Bootstrap for mortality projections on dependent data

V. D'Amato[1], S. Haberman[2],G. Piscopo[3], M. Russolillo[1]

***1** Department of Economics and Statistics, University of Salerno – Italy -e-mail: vdamato@unisa.it, mrussolillo@unisa.it*

***2** Faculty of Actuarial Science and Insurance, Cass Business School, City University, London – UK - e-mail: s.haberman@city.ac.uk*

***3** Department of Mathematics for Decisions, University of Florence, Italy - e-mail: gabriella.piscopo@unifi.it*

AGENDA

- The motivation
- The Lee Carter Model
- The Lee Carter Sieve Bootstrap
- Numerical Applications

Motivation

Because of the nonlinear nature of the quantities of interest, such as life expectancy, annuity premiums and so on, an analytic approach to the calculation of prediction intervals is intractable, so that it is necessary to resort to a *simulation* approach.

Motivation

The presence of *dependence* across time leads to systematic over-estimation or under-estimation of uncertainty in the mortality estimates, caused by whether negative or positive dependence dominates.

Motivation

The correlation structure between the residuals has to be tackled. Otherwise prediction intervals for projections underestimate the actual longevity risk.

In other words, it is necessary to assess a significant and further source of risk: a sort of *dependency risk*.

The Lee Carter model

Lee and Carter (1992) suggested a log-bilinear form for the force of mortality:

$$m_{xt} = \exp(\alpha_x + \beta_x k_t + u_{xt})$$

$$\ln(m_{xt}) = \alpha_x + \beta_x k_t + u_{xt}$$

$$\sum_t k_t = 0 \qquad \sum_x \beta_x = 1$$

The LC Sieve Bootstrap

In the literature, there is more than one **bootstrap** method for **dependent data** as for example block, local, wild, Markov bootstrap, sub-sampling and **sieve**.

Choi and Hall (2000) show that the sieve bootstrap has substantial advantages over blocking methods, to such an extent that block–based methods are not really competitive. In particular, other authors show that the sieve bootstrap outperforms the block bootstrap (Hardle et al. 2003).

The LC Sieve bootstrap

Notation:

u_{xt} error term

ε_{xt} innovation term

r_{xt} estimated innovation or residual

$\bar{r}_{xt}$ mean value of the residuals

$r_{xt} - \bar{r}_{xt}$ centred residuals

$\hat{F}_r$ ecdf of residuals

u^*_{xt} bootstrap error ε^*_{xt} IID term from $\hat{F}_r$

The LC Sieve Bootstrap

The Scheme:

The error term is approximated by an $AR(\infty)$ representation:

$$u_{xt} = \sum_{j=1}^{\infty} \varphi_j u_{xt-j} + \varepsilon_{xt} \qquad x = 1, 2, \ldots, m$$

The LC Sieve Bootstrap

The Steps:

1. *<u>Fit the model and obtain the OLS estimates</u>* :

$$\hat{u}_{xt} = \sum_{j=1}^{\hat{p}(n)} \varphi_j \hat{u}_{xt-j} + \varepsilon_{xt} \qquad x = 1, 2, \ldots, m$$

The Lc Sieve Bootstrap

The Steps:

2. ***Specify the lag length*** $\hat{p}(n)$
   by BIC, AIC, etc

The LC Sieve Bootstrap

The Steps:

3. <u>Calculate the autoregressive coefficients</u> by the Ordinary Least Squares or by using the Yule-Walker method

$$\hat{\phi}_j, \qquad j = 1, \ldots, \hat{p}(n)$$

The LC Sieve Bootstrap

The Steps:

4. <u>Calculate the residuals (or estimated innovations)</u> associated with $\hat{\phi}_j$ according the following formula:

$$r_{xt} = \hat{u}_{xt} - \sum_{j=1}^{\hat{p}(n)} \hat{\phi}_j \hat{u}_{x\,t-j} \qquad x = 1,2,\ldots,m$$

$$t = \hat{p}(n)+1,\ldots,n$$

The LC Sieve Bootstrap

The Steps:

5. ***<u>Calculate the centred residuals</u>***

$$\tilde{r}_{xt} = r_{xt} - \bar{r}_{xt}$$

The LC Sieve Bootstrap

The Steps:

6. <u>Define the empirical distribution function</u> of the centred residuals

$$\hat{F}_{xr}(y) = \frac{1}{n-p} \sum_{t=p+1}^{n} 1_{\{\tilde{r}_{xt} \leq y\}}$$

The LC Sieve Bootstrap

The Steps:

7. <u>Draw</u> ε^{*}_{xt} ***<u>IID terms from</u>*** $\hat{\mathrm{F}}_{x\,r}$ ***<u>with replacement</u>***

The LC Sieve Bootstrap

The Steps:

8. <u>Bootstrap</u> u^{*}_{xt} ***<u>are simulated</u>*** by recursion according to the bootstrap regression model:

$$u^{*}_{xt} = \sum_{j=1}^{\hat{p}(n)} \hat{\varphi}_{j} u^{*}_{xt-y} + \varepsilon^{*}_{xt} \qquad x = 1, 2, \ldots, m$$

The LC Sieve Bootstrap

Summary:

In other words, the values of ε^{*}_{xt} are obtained by randomly sampling with replacement from $\hat{F}_{xr}$ and consequently the simulated u^{*}_{xt} are computed and the m^{*}_{xt} are mapped. Finally the estimates $\hat{\alpha}^{*}_{x}$, $\hat{\beta}^{*}_{x}$, $\hat{\kappa}^{*}_{t}$ are obtained by fitting the log-bilinear structure to the m^{*}_{xt}

Numerical Application

Application scheme:

-Model Fitting

-Analysis of residuals

-Simulation algorithm

-Comparison of the results

Numerical Application

Dataset:

The population data is composed by the Italian male from 1980 up to 2006 from 0 up to 100 years, collected from Human Mortality Database. The death rates above age 100 have been aggregated in an open age group 100+.

