

Quantifying the Trendiness of Trends (or how I learned to stop worrying and love the uncertainty of changes)

*Andreas Kryger Jensen and Claus Ekstrøm
Biostatistics, Institute of Public Health, University of Copenhagen*

14 April, 2019

Abstract

A statement often seen in the news concerning some public health outcome is that some trend has changed or been broken. Such statements are often based on longitudinal data from e.g., surveys, and the change in the trend is claimed to have occurred at the time of the latest data collection. These types of statistical assessments are very important as they may potentially influence public health decisions on a national level.

Instead we propose a statistical model using methods from Functional Data Analysis under the assumption that reality evolves in continuous time. Under this assumption we define what constitutes a trend and a change in a trend, and we introduce a probabilistic Trend Direction Index. This index has the intuitive interpretation of the probability that an underlying trend has changed direction at any given time conditional on observed data. We also define a global index of Expected Trend Instability that quantifies the expected number of times a trend changes on an interval.

We show how the Trend Direction Index and the Expected Trend Instability can be estimated from data in a Bayesian framework and give an application to development of the proportion of smokers in Denmark during the last 20 years.

Keywords: Functional Data Analysis, Gaussian Processes, Trends, Bayesian Statistics

1 Introduction

This manuscript is concerned with the fundamental problem of estimating an underlying trend based on random variables observed repeatedly over time. In addition to this problem we also wish to assess points in time where it is possible that such a trend is changing. Our motivation is a recent example from the news in Denmark where it was stated that the trend in the proportion of smokers in Denmark has changed at the end of the year 2018. This statement was based on survey data collected yearly since 1998 and reported by the Danish Health Authority (The Danish Health Authority 2019). It is not immediately obvious what constitutes a trend and more so what a change in a trend is. To elucidate these questions and provide a clear specification we start from the ground by defining the problem.

[**TODO:** Claus tilbyder at b tte sig igennem resten af introduktionen].

[**TODO:** What do we propose? We work in continuous time and introduce TDI and ETI informally at this point. TDI gives the local probability of the direction of f and can answer the question „What is the probability that f is increasing?“. ETI gives the expected number of changes in the direction of f on an interval and can answer the question „Is it the first time in 20 years that f has changed trend?].

[**TODO:** We are not doing a hypothesis test and compute a p-value. I.e., we do not calculate the probability that an observed change in trend is different than would have been expected if there was

no change in trend. We estimate the actual probability that the trend is changing.]

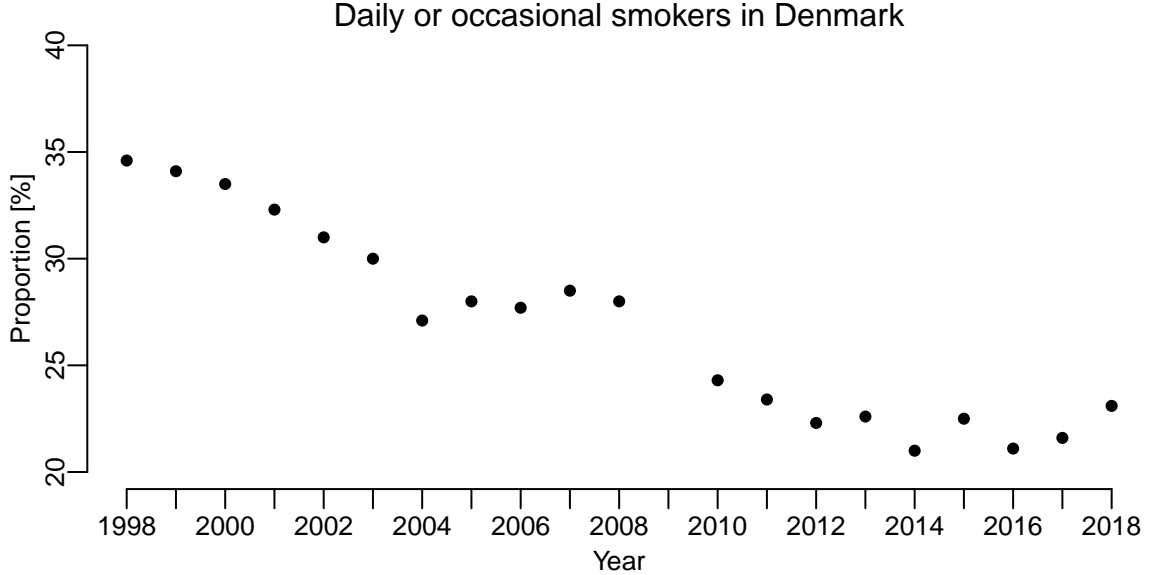


Figure 1: The proportion of daily or occasional smokers in Denmark during the last 20 years estimated from survey data and reported by the Danish Health Authority. [TODO: Mention in text why 2009 is missing]

[TODO: Assert that both TDI and EDI are scale free. They don't tell anything about the magnitude of a trend. They should therefore always be accompanied by plots of the posterior of df].

[TODO: Reference some previous work]. Gottlieb and Müller (2012) define a stickiness coefficient for longitudinal data. Our ETI has some similarities to the tracking index of Foulkes and Davis (1981).

The manuscript is structured as follows: In section 2 we present our statistical model based on a latent Gaussian Process formulation giving rise to analytic expressions for the Trend Direction Index and the Expected Trend Instability index conditional on observed data. We then extend this basic model to enable a data-driven adaptation to non-stationarity in the underlying dynamics. Section 2.4 is concerned with estimating the models parameters, and we give an application to the proportion of smokers in Denmark during the last 20 years in section 3, and we conclude with a discussion.

2 Methods

We assume that reality evolves in continuous time $t \in \mathcal{T} \subset \mathbb{R}$ and that there exists a latent, random function $f = \{f(t) : t \in \mathcal{T}\}$ that governs the underlying evolution of some observable characteristic in a population. We are able to observe a noisy version of this latent characteristic by sampling f at discrete time points according to the additive model $Y_i = f(t_i) + \varepsilon_i$ where Y_i is the i 'th observation at time $t_i \in \mathcal{T}$ and ε_i is a random variable with $E[\varepsilon_i | t_i] = 0$ [TODO: Explain why?]. Given observations of the form $(Y_i, t_i)_{i=1}^n$ we are interested in estimating the dynamical properties of the latent f .

The trend of f is defined as its instantaneous slope given by the function $df(t) = \left(\frac{df(s)}{ds}\right)(t)$, and we say that f exhibits a positive trend at t if $df(t) > 0$ and a negative trend at t if $df(t) < 0$. A change in trend is defined as a point in time where df changes sign, i.e., a positive trend becomes a negative trend meaning that f goes from being increasing to becoming decreasing or vice versa. As f is a random function there are no exact points in time where df changes sign. This is instead a gradual and continuous concept and the assessment must be defined in terms of the probability that a change in trend is occurring. This stands in contrast to traditional change-point models which assume that there are one or more exact time points where sudden a change in a regression parameter occurs (Picard 1985).

