A statistical analysis of time trends in atmospheric ethane

Marina Friedrich, E. Beutner, H. Reuvers, S. Smeekes, J.-P. Urbain

Climate Econometrics Seminar – 21.09.2021

Paper and co-authors

- Talk is based on the paper A statistical analysis of trends in atmospheric ethane, *Climatic Change* 162, 105-125, 2020

  `https://link.springer.com/article/10.1007/s10584-020-02806-2`

- Additional co-authors from the geo- and astrophysics community: W. Bader, B. Franco, B. Lejeune, E. Mahieu

- New co-author for the extension / multivariate analysis S.J. Koopman

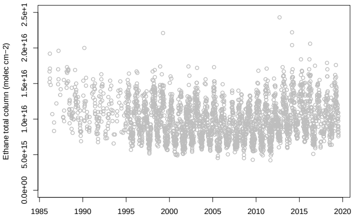
Outline

1. Motivation
2. The data
3. Trend analysis
   3.1 broken linear trend
   3.2 nonparametric trend
   3.3 inference on trend shapes
4. (Multivariate) Extensions
5. Conclusion

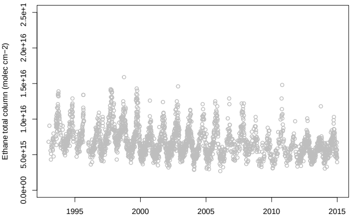
Motivation

- Trend analysis tool for atmospheric time series (ethane)
- What is ethane?
  - ✓ after methane, it is the most abundant hydrocarbon gas
  - ✓ from anthropogenic activities in Northern Hemisphere
  - ✓ useful indicator of atmospheric pollution
- Why study ethane?
  - ✓ used to measure anthropogenic methane emissions
  - ✓ indirect greenhouse gas, increasing lifetime of methane
  - ✓ contributes to the formation of 'bad' (ground-level) ozone
  - ✓ emitted during shale gas extraction

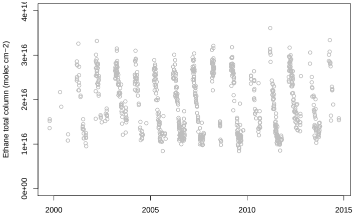
The data



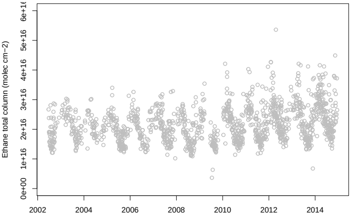
(a) Jungfraujoch (CH)



(b) Lauder (NZ)



(c) Thule (Greenland)



(d) Toronto (CA)

The data

- Northern Hemisphere
  - ✓ Jungfraujoch: 1986-2019, 2935 data points, 89.9 per year
  - ✓ Thule: 1999-2014, 814 data points, 54.4 per year
  - ✓ Toronto: 2002-2014, 1399 data points, 112.1 per year
- Southern Hemisphere
  - ✓ Lauder: 1992-2014, 2550 data points, 115.9 per year

The model

We consider the **general model**:

$$y_t = d_t + s_t + u_t,$$

where d_t is the long-run trend, our object of interest, and $u_t = \sigma_t v_t$ with v_t a linear process,

with **seasonal pattern**:

$$s_t = \sum_{j=1}^{S} a_j \cos(2j\pi t) + b_j \sin(2j\pi t),$$

and **missing observations**:

$$M_t = \begin{cases} 1 & \text{if } y_t \text{ is observed} \\ 0 & \text{if } y_t \text{ is missing} \end{cases} \qquad t = 1, \dots, T.$$

The model

We consider two **trend specifications**:

(1) A broken linear trend of the form

$$d_t = \alpha + \beta t + \delta D_{t,T_1},$$

where

$$D_{t,T_1} = \begin{cases} 0 & \text{if } t \leq T_1, \\ t - T_1 & \text{if } t > T_1. \end{cases}$$

(2) A nonparametric trend of the form

$$d_t = g(t/T), \qquad t = 1, ..., T,$$

where $g(\cdot)$ denotes a smooth (i.e. twice-differentiable) function defined on the unit interval.

