A COMPARISON OF A BIDIMENSIONAL SDE AND A VARMA MODEL FOR FORECASTING MORTALITY RATES

Sandra Lagarto, Carlos A. Braumann, Dulce Gomes

CIMA - Research Center in Mathematics and Applications of the University of Évora

Fourth Workshop
Dynamical Systems Applied to Biology and Natural Sciences
Lisbon, 02/15/2013

UNIVERSIDADE DE ÉVORA cima-UE

Plan for talk

Original Problem

Considering increasing life expectancy on last decades,

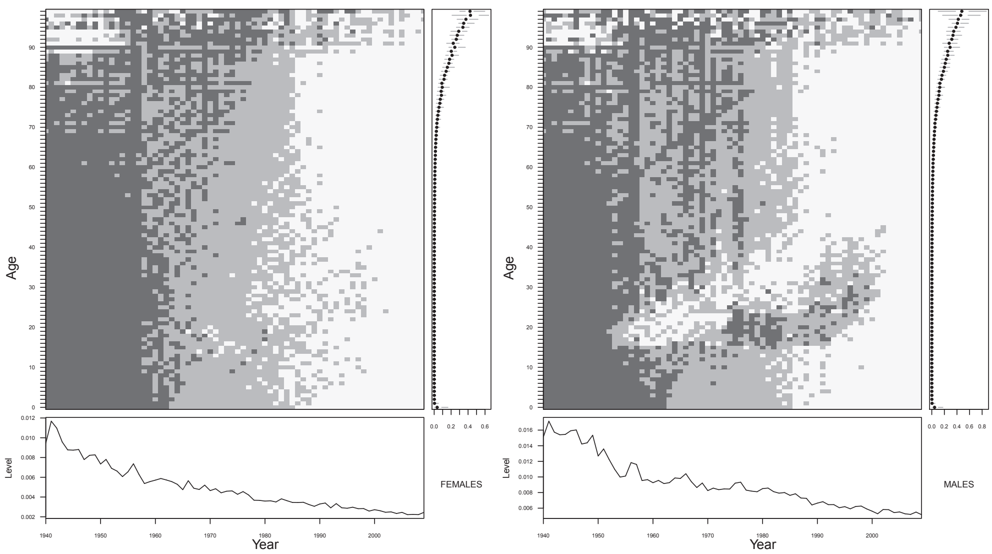
what are the short/long term mortality trends?

Current motivation

- **Problem**
  There are environmental fluctuations that may influence in the same way death rates on both males and females;
- **Goals**
  Apply multivariate stochastic differential equations models (SDE) and vector autoregressive and moving averages (VARMA) models, to *CDR* of Portuguese population;
  - Compare forecasts, by age.

Data

Annual (1940-2009) CDR(%..) of Portuguese population by age (age 0 to 99) and sex are available at `http://www.mortality.org/`



Bi-dimensional Stochastic Gompertz Model

For each age, the model can be written as a system of autonomous SDE

$$\begin{cases} dY_1(t) = b_1(A_1 - Y_1(t))dt + \sigma_1 dW_1^*(t) \\ dY_2(t) = b_2(A_2 - Y_2(t))dt + \sigma_2 dW_2^*(t), \end{cases}$$

with $Y_i(t) = \ln(X_i(t))$ and $Y_i(t_0) = y_{i,t_0}$, $X_i(t)$ the *CDR* of Portuguese population at time t, by age and sex (i=1 females; i=2 males), $W_i^*(t)$ correlated standard *Wiener processes* and parameter $A_i = \ln(a_i)$ (a_i, asymptotic death rates), b_i (approach rate to asymptotic regime), σ_i (intensity of environmental fluctuations).

The solutions of the SDE system, for each age, at time t, are

$$Y_1(t) = A_1 + (y_{1,t_0} - A_1)exp\{-b_1(t-t_0)\} + \sigma_1 exp\{-b_1 t\} \int_{t_0}^{t} exp\{b_1 s\}\, dW_1^*(s)$$
$$Y_2(t) = A_2 + (y_{2,t_0} - A_2)exp\{-b_2(t-t_0)\} + \sigma_2 exp\{-b_2 t\} \int_{t_0}^{t} exp\{b_2 s\}\, dW_2^*(s)$$

We can use to forecasts, for each age (with $t > t_k$)

$$\widehat{Y}_i(t) = \widehat{E}[Y_i(t)|Y_{i,t_k}] = \widehat{A}_i + (y_{i,t_k} - \widehat{A}_i)exp\{-\widehat{b}_i(t-t_k)\}$$

Maximum Likelihood Estimation.

