The Trendiness of Trends

Andreas Kryger Jensen (joint with Claus Ekstrøm)

April 9$^{\text{th}}$ 2019, DSTS meeting

Biostatistics, Institute of Public Health
University of Copenhagen

The problem

Changes in trends

A statement often seen in the news is that at this very moment we see a significant change in the trend of something.

„The trend has broken"

Recent examples from the Danish public news:

- Proportion of injuries from fireworks on New Year's Eve
- Proportion of children being baptized
- Average price of a one-family house
- Proportion of smokers

Fundamental questions

It is trendy to talk about changes in trends.

But...

- What is a trend?
- What is a change in a trend?
- What is the „trendiness” of a trend?
- How many times has the trend changed?

Can we quantify and estimate answers to these questions?

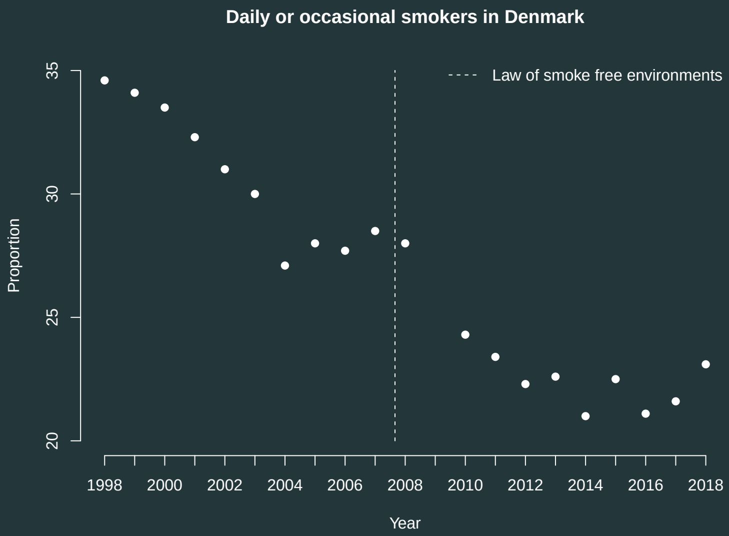
Politiken, January 02, 2019

Første gang i 20 år: Flere danskere ryger

Prisen på cigaretter skal op. Sådan siger fagfolk på baggrund af årets undersøgelse af rygevaner, der viser, at andelen af rygere er steget markant på to år.

For første gang i to årtier er der en signifikant stigning i andelen af danskere, der ryger. Det viser den årlige undersøgelse af rygevaner blandt 5.017 danskere, som Sundhedsstyrelsen og tre organisationer offentliggør torsdag.

Proportion of smokers in Denmark



Data analysis objectives

Focusing on the proportion of smokers in Denmark we wish to:

1. Quantify the certainty by which the proportion of smokers is increasing in 2018.
2. Find out when the proportion started to increase (if it is currently increasing).
3. Assess if it is the first time in 20 years that the proportion has increased.

Six definitions

Definition 1
Reality evolves in continuous time $t \in \mathcal{T} \subset \mathbb{R}$.

Definition 2
There exists a latent function $f = \{f(t) : t \in \mathcal{T}\}$ governing the time evolution of some outcome. We observe random variables sampled from f at discrete time points with additive measurement noise

$$Y_i = f(t_i) + \varepsilon_i, \quad t_i \in \mathcal{T}, \quad \mathbb{E}[\varepsilon_i \mid t_i] = 0$$

We wish to understanding the dynamics of f conditional on observed data $(Y_i, t_i)_{i=1}^n$.

Definition 3
The trend of f is its instantaneous slope

$$df(t) = \left(\frac{\mathrm{d}f(s)}{\mathrm{d}s}\right)(t)$$

- $df(t) > 0$: f is increasing at t and has a positive trend
- $df(t) < 0$: f is decreasing at t and has a negative trend

Definition 4

Definition 4
A change in trend of f is when df changes sign.

- f goes from increasing to decreasing or vice versa
- The trend, df, goes from positive to negative or vice versa

Definition 4
A change in trend of f is when df changes sign.

- f goes from increasing to decreasing or vice versa
- The trend, df, goes from positive to negative or vice versa

A change in trend does not occur at a unique point in time.

The question is:

> „*What is the probability that the trend is changing at t given everything we know?*"

Definition 5
The Trend Direction Index (called Teddy) is the conditional probability

$$\mathrm{TDI}(t, \delta) = P(df(t + \delta) > 0 \mid \mathcal{F}_t)$$

where $\mathcal{F}_t$ is the sigma algebra of all available information up until time t.

