

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*

20 March, 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 kinds of statistical assessments are very important as they may influence public health decisions on a national level.

Standard change-point models are not applicable since there are no observations beyond the time of the stipulated change-point. 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 and can be used for both estimation and forecasting. [TODO: Also mention ETI.]

We show how the Trend Direction Index 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. We also show how our model can be extended to include a time-varying latent class variable enabling data-driven adaptation to underlying non-stationarity.

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 børte 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?“].

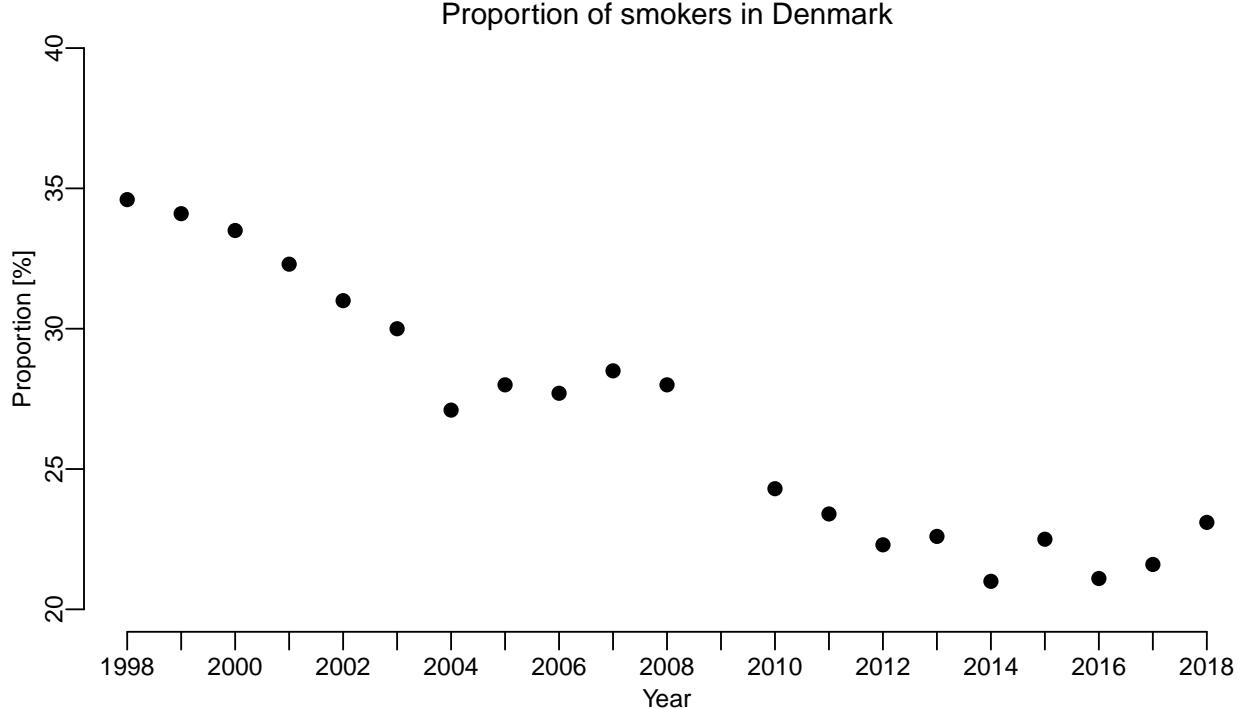


Figure 1: [TODO: add title]

[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). Empirical dynamics for longitudinal data has been studied by Müller and Yao (2010).

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 3 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 4, 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. 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 fixed time points where sudden a change happens (see e.g., 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 a standard change-point model as there are no observations beyond the stipulated change-point.

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. We quantify the trend instability by the expected number of times f changes direction on the interval, which is equivalent to either counting the expected number local maxima of f or the number of zero crossings by df . 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 sigma algebra \mathcal{F} . The value of ETI is a finite [TODO: finite because of assumptions] real positive number, and the lower $\text{ETI}(\mathcal{I})$ is, the more trend stable f is on \mathcal{I} and vice versa. [TODO: Maybe something about normalizing by \mathcal{I} or the interpretation.]