VARMA(p,q) Model

Let $\mathbf{Y}_t = (y_{1t}, ...y_{kt}, ..., y_{Kt})$ be a set of K variables with $k = 1, ..., K$, a stationary VARMA(p,q) process might be defined as:

$$\mathbf{Y}_t = \boldsymbol{\nu} + \mathbf{A}_1\mathbf{Y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{Y}_{t-p} + \boldsymbol{\epsilon}_t - \mathbf{M}_1\boldsymbol{\epsilon}_{t-1} - \cdots - \mathbf{M}_q\boldsymbol{\epsilon}_{t-q},$$

with $t = 0, \pm1, \pm2, ...$, $\mathbf{A}_i$ $(i = 1, ..., p)$ and $\mathbf{M}_j$ $(j = 1, ..., q)$, both $(K \times K)$, are coefficient matrices associated, respectively, to autoregressive and moving average compounds and $\boldsymbol{\epsilon}_t$ is a K-dimensional WN, process with non singular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')$.

For a VAR(1) process we obtain:

$$\mathbf{Y}_t = \boldsymbol{\nu} + \mathbf{A}_1\mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t.$$

Assuming the process is stable ($det(I_K - A_1z) \neq 0$ for $|z| \leq 1$) it can be represented in the general form

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}_1{}^i\boldsymbol{\epsilon}_{t-i}, \;\; \boldsymbol{\mu} = (\mathbf{I}_k - \mathbf{A}_1)^{-1}\boldsymbol{\nu}.$$

A forecast $\mathbf{Y}_{t+l}$ that minimize the MSE can be obtained, for $t + l > t$, by

$$\widehat{\mathbf{Y}}_t(l) = E(\mathbf{Y}_{t+l}|\mathbf{Y}_t, \mathbf{Y}_{t-l}, ...)$$

Bivariate VAR(1) model

A VAR(1) model, with 2 variables, is given as:

$$\begin{cases} Z_1(t) = \nu_1 + \alpha_{11}Z_1(t-1) + \alpha_{12}Z_2(t-1) + \epsilon_1(t) \\ Z_2(t) = \nu_2 + \alpha_{21}Z_1(t-1) + \alpha_{22}Z_2(t-1) + \epsilon_2(t), \end{cases}$$

with $Z_i(t) = \ln(X_i(t)/X_i(t-1))$ (with $X_i(t)$ the *CDR* of Portuguese population (i=1 females; i=2 males), α coefficients associated to the autoregressive polynomium, assuming error term $\epsilon_i(t)$: $E[\epsilon_i(t)] = 0$ and $E[\epsilon_i(t) * \epsilon_j(\tau)] = 0$ (with $t \neq \tau$).

Least squares estimation.

Model selection

- Test series stationarity;
- Select lag length;
- Test for cointegration;
- Estimate parameters and perform diagnostic tests.

Structural Analysis

What kind of information can we extract from the fitted-model?

- **Granger causality tests**
  Verify if exists causality and determines direction of relationship (if a time series contribute to forecast another series);
- **Impulse-response functions analysis**
  Study the effects of a simulated exogenous shock on the process and detect dynamic relationships over time;
- **Forecast error variance decompositions**
  Variance decomposition purpose is to achieve information about the forecast ability (what proportions of errors are due to shocks).

Fitting bivariate VAR(1) model to CDR of Portuguese population

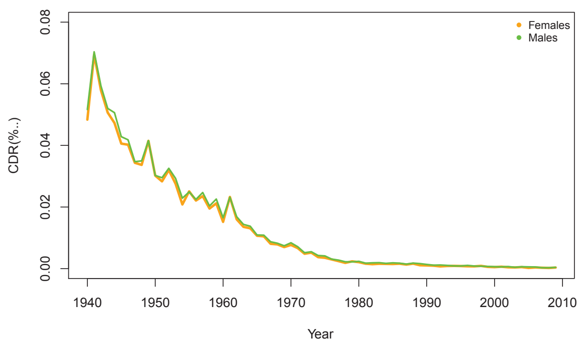
For each time series (per age), we use:

- 60 observations to fit the model (1940-1999);
- 10 observations left for prediction (2000-2009).