To quantify the probability that a trend is changing we introduce the following local probabilistic Trend Direction Index

$$\text{TDI}(t, \delta) = P(df(t + \delta) > 0 \mid \mathcal{F}_t) \quad (1)$$

where \mathcal{F}_t is the σ -algebra of all available information observed up until time t . The value of $\text{TDI}(t, \delta)$ is equal to the probability that f is increasing at time $t + \delta$ given all we know about the data generating process up until time t . A similar definition could be given for a negative trend but that is equal to $1 - \text{TDI}(t, \delta)$ and therefore redundant. The sign of δ determines whether we wish to estimate the past ($\delta \leq 0$) or forecast into the future ($\delta > 0$). Most of the examples seen in the news concerning public health outcomes is concerned with t being equal to the current calendar time and $\delta = 0$. This excludes using both change-point or segmented regression models (Quandt 1958) as there are no observations beyond the stipulated change-point. A useful parametrization of the Trend Direction Index is $\text{TDI}(\max \mathcal{T}, t - \max \mathcal{T})$ with $t \leq \max \mathcal{T}$ which conditions on the full observation period and looks back in time. In this parametrization $t = \max \mathcal{T}$ corresponds to the Trend Direction Index at the end of the observation period. A natural threshold for trendiness is $\text{TDI} = 0.5$ corresponding to a situation where the trend is symmetrically distributed around zero. Depending on the application, different thresholds can be used based on external loss or utility functions.

In addition to the Trend Direction Index we also define a global index of trend instability. We say that a function f is *trend stable* on an interval $\mathcal{I} \subseteq \mathcal{T}$ if it remains either increasing or decreasing so that its trend maintains its sign. As a quantification of the trend instability of a function f we propose to use the expected number of times that the function changes direction on the interval. This is equivalent to counting the number of zero crossings by df on \mathcal{I} . Informally at this point, we define the Expected Trend Instability as

$$\text{ETI}(\mathcal{I}) = E[\#\{t \in \mathcal{I} : df(t) = 0\} \mid \mathcal{F}] \quad (2)$$

which is the expected value of the size of the random point set of zero crossings by df on \mathcal{I} conditional on a suitable σ -algebra \mathcal{F} . An Expected Trend Instability of zero implies that $f \dots$ The lower $\text{ETI}(\mathcal{I})$ is, the more trend stable f is on \mathcal{I} and vice versa.

2.1 Latent Gaussian Process Model

The definitions in the previous section impose the existence of an infinite dimensional latent variable f and its derivative df . In the following we assume that f is a latent Gaussian Process which is equivalent to imposing a functional prior distribution on the latent characteristic [**TODO: something about flexible**]. Our knowledge about f and df will then be updated based on the observed data giving rise to a joint posterior distribution over f and df .

A random function f is a Gaussian Process if and only if the vector $(f(t_1), \dots, f(t_n))$ has a multivariate normal distribution for every set of evaluation points $(t_1, \dots, t_n) \subset \mathcal{T}$ with $n < \infty$, and we write $f \sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))$ where $\mu: \mathcal{T} \mapsto \mathbb{R}$ is the mean function and $C: \mathcal{T} \times \mathcal{T} \mapsto \mathbb{R}$ is a symmetric, positive definite covariance function (Cramer and Leadbetter 1967). Given observed data of the form $(Y_i, t_i)_{i=1}^n$ we lay out the data generating process by the following hierarchical model

$$\begin{aligned} f \mid \beta, \theta &\sim \mathcal{GP}(\mu_\beta(\cdot), C_\theta(\cdot, \cdot)) \\ Y_i \mid t_i, f(t_i), \Theta &\stackrel{iid}{\sim} N(f(t_i), \sigma^2), \quad \Theta = (\beta, \theta, \sigma) \end{aligned} \quad (3)$$

where β is a vector of parameters for the mean function of f , θ is a vector of parameters governing the covariance of f , and σ^2 is the observation variance. We assume the following regularity conditions.

Assumption 1. *We assume that f is a separable Gaussian process and that*

A1: ?

A2: ?

This ensures that the sample paths of f are in $C^r(\mathcal{T})$. We require that $r \geq 1$ for the Trend Direction Index and $r \geq 2$ for the index of Expected Trend Instability.

[@scheuerer2010regularity]

Under the above assumptions, an important property of a Gaussian process is that it together with its derivative functions are multivariate Gaussian Process (Cramer and Leadbetter 1967). This implies that the joint distribution of f and df is given by the bivariate Gaussian Process

$$\begin{bmatrix} f(s) \\ df(t) \end{bmatrix} \mid \beta, \theta \sim \mathcal{GP} \left(\begin{bmatrix} \mu_\beta(s) \\ d\mu_\beta(t) \end{bmatrix}, \begin{bmatrix} C_\theta(s, s') & \partial_2 C_\theta(s, t) \\ \partial_1 C_\theta(t, s) & \partial_1 \partial_2 C_\theta(t, t') \end{bmatrix} \right) \quad (4)$$

where $d\mu_\beta$ is the derivative function of μ_β and ∂_j denotes the partial derivatives with respect to the j 'th variable. This generalizes directly to higher order derivatives assuming that necessary regularity conditions are satisfied. Let $\mathbf{Y} = (Y_1, \dots, Y_n)$ be the vector of outcomes observed at times $\mathbf{t} = (t_1, \dots, t_n) \subset \mathbf{T}$ and \mathbf{t}^* any finite dimensional vector of time points not necessarily equal to \mathbf{t} . Formulas (3) and (4) imply that joint distribution of $(f(\mathbf{t}^*), df(\mathbf{t}^*), \mathbf{Y}) \mid \mathbf{t}, \Theta$ is multivariate normal, and that the marginal distribution of the observed data is $\mathbf{Y} \mid \mathbf{t}, \Theta = N(\mu_\beta(\mathbf{t}), C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)$ where $C_\theta(\mathbf{t}, \mathbf{t})$ denotes a matrix consisting of the values of the covariance function at $\{(s, t) \in \mathbf{t} \times \mathbf{t}\}$, and I is an identity matrix of appropriate dimensions. Predictions of the latent function and its derivative at \mathbf{t}^* are given by the posterior distribution $(f(\mathbf{t}^*), df(\mathbf{t}^*)) \mid \mathbf{Y}, \mathbf{t}, \Theta$. This distribution is also multivariate normal and its moments are given in Proposition 1.

