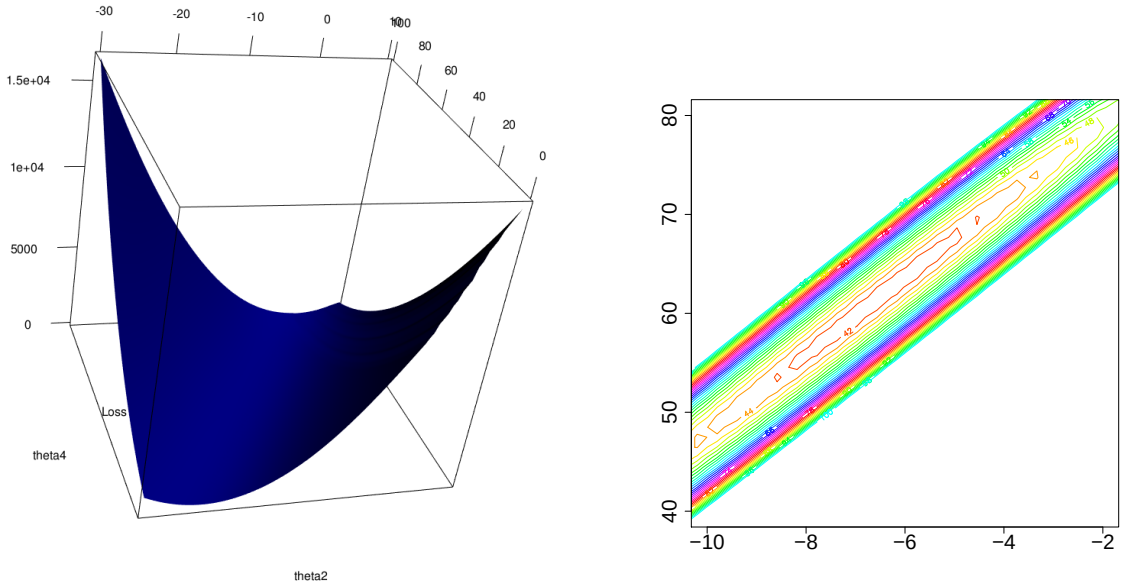


Figure 5: *Loss surface of θ_2 and θ_4 (left) and Contour of θ_2 and θ_4 (right) with $\theta_1 = \theta_3 = 1$.*

MuPoMo

Initial choice of θ_2 and θ_4 From previous Fig. 3 and in our initial analysis, we see that there is a potential ambiguity between θ_2 and θ_4. It is also clear that we can find the replacement relationship from Fig.4. To investigate further, we show the criterion function in Fig.5 as a function of (θ_2, θ_4). There is a valley area in the loss surface function of θ_2 and θ_4 in the left plot and also in the contour of θ_2 and θ_4 in the right one, which suggests that there exists an approximate linear combination of θ_2 and θ_4 in searching θ for an optimal solution. It will bring about difficulties in finding the optimal parameters θ in numerical optimization.

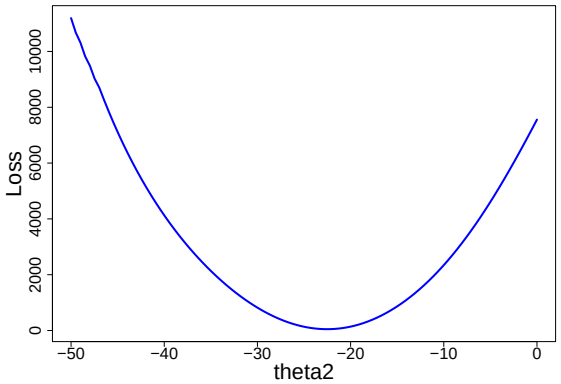


Figure 6: *Loss function of θ_2 with $(\theta_1, \theta_3, \theta_4)^\top = (1, 1, 0)^\top$.*

MuPoMo

In order to find the optimal value, we should be very careful with selecting initial values of

θ, so we need to decide whether the analysis concentrates on time delay or vertical shift. In our analysis we stick with time delay influence θ_2 since it is more valuable in prediction perspective, and thus the initial value of θ is determined as $(\theta_1, \theta_3)^\top = (1, 1)^\top$ and remove the parameter θ_4 because it will also bring about ambiguity in optimization. Hence the optimal initial θ_2 is obtained around -23, see Fig. 6.

Goodness of Fit Based on the initial $\theta^{(0)} = (\theta_1, \theta_2, \theta_3)^\top = (1, -23, 1)^\top$ and algorithm, the optimal parameter θ is reached at $\hat{\theta} = (1.205, -22.621, 1.000)^\top$. From Fig.7, one sees that after the curve shifts (based on the optimal value of θ) the k_t of China fits quite well in the k_t of Japan of years around from 1970 to 1990.

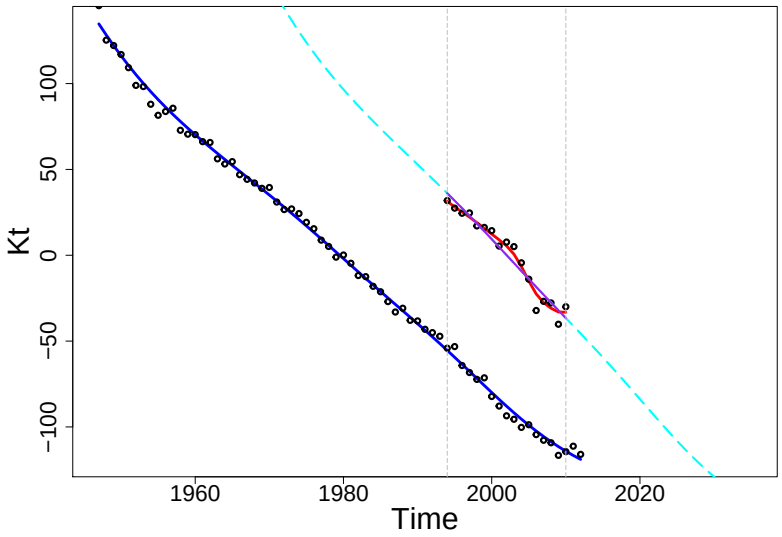


Figure 7: *Goodness of Fit and forecast of China's mortality trend from 2011 to 2030 via Japan's historical data: black dots represent the original k_t from Japan and China, and Japan smoothed trend, China smoothed trend are displayed as well; the fitted trend is plotted as light blue dashed line, while the overlapping part is colored in purple; the light blue dashed line after year 2011 is the forecast part.*

MuPoMo

2.3 Forecast

Afterwards we can forecast k_t for China via the data from Japan and the optimal estimated shape deviation parameter $\hat{\theta}$ to extend the forecasting horizon.

$$k_c(t+h) = \hat{\theta}_1 k_j \left\{ \frac{(t+h) - \hat{\theta}_2}{\hat{\theta}_3} \right\}, \tag{6}$$

where $\hat{\theta} = (1.205, -22.621, 1.000)^\top$ and $t = 1994, 1995, ..., 2010; h = 1, 2, ..., 20$.

Compared with the traditional forecasting method with time series analysis, our proposed method can extend the forecasting horizon from 5 years to 23 years, which is a big advantage from

this semi-parametric comparison technique of regression curves, see Fig. 7. However, searching for numerical optimal solutions for seemingly linear regression curves is still a challenge and a problem. That is a motivation why multi-populations mortality comes as next stage.

3. Multi-Countries Case

In the light of the study of two countries under sparse data, investigation of multi-populations becomes promising and necessary since more information on nonlinear trend will be provided in case of multi-countries. At the same time it helps to study global mortality trend in the past century and its future.

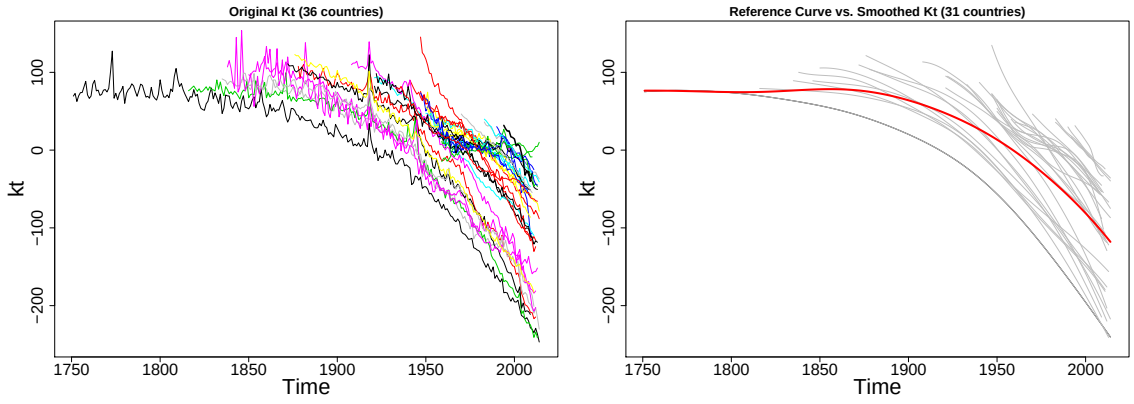


Figure 8: *Similarities of mortality trend among countries: different colors represent different countries (left), and the red thick curve in the right plot stands for reference curve while grey ones are smoothed curves.*

MuPoMo

Denote by k_i the derived time-varying mortality indicator for country i, with $i \in \{1, ..., N\}$. Fig. 8 displays the estimates from 36 countries. On the left, the 36 curves display originally estimated k_t from the LC model without any further nonparametric smoothing, while on the right it shows the smoothed k_t with an initial estimate of the reference curve overlayed. Later on, we will discuss why 31 countries are selected for analysis. By design, the available time measurement varies among countries. Nevertheless, we notice remarkable similarities in the trend across the countries, subject to individual variability.