Definition 5
The Trend Direction Index (called Teddy) is the conditional probability

$$\mathrm{TDI}(t, \delta) = P(df(t+\delta) > 0 \mid \mathcal{F}_t)$$

where $\mathcal{F}_t$ is the sigma algebra of all available information up until time t.

- TDI is a local probability.
- $\delta \leq 0$ is estimation. $\delta > 0$ is forecasting.
- Alternatively: $df(t+\delta) < 0$ but that is just $1 - \mathrm{TDI}(t, \delta)$.
- Common example: $t = \text{now}$ and $\delta = 0$ (change-point models are impossible).

Definition 6
The Expected Trend Instability (called Eddy) on an interval $\mathcal{I}$ is

$$\mathrm{ETI}(\mathcal{I}) = \mathbb{E}[\#\{t \in \mathcal{I} : df(t) = 0\} \mid \mathcal{F}]$$

The value is equal to:

- The expected number of trend changes in $\mathcal{I}$
- The expected number of zero-crossings by df in $\mathcal{I}$

ETI is a global measure and the lower ETI is, the more stable the trend is on an interval.

A Functional Data Approach for Assessing the Trendiness of Trends

Teddy and Eddy

We wish to estimate Teddy and Eddy from data. This requires:

1. A probabilistic model linking observed data to the latent function f.
2. The distribution of df conditional on data (for Teddy).
3. The joint conditional distribution of (df, d^2f) (for Eddy).

Using latent Gaussian Processes we can easily derive these and estimate them from data.

Latent Gaussian Process model

Gaussian Process

A random function $\{f(t) : t \in \mathcal{T}\}$ is a Gaussian Process if and only if $(f(t_1), \ldots, f(t_n))$ is multivariate normal distributed for every $(t_1, \ldots, t_n) \subset \mathcal{T}$ with $n < \infty$.

We write $f \sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))$ where $\mu\colon \mathcal{T} \mapsto \mathbb{R}$ and $C\colon \mathcal{T} \times \mathcal{T} \mapsto \mathbb{R}$ are the mean and covariance functions.

Gaussian Process

A random function $\{f(t) : t \in \mathcal{T}\}$ is a Gaussian Process if and only if $(f(t_1), \ldots, f(t_n))$ is multivariate normal distributed for every $(t_1, \ldots, t_n) \subset \mathcal{T}$ with $n < \infty$.

We write $f \sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))$ where $\mu\colon \mathcal{T} \mapsto \mathbb{R}$ and $C\colon \mathcal{T} \times \mathcal{T} \mapsto \mathbb{R}$ are the mean and covariance functions.

We assume that the observed data is generated by the following hierarchical model:

$$f \mid \beta, \theta \sim \mathcal{GP}(\mu_\beta(\cdot), C_\theta(\cdot, \cdot))$$

$$Y_i \mid f(t_i), \Theta, t_i \overset{iid}{\sim} N(f(t_i), \sigma^2), \quad \Theta = (\beta, \theta, \sigma^2)$$

Joint distribution

Let $\mathbf{Y}$ be the vector of observed data at time points $\mathbf{t}$ and $\mathbf{t}^*$ any finite subset of $\mathcal{T}$.

The model implies the following joint distribution:

$$\begin{bmatrix} f(\mathbf{t}^*) \\ df(\mathbf{t}^*) \\ \mathbf{Y} \end{bmatrix} \mid \mathbf{t}, \Theta \sim N\left(\begin{bmatrix} \mu_\beta(\mathbf{t}^*) \\ d\mu_\beta(\mathbf{t}^*) \\ \mu_\beta(\mathbf{t}) \end{bmatrix}, \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \\ C_\theta(\mathbf{t}, \mathbf{t}^*) & \partial_2 C_\theta(\mathbf{t}, \mathbf{t}^*) & C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \end{bmatrix} \right)$$

- $\mathbf{t}^*$ can be any finite vector in continuous time
- Directly extendable to include $d^2 f$ (required for ETI)
- (sample path regularity conditions are implicitly assumed)

