THE COINTEGRATED VECTOR AUTOREGRESSIVE MODEL WITH AN APPLICATION TO THE ANALYSIS OF SEA LEVEL AND TEMPERATURE

by Søren Johansen

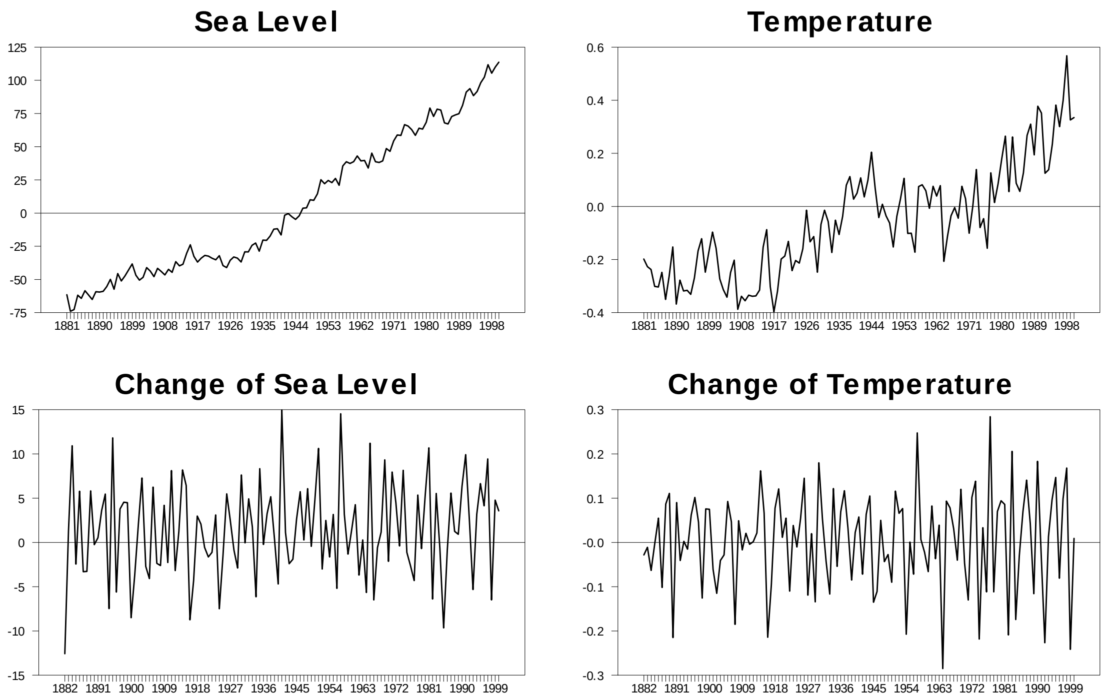
Department of Economics, University of Copenhagen
and
CREATES, Aarhus University

Content of Lectures

1. An Example: Sea level and temperature 1881-1995
2. A simple example and some literature
3. The vector autoregressive model and its solution
4. Hypotheses of interest in the I(1) model
5. The Statistical Analysis
6. The analysis of the data
7. Asymptotic Analysis
8. Conclusion

1. An Example: Sea level and temperature 1881-1995

Data in levels and differences



1.Data : 1881:01 to 1995:01, Hansen, J. et al *J. Geophys. Res. Atmos.* **106** 23947 (2001)

2. A SIMPLE EXAMPLE AND SOME LITERATURE

The mathematical formulation of cointegration by a simple example.
Take two stochastic processes which have a stochastic trend (random walk)

$$\begin{aligned} X_{1t} &= a\sum_{i=1}^{t}\varepsilon_{1i} + \varepsilon_{2t} \sim I(1) \\ X_{2t} &= b\sum_{i=1}^{t}\varepsilon_{1i} + \varepsilon_{3t} \sim I(1) \\ bX_{1t} - aX_{2t} &= b\varepsilon_{2t} - a\varepsilon_{3t} \sim I(0) \end{aligned}$$

1. $X(t)$ is nonstationary and ΔX_t is stationary: X_t is $I(1)$
2. $\beta' X_t$ is stationary with $\beta = (b, -a)'$
3. The common stochastic trend is $\sum_{i=1}^{t} \varepsilon_{1i}$

The cointegrated vector autoregressive model

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

Literature

1. Johansen S. (1996) ‘Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.’ Oxford University Press, Oxford. www.math.ku.dk/~sjo

2. Juselius, K. (2006) ‘The Cointegrated VAR Model: Methodology and Applications.’ Oxford University Press, Oxford. www.econ.ku.dk/~katarina.juselius

3. Dennis, J., Johansen, S. and Juselius, K. (2006), ‘CATS for RATS: Manual to Cointegration Analysis of Time Series.’ Estima, Illinois.

See also

Johansen, S. (2006) Cointegration: a survey. In: T.C. Mills and K. Patterson (eds.) *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*, Basingstoke, Palgrave Macmillan.