In order to investigate this structural similarity and to borrow the information across the countries, we consider the shape invariant model introduced in Section 4 of methodology part.

3.1 Model

Specifically we assume that the curves share some common trend and can be represented in the form

$$k_i(t) = \theta_{i1} k_0 \left(\frac{t - \theta_{i2}}{\theta_{i3}} \right) + \theta_{i4}, \tag{7}$$

where $k_0(t)$ is a reference curve, understood as common trend and $\theta_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}, \theta_{i4})^\top$ are shape deviation parameters. In order to be able to interpret the reference curve k_0 as a mean trend, we can use the normalizing constraints on the parameter θ_i as

$$N^{-1} \sum_{i=1}^{N} \theta_{i1} = N^{-1} \sum_{i=1}^{N} \theta_{i3} = 1, \tag{8}$$

$$N^{-1} \sum_{i=1}^{N} \theta_{i2} = N^{-1} \sum_{i=1}^{N} \theta_{i4} = 0\,. \tag{9}$$

Alternatively, we can use any country as a reference curve, for example, Sweden as the longest record holder could be a reasonable choice, in which case the reference curve is set to be k_t of Sweden with $\theta_0 = (1, 0, 1, 0)$ and θ_i will measure the deviation with respect to the reference curve. In this work we will consider the mean curve as a reference curve and use the above normalization constraints.

3.2 Estimation

Estimation of parameters Suppose that k_0 and k_i are given. Then for each country i, the parameter θ_i can be determined by minimizing the least squares criterion as

$$\int \Big\{ k_i(t) - \theta_{i1} k_0\Big(\frac{t - \theta_{i2}}{\theta_{i3}} \Big) - \theta_{i4} \Big\}^2 w_i(t)\, dt \tag{10}$$

where w_i is chosen to ensure that the two functions are evaluated over the common domain as in the case of two countries.

Estimation of Common Trend For given parameters $\theta_i, i = 1, \ldots, n$, the functional relationship in (7) implies that

$$k_i(\theta_{i3} t + \theta_{i2}) = \theta_{i1} k_0(t) + \theta_{i4}, \tag{11}$$

Thanks to the normalizing conditions on θ_{i1} and θ_{i4}, this implies that

$$k_0(t) = N^{-1} \sum_{i=1}^{N} k_i(\theta_{i3} t + \theta_{i2})\,. \tag{12}$$

That is, if k_i is appropriately transformed with respect to the individual parameters θ_i, then k_0 is simply the average. In practice, k_i can have measurement errors, and also are available at different number of time points. Then the functional mean can be estimated more efficiently with nonparametric smoothing, which essentially gives rises to a weighted average estimate.

Estimation algorithm Combining the above two steps leads to the following iterative algorithm for estimation of the parameters.

(a) Given $\hat{k}_i$, obtain an initial estimate of k_0 based on all country-level mortality rates

(b) Given k_0, update θ_i by minimizing the nonlinear least squares criterion in (10) for each $i = 1, \ldots, N$.

(c) Normalize the parameters to satisfy the constraints.

(d) Given $\theta_i, i = 1, \ldots, N$, update k_0 by (12).

(e) Iterate (b)-(d) until convergence.

Choice of Smoothing Parameters To initialize k_0, we choose the trimmed mean of the sample estimates, based on the middle 50% of the countries in terms of the length of the recording period. The estimation of k_0 and k_i is done with local linear kernel smoothing method to account for measurement error.

Computational Issues with the Parameterization The shape invariant model implicitly assumes that there are identifiable features that are common across the sample. It is easy to check for densely observed curves (with non-monotone functions) by means of the derivative estimation, but for sparsely observed curves, there could be an ambiguity in identifying the parameters. In the case of the mortality curves, due to the limited measurements available, the ambiguity occurs in distinguishing the differential effect of vertical shift (θ_4) and horizontal shift (θ_2) in time. In this case, we choose to attribute the effect as horizontal shift and set $\theta_4 = 0$, as this is more amenable to interpretation and meanwhile promising to extend forecast horizon.

In following empirical section, comparison of these two parameterization cases will be illustrated.

Bootstrap for Prediction Interval The bootstrap is a simulation-based method to quantify the uncertainty in situations where the traditional methods failed to provide an adequate approximation or variance estimates are complicated to obtain.

In general bootstrap relies on identically independent distributed (i.i.d.) observations. But time series data is another story of containing structures of dependent data, where the dependence data arrangement should be kept during the re-sampling scheme. Recently re-sample methods for dependent data have considered several options: bootstrap with i.i.d. innovations, bootstrap with block segments and model-based bootstrap.

In our section of comparing China with global trend, we use bootstrap method to construct the prediction interval of China's mortality trend. Due to limited sample of China's mortality time series, bootstrap with block segments do not present ideal re-sampled time series. Alternatively we bootstrap the mortality data based on i.i.d. innovations obtained from fitting time series model,

and afterwards we carry estimation on the re-sampled data and generate prediction interval at different levels.

Suppose for the time series we have observed $k_1, ..., k_n$ and some fixed $p \in N$, there exists a parametric estimator of the conditional expectation $E(k_t|k_{t-1}, ..., k_{t-p})$ denoted by $\widehat{m}_n(k_{t-1}, ..., k_{t-p})$. This estimator leads to residuals

$$\widehat{e_t} := k_t - \widehat{m}_n(k_{t-1}, ..., k_{t-p}), t = p+1, ..., n, \tag{13}$$

and in following step to a bootstrap time series

$$k_t^* = \widehat{m}_n(k_{t-1}^*, ..., k_{t-p}^*) + e_t^*, t = 1, ..., n. \tag{14}$$

The bootstrap innovations $e_1^*, ..., e_n^*$ are assumed to share more or less the same variance and identically independent distributed from each other, so that bootstrap time series can be realized via bootstrapping independent innovations or residuals of a parametric conditional predictor of k_t. The idea of parametric fit to the conditional expectation can be executed by ARIMA models. Kreiss and Lahiri (2012) discussed different situations with parametric and nonparametric modeling predictor of k_t and respective asymptotic consistence properties.

3.3 Global Trend

In the light of similarities across countries, seeking global mortality trend becomes very reasonable and natural. Some research also find out mortality trend has connection with economic development level, GDP for instance, see Hanewald (2011). In the remaining section, we are going through the empirical analysis of common trend in different groups via slightly different models.

Figure 9: *Five outlying countries: geographically neighbors in east Europe. Source: Google Map.*

Outlying Countries Evidence tells us that the mortality rate is decreasing as time evolves, due to medical improvement, economic development and social stability. However, there are some remarkable outliers like Russia, Estonia, Latvia, Lithuania and Belarus.

As shown in Fig. 9, all these five countries locate themselves in east Europe and are used to be members of the Soviet Union. They share similar geographic characteristics and meanwhile experienced parallel economic and social progress, and within our expectation they reflect comparable mortality moving path as well, as displayed in Fig. 10 and Appendix 1. Note that solid (blue) curves represent global mortality trend, dashed (cyan) curves are representing estimated individual country-level mortality trends based on global trend, the short solid (red) curves are smoothed original individual country-level mortality trends and (black) circles are original individual country-level mortality trends.

Surprisingly, they exhibit a quite opposite tendency in contrast with other 31 countries. The mortality rates go through a short period of decrease, then stay stable or slightly increase for several years and afterwards go back to declining path again. One possible reason on this different phenomenon perhaps is connected with political event of dissolution of the soviet union.