Numerical Application

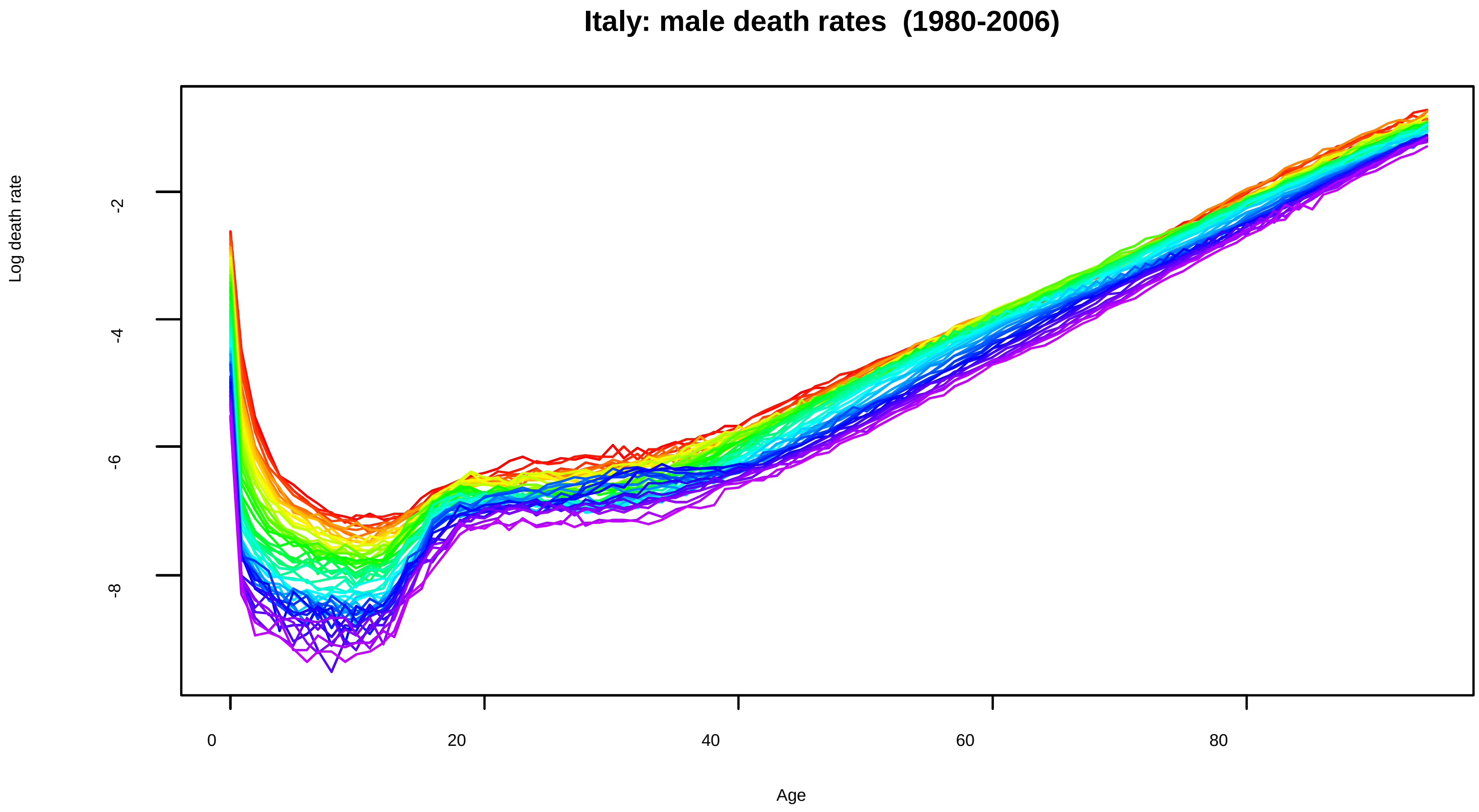

Figure 1- log death rates - Italian male population, age from 0 to 100

Numerical Application

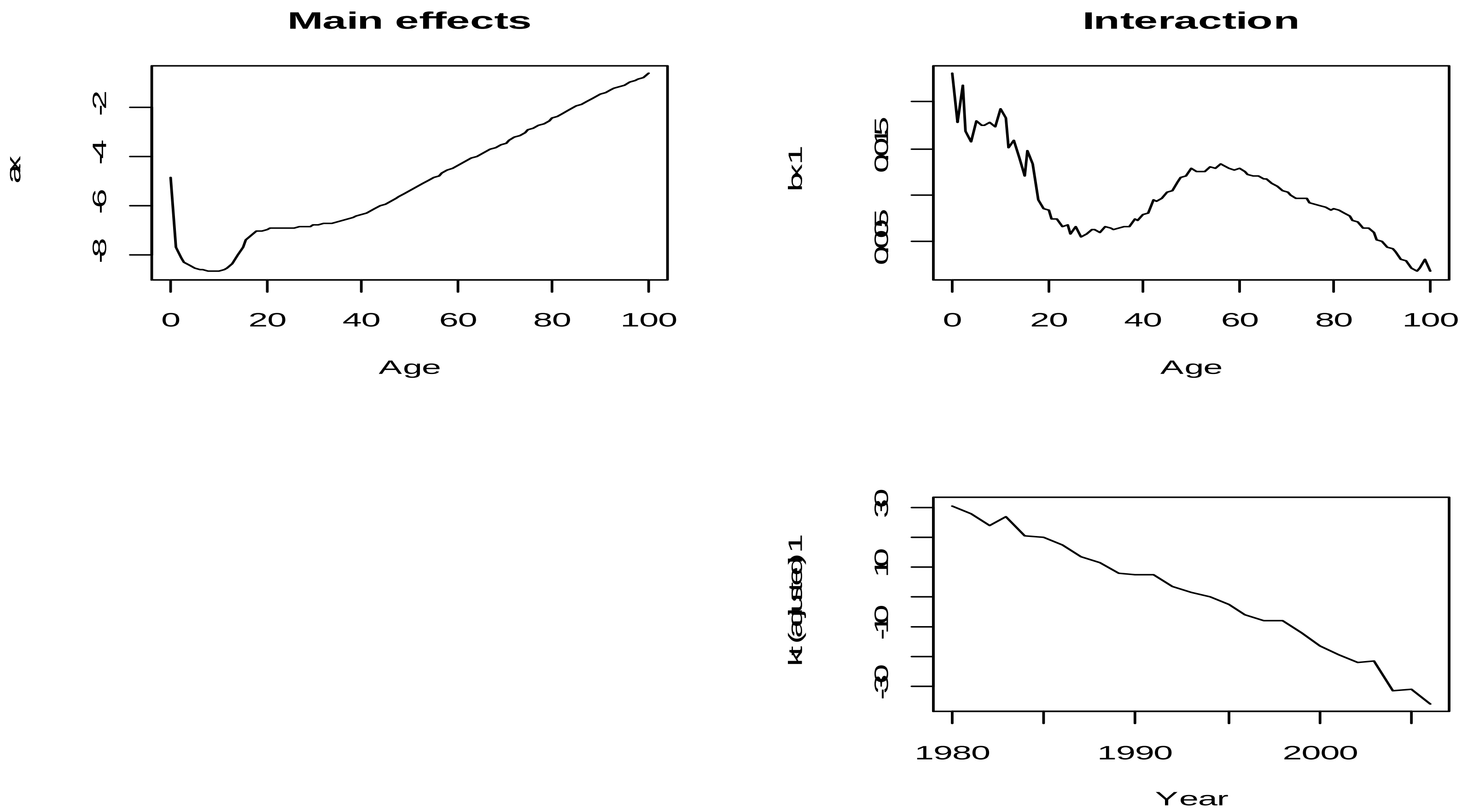

Figure 2- ax, bx, kt, basic LC model - Italian male population, age from 0 to 100

Numerical Application

ERROR MEASURES BASED ON MORTALITY RATES			
Averages across ages:			
ME Mean error	MSE Mean Squared Error	MPE Mean Percentage Error	MAPE Mean Absolute Percentage Error
-0.00008	0.00028	0.01102	0.08361
Averages across years:			
IE Integrated Error	ISE Integrated Squared Error	IPE Integrated Percentage Error	IAPE Integrated Absolute Percentage Error
-0.00455	0.01930	1.10273	8.19838

Numerical Application

ERROR MEASURES BASED ON LOG MORTALITY RATES			
Averages across ages:			
ME **Mean error**	**MSE** **Mean Squared Error**	**MPE** **Mean Percentage Error**	**MAPE** **Mean Absolute Percentage Error**
0.00367	**0.01487**	**-0.01132**	**0.03596**
Averages across years:			
IE **Integrated Error**	**ISE** **Integrated Squared Error**	**IPE** **Integrated Percentage Error**	**IAPE** **Integrated Absolute Percentage Error**
0.36583	**1.43059**	**-1.00498**	**3.40482**