Joint posterior distribution

The joint posterior distribution of the latent functions is

$$\begin{bmatrix} f(\mathbf{t}^*) \\ df(\mathbf{t}^*) \end{bmatrix} \mid \mathbf{Y}, \mathbf{t}, \Theta \sim N\left(\begin{bmatrix} \mu_f(\mathbf{t}^*) \\ \mu_{df}(\mathbf{t}^*) \end{bmatrix}, \begin{bmatrix} \Sigma_{f,f}(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*) \\ \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*)^T & \Sigma_{df,df}(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix} \right)$$

$$\begin{aligned}
\mu_f(\mathbf{t}^*) &= \mu_\beta(\mathbf{t}^*) + C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right]^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\mu_{df}(\mathbf{t}^*) &= d\mu_\beta(\mathbf{t}^*) + \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right]^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\Sigma_{f,f}(\mathbf{t}^*, \mathbf{t}^*) &= C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right]^{-1} C_\theta(\mathbf{t}^*, \mathbf{t})^T \\
\Sigma_{df,df}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right]^{-1} \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t})^T \\
\Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right]^{-1} \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t})^T
\end{aligned}$$

Trend Direction Index

Recall the definition of the Trend Direction Index

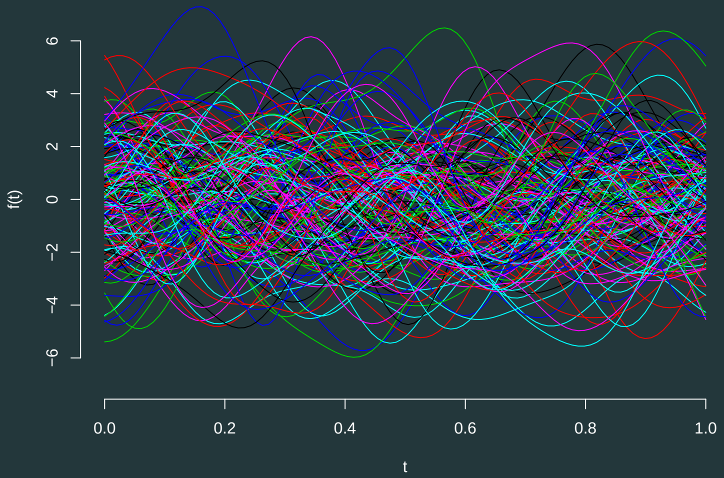
$$\mathrm{TDI}(t, \delta) = P(df(t+\delta) > 0 \mid \mathcal{F}_t)$$

Letting $\mathcal{F}_t = \sigma\left\{\mathbf{Y}, \mathbf{t}\right\}$, we may express TDI through the posterior of df as

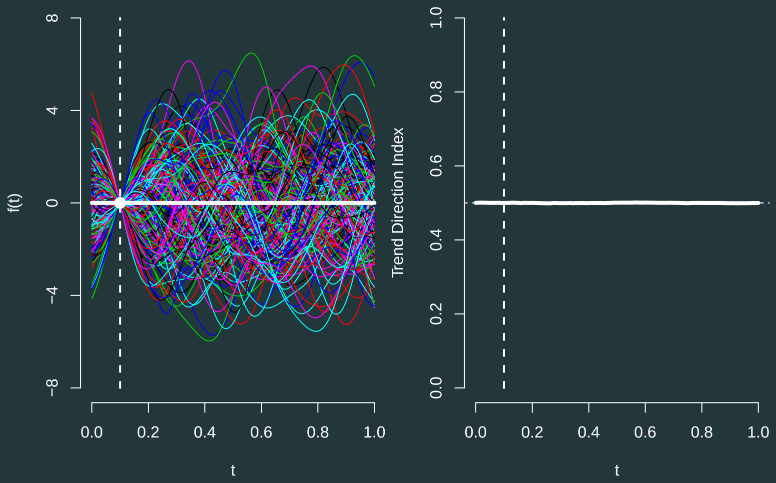
$$\begin{aligned}\mathrm{TDI}(t, \delta \mid \Theta) &= P\left(df(t+\delta) > 0 \mid \mathbf{Y}, \mathbf{t}, \Theta\right) \\ &= \int_0^\infty N\left(u, \mu_{df}(t+\delta), \Sigma_{df,df}(t+\delta, t+\delta)^{1/2}\right) \mathrm{d}u\end{aligned}$$