2.1 Latent 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 semi-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

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

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*

1. a
2. b

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

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

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 $(f(\mathbf{t}^*), df(\mathbf{t}^*), \mathbf{Y}) | \mathbf{t}, \Theta$ is multivariate normal, and that the marginal distribution of the observed data is $\mathbf{Y} | \mathbf{t}, \Theta = N(\mu_\beta(\mathbf{t}), C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)$ 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 $(f(\mathbf{t}^*), df(\mathbf{t}^*)) | \mathbf{Y}, \mathbf{t}, \Theta$. 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*

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

where

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

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

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 states 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. We have set the mean of the Gaussian Process to $\mu_\beta(t) = 0$ and used the squared exponential (SE) covariance function. [TODO: Move this to estimation] This is an often used covariance function in Gaussian Process regression (C. E. Rasmussen and Williams 2006) and it has the following form

$$C_\theta(s, t) = \alpha^2 \exp\left(-\frac{(s-t)^2}{2\rho^2}\right), \quad \theta = (\alpha, \rho) \tag{5}$$

The two parameters have the interpretation that α is the standard deviation of the latent function and ρ is its length-scale. For the example we have used $\alpha = 2$ and $\rho = 0.1$ and consider the noise free case with $\sigma^2 = 0$. 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

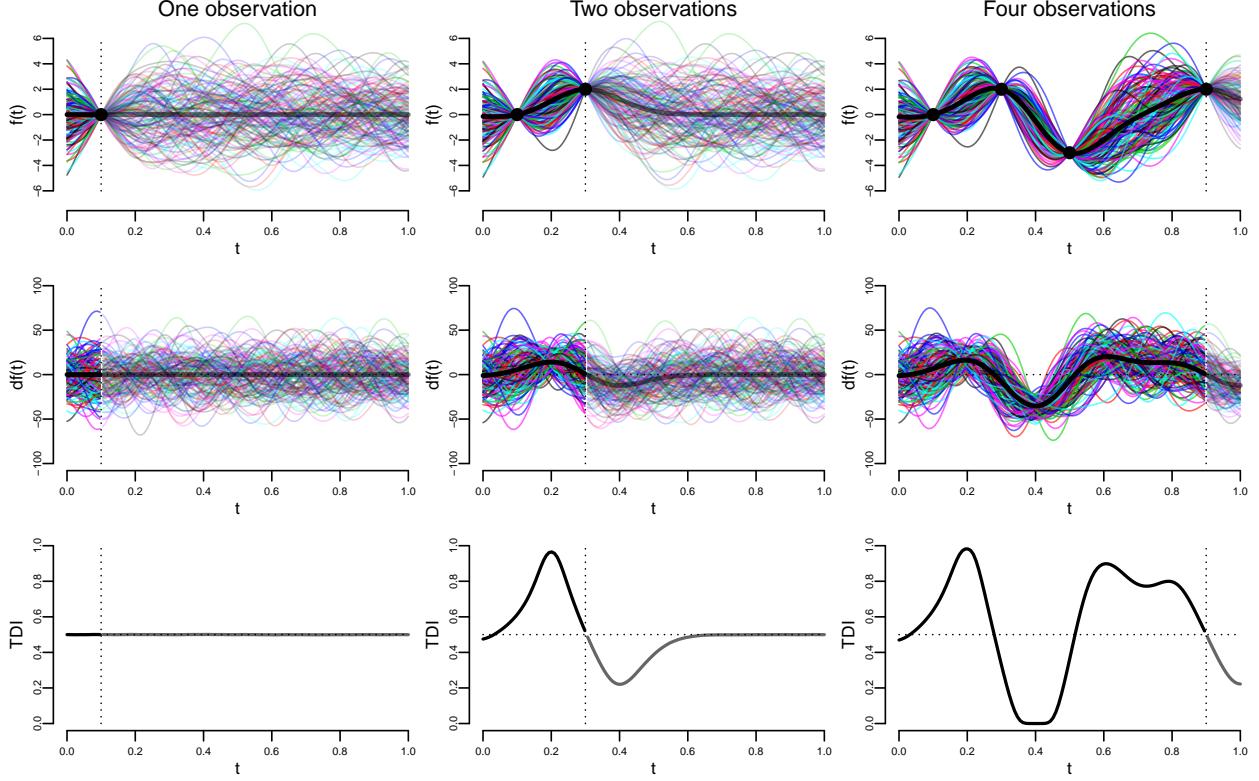


Figure 2: 150 realizations from the posterior distribution of f (top row), df (middle row) with expected values in bold and the Trend Direction Index (bottom row) conditional on one, two and four noise free observations. Dotted vertical lines show the points in time after which forecasting occurs.

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]

2.3 Expected Trend Instability

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

Proposition 3.

