

Margvís títaraðalíkön í samfelldum tíma

Helgi Tómasson

28.nóvember 2019

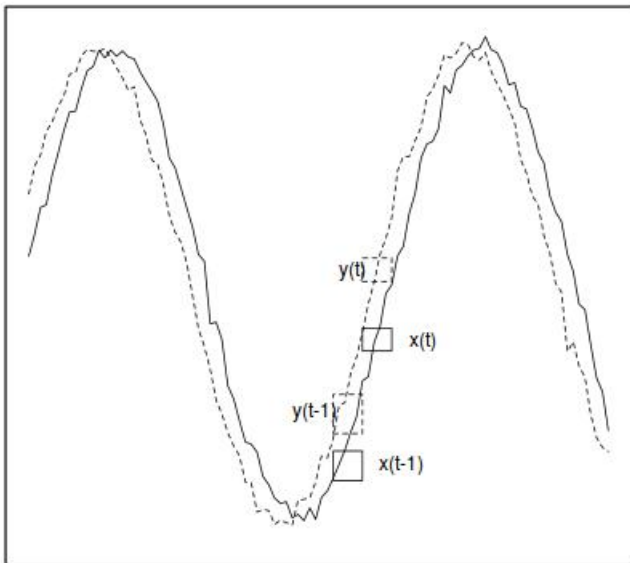
Skipulag fyrirlestrar

- Hvers vegna tímaraðalíkan? Tengsl við eldri hagrannsóknaraðferðir. Minn bakgrunnur.
- Strjáll tími eða samfelldur?
- Einvítt líkan í samfelldum tíma. Ýmis tæknileg atriði.
- Frá einni vídd í margar, AR yfir í VAR í strjálum tíma, hliðstæða í mörgum víddum.
- Upprifjun á nonsense-correlation/spurious-regression.
- Punktur um samfelldan tíma og spectur
- Hvaða segja tímaraðamenn um global-warming? Smooth-trend eða fractional integration?
- Framtíðarplön

Hvers vegna tímaraðalíkan? Tengsl við eldri hagrannsóknaraðferðir. Minn bakgrunnur.

- Ég læri fyrst um tímaraðagreiningu 1976-1977. Box-Jenkins aðferðafræðin (Box & Jenkins, 1970). Setti hana upp í Seðlabankanum 1978-1980.
- Hagrannsóknir þess tíma einkenndust af deilum milli þeirra sem vildu setja upp flókin líkön af öllu hagkerfinu sem tengdu saman mjög margar breytur (t.d. 200) og þeirra sem höfðu uppgvötvað þá empírísku staðreynd að einföld tímaraðalíkön spáðu betur. Frammistaða stóru líkananna leit út eins og brandari.
- Einföld skýring á góðri frammistöðu tímaraðalíkna gæti t.d. verið einhvers konar „omitted-variable“ vandi.
- Fylgni milli $X(t)$ og $Y(t)$ getur verið sýndarfylgni ef raunverulega samböndin eru á milli $X(t)$ og $X(t-1)$ eða $Y(t)$ og $Y(t-1)$. Hreyfimyndir er mikilvægt atriði.
- Tímaröð þarf að sundurliða í:

trend + cycle + season + irregular



- Í tímaraðgreiningu fæst bara eitt tækifæri og því þarf stöðugleikahugtök, stationarity og ergodicity.
- Stationary hugtakið sem notast er við er: $E(X(t) = \mu, V(X(t)) \leq \infty$ og $\gamma(\tau) = E(X(t) - \mu)(X(t - \tau) - \mu)$ bara fall af τ .
- Sjálffylgni= $\gamma(\tau)/\gamma(0)$.
- Grundvallar viðmiðunarferli eru „white-noise“, ε_t ,

$$\varepsilon_t, \quad E(\varepsilon_t) = 0, \quad V(\varepsilon_t) = \sigma^2, \gamma_\varepsilon(\tau) = 0 \text{ ef } \tau \neq 0.$$

og random-walk, $X_t = X_{t-1} + \varepsilon_t = \sum^t \varepsilon_j$,

$$E(X_t | X_0 = 0) = 0, \quad V(X_t | X_0 = x_0) = t\sigma^2,$$

$$E(X_t X_s | X_0 = 0) = \frac{\sigma^2}{2} (t + s - |t - s|) = \sigma^2 \min(t, s).$$

p.e. variance og sjálffylgnifall háð t .

- Ef horft er á tímaröð þarf að taka afstöðu til þess hvor hún er líkari „white-noise“ eða random-walk.

ARMA líkön

- ARMA líkön ganga út frá að núverandi gildi sé skýrt með nýliðnum gildu og óvæntum samtímaskelli. Einfaldasta gerð er AR(1):

$$X_t = \phi X_{t-1} + \varepsilon_t,$$

$$\Delta X_t = X_t - X_{t-1} = (\phi - 1)X_t + \varepsilon_t.$$

X_t er stationary ferill ef $|\phi| < 1$.

- Yfirfært í samfelldan tíma væri þetta:

$$dX(t) = \alpha X(t) + \text{"white-noise"}$$

- Hvað þýðir „white-noise“ í samfelldum tíma? Ein hugmynd er að nota Wiener-ferli:

$$W(t) = \int_0^t dW(s), \quad W(0) = 0, E(W(t)W(s)) = \min(t,s).$$

Ef $W(t)$ er samfelldur ferill, með „independent-increment“ þá er $W(t) \sim N(0, t\sigma^2)$.