3. The vector autoregressive model and its solution

Error correction formulation of the vector autoregressive model

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

$$\Pi(z) = (1-z)I_p - \alpha\beta' z - \sum_{i=1}^{k-1}(1-z)z^i\Gamma_i$$

$$\det \Pi(1) = \det \alpha\beta' = 0 \Longrightarrow z = 1 \text{ is a root of } \det \Pi(z) = 0$$

Question: If the VAR has unit roots and the other roots larger than one, what is the moving average representation and what are the properties of the process?

$I(1)$ condition : $\det(\Pi(z)) = 0 \Longrightarrow z = 1$ or $|z| > 1$ and

$$\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i, \quad \text{det}(\alpha'_\perp \Gamma \beta_\perp) \neq 0$$

$$
\begin{aligned}
\Delta X_t &= \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \\
\Pi(z) &= (1-z)I_p - \alpha\beta' z - \sum_{i=1}^{k-1} (1-z) z^i \Gamma_i
\end{aligned}
$$

THEOREM If $\det(\Pi(z)) = 0 \Longrightarrow z = 1$ or $|z| > 1$ and det$(\alpha'_\perp \Gamma \beta_\perp) \neq 0$, then

$$
\begin{aligned}
\Pi(z)^{-1} &= C\frac{1}{1-z} + \sum_{i=0}^{\infty} C_i^* z^i \\
X_t &= C\sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{\infty} C_i^* \varepsilon_{t-i} + A, \quad \beta' A = 0 \\
C &= \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp
\end{aligned}
$$

1. X_t is nonstationary, ΔX_t is stationary: X_t is called $I(1)$
2. $\beta' X_t$ is stationary: X_t is cointegrating (r relations)
3. The common trends are $\alpha'_\perp \sum_{i=1}^{t} \varepsilon_i$ ($p - r$ trends)

Example

$$
\begin{aligned}
\Delta X_{1t} &= -\frac{1}{4}(X_{1t-1} - X_{2t-1}) + \varepsilon_{1t} \\
\Delta X_{2t} &= \frac{1}{4}(X_{1t-1} - X_{2t-1}) + \varepsilon_{2t}
\end{aligned}
$$

$$
\begin{aligned}
\Delta(X_{1t} - X_{2t}) &= -\frac{1}{2}(X_{1t-1} - X_{2t-1}) + \varepsilon_{1t} - \varepsilon_{2t} \Longrightarrow X_{1t} - X_{2t} \text{ stationary } (= y_t) \\
\Delta(X_{1t} + X_{2t}) &= \varepsilon_{1t} + \varepsilon_{2t} \Longrightarrow X_{1t} + X_{2t} \text{ nonstationary random walk } (= S_t)
\end{aligned}
$$

Granger Representation Theorem

$$
\begin{aligned}
X_{1t} &= \frac{1}{2}(S_t + y_t) \\
X_{2t} &= \frac{1}{2}(S_t - y_t)
\end{aligned}
$$

1. X_t is nonstationary, ΔX_t is stationary
2. $\beta' X_t$ is stationary with cointegration vector $\beta = (1, -1)'$
3. $\sum_{i=1}^{t}(\varepsilon_{1i} + \varepsilon_{2i})$ is a common trend

Two applications of the Granger Representation Theorem

1. The role of deterministic terms

$$\begin{aligned}
\Delta X_t &= \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \mu + \varepsilon_t \\
X_t &= C \sum_{i=1}^{t} (\varepsilon_i + \mu) + \sum_{i=0}^{\infty} C_i^* (\varepsilon_{t-i} + \mu) + A \\
X_t &= C \sum_{i=1}^{t} \varepsilon_i + C\mu t + \text{stationary process}
\end{aligned}$$

Thus

1. Linear trend in general
2. If $C\mu = \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}\alpha'_\perp \mu = 0$ (or $\alpha'_\perp \mu = 0)$: only constant term

Other deterministics.

2. Asymptotics

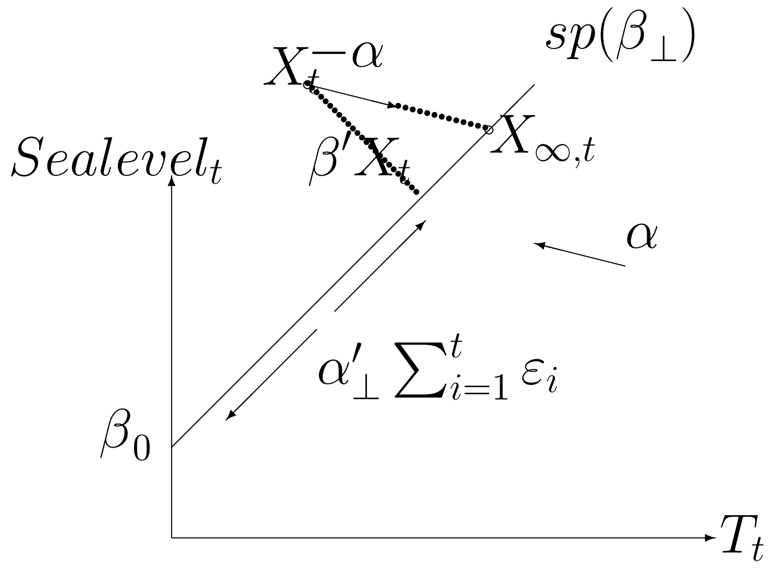
$$X_t = C\sum_{i=1}^{t}\varepsilon_i + C\mu t + \text{stationary process}$$

$$\frac{X_{[Tu]}}{\sqrt{T}} = C\frac{\sum_{i=1}^{[Tu]}\varepsilon_i}{\sqrt{T}} + C\mu\frac{[Tu]}{\sqrt{T}} + \frac{\text{stationary process}}{\sqrt{T}}$$

$$\downarrow d \qquad\qquad\qquad\qquad \downarrow P$$

$$CW(u) \qquad\qquad\qquad\qquad 0$$

The process $X_{[Tu]}$ is proportional to T in the direction $C\mu$, but orthogonal to this direction it behaves like $T^{1/2}$.