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

where ϕ is the standard normal density function, Erf is the error function with $\frac{1}{2} (\text{Erf}(2^{-1/2}x) + 1)$ equal to the standard normal cumulative distribution function, and ρ and η are functions defined by

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

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.

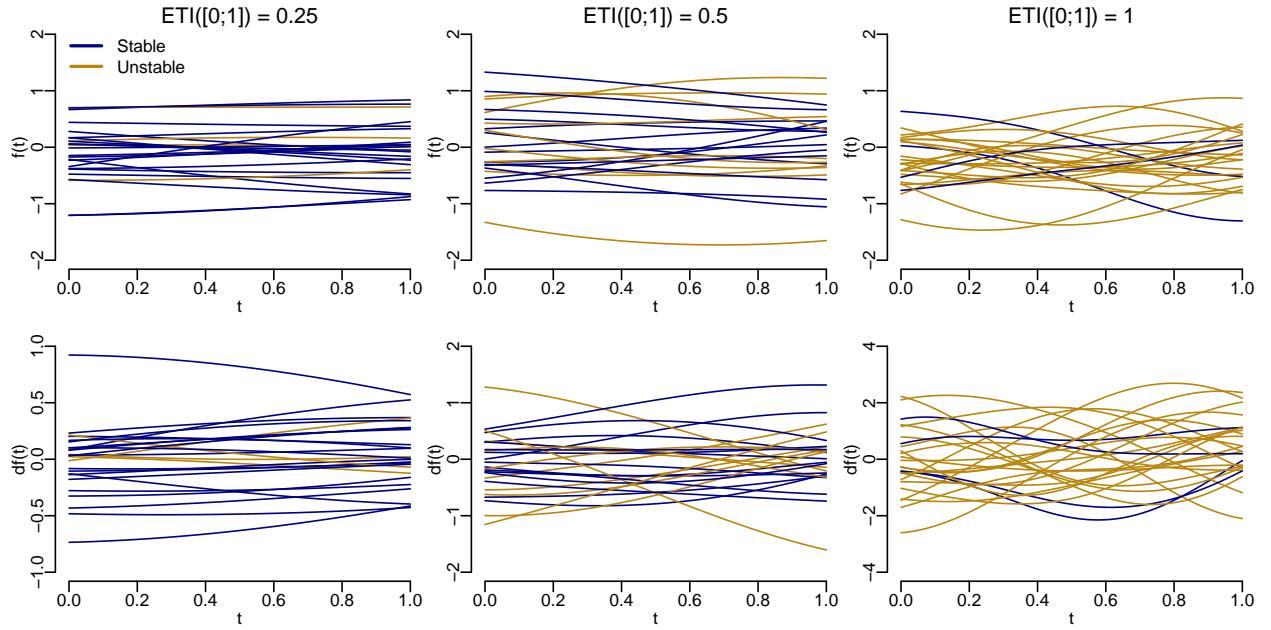


Figure 3: Realizations from the joint distribution of (f, df) using the squared exponential covariance function under different values of the Expected Trend Instability (ETI). Sample paths that are trend stable are shown with blue lines, and unstable sample paths are gold colored.

2.4 Data-driven non-stationarity (better title needed)

By construction, the distribution of the derivative at time t conditional on the value of the function at the same time is a linear function, and the latent dynamics is therefore governed by linear first-order differential equation with time-varying coefficients that can be written as

$$df(t) \mid f(t), \beta, \theta = a(t) + b(t)f(t) + \epsilon(t)$$

where $b(t) = \text{Cov}[df(t), f(t) \mid \beta, \theta] \text{Cov}[f(t), f(t) \mid \beta, \theta]^{-1}$.

[TODO: Fix from here] The generating model described hitherto depends on a parametric covariance function $C_\theta(\cdot, \cdot)$ governing the dynamics of the latent function. Usually in Gaussian Process Regression a stationary covariance function, such as the SE function in Equation (5) is chosen out of convenience (C. E. Rasmussen and Williams 2006). Choosing a stationary covariance function, however, implies an important model assumption. It can be shown that for all stationary covariance

functions $C_\theta(s, t) = C_\theta(|s - t|)$, it follows that $\partial_2 C_\theta(t, t) = 0$ and therefore $f(t) \perp\!\!\!\perp df(t)$ (Cramer and Leadbetter 1967). Using a model with a stationary covariance function implies that $b(t) = 0$, and the model must therefore require non-stationarity in order to have non-trivial latent dynamics with $b(t) \neq 0$.