Example - prior distribution of f

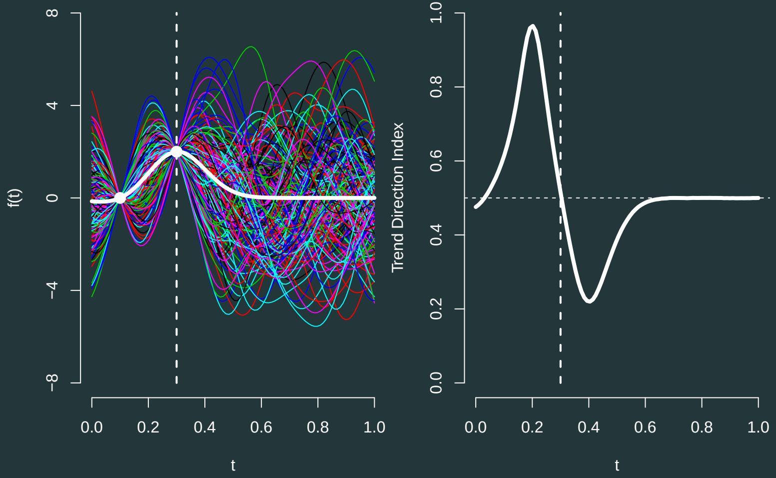
What the world could look like with eyes closed



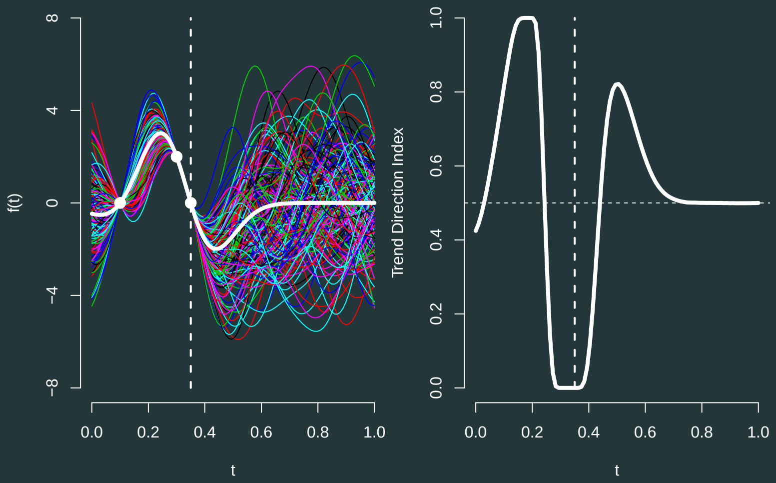
Example - posterior distribution of f and TDI



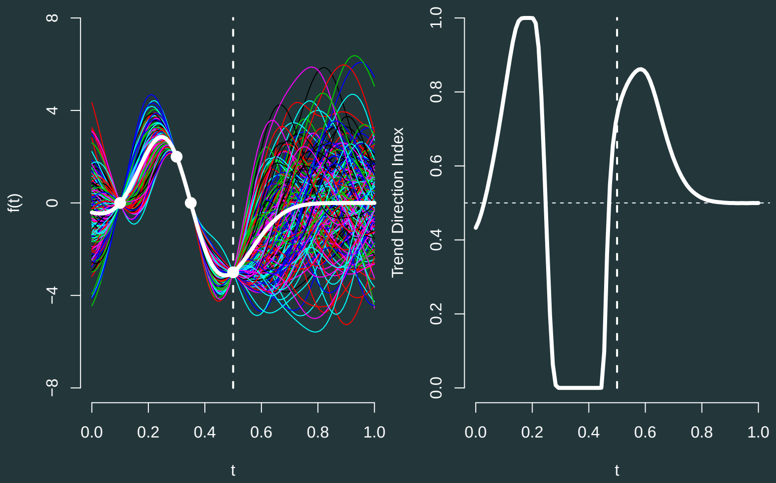
Example - posterior distribution of f and TDI



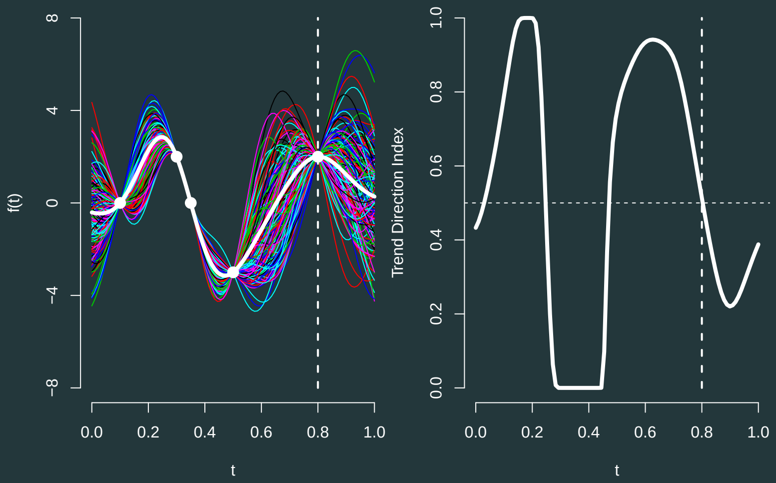
Example - posterior distribution of f and TDI