2.The process $X_t' = [SeaLevel_t, T_t]$ is pushed along the attractor set by the common trends and pulled towards the attractor set by the adjustment coefficients

4. Hypotheses of interest in the I(1) model

1. **Hypotheses on the rank**

$$\begin{aligned} \mathcal{H}_r \quad &: \quad \Delta X_t = \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Phi D_t + \varepsilon_t, \quad \alpha_{p\times r}, \beta_{p\times r} \\ \mathcal{H}_0 \; &\subset \; \ldots \subset \mathcal{H}_r \subset \ldots \subset \mathcal{H}_p \end{aligned}$$

2. **Hypotheses on the long run relations** β

3. **Hypotheses on** α

5. The statistical analysis

Test for misspecification of the VAR

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t$$

The VAR model assumes

1. Linear conditional mean explained by the past observations and deterministic terms (Check for unmodelled systematic variation, the choice of lag length, choice of information set (data), possible outliers, nonlinearity, **constant parameters**)

2. Constant conditional variance
(Check for ARCH effects, but also for regime shifts in the variance)

3. Independent Normal errors, mean zero, variance Ω
(Check for **lack of autocorrelation**, distributional form)

Estimation of the I(1) model

$$\mathcal{H}_r : \Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t,$$

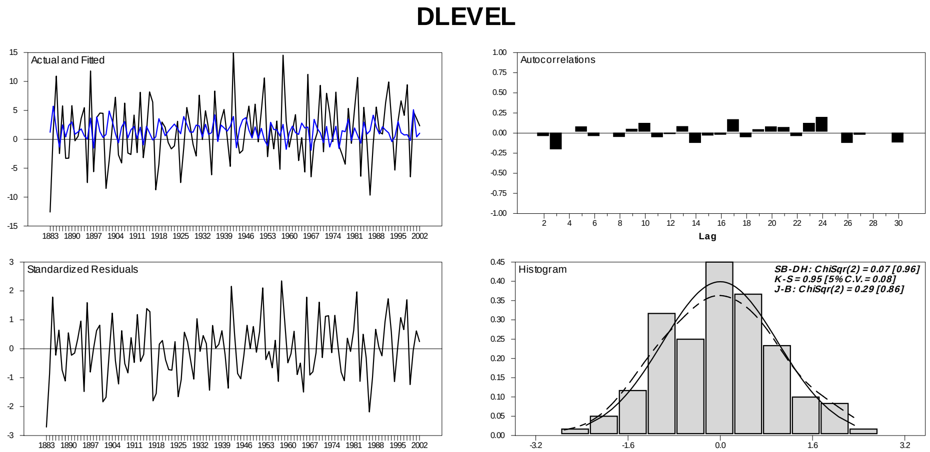
where ε_t i.i.d. $N_p(0, \Omega)$ and α and β are $(p \times r)$.
Maximum likelihood is calculated by reduced rank regression of ΔX_t on X_{t-1} corrected for

$$\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}, D_t$$

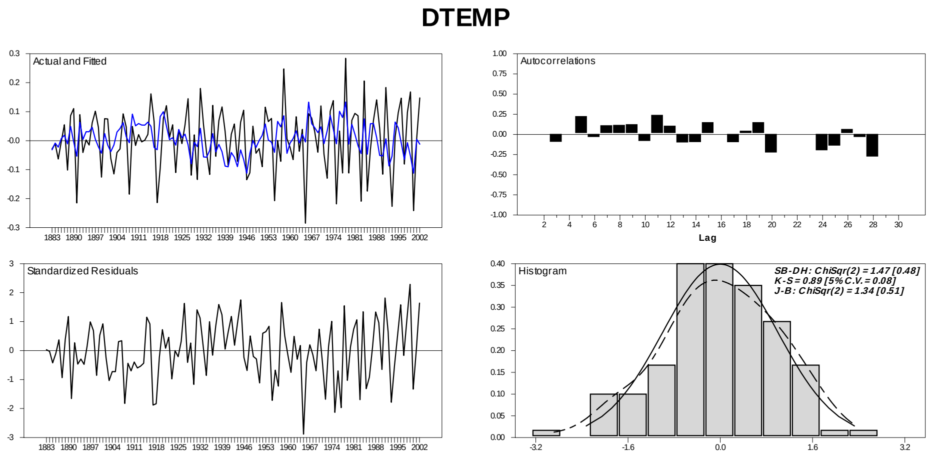
(T. W. Anderson, 1951). The estimate of β are the r linear combinations of the data which have the largest empirical correlation with the stationary process ΔX_t. (Canonical variates and canonical correlations)
If λ_i, are the squared canonical correlations, then

$$L_{\max}^{-2/T}(\mathcal{H}_r) \propto |\hat{\Omega}| \propto \prod_{i=1}^{r}(1 - \hat{\lambda}_i), \quad -2\log LR(\mathcal{H}_r|\mathcal{H}_p) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i)$$