It can be argued that a priori selecting a parametric non-stationary covariance function that fits data is a very difficult problem. We now present an extension to the previous model that facilitates a data-driven adaptation to non-stationarity in f . We do this by introducing a latent time-varying categorical variable $G(t)$ taking values in the discrete set $\{1, \dots, K\}$ at every time t . We then assume that f conditionally on $G(t) = k$ is a Gaussian Process with a stationary covariance function indexed by a parameter θ_k . This corresponds to the following functional latent class model

$$\begin{aligned} w(t) | \phi &\sim \mathcal{S}_{K-1}^\infty(\phi) \\ G(t) | w(t), \phi &\sim \text{Multinomial}(w(t)) \\ f(t) | G(t) = k, \theta_k, w, \phi &\sim \mathcal{GP}(m, C_{\theta_k}(\cdot, \cdot)) \end{aligned} \tag{6}$$

where w is a K -dimensional probability function and \mathcal{S}_{K-1}^∞ is a distribution on the interior of the functional $K - 1$ dimensional simplex

$$\left\{ w_k(t), k = 1, \dots, K, t \in \mathcal{T}, w_k(t) > 0, \sum_{k=1}^K w_k(t) = 1 \right\}$$

Marginalizing out the distribution of $G(t)$ in $f(t) | G(t) = k, \theta_k, w, \phi$ we may write the joint distribution of f and df conditional on the probability weights in the samme form as in Equation (4) but with the following covariance and cross-covariance functions

$$\begin{aligned} C(s, t) &= \sum_{k=1}^K w_k(s) C_{\theta_k}(s, t) w_k(t) \\ \partial_1 C(t, s) &= \sum_{k=1}^K dw_k(s) C_{\theta_k}(s, t) w_k(t) + \sum_{k=1}^K w_k(s) \partial_1 C_{\theta_k}(s, t) w_k(t) \\ \partial_2 C(s, t) &= \sum_{k=1}^K w_k(s) C_{\theta_k}(s, t) dw_k(t) + \sum_{k=1}^K w_k(s) \partial_2 C_{\theta_k}(s, t) w_k(t) \\ \partial_1 \partial_2 C(s, t) &= \sum_{k=1}^K dw_k(s) C_{\theta_k}(s, t) dw_k(t) + \sum_{k=1}^K w_k(s) \partial_1 C_{\theta_k}(s, t) dw_k(t) \\ &\quad + \sum_{k=1}^K dw_k(s) \partial_2 C_{\theta_k}(s, t) w_k(t) + \sum_{k=1}^K w_k(s) \partial_1 \partial_2 C_{\theta_k}(s, t) w_k(t) \end{aligned} \tag{7}$$

The construction with a latent class implies that the covariance function of f , C , becomes a time-varying convex sum of stationary covariance functions with separate parameters. The expressions for the covariance function of df and its cross-covariance with f follows according to Equation (4)

and the chain rule. From the expression of cross-covariance it can be seen that $\partial_2 C(t, t)$ is generally non-zero due to the existence of the terms in the first sum. The terms in the second sum will, however, still equal to zero on the diagonal when each class uses a stationary covariance function. The joint posterior distribution for this extended model is then given similarly to the expressions in Equation (??) with the appropriate covariance substitutions.

To model the functional class probabilities, $w(t)$, in Equation (6) we suggest mapping $K - 1$ linear basis expansions onto the simplex. The basis could for example be polynomial splines where we require that the degree is chosen such they are at least once continuously differentiable on \mathcal{T} . Let $\{B_p(t) : t \in \mathcal{T}, p = 1, \dots, P\}$ be a set of basis functions on \mathcal{T} with P degrees of freedom and $(\phi_0^{(k)}, \dots, \phi_P^{(k)})$ a vector of coefficients for each $k = 1, \dots, K - 1$ expansion. We then construct the $K - 1$ functions as

$$\eta_k(t) = \phi_0^{(k)} + \sum_{p=1}^P \phi_p^{(k)} B_p(t), \quad k = 1, \dots, K - 1 \quad (8)$$

which we map into the class probabilities by the inverse log-additive ratio transform

$$w_k(t) = \frac{\exp(\eta_k(t))}{1 + \sum_{j=1}^{K-1} \exp(\eta_j(t))}, \quad k = 1, \dots, K - 1, \quad w_K(t) = \frac{1}{1 + \sum_{j=1}^{K-1} \exp(\eta_j(t))}$$

Note that the derivatives of w_k are required in the calculation of the covariance matrices in Equation (7). These can be derived analytically using the chain rule and pre-computed derivatives of the basis functions.