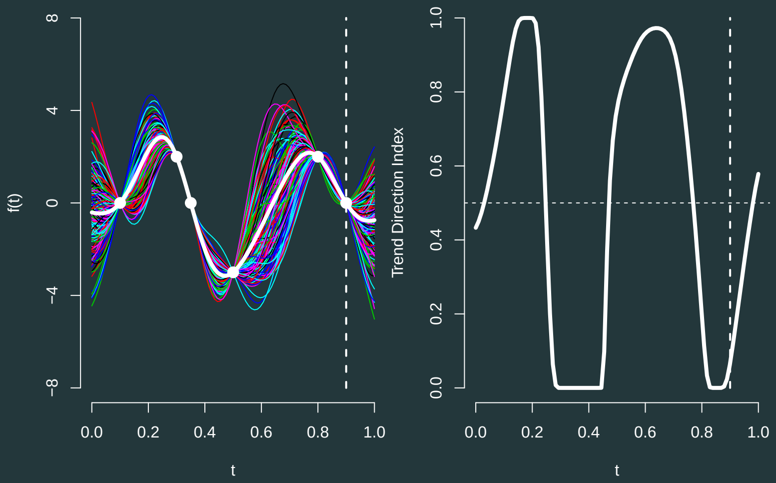
Example - posterior distribution of f and TDI



Example - posterior distribution of f and TDI



Example - posterior distribution of f and TDI



Expected Trend Instability

Recall the definition of the Expected Trend Instability

$$\mathrm{ETI}(\mathcal{I}) = \mathbb{E}[\#\{t \in \mathcal{I} : df(t) = 0\} \mid \mathcal{F}]$$

Rice (1944) showed that the expected number of 0-crossings of a process X under suitable regularity conditions is equal to

$$\mathrm{ETI}(\mathcal{I}) = \int_{\mathcal{I}} \int_{-\infty}^{\infty} |v| f_{X(t), dX(t)}(0, v) \mathrm{d}v \mathrm{d}t$$

where the integrand is the local crossing intensity.

Expected Trend Instability

Recall the definition of the Expected Trend Instability

$$\mathrm{ETI}(\mathcal{I}) = \mathbb{E}[\#\{t \in \mathcal{I} : df(t) = 0\} \mid \mathcal{F}]$$

Rice (1944) showed that the expected number of 0-crossings of a process X under suitable regularity conditions is equal to

$$\mathrm{ETI}(\mathcal{I}) = \int_{\mathcal{I}} \int_{-\infty}^{\infty} |v| f_{X(t), dX(t)}(0, v) \mathrm{d}v \mathrm{d}t$$

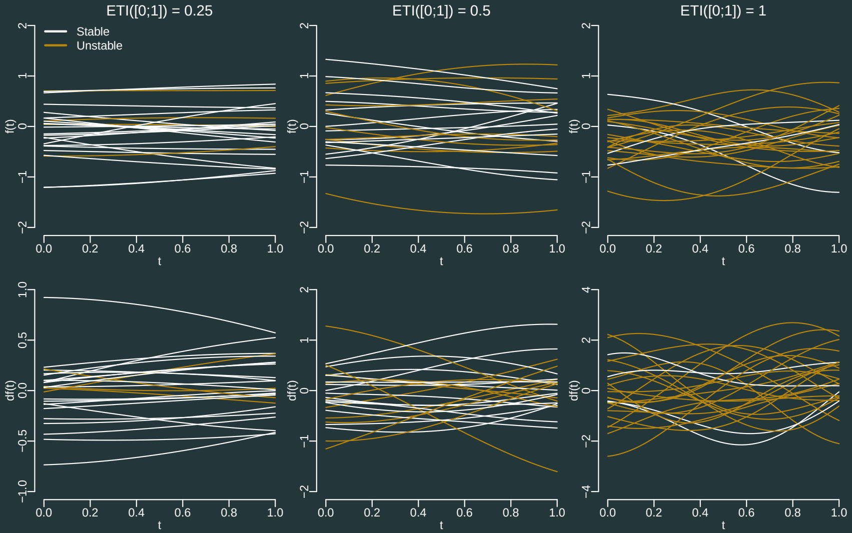
where the integrand is the local crossing intensity.

