# Statistical Inference for Traffic Safety Analysis Using the Generalized Pareto Distribution

**Filip Vestin and Sara Emilsson**

*Supervisor:* Nader Tajvidi

Masters Thesis


LUNDS UNIVERSITET

Department of Mathematical Statistics

Lund University

Sweden

August, 2020

## Acknowledgements

We would like to offer our sincere thanks to our supervisor Nader Tajvidi for his valuable insights and suggestions. His help and contributions to making this thesis has been most appreciated.

We also offer our gratitude to Aliaksei Laureshyn and Carl Johnson for providing us with the data sets used in this thesis, and for the patience and many mail conversations that ensued.

## Abstract

This thesis investigates the possibility of computing interval estimates for metrics that pertain to traffic safety based on surrogate measures of safety. A probabilistic model of the (near) crash count is defined using the generalized Pareto distribution and three different methods for calculating confidence intervals for the corresponding intensity parameter are proposed. We consider the delta method, the profile likelihood and a modification to the profile likelihood. Using the same methods we construct statistical tests in order to compare the safety of two infrastructure designs controlling for difference in traffic flow. These methods are then applied to three data sets from intersections in Denmark and three intersections in the Netherlands. Our findings show that the profile likelihood method yields satisfactory results in comparison to the other two methods; while also being relatively straightforward to implement.

**Keywords:** *Extreme value theory, traffic safety analysis, surrogate measure of safety, maximum likelihood, delta method, profile likelihood, modified profile likelihood.*

# Nomenclature

### Abbreviations

| | |
|---|---|
| Dk1 | First intersection in Denmark |
| Dk2 | Second intersection in Denmark |
| Dk3 | Third intersection in Denmark |
| GPD | Generalized Pareto Distribution |
| ML | Maximum likelihood |
| MPL | Modified profile likelihood |
| Ne1 | First intersection in the Netherlands |
| Ne2 | Second intersection in the Netherlands |
| Ne3 | Third intersection in the Netherlands |
| PDF | Probability density function |
| PET | Post-encroachment time |
| PL | Profile likelihood |
| PMF | Probability mass function |
| POT | Points over Threshold |
| SMoS | Surrogate Measures of Safety |
| TCT | Traffic Conflict Technique |
| TTC | Time-to-Collision |

### Notations

| | |
|---|---|
| $\tilde{S}$ | Random variable for the negated surrogate measures of safety, where $\tilde{S} = -S$ |
| $S$ | Random variable for surrogate measures of safety |

# Contents

# 1 Introduction

The first part of this section provides an introduction to the importance of traffic safety as well as previous research on the subject. The second part will present the purpose of this thesis and a description of the data that has been used for the analysis.

## 1.1 The Importance of Traffic Safety Analysis

In the European Union, more than 35000 people died in 2009. As part of the White Paper, the EU has adapted "Vision Zero" in an attempt to curb this problem, a vision where the number of road deaths are to be reduced to almost zero by 2050. A sub goal is that the number is halved between the year 2011 and 2020 [1].

As an effect of this vision, a number of countries have seen a reduced road accident fatality rate, however, Sweden and the Netherlands have some of the slowest progress. In the past few years, Sweden has even seen an increase in the road accident fatality rate. During 2018, 324 people died in road traffic accidents (defined as "an accident that occurred in traffic on a road generally used for traffic with motor vehicles, whereas at least one vehicle were in motion and resulted in person injury" [2]) in Sweden, which is an increase of 29% since the year before.

Internationally speaking, however, Sweden has a fairly low fatality rate. In 2017, Sweden had 25 fatalities due to road accidents per million inhabitants, while Denmark had 32 and the Netherlands had 36. By 2018, Sweden had increased that number to 32 [2].

Assessing the traffic safety of roads can be challenging. Not all accidents are reported, especially non-fatal accidents suffer from being under reported. Thus, obtaining data can sometimes be challenging, and not to mention, the data received often contain very little additional information of the accident. As such, even accident types with a high frequency often needs years of observation before any analysis of safety can be done. Consequently, the use for different surrogate measures, that measure how close an interaction of road users are to collision, is unquestionably useful [3]. The practicality of these measures is the added information, such as the type of road user, evasive manoeuvres, and type of intersection at which the near-crashes occured, as well as a higher frequency of data points.