Numerical Application

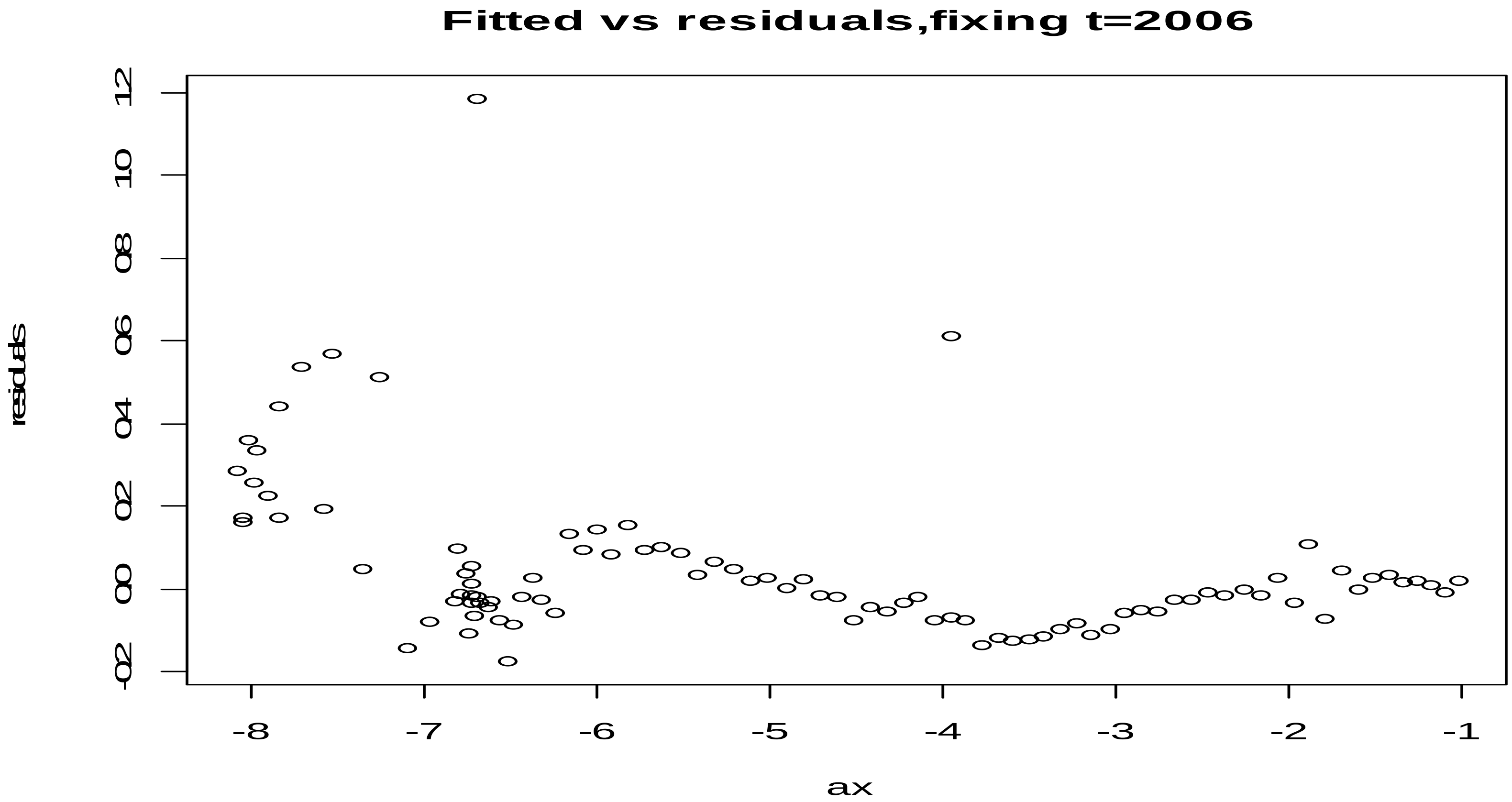

Figure 3 - Fitted ax vs residuals

Numerical Application

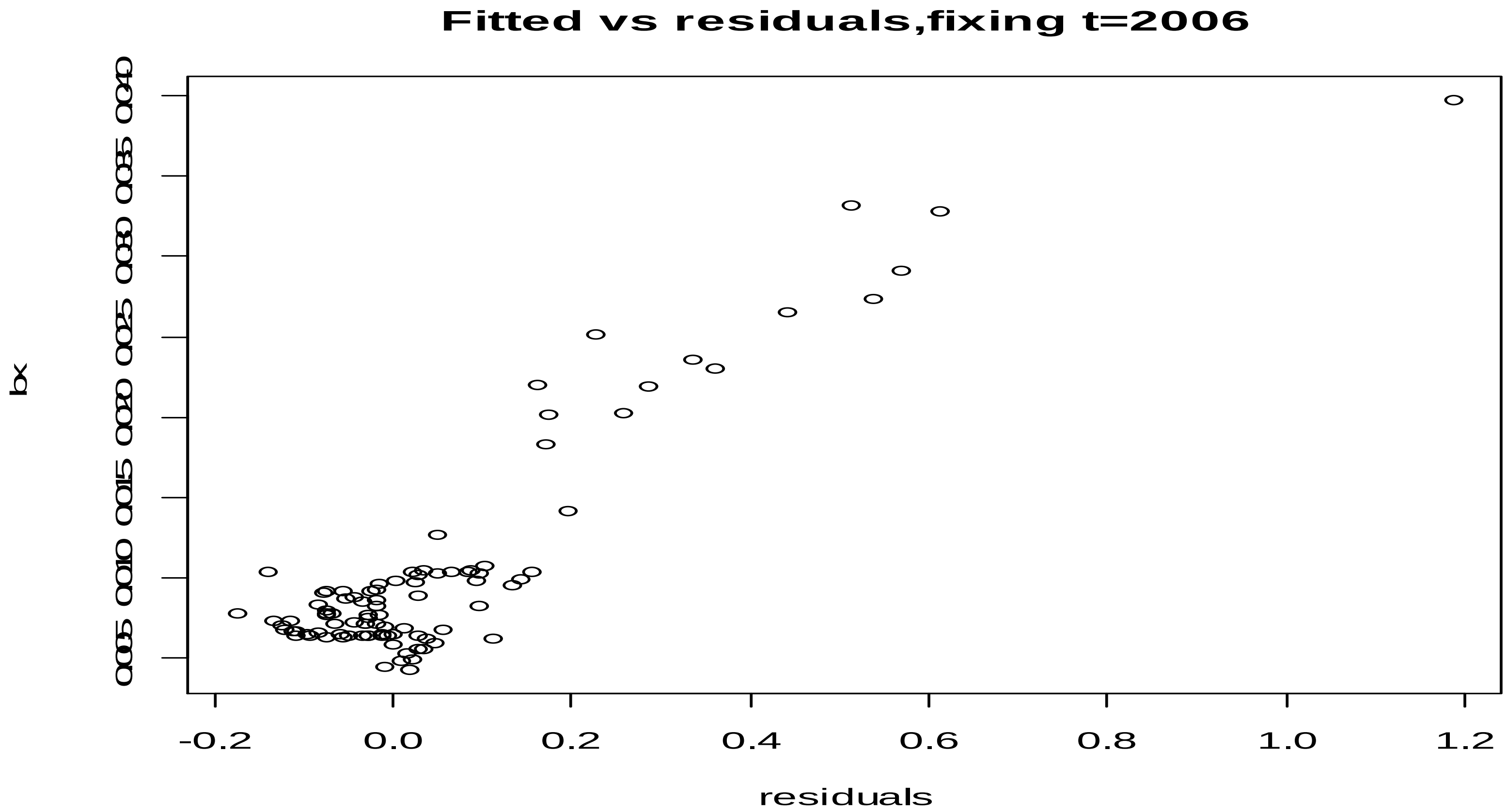

Figure 4 - Fitted bx vs residuals

Numerical Application

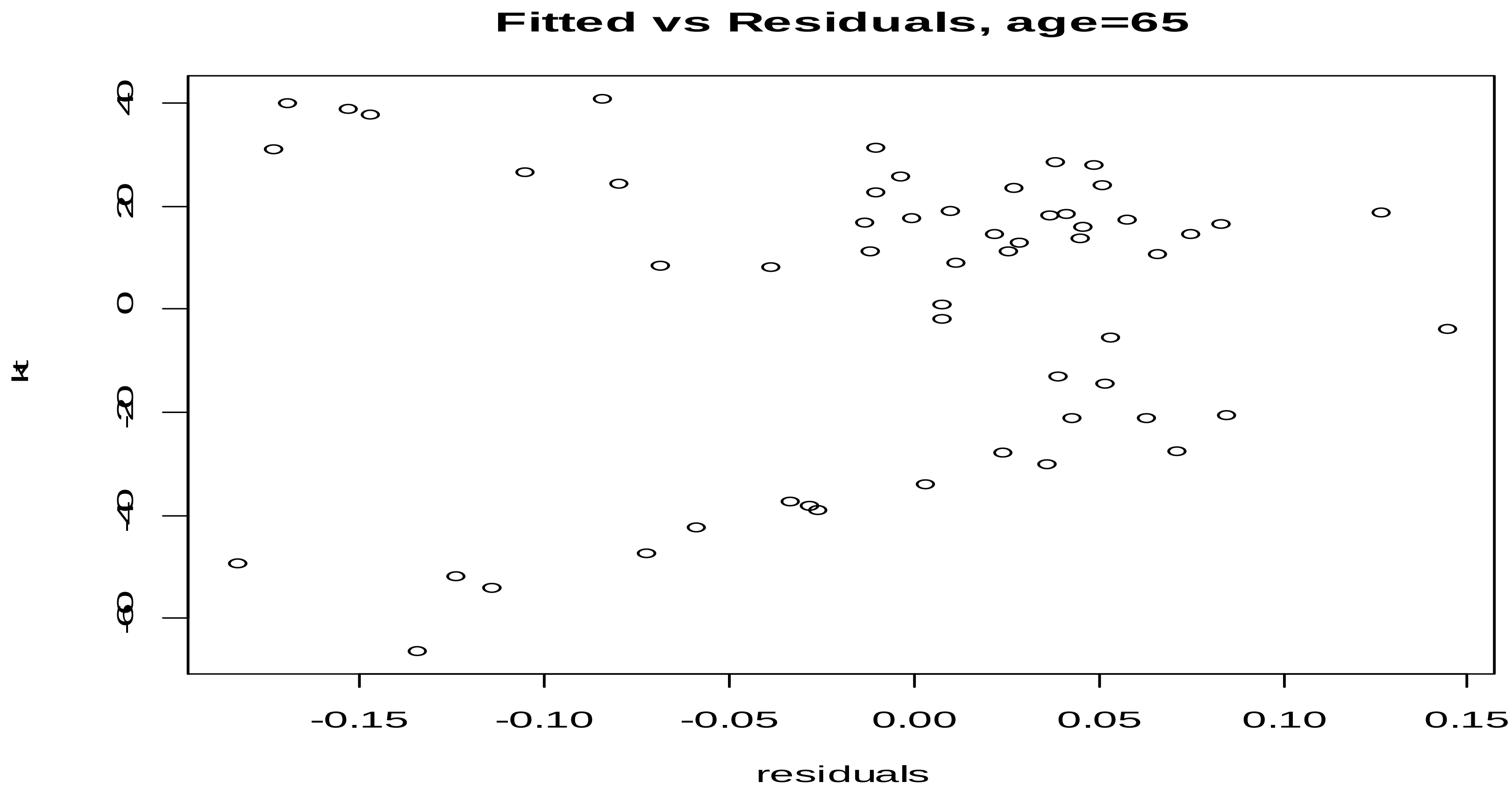

Figure 5 - Fitted kt vs residuals

Numerical Application

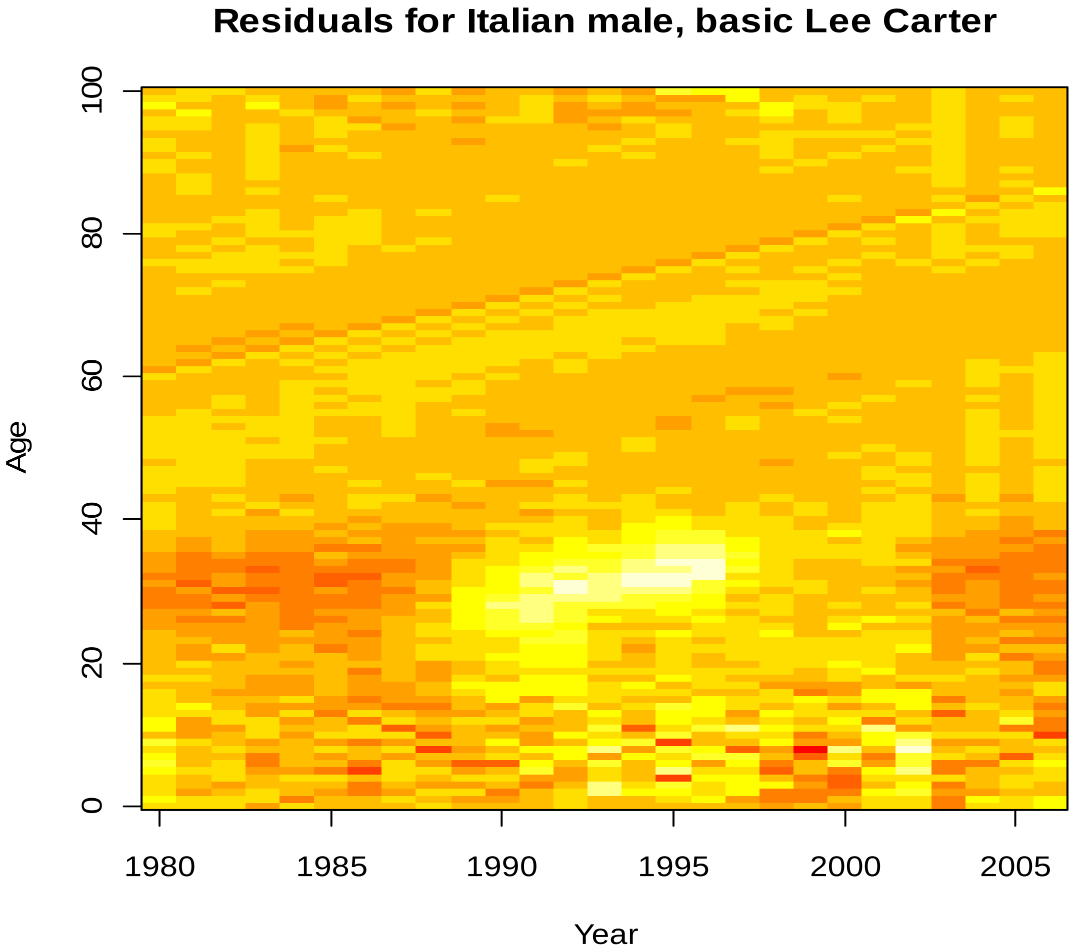

Figure 6 – Residuals years vs age – basic LC on Italian (

Numerical Application

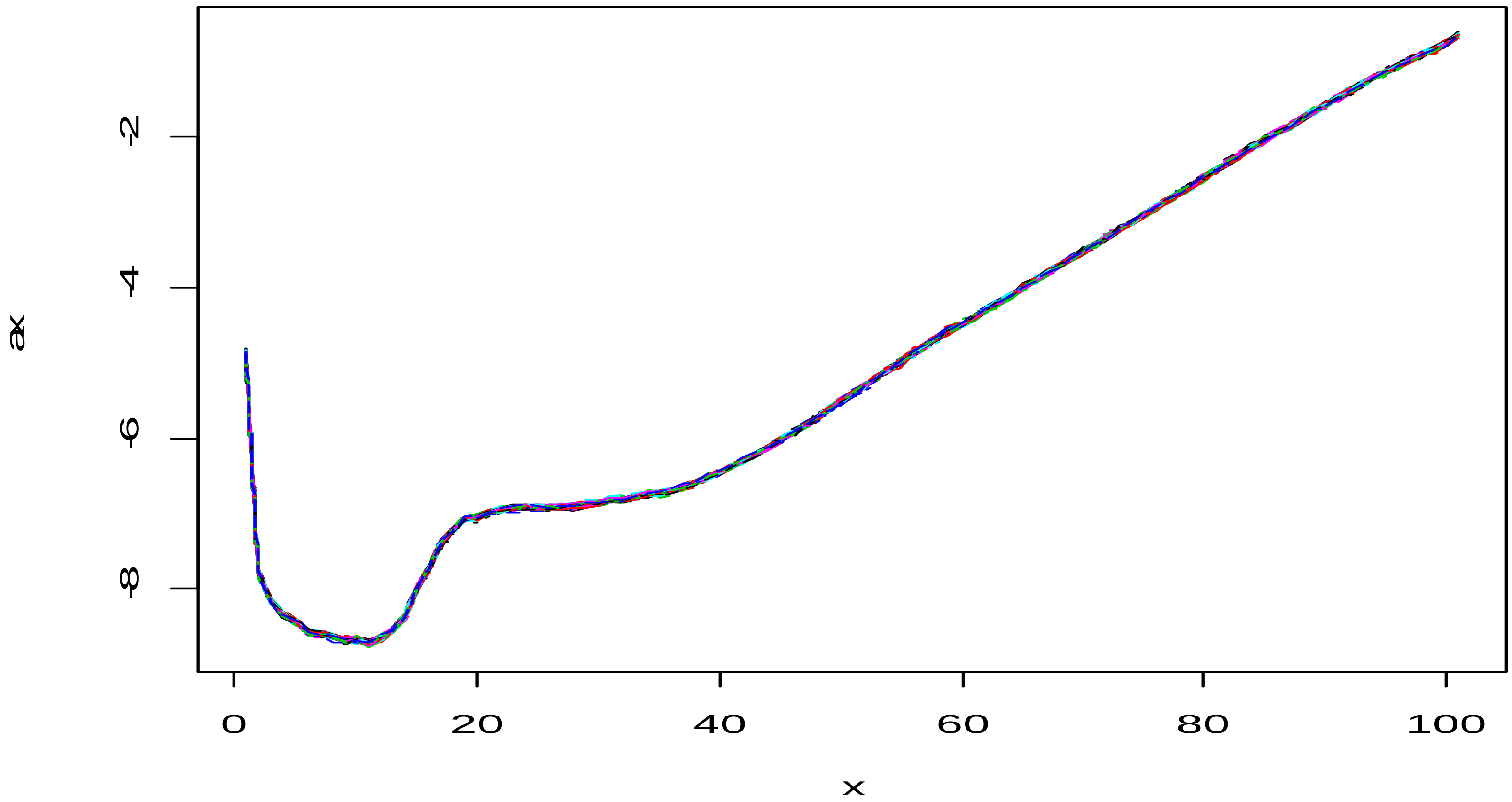

Figure 7 –Paths for ax – Sieve Bootstrap

Numerical Application

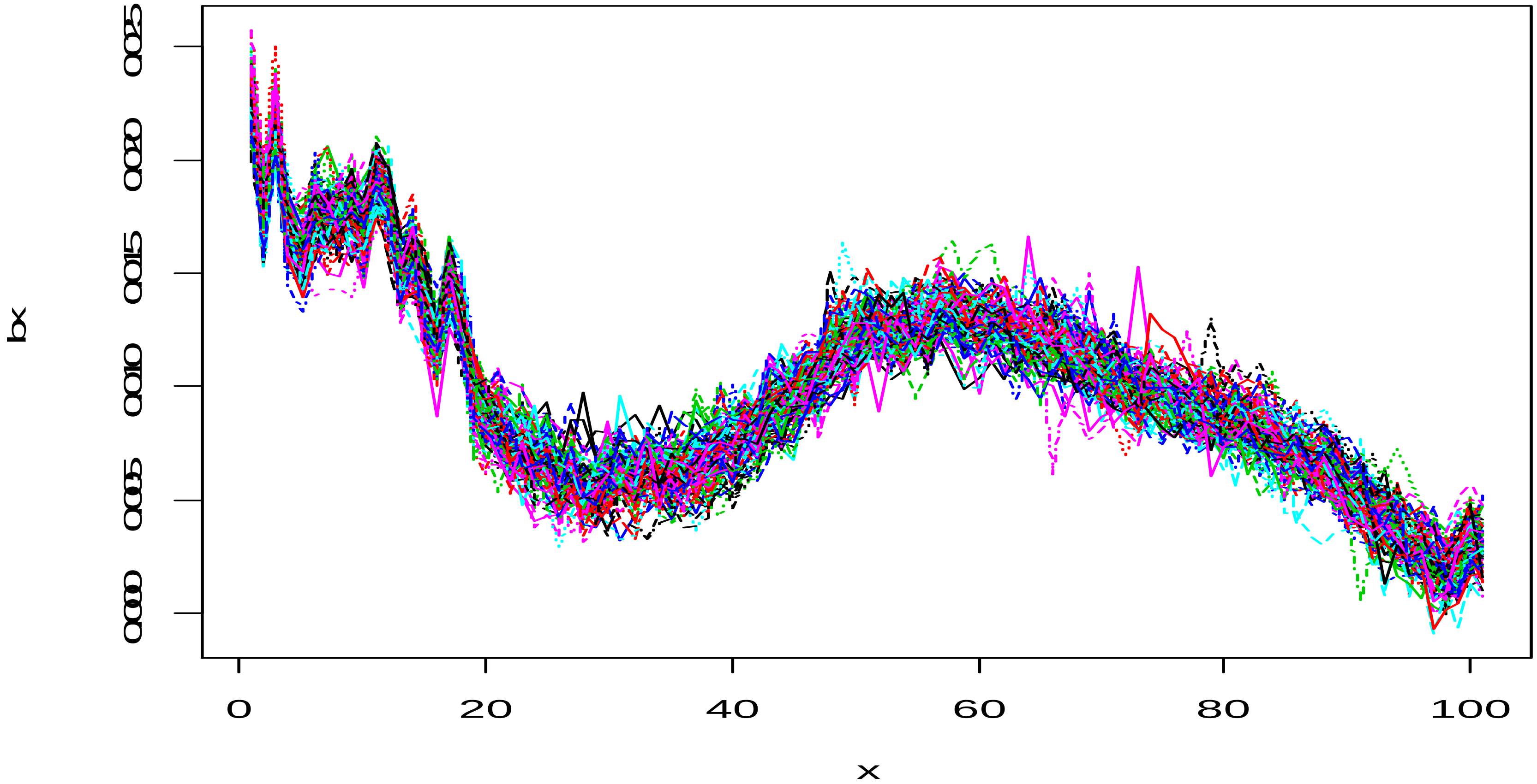

Figure 8 - Simulated paths for bx – Sieve Bootstrap

Numerical Application

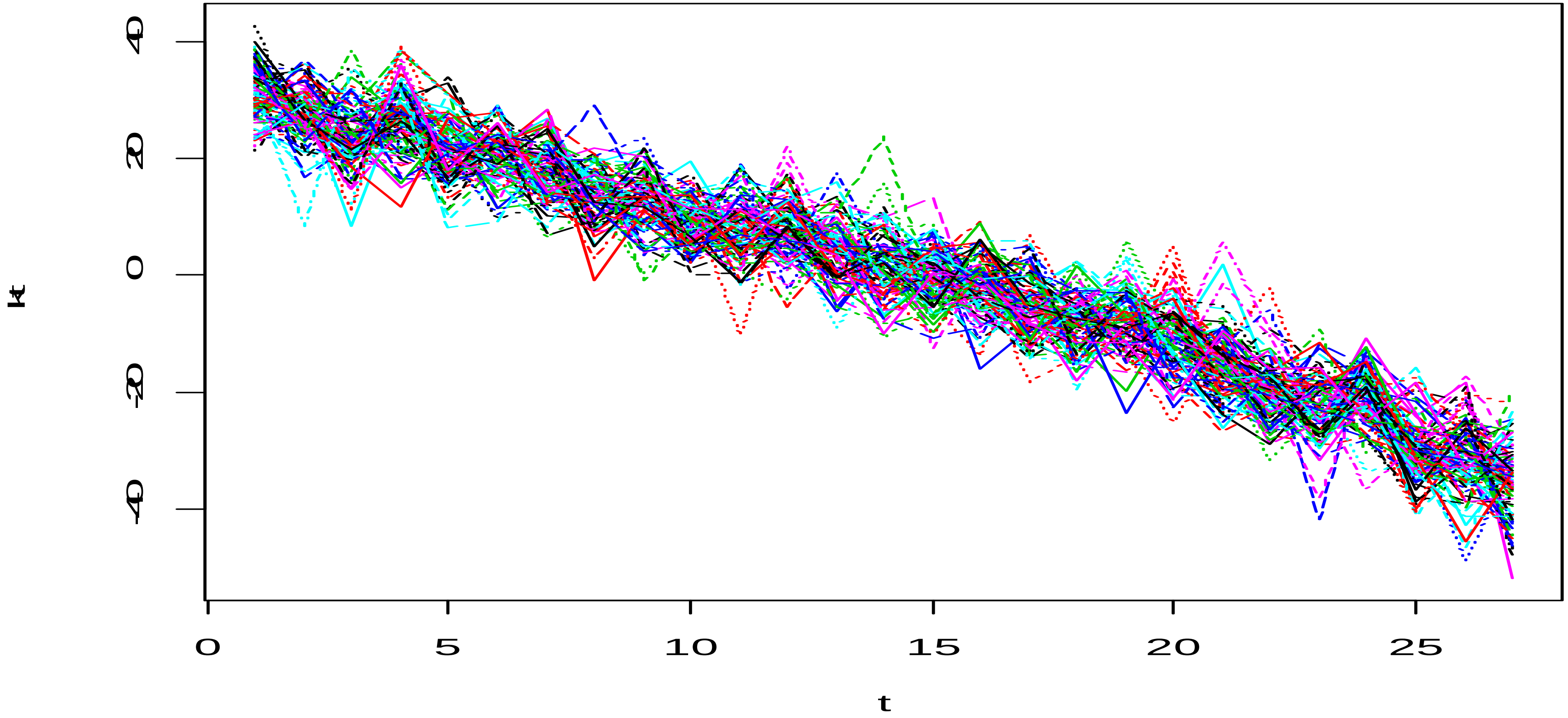

Figure 9 - Simulated paths for kt – Sieve Bootstrap

Numerical Application

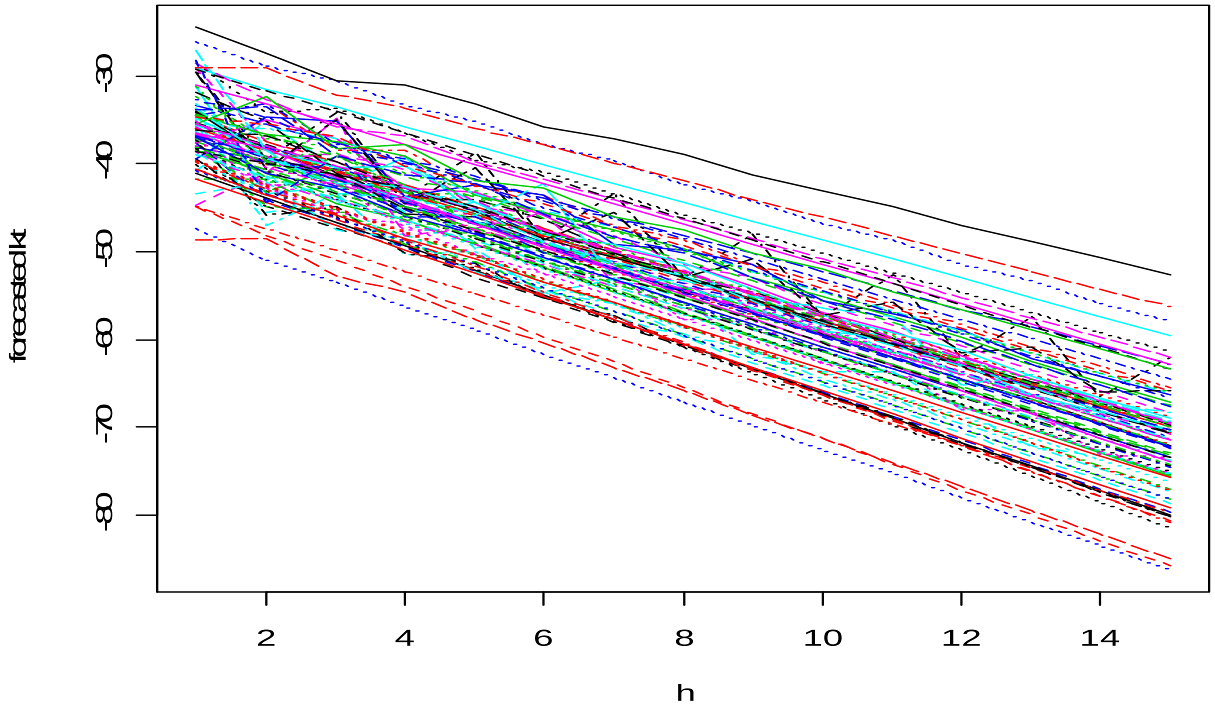