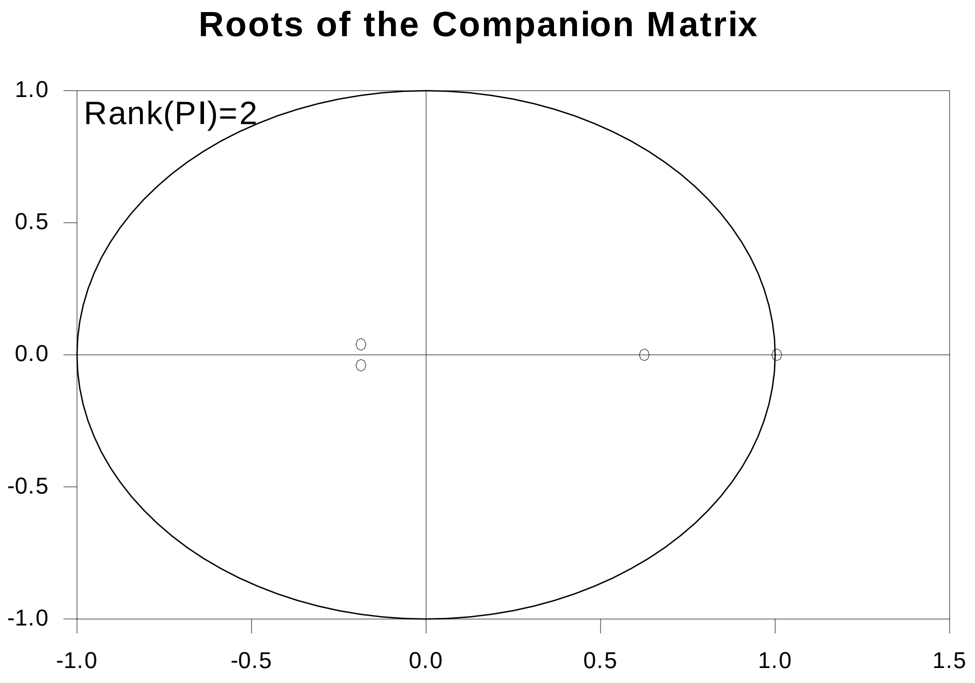
6. The analysis of the data

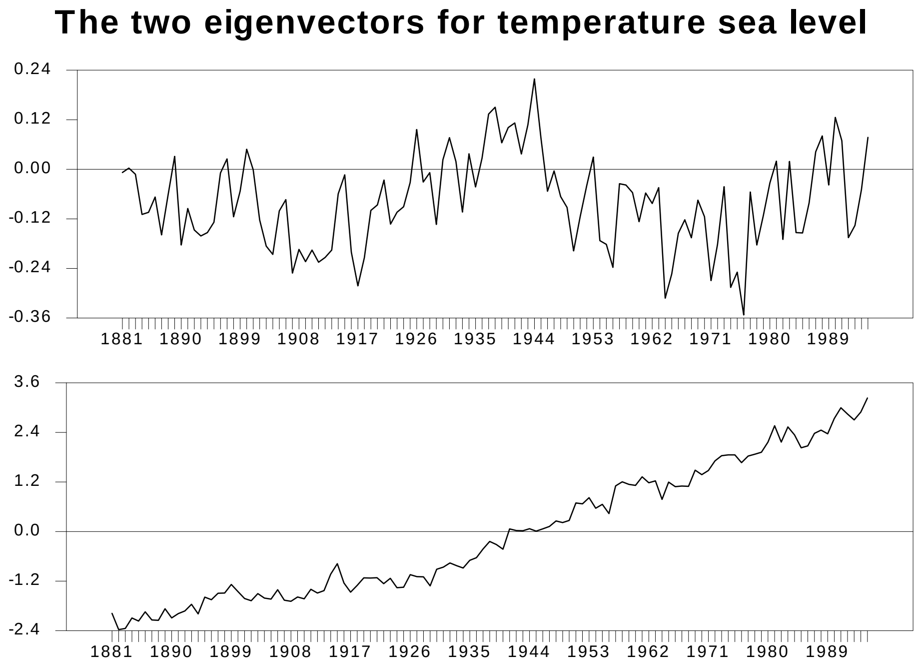


3.Residual analysis: Plot of $\Delta SeaLevel_t$, and fitted value, normalized residuals, auto-correlations function of residuals and histogram