Broken linear trend

We test the null hypothesis of no break vs. one break using

$$F_T = \min_{\alpha,\beta,s_t} \sum_{t=1}^{T} M_t \left(y_t - \alpha - \beta t - s_t\right)^2$$
$$- \inf_{T_c \in \Lambda} \min_{\alpha,\beta,\delta,s_t} \sum_{t=1}^{T} M_t \left(y_t - \alpha - \beta t - \delta D_{t,T_c} - s_t\right)^2,$$

as the *"usual"* test statistic (Bai and Perron, 1998) where for some $0 < \lambda < \frac{1}{2}$, we specify $\Lambda = [\lambda T, (1-\lambda)T]$.

We use the **autoregressive (AR) wild bootstrap** for

- critical values of break test
- confidence intervals around parameter estimates
- confidence intervals around break location

AR wild bootstrap – general idea

- The AR wild bootstrap is a modified version of the wild bootstrap which can handle autocorrelation
- We obtain bootstrap errors as $u_t^* = \xi_t^* \hat{u}_t$
- Usually, the ξ_t^*'s are i.i.d. random variables with $\mathbb{E}^*(\xi_t^*) = 0$ and $\mathbb{E}^*(\xi_t^*)^2 = 1$
- Here, the ξ_t^*'s are allowed to be dependent: they are generated by an AR(1) model
- The residuals are not resampled as in many other bootstrap methods which helps us keep the missing data pattern intact

AR wild bootstrap algorithm – break test

1. Calculate the following residuals, for $t = 1, ..., T$,

$$\hat{u}_t = M_t \left(y_t - \hat{\alpha} - \hat{\beta} t - \hat{s}_t \right).$$

2. Generate $\xi_t^* = \gamma \xi_{t-1}^* + \nu_t^*$ with $\nu_1^*, \ldots, \nu_n^*$ as i.i.d. $\mathcal{N}(0, 1 - \gamma^2)$.

3. Calculate the bootstrap errors $u_t^* = M_t \xi_t^* \hat{u}_t$ and generate the bootstrap sample as

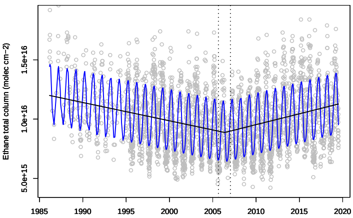
$$y_t^* = M_t \left(\hat{\alpha} + \hat{\beta} t + \hat{s}_t + u_t^* \right)$$

   for $t = 1, ..., T$.

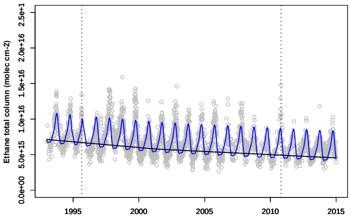
4. Obtain the break test statistic F_T^* from y_t^*.

5. Repeat Steps 2 to 4 B times.

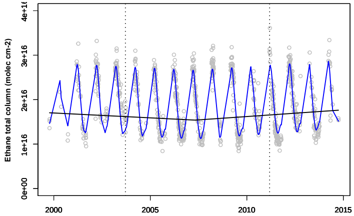
Results



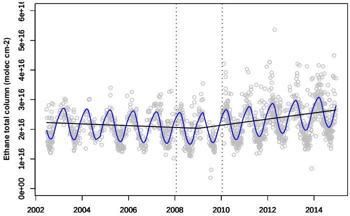
(e) Jungfraujoch



(f) Lauder



(g) Thule



(h) Toronto

Nonparametric trends

- To allow for more flexible trend shapes, we model the trend by

$$d_t = g(t/T), \qquad t = 1, ..., T,$$

as a smooth function of (rescaled) time.

- Estimation is done using a nonparametric kernel smoother (Nadaraya-Watson estimator).

- We use the AR wild bootstrap to construct (simultaneous) confidence intervals around the estimated trend.

Trend estimation

We focus on the **local constant** estimator for $\tau \in (0,1)$:

$$\hat{g}(\tau) = \arg\min_{g(\tau)} \sum_{t=1}^{T} K\left(\frac{t/T-\tau}{h}\right) M_t \{y_t - g(\tau)\}^2$$
$$= \left[\sum_{t=1}^{T} K\left(\frac{t/T-\tau}{h}\right) M_t\right]^{-1} \sum_{t=1}^{T} K\left(\frac{t/T-\tau}{h}\right) M_t y_t,$$

where $K(\cdot)$ is a kernel function and $h > 0$ is a bandwidth.