Proposition 1. Let the data generating model be defined as in Equation (3) and \mathbf{t}^* any finite vector of time points. The posterior distribution of (f, df) evaluated at \mathbf{t}^* is then given by

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

where

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{t}^*) &= \begin{bmatrix} \mu_\beta(\mathbf{t}^*) \\ d\mu_\beta(\mathbf{t}^*) \end{bmatrix} + \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \end{bmatrix} (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\ \boldsymbol{\Sigma}(\mathbf{t}^*, \mathbf{t}^*) &= \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix} - \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \end{bmatrix} (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} \begin{bmatrix} C_\theta(\mathbf{t}, \mathbf{t}^*) \\ \partial_2 C_\theta(\mathbf{t}, \mathbf{t}^*) \end{bmatrix} \end{aligned}$$

The results from Proposition 1 enables us to simulate random realizations from the posterior distribution of (f, df) at any finite number of time points given values of Θ . Further, (μ_f, μ_{df}) is the Best Linear Unbiased Predictor of (f, df) .

2.2 Trend Direction Index

The Trend Direction Index was defined generally in Equation (1). Letting the filtration be the sigma algebra of the observed data and their sampling times, we may express the Trend Direction Index through the posterior distribution of df . The following proposition sates this result.

Proposition 2. Let \mathcal{F}_t be the sigma algebra generated by (\mathbf{Y}, \mathbf{t}) . The Trend Direction Index in Equation (1) can then be written in terms of the posterior distribution of df as

$$\begin{aligned} \text{TDI}(t, \delta | \Theta) &= P(df(t + \delta) > 0 | \mathbf{Y}, \mathbf{t}, \Theta) \\ &= \int_0^\infty N(u; \mu_{df}(t + \delta), \Sigma_{df}(t + \delta, t + \delta)^{1/2}) du \end{aligned}$$

where μ_{df} and Σ_{df} are given in Proposition 1 and we condition on the model parameters Θ .

We note that the Trend Direction Index depends on the parameters of the model, $\Theta = (\beta, \theta, \sigma^2)$, which in practice must be estimated from data. When taking their uncertainties into account TDI becomes a functional distribution over probabilities.

To illustrate the Trend Direction Index we show an example in the noise free case with known model parameters in Figure 2. The first and second rows of the plot show 150 random realizations from the posterior distribution of f and df with the expected functions in bold lines, and the third row shows the Trend Direction Index. In this example we simulate using known model parameters and therefore the Trend Direction Index in Proposition 2 is a deterministic function. The three columns of the plot show how the posterior of f and df and the Trend Direction Index are updated after one, two or four observed data points. It is seen how updating the Gaussian Process prior with observed data results in a continuous change in the posterior distribution of f and its derivative df . The posterior knowledge about the sign of the derivative is correspondingly reflected in the Trend Direction Index. Going further and further away from the last observed data point it is seen that the posterior of f becomes dominated by its prior distribution. In the same sense df converges to a point-wise symmetric distribution around zero and therefore the Trend Direction Index stabilizes around 50%. This is clearly seen in the second column of the figure. [TODO: More explaining]

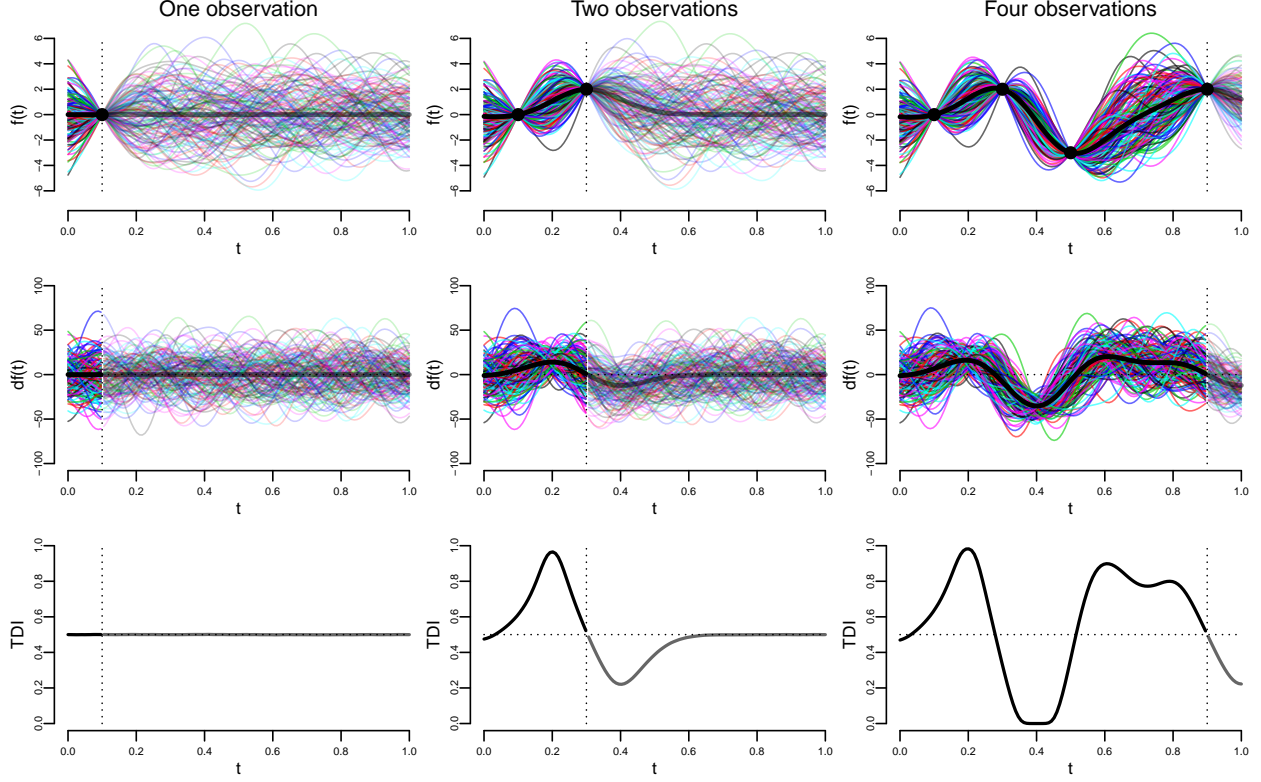


Figure 2: 150 realizations from the posterior distribution of f (top row), df (middle row) with expected values in bold and the Trend Direction Index (bottom row) conditional on one, two and four noise free observations. Dotted vertical lines show the points in time after which forecasting occurs.

2.3 Expected Trend Instability

The general form of the Expected Trend Instability index was defined in Equation (2).