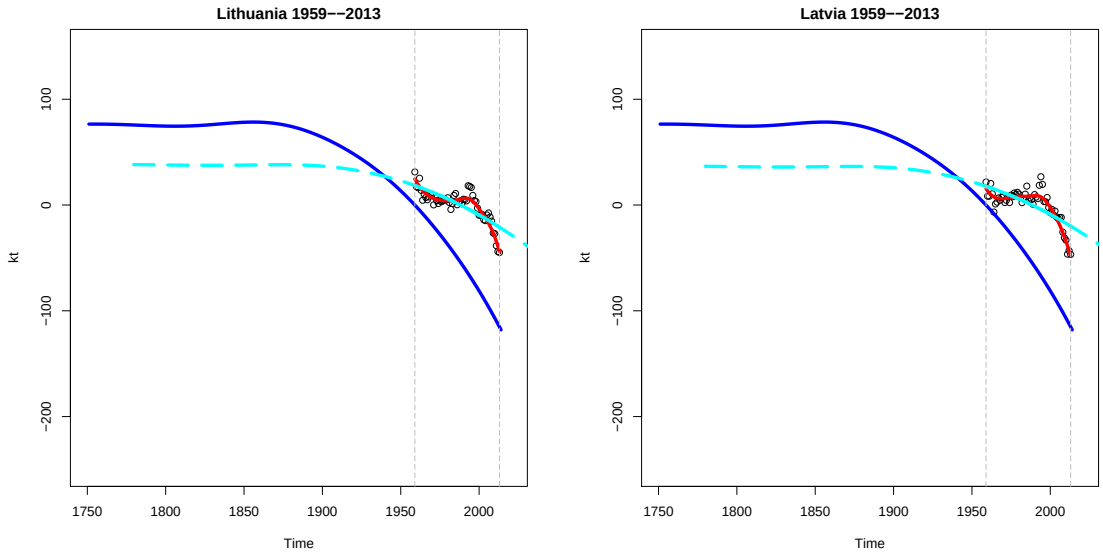


Figure 10: *Different mortality movements of Lithuania and Latvia: blue curves represent global mortality trend, light blue curves are representing estimated individual country-level mortality trends based on global trend, red curves are smoothed original individual country-level mortality trends and black dots are original individual country-level mortality trends.*

MuPoMo

To ideally demonstrate global mortality movements in majority of countries, we remove these five countries for remaining study to reduce influences from minor outlying ones.

Mortality Trend among Majority As discussed previously, for sparsely observed curves there could be an ambiguity in identifying all of the four parameters. Therefore, we choose to compare the original 4-parameters model with a simplified 3-parameter model of setting $\theta_4 = 0$,

$$k_i(t) = \theta_{i1} k_0 \left(\frac{t - \theta_{i2}}{\theta_{i3}} \right).$$

Under two different parameterization, we estimate the parameters and reference curve respectively. In Fig. 11, the reference curves and common trend for these two cases are plotted: the left one is calculated from the 3-parameters model and the right one is generated from the 4-parameters one. The red curve is the initial reference curve, blue ones are one-step ahead updated reference curves.

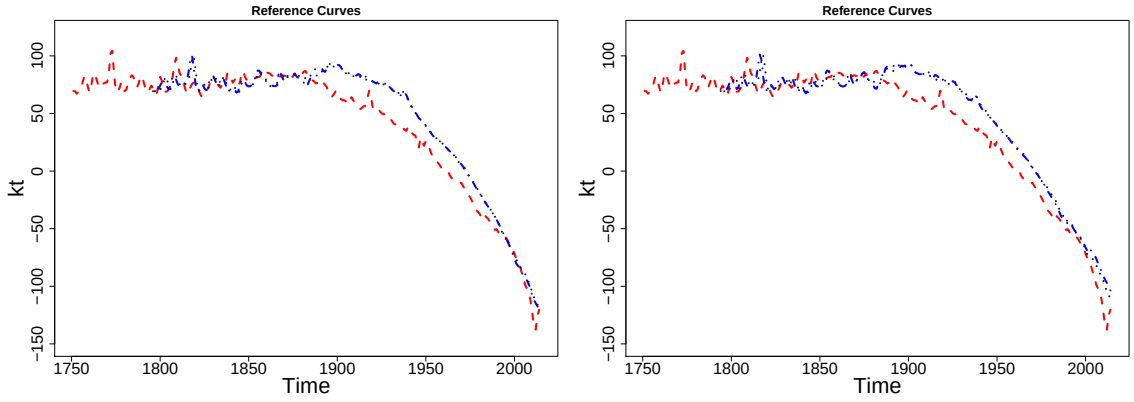


Figure 11: *Common mortality trends estimated by 3-parameters model (left) vs. 4-parameters model (right).*

MuPoMo

From these two plots, no clear and obvious difference can be read out. But from the viewpoint of analytic thinking, we choose to attribute the effect as horizontal shift and set $\theta_4 = 0$, as this is more amenable to interpretation and meanwhile promising to extend forecast horizon.

Fig. 12 explains the common mortality trend generated from 3-parameters model compared with individual nation-level mortality trend. In this graph, the (black) solid curve is the initial reference curve, cyan, green, blue and red ones represent the updated ones at different iteration stage while the grey ones are the non-smoothed mortality trend from each country. It is obvious that after one step optimization, reference curves are already showing a quite similar pattern. The common mortality trend is adjusted to an upper level, mostly because more developing countries (Czech Republic, Hungary and China, for example) started collecting demographic data at a later time period in contrast with more developed countries (such as Sweden, Norway and France), and also developing countries have higher mortality rates generally. The figures on illustrating individual case are provided in Appendix 2.

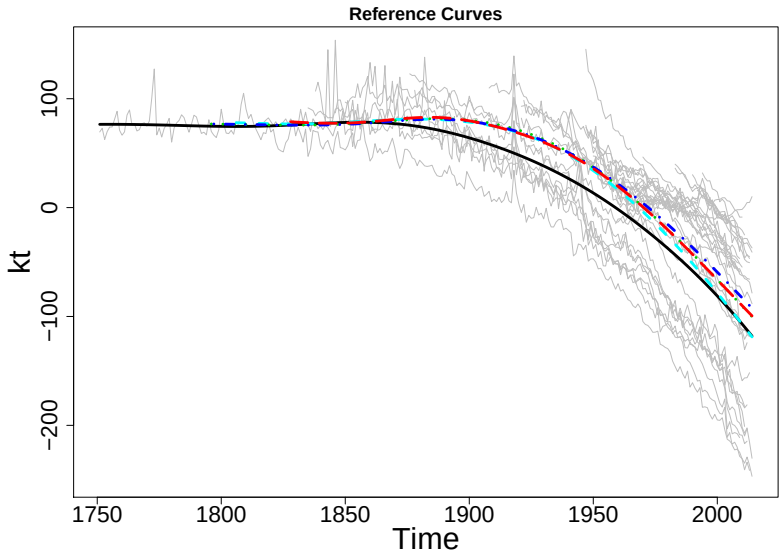


Figure 12: *Common mortality trend compared with individual nation-level mortality trends: red curve is the initial reference curve, blue and black ones are updated reference curves convergent to common trend, while the grey lines represent individual country.*

MuPoMo

3.4 China and Global Mortality Trend

Since a common mortality trend is available, it could be applied to help improve estimation and forecasting of individual case. Especially when sample size from individual country is relatively smaller than anticipated, semi-parametric comparison of common mortality trend with each individual nation-level one will be a promising way with respect to forecasting. At the same time, it helps to reduce ambiguity in identifying the parameters in case of seemingly linear co-movement between regression curves, like the case of comparing China and Japan.

In the following Fig. 13, it displays newly estimated China mortality trend via semi-parametric comparison with common trend. The (blue) solid line is the common trend or updated reference curve and the (cyan) dashed line is the estimated China mortality trend based on the common trend. In comparison, the raw China mortality curve estimate is marked by black circles, with the individual smoothing estimate overlayed in short (red) line. Thanks to parameters deviation on time axis $\frac{t-\theta_2}{\theta_3}$, we could extend forecasting horizon of China approximately 40 years through the information from common trend, see Fig. 13. Referring to estimation and forecasting of the other 30 countries, they are arranged in Appendix 2.

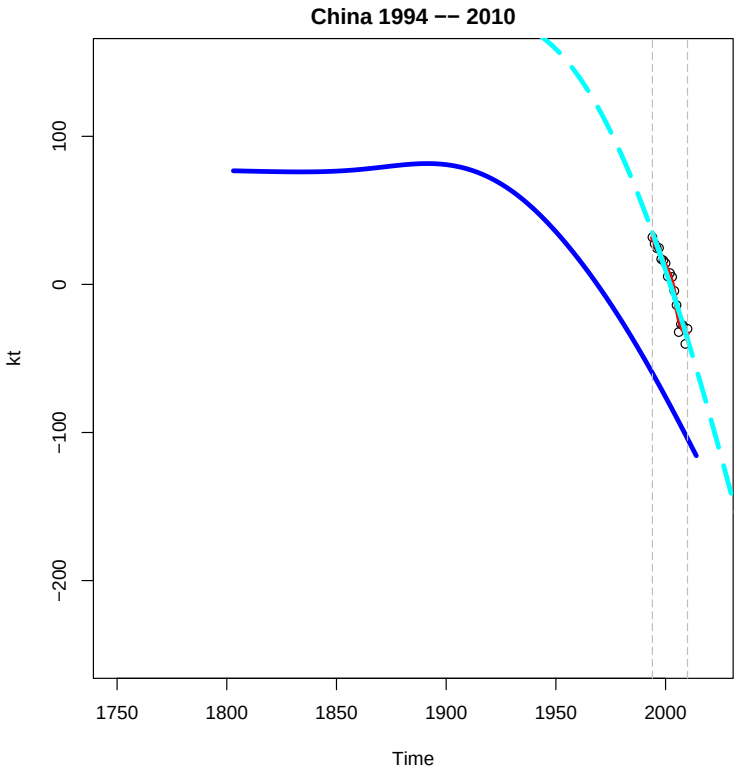


Figure 13: *Common mortality trend and estimated China mortality trend based on common trend.*

MuPoMo

With model-based bootstrap approach, we simulate 500 re-sampled China's mortality time series from 1994 to 2010 based on ARIMA model. From each simulation, we estimate optimal shape deviation parameters θ and accordingly calculate estimated China mortality with longer time horizon based on common trend.