- Ferlið, $X(t)$, CAR(1), Ornstein-Uhlenbeck (stationary ef $\alpha \leq 0$),

$$dX(t) = \alpha X(t)dt + \sigma dW(t),$$

$$X(t) = \exp(\alpha t)X(0) + \sigma \exp(\alpha t) \int_0^t \exp(-\alpha s)dW(s)$$

- ARMA(p,q)

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

$$E(\varepsilon_t) = 0, \text{ when } t \neq s \quad E(\varepsilon_t \varepsilon_s) = 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2.$$

- ARMA(p,q) er stationary ef rætur margliðunnar $1 - \phi_1 z - \dots - \phi_p z^p$ eru utan einingahrings.
- Ferli, X_t er sagt ARIMA(p,d,q) ef:

$$\Delta^d X_t \text{ er stationary ARMA}(p,q), \text{ þar sem } \Delta X_t = X_t - X_{t-1}.$$

- ARIMA(p,d,q) eru non-stationary ferlar. Slík ferli eru söguð I(d) (integrated of order d). Random-walk er ARIMA(0,1,0).

- CAR(1) er stationary ef $\alpha < 0$.
- ARMA er útvíkkun á AR(1),

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \varepsilon_t,$$

þetta er AR(1) ef $\phi_2 = 0$. $AR(1) \subseteq AR(2)$.

- Svona gildir ekki fyrir samfelldan tíma. Þar gildir $CAR(1) \subseteq CARMA(2,1)$, en $CAR(1) \not\subseteq CAR(2,0)$.
- ARMA(p,q) líkan í samfelldum tíma er skrifað t.d.:

$$Y^{(p)}(t) + \alpha_1 Y(t)^{(p-1)} + \dots + \alpha_p Y(t) = \\ \sigma(W(t)^{(1)} + \beta_1 W(t)^{(2)} + \dots + \beta_q W(t)^{(q)})$$

- $Y(t)$ er stationary ef rætur: $z^p + \alpha_1 z^{p-1} + \dots + \alpha_p$ hafa neikvæðan rauntöluhluta. $Y^{(p)}(t)$ stendur fyrir p -tu afleiðu.
- Sjálfyfyllingfall ARMA líkana fellur exponentialt.

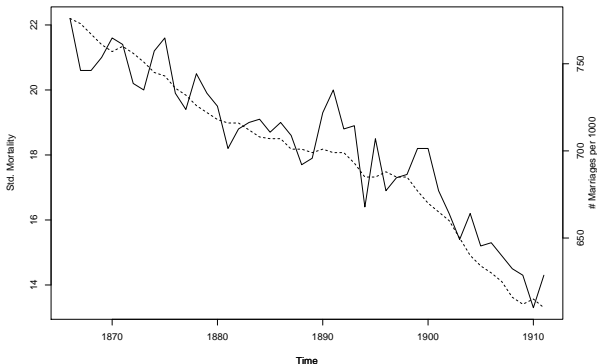
Sjónhverfing: Enska biskupakirjan stórhættuleg

- Engendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.
- Þessar stofnanir safna vönduðum mælingum.
- Til dæmis hefur landlæknisembættið safna gögnum um „mortalitet“, dánir á ári per 1000 íbúa og kirkjan hefur talið hver mörg hjónabönd á ári per 1000 hjónabönd eru á vegum kirkjunnar.
- Hver skyldu tengsl þessara hagstærða vera?
- Reiknum fylgnistuðul og fáum:
Svar: 0.95
- Ályktum, **sterkt samband** eða jafnvel **sterkt marktækt samband** milli markaðshlutdeildar kirkjunnar í brúðkaupum og „mortalitets“.

- **Varúð:** Hugtakið marktækt hefur ekkert með mikilvægi að gera.
- Marktækt þýðir eingöngu: $p < \alpha$, þar sem p þýðir líkur á fenginni útkomu eða ótrúlegri gefið að tiltekið tölfræðilegt líkan sé rétt.
- Ekkert vit í gagnagreiningu (ályktunum) án tölfræðilegs líkans. Ég tel fullt vit í reikningshaldi/bókhaldi/bókasafnsfræði og annari skráningu staðreynda.
- Í þessum útreikningi hefur það gleymst að gögnin eru tímaraðir og því kallar tölfræðileg greining á tímaraðalíkön/tímaraðaaðferðir.
- Mjög mikið af haggögnum eru tímaraðir.
- Á tímabilinu 1970-1985 voru miklar deilur í ekónometrúnni milli fylgismanna stórra flókinna líkana og einfaldra tímaraðalíkana.

- Yule (1926): Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series.

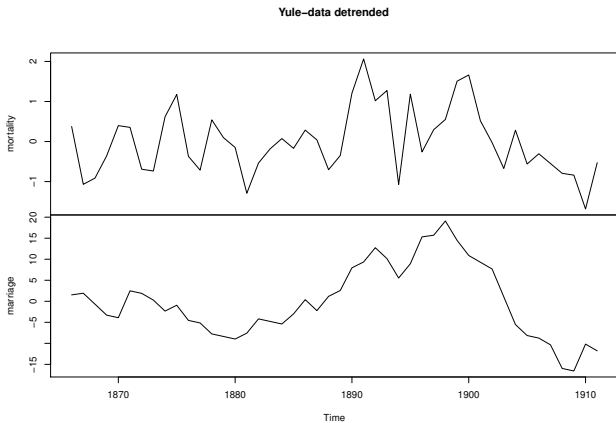
Mortality and marriage rates



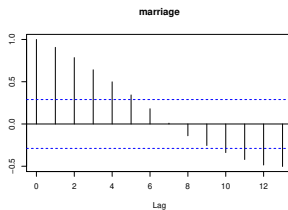
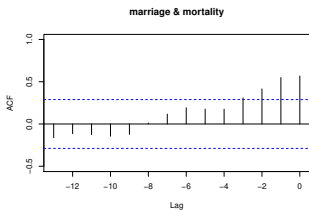
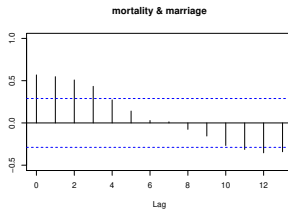
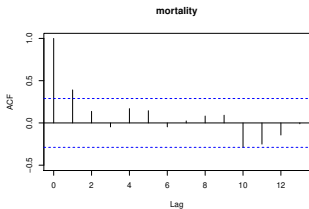
Mynd: Yule(1926) data on mortality and marriages.