Proposition 3. *The local Expected Trend Instability is*

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

where $\phi: x \mapsto 2^{-1/2}\pi^{-1/2}\exp(-\frac{1}{2}x^2)$ is the standard normal density function, $\text{Erf}: x \mapsto 2\pi^{-1/2} \int_0^x \exp(-u^2)du$ is the error function, and ω and ζ are given by

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

where μ_{df} , μ_{d^2f} , Σ_{df} , Σ_{d^2f} and Σ_{df, d^2f} are the posterior moments given in ???. The Expected Trend Instability is then given by integrating the local Expected Trend Instability over the interval

$$\text{ETI}(\mathcal{I} \mid \Theta) = \int_{\mathcal{I}} \text{dETI}(s, \mathcal{I} \mid \Theta) ds$$

The value of ETI is a finite [TODO: finite because of assumptions] real positive value.

Figure 3 shows that Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam vitae vulputate tellus. Cras vitae tempor nibh. Donec dignissim ex ex, nec dignissim magna hendrerit in. Aenean orci nunc, rhoncus quis dictum vitae, tristique non lorem. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nulla facilisi. Vivamus eget ante consectetur, bibendum nulla non, ultrices risus. Nulla neque mauris, tempus quis dui eu, eleifend gravida sem. Nulla blandit non tortor pellentesque varius.

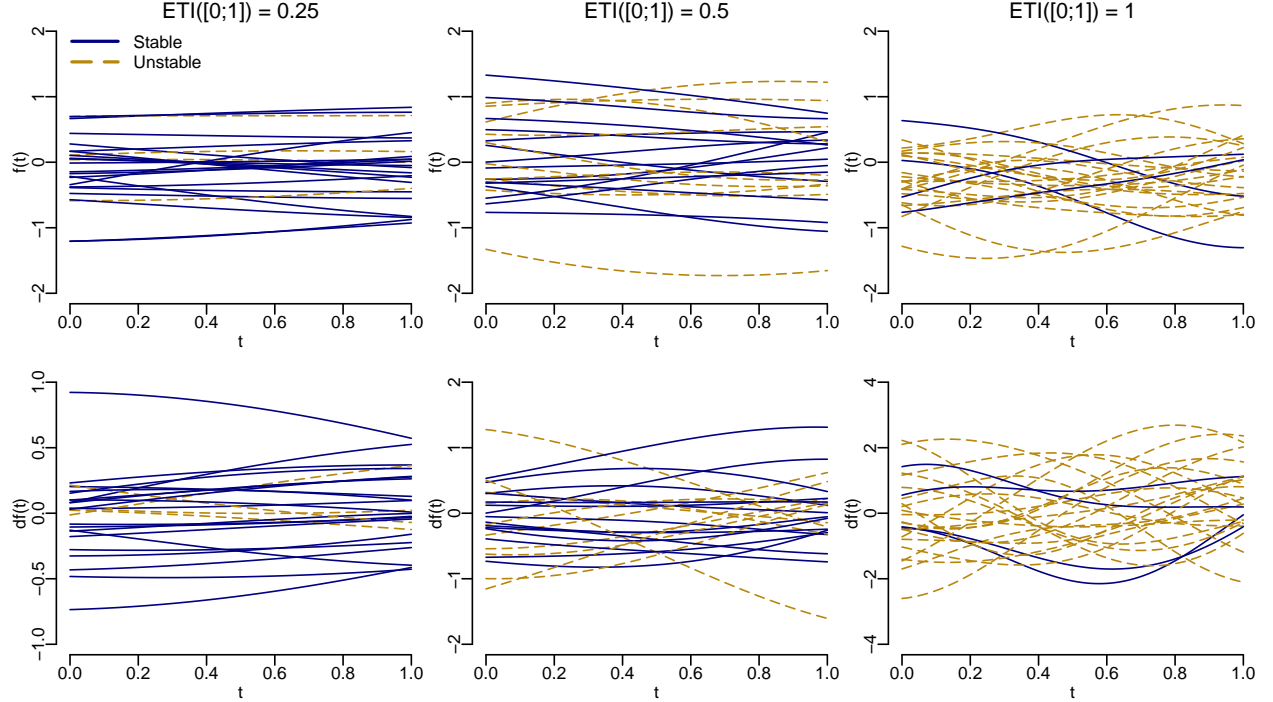


Figure 3: 25 random pairs sampled from the joint distribution of a Gaussian Process (f) and its derivative (df) with different values of Expected Trend Instability (ETI). The first row shows samples from f , and the second row shows samples from df . The columns define different values of ETI. Sample paths that are trend stable are shown by solid blue lines, and unstable sample paths are shown by dashed gold colored lines.

2.4 Estimation

The latent Gaussian Process model defined in Equation (3) requires estimates of the parameter vector Θ in order to calculate the joint posterior distribution required for the trend indices. Different methods for fitting models with latent Gaussian Processes such as maximum likelihood (also called Empirical Bayes), fully Bayesian estimation and variational Bayes have been considered in the literature.

Maximum marginal likelihood:

$$L(\Theta \mid \mathbf{Y}, \mathbf{t}) = \frac{1}{2} \log |C_{\theta}(\mathbf{t}, \mathbf{t})| - \frac{1}{2} (\mathbf{Y} - \mu_{\beta}(\mathbf{t})) C_{\theta}(\mathbf{t}, \mathbf{t})^{-1} (\mathbf{Y} - \mu_{\beta}(\mathbf{t}))^T \quad (5)$$

Bayesian general prior:

$$\Theta_k \sim \text{Half-T} \left(\widehat{\Theta}_k^{\text{ML}}, \tau_k^2, \kappa_k \right)$$

We have implemented both our model in Stan (Carpenter et al. 2017). Stan is a probabilistic programming language that enables full Bayesian statistical inference using Markov chain Monte Carlo sampling and can be used together with R through the `rstan` package (Stan Development Team 2018). The Stan programs are available at the first author’s Github repository (Jensen 2019).

Given observed data $(Y_i, t_i)_{i=1}^n$ there are different approach for estimating models with latent Gaussian Processes. We believe that when doing a probabilistic assessment of a latent variable, such as the Trend Direction Index, it is important to account for all the uncertainty in the model. This is best achieved by using a fully Bayesian estimation. This requires placing prior distributions on the model parameters Θ . These will then be updated based on the observed data and the Trend Direction Index in Proposition 2 is then conditional on these posterior distributions.

3 Application

A report published by The Danish Health Authority in January 2019 updated the estimated proportion of smokers in Denmark with new data from 2018 (The Danish Health Authority 2019). The data was based on an internet survey including 5017 participants. The report also included data on the proportion of smokers in Denmark during the last 20 years. These data were shown in Figure 1. The report was picked up by several news papers under headlines stating that the proportion of smokers in Denmark had significantly increased for the first time in two decades (Navne, Schmidt, and Rasmussen 2019). The report published no statistical analyses for this statement but stated in a note that because the study population is so large, then more or less all differences become statistically significant at the 5% level (this was erroneously written as a 95% significance level in the report).