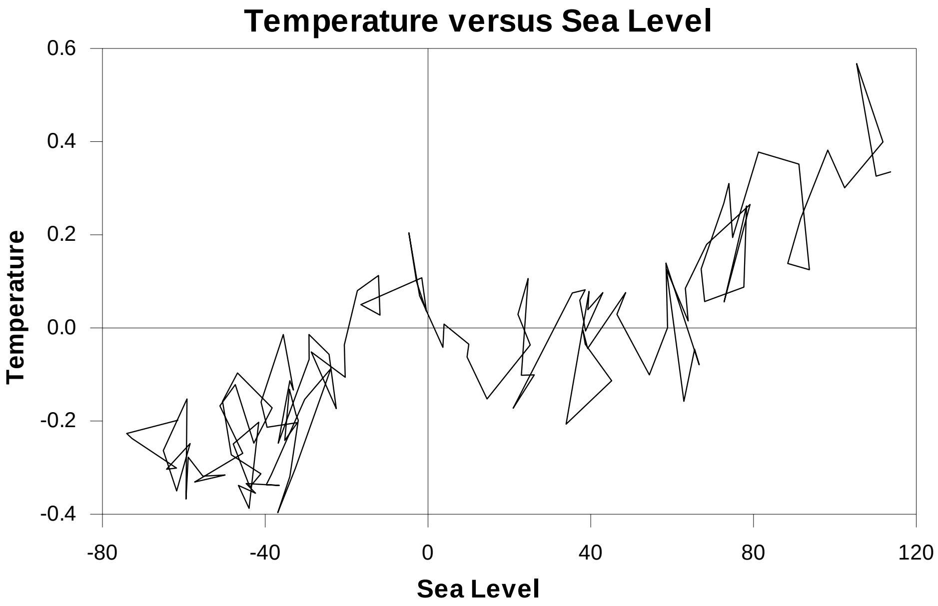


4.Residual analysis: Plot of $\Delta Temperature_t$, and fitted value, normalized residuals, autocorrelations function of residuals and histogram


Roots of the Companion Matrix
Rank(PI)=2
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0
1.5



5.The first canonical variate is $T_{t-1} - \underset{(t=-7.37)}{0.0031}\, h_{t-1}$



Temperature versus Sea Level
0.6
0.4
0.2
0.0
-0.2
-0.4
Temperature
-80
-40
0
40
80
120
Sea Level

Rank determination for temperature and sea level data.					
Rank $r=0$ is rejected and rank $r=1$ is accepted					
$p-r$	r	EigVal	$-2\log LR(\mathcal{H}_r \vert \mathcal{H}_p)$	95%Fract	p-val
2	0	0.168	20.76	15.41	0.005
1	1	0.003	0.36	3.84	0.54

The fitted ECM model for temperature and sea level data

$$\begin{aligned}
\Delta h_t &= \underset{(t=0.86)}{4.15}(T_{t-1} - \underset{(t=-7.37)}{0.0031}\, h_{t-1}) - \underset{(t=-3.11)}{0.2805}\,\Delta h_{t-1} + \underset{(t=0.60)}{3.04}\,\Delta T_{t-1} + \underset{(t=3.55)}{2.22} \\
\Delta T_t &= \underset{(t=-4.26)}{-0.40}(T_{t-1} - \underset{(t=-7.37)}{0.0031}\, h_{t-1}) - \underset{(t=-1.40)}{0.0024}\,\Delta h_{t-1} - \underset{(t=-0.54)}{0.053}\,\Delta T_{t-1} - \underset{(t=-1.91)}{0.023}
\end{aligned}$$

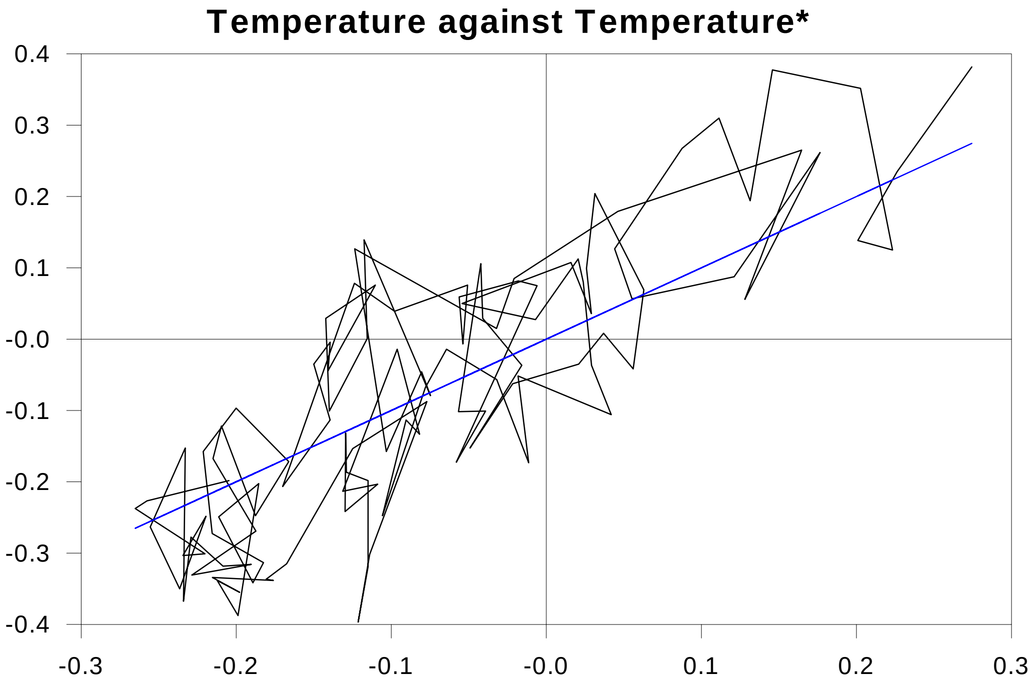
Note h_t weakly and strongly exogenous because the coefficients 4.15 to $(T_{t-1} - \underset{(t=-7.37)}{0.0031}\, h_{t-1})$ and 3.04 to ΔT_{t-1} are insignificant.