Lausn Box og Jenkins 1970

- Við kunnum reglu Bartlett um dreifingu úrtaksfylgnistuðla í stationary ferlum.
- Nálgum lausnaferil með ARIMA(p,d,q).
- - 1 Hreinsum burt trend og deterministíska þætti.
 - 2 Nálgum trendhreinsaða röð með ARMA(p,q) líkani.
 - 3 Skoðum spáskekkur.
 - 4 Reiknum spá.
- Í margvíðum röðum trend- og sveifluhreinum hverja vídd (pre-whitening) og skoðum cross-fylgni (cross-spectur) hreinsuðu raðanna.
- Granger & Newbold (1974) kynntu hugmyndir Yule, Bartlett ofl., og markaðsettu í eknómétríunni.
- Granger & Newbold (1974) rifjuðu upp reglu Bartlett (1946) sem segir að varíans úrtaksfylgni tveggja óháðra stationary ferla er flókið fall af sjálfylnifalli þeirra. Ef sjálfylnin er mikil og röðin stutt getur dreifing úrtaksfylgnistuðulsins orðið U-laga. (Nemendur mínir eiga að kannast við þessi dæmi)

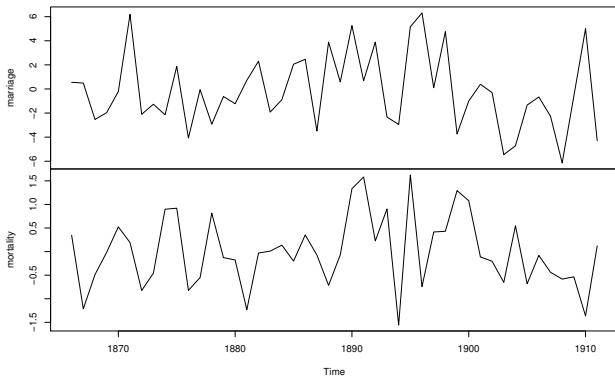


Mynd: Trendhreinsuð Yule-gögn.

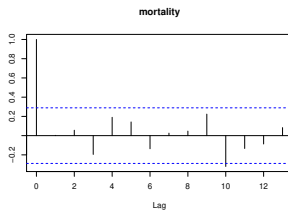
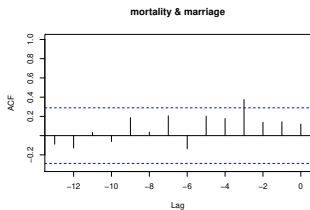
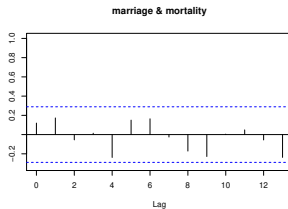
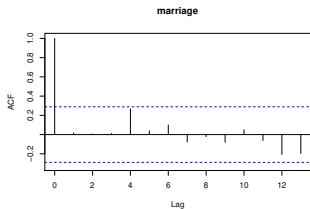


Mynd: Sjálffylgni í trendhreinsuðum Yule-gögnum.

Yule-data filtered



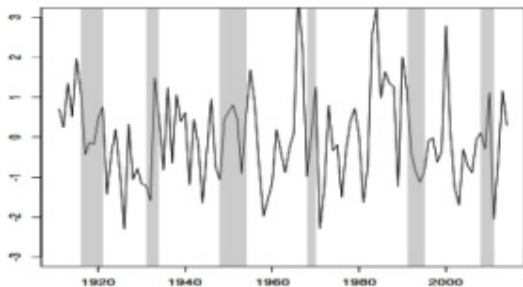
Mynd: Trend og sjálffylgnihreinsuð Yule-gögn.



Mynd: Sjálffylgni í trend- og sjálffylgnihreinsuðum Yule-gögnum.

- Niðurstaða: Ekkert sem bendir til þess að samband sé milli mortalitets og markaðshlutdeildar Ensku biskupakirkjunnar í brúðkaupum.
- Nútímamenn myndu e.t.v. meta tvívítt VAR með trendum og prófa fyrir Granger causality.
- Ekónómetrían var slæm með svona fram yfir 1980. Hugsanlega vegna þess að Havelmaa sló í gegn með simultan líkönum. Hermann Wold sagði: simultan líkөн skemmtileg en hagfræði ekki simultan vísindagrein.
- Granger cointegration kúltur gjörbylti hvernig ekónómetrían tekur á tímaraðagögnum.
- Mér vitanlega er ekónómetrían eina tölfræðigreinin sem hefur tekið formlega á því hvernig bera að álykta um non-stationary tímaraðir (I(1), vector ARIMA(p,1,q)/unit-root).

Nýleg grein í Læknablaðinu Oskarsson et al. (2019) sýnir að sjálfsmorð geta ekki tengst efnahagskreppum



Mynd 1: Sköluð frávik frá væntanlegu nýgengi sjálfsmorða. Kreppur eru skyggðar.

Mynd: Sjálfyllgni í trend- og sjálfyllgnihreinsuðum Yule-gögnum.

Hvað þýðir spectur ferlis?