This data set provides an instrumental way of exemplifying our two proposed trend quantification measures. In this application we with to assess the statistical properties of following questions:

- Q1: Is the proportion of smokers increasing in the year 2018 conditional on data from the last 20 years?
- Q2: If the proportion of smokers is currently increasing, when did this increase probably start?
- Q3: Is it the first time during the last 20 years that the trend in the proportion of smokers has changed?

A simple approach for trying to answer questions Q1 and Q2 would be to apply a χ^2 -test in a 2×2 table. Table 1 shows the p-values for the χ^2 -test of independence between the proportion of smokers in 2018 and each of the five previous years. Using a significance level of 5% the conclusion is ambiguous. Compared to the previous year, there was no significant change in the proportion in 2018. Three out of these five comparisons fail to show a significant change in proportions. It is therefore clear that such point-wise testing is not sufficiently perceptive to catch the underlying continuous development.

2018	2017	2016	2015	2014	2013
p-value	0.074	0.020	0.495	0.012	0.576

Table 1: p-values obtained from χ^2 -tests of independence between the proportion of smokers in 2018 and the five previous years. Numbers in bold are significant differences at the 5% level.

Similarly, a simple approach for trying to answer question Q3 would be to look at the cumulative number of times that the difference in proportion between consecutive years changes sign. In the data set there were nine changes in the sign of the difference between the proportion in each year and the proportion in the previous year giving this very crude estimate of the number of times, that the trend has changed. This approach suffers from the facts that it is based on a finite difference approximation at the observed time points to the continuous derivative, and that it uses the noisy measurements instead of the values of the latent function.

We now present an analysis of the data set using our proposed model. To complete the model specification in Equation (3) we must decide on the functional forms of the mean and covariance functions of the latent Gaussian Process. We considered three different mean functions: a constant mean, $\mu_\beta(t) = \beta_0$, a linear mean, $\mu_\beta(t) = \beta_0 + \beta_1 t$, and a quadratic mean function, $\mu_\beta(t) = \beta_0 + \beta_1 t + \beta_2 t^2$. For the covariance function we consider the Squared Exponential (SE), the Rational Quadratic (RQ), the Matern 3/2 (M3/2), and the Matern 5/2 (M5/2) covariance functions. The expressions for these covariance functions are given by

$$\begin{aligned}
C_\theta^{\text{SE}}(s, t) &= \alpha^2 \exp\left(-\frac{(s-t)^2}{2\rho^2}\right), \quad \theta = (\alpha, \rho) \\
C_\theta^{\text{RQ}}(s, t) &= \alpha^2 \left(1 + \frac{(s-t)^2}{2\rho^2\nu}\right)^{-\nu}, \quad \theta = (\alpha, \rho, \nu) \\
C_\theta^{\text{M3/2}}(s, t) &= \alpha^2 \left(1 + \frac{\sqrt{3}\sqrt{(s-t)^2}}{\rho}\right) \exp\left(-\frac{\sqrt{3}\sqrt{(s-t)^2}}{\rho}\right), \quad \theta = (\alpha, \rho) \\
C_\theta^{\text{M5/2}}(s, t) &= \alpha^2 \left(1 + \frac{\sqrt{5}\sqrt{(s-t)^2}}{\rho} + \frac{5(s-t)^2}{3\rho^2}\right) \exp\left(-\frac{\sqrt{5}\sqrt{(s-t)^2}}{\rho}\right), \quad \theta = (\alpha, \rho)
\end{aligned} \tag{6}$$

and we note that the Rational Quadratic covariance function converges to the Squared Exponential covariance function for $\nu \rightarrow \infty$, $\lim_{\nu \rightarrow \infty} C_\theta^{\text{RQ}}(s, t) = C_\theta^{\text{SE}}(s, t)$. [**TODO:** Something about the sample path regularity for these covariance functions. M3/2 cannot be used for ETI but okay for TDI].

The total combination of mean and covariance functions results in 12 candidate models, and to select the model providing the best fit to data we used leave-one-out cross-validation based on maximum likelihood estimation. For each model, \mathcal{M} , we turn by turn excluded a single observation (t_i, Y_i) from the vectors of observed outcomes and time points, and we let the leave-one-out data be denoted by $(\mathbf{Y}_{-i}, \mathbf{t}_{-i})$. Based on the leave-one-out data sets we estimated the parameters, $\Theta_{-i}^{\mathcal{M}} = (\beta, \theta, \sigma^2)$, of the latent Gaussian Process for each model by maximizing the marginal log-likelihood function of the model in Equation (3). These estimates were given by

$$\hat{\Theta}_{-i}^{\mathcal{M}} = \arg \sup_{\Theta} L(\Theta \mid \mathbf{Y}_{-i}, \mathbf{t}_{-i})$$

where the log-likelihood is given by Equation (5). These estimates are then plugged into the expression for the posterior expectation of f in Proposition 1 to obtain the leave-out-out predictions. The mean squared prediction error was then calculated from the squared errors between the leave-one-out predictions and the observed values averaged across all data points:

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

and the model providing the best fit to data was selected as $\mathcal{M}_{\text{opt}} = \arg \min_{\mathcal{M}} \text{MSPE}_{\text{LOO}}^{\mathcal{M}}$.

	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

Table 2: Leave-one-out cross-validated mean squared error of prediction for each of the 12 candidate models. \Leftarrow indicates numerical convergence to the SE covariance function.

Table 2 shows the mean squared prediction error for each of the models. For the models with a Rational Quadratic covariance function and a linear and a quadratic mean function the parameter ν diverged numerically implying convergence to the Squared Exponential covariance function. Comparing the leave-one-out mean squared prediction errors for the candidate models, the optimal model has a constant mean function and a Rational Quadratic covariance function. The marginal maximum likelihood estimates of the parameters of the optimal model was

$$\widehat{\beta}_0^{\text{ML}} = 28.001, \quad \widehat{\alpha}^{\text{ML}} = 4.543, \quad \widehat{\rho}^{\text{ML}} = 4.438, \quad \widehat{\nu}^{\text{ML}} = 1.020, \quad \widehat{\sigma}^{\text{ML}} = 0.622$$

Figure 4 shows the fit of the model by plugging in the values of the marginal maximum likelihood estimates into the expressions for the posterior distributions of f and df defined in Proposition 1, the Trend Direction Index in Proposition 2, and the Local Expected Trend Instability in Proposition 3.

$$\begin{aligned} \beta_0 &\sim T\left(\widehat{\beta}_0^{\text{ML}}, 3, 3\right) & \alpha &\sim \text{Half-}T\left(\widehat{\alpha}^{\text{ML}}, 3, 3\right) & \rho &\sim \text{Half-}N\left(\widehat{\rho}^{\text{ML}}, 1\right) \\ \nu &\sim \text{Half-}T\left(\widehat{\nu}^{\text{ML}}, 3, 3\right) & \sigma &\sim \text{Half-}T\left(\widehat{\sigma}^{\text{ML}}, 3, 3\right) \end{aligned}$$

We ran four chains for 25,000 iterations each with half of the iterations used for warm-up. Convergence was assessed by the potential scale reduction factor (\hat{R}) of Gelman and Rubin (1992).