We need to calculate this for the posterior distribution of (df, d^2f). We obtain:

$$\mathrm{ETI}(\mathcal{I} \mid \Theta) = \int_{\mathcal{I}} \frac{\Sigma_{d^2f}(t,t)^{1/2}}{\Sigma_{df}(t,t)^{1/2}} \sqrt{1 - \omega(t)^2} \phi\left(\frac{\mu_{df}(t)}{\Sigma_{df}(t,t)^{1/2}}\right) \left[2\phi(\zeta(t)) + \zeta(t) \operatorname{Erf}\left(\frac{\zeta(t)}{\sqrt{2}}\right)\right] \mathrm{d}t$$

$$\omega(t) = \frac{\Sigma_{df,d^2f}(t,t)}{\Sigma_{df}(t,t)^{1/2} \Sigma_{d^2f}(t,t)^{1/2}}, \quad \zeta(t) = \frac{\mu_{d^2f}(t) - \mu_{df}(t) \Sigma_{d^2f}(t,t)^{1/2} \omega(t) \Sigma_{df}(t,t)^{-1/2}}{\Sigma_{d^2f}(t,t)^{1/2} (1 - \omega(t)^2)^{1/2}}$$

Example - Expected Trend Instability



Estimation

Maximum Marginal Likelihood

The marginal likelihood is analytically available by integrating out the latent process

$$L(\Theta \mid \mathbf{Y}, \mathbf{t}) = N(\mathbf{Y}; \mu_\beta(\mathbf{t}), C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)$$

leading to the estimate $\widehat{\Theta} = \arg\sup_\Theta L(\Theta \mid \mathbf{Y}, \mathbf{t})$.

Maximum Marginal Likelihood

The marginal likelihood is analytically available by integrating out the latent process

$$L(\Theta \mid \mathbf{Y}, \mathbf{t}) = N(\mathbf{Y}; \mu_\beta(\mathbf{t}), C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)$$

leading to the estimate $\widehat{\Theta} = \arg\sup_\Theta L(\Theta \mid \mathbf{Y}, \mathbf{t})$.

This gives us the point estimates of Eddy and Teddy:

$$\mathrm{TDI}(t, \delta \mid \widehat{\Theta}), \quad \mathrm{ETI}(\mathcal{T} \mid \widehat{\Theta})$$

But it is difficult to obtain their distributions which are required in order to assess their uncertainties.

Bayesian estimation

Another approach is a fully Bayesian model with a prior distribution on the parameters of the latent function, $\Theta \sim G$.

This enables posterior distributions of $\mathrm{TDI}(t, \delta \mid \Theta)$ and $\mathrm{ETI}(\mathcal{T} \mid \Theta)$ using Markov-chain Monte Carlo simulation.

We have implemented the model in Stan.

Bayesian estimation

Another approach is a fully Bayesian model with a prior distribution on the parameters of the latent function, $\Theta \sim G$.

This enables posterior distributions of $\mathrm{TDI}(t, \delta \mid \Theta)$ and $\mathrm{ETI}(\mathcal{T} \mid \Theta)$ using Markov-chain Monte Carlo simulation.

We have implemented the model in Stan.

We use moderately informative priors for Θ centered at the marginal maximum likelihood estimates but with heavy tails and large variances.

We ran 4 chains for 25,000 iterations (half for warm-up).

Application

From the Danish Health Authority[1]

Danskernes rygevaner 2018

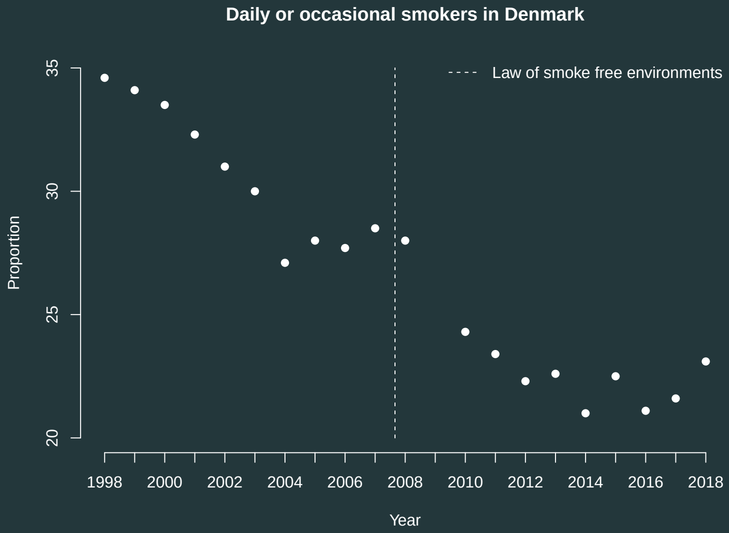
Oprettet 3. januar 2019

Hjerteforeningen, Sundhedsstyrelsen, Kræftens Bekæmpelse og Lungeforeningen har undersøgt danskernes rygevaner. Undersøgelsen er baseret på et repræsentativt udsnit af danskerne i forhold til alder, køn, religion og uddannelse. I alt har 5.017 danskere deltaget i undersøgelsen. Dataindsamlingen er foretaget af TNS Gallup via deres internetpanel i november 2018.