## 1.2 Previous Research of Traffic Safety

The Safety Pyramid, Figure 1, was introduced in 1987 by Hydén [4] to illustrate the relationship between severity and frequency of interactions between road users. At the top of the pyramid are the most severe collisions (the fatal) which have the lowest frequency of all interactions, and at the bottom are the undisturbed passages, that constitute the highest frequency. By analyzing near-crashes, we will be focusing on the top of the pyramid, including potential conflicts.

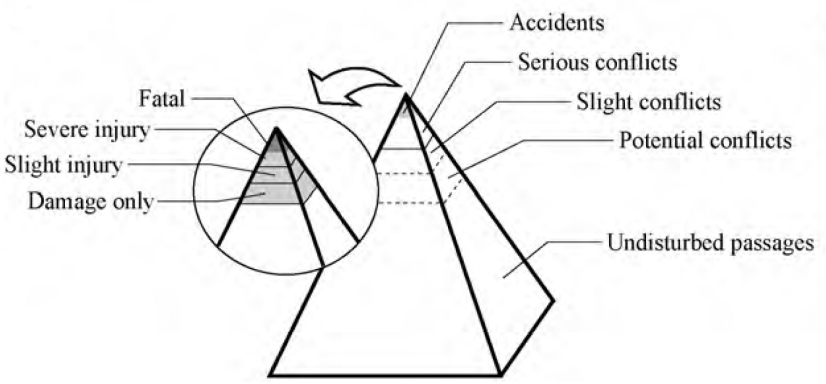


Figure 1: The Safety Pyramid as introduced by Hydén in 1987

Source: Laureshyn et al. (2010) [5]

Traffic Conflict Technique (TCT) is the use of surrogate measures to to assess levels of conflicts. These are related to accidents, and were primarily used to evaluate individual intersections. The first practical application of TCT was suggested by Harris and Perkins in 1967 [6], and since, many surrogate measures of safety (SMoS) have emerged, Time-to-Collision (TTC) being among the more popular which was introduced by Hayward in 1972 [7].

Extreme value theory is used in various fields, but the pioneers to apply it to traffic safety was Campbell, Joksch and Green whom in 1996 fit the Weibull distribution to the extreme values of TTC to "estimate the safety benefits of active safety technology" [8]. In 2006, Songchitruksa and Tarko let the extreme value distributions vary with the traffic flow and predicted crash frequencies which were consistent with the observed crash frequencies when analysing post-encroachment times, another surrogate measure of safety, from video recordings [9].

## 1.3 Surrogate Measures of Safety (SMoS)

As mentioned in Section 1.1, surrogate measures are a useful tool for statistical analysis of traffic safety. By utilising the additional information and data of near-crashes rather than pre-existing crash data, one can work pro-actively as opposed to reactively [3]. However, this approach tend to be somewhat labour intensive.

In this thesis, the following three SMoS will be used in the analysis.

- $\text{TTC}_{min}$ (Time-to-Collision): The minimum time in seconds it takes for the two road users to collide if they continue with their same speed and paths. The road users must be on a collision course for this to be measured [7].
- PET (Post-Encroachment Time) is the time in seconds it takes for the second road user to reach the place of the first road user [5].

- $T2_{min}$ is an extension of PET, measuring the PET for each given moment, given that both road users continue with the same speeds and paths, and taking the minimum value [5]. Also measured in seconds.

## 1.4 Purpose and Problems

As was previously described, there is an established approach within traffic safety analysis to use SMoS and extreme value distributions to assess the safety of a particular piece of road infrastructure. Generally, it is assumed that some extreme value distribution can be used to approximate the lower tail distribution of some SMoS, from which a crash probability can be computed. With some additional assumptions one can then use the crash probability to estimate the intensity at which a crash occurs per unit of time. The two primary objectives of this thesis can summarized as:

- The first goal is to develop and evaluate different methods for computing confidence intervals for the (near) crash intensity at an intersection.
- The second goal is to develop hypothesis tests in order to assess the relative safety between two different intersections while accounting for differences in traffic flows.