Presence of $\{M_t\}$ ensures that the estimator only depends on the actually observed data.

Data driven bandwidth selection

The bandwidth determines the smoothness of the trend estimate.

We consider a time series version of cross-validation, called **modified cross-validation** (Chu and Marron (1991)).

It is based on minimizing the criterion function $\frac{1}{T}\sum_{t=1}^{T} M_t \left(\hat{g}_{k,h}\left(\frac{t}{T}\right) - y_t\right)^2$ with respect to h, where

$$\hat{g}_{k,h}(\tau) = \frac{(T-2k-1)^{-1}\sum_{t:|t-\tau T|>k} K\left(\frac{t/T-\tau}{h}\right) M_t y_t}{(T-2k-1)^{-1}\sum_{t:|t-\tau T|>k} K\left(\frac{t/T-\tau}{h}\right) M_t}$$

is a leave-$(2k+1)$-out version of the leave-one-out estimator of ordinary cross-validation.

Simultaneous confidence bands

1. For all $\tau \in (0,1)$, obtain pointwise quantiles $\hat{q}_{\alpha_p/2}(\tau), \hat{q}_{1-\alpha_p/2}(\tau)$ for varying $\alpha_p \in [1/B, \alpha]$.

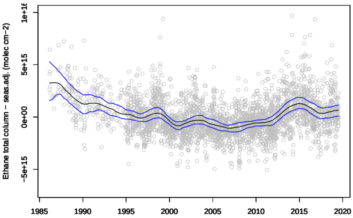
2. Choose α_s such that

$$\alpha_s = \underset{\alpha_p \in [1/B, \alpha]}{\arg\min} \left| \mathbb{P}^* \left[\hat{q}_{\alpha_p/2}(\tau) \leq \hat{g}^*(\tau) - \tilde{g}(\tau) \leq \hat{q}_{1-\alpha_p/2}(\tau) \right.\right.$$
$$\left.\left. \forall \tau \in (0,1) \right] - (1-\alpha) \right| .$$

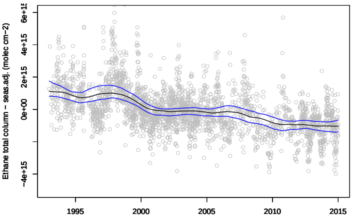
3. Construct the simultaneous confidence bands as

$$I_{n,\alpha}(\tau) = \left[\hat{g}(\tau) - \hat{q}_{1-\alpha_s/2}(\tau), \hat{g}(\tau) - \hat{q}_{\alpha_s/2}(\tau) \right] \qquad \tau \in (0,1) \,.$$

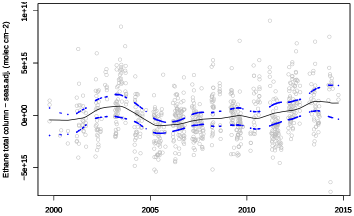
Results



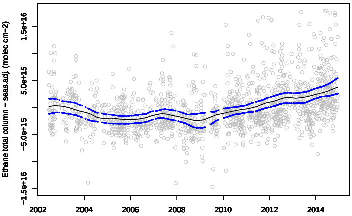
(i) Jungfraujoch



(j) Lauder



(k) Thule



(l) Toronto

Inference on trend shapes

- We analyze some features of the nonparametric trend estimates:
  - ✓ (location of local extrema)
  - ✓ specification test: linearity
  - ✓ (monotonicity)

- We construct confidence intervals around the minimum of the nonparametric trend for the NH series.

- We analyze the post-minimum upward trend and test whether we find evidence for linearity and monotonicity.

Bootstrap-based specification test

The test is based on the following null hypothesis

$$\mathsf{H}_0 : g(t) = g_0(\boldsymbol{\theta}, t) \quad \forall t \in \mathcal{G}_m = \{t_1, t_2, ..., t_m\},$$

where $g_0(\boldsymbol{\theta}, \cdot)$ belongs to a parametric family

$$\mathsf{G} = \{g(\boldsymbol{\theta}, \cdot); \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$$

We consider the linear trend function

$$g_0(t) = \alpha + \beta t$$

Under the alternative, the trend can be modeled by the nonparametric trend function $g(t/T)$.