Introduction SDE Models VARMA models Results Conclusions/ Work in progress

Case study: age 3

$Z_1 = f$; $Z_2 = m$



Test variables stationarity

Augmented Dickey-Fuller Test Unit Root Test
Test regression trend

lm(formula = z.diff ≈ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
Min 1Q Median 3Q Max
-0.34550 -0.09113 0.01037 0.10535 0.32486

Coefficients:
Estimate Std. Error t value Pr(> |t|)
(Intercept) -0.1799256 0.0505949 -3.556 0.000823 ***
z.lag.1 -2.3039197 0.3055430 -7.540 7.64e-10 ***
tt 0.0008916 0.0013014 0.685 0.496376
z.diff.lag1 0.8013319 0.2219408 3.611 0.000697 ***
z.diff.lag2 0.3103135 0.1303748 2.380 0.021081 *
—
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.1563 on 51 degrees of freedom
Multiple R-squared: 0.7196, Adjusted R-squared: 0.6976
F-statistic: 32.72 on 4 and 51 DF, p-value: 1.6e-13

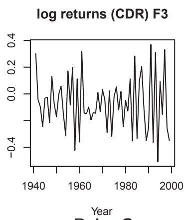
Augmented Dickey-Fuller Test Unit Root Test
Test regression drift

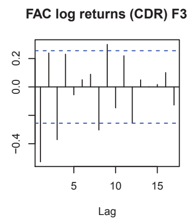
lm(formula = z.diff ≈ z.lag.1 + 1 + z.diff.lag)

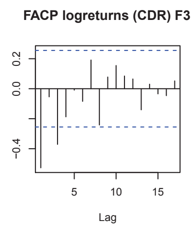
Residuals:
Min 1Q Median 3Q Max
-0.3228 -0.1001 0.0075 0.1016 0.3166

Coefficients:
Estimate Std. Error t value Pr(> |t|)
(Intercept) -0.1513 0.0284 -5.327 2.17e-06 ***
z.lag.1 -2.2818 0.3023 -7.549 6.61e-10 ***
z.diff.lag1 0.7862 0.2197 3.578 0.000759 ***
z.diff.lag2 0.3058 0.1295 2.361 0.022018 *
—
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.1555 on 52 degrees of freedom
Multiple R-squared: 0.717, Adjusted R-squared: 0.7007
F-statistic: 43.93 on 3 and 52 DF, p-value: 2.765e-14







Select the lag length

Select order p := 1..8

VARselect(Z.2,lag.max=8,type=´both´)

$selection

AIC(n)	HQ(n)	SC(n)	FPE(n)
3	3	1	3

$criteria

	1	2	3	4	5	6	7	8
AIC(n)	-7.1301646	-7.2073735	-7.2913695	-7.285747	-7.2107563	-7.253720	-7.2610770	-7.2530284
HQ(n)	-7.0143673	-7.0336776	-7.0597749	-6.996254	-6.8633645	-6.848430	-6.7978878	-6.7319406
SC(n)	-6.8271331	-6.7528263	-6.6853065	-6.528169	-6.3016619	-6.193110	-6.0489511	-5.889386
FPE(n)	0.0008011	0.0007427	0.0006849	0.000692	0.0007519	0.0007282	0.0007337	0.0007543

Test for cointegration

```
coint.test := ca.jo(Z.2, type = ´trace´, ecdet = ´trend´, K = 2, spec = ´transitory´)
summary(coint.test)
```

Johansen-Procedure
============

Test type: trace statistic, with linear trend in cointegration

Eigenvalues (lambda):
6.448363e-01 5.130834e-01 2.220446e-16

Values of test statistic and critical values of test:

test 10pct 5pct 1pct
r <= 1 | 41.02 10.49 12.25 16.26
r = 0 | 100.03 22.76 25.32 30.45

Eigenvectors, normalised to first column:
(These are the cointegration relations)

$f.l1 m.l1 trend.l1$
$f.l1$ 1.000000e+00 1.0000000 1.00000000
$m.l1$ -1.2140173198 0.1257628803 -0.7513293
$trend.l1$ 0.0002187754 -0.0001801689 -0.1688358

Weights W:
(This is the loading matrix)

$f.l1 m.l1 trend.l1$
$f.d$ -0.4394603 -1.560226 -9.208644e-19
$m.d$ 1.4294023 -1.158921 8.679284e-19