- Gróft séð er það varíans Fourier-transforms ferlis.

$$\mathcal{F}_{\mathbf{X}}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2T}} \int_{-T}^T \mathbf{X}(t) \exp(-i\omega t) dt,$$

and the corresponding spectral density as:

$$S_{\mathbf{X}}(\omega) = \lim_{T \rightarrow \infty} E(\mathcal{F}_{\mathbf{X}}(\omega) \overline{\mathcal{F}_{\mathbf{X}}(\omega)}),$$

$$S_{\mathbf{X}}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega\tau) \Gamma_{\mathbf{X}}(\tau) d\tau.$$

- Hér er $\Gamma_{\mathbf{X}}(\tau) = E(\mathbf{X}(t)\mathbf{X}(t)')$, auto-cross-covariance fylki margvíða ferlisins $\mathbf{X}(t)$.

The diagonal of the matrix $S_{\mathbf{X}}(\omega)$ contains the spectral function of each series. The off-diagonal element, the cross-spectrum, is a measure of covariance in the frequency domain. If $S_{ij}(\omega)$ is the cross-spectrum between coordinate i and j the coherence is given by:

$$coh_{ij}(\omega) = \frac{|S_{ij}(\omega)|^2}{S_{ii}(\omega)S_{jj}(\omega)},$$

and the phase is given by:

$$phase_{ij}(\omega) = atan(-Im(S_{ij}(\omega)/Re(S_{ij}(\omega))).$$

Where Im and Re give the imaginary and real part of a complex number, respectively. Elementary Fourier theory shows that if $X_2(t)$ is a lagged version of $X_1(t)$, i.e., $X_2(t) = X_1(t-l)$, then

$$S_{12}(\omega) = \exp(-il\omega)S_{11}(\omega).$$

It is therefore clear that if one coordinate is the lagged version of the other the coherence is one and the phase is $l\omega$.

- Hallatala phase gefur til kynna lag length og coherence fylgni per tíðni.

A two dimensional Ornstein-Uhlenbeck (AR(1),VAR(2)) can be written as:

$$dY_1(t) = \alpha_{11}Y_1(t)dt + \alpha_{12}Y_2(t)dt + \sigma_1dW_1(t),$$

$$dY_2(t) = \alpha_{21}Y_1(t)dt + \alpha_{22}Y_2(t)dt + \sigma_2dW_2(t),$$

$$E(dW_1(t)dW_2(t)) = \rho\sigma_1\sigma_2dt.$$

Equation system (21) is more conveniently written in matrix form:

$$d\mathbf{Y}(t) = \mathbf{A}\mathbf{Y}(t)dt + d\mathbf{W}(t), \quad E(d\mathbf{W}(t)d\mathbf{W}(t)') = \Sigma dt,$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}.$$

The dependency structure, the auto-correlation, in a two-dimensional CAR(1) depends on the causal parameters, α_{12} , α_{21} , and the correlation, ρ , between the innovations.

- Hér má t.d. álykta um Granger-causality.

A generalisation of the univariate CARMA model to a multivariate CARMA can be written formally as:

$$\mathbf{y}^{(p)}(t) = A_1 \mathbf{y}^{(p-1)}(t) + \dots + A_p \mathbf{y}(t) + d\mathbf{W}(t) + B_1 d\mathbf{W}^{(2)}(t) + \dots + B_q d\mathbf{W}^{(q+1)}(t),$$

$$V(d\mathbf{W}(t)d\mathbf{W}(t)') = \Sigma dt,$$

$$A_i = \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} & \cdots & \alpha_{1d,i} \\ \alpha_{21,i} & \alpha_{22,i} & \cdots & \alpha_{2d,i} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{d1,i} & \cdots & \cdots & \alpha_{dd,i} \end{bmatrix}, \quad i = 1, \dots, p,$$

$$B_j = \begin{bmatrix} \beta_{11,j} & \cdots & \cdots & \beta_{1d,j} \\ \beta_{21,j} & \cdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{d1,j} & \cdots & \cdots & \beta_{dd,j} \end{bmatrix}, \quad j = 1, \dots, q.$$

- The parameters are $p + q$ $d \times d$ matrices, the A_i 's and B_j 's and $(d(d+1)/2)$ parameters in the covariance matrix of the innovations Σ . The number of parameters that describe the covariance structure in the system is thus $pd^2 + qd^2 + d(d+1)/2$.

State-space framsetning

The representation in Zadrozny (1988) is:

$$\mathbf{y}(t) = \mathbf{C}\mathbf{X}(t), \quad \mathbf{C} \quad pd \times d$$

$$d\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t)dt + \mathbf{R}d\mathbf{W}, \quad \mathbf{C} = \begin{bmatrix} I_d \\ 0_d \\ \vdots \\ 0_d \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} A_1 & I_{d \times d} & 0_{d \times d} & \cdots & \cdots & 0_{d \times d} \\ A_2 & 0_{d \times d} & I_{d \times d} & 0_{d \times d} & \cdots & 0_{d \times d} \\ \vdots & 0_{d \times d} & 0_{d \times d} & \ddots & \vdots & \vdots \\ A_{p-1} & 0_{d \times d} & 0_{d \times d} & \cdots & \cdots & I_{d \times d} \\ A_p & 0_{d \times d} & 0_{d \times d} & \cdots & \cdots & 0_{d \times d} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} B_q \\ \vdots \\ B_1 \\ I_{d \times d} \end{bmatrix}.$$

- Spektur fyrir $\mathbf{y}(t)$ er:

$$S_{\mathbf{y}}(\omega) = \mathbf{C}S_{\mathbf{X}}(\omega)\mathbf{C}'.$$

$$S_{\mathbf{X}}(\omega) = (i\omega I_{pd \times pd} - \mathbf{A})^{-1} \frac{\mathbf{R}\Sigma\mathbf{R}'}{2\pi} (-i\omega I_{pd \times pd} - \mathbf{A}')^{-1}$$