I denne undersøgelse er studiepopulationen imidlertid så stor, at stort set alle forskellene bliver statistisk signifikant på et 95% signifikansniveau.

[1] Nøgletal - Danskernes rygevaner 2018, www.sst.dk/da/udgivelser/2019/danskernes-rygevaner-2018

Proportion of smokers in Denmark



Simple analysis - Has the proportion in 2018 changed?

A simple analysis is to test whether the proportion of smokers in 2018 has changed compared to a previous year. Just a χ^2-test.

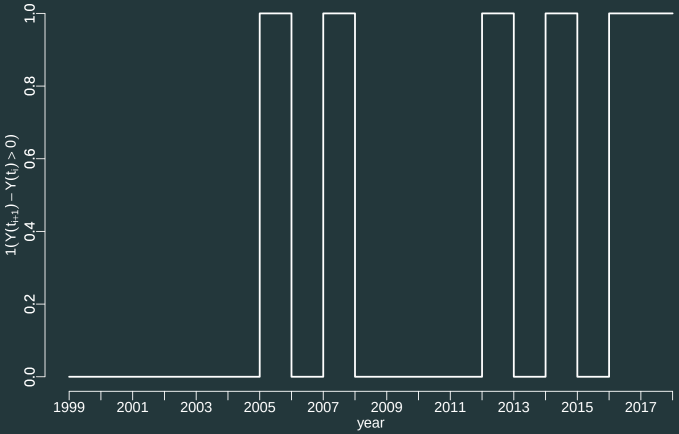
Comparison	p-value
2018 vs. 2017	0.074
2018 vs. 2016	**0.020**
2018 vs. 2015	0.495
2018 vs. 2014	**0.012**
2018 vs. 2013	0.576

Conclusion?

Simple analysis - It is the first time the trend has changed?

A simple approach is to look at how often $1(Y(t_{i+1}) - Y_{t_i} > 0)$ jumps.

9 jumps. Many of them are probably just noise.



Trendiness analysis - Model selection

Before fitting the model we need to select a prior mean and covariance function for the Gaussian Process.

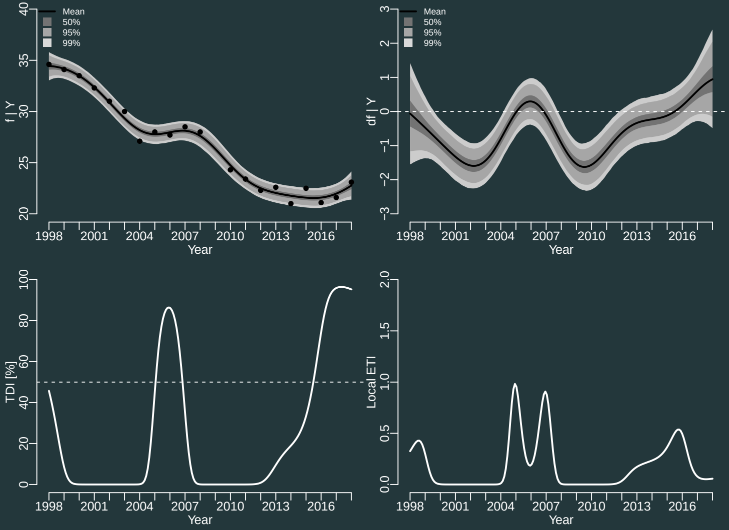
We consider 12 different models and compare them by Maximum Marginal Likelihood based LOO MSEP.