Bootstrap-based specification test

The test statistic is

$$Q_t = \left(\hat{g}(t/n) - g_0(\widehat{\boldsymbol{\theta}}, t)\right)^2,$$

where $\hat{g}(t/n)$ denotes the nonparametric kernel estimator and $\widehat{\boldsymbol{\theta}}$ denotes the parameter estimates under the null hypothesis.

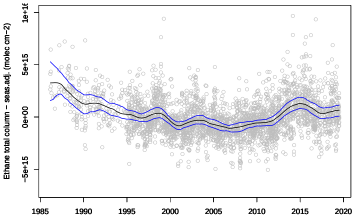
We consider the summary versions for the set $\mathcal{G}_m = \{t_1, t_2, ..., t_m\}$:

$$Q_{ave} = \frac{1}{m} \sum_{j=1}^{m} Q_{t_j}$$

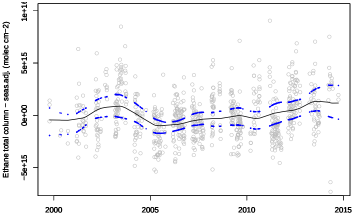
$$Q_{sup} = \sup_j Q_{t_j}.$$

We obtain critical values using the AR wild bootstrap.

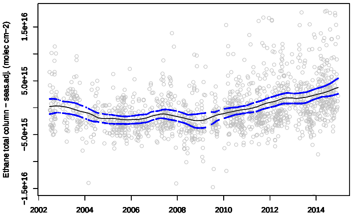
Results



(m) Jungfraujoch



(n) Thule



(o) Toronto

Results

- Jungfraujoch
  - ✓ Minimum in Nov. 2006 (cf. break in May 2006)
  - ✓ Linearity rejected
  - ✓ Monotonicity rejected
- Thule
  - ✓ Minimum in June 2005 (cf. break in April 2007)
  - ✓ Linearity not rejected
  - ✓ Monotonicity not rejected
- Toronto
  - ✓ Minimum in Oct. 2008 (cf. break in Dec. 2008)
  - ✓ Linearity not rejected
  - ✓ Monotonicity not rejected

Multivariate extensions

- Broken linear trends:
  - ✓ Locate a common break using approach by Kim (2011)
  - ✓ Direct extension of our univariate approach
  - ✓ Estimate a common break among NH ethane trends
  - ✓ Confidence intervals (CI) using extended ARW bootstrap
  - ✓ Break is located in 2008.47 (June 2008)
  - ✓ 95% CI: [2007.75; 2009.15] (Oct. 2007 - Feb. 2009)

- Smooth trends:
  - + Modeling smooth (common) trends with an unobserved components model
  - + Extracting separate common trend components from NH and SH series
  - + Testing for/locating common trend reversal patterns

Unobserved Components Time Series (UCTS) model

Classical time series decomposition: Trend + Seasonal + Irregular

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

where y_t is the *ethane* time series, with Trend μ_t, Seasonal γ_t and Noise ε_t, for $t = 1, \ldots, T$.

- UCTS model can represented in state space form
- Loglikelihood evaluation via prediction error decomposition and Kalman filter
- Maximum Likelihood estimation of parameters via numerical optimisation
- Signal extraction of trend, seasonal and irregular using Kalman filter and smoothing
- Diagnostic checking and testing based on prediction errors
- Software: OxMetrics/STAMP and TSL/State Space Edition

Unobserved Components Time Series model

Classical time series decomposition for ethane time series:

$$y_t = \mu_t + \gamma_t + \varepsilon_t,$$

for $t = 1, \ldots, T$ with signal of trend μ_t plus seasonal γ_t, and with noise ε_t.

The dynamic equations for trend and seasonal are

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma_\eta^2)$$

$$\beta_{t+1} = \beta_t + \zeta_t \qquad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2)$$

with trend μ_t, growth (or drift) β_t, disturbances η_t and ζ_t, and with seasonal $\gamma_t = \gamma_{1t}^* + \gamma_{2t}^*$ where γ_{it}^* is a time-varying i-yearly persistent cycle process, for $i = 1, 2$.