Another way of deriving the spectrum is to use the fact that

$E(\mathbf{X}(t+h)\mathbf{X}(t)') = \mathbf{e}^{Ah}\Gamma_{\mathbf{X}}(0)$, and $E(\mathbf{X}(t-h)\mathbf{X}(t)') = \Gamma_{\mathbf{X}}(0)\mathbf{e}^{Ah'}$, for $h > 0$. The Laplace transform of $\mathbf{e}^{\mathbf{A}t}$, is $\int_0^{\infty} e^{-st} \mathbf{e}^{\mathbf{A}t} dt = (sI_{pd \times pd} - \mathbf{A})^{-1}$. It is then straightforward to show that Fourier-transform of the auto-covariance function is:

$$S_{\mathbf{X}}(\omega) = \frac{1}{2\pi} \left((-i\omega I_{pd \times pd} - \mathbf{A})^{-1} \Gamma_{\mathbf{X}}(0) + \Gamma_{\mathbf{X}}(0) (i\omega I_{pd \times pd} - \mathbf{A}')^{-1} \right).$$

The diagonal of the matrix $S_{\mathbf{X}}(\omega)$ contains the spectral function of each series. The off-diagonal element, the cross-spectrum, is a measure of covariance in the frequency domain.

- The unconditional stationary variance, $\Gamma_{\mathbf{X}}(0)$ solves the equation system:

$$\mathbf{A}\Gamma_{\mathbf{X}}(0) + \Gamma_{\mathbf{X}}(0)\mathbf{A}' = -\mathbf{R}\Sigma\mathbf{R},$$

which, by using standard results on Kronecker products, can be transformed into;

$$((I \otimes \mathbf{A}) + (\mathbf{A} \otimes I)) \text{vec}(\Gamma_{\mathbf{X}}(0)) = -\text{vec}(\mathbf{R}\Sigma\mathbf{R}').$$

Non-synchronous sampling

- The non-synchronous case is solved by letting the \mathbf{C} matrix depend on the sampling. I.e. if coordinate i is observed the \mathbf{C} matrix consists of nothing but zeros except for coordinate ii which is set to one. The measurement equation with that particular \mathbf{C} reflects that coordinate i is measured at that particular time-point.
- With this setup the implementation of the Kalman-filter is straightforward. Solving the above equation, for $\Gamma_{\mathbf{X}}(0)$, can be numerically problematic. In some cases it is possible to derive explicit algorithms to calculate $\Gamma_{\mathbf{X}}(0)$, see, e.g. Tsai & Chan (2000).
- In practice it is worthwhile to observe that \mathbf{A} and I are both sparse matrices, so sparse-matrix algorithms can be used to invert $(I \times \mathbf{A}) + (\mathbf{A} \times I)$.
- Mörg númerísk erfið vandamál, matrix-exponent, númerísk háþörkun o.s.frv. (í process analýtísk diffrun á líkelihood falli). Pararellisation (t.d. CUDA) gæti nýst.
- Hef hannað aðferða fræði og pakkað inn í R-pakka, ctarmaRcpp (fyrir einvítt, arftaki ctarma pakkans sem var bannfærður) og mctarmaRcpp fyrir margvítt.

Smá illustration

```
set.seed(12345689)
a=log(2)
sigma=1
nn=1000
tt1=cumsum(rexp(nn, 1))/10
b=1
y1=carma.sim.timedomain(tt1, a, b, sigma)
y2=y1
tt2=tt1+1
my1=ctarma(ctarmalist(y1, tt1, a, b, sigma))
my2=ctarma(ctarmalist(y2, tt2, a, b, sigma))
my1e=ctarma.maxlik(my1)
my2e=ctarma.maxlik(my2)
sigma=1
nn=1000
tt1=cumsum(rexp(nn, 1))/10
b=1
y1=carma.sim.timedomain(tt1, a, b, sigma)
y2=y1
tt2=tt1+1
my1=ctarma(ctarmalist(y1, tt1, a, b, sigma))
my2=ctarma(ctarmalist(y2, tt2, a, b, sigma))
```

Now the objects `my1e` and `my2e` contain the maximum-likelihood estimates of the `CARMA(1,0)` model. The log-likelihood is calculated by:

```
> ctarma.loglik(my1e)
[1] 52.61891
```

There are many ways such that the `CARMA(1,0)` can be embedded in a higher order `CARMA`. The function `ctarma.new` creates an equivalent object with one more AR term and one more MA term.

```
> ctarma.loglik(ctarma.new(my1e))
[1] 52.61891
```

Using this feature repeatedly,

```
my2=ctarma.new(my1e)
my2e=ctarma.maxlik(my2)
my3=ctarma.new(my2e)
my3e=ctarma.maxlik(my3)
my4=ctarma.new(my3e)
my4e=ctarma.maxlik(my4)
```

higher order `CARMA` models can be fitted.

The estimated ARMA parameters and log-likelihood for CARMA(4,3) are:

```
> my4e$ahat
```

```
[1] 3.237786 5.369379 6.158908 1.837239
```

```
> my4e$bhat
```

```
[1] 1.0000000 1.1202398 0.7557877 0.2705627
```

```
> my4e$sigma
```

```
[1] 3.673917
```

```
> ctarma.loglik(my4e)
```

```
[1] 53.04228
```

The two univariate CARMA(4,3), my14e and my24e objects are merged into a bivariate non-synchronous CARMA(4,3) object by:

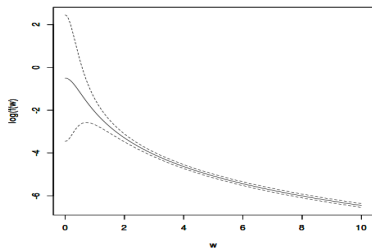
```
cc=cbind(as.list(my14e),as.list(my24e))
library(mctarmaRcpp)
mct4=mctobjnonsync(cc)
```

