arXiv:1911.06722v2 [stat.ME] 16 Jul 2020

Causal inference using Bayesian non-parametric quasi-experimental design

Max Hinne, Marcel A. J. van Gerven and Luca Ambrogioni

Radboud University, Nijmegen, The Netherlands
m.hinne@donders.ru.nl

July 17, 2020

Abstract

The de facto standard for causal inference is the randomized controlled trial, where one compares an manipulated group with a control group in order to determine the effect of an intervention. However, this research design is not always realistically possible due to pragmatic or ethical concerns. In these situations, quasi-experimental designs may provide a solution, as these allow for causal conclusions at the cost of additional design assumptions. In this paper, we provide a framework for quasi-experimental design using Bayesian model comparison. We provide a theoretical motivation for a Gaussian process based approach, and demonstrate its convenient use in a number of simulations. Finally, we apply the framework to determine the effect the 2005 smoking ban in Sicily on the number of acute coronary events, and of the effect of an alleged historical phantom border in the Netherlands on Dutch voting behaviour.

1 Introduction

The bread and butter of scientific research is the randomized-controlled trial (RCT) [1]. In this design, the sample population is randomly divided into two groups; one that is manipulated (e.g. a drug is administered or a treatment is performed), while the other is left unchanged. Because of the random group assignment, we may assume (if the sample size is large enough) that the distributions of all irrelevant variables are the same between the two groups, so that if we measure a difference, this can only be due to the intervention. The consequence is that the RCT allows us to perform *causal inference*, that is, via this approach we can learn about the causal effect of the intervention [2].

While the RCT is a staple tool in the toolbox of any scientist, in practice there may be several insurmountable hurdles that deter one from using it. For instance, ethical considerations may force a study to halt midway when the intervention appears to cause serious adverse effects. Other objections may be pragmatic; it may be impossible to obtain sufficiently equivalent groups of samples if these samples represent for instance cities or species. Luckily, even in these cases all is not lost for causal inference. There exist several *quasi*-experimental designs (QEDs) that replace random assignment with deterministic assignment, that still allow for valid causal inferences at the cost of some additional assumptions. For example, regression discontinuity (RD) design [3–5] and difference-within-differences (DiD) [6] assign a sample to one of the two groups based on it passing or failing a threshold on an assignment variable. For instance, a patient may be assigned to the intervention group if their blood cholesterol levels exceed a certain predetermined limit. A closely related quasi-experimental design is interrupted time series (ITS) analysis, which is essentially RD, but with time as its assignment variable [7].

The idea behind these approaches is that, around the assignment threshold, data points are distributed essentially randomly, so that locally the conditions of RCT are recreated [8, 9]. Crucially, for this to work, these methods assume that the sampled population cannot directly influence its score on the assignment variable. For example, if a patient can actively manipulate their own cholesterol level they can determine whether they are in the intervention or the control group, which would render the conclusion invalid. The methodological pipeline of quasi-experimental designs like these generally consists of three

steps [10]. First, a regression is fit to each of the two groups individually. Typically, one parametrizes this as linear regression. Next, the fits are extrapolated to the threshold. Finally, the difference between the extrapolations of the two groups at the boundary is taken as the effect size of the intervention. A straightforward statistical test can be applied to check whether the effect is present.

Although RD, ITS and DiD have been applied frequently in education, econometrics and political science, recent studies have pointed out that they are underutilized in other domains, such as neuroscience, behavioural science and medicine [11, 12]. In this paper we provide a novel Bayesian non-parametric framework for quasi-experimental design, which we refer to as BNQD. BNQD is a Bayesian approach to quasi-experimental design that offers several advantages to existing approaches for RD and ITS.

First, we frame the problem of detecting a difference between the two groups as Bayesian model comparison. Instead of comparing the pre- and post-intervention regressions, we introduce a continuous model and a discontinuous model. In the discontinuous model, observations before and after the intervention are assumed to be independent,while in the continuous model this assumption is omitted. We then quantify the evidence in favor of either model, rather than only for the alternative model, via the Bayes factor [13]. Furthermore, our approach makes it possible to compute the Bayesian model averaged effect size, which provides a more nuanced estimate compared to implicitly conditioning on the alternative model [14, 15].

Second, our approach is non-parametric and is based on Gaussian process (GP) regression. This means that we do not impose strong constraints on the functional form of the regression in our models. Instead, we specify a covariance function that describes how similar any two points of the response are, depending on the distance between their predictors. The result is a flexible model that can capture all sorts of nonlinear interactions between the assignment and the outcome variable. By appropriate selection of the GP covariance function, BNQD serves as an alternative for several traditional quasi-experimental designs. For instance, with a linear covariance function, BNQD can be used for linear RD. Alternatively, with an exponential covariance function BNQD serves as a non-parametric form of RD design. Finally, BNQD can just as easily be applied to multiple assignment variables [16–18], as we demonstrate in one of our examples.

Third, in most methods for quasi-experimental design, a bandwidth parameter determines the trade-off between estimation reliability and the local randomness assumptions that are needed to draw causal inferences [19]. In BNQD, all observations are used to estimate both the continuous and discontinuous model, but by optimizing the length scale parameter of the GP covariance functions we control the sensitivity to different types of discontinuities and adherence to locality assumptions.

The paper is organized as follows. First, we explain the proposed method. We then assess its performance in simulations, and demonstrate how it can be applied in practice. We proceed with a discussion of related work and future directions for Bayesian quasi-experimental designs.