Estimated parameters

```
ajustVAR1FM:=VAR(Z,p=1,type=´both´)
summary(ajustVAR1FM,equation=´f´)
```

VAR Estimation Results:
==============

Estimated coefficients for equation f:
======================
Call:
$f = f.l1 + m.l1 + const + trend$

Estimate Std. Error t value Pr(> |t|)
$f.l1$ -0.6890953 0.1220413 -5.646 6.25e-07 ***
$m.l1$ 0.3402189 0.1397565 2.434 0.0183 *
const -0.0974053 0.0460838 -2.114 0.0392 *
trend -0.0001752 0.0013091 -0.134 0.8940
—

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.1668 on 54 degrees of freedom
Multiple R-Squared: 0.3723, Adjusted R-squared: 0.3374
F-statistic: 10.68 on 3 and 54 DF, p-value: 1.291e-05

```
ajustVAR1FM:=VAR(Z,p=1,type=´both´)
summary(ajustVAR1FM,equation=´m´)
```

VAR Estimation Results:
==============

Estimation results for equation m:
=====================
Call:
$m = f.l1 + m.l1 + const + trend$

Estimate Std. Error t value Pr(¿|t|)
$f.l1$ -0.0061819 0.1262455 -0.049 0.9611
$m.l1$ -0.2540421 0.1445710 -1.757 0.0845 .
const -0.0913414 0.0476714 -1.916 0.0607 .
trend 0.0001964 0.0013541 0.145 0.8852
—

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.1725 on 54 degrees of freedom
Multiple R-Squared: 0.07249, Adjusted R-squared: 0.02096
F-statistic: 1.407 on 3 and 54 DF, p-value: 0.2508

Diagnostic tests

ser11:=serial.test(ajustVAR1FM,lags.pt=16,type=´PT.asymptotic´)

Portmanteau Test (asymptotic)

data: Residuals of VAR object ajustVAR1FM
Chi-squared = 68.8094, df = 60, p-value = 0.2038

norm1 := normality.test(ajustVAR1FM)

JB-Test (multivariate)

data: Residuals of VAR object ajustVAR1FM
Chi-squared = 2.6052, df = 4, p-value = 0.6259

Skewness only (multivariate)

data: Residuals of VAR object ajustVAR1FM
Chi-squared = 0.1727, df = 2, p-value = 0.9173

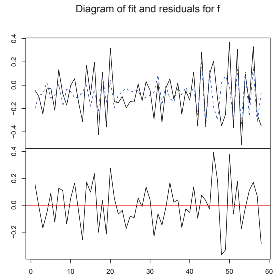
Kurtosis only (multivariate)

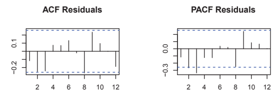
data: Residuals of VAR object ajustVAR1FM
Chi-squared = 2.4325, df = 2, p-value = 0.2963

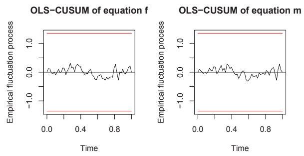
arch1 := arch.test(ajustVAR1FM,lags.multi=5)

ARCH (multivariate)

data: Residuals of VAR object ajustVAR1FM
Chi-squared = 63.6835, df = 45, p-value = 0.03465



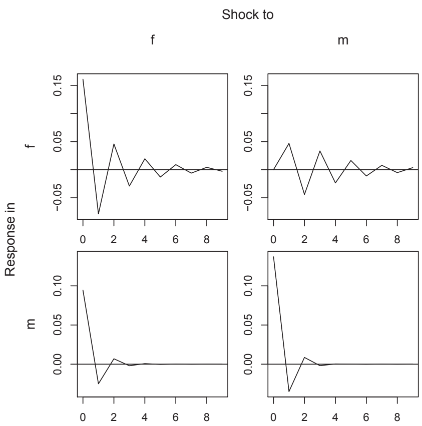




Impulse-response functions analysis

irf.FM := reduced.form.var(Z.2,p=1)
irf.FM

[,1] [,2]
-0.07871481 -0.0349111
0.16096526 0.1371232



Testing for Granger Causality

var.causal.f := causality(ajustVAR1FM, cause = ´f´)
$Granger
Granger causality H0: f do not Granger-cause m

data: VAR object ajustVAR1FM
F-Test = 0.0024, df1 = 1, df2 = 108, p-value = 0.961