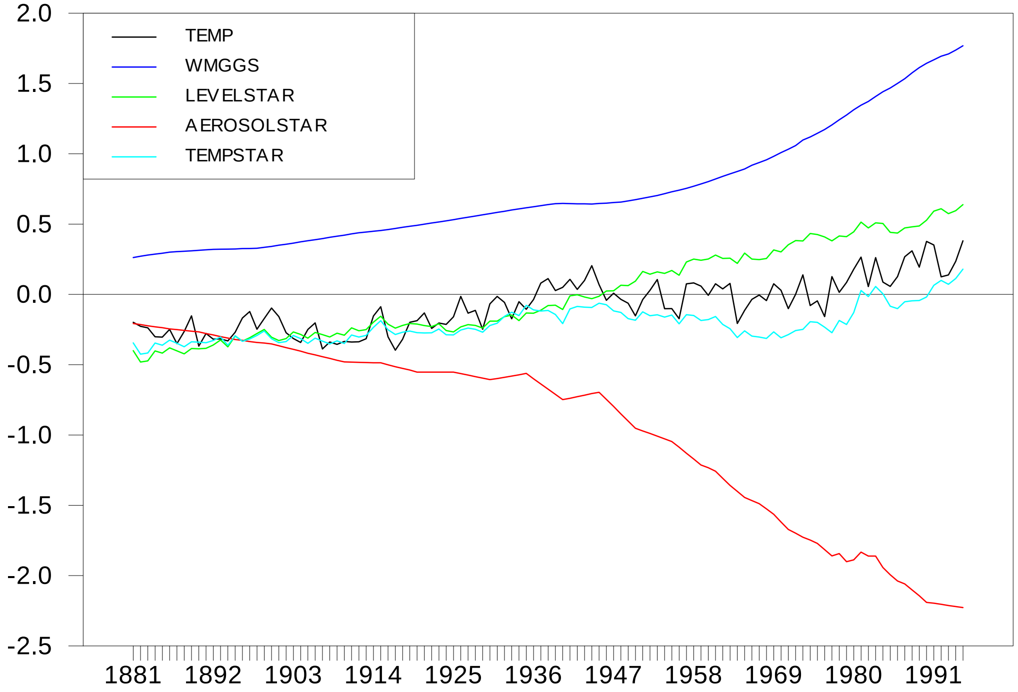
A partial model for (T_t, h_t) **conditional** on the forcing (weakly exogenous) variables $Wmgg\ (CO2\ and\ Methan)$ and $Aerosols\ (Sulphate)$

$$\Delta \begin{pmatrix} T_t \\ h_t \end{pmatrix} = \alpha \left(T + \gamma_1 h + \gamma_2 Wmgg + \gamma_3 Aerosol\right)_{t-1} + \cdots + \varepsilon_t$$

Cointegrating relation

$$\hat{\beta}' X_t = T_t - \underset{(t=-3.52)}{0.0065}\, h_t - \underset{(t=4.68)}{0.768}\, Wmgg_t - \underset{(t=3.51)}{1.478}\, Aerosol_t$$


Temperature against Temperature*
0.4
0.3
0.2
0.1
-0.0
-0.1
-0.2
-0.3
-0.4
-0.3
-0.2
-0.1
-0.0
0.1
0.2
0.3



6.Plot of the components of the fitted Temperature for simplified forcing variables

7. Asymptotic analysis

Asymptotic properties of the process X_t and its product moments for model without deterministic terms.
Three basic results

$$T^{-1/2}X_{[Tu]} = CT^{-1/2}\sum_{i=1}^{[Tu]}\varepsilon_i + T^{-1/2}\sum_{i=0}^{\infty}C_i^*\varepsilon_{t-i} \xrightarrow{d} CW(u)$$

$$T^{-2}\sum_{t=1}^{T}X_{t-1}X_{t-1}' \xrightarrow{d} \int_0^1 W(u)W(u)'du$$

$$T^{-1}\sum_{t=1}^{T}X_{t-1}\varepsilon_t' \xrightarrow{d} \int_0^1 W(dW)'$$

where W is Brownian motion with variance Ω.
Test for rank

$$-2\log LR(\mathcal{H}_r|\mathcal{H}_p) = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i) \xrightarrow{d} tr\{\int_0^1(dB)B'\left(\int_0^1 BB'\right)^{-1}\int_0^1 B(dB)'\}$$

where B is standard Brownian motion. Limit invariant to distribution of i.i.d. $(0,\Omega)$ errors but depends on the choice of deterministic terms