2 Discontinuity-based causal inference

We use the potential outcomes framework [20] to define, for an observation i and predictor x_i, the two potential outcomes $y_i(0)$ and $y_i(1)$ that represent the responses for i in the untreated and treated conditions, respectively. Furthermore, let $\ell_i = 1$ indicate that the intervention is applied to unit i, and $\ell_i = 0$ if it has not. In (sharp) regression discontinuity design, ℓ is a simple function of x_i, namely

$$\ell(x_i) = \begin{cases} 1 & \text{if } x_i \geq x_0 \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

where x_0 is the threshold that determines above which values for the forcing variable x the intervention is applied. The effect size (also referred to as treatment effect [21] or causal effect [22]) can now be defined as

$$d_{\text{RD}} = \mathbb{E}[y_i(1) - y_i(0) \mid x_i = x_0] \ .$$

Now, if the distributions of the response y_i given the predictor variable x_i are continuous in x, and the conditional expectations $\mathbb{E}[y_i(1) \mid x_i]$ and $\mathbb{E}[y_i(0) \mid x_i]$ exist, then the effect size can be computed as [8, 19, 23]

$$d_{\text{RD}} = \lim_{x \downarrow 0} \mathbb{E}[y_i \mid x_i = x] - \lim_{x \uparrow 0} \mathbb{E}[y_i \mid x_i = x] \ .$$

To learn d, one can estimate the regressions on either side of the threshold, and evaluate their difference at x_0, $d = f_I(x_0) - f_C(x_0)$, where f_I and f_C are the intervention and control regression functions, respectively [23]. To learn whether the treatment has an effect, one simply tests for $d_{\mathrm{RD}} = 0$.

Importantly, this approach assumes that observations (e.g. participants) cannot control the value of x_i, so that near the threshold x_0 observations are assigned approximately randomly, thus locally recreating the conditions for randomized design. Furthermore, the approach assumes that there are no confounding variables that affect the relationship between the forcing variable x and outcome y (for a more in-depth discussion of the assumptions for valid RD design, see Geneletti et al. [19]).

3 Bayesian non-parametric quasi-experimental design

We approach discontinuity-based quasi-experimental designs in a Bayesian way. Rather than focusing primarily on the estimation of the effect size d, we perform Bayesian model comparison to see whether the data are supported better by the alternative model $\mathcal{M}_1$, that claims an effect is present, than by the null model $\mathcal{M}_0$, in which such an effect is not present.

Let $D = \{(x_i, y_i)\}_{i=1}^n$, with $x_i \in \mathbb{R}^p$and $y_i \in \mathbb{R}$ (for this exposition we will assume $p = 1$, but extension to multidimensional input is straightforward, as we will show in an example later on). The result of the model comparison is quantified by the Bayes factor [24, 25]:

$$\mathrm{BF}_{10} = \frac{p(D \mid \mathcal{M}_1)}{p(D \mid \mathcal{M}_0)} \ . \tag{1}$$

Here, $p(D \mid \mathcal{M}_1)$ and $p(D \mid \mathcal{M}_0)$ are the model marginal likelihoods of the two models that have integrated out their respective (hyper)parameters. The Bayes factor indicates how much more likely the data are given the discontinuous model, compared to the continuous model [26]. Unlike a p-value, it can provide evidence for either model, so that it is possible to find evidence supporting the *absence* of a discontinuity [13, 27].

Both models imply a distribution over the effect size. In the null model, all probability mass of $p(d \mid D, \mathcal{M}_0)$ is concentrated at $d = 0$, while for the alternative model we can infer the posterior distribution of the effect size $p(d \mid D, \mathcal{M}_1)$. Importantly, existing regression discontinuity methods that focus on the inference of d, implicitly condition on $\mathcal{M}_1$. This results in an overconfident estimate of the effect size, and consequently of too optimistic conclusions of the efficacy of an intervention, as this approach ignores the uncertainty in the model posterior $p(\mathcal{M} \mid D)$. This uncertainty can be accounted for via the Bayesian model average estimate of d, which is given by

$$p(d \mid D) = \sum_j p(d \mid D, \mathcal{M}_j) p(\mathcal{M}_j \mid D) \ . \tag{2}$$

The resulting mixture distribution integrates out the uncertainty of the model, which has been shown to lead to optimal predictive performance [14, 15]. Below, we proceed to explain the distributions implied by the two models in more detail.

3.1 The continuous model

In the continuous model, it is assumed that the treatment has no effect, which implies that the regression does not depend on the threshold. This leaves us with a single regression for all data points. We assume Gaussian observation noise:

$$y_i \sim \mathcal{N}\left(f(x_i), \sigma_n^2\right) \ .$$

Here, σ_n^2 is the observation noise variance, and $f(x_i)$ captures the relationship between the predictor and the expected response. We make no parametric assumptions about f, and instead assume f follows a Gaussian process (GP). What results is simply Gaussian process regression [28]:

$$f_0(x) \mid \mathcal{M}_0 \sim \mathcal{GP}\left(\mu(x), k(x, x'; \theta_0)\right) \ .$$

The parameters of the GP, $\mu(x)$ and $k(x, x'; \theta)$, are again functions and have the following meaning. The mean function $\mu(x)$ specifies the expected mean of the functions that may be drawn from the process.

Typically, a constant zero function is assumed. The essential component of the Gaussian process is the *covariance function* $k(x, x'; \theta)$, which specifies for two distinct points x and x' how similar the function values $f(x)$ and $f(x')$ are, depending on a set of hyperparameters θ (in terms of parameter estimation, the observation noise σ^2 is considered an element of θ).

Using textbook results of Gaussian distributions, we can conveniently marginalize out f_0, that is, with $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^T$,

$$\begin{aligned} \log p(\boldsymbol{y} \mid \boldsymbol{x}, \theta_0, \mathcal{M}_0) &= \log \left[\int p(\boldsymbol{y} \mid \boldsymbol{x}, f_0, \theta, \mathcal{M}_0) p(f_0 \mid \theta, \mathcal{M}_0) \, \mathrm{d} f_0 \right] \\ &= -\frac{1}{2} \boldsymbol{y}^T (K + \sigma_n^2 I)^{-1} \boldsymbol{y} - \frac{1}{2} \log \left| K + \sigma_n^2 I \right| - \frac{n}{2} \log 2\pi \ , \end{aligned} \tag{3}$$

with K the matrix of pairwise covariances between all training points, and I the n-element identity matrix [28]. However, the hyperparameters θ_0 cannot be integrated out so easily. We adopt a strategy common in machine learning, which is to approximate the model marginal likelihood by the Bayesian information criterion (BIC) instead, defined as

$$\begin{aligned} \log p(D \mid \mathcal{M}_0) = \log p(\boldsymbol{y} \mid \boldsymbol{x}, \mathcal{M}_0) &= \int p(\boldsymbol{y} \mid \boldsymbol{x}, \theta_0, \mathcal{M}_0) p(\theta_0 \mid \mathcal{M}_0) \, \mathrm{d}\theta_0 \\ &\approx \log p(\boldsymbol{y} \mid \boldsymbol{x}, \hat{\theta}_0, \mathcal{M}_0) - \frac{k}{2} \log n \end{aligned} \ ,$$

where k is the number of hyperparameters and

$$\hat{\theta}_0 = \arg\max_{\theta_0} p(\boldsymbol{y} \mid \boldsymbol{x}, \theta_0, \mathcal{M}_0)$$

are the optimized hyperparameters. This approximation is warranted as we typically have $n \gg k$.