	SE	RQ	Matern 3/2	Matern 5/2
Constant	0.682	**0.651**	0.687	0.660
Linear	0.806	$\Leftarrow$	0.896	0.865
Quadratic	0.736	$\Leftarrow$	0.800	0.785

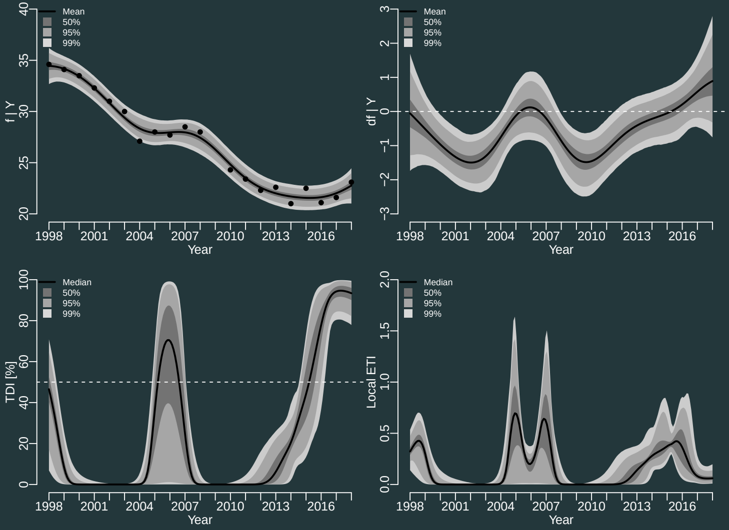
$$\Theta_{-i}^{\mathcal{M}} = \arg\sup_{\Theta} N\left(\mathbf{Y}_{-i}; \mu_{\beta}^{\mathcal{M}}(\mathbf{t}_{-i}), C_{\theta}^{\mathcal{M}}(\mathbf{t}_{-i}, \mathbf{t}_{-i}) + \sigma^2 I\right)$$

$$\mathrm{MSPE}_{\mathrm{LOO}}^{\mathcal{M}} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \mathbb{E}[f(t_i) \mid \mathbf{Y}_{-i}, \mathbf{t}_{-i}\Theta_{-i}^{\mathcal{M}}]\right)^2$$

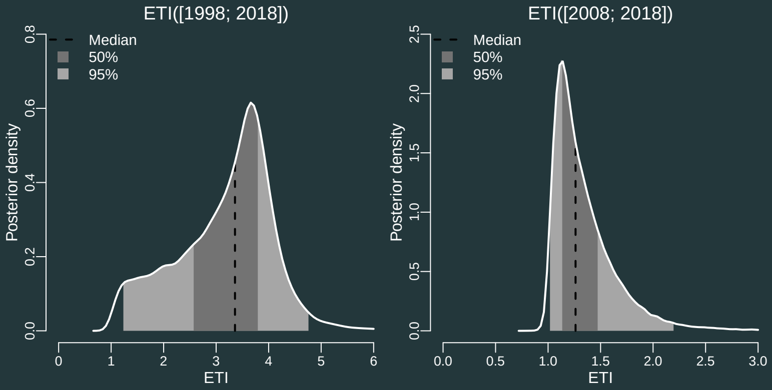
Trendiness analysis - Maximum Likelihood fit



Trendiness analysis - Bayesian fit



Trendiness analysis - Posterior ETI



Conclusions

Maximum Marginal Likelihood analysis:

- TDI(2018) = 95.24%
- ETI([1998; 2018]) = 3.68
- ETI([2008; 2018]) = 1.39

Bayesian analysis:

- Mean TDI in 2018 = 93.37% (95% CI = [82.22%; 98.87%])
- Median ETI([1998; 2018]) = 3.36 (95% CI = [1.22; 4.76])
- Median ETI([2008; 2018]) = 1.26 (95% CI = [1.02; 2.20])

Conclusions

Maximum Marginal Likelihood analysis:

- $\text{TDI}(2018) = 95.24\%$
- $\text{ETI}([1998; 2018]) = 3.68$
- $\text{ETI}([2008; 2018]) = 1.39$

Bayesian analysis:

- Mean TDI in 2018 $= 93.37\%$ (95% CI $= [82.22\%; 98.87\%]$)
- Median $\text{ETI}([1998; 2018]) = 3.36$ (95% CI $= [1.22; 4.76]$)
- Median $\text{ETI}([2008; 2018]) = 1.26$ (95% CI $= [1.02; 2.20]$)

Conclusions:

- We are currently (positively) trending with a very high probability.
- We have been trending with probability $> 50\%$ since sometime between 2015 and 2016.
- The expected number of changes in trend during the last 20 years is higher than stipulated.

Summary

We have tried to answer the questions:

1. What is the probability of having positive trend at a given time?
2. What is the expected number of times that a trend has changed on an interval?

Functional Data Analysis using latent Gaussian Processes is a flexible way to obtain such answers.

Thank you!