The only difference between the objects, my14e and my24e, is that the time has been translated one time-unit, as X_2 is just a lagged copy of X_1 . This bivariate CARMA(4,3) object, mct4, is composed of two estimated univariate CARMA(4,3) objects. The parameterization of this bivariate CARMA(4,3) object reflects the assumption of independence of the two coordinates (X_1, X_2). The log-likelihood value of the bivariate object is therefore the sum of the log-likelihood of univariate objects:

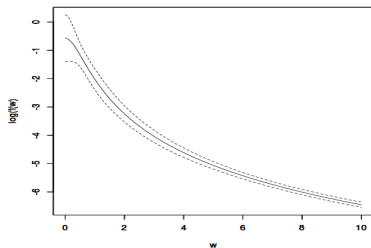
```
> mctarmanonsync.loglik(mct4)/2
[1] 53.04228

> cbind(mct4e$y,mct4e$tt,mct4e$z)[13:19,]
      [,1]      [,2] [,3]
[1,] -0.9942879  1.391931  2
[2,] -1.1332209  1.536528  2
[3,]  0.6258719  1.564668  1
[4,] -0.9815641  1.564864  2
[5,] -0.6691930  1.608199  2
[6,]  0.5880817  1.610724  1
[7,]  0.6950109  1.739455  1
```

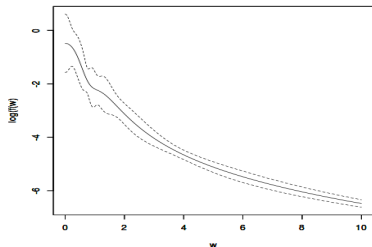
Log-spectrum of an estimated CARMA(1,0)



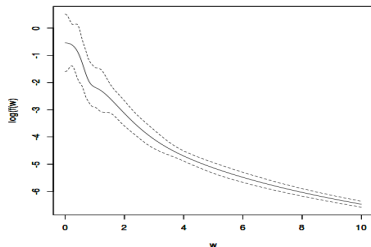
Log-spectrum of an estimated CARMA(2,1)



Log-spectrum of an estimated CARMA(5,4)

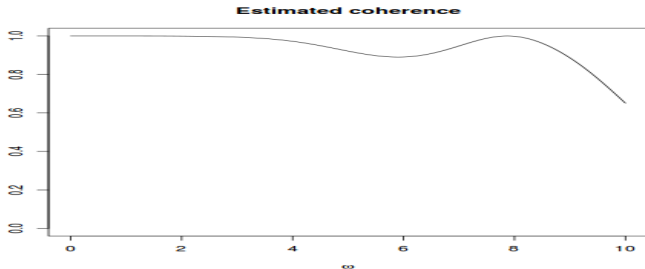


Log-spectrum of an estimated CARMA(6,5)

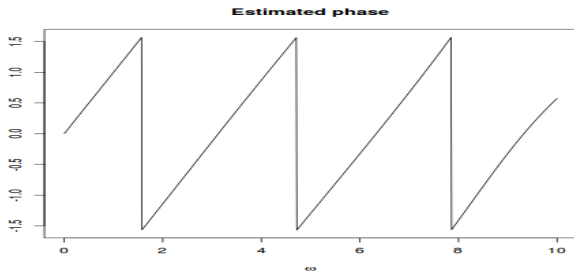


Mynd: The log-spectrum and approximate confidence intervals based on estimates of several CARMA models for a simulated Ornstein-Uhlenbeck process.

In the low frequencies the slope of the phase curve is close to the theoretical value 1, reflecting (the fact) that $X_2(t)$ is lagged version of $X_1(t)$.



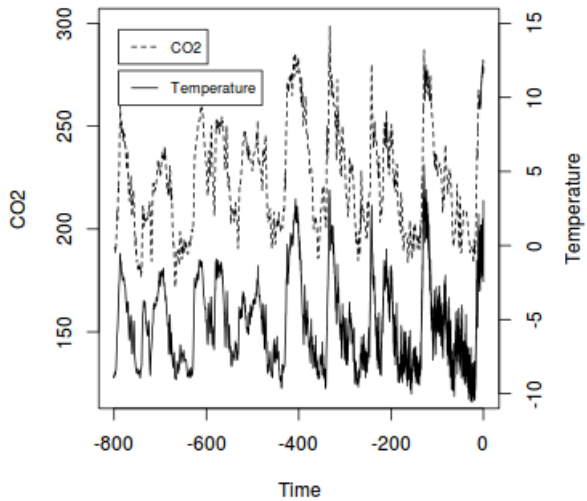
Mynd: Estimated coherence of a CARMA(1,0) and its lagged version. The estimates are based on estimates of a CARMA(4,3).

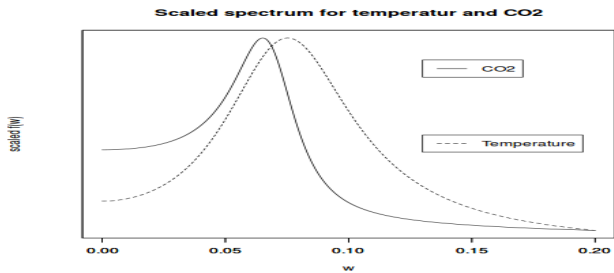


Mynd: The estimated phase of CARMA(1,0) and its lagged version. The estimates are based on estimates of a CARMA(4,3).

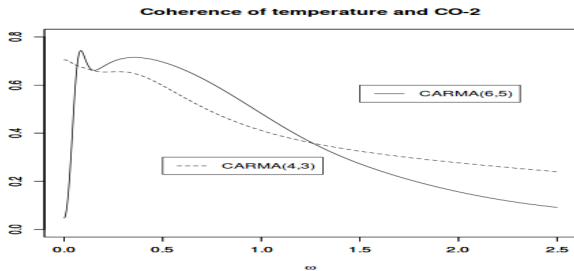
Hiti og CO_2

The temperature series consists of 5788 observations. The average time between measurements is 138 years and the standard deviations is 173 years. For the CO_2 we have 1095 observations. The average time between observations is 730 years and the standard deviations is 672 years. Both series were detrended with a regression line. The temperature series has virtually no trend, but there is a slight trend in the CO_2 series.