3.2 The discontinuous model

In the alternative model, we assume that a treatment effect is present. That is, the latent processes below and above x_0 differ by the effect d, measurable at $x = x_0$. The consequence of the assumption of treatment effect is that we can factorize the model marginal likelihood as $p(D \mid \mathcal{M}_1) = p(D_C \mid \mathcal{M}_1) p(D_I \mid \mathcal{M}_1)$, where $D_C = \{(x, y) \in D \mid x < x_0\}$ and $D_I = \{(x, y) \in D \mid x \geq x_0\}$ ($D = D_C \cup D_I$). We proceed by assuming each subset of the data follows its own regression:

$$\begin{aligned} f_C(x) \mid \mathcal{M}_1 &\sim \mathcal{GP}\left(\mu(x), k(x, x'; \theta_1)\right) \\ f_I(x) \mid \mathcal{M}_1 &\sim \mathcal{GP}\left(\mu(x), k(x, x'; \theta_1)\right) \ . \end{aligned}$$

Here, the first function f_C describes the latent process of the control condition for data D_C, and the second function f_I the process of the intervention condition for data D_I. The model marginal likelihood terms are computed in the same way as for the continuous model. Once again, the shared set of hyperparameters θ_1 is optimized using the (combined) marginal likelihood, which is approximated via BIC:

$$\log p(D \mid \mathcal{M}_1) = \log p(\boldsymbol{y}_C \mid \boldsymbol{x}_C, \hat{\theta}_1, \mathcal{M}_1) + \log p(\boldsymbol{y}_I \mid \boldsymbol{x}_I, \hat{\theta}_1, \mathcal{M}_1) - \frac{k}{2} \log n \ .$$

Note that the discontinuous model and the continuous model have identical sets of hyperparameters, although the optimization can lead to different values.

3.2.1 Estimating the effect size

In addition to their flexibility, a further attractive property of GPs is that the predictive posterior of f is available in closed-form. That is, its posterior mean and variance are given by

$$\begin{aligned} f_* &= \boldsymbol{k}_*^T \left(K + \sigma_n^2 I\right)^{-1} \boldsymbol{y} \\ \mathbb{V}[f_*] &= k(x_*, x_*) - \boldsymbol{k}_*^T \left(K + \sigma_n^2 I\right)^{-1} \boldsymbol{k}_* \ , \end{aligned} \tag{4}$$

where $\boldsymbol{k}_*$ is the vector of covariances between the test point x_* and the n training points $(x_i)_{i=1}^n$. This allows us to compute the posterior of the effect size d , which is the difference in the predictions of the two latent functions in $\mathcal{M}_1$, that is $d = f_I(x_0) - f_C(x_0)$. Using standard properties of Gaussian distributions, this implies the following distribution of the effect size:

$$p(d \mid D, \mathcal{M}_1) = \mathcal{N}\left(f_I(x_0) - f_C(x_0), \mathbb{V}[f_I(x_0)] + \mathbb{V}[f_C(x_0)]\right) \tag{5}$$

As a result, the BMA in (2) is a mixture distribution that combines a spike at $d = 0$ with a Gaussian distribution centered around $f_I(x_0) - f_C(x_0)$. Compared to the overconfident estimation of d conditioned only on $\mathcal{M}_1$, this has a regularizing effect [29], shrinking small effect size estimates towards zero.

3.3 Covariance functions

Depending on what covariance function we specify, widely different behaviour for f, f_C and f_I can be achieved. For instance, we may specify that we expect functions to be smooth in x, or that functions exhibit periodicity. The properties of these covariance functions, like their length scale or frequency, are captured by the hyperparameters θ.

3.3.1 Model averaging across GP covariance functions

The pragmatic researcher may be agnostic about which covariance function best describes their observations. In this case, multiple covariance functions may be considered, and a final Bayes factor may be obtained by Bayesian model averaging across all covariance functions. Consider the following example in which we entertain both a rigid linear covariance function and a flexible squared-exponential (SE) covariance function, so that $\mathcal{K} = \{k_{\text{LIN}}, k_{\text{SE}}\}$. We can compute the Bayes factor BF_{DC} for each covariance function, but we can also obtain the total Bayes factor given the set of covariance functions:

$$\text{BF}_{10}^{\text{total}} = \frac{p(D \mid \mathcal{M}_1)}{p(D \mid \mathcal{M}_0)} = \frac{\sum_{k \in \mathcal{K}} p(D \mid k) p(k \mid \mathcal{M}_1)}{\sum_{k \in \mathcal{K}} p(D \mid k) p(k \mid \mathcal{M}_0)} \ . \tag{6}$$

The quantity $\text{BF}_{10}^{\text{total}}$ serves as a final decision metric to determine an effect in a quasi-experimental design, while a detailed report is provided by inspecting the Bayes factors corresponding to each considered covariance function. In practice, the evidence of one covariance function can dominate all others, i.e. $p(D \mid k_i) \gg p(D \mid k_j), \forall j \neq i$, in which case the BMA procedure is effectively equivalent to performing the analysis with the best covariance function only.