Figure 10 - Forecasted kt – Sieve Bootstrap

Numerical Application

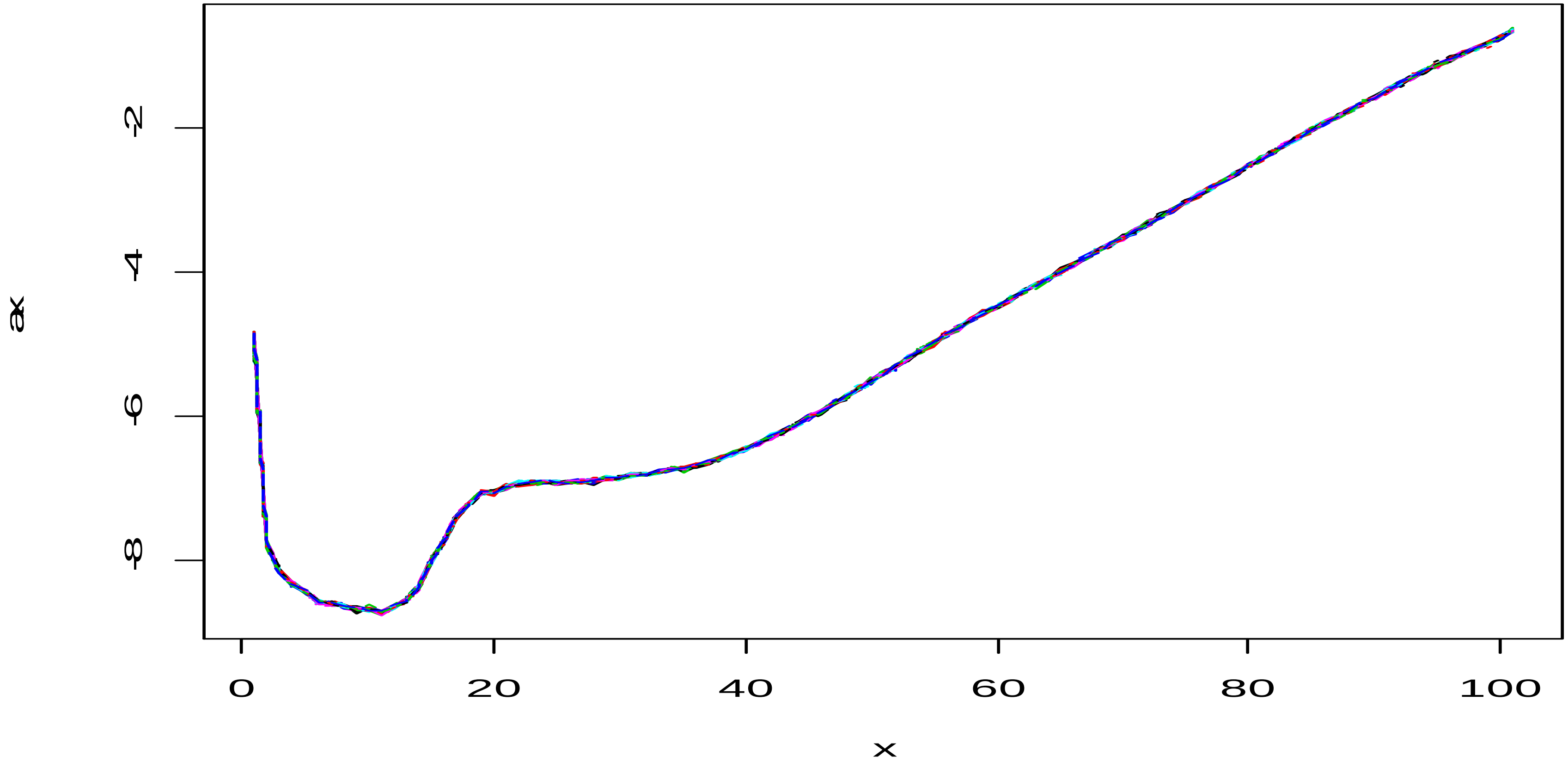

Figure 11 – Paths for ax – Residual standard bootstrap

Numerical Application

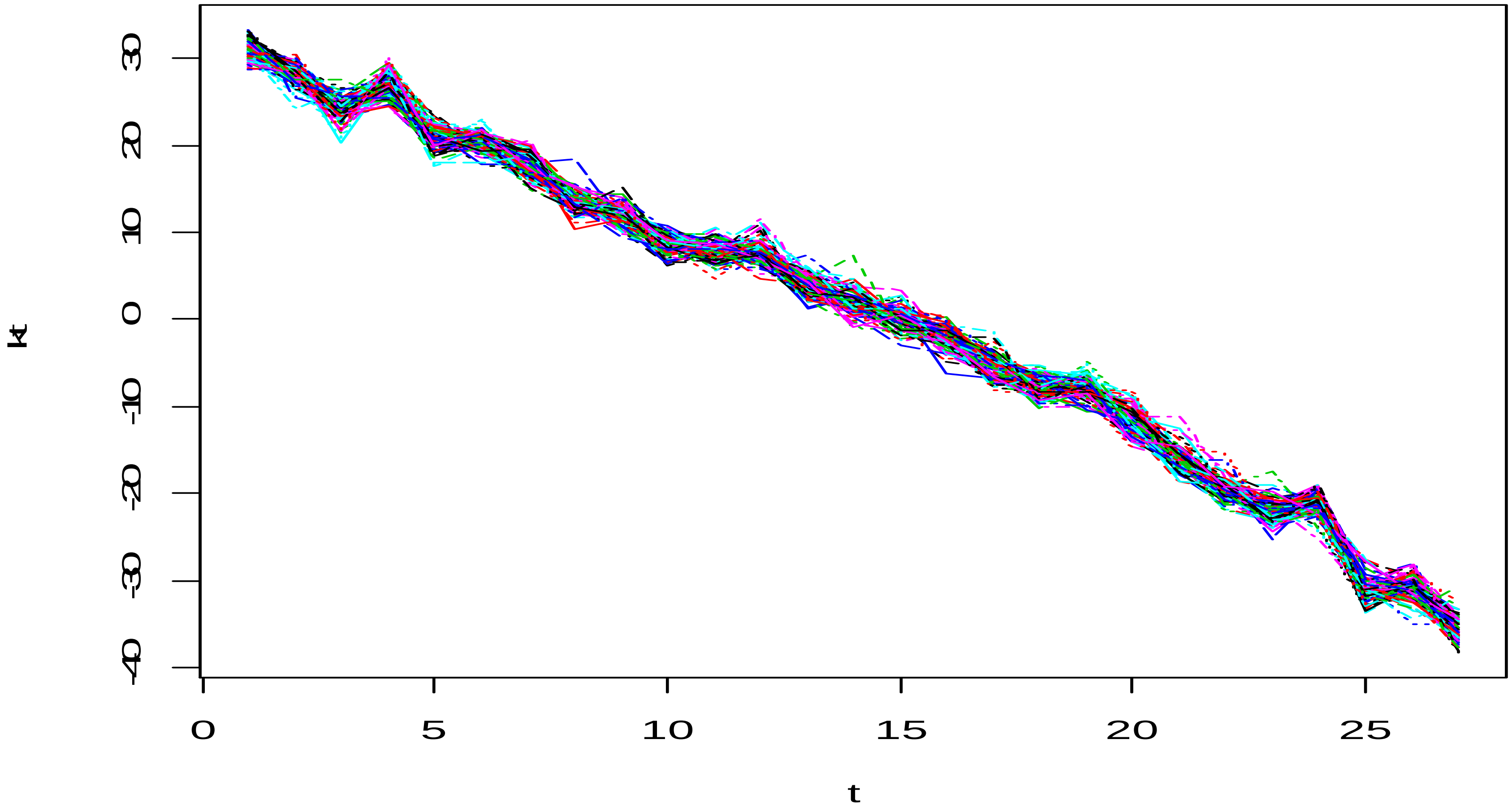

Figure 12 - Simulated paths for bx – Residual standard bootstrap

Numerical Application

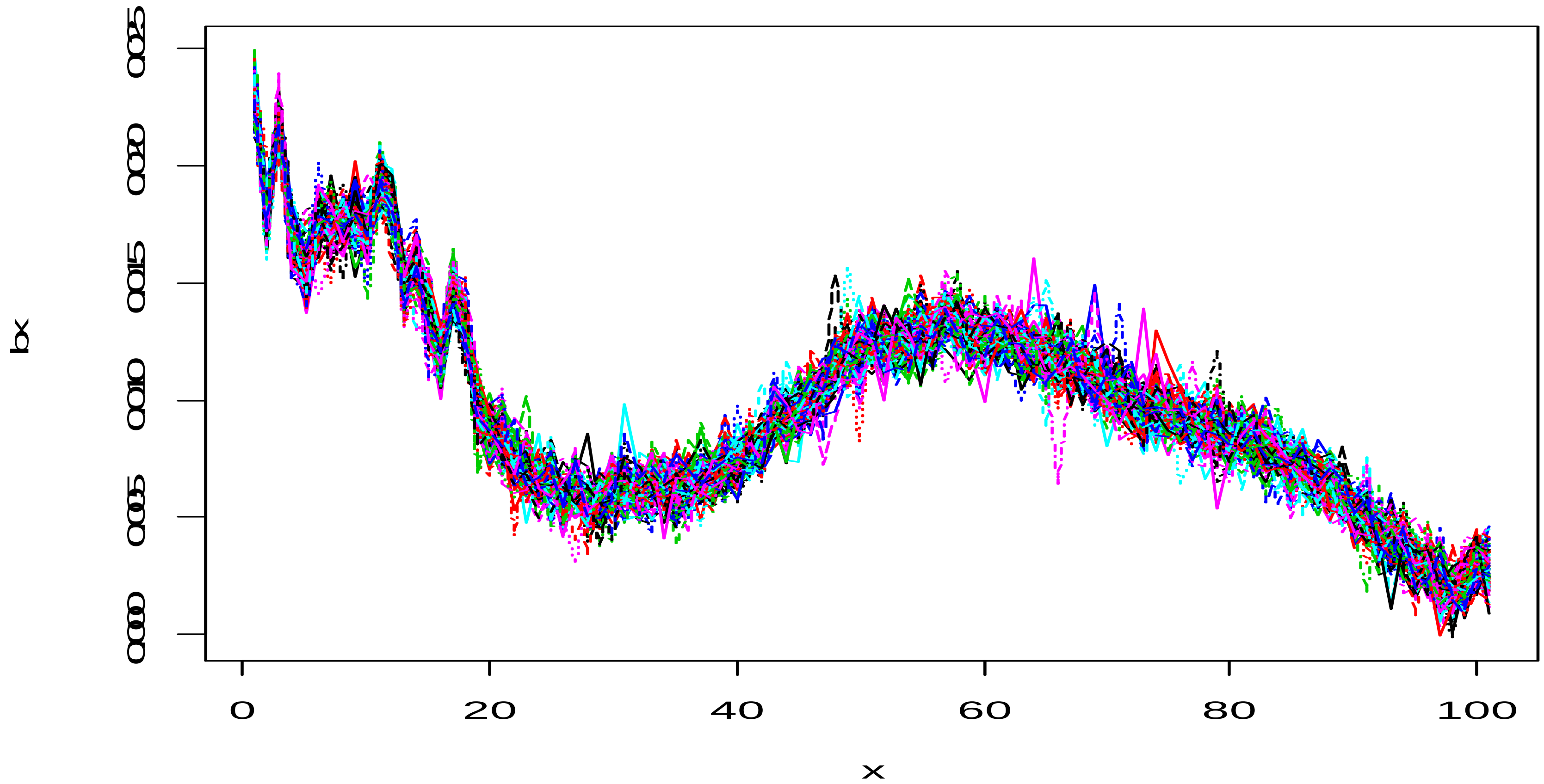

Figure 13 - Forecasted kt – Residual standard bootstrap

Numerical Application

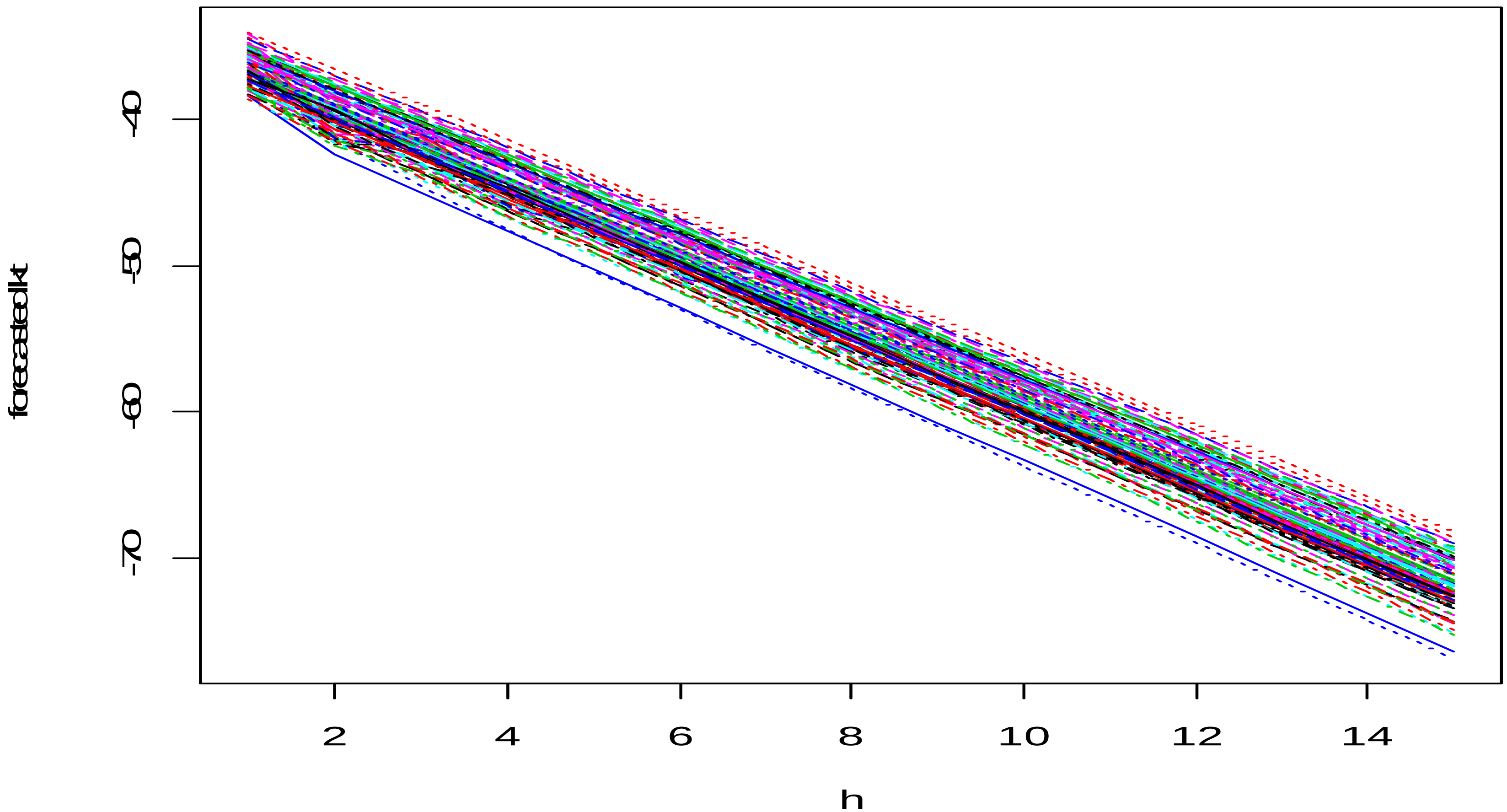

Figure 14 - Forecasted kt – Residual standard bootstrap

Numerical Application

	Residual Bootstrap		Sieve Bootstrap	
h	**5%**	**95%**	**5%**	**95%**
1	-38.10	-34.81	-41.75	-28.37
2	-41.45	-37.53	-45.66	-32.52
3	-43.66	-39.89	-47.82	-34.01
4	-46.30	-42.34	-49.82	-37.49
5	-48.82	-44.78	-52.95	-40.37
6	-51.40	-47.22	-55.07	-41.89
7	-53.95	-49.67	-57.69	-43.44