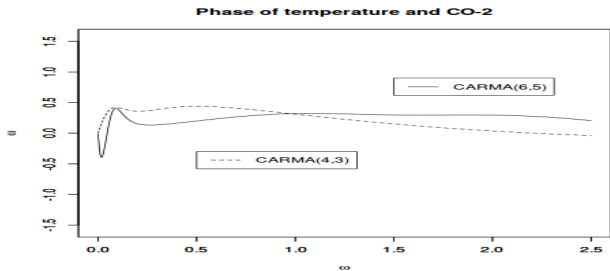




Mynd: A low frequency view of the scaled spectrum of temperature and CO₂.



Mynd: Coherence between temperature and CO_2 estimated with CARMA(4,3) and CARMA(6,5) models.



Mynd: Phase between temperature and CO₂ estimated with CARMA(4,3) and CARMA(6,5) models.

The main features of both series is the high spectrum at low frequency. Taking a closer look at the scaled estimated spectrum at low frequencies, figure 11 suggests that the variation of the CO_2 is more concentrated at lower frequencies than of the temperature. The peak of the CO_2 suggests a cycle of about 100.000 years. Similarly the figure suggests that the cycle of the temperature is about 80.000 years. Not only is the peak of the spectrum for CO_2 at a lower frequency, but a relatively larger share of the variance of the CO_2 seems due to frequencies close to zero. That might be a result of improper modeling of the trend in the CO_2 .

The most prominent features of the estimated multivariate model reflect those of the two univariate series. The high values of the spectrum are concentrated to the low frequencies. The coherence suggests some correlation at low frequencies, but the phase does not give a clear impression on which series is leading. Taking the slope of estimated phase at the frequency range, $0.5 < \omega < 2.5$ (cycle of 2.500 years to 12.500 years) suggests temperature lags 200 years (the slope is 0.2) based on the CARMA(4,3) and that CO_2 lags 2 (virtually zero) years based on the CARMA(6,5). At this frequency band the coherence is low so that the possible impact of one series on the other at those frequencies is little. The high coherency is concentrated at the very low frequencies. The estimated correlation of the innovations is 0.6 based on the CARMA(4,3) and 0.1 based on the CARMA(6,5).

The physical interpretation is that these data reveal very little information on an eventual causality in the relation between CO_2 and temperature.

- **Á heiðarlegri íslensku: Engin vísbending um lag eða causality milli þessara breyta. Innovationfylgni ca. 0.6.**

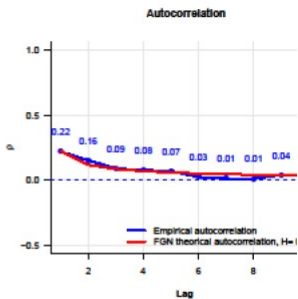
Hvað segja aðrir um hita og CO_2 ?

- Davidson et al. (2015) nota nákvæmlega sömu gögn en með discrete-tíma nálgun. Ályktunin er eins: Ekkert samband milli hita og CO_2 , en þeir gæta þess að móðga ekki bókstafstrúarfólkið með því að segja að ályktunin eigi ekki við um hækkun CO_2 af mannavöldum.
- Snögg skoðun með aðferðum Box-Jenkins, sbr. biskupakirkjuna/lífshættu, leiðir strax í ljós að tengslin geta ekki verið mjög sterk. Besta spámódel fyrir hita er t.d. ARIMA(0,1,3) og afgangslíðir hafa lítil tengsl við CO_2 .
- ARIMA ferli eru non-stationary og því hugsanlega eðlisfræðilega ófýsileg fyrir breytu eins og hita. Þetta segir mér að tal um co-integration eigi ekki við.
- Mills (2007) Veltir því fyrir sér hvernig eigi að taka á trend-strúktúrnum, smoothly changing trend eða long-memory process.
- Mills sér að ARIMA(p,1,q), þ.e. einhvers konar random-walk er ekki rýmilegt og að sjálfyfyllnifallið virðist deyja hægar en exponentielt út.
- Þess vegna stingur hann upp á ARFIMA, þ.e. ARIMA(p,d,q) þar sem d er ekki heiltala.
- Þetta eru kölluð fractionally-integrated líkön. Þau eru stundum kölluð long-memory líkön.

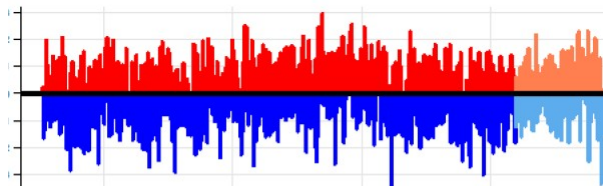
Aðeins um long-memory líkön

- Auto-variance minnkar hægar en exponentialt (sbr Pareto-tail). Standard reference er Beran (1994). Standard dæmið er flóðahæð í Níl.
- Dagsvik et al. (2015) nota hita gögn frá 96 veðurstöðvum í ca. 200 ár, meta fractal-gaussian-noise líkan (ARIMA(0,d,0)) og álykta að hiti sé stationary ferli. Þau nota líka ca. 2000 ára röð byggða á áhringjum trjáa.
- Hefðbundnari ARIMA/smooth-trend greiningu má t.d. sjá í Mills (2016).
- Sumir líkja umræðunni við deilur spálíkön versus stór ekónometrísk líkön: *A new paper published today by the Global Warming Policy Foundation explains how statistical forecasting methods can provide an important contrast to climate model-based predictions of future global warming. The repeated failures of economic models to generate accurate predictions has taught many economists a healthy scepticism about the ability of their own models, regardless of how complex, to provide reliable forecasts. Statistical forecasting has proven in many cases to be a superior alternative. Like the economy, the climate is a deeply complex system that defies simple representation. Climate modelling thus faces similar problems.*