3.3.2 Theoretical motivation of covariance function choices

The choice of the Gaussian process covariance functions plays two conceptually distinct roles in BNQD. First, our choice of covariance function reflects our beliefs about the latent process that generated the observations. In traditional RD designs, one assumes a parametrized model such as linear regression. In BNQD, this explicit parametric form is replaced by a GP prior that assigns a probability distribution to the space of functions. However, BNQD can still replicate parametrized models by selecting degenerate covariance functions, such as a polynomial covariance function. These modeling choices are crucial in RD design as model misspecification can lead to incorrect inference. In the case that we do not have clear prior beliefs about an appropriate covariance function, we can look at the weighted average of several covariance functions via the Bayesian model average.

The second role of the covariance function choice is that it determines to which types of discontinuities BNQD is sensitive. Importantly, different covariance functions can be used to test fundamentally different hypotheses, as the covariance function determines which features of the latent function are part of the alleged effect. For example, the simplest (degenerate) covariance function, the constant covariance function, is sensitive only to differences in the means of the two groups, while the linear covariance function is sensitive to both the difference in mean as well as the difference in slope. Similarly in the nonparametric approach, the Matérn covariance function with parameter $\nu = p + 1/2$ can detect discontinuities in up to the pth derivative. There are two interesting special cases: one is the exponential covariance function which corresponds to Matérn with $p = 1/2$, which detects only discontinuities in the function itself (and not in its derivatives). This is the nonparametric counterpart of traditional linear

regression discontinuity. On the other end of the spectrum is the squared-exponential covariance function which corresponds to a Matérn kernel with $\nu = \infty$. This allows it to detect discontinuities of any order, although the amount of data requires to detect these subtle effects can become prohibitively high.

3.3.3 Bandwidth selection and length scales

In order to infer causality from QED, one assumes that the alleged change occurs at the threshold, but that the latent process is otherwise stationary. Consequently, the behaviour of the two groups changes sharply around the intervention instead of gradually. In standard RD studies, this locality is controlled via a bandwidth parameter that determines the sensitivity of the detection approach [30], but this requires the availability of sufficient data around the threshold, and conclusions depend upon this bandwidth parameter. In BNQD, when using stationary nonparametric covariance functions, the bandwidth is replaced by the length scale hyperparameter, which we optimize using the model marginal likelihood. The length scale regulates how fast the correlations between consecutive points decays with their distance, and thus how sensitive BNQD is around the threshold [31]. For example, if the estimated length scale is very short, the continuous model will be able to capture a discontinuity almost as well as the discontinuous model. Conversely, for a large length scale the discontinuity cannot be captured by the continuous model at all, resulting in a poor fit to the data and hence a Bayes factor in favor of the discontinuous model.

3.4 Model training

The BNQD analysis is implemented in Python using GPflow 2.0. Training consists of first optimizing the hyperparameters using the closed-form marginal likelihood, and then computing the posterior mean and variance regressions in both models, using Eq. (4).

We assume a constant mean function $\mu(x) = c$, where $c = \frac{1}{n}\sum_{i=1}^{n} y_i$. For all hyperparameters constrained to be positive we assume vague Gamma$(0.01, 0.01)$ priors, and for unconstrained priors we assume a Gaussian$(0.0, 1.0)$ prior. The same priors are used in all models. Note that these priors have no noticeable effect on the results, as they tend to get completely dominated by the marginal likelihoods. We assume a priori that $p(\mathcal{M}_0) = p(\mathcal{M}_1)$. The BMA distribution is approximated via Monte Carlo, and visualized with kernel density estimation.

All materials are available at Github.

4 Simulations

We evaluate the performance of BNQD using simulations. To allow for a varied range of potential true latent functions, we use the simulations that were discussed by Branson et al. [32], which consist of several different polynomials up to order 5. In addition, we consider a simple linear function, which is the assumption in traditional RD design, and a trigonometric function. The function definitions are provided in the Supplementary Material.

For each latent function f, we generate n observations (x_i, y_i) according to the following procedure:

$$\begin{aligned} x_i &\sim \mathcal{U}(-1, 1) \\ y_i \mid x_i, \sigma, d, f &\sim \mathcal{N}\left(f(x_i) + d[x_i \geq x_0], \sigma^2\right) \quad , \end{aligned}$$

where the threshold $x_0 = 0$. We fix $\sigma = 1.0$ and vary $d \in \{2^{-2}, \ldots, 2^2\}$. Next, we subject the simulated data to analysis by BNQD, using a first-order polynomial, an exponential, a Matérn ($\nu = 3/2$) and a squared-exponential covariance function. For each covariance function, we compute the Bayes factor for the presence of a discontinuity, and we estimate the effect size given the discontinuous model, $p(d \mid D, \mathcal{M}_1)$ (see Eq. (5)), as well as via the Bayesian model average, $p(d \mid D)$ (see Eq. (2)). In addition, we compute the total Bayes factor that averages across covariance functions, as in Eq. (6). One example simulation run is shown in Fig. 1. Here, the different functions are shown, as well as the regressions by both the continuous and discontinuous models, for each of the four considered covariance functions. The vertical bars in the figure show the expectation of the estimated effect size $p(d \mid D, \mathcal{M}_1)$.

Figure 1 shows that indeed the discontinuous model overestimates the effect size when the discontinuity is small compared to the noise level. This effect is a consequence from the conditioning on $\mathcal{M}_1$, which

Figure 1: Example of one simulation run for effect sizes $d \in \{0.25, 1.0, 4.0\}$ and $\sigma = 1.0$. The covariance functions used here are linear, exponential, Matérn ($\nu = 3/2$) and squared exponential. The vertical bars indicate the true effect size (black) and the estimated effect sizes by the discontinuous models for the different covariance functions. As the figure shows, the linear covariance function tends to have the strongest bias, in particular in the low signal-to-noise regime.

Figure 2: The recovery of the true effect size via $p(d \mid D, \mathcal{M}_1)$ as well as via the Bayesian model average estimate $p(d \mid D)$, as a function of increasing signal-to-noise ratios. The covariance functions used here are linear, exponential, Matérn ($\nu = 3/2$) and squared exponential. **A**. The expectation of the posterior effect size distributions. **B**. The RMSE between the posterior distribution and the true effect size d, approximated via Monte Carlo. Note how the RMSE is reduced by the BMA approach for low signal-to-noise regimes. In both panels, plots show the mean and one standard error over 100 simulations.

most existing RD analyses implicitly do. The polynomial covariance function, which is the Gaussian process equivalent to linear regression, results in the greatest bias of the effect size estimate, due to its poor fit to the data.