The parameters of this extension to the model are therefore $\Theta = (m, (\phi_0^{(k)}, \dots, \phi_P^{(k)})_{k=1}^{K-1}, (\theta_k)_{k=1}^K, \sigma^2)$ and the Trend Direction Index in Proposition 2 is as before defined conditionally on these variables.

3 Estimation

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.

[TODO: Write about the covariance function here]

In our application we use the squared exponential covariance function defined in Equation (5) which means that our model without the latent class has the following parameters $\Theta = (\beta, \alpha, \rho, \sigma)$. We suggest using the following prior distributions

$$\beta \stackrel{iid}{\sim} N(m_\beta, 3), \quad \alpha \sim \text{Half-Normal}(0, 3), \quad \rho \sim \Gamma^{-1}(\rho_a, \rho_b), \quad \sigma \sim \text{Half-Normal}(0, 3)$$

where Half-Normal denotes the half-normal distribution due to the requirement of positivity of α and σ and Γ^{-1} is the inverse Gamma distribution (**TODO:** This is not scale-free. Fix). These provide weakly informative priors for the model parameters. We set the hyper-parameter μ_m in the prior for m as the empirical average of the observed outcomes, $\mu_m = \frac{1}{n} \sum_{i=1}^n Y_i$. The hyper-parameters for the length-scale parameter ρ deserve special attention. The sampling design of the data, i.e., the values of t_1, \dots, t_n , restricts the possible values of estimable length-scales. Specifically, it is impossible to have a length-scale of the latent function smaller than smallest distance between any two sampling times. Similarly, it is impossible to have a length-scale that is greater than the range of the sampling times. To include this information in the prior distribution of ρ while simultaneously not enforcing hard constraints we find the values of ρ_a and ρ_b such that $P(\rho_{\min} \leq \rho \leq \rho_{\max}) = 1 - \gamma$ where $\rho_{\min} = \min_{i,i'} |t_i - t'_i|$, $\rho_{\max} = \max_{i,i'} |t_i - t'_i|$ and γ is a small probability, e.g., 5% or 1%. The values of ρ_a and ρ_b can be found using numerical optimization.

To estimate the extended model with a latent class we further need to impose prior distributions on the spline coefficients for the probability functions w_k . A well-known problem when using spline expansions is selecting the number of knots or equivalently the degrees of freedom of the basis. To mitigate this problem we suggest using a large number of knots and then regularize the splines through the prior distributions on their coefficients. It can be shown that if all spline coefficients are equal, then the resulting expansion is a constant functions. This motivates imposing a random walk prior on the coefficients of each η_k in Equation (8) in order to penalize local variability. For each $k = 1, \dots, K - 1$ spline expansion we therefore use the following prior distribution

$$\phi_0^{(k)} \sim N(0, 1), \quad \phi_1^{(k)} \sim N(0, 1), \quad \{\phi_p^{(k)} \sim N(\phi_{p-1}^{(k)}, \tau^{(k)})\}_{p=2}^P, \quad \tau^{(k)} \sim \text{Half-Normal}(0, 1)$$

We have implemented both our model with and without a latent class 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 for our two models are available at the first author's Github repository (Jensen 2019).

4 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. The result 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 inference on this statement but wrote in a note that because the study population is so large, then more or less all differences become statistically significant at the 95% level. This is an illustrative example of our proposed Trend Direction Index. If it really is so that the proportion of smokers have increased in 2018 we should expect that the Trend Direction Index in 2018 conditional on data from the last 20 years would have a high value (or at least larger than 50%) when fitting our model to this data.

[**TODO:** Since the outcomes are proportions calculated from known sample size we know the marginal variance of the outcomes by the formula for the variance of a sample proportion. In the

stationary case with a standardized covariance function such that $C_\theta(t, t) = \alpha^2$ we obtain according to Equation (3) the following equality constraint between the variance of the latent function and the residual variance

$$\text{Var}[Y_i, | t_i, \Theta] = \alpha^2 + \sigma^2 = \frac{Y_i(1 - Y_i)}{n_i}$$

where n_i is the number of responders at time t_i . How do we best enforce this constraint? **TODO:** Maybe all of this is an unnecessary complication that we don't need to mention.]