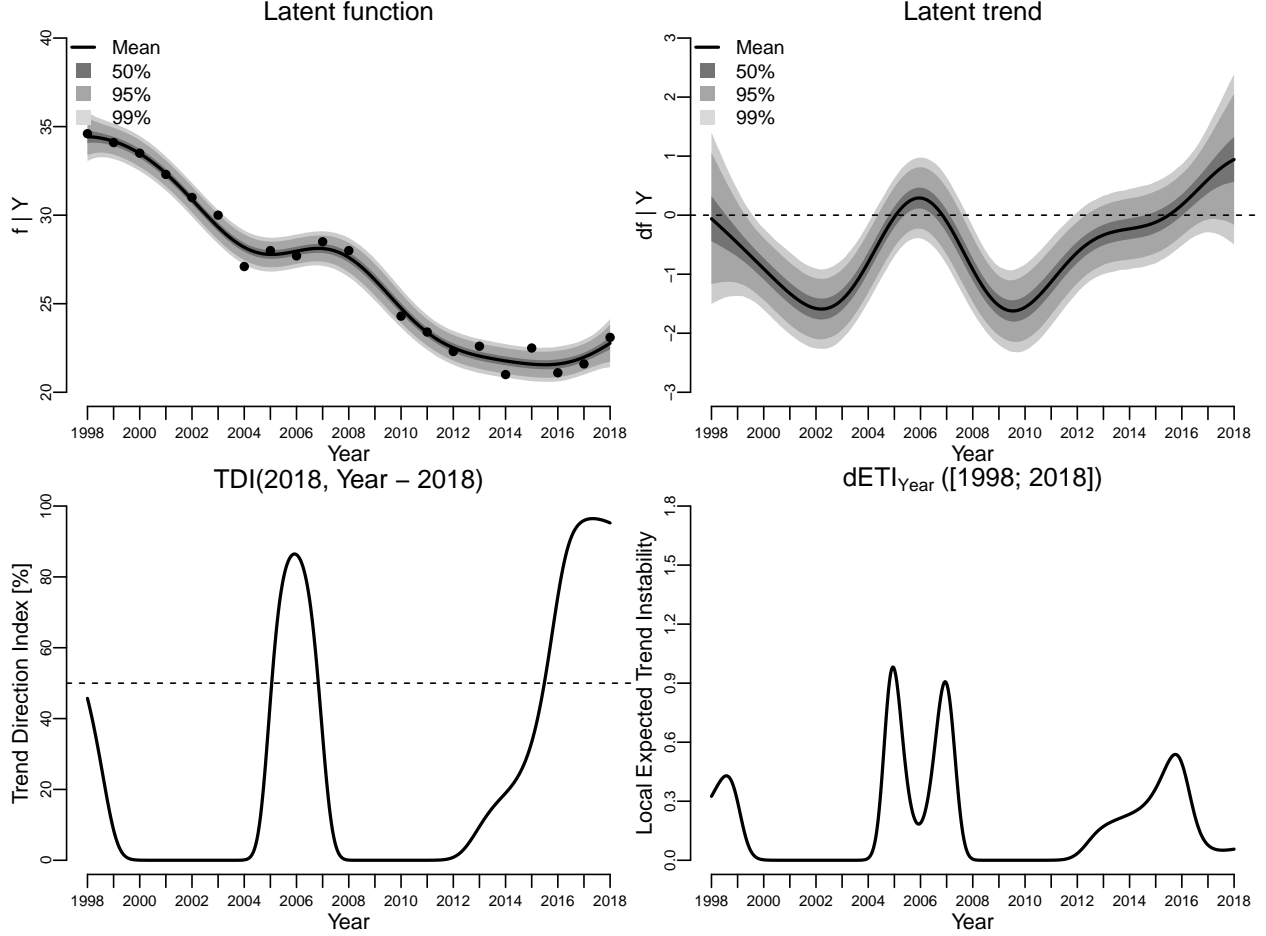


Figure 4: Results from fitting the latent Gaussian Process model by maximum likelihood. The first row shows the posterior distributions of f (left) and df (right) with the posterior means in bold and gray areas showing point-wise probability intervals for the posterior distribution. The second row shows the estimated Trend Direction Index (left) and the local Expected Trend Instability (right).

The top row of Figure 5 shows the posterior distributions of the Trend Direction Index and the Local Expected Trend Instability from the Bayesian model. The bottom row shows the posterior distributions of the Expected Trend Instability during the last twenty years and the last ten years.

Cross-point for 50% for MLE: year = 2015.48.

Cross-point for 50% for Bayes: [2014.62; 2015.96]

Figures 4 and 5 and the summary statistics in Table 3 indicate that the Bayesian estimates generally produce a high degree of smoothness in the model fit. This can be seen from the plot of the median local Expected Trend Instability in Figure 5 which is generally lower than its corresponding maximum likelihood point estimates in Figure 4. This is similarly reflected in the median ETI estimates which are lower than their values under maximum likelihood in Table 3. Looking at the posterior distributions of the covariance parameters θ (not shown), we see that this is mainly a result of not restricting the parameter ν to its maximum likelihood value. The 95% probability interval of the posterior distribution of ν was [0.328; 10.743] which is highly right-skewed compared to the maximum likelihood estimate of $\nu^{\text{ML}} = 1.020$.