Figures 2A quantifies these observations by showing the expectation of the posterior effect size distributions (given $\mathcal{M}_1$; solid lines, or given the BMA; dashed lines). Due to the shrinkage, the expectation of these distributions is slightly lower for the model-averaged approach. The effect of BMA becomes more apparent in the RMSE as shown in Fig. 2B. In contrast to the expectation, the RMSE captures the performance of the entire posterior distribution of the effect size, rather than its expectation. As a result, we see that for low noise levels the BMA is clearly advantageous, as it has substantially lower RMSE. Once $d = 2\sigma^2$, the discontinuity is easy to detect, so that $\mathcal{M}_1$ dominates the BMA and the two approaches converge.

The detection of an effect is quantified using the (log) Bayes factor, as shown for the same simulations in Fig. 3. Here, we see somewhat different behaviour depending on the true latent function f. For Quad, Cubic, CATE1, CATE2, Curvature and Sine, the log Bayes factor is negative for low signal-to-noise ratios, correctly indicating that there is evidence for an absence of an effect. For the Linear function, the log Bayes factor is approximately 0 in this regime, which indicates that the data has no preference for either model. Interesting special cases are the Lee and Ludwig functions. Both of these function feature a discontinuity in their derivative [32], which is correctly picked up by the Bayes factor, even when the magnitude of the effect size is small. This confirms BNQD's ability to detect discontinuities of higher orders.

5 Applications

5.1 Sicilian smoking ban effect on acute coronary events

On January 10th 2005, the Italian government installed a national smoking ban in public places. One of the main reasons for this decision is the mounting evidence for the adverse effects of secondhand smoking [33]. Several studies indeed report a reduction in rates of hospital admissions for acute coronary events (ACEs) after similar bans have been instantiated [34, 35], but the benefits of this policy are not undisputed. Widely different effect sizes have been reported, ranging from large to practically absent [36]. The effect of the smoking ban in Sicily was investigated by Barone-Adesi et al. [36] who collected data on the number of ACEs in the four years prior to the ban as well as up to two years after the ban using interrupted time series design [36, 37]. Here, we apply BNQD to the same data for comparison. The goal of this example is not to provide definitive evidence in favor or against the effects of a smoking ban,

Figure 3: The log Bayes factor of the detection of a discontinuity. Note how the linearity assumption, captured by the linear covariance function, leads to extremely overconfident results. The nonparametric covariance functions do detect the discontinuity, but with much more moderate confidence. The total Bayes factor that averages over all considered covariance functions (here linear, exponential, Matérn and squared exponential) seems to follow the Matérn covariance function, and is hardly influenced by the linear covariance function, as that tends to fit the data poorly.

Table 1: The analysis of the Sicilean smoking ban effects on the number of acute coronary events. Shown are the log Bayes factor, the expectation of the effect size under $\mathcal{M}_1$ and the Bayesian model average effect size.

Covariance function	**log BF**$_{10}$	$\mathbb{E}[p(d \mid D, \mathcal{M}_1)]$	$\mathbb{E}[p(d \mid D)]$
Linear	-3.7549	-14.1025	-0.3225
Matérn ($\nu = 3/2$)	-3.6435	-8.1277	-0.2072
BMA	-3.6436	-8.1334	-0.2073

but rather to illustrate how BNQD is applicable in similar situations.

The data are shown in Fig. 4 with y_i the age-standardized number of ACEs per 100 000 citizens at timepoint x_i. The data set shows strong seasonal effects, for which both aforementioned studies corrected a priori. In the GP framework, such adjustment is unnecessary when using nonparametric covariance functions. Here, we perform the model comparison using linear and Matérn ($\nu = 3/2$) covariance functions. For both covariance functions, the corresponding hyperparameters are optimized using maximum likelihood. The results of the Bayesian non-parametric analysis are shown in Fig. 4. For both covariance functions, we find evidence *against* an effect; the data are approximately 40 times more likely under $\mathcal{M}_0$. A summary of the analysis is shown in Table 1. The example demonstrates the importance of model averaging: If we condition on $\mathcal{M}_1$, we find a reduction in population-standardized acute coronary events of 14.1 (linear) and 8.1 (Matérn), but when taking the high probability of the null model into account, this effect drastically reduces to 0.3 and 0.2, respectively.

5.2 Phantom border effect on Dutch government elections

In 2017, the Dutch parliament elections were held, and soon after the results were published online by the Dutch government [38]. According to Dutch electorate geographer De Voogd, the share of votes that go to populist parties is different between the Dutch areas north and south of a so-called 'phantom border', a

Figure 4: Discontinuity analysis of the Sicilean smoking ban instantiation in January 2005. The left column shows the fitted regressions using either the linear (top) or the Matérn ($\nu = 3/2$) covariance function (bottom). Shown are both the continuous and the discontinuous models. The lines and shaded intervals represent the mean and one standard deviation of the posterior Gaussian process. The right column shows the distributions of the estimated effect size. The dashed line corresponds to $d = 0$, as dictated by the null model, while the purple shaded distribution represents $p(d \mid D, \mathcal{M}_1)$. The dark shaded distribution is a kernel density estimate of the Bayesian model average mixture distribution, which in both cases is dominated by the null model. The inset pie-charts indicate the posterior model probabilities $p(\mathcal{M}_1 \mid D)$ (dark) and $p(\mathcal{M}_0) \mid D)$ (purple).

line that traditionally divided the catholic south of the Netherlands from the protestant north, and which originated in the Eighty Years' War (1568–1648) [39, 40] (As this example merely serves to illustrate a two-dimensional quasi-experimental design, we refrain from an extensive discussion of the definition of populism. Instead, we use part of the definition by Müller [41] and refer to populist parties as those parties that emphasize an alleged chasm between the elite and the general population. In the Dutch 2017 elections, parties that fit this description were PVV, SP, 50Plus and FvD.). This border serves as a two-dimensional threshold along which one can apply RD design. This special case of RD design where the assignment threshold is a geographical boundary is sometimes referred to as GeoRDD [10]. Here, we test the hypothesis by De Voogd and demonstrate that BNQD is straightforward to apply a GeoRDD context.