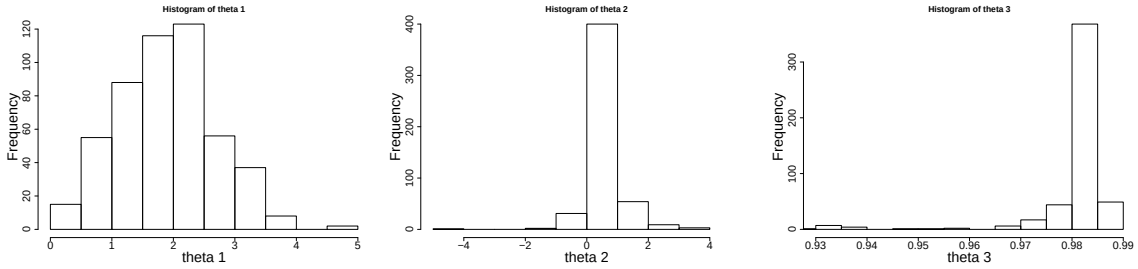


Figure 14: *Histograms of* $\hat{\theta}_1$,$\hat{\theta}_2$ *and* $\hat{\theta}_3$.

MuPoMo

In Fig. 14, we display the variation of $\hat{\theta}_1$,$\hat{\theta}_2$ and $\hat{\theta}_3$ across the countries. From histogram of $\hat{\theta}_1$, 50% of $\hat{\theta}_1$ lies between 1.5 and 2.5, which indicates the overall accelerating declining mortality trend of China compared with global trend. More than 95% of $\hat{\theta}_2$, the parameter describing time delay, falls into the interval of $(0, 10)$, which further confirms there exists a time delay of China's mortality trend around 10 years later than global situation. Majority of $\hat{\theta}_3$ ranges from 0.98 to 1, which reveals a little time acceleration in China's mortality trend.

In Fig. 15, confidence intervals at different levels are displayed. On the left part, confidence intervals at 80% and 90% are plotted in grey zone and blue zone respectively, while yellow zone highlights the central area of possible forecast path. On the right one, only 90% confidence interval is presented. Black line stands for global mortality trend, red one is original China's mortality trend and light blue curve shows the estimated China's mortality trend based common trend and original China's data.

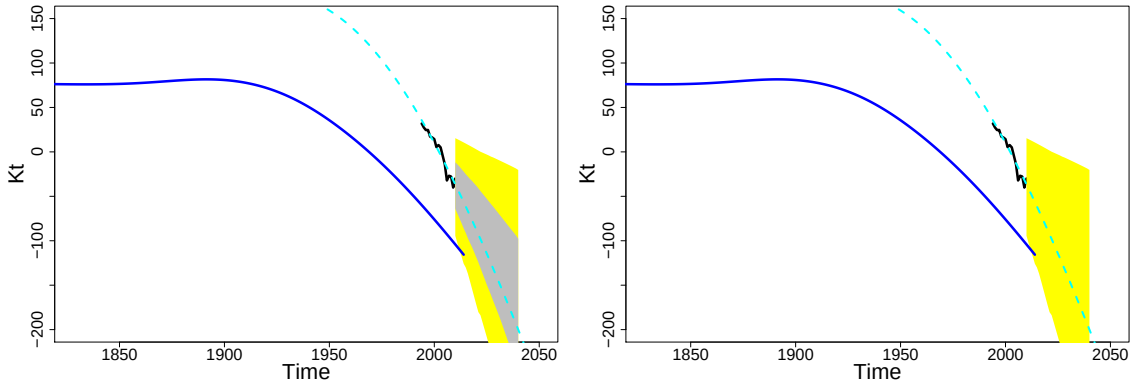


Figure 15: *Confidence intervals at different levels (left) vs. 90% confidence interval (right).*

MuPoMo

IV. Discussion

The global mortality, as expected, is undergoing a shift toward exhibiting declining tendency, and with a dramatic decreasing movement in the last several decades in contrast with hundreds years ago. It also depicts that most of countries are converging with a similar mortality pattern of decreasing over time. The improvement possibly results from economic development and medical improvement, and it is also not difficult to imagine that there will be gradually declining mortality rate in the near future due to technology progress.

In addition to the global mortality trend, each country still behaves differently from others to some extent. From this perspective, it might be possible for life insurance companies to design insurance products among different countries to hedge global longevity risk. That is, if wider

range of countries is covered by a particular insurance company, it is possible to redistribute longevity risk among them.

Another advantage from this research is to establish a better forecasting regime to foresee mortality change in longer time horizon, particularly for countries with limited historical mortality data, such as China and Chile.

References

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22:547–581.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science*, 3:3–43.

Cairn, Blake, and Dowd (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, 108(2-3):79–113.

Chan, W., Li, S., and Cheung, S. (2008). Testing deterministic versus stochastic trends in the lee-carter mortality indexes and its implications for projecting mortality improvements at advanced ages. *Living to 100*.

Fang, L. and Härdle, W. (2015). Stochastic population analysis: a functional data analysis. Technical report, SFB 649 Discussion Paper Series 2015.

Hanewald, K. (2011). Explaining mortality dynamics: the role of macroeconomic fluctuations and cause of death trends. *North American Actuarial Journal*, 15(2).

Hyndman and Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, 51:4942–4956.

Härdle, W. and Marron, J. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, 18:63–89.

Kneip, A. and Engel, J. (1995). Model estimation in nonlinear regression under shape invariance. *Annuals of Statistics*, 23:551–570.

Kreiss, J. and Lahiri, S. (2012). Bootstrap methods for time series. *Handbook of Statistics*, 30:3–26.

Lawton, W. H., Sylvestre, E. A., and Maggio, M. S. (1972). Self modeling nonlinear regression. *Technometrics*, 14:513–532.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting u.s. mortality. *Journal of the American Statistical Association*, 87:659–671.

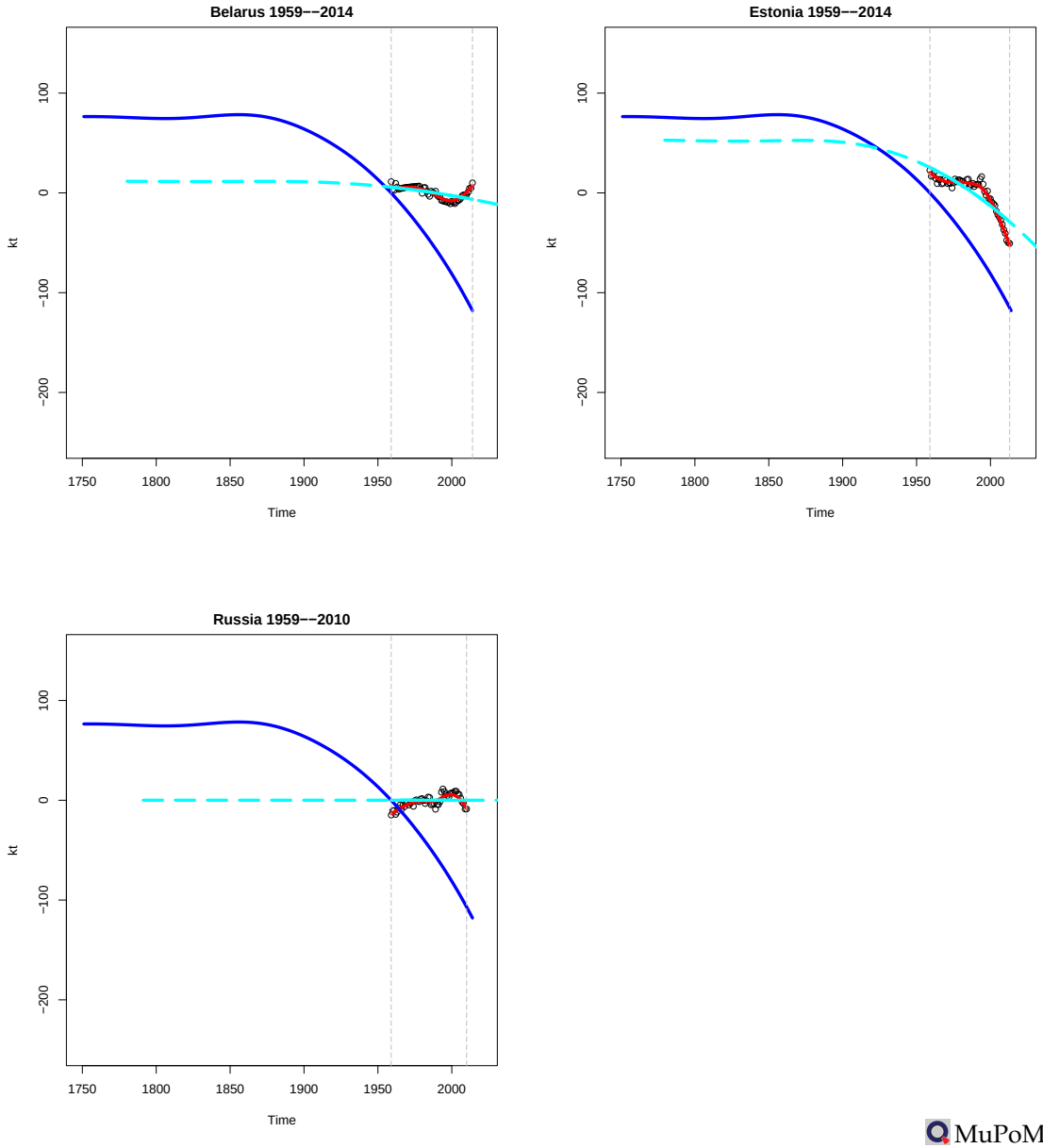
Lee, R. D. and Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, 38:537–549.