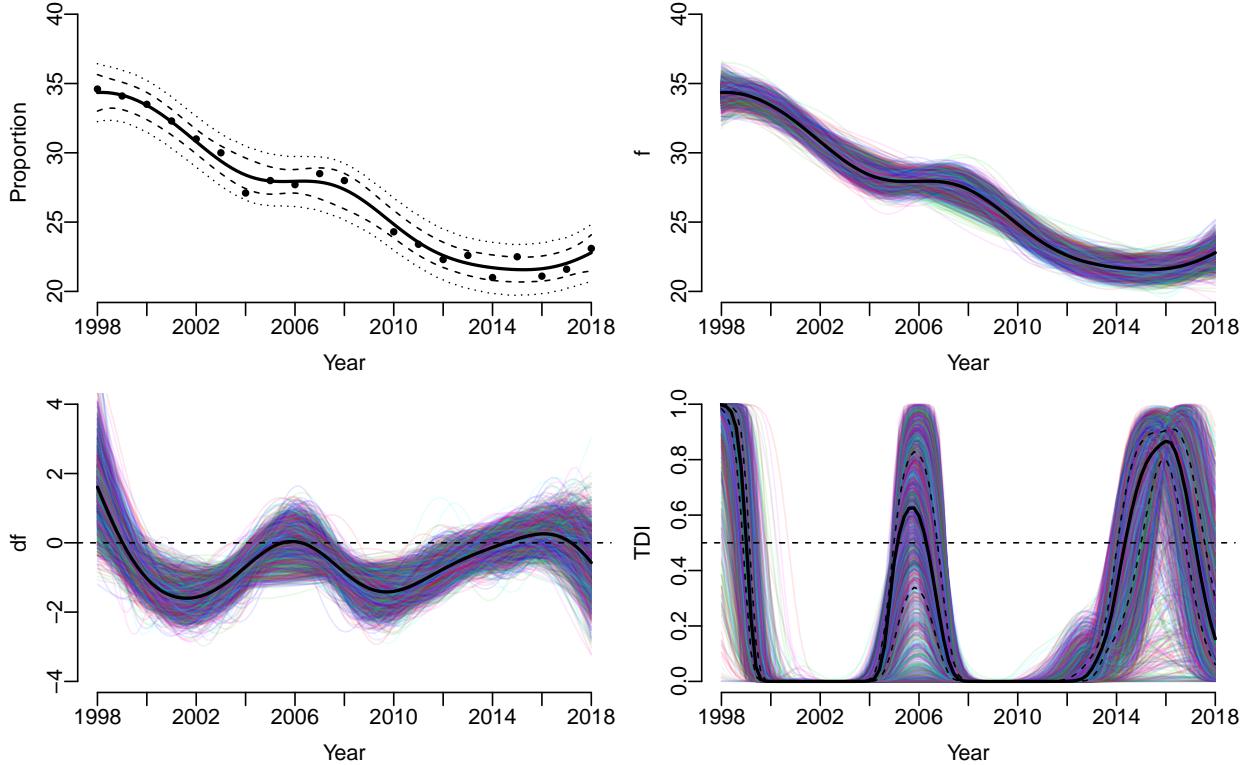


Figure 4: hejhej

We fitted the model without a latent class to the data using the Stan implementation with the squared exponential covariance function. We ran four chains for 25000 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). The results from the fitted model are seen in Figure 4.

The upper left panel of Figure 4 shows a plot of the raw data points along with the posterior mean of the latent function as the bold line. The dashed lines are lower and upper 95% probability intervals for posterior distribution of f and the dotted lines are lower and upper 95% probability intervals for the posterior predictive distribution of the proportions. The upper right panel shows 3000 random realizations from the posterior distribution of f and again the posterior mean as the bold line. The lower left panel shows 3000 random realizations from the posterior distribution of the derivative of the latent function with the posterior mean as a bold line, at the lower right panel shows 3000 random realizations from the distribution of the Trend Direction Index along with the mean as the bold line.

The figure shows some interesting things. First, around the year 2006 a increasing trend was also seen even though the average Trend Direction Index was only around 60%. This was followed by a long period of close to zero probability of a positive derivative. Secondly, the current reported increase in the proportion of smokers appears to have started already four years ago in 2014 when the average Trend Direction Index started to climb above 50%. This peaked around 2016 with a Trend Direction Index of 93%. Surprisingly, the Trend Direction Index immediately decreases and in 2018 at the latest survey results it is only 14%. This suggests with a high probability that there has been an increase in the proportion of smokers in Denmark but that began four years ago and currently there is no strong evidence for this increase to still occur. The value of the Trend Direction Index in 2018 even suggests that it is decreasing again.