Mynd frá Dagvik



Deviation from the mean



- *Unfortunately the experts gathered together for the IPCC's work did not include statistical experts. This has resulted in some potentially serious flaws in the statistical work of the IPCC which I will describe shortly. These flaws will tend to exaggerate the extent of future climate change.* Trewin (2008)
- Scott Armstrong stofnandi, Journal of Forecasting og International Journal of Forecasting er ekki impressed. Hann talar um anti-scientific political movement Armstrong et al. (2011).
- Meðal tímaraðafólks er (hefur verið?) í gangi veðmál, T1000.zip. Þ.e. teknar hafa verið hitatölur í 135 ár, trend sigtuð frá en skammtímasveiflur skildar eftir. Síðan eru simuleraðar 1000 raðir með sama sveiflustrúktúr. Síðan er trend sem er 1-gráða/öld, bætt við 1/3 af röðunum, 0 bætti við 1/3 og síðan -1 við 1/3. Síðan er keppni reynið að flokka raðirnar rétt, verðlaun 100.000 dollarar. (Observeruð hlýnun á öld er ca. 0.6 gráður).

Hvað hafa aðrir gert?

- R-pakkinn YUIMA metur ýmsar gerðir af stókastískum diffurjöfnum.
- Til dæmis hafa verið sett í hann ýmsar gerðir af ARMA módelnum með Lévy innvation processum, compound-poisson, variance-gamma, inverse-gaussian, ofl. Iacus & Mercuri (2015).
- Í YUIMA er líka hægt að vinna með fractionally-integrated/long-memory ferli. Ékki trivial því diffurjafnan er:

$$dY(t) = \alpha Y(t)dt + \sigma dW^H, \text{ og lausnin er}$$

$$Y(t) = Y(0) + \alpha \int_0^t Y(s)ds + W^H(t), \text{ H er Hurst index}$$

$$E(W^H(t)W^H(s)|W^H(0) = 0) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})$$

- Ný bók um YUIMA pakkann er Iacus & Yoshida (2018).
- Dæmi um hagnýtingar um orkumál eru: Garcia et al. (2011), stable-distribution ARMA og spot-markaður í Singapore. Lingohr (2019) notar svipaða aðferðafræði (Rcpp R o.s.frv) við líkanagerð um spot-markað og verð á endurnýjanlegri orku í Þýskalandi.

Hvað nú?

- Fullkomna `ctarmaRcpp` og `mctarmaRcpp` fyrir stationary ARMA og reyna að koma í dreifingu.
- Kannski einhvers konar hagnýtinga?
- Bæta við non-stationary atriðum, `rwnoise`. Margvíður non-sync pakki fyrir random-walk.
- Meiri bayesismi.
- Symbolic diffrun á log-likelihood falli.
- Verðbréfamarkaðir frábrugðnir orkumörkuðum.
- Einföld Box-Jenkins aðferðafræði (transfer-function) ekki fær í mörgum víddum af non-synchronous mælingum
- Spennandi reiknings- og túlkunarverkefni. En hafið í huga „*All models are wrong but some are useful*“ Box (1979).

TAKK FYRIR

- Armstrong, J. S., Green, K. C., & Soon, W. (2011). Research on forecasting for the manmade global warming alarm. *Energy & Environment*, 22(8), 1091–1104.
- Bartlett, M. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal the Royal Statistical Society, series B*, 8, 27–41.
- Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall.
- Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & E. G. N. Wilkinson (Eds.), *Robustness in Statistics*. Academic Press: New York.
- Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco.
- Dagsvik, J. K., Fortuna, M., & Moen, S. H. (2015). How does the temperature vary over time? evidence on the stationary and fractal nature of temperature fluctuations.
- Davidson, J. E., Stephenson, D. B., & Turasie, A. A. (2015). Time series modeling of paleoclimate data. *Environmetrics*, 27(1), 55–65.
- Garcia, I., Klüppelberg, C., & Müller, G. (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Statistical Modelling*, 11(5), 447–470.
- Granger, C. & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 38, 189–203.
- Iacus, S. M. & Mercuri, L. (2015). Implementation of lévy carma model in yuima package. *Comput. Stat.*, 30(4), 1111–1141.
- Iacus, S. M. & Yoshida, N. (2018). *Simulation and Inference for Stochastic Process with YUIMA*. Springer.

- Lingohr, D. (2019). Continuous-time autoregressive processes for modeling electricity prices and renewable energies. Master's thesis, University of Augsburg.
- Mills, T. C. (2007). Time series modelling of two millennia of northern hemisphere temperatures: Long memory or shifting trends? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 170(1), 83–94.
- Mills, T. C. (2016). Statistical forecasting: How fast will future warming be? GWPF Technical Paper 1, The Global Warming Policy Foundation.
- Oskarsson, H., Tomasson, K., Palsson, S., & Tomasson, H. (2019). Economic crises and incidence of suicide in iceland 1911-2017. *Læknablaðið*, 105(11), 483–488.
- Trewin, D. (2008). How reliable are climate change projections – a statistician's perspective. In *Statistics, Knowledge and Policy 2007 Measuring and Fostering the Progress of Societies*.
- Tsai, H. & Chan, K. (2000). A note on the covariance structure of a continuous-time ARMA process. *Statistica Sinica*, 10, 989–998.
- Yule, G. (1926). Why do we sometimes get nonsense-correlations between time-series? a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1), 1–63.
- Zadrozny, P. (1988). Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory*, 4(1), 108–124.