8	-56.52	-52.13	-61.43	-47.03
9	-59.08	-54.60	-63.52	-50.13
10	-61.66	-57.07	-66.63	-51.33
11	-59.55	-64.23	-68.97	-53.24
12	-62.02	-66.80	-72.17	-55.72
13	-69.37	-64.47	-74.42	-57.84
14	-71.94	-66.91	-77.69	-60.33
15	-74.50	-69.36	-80.27	-62.98

Table 3- Non parametric standard bootstrap and Sieve bootstrap 5% and 95% Confidence Intervals for k_{t+h}

Concluding Remarks

Our research proposes a particular bootstrap methodology, the LC Sieve Bootstrap, for capturing the ***dependence*** in deriving ***prediction intervals***, thus avoiding a systematic over-estimation or under-estimation of the amount of uncertainty in the parameter estimates, respectively if negative or positive dependence dominates.

Concluding Remarks

- The ***standard residual bootstrap*** procedure does not preserve the correlation structure in the data.

- The ***sieve bootstrap,*** on the other hand, captures the dependency structure, leading to more reliable uncertainty measurement

Main References

- *Alonso, A. M., Pena D., Romo, J., 2002, Forecasting time series with sieve boostrap, Journal of Statistical planning and inference, n.100*

- *Bühlmann, P.,1997, Sieve bootstrap for time series. Bernoulli 3, 123-148.*
-
- *Bühlmann, P., 1998, Sieve bootstrap for smoothing in nonstationary time series. The Annals of Statistics 26 (1), 48-82.*

- *Bühlmann, P., 2002, Bootstrap for time series. Statistical science 17 (1), 52-72.*

Main References

- *Choi, E., Hall, P., 2000, Bootstrap confidence regions computed from autoregressions of arbitratry order, Biometrika, 62: 461-477.*