Multivariate Unobserved Components Time Series model

Multivariate time series decomposition with common trend:

$$y_{it} = \lambda_i \mu_t + \gamma_{it} + \varepsilon_{it},$$

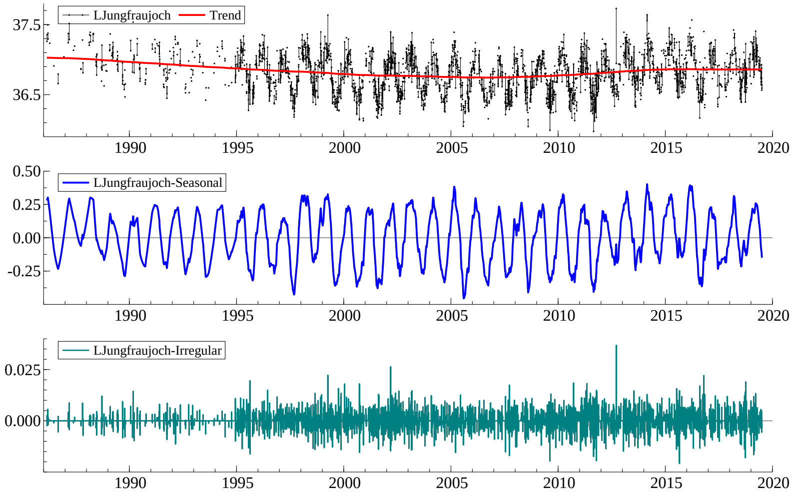
with $i \in \{\text{Jungfraujoch, Thule, Toronto}\}$, for $t = 1, \ldots, T$, and with loading λ_i, common trend μ_t, and the idiosyncratic seasonal γ_{it} and noise ε_{it} terms.

The common trend remains

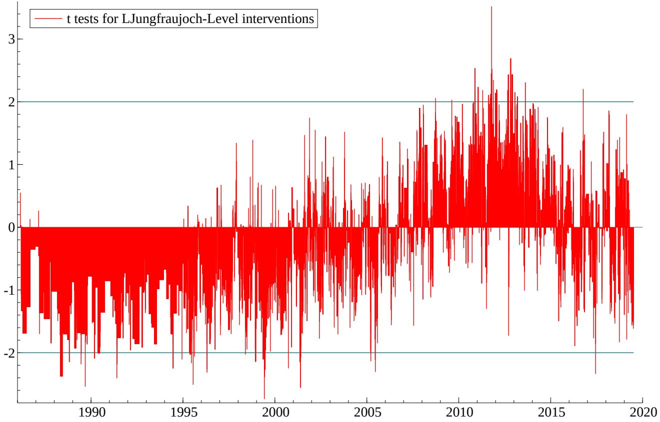
$$\mu_{t+1} = \mu_t + \beta_t + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma_\eta^2)$$

while breaks in trends can be detected from estimates of η_t.

Ethane Jungfraujoch (in logs): classical decomposition



Level break detection: sign t-test changes (June 2008)



Tests for $H_0: \ \delta = 0$ in $\mu_{t+1} = \mu_t + \beta_t + \delta x_t + \eta_t$ where

$$x_t = \begin{cases} 1 & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \qquad t = 1, \dots, T \quad \text{and} \quad \tau = 1, \dots, T.$$

Conclusion

- We propose a toolbox for flexible trend analysis to apply to atmospheric time series
- The challenge is to deal with missing data, autocorrelation, heteroskedasticity which can be achieved using the AR wild bootstrap
- We find a break in trend in all four time series of atmospheric ethane (NH: downward–upward, SH: downward–downward)
- Interesting extension: analyzing common trends


THANK YOU

References

- ✓ Chu, C.-K. and J. S. Marron (1991). Comparison of two bandwidths selectors with dependent errors. Annals of Statistics 19, 1906-1918.
- ✓ Franco, B., Bader W., Toon G.C., Bray C., Perrin A., Fischer E.V., Sudo K., Boone C.D., Bovya B., Lejeune B., Servais C. and E. Mahieu (2015). Retrieval of ethane from ground-based FTIR solar spectra using improved spectroscopy: Recent burden increase above Jungfraujoch. Journal of Quantitative Spectroscopy and Radiative Transfer 160, 36-49.
- ✓ Friedrich, M., Smeekes, J.-P. and J.-P. Urbain (2020). Autoregressive wild bootstrap inference for nonparametric trends. Journal of Econometrics 214, 81-109.
- ✓ Kim, D. (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. Journal of Econometrics 164, 310-330.