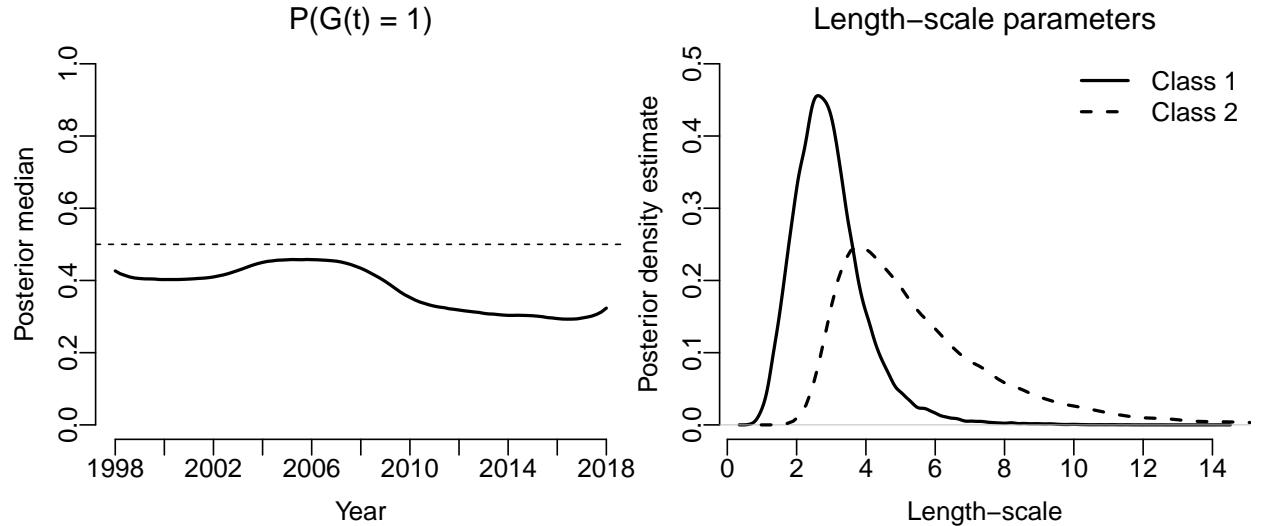


Figure 5: Median posterior class probability (left) and kernel density estimates of the posterior length-scales (right) for the non-stationary model fitted with two latent classes.

We also fitted a model with $K = 2$ latent classes to the data. We used a cubic spline basis with 15 degrees of freedom for the functional class probabilities. The left panel of Figure 5 shows the posterior median of the probability that the latent class is equal to class 1 at any given time. The right panel shows kernel density estimates of the posterior distributions of the length-scale parameter for each latent class. The plot suggests that some degree of non-stationarity is present in the latent dynamics. The latent function seems to develop at two length-scales, one with a posterior mode of 2.78 and another with a posterior mode of 3.79. The smaller length-scale has a higher influence during the first ten years with a bump around 2004 corresponding to the close to constant region seen in the upper left panel of Figure 4. Figure 6 shows results similar to Figure 4 but for the latent class model. The conclusions from this model fit are similar to the ones from before.

5 Discussion

[TODO: Her må der båttes løs]