- *Deaton, A. and Paxson C., 2004, Mortality, Income, and Income Inequality Over Time in the Britain and the United States. Technical Report 8534 National Bureau of Economic Research Cambridge, MA: . http://www.nber.org/papers/w8534.*
- *Denton, F.T, Feaver C.H., Spencer B.G., 2005, Time series analysis and stochastic forecasting: An econometric study of mortality and life expectancy, Journal of Popul. Econ 18:203–227*
- *Efron, B., Tibshirani, R.J., 1994, An introduction to the bootstrap. Chapman & Hall*
- *Kreiss, J. (1992). Bootstrap procedures for AR(1)-process. Springer: Heidelberg.*

Main References

- *D'Amato, V., S. Haberman, M. Russolillo, 2009, Efficient Bootstrap applied to the Poisson Log-Bilinear Lee Carter Model, Applied Stochastic Models and Data Analysis – ASMDA 2009 Selected Papers*

- *D'Amato V., Haberman S., Russolillo M., 2011a, The Stratified Sampling Bootstrap: an algorithm for forecasting mortality rates in the Poisson Lee Carter setting, Methodology and Computing in Applied Probability*

- *D'Amato V., Di Lorenzo E., Haberman S., Russolillo M., Sibillo M., 2011b, The Poisson log-bilinear Lee Carter model: Applications of efficient bootstrap methods to annuity analyses, North American Actuarial Journal*

Main References

- *Koissi, M.C., A.F. Shapiro, G. Hognas, 2006, Evaluating and Extending the Lee–Carter Model for Mortality Forecasting: Bootstrap Confidence Interval. Insurance: Mathematics and Economics 26: 1-20*

- *Kunsch, H.R., 1989, The jackknife and the bootstrap for general stationary observations, Ann. Statist. 17 1217-1241.*

- *Lahiri, S.N., 2003, Resampling Methods for Dependent Data, Springer*

- *Lee, R.D., L. R. Carter, 1992, Modelling and Forecasting U.S. Mortality, Journal of the American Statistical Association, 87, 659-671.*