Li, N., Lee, R., and Tuljapurkar, S. (2004). Using the lee-carter method to forecast mortality for populations with limited data. *International Statistical Review*, 72:19–36.

Raftery, A., Li, N., Ševčíková, H., Gerland, P., and Heilig, G. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109:13915–13921.

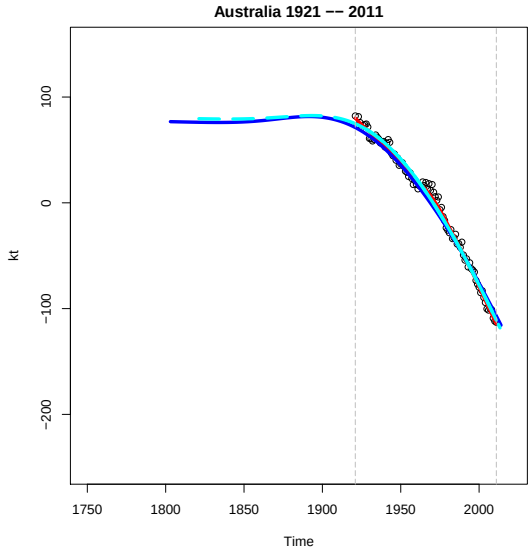
V. Appendices

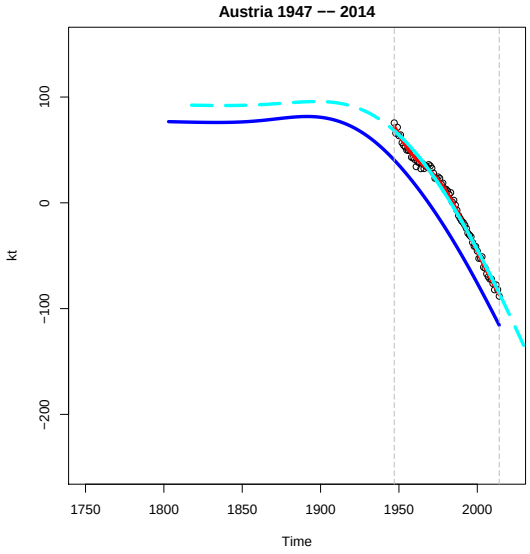
1. Appendix 1

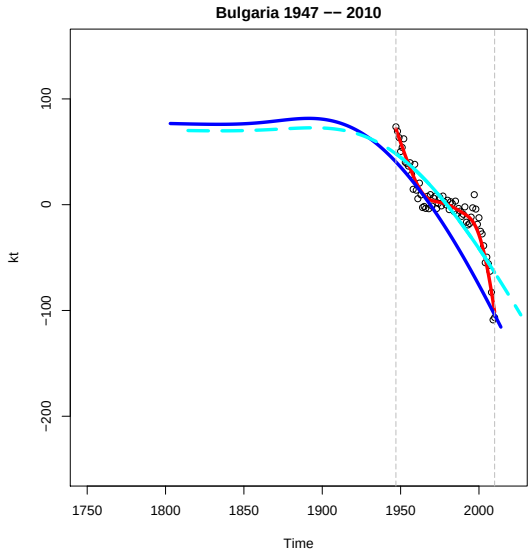


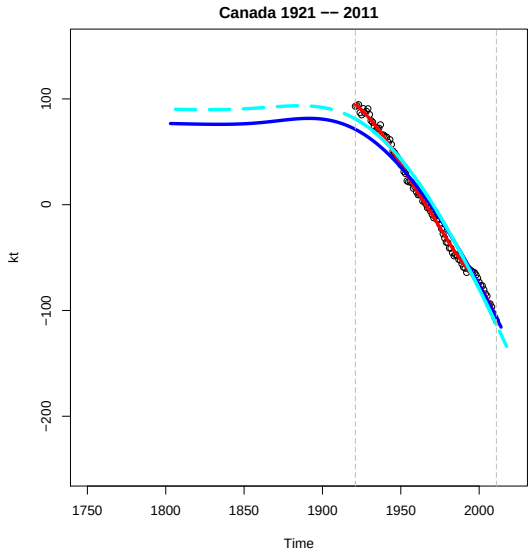
MuPoMo

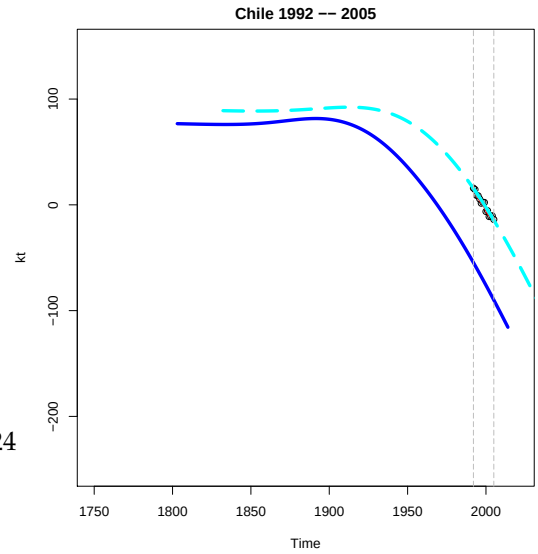
2. Appendix 2

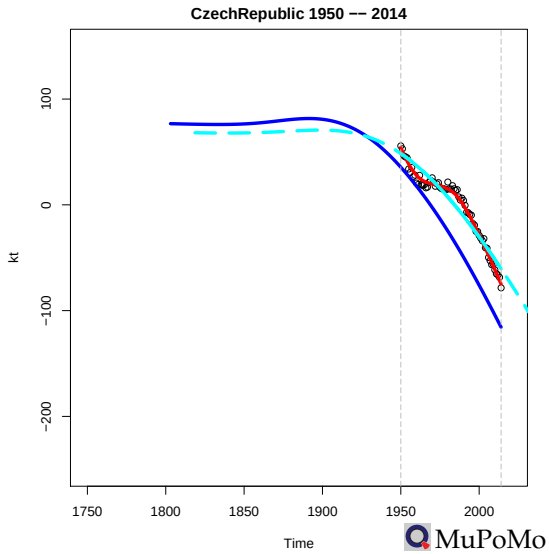

Australia 1921 −− 2011
kt
100
0
−100
−200
1750
1800
1850
1900
1950
2000
Time

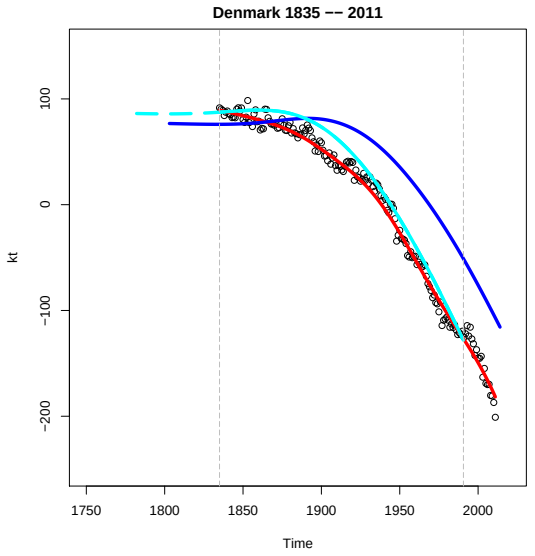

Austria 1947 −− 2014
kt
100
0
−100
−200
1750
1800
1850
1900
1950
2000
Time

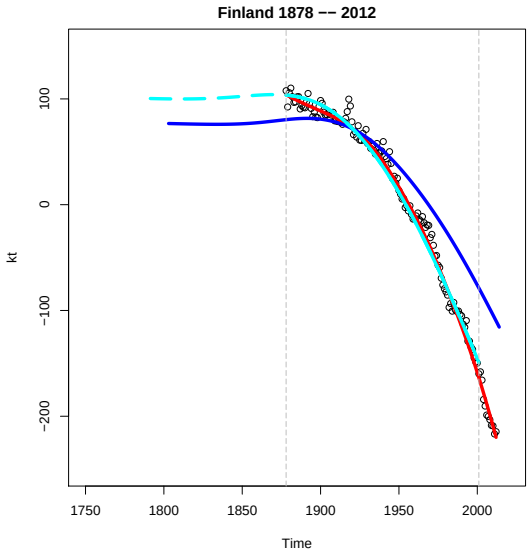

Bulgaria 1947 −− 2010
kt
100
0
−100
−200
1750
1800
1850
1900
1950
2000
Time

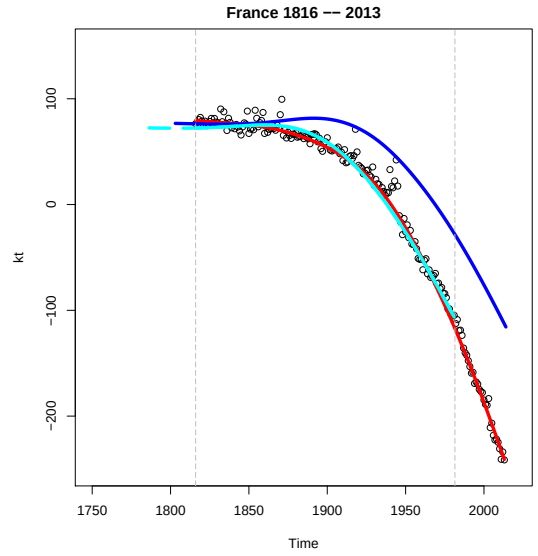

Canada 1921 −− 2011
kt
100
0
−100
−200
1750
1800
1850
1900
1950
2000
Time

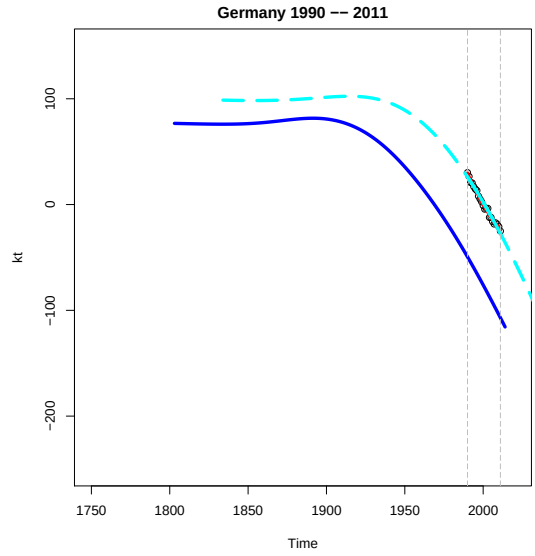

Chile 1992 −− 2005
kt
100
0
−100
−200
1750
1800
1850
1900
1950
2000
Time

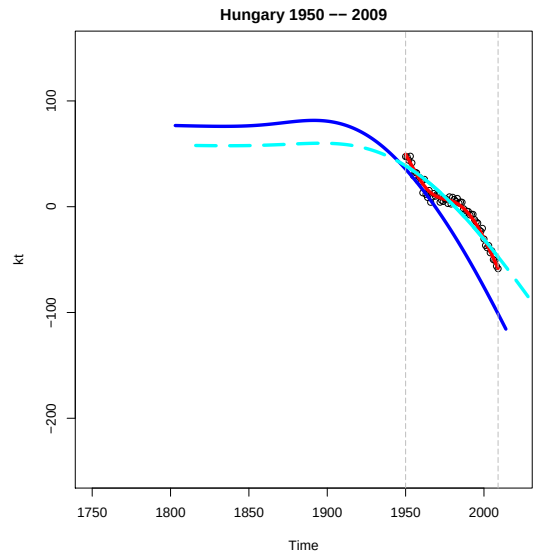

CzechRepublic 1950 −− 2014
kt
100
0
−100
−200
1750
1800
1850
1900
1950
2000
Time

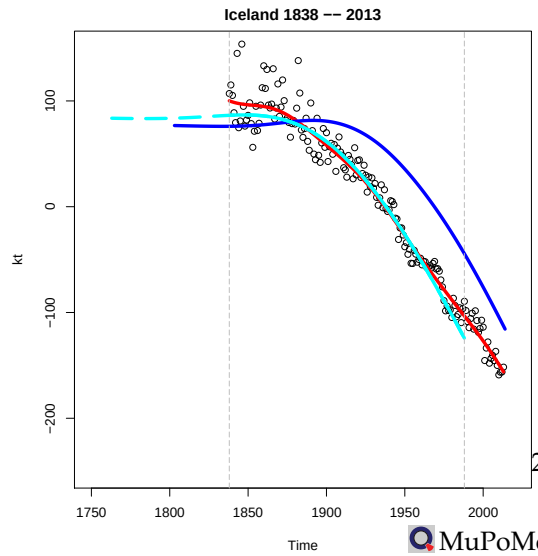

Denmark 1835 −− 2011
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

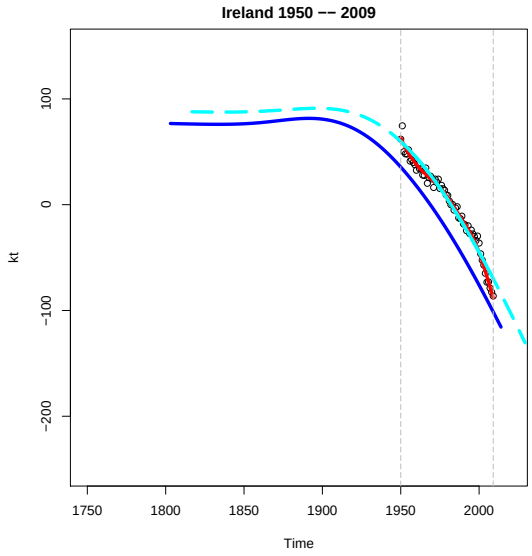

Finland 1878 −− 2012
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

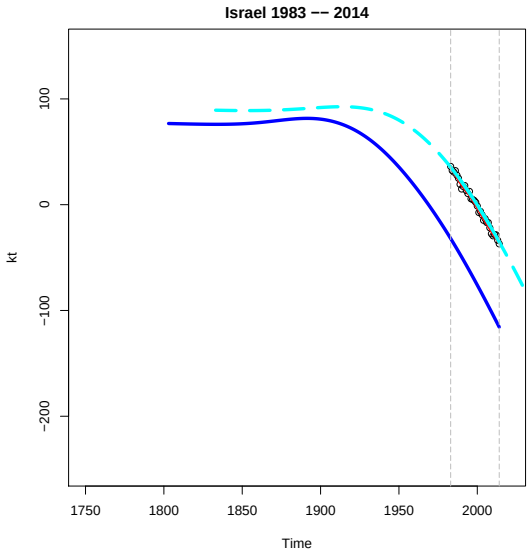

France 1816 −− 2013
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

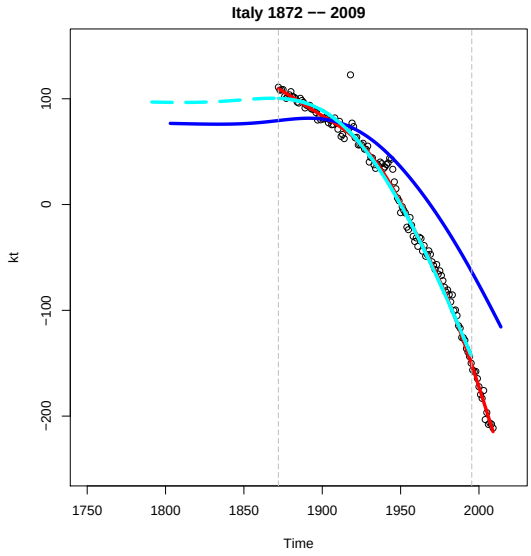

Germany 1990 −− 2011
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

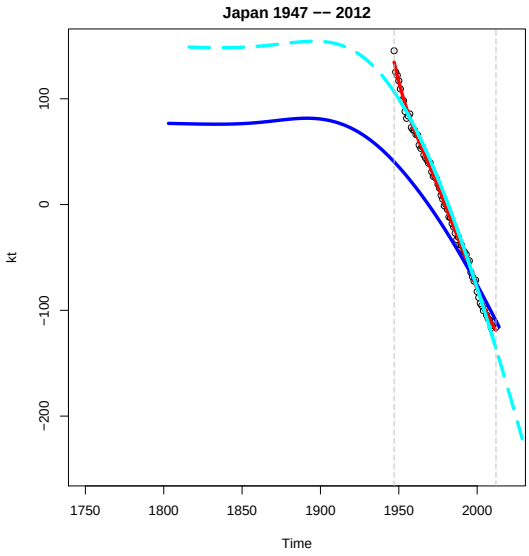

Hungary 1950 −− 2009