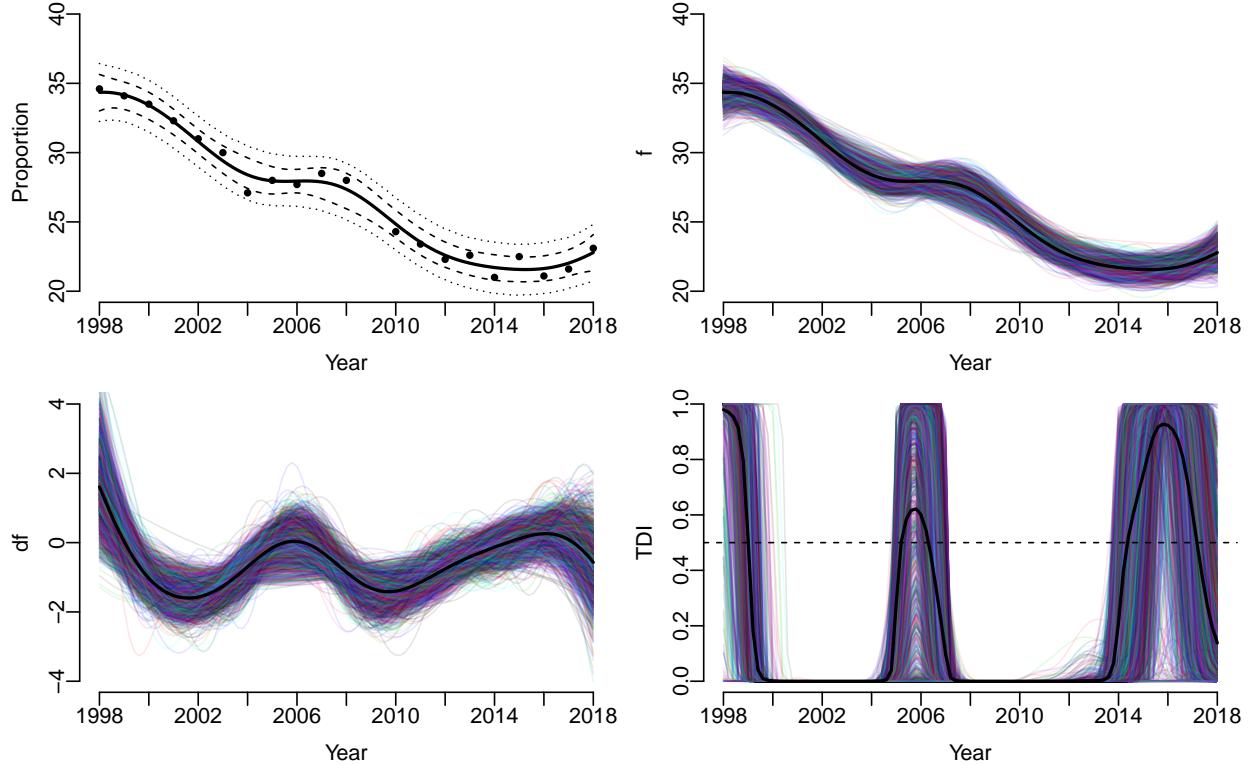


Figure 6: THIS IS NOT THE CORRECT FIGURE YET (men den er god nok til jazz...)

6 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>.

Müller, Hans-Georg, and Fang Yao. 2010. “Empirical Dynamics for Longitudinal Data.” *The Annals of Statistics* 38 (6): 3458–86.

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.

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes in Machine Learning*. MIT Press.

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. “Danskerne Rygevaner 2018.” www.sst.dk/da/udgivelser/2019/danskerne-rygevaner-2018.

7 Appendix

7.1 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 $\mathbf{Y} | \mathbf{t}, \Theta$ 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

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

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

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

$$\Sigma(\mathbf{t}^*, \mathbf{t}^*) = \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \\ \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1^2 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) & \partial_1^2 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix} - \begin{bmatrix} C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \\ \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \end{bmatrix} K(\mathbf{t}, \mathbf{t})^{-1} \begin{bmatrix} C_\theta(\mathbf{t}, \mathbf{t}^*) \\ \partial_2 C_\theta(\mathbf{t}, \mathbf{t}^*) \\ \partial_2^2 C_\theta(\mathbf{t}, \mathbf{t}^*) \end{bmatrix}$$

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

$$\mu_f(\mathbf{t}^*) = \mu_\beta(\mathbf{t}^*) + C_\theta(\mathbf{t}^*, \mathbf{t}) (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t}))$$

$$\Sigma_f(\mathbf{t}^*, \mathbf{t}^*) = C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} C_\theta(\mathbf{t}^*, \mathbf{t})$$

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

$$\mu_{df}(\mathbf{t}^*) = d\mu_\beta(\mathbf{t}^*) + \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \quad (9)$$

$$\Sigma_{df}(\mathbf{t}^*, \mathbf{t}^*) = \partial_1 \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t})$$

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

$$\mu_{d^2f}(\mathbf{t}^*) = d^2\mu_\beta(\mathbf{t}^*) + \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \quad (10)$$

$$\Sigma_{d^2f}(\mathbf{t}^*, \mathbf{t}^*) = \partial_1^2 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t})$$

$$\Sigma_{df, d^2f}(\mathbf{t}^*, \mathbf{t}^*) = \partial_1 \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) (C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1} \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t})$$

7.2 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 $(df, d^2f) | \Theta$ follows a bivariate Gaussian process. Under the conditions of [TODO: ??] we may insert the posterior moments of $(df, d^2f) | \Theta$ 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

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

where ϕ is the standard normal density function, Erf is the error function with $\frac{1}{2} (\operatorname{Erf}(2^{-1/2}x) + 1)$ equal to the standard normal cumulative distribution function, ρ 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 (9) and (10).

7.3 Remove later

Proposition 4. Let the data generating model be defined as in Equation (3) and \mathbf{t}^* any finite vector of time points. Then the posterior distribution of (f, df) at \mathbf{t}^* is given by

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