$Instant
H0: No instantaneous causality between: f and m

data: VAR object ajustVAR1FM

Chi-squared = 14.1439, df = 1, p-value = 0.0001693

var.causal := causality(ajustVAR1FM, cause = ´m´)
$Granger
Granger causality H0: m do not Granger-cause f

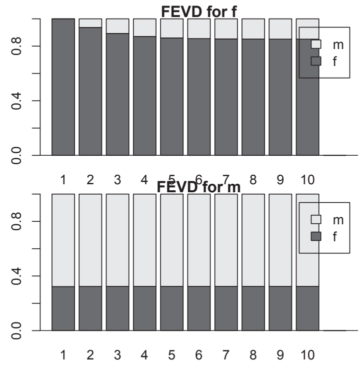
data: VAR object ajustVAR1FM
F-Test = 5.9262, df1 = 1, df2 = 108, p-value = 0.01656

$Instant
H0: No instantaneous causality between: m and f

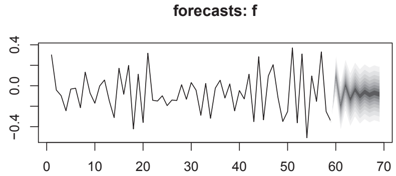
data: VAR object ajustVAR1FM

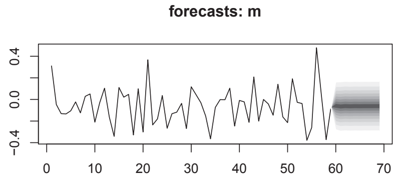
Chi-squared = 14.1439, df = 1, p-value = 0.0001693

Variance decomposition of forecast errors

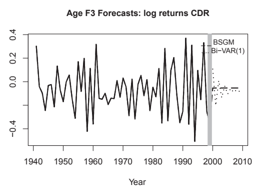


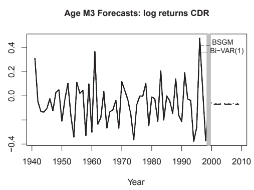
Bi-VAR(1) model forecasts





Bi-VAR(1) model vs BSGM





Parameter estimates ($p < 0.05$) and 95% semi-amplitude CI

BSGM	f	m
$\widehat{A}$	-9.29764 $\pm$7.486291	-9.49275 $\pm$9.372
$\widehat{b}$	0.020124 $\pm$0.04741	0.016465 $\pm$0.044205
$\widehat{\sigma}$	0.201043 $\pm$0.03685	0.171037 $\pm$0.03136156
$\widehat{r}$	0.434236 $\pm$0.209347	

Goodness-of-fit: MSE

Models	f	m
Bi-VAR(1)	0.02547221	0.01706095
BSGM	0.01388476	0.007294422

Conclusions/Work in Progress

Both models are well-fitted to the dataset; to most of the fitted ages, forecasts for CDR log-returns don´t show a linear pattern, which make it more realistic, however with a higher associated MSE when compared to SDE models.

VAR(p) models advantages face to BSGM:

- Allow to explore dynamical relations between variables;
- Provide good forecasts;
- ....

VAR(p) models disadvantages:

- There´s an increase in the number of parameters with the increase of the time lag lenght;
- If p is small, fitted models could be less precise;
- Higher MSE estimates when comparing to SDE models;
- Still present several computational restrictions and scarce numerical methods.

Work in progress:

- Fit VARMA(p,q) or dynamical models to CDR, considering correlations by age and compare its performances with a multivariate SDE model which includes the same effect;
- Introduce some exogenous variables (to the CDR system) in the SDE multivariate model.

References...

- Filipe, P.A., Braumann, C.A., Roquete, C.J. (2010). Multiphasic Individual Growth Models in Random Environments. *Methodology and Computing in Applied Probability*, DOI 10.1007/s11009-010-9172-0
- Lagarto, S., Braumann, C. (2012). Modeling human population death rates: A bi-dimensional stochastic Gompertz model with correlated Wiener processes. [in print]
- Lütkepohl, H. (2005). An Introduction to Multiple Time Series Analysis. Springer
- Tong, Y. L. (1989). Multivariate Normal Distribution. Springer Series in Statistics

smdl@uevora.pt
braumann@uevora.pt
dmog@uevora.pt

Thank You!