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

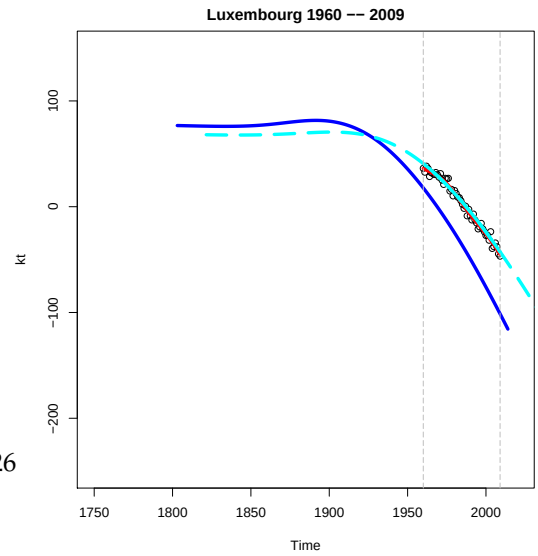

Iceland 1838 −− 2013
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

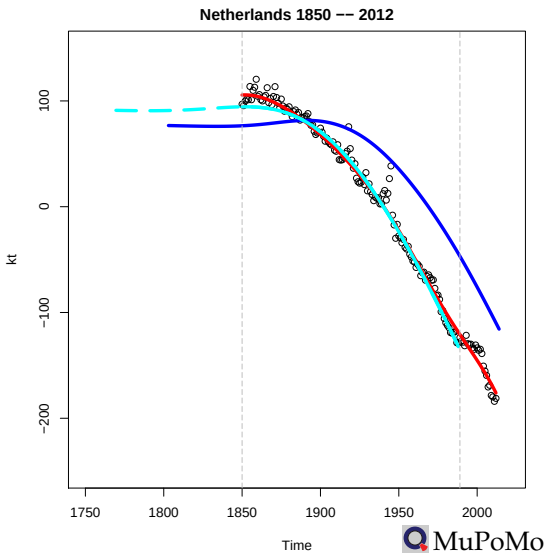

Ireland 1950 −− 2009
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

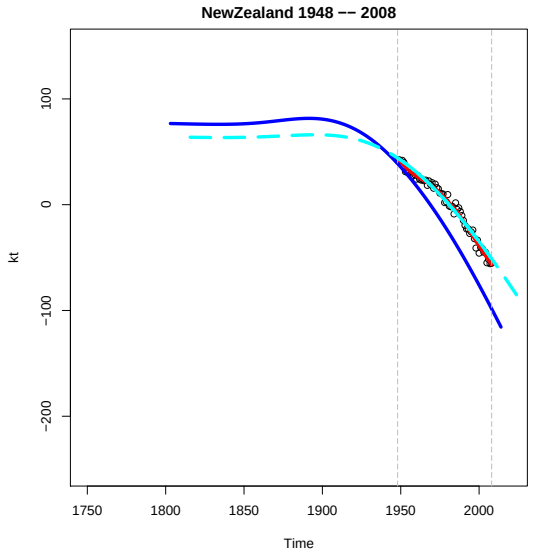

Israel 1983 −− 2014
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

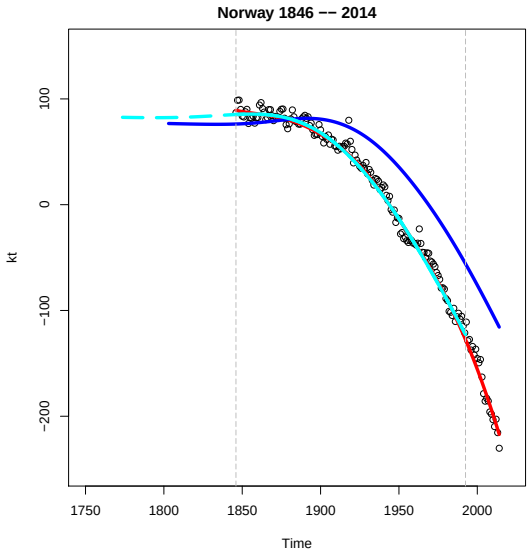

Italy 1872 −− 2009
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

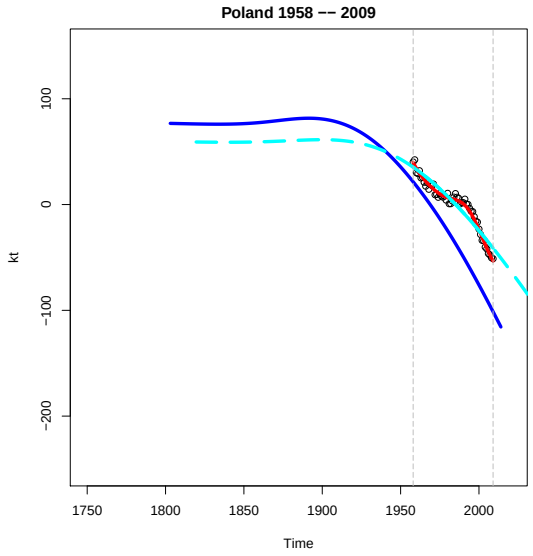

Japan 1947 −− 2012
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

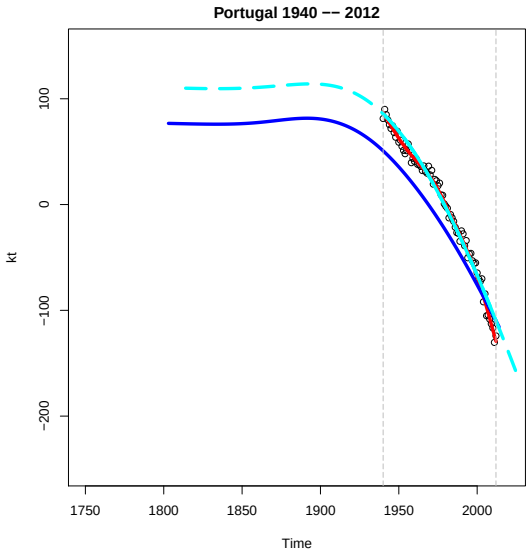

Luxembourg 1960 −− 2009
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

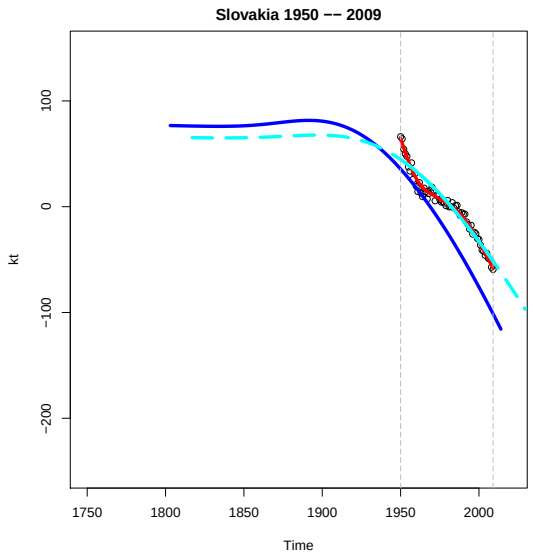

Netherlands 1850 −− 2012
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

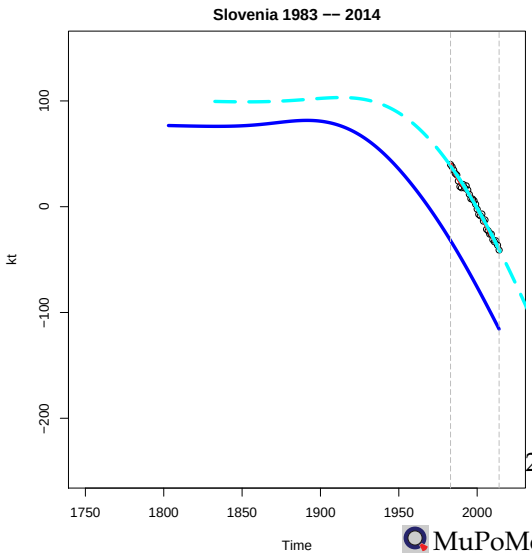

NewZealand 1948 −− 2008
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

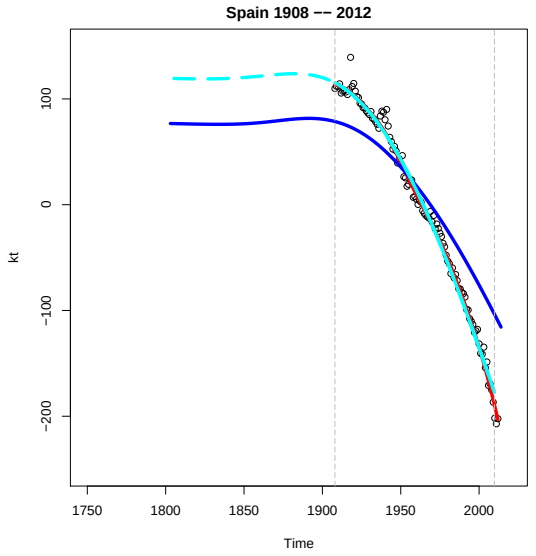

Norway 1846 −− 2014
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

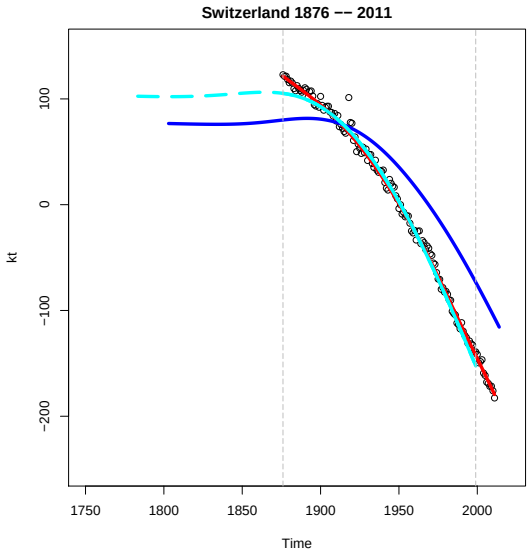

Poland 1958 −− 2009
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

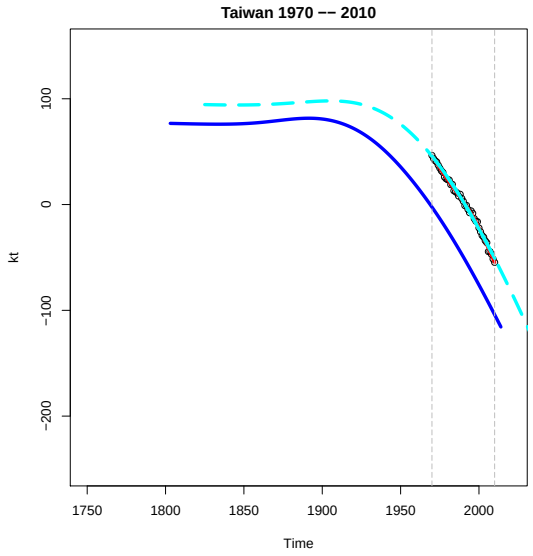

Portugal 1940 −− 2012
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

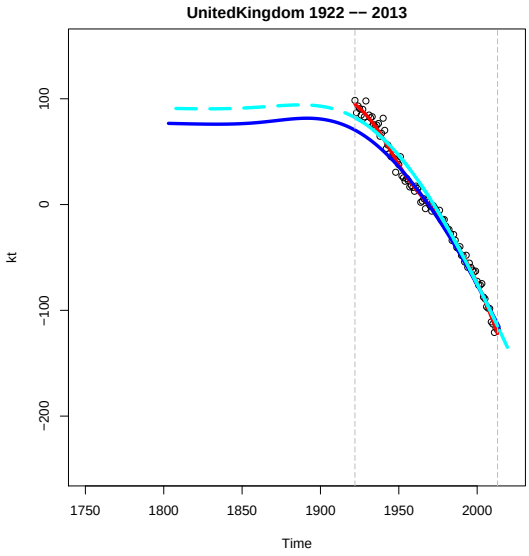

Slovakia 1950 −− 2009
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

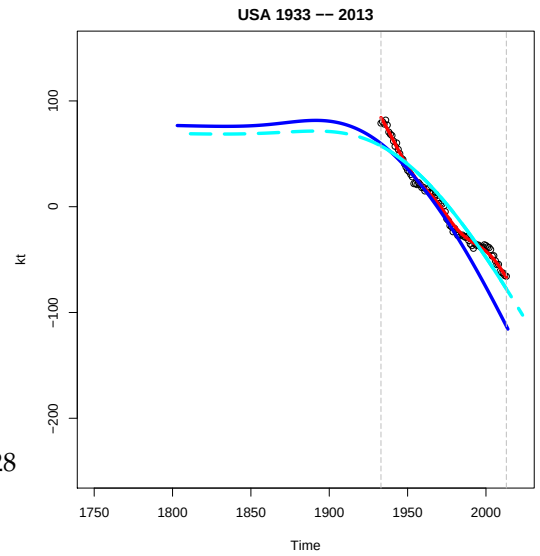

Slovenia 1983 −− 2014
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

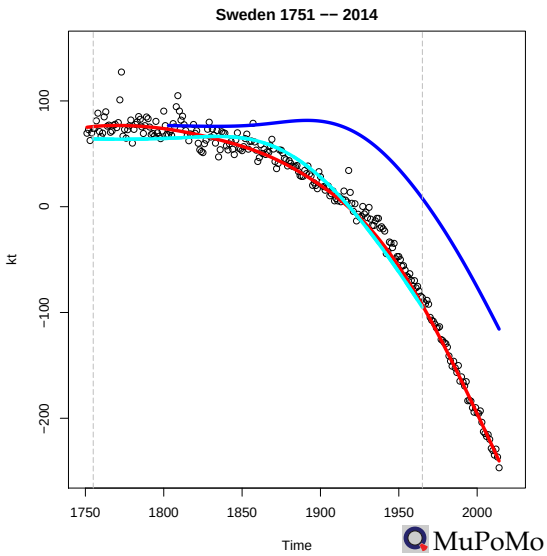

Spain 1908 −− 2012
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000


Switzerland 1876 −− 2011
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000