We apply BNQD using the linear, and first-order Matérn covariance functions. The results of the analysis are shown in Fig. 5. The first figure shows the map of the Netherlands with the fraction of populist votes per municipality superimposed. In addition, each map shows a hand-drawn approximate phantom border representing the supposed divide in voting behaviour. The top row of Fig. 5A shows the results of a linear covariance function, which in two dimensions fits a plane (or two planes for the discontinuous model) to the data. The bottom row shows the results using the Matérn ($\nu = 3/2$) covariance function. Figure 5B shows the effect sizes conditioned on the discontinuous model, which are distributions at each point of the phantom border (the shaded interval indicates one standard deviation around the mean).

If we assume a linear underlying process, there is strong evidence for a discontinuity ($\log \mathrm{BF} = 23.3$), confirming the hypothesis by De Voogd. Visually however, the data do not appear to follow these linear trends strongly. The nonparametric Matérn covariance function results in evidence against an effect ($\log \mathrm{BF} = -3.7$). As the Matérn covariance function fits the data much more accurately than the linear covariance function, the Bayesian model average is completely dominated by the former. Note that Fig. 5B shows the effect size conditioned on the discontinuous model, but for the Matérn covariance function the continuous model is actually favoured. The BMA of the effect size is therefore approximately zero across the phantom border.

Figure 5: Discontinuity analysis along a two-dimensional boundary (indicated by the dashed line). **A**. Circles indicate the observed fraction of populist votes; municipalities are shaded according to the Gaussian process predictions. **B**. The effect size $p(d \mid D, \mathcal{M}_1)$ along the phantom border.

6 Discussion

While quasi-experimental designs such as regression discontinuity design, have been around since the 1960s [42], recently there has been a renewed interest in this class of methods [9, 43, 44], in particular in epidemiology [45] and education [30]. At the same time, researchers from different domains are actively

promoting the use of QED [11, 12]. This has prompted researchers to extend classical QED in several ways. For instance, several authors have proposed to use Bayesian models for designs such as RD [19], DiD [46], ITS [47], single-subject designs [48] and stepped-wedge trials [49]. In these studies, parameters such as (linear) regression coefficients are associated with prior probability distributions, and posterior distributions are estimated using Bayesian inference. Compared to classical implementations of such study designs, these approaches provide explicit descriptions of the uncertainty that is associated with their parameter estimates. In contrast to our work however, these methods focus on the estimation of the treatment effect, instead of model comparison, and typically assume particular parametric forms.

Other studies have explored non-parametric models for QED, in particular for RD, to alleviate the reliance on a linear model. For example, one can use locally linear non-parametric regression as an alternative to linear regression [50]. Alternatively, one can use kernel methods that compute a smoothly weighted average of the data points to create an interpolated regression that does not depend on a specific parametric form [30] (for a review on the similarities between frequentist kernel-based methods and Gaussian processes we refer the interested reader to Kanagawa et al. [51]). However, for these approaches as well as for BNQD, we assume that the identified discontinuity indeed occurred at the threshold, and that the underlying processes are stationary. Here, the presence of change points away from the intervention threshold can lead to false alarms. If this cannot be verified, or one is interesting in using non-local covariance functions such as a periodic covariance, BNQD should be performed in a sliding-window fashion in order to ensure that the highest Bayes factor is actually at the intervention threshold.

Most similar to BNQD are the Gaussian process regression discontinuity models by Rischard et al. [10], Branson et al. [32]. In both cases, the authors fit a Gaussian process to the intervention and the control group, and compare the difference in their prediction at the intervention point. The primary difference with our work is that these studies focus on inference of the treatment effect $f_I(x_0) - f_C(x_0)$, where f_I and f_C correspond to the intervened and the controlled processes. Instead, we remain within the Bayesian paradigm, which enables us to quantify evidence in favor of the null model [13], and to use Bayesian model averaging [14, 15]. This reduces the overconfidence that is present in many existing methods that condition on $\mathcal{M}_1$. Furthermore, we can integrate over different covariance functions c.q. assumptions on the regression. Because of our focus on model comparison, our approach is able to detect discontinuities in derivatives of the latent function. Furthermore, BNQD extends naturally to the setting of multiple assignment variables[16–18, 43, 52]. A special case of such multivariate QED is GeoRDD, in which the two-dimensional assignment variable represents a spatial location [10, 53, 54]. Our approach does not assume a univariate threshold to determine the assignment to intervention and control group, and can work on arbitrary complex label functions. This can be a geographical border as in our second application, but also more complex shapes such as, for example, one region versus the rest of a country, or a particular regime of diagnostic variables.

BNQD can be extended in several ways to make it more widely applicable. For instance, we do not currently account for covariates that may serve as confounds for causal inference [45]. However, such covariates can be explicitly taken into account in the regression models, or even be learned from the observations [55]. Covariate selection can be performed using automatic relevance determination [56], where we learn separate length scales for each covariate. Those features that do not affect the prediction correspond to diverging length scales.

Other extensions include the fuzzy application of a threshold function (for instance, a drug may be prescribed according to a certain policy, but that policy is not always adhered to). Another extension is the modelling of partial effects [43, 52]. In our multivariate example, a single assignment function is used to label each 2D point as above or below the phantom border. Alternatively, a threshold or other decision function can be applied to either of these two dimensions, which can then be used to test whether each dimension individually results in a discontinuity, in addition to the multivariate test. Finally, more sophisticated covariance functions allows one to test discontinuities in global properties of the signal. For example, periodic covariance functions can detect discontinuities in oscillatory signals such as heart rate and respiration, by identifying differences in frequency, the amplitude or waveform. We explore these extensions in future work.