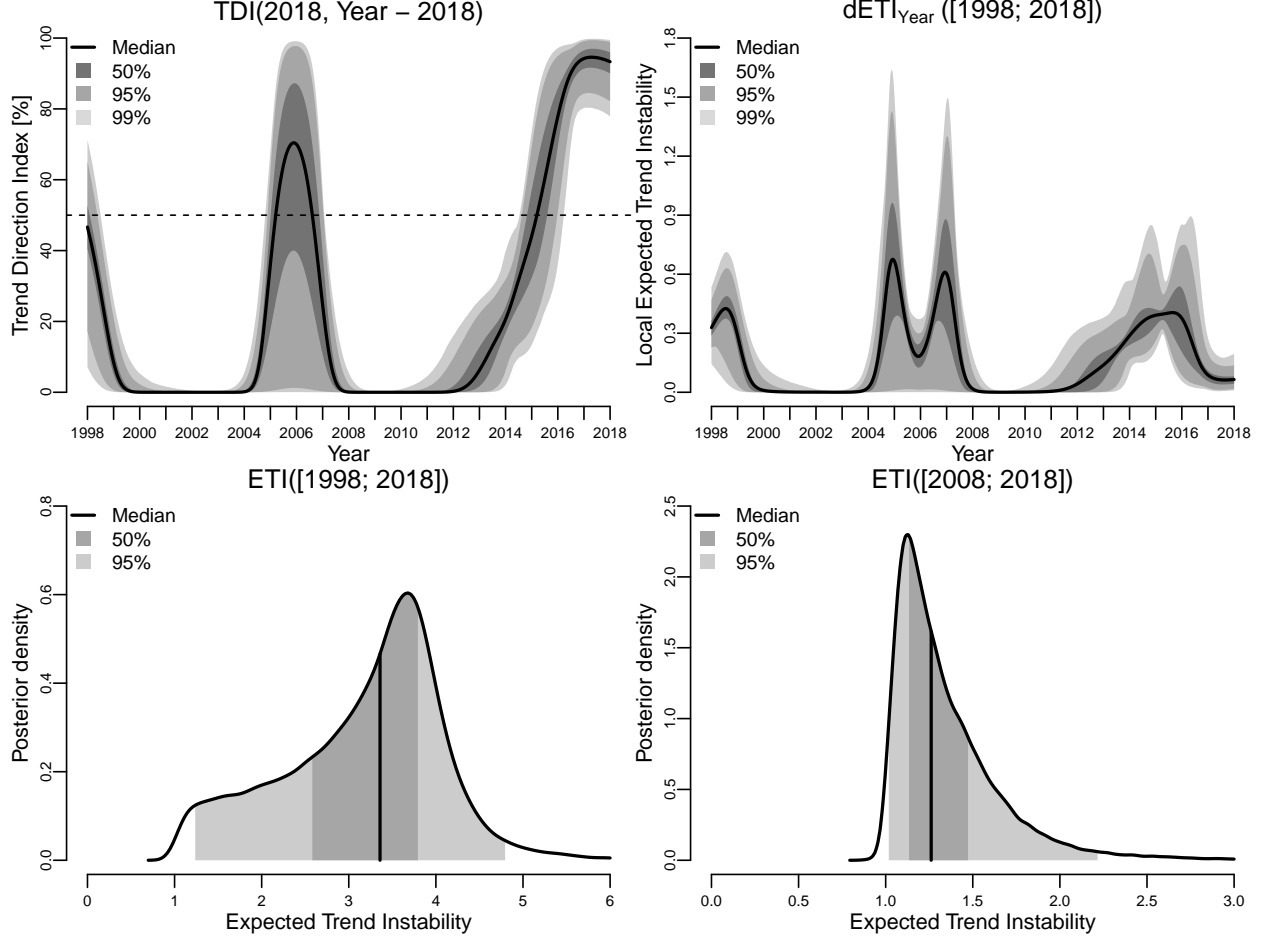


Figure 5: hejhej

To understand the effect of ν on the smoothness of the fitted models can compare the local expected number of crossings by a Gaussian Process and its derivative at their mean values in the simple case with a zero mean and either the Rational Quadratic or the Squared Exponential covariance function. In this case the formula in Proposition 3 simplifies immensely so that the local expected number of mean-crossings by f is equal to $\pi^{-1}(\frac{\partial_1 \partial_2 C_\theta(t,t)}{C_\theta(t,t)})^{1/2}$ and for df it is equal to $\pi^{-1}(\frac{\partial_1^2 \partial_2^2 C_\theta(t,t)}{\partial_1 \partial_2 C_\theta(t,t)})^{1/2}$. From the expressions in Equation (6) we see that the local expected number of mean-crossings by f is the same for both the Squared Exponential and the Rational Quadratic covariance functions and is equal to $\pi^{-1}\rho^{-1}$. However, for df we see that its local expected number of mean-crossings is equal to $3^{1/2}\pi^{-1}\rho^{-1}$ for the Squared Exponential covariance function and $3^{1/2}\pi^{-1}\rho^{-1}(1 + \nu^{-1})^{1/2}$ for the Rational Quadratic covariance function. As $1 < (1 + \nu^{-1})^{1/2} < \infty$ for $0 < \nu < \infty$ and monotonically decreasing for $\nu \rightarrow \infty$ with a limit value of one, we see that the local expected number of mean-crossings by df is always larger for the Rational Quadratic covariance functions compared to the Squared Exponential covariance function with equality in the limit. A right-skewed posterior distribution of ν therefore generally favors fewer crossings of the trend leading to a more stable trend and a smaller value of the Expected Trend Instability.

	Maximum Likelihood	Bayesian Posterior
TDI(2018, 0)	95.24%	93.32% [82.15%; 98.86%]
TDI(2018, -1)	95.92%	94.21% [84.28%; 99.11%]
TDI(2018, -2)	74.41%	77.87% [51.02%; 94.94%]
TDI(2018, -3)	33.36%	44.11% [18.23%; 69.19%]
TDI(2018, -4)	18.96%	20.60% [6.05%; 31.82%]
TDI(2018, -5)	9.50%	6.21% [0.03%; 22.21%]
ETI([2008, 2018])	1.39	1.25 [1.02; 2.22]
ETI([1998, 2018])	3.68	3.36 [1.24; 4.79]

Table 3: Summary measures from the Maximum Likelihood and Bayesian analyses. The rows show the estimated Trend Direction Index for 2013 to 2018 and the Expected Trend Instability for the last 10 and 20 years all conditional on data from 1998 to 2018. For the Bayesian analysis posterior medians and 95% posterior probability intervals are given.

4 Discussion

[**TODO:** Her må der bættess løs]

Something about that the indices are scale-free hence TDI should be interpreted along with $df \mid Y$ at the same time point. We could also define

$$\text{TDI}_u(t, \delta) = P(df(t + \delta) > u \mid \mathcal{F}_t)$$

where u is some trend value.

We could also have defined

$$\text{ETI}^\uparrow(\mathcal{I}) = \mathbb{E} \left[\# \left\{ t \in \mathcal{I} : df(t) = 0, d^2f(t) > 0 \right\} \mid \mathcal{F} \right]$$

counting the expected number of up-crossings at zero by df (the number of times f has gone from being decreasing to increasing) on \mathcal{T} .

Also something about stationarity of C_θ .

Different distributions of $Y \mid f$ could be used (e.g. binary outcomes) but then maximum likelihood estimation is not analytically tractable.