Taiwan 1970 −− 2010
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000


UnitedKingdom 1922 −− 2013
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000


USA 1933 −− 2013
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000


Sweden 1751 −− 2014
kt
Time
100
0
−100
−200
1750
1800
1850
1900
1950
2000

SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Downside risk and stock returns: An empirical analysis of the long-run and short-run dynamics from the G-7 Countries" by Cathy Yi-Hsuan Chen, Thomas C. Chiang and Wolfgang Karl Härdle, January 2016.
002 "Uncertainty and Employment Dynamics in the Euro Area and the US" by Aleksei Netsunajev and Katharina Glass, January 2016.
003 "College Admissions with Entrance Exams: Centralized versus Decentralized" by Isa E. Hafalir, Rustamdjan Hakimov, Dorothea Kübler and Morimitsu Kurino, January 2016.
004 "Leveraged ETF options implied volatility paradox: a statistical study" by Wolfgang Karl Härdle, Sergey Nasekin and Zhiwu Hong, February 2016.
005 "The German Labor Market Miracle, 2003 -2015: An Assessment" by Michael C. Burda, February 2016.
006 "What Derives the Bond Portfolio Value-at-Risk: Information Roles of Macroeconomic and Financial Stress Factors" by Anthony H. Tu and Cathy Yi-Hsuan Chen, February 2016.
007 "Budget-neutral fiscal rules targeting inflation differentials" by Maren Brede, February 2016.
008 "Measuring the benefit from reducing income inequality in terms of GDP" by Simon Voigts, February 2016.
009 "Solving DSGE Portfolio Choice Models with Asymmetric Countries" by Grzegorz R. Dlugoszek, February 2016.
010 "No Role for the Hartz Reforms? Demand and Supply Factors in the German Labor Market, 1993-2014" by Michael C. Burda and Stefanie Seele, February 2016.