Our approach is based on Gaussian process regression, which has a number of computationally attractive features. In particular, with a Gaussian observation model, many of the required computations become analytically tractable. However, via for instance the Laplace approximation, Gaussian process regression can be used in tandem with non-Gaussian observation models [28]. For instance, one obser-

vation model could be Poisson to model count data [57, 58], which is arguably more applicable in the Sicilian smoking ban example, or Bernoulli for binary outcomes.

7 Conclusion

We have presented BNQD, a Bayesian non-parametric approach for causal inference in quasi-experimental designs. By selecting the appropriate covariance function, one has precise control over the type of discontinuity that can be detected. For example, using BNQD with an exponential covariance function results in a non-parametric quasi-experimental setup that is only sensitive to discontinuities in the latent process itself, while a squared-exponential covariance function can detect discontinuities in any derivative of the latent process, given sufficient observations.

References

[1] A. B. Hill. The clinical trial. *N Engl J Med*, 247:113–119, 1952.

[2] J. Pearl and D. Mackenzie. *The book of Why: The new science of cause and effect.* Basic Books, 2018.

[3] Donald T. Campbell and Julian C. Stanley. *Experimental and quasi-experimental designs for research.* Rand McNally College Publishing, 1963.

[4] Steffen L Lauritzen. Causal inference from graphical models. *Complex stochastic systems*, pages 63–107, 2001.

[5] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009.

[6] Michael Lechner et al. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, 4(3):165–224, 2011.

[7] David McDowall, Richard McCleary, Errol Meidinger, and Richard Hay. *Interrupted time series analysis.* Thousand Oaks, CA, USA, 1980.

[8] David S. Lee and Thomas Lemieux. Regression discontinuity designs in econometrics. *Journal of Economic Literature*, 48(June):281–355, 2010.

[9] Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

[10] M. Rischard, Z. Branson, L. Miratrix, and L. Bornn. A Bayesian nonparametric approach to geographic regression discontinuity designs: Do school districts affect NYC house prices? *arXiv preprint arXiv:1807.04516*, 2018.

[11] Ioana E Marinescu, Patrick N Lawlor, and Konrad P Kording. Quasi-experimental causality in neuroscience and behavioural research. *Nature Human Behaviour*, pages 1–11, 2018.

[12] Ellen Moscoe, Jacob Bor, and Till Bärnighausen. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice. *Journal of Clinical Epidemiology*, 68(2):132–143, 2015.

[13] Eric-Jan Wagenmakers. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5):779–804, 2007.

[14] Jennifer A. Hoeting, David Madigan, Adrian E. Raftery, and Chris T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–401, 1999.

[15] Max Hinne, Quentin F. Gronau, Don van den Bergh, and Eric-Jan Wagenmakers. A conceptual introduction to bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, 3(2):200–215, 2020.

[16] Sean F. Reardon and Joseph P. Robinson. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, 5(1):83–104, 2012.

[17] John P. Papay, John B. Willett, and Richard J. Murnane. Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, 161(2):203–207, 2011.

[18] V. C. Wong, P. M. Steiner, and T. D. Cook. Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, 38(2):107–141, 2013.

[19] Sara Geneletti, Aidan G. O'Keeffe, Linda D Sharples, Sylvia Richardson, and Gianluca Baio. Bayesian regression discontinuity designs: Incorporating clinical knowledge in the causal analysis of primary care data. *Statistics in Medicine*, 34:2334–2352, 2015.

[20] Donald B Rubin. Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[21] Y. Kim and P. Steiner. Quasi-experimental designs for causal inference. *Educational Psychologist*, 51(3-4):395–405, 2016.

[22] Judea Pearl. *Causality: models, reasoning and inference.* Cambridge University Press, New York, NY, USA, 2nd edition, 2009.

[23] Gguido Imbens and Karthik Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79(3):933–959, 2012.

[24] Steven N. Goodman. Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, 130(12):1005–1013, 06 1999.

[25] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[26] Andrew F Jarosz and Jennifer Wiley. What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving*, 7:2–9, 2014.

[27] Steven N. Goodman. Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of Internal Medicine*, 130(12):995–1004, 06 1999.

[28] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning.* The MIT Press, 2005. ISBN 026218253X.

[29] Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 9(1):247–274, 2015.

[30] Howard S. Bloom. Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, 5(1):43–82, 2012.

[31] David Duvenaud. *Automatic model construction with Gaussian processes.* PhD thesis, Computational and biological learning laboratory, University of Cambridge, 2014.

[32] Zach Branson, Maxime Rischard, Luke Bornn, and W. Miratrix, Luke. A nonparametric Bayesian methodology for regression discontinuity designs. *Journal of Statistical Planning and Inference*, 202: 14–30, 2019.

[33] Lorenzo Richiardi, Loredana Vizzini, Franco Merletti, and Francesco Barone-Adesi. Cardiovascular benefits of smoking regulations: The effect of decreased exposure to passive smoking. *Preventive Medicine*, 48(2):167–172, 2009.

[34] David G Meyers, John S Neuberger, and Jianghua He. Cardiovascular effect of bans on smoking in public places: a systematic review and meta-analysis. *Journal of the American College of Cardiology*, 54(14):1249–1255, 2009.

[35] Antonio Gasparrini, Giuseppe Gorini, and Alessandro Barchielli. On the relationship between smoking bans and incidence of acute myocardial infarction. *European journal of epidemiology*, 24(10): 597–602, 2009.

[36] Francesco Barone-Adesi, Antonio Gasparrini, Loredana Vizzini, Franco Merletti, and Lorenzo Richiardi. Effects of Italian smoking regulation on rates of hospital admission for acute coronary events: A country-wide study. *PLoS ONE*, 6(3):1–6, 2011.

[37] James Lopez Bernal, Steven Cummins, and Antonio Gasparrini. Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, 46(1):348–355, 2017.

[38] Kennis- en Exploitatiecentrum Officiële Overheidspublicaties (KOOP). Verkiezingsuitslag Tweede Kamer 2017. `https://data.overheid.nl/data/dataset/verkiezingsuitslag-tweede-kamer-2017/`, 2018.

[39] Josse De Voogd. Van Volendam tot Vinkeveen: de electorale geografie van de PVV, December 22nd 2016. `http://www.socialevraagstukken.nl/volendam-tot-vinkeveen-de-electorale-geografie-van-de-pvv/` (visited: 21-09-2019).