4. Asymptotic distribution of $\hat{\beta}$

Consider the model with no deterministic terms, $r = 2$, and β identified by

$$\beta = (\underset{p\times 1}{h_1} + \underset{p\times m_1}{H_1}\underset{m_1\times 1}{\varphi_1}, \underset{p\times 1}{h_2} + \underset{p\times m_2}{H_2}\underset{m_2\times 1}{\varphi_2}).$$

THEOREM If ε_t i.i.d. $(0, \Omega)$, then

$$T^{-1/2}x_{[Tu]} \xrightarrow{d} CW = G, \text{ and } T^{-1}S_{11} = T^{-2}\sum_{t=1}^{T} x_{t-1}x_{t-1}' \xrightarrow{d} C\int_0^1 WW'duC' = \mathcal{G}$$

$$T\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \rho_{11}H_1'\mathcal{G}H_1 & \rho_{12}H_1'\mathcal{G}H_2 \\ \rho_{21}H_2'\mathcal{G}H_1 & \rho_{22}H_2'\mathcal{G}H_2 \end{pmatrix}^{-1} \begin{pmatrix} H_1'\int_0^1 G(dV_1) \\ H_2'\int_0^1 G(dV_2) \end{pmatrix},$$

where $\rho_{ij} = \alpha_j'\Omega^{-1}\alpha_j$, and $V = \alpha'\Omega^{-1}W = (V_1, V_2)'$.
Note that $V = \alpha'\Omega^{-1}W$ is independent of $CW = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' W$. Hence limit is mixed Gaussian. The estimators of the remaining parameters are asymptotically Gaussian and asymptotically independent of $\hat{\varphi}$ and $\hat{\beta}$.

Example

The simplest example is a cointegrating regression

$$x_{1t} = \theta x_{2t-1} + \varepsilon_{1t}, \text{ and } \Delta x_{2t} = \varepsilon_{2t}$$

where ε_t are i.i.d. Gaussian and $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ are independent. Then

$$T(\hat{\theta} - \theta) = \frac{T^{-1}\sum_{t=1}^{T} x_{2t-1}\varepsilon_{1t}}{T^{-2}\sum_{t=1}^{T} x_{2t-1}^2} \xrightarrow{d} \frac{\int_0^1 W_2 dW_1}{\int_0^1 W_2^2(u)du} \approx \text{ Mixed Gaussian}$$

For given $\{x_{2t}, t = 1, \ldots, T\}$

$$\hat{\theta}|_{\{x_{2t}\}} \sim N(\theta, \frac{\sigma_1^2}{\sum_{t=1}^{T} x_{2t-1}^2}) \text{ and } \frac{(\hat{\theta} - \theta)}{\sigma_1}(\sum_{t=1}^{T} x_{2t-1}^2)^{1/2}|_{\{x_{2t}\}} \sim N(0, 1)$$

This implies that the marginal distributions satisfy

$$\frac{\hat{\theta} - \theta}{\sqrt{Var(\hat{\theta})}} \text{ not Gaussian and } \frac{\hat{\theta} - \theta}{\sigma_1}(\sum_{t=1}^{T} x_{2t-1}^2)^{1/2} \text{ is Gaussian}$$

The mixed Gaussian distribution.
A simple simulation of the model

$$\Delta X_{1t} = \alpha_1(X_{1t-1} - \gamma X_{2t-1}) + \varepsilon_{1t} \text{ and } \Delta X_{2t} = \varepsilon_{2t}$$

where ε_t are i.i.d. $N_2(0, diag(\sigma_1^2, \sigma_2^2))$. DGP $(\alpha_1 = -1, \gamma = 1, \sigma_1^2 = \sigma_2^2 = 1) : X_{1t} = X_{2t-1} + \varepsilon_{1t}, \Delta X_t = \varepsilon_{2t}$. MLE by regression:

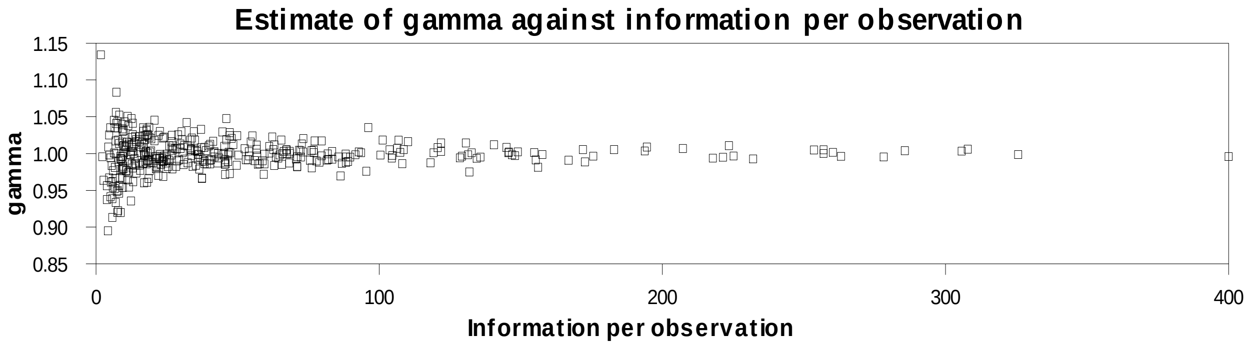
$$\Delta X_{1t} = \xi_1 X_{1t-1} + \xi_2 X_{2t-1} + \varepsilon_{1t}, \quad \hat{\alpha}_1 = \hat{\xi}_1, \hat{\gamma} = -\hat{\xi}_2/\hat{\xi}_1$$