011 "Cognitive Load Increases Risk Aversion" by Holger Gerhardt, Guido P. Biele, Hauke R. Heekeren, and Harald Uhlig, March 2016.
012 "Neighborhood Effects in Wind Farm Performance: An Econometric Approach" by Matthias Ritter, Simone Pieralli and Martin Odening, March 2016.
013 "The importance of time-varying parameters in new Keynesian models with zero lower bound" by Julien Albertini and Hong Lan, March 2016.
014 "Aggregate Employment, Job Polarization and Inequalities: A Transatlantic Perspective" by Julien Albertini and Jean Olivier Hairault, March 2016.
015 "The Anchoring of Inflation Expectations in the Short and in the Long Run" by Dieter Nautz, Aleksei Netsunajev and Till Strohsal, March 2016.
016 "Irrational Exuberance and Herding in Financial Markets" by Christopher Boortz, March 2016.
017 "Calculating Joint Confidence Bands for Impulse Response Functions using Highest Density Regions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, March 2016.
018 "Factorisable Sparse Tail Event Curves with Expectiles" by Wolfgang K. Härdle, Chen Huang and Shih-Kang Chao, March 2016.
019 "International dynamics of inflation expectations" by Aleksei Netšunajev and Lars Winkelmann, May 2016.
020 "Academic Ranking Scales in Economics: Prediction and Imdputation" by Alona Zharova, Andrija Mihoci and Wolfgang Karl Härdle, May 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

021 "CRIX or evaluating blockchain based currencies" by Simon Trimborn and Wolfgang Karl Härdle, May 2016.

022 "Towards a national indicator for urban green space provision and environmental inequalities in Germany: Method and findings" by Henry Wüstemann, Dennis Kalisch, June 2016.

023 "A Mortality Model for Multi-populations: A Semi-Parametric Approach" by Lei Fang, Wolfgang K. Härdle and Juhyun Park, June 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".