5 Bibliography

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. “Stan: A Probabilistic

- Programming Language.” *Journal of Statistical Software* 76 (1).
- Cramer, Harald, and M. R. Leadbetter. 1967. *Stationary and Related Stochastic Processes – Sample Function Properties and Their Applications*. John Wiley & Sons, Inc.
- Foulkes, Mary Ann, and CE Davis. 1981. “An Index of Tracking for Longitudinal Data.” *Biometrics*. JSTOR, 439–46.
- Gelman, Andrew, and Donald B. Rubin. 1992. “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science* 7 (4): 457–72.
- Gottlieb, Andrea, and Hans-Georg Müller. 2012. “A Stickiness Coefficient for Longitudinal Data.” *Computational Statistics & Data Analysis* 56 (12): 4000–4010.
- Jensen, Andreas Kryger. 2019. “Stan Implementations of the Gaussian Process Trend Direction Index.” <https://github.com/aejensen/TrendinessOfTrends>.
- Navne, Helene, Anders Legarth Schmidt, and Lars Igum Rasmussen. 2019. “Første Gang I 20 år: Flere Danskere Ryger.” <https://politiken.dk/forbrugogliv/sundhedogmotion/art6938627/Flere-danskere-ryger>.
- Picard, Dominique. 1985. “Testing and Estimating Change-Points in Time Series.” *Advances in Applied Probability* 17 (4): 841–67.
- Quandt, Richard E. 1958. “The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes.” *Journal of the American Statistical Association* 53 (284): 873–80.
- Rice, Stephen O. 1944. “Mathematical Analysis of Random Noise.” *Bell System Technical Journal* 23 (3): 282–332.
- Stan Development Team. 2018. “RStan: The R Interface to Stan.” <http://mc-stan.org/>.
- The Danish Health Authority. 2019. “Danskernes Rygevaner 2018.” www.sst.dk/da/udgivelser/2019/danskernes-rygevaner-2018.

6 Appendix

6.1 Proof of Proposition 1

We derive the joint posterior distribution of the latent process (f, df, d^2f) under the model in Equation (3). We start by noting that $\mathbf{Y} \mid \mathbf{t}, \Theta$ is jointly normal distributed with the latent processes with marginal mean $\mu_\beta(\mathbf{t})$ and covariance matrix

$$K(\mathbf{t}, \mathbf{t}) = C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I$$

where $C_\theta(\mathbf{t}, \mathbf{t})$ denotes a matrix consisting of the values of the covariance function at $\{(s, t) \in \mathbf{t} \times \mathbf{t}\}$, and I is an identity matrix of appropriate dimensions. Combining this with the result from Equation (4) generalized to also include the second-order derivative of f we obtain the following joint distribution for the latent processes and the observed data

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

where ∂_j^k denotes the k 'th derivative of the j 'th variable. By the joint normality, the joint posterior distribution of (f, df, d^2f) conditional on the parameters and the sampling times is equal to the following multivariate normal distribution

$$\begin{bmatrix} f(\mathbf{t}^*) \\ df(\mathbf{t}^*) \\ d^2f(\mathbf{t}^*) \end{bmatrix} | \mathbf{Y}, \mathbf{t}, \Theta \sim N(\boldsymbol{\mu}(\mathbf{t}^*), \boldsymbol{\Sigma}(\mathbf{t}^*, \mathbf{t}^*))$$

where the moments are partitioned as

$$\boldsymbol{\mu}(\mathbf{t}^*) = \begin{bmatrix} \mu_f(\mathbf{t}^*) \\ \mu_{df}(\mathbf{t}^*) \\ \mu_{d^2f}(\mathbf{t}^*) \end{bmatrix}, \quad \boldsymbol{\Sigma}(\mathbf{t}^*, \mathbf{t}^*) = \begin{bmatrix} \Sigma_f(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{f,d^2f}(\mathbf{t}^*, \mathbf{t}^*) \\ \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{df}(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{df,d^2f}(\mathbf{t}^*, \mathbf{t}^*) \\ \Sigma_{d^2f,f}(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{d^2f,df}(\mathbf{t}^*, \mathbf{t}^*) & \Sigma_{d^2f}(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix}$$

Using the standard formulas for conditional distributions in a multivariate normal model we obtain the following expressions for the joint posterior moments

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{t}^*) &= \begin{bmatrix} \mu_\beta(\mathbf{t}^*) \\ d\mu_\beta(\mathbf{t}^*) \\ d^2\mu_\beta(\mathbf{t}^*) \end{bmatrix} + \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \end{bmatrix} K(\mathbf{t}, \mathbf{t})^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\ \boldsymbol{\Sigma}(\mathbf{t}^*, \mathbf{t}^*) &= \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \\ \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1^2 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1^2 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix} - \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \end{bmatrix} K(\mathbf{t}, \mathbf{t})^{-1} \begin{bmatrix} C_\theta(\mathbf{t}, \mathbf{t}^*) \\ \partial_2 C_\theta(\mathbf{t}, \mathbf{t}^*) \\ \partial_2^2 C_\theta(\mathbf{t}, \mathbf{t}^*) \end{bmatrix} \end{aligned}$$

The posterior moments of f are useful for assessing the fit of the model. By completing the expressions in the previous display we obtain these as

$$\begin{aligned} \mu_f(\mathbf{t}^*) &= \mu_\beta(\mathbf{t}^*) + C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\ \Sigma_f(\mathbf{t}^*, \mathbf{t}^*) &= C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} C_\theta(\mathbf{t}, \mathbf{t}^*) \end{aligned}$$

The posterior moments of df are required in order to calculate the Trend Direction Index in Proposition 2, and these are given by

$$\begin{aligned}\mu_{df}(\mathbf{t}^*) &= d\mu_\beta(\mathbf{t}^*) + \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\ \Sigma_{df}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t})\end{aligned}\tag{7}$$

In order to calculate the Expected Trend Instability index in Proposition 3 we in addition require the posterior moments of d^2f and its cross-covariance with df . These are given by

$$\begin{aligned}\mu_{d^2f}(\mathbf{t}^*) &= d^2\mu_\beta(\mathbf{t}^*) + \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\ \Sigma_{d^2f}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1^2 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \Sigma_{df, d^2f}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t})\end{aligned}\tag{8}$$

6.2 Proof of Proposition 3

The generalized Rice formula for the expected number of zero crossings of a Gaussian process X [TODO: Conditions] on an interval \mathcal{I} is

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

where $f_{X(t), dX(t)}$ is the joint density function of X and dX at time t (Rice 1944).

From Proposition 1 we have that $(df, d^2f) \mid \Theta$ follows a bivariate Gaussian process. Under the conditions of [TODO: ??] we may insert the posterior moments of $(df, d^2f) \mid \Theta$ and through a direct integration of the expression in the previous display we obtain that the Expected Trend Instability index generally defined in Equation (2) is given by

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

where $\phi(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2)$ is the standard normal density function, $\text{Erf}(x) = 2\pi^{-1/2} \int_0^x \exp(-u^2) du$ is the error function, and ω and η are functions defined by

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

and μ_{df} , μ_{d^2f} , Σ_{df} , Σ_{d^2f} and Σ_{df, d^2f} are defined in Equations (7) and (8). [TODO: Check formula for η]