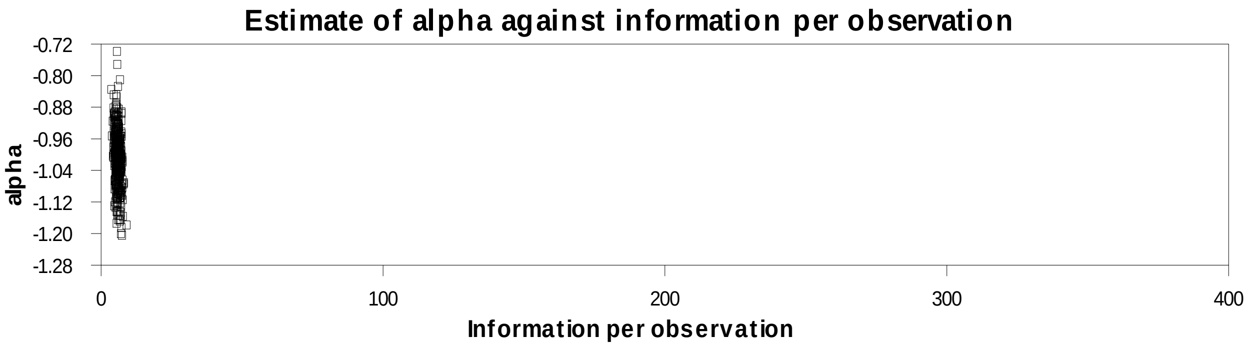
Asymptotic results for testing using the Wald test

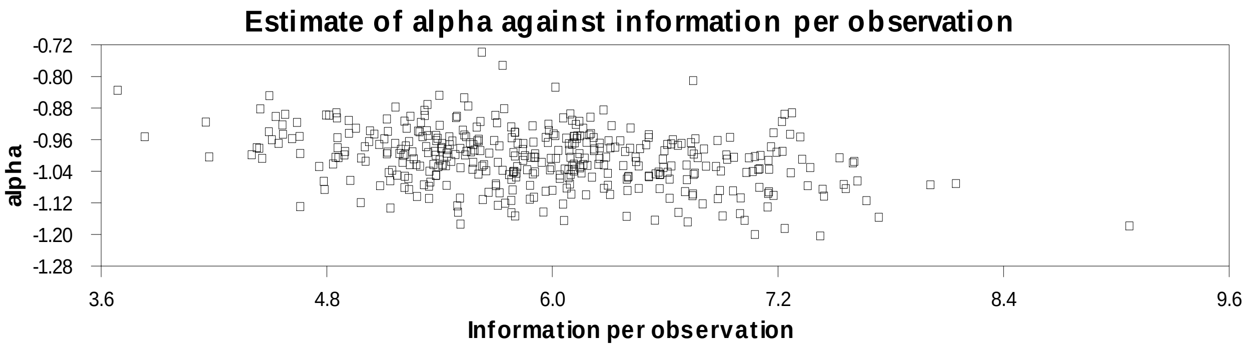
$$\frac{1}{\hat{\sigma}_1}(\sum_{t=1}^{T}(X_{1t-1} - \hat{\gamma}X_{2t-1})^2)^{1/2}(\hat{\alpha}_1 - \alpha_1) \xrightarrow{d} N(0,1); \text{ or } \frac{(1+\hat{\gamma}^2)^{1/2}}{\hat{\sigma}_1}T^{1/2}(\hat{\alpha} - \alpha)$$

$$\frac{|\hat{\alpha}_1|}{\hat{\sigma}_2}(\sum_{t=1}^{T} X_{2t-1}^2)^{1/2}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0,1), \text{ or } \frac{|\hat{\alpha}_1|}{\hat{\sigma}_2}(\int_0^1 W_2(u)^2 du)^{1/2} T(\hat{\gamma} - \gamma) \xrightarrow{d} N(0,1),$$

We need the distribution of $\hat{\alpha}_1$ and the joint distribution of $\hat{\gamma}, T^{-2}\sum_{t=1}^{T} X_{2t-1}^2$ to conduct inference.







7.From the model $\Delta x_{1t} = \alpha(x_{1t-1} - \gamma x_{2t-1}) + \varepsilon_{1t}, \quad \Delta x_t = \varepsilon_{2t}, T = 100, \quad SIM = 400$

8. Conclusion

When modelling stochastically trending variables, the usual regression methods need to be modified, in order to guarantee valid inference.

The vector autoregressive model allowing for $I(1)$ variables and cointegration has been analysed in detail, and the challenge is to apply it to nonstationary climate data, in order to avoid the fallacies involved in spurious correlation and regression.