[40] Josse De Voogd. Deze eeuwenoude grenzen kleuren de verkiezingen nog altijd, March 1st 2017. `https://decorrespondent.nl/6298/deze-eeuwenoude-grenzen-kleuren-de-verkiezingen-nog-altijd/` (visited: 21-09-2019).

[41] Jan-Werner Müller. *What is populism?* Philadelphia, PA, U.S.A., 2016.

[42] Thomas D. Cook. "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2):636–654, 2008.

[43] Jin-young Choi and Myoung-jae Lee. Regression discontinuity: review with extensions. *Statistical Papers*, 58(4):1217–1246, 2017.

[44] W R Shadish, Thomas D Cook, and D T Campbell. *Experimental and quasi-experimental designs for generalized causal inference*. Cencage Learning, Inc., 2nd edition, 2002.

[45] Anthony D. Harris, Douglas D. Bradham, Mona Baumgarten, Ilene H. Zuckerman, Jeffrey C. Fink, and Eli N. Perencevich. The use and interpretation of quasi-experimental studies in infectious diseases. *Antimicrobial resistance*, 38:1586–1591, 2004.

[46] James Normington, Eric Lock, Caroline Carlin, Kevin Peterson, and Bradley Carlin. A Bayesian difference-in-difference framework for the impact of primary care redesign on diabetes outcomes. *Statistics and Public Policy*, 6(1):55–66, 2019.

[47] A. Freni-Sterrantino, R. E. Ghosh, D. Fecht, M. B. Toledano, P. Elliott, A. L. Hansell, and M. Blangiardo. Bayesian spatial modelling for quasi-experimental designs: An interrupted time series study of the opening of municipal waste incinerators in relation to infant mortality and sex ratio. *Environment International*, 128(November 2018):109–115, 2019.

[48] Rivka M. De Vries and Richard Donald Morey. Bayesian hypothesis testing for single-subject designs. *Psychological Methods*, 18(2):165–185, 2013.

[49] Kristen M. Cunanan, Bradley P. Carlin, and Kevin A. Peterson. A practical Bayesian stepped wedge design for community-based cluster-randomized clinical trials: the British Columbia Telehealth Trial. *Clinical Trials*, 13(6):641–650, 2016.

[50] B Y Jinyong Hahn, Petra Todd, and Wilbert van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.

[51] Motonobu Kanagawa, Philipp Hennig, Dino Sejdinovic, and Bharath K Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences, 2018.

[52] Jin-young Choi and Myoung-jae Lee. Regression discontinuity with multiple running variables allowing partial effects. *Political Analysis*, 26(3):258—274, 2018.

[53] L. J. Keele and R. Titiunik. Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155, 2015.

[54] L. Keele, S. Lorch, M. Passarella, D. Small, and R. Titiunik. *An overview of geographically discontinuous treatment assignments with an application to children's health insurance*, chapter 4, pages 147–194. Emerald Publishing Limited, 2017.

[55] Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NeurIPS) 30*, pages 7018–7028. 2017.

[56] David P. Wipf and Srikantan S. Nagarajan. A new view of automatic relevance determination. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1625–1632. 2008.

[57] Ryan P. Adams, Iain Murray, and David J. C. MacKay. Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In Léon Bottou and Michael Littman, editors, *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pages 9–16, Montreal, June 2009. Omnipress.

[58] C. K. I. Williams and D. Barber. Bayesian classification with Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1342–1351, 1998.

A Supporting information

A.1 Covariance functions

Throughout this paper we have made use of the following covariance functions:

$$k(x,x') = (\sigma_v^2 x x' + \gamma)^k \qquad \text{Polynomial}$$

$$k(x,x') = \sigma_v^2 \exp\left(-\frac{x-x'}{l}\right) \qquad \text{Exponential}$$

$$k(x,x') = \sigma_v^2 \left(1 + \frac{\sqrt{3}(x-x')}{l}\right) \exp\left(-\frac{\sqrt{3}(x-x')}{l}\right) \qquad \text{Matérn } (\nu = 3/2)$$

$$k(x,x') = \sigma_v^2 \exp\left(-\frac{(x-x')^2}{l}\right) \ , \qquad \text{Squared-exponential}$$

where σ_v^2 is the output variance of the Gaussian process, k and γ are the degree and offset of the polynomial, and l is the length scale parameter.

A.2 Simulation polynomials

The following function definitions are used in the simulations:

$$f(x) = 0.23 + 0.89x \qquad \text{Linear}$$

$$f(x) = \begin{cases} 3x^2 & \text{if } x < x_0, \\ 4x^2 & \text{otherwise.} \end{cases} \qquad \text{Quad}$$

$$f(x) = \begin{cases} 3x^3 & \text{if } x < x_0, \\ 4x^3 & \text{otherwise.} \end{cases} \qquad \text{Cubic}$$

$$f(x) = \begin{cases} 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < x_0, \\ 0.48 + 0.84x - 3.0x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 & \text{otherwise.} \end{cases} \qquad \text{Lee}$$

$$f(x) = 0.42 + 0.84x - 3.0x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 \qquad \text{CATE1}$$

$$f(x) = 0.42 + 0.84x + 7.99x^3 - 9.01x^4 + 3.56x^5 \qquad \text{CATE2}$$

$$f(x) = \begin{cases} 3.71 + 2.3x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5 & \text{if } x < x_0, \\ 3.71 + 18.49x - 54.81x^2 + 74.3x^3 - 45.02x^4 + 9.83x^5 & \text{otherwise.} \end{cases} \qquad \text{Ludwig}$$

$$f(x) = \begin{cases} 0.48 + 1.27x - 3.44x^2 + 14.147x^3 + 23.694x^4 + 10.995x^5 & \text{if } x < x_0, \\ 0.48 + 0.84x - 0.3x^2 - 2.397x^3 - 0,901x^4 + 3.56x^5 & \text{otherwise.} \end{cases} \qquad \text{Curvature}$$

$$f(x) = \sin x \qquad \text{Sine}$$