A lot of previous research has focused on producing adequate point estimates of the crash intensity. In this thesis, we put more consideration into constructing adequate interval estimates of the crash intensity. Incorporating confidence intervals will allow the researcher to get a sense of the uncertainty of their point estimates. This is not necessarily something novel since most studies from our literature review also compute confidence intervals. However, from our research it appears that the intervals are computed through parametric bootstrapping schemes; see for example Zheng et. al (2014) [10], Tarko (2012) [11], Songchitruksa and Tarko (2006) [9], and Farah and Azevedo (2017) [12]. Our contention with this approach is that it has been shown that parametric bootstrapping methods perform poorly for extreme value distributions in the sense that it produces excessively wide confidence intervals (see e.g. [13]). We propose some alternative methods of computing confidence intervals for the crash intensity based on some common principles within extreme value theory and statistical inference theory.

The second aim is to assess the relative safety between two intersections. Although we may not be able to model the exact underlying process generating the SMoS, we may still be able to make correct decisions as to which intersection is safer. As a consequence, these methods could be used to identify optimal infrastructure designs from a safety perspective, even if we have inaccurate representations of the variability in crash frequencies in the form of a model. This is potentially also an alternative method to validating proposed models of the crash frequency. If we know that a particular infrastructure design in theory is safer than another, and the proposed model identifies this design as safer than some other design, then this is evidence that the model is valid;

at least for assessing the relative safety of two designs. We apply our proposed methods on three data sets from intersections in Denmark and the Netherlands.

For this particular application there is some theoretical evidence that the intersection design in Denmark is generally safer than in the Netherlands. This proposition is mostly based on a literature review by Prati et al (2018) [14], which found evidence that the intersection design in Denmark is safer than the corresponding design in the Netherlands for interactions between motor vehicles and cyclists [14]. The bicycle crossing at the intersections in the Netherlands consist of a separate cyclist track together with pedestrians in which the cyclist lane is separate from the motor vehicle lane. The intersections in Denmark are characterized by having bike paths that are separated from the motor vehicle lane only by painted lines on the road before and after the intersection. So there is some justification for the proposition that the intersections in Denmark are generally safer. To evaluate the relative safety, our approach is to define hypothesis tests for the difference in crash probabilities at the respective intersections using some asymptotic results from inference theory (see Section 3).

## 1.5 Data Description

The data used in this thesis has been collected by the Transport and Roads Division of the Department of Technology and Society at LTH, by filming intersections in several European countries. The videos have been processed by using a combination of the two softwares T-analyst and Road User Behavioural Analysis (RUBA), in addition to using human observers.

In this thesis, we are analysing a subset of this data, consisting of three intersections in Denmark and three in the Netherlands, hereby referred to as Denmark 1 - 3 and Netherlands 1 - 3. We are only focusing on conflicts between right turning cars and straight going bicycles, using the three surrogate measures of safety $\text{TTC}_{min}$, $\text{T}_{2min}$ and PET (see Section 1.3 for definitions of the surrogate measures). In total, three outliers were removed from the data: one observation for $\text{T2}_{min}$ from Denmark 1 with a value of 56.9128 seconds, one observation for $\text{T2}_{min}$ from the Netherlands 1 with a value of 21.1210 seconds , and one from $\text{T2}_{min}$ from the Netherlands 3 with a value of 21.8707 seconds. All analysis below are done after these outliers were removed.

All intersections have been filmed from 00:00 - 23:59 during 2016. The intersections in the Netherlands have all been filmed for a period of 24 hours on the 16/6, while two of the ones in Denmark have been filmed for more than one day. The data for Denmark 1 have been collected from 18/4 - 21/4, Denmark 2 from 19/5, and Denmark 3 from 23/8 - 24/8. The intersections Netherlands 1 - 2 have a higher number of interactions than Denmark 1 and 3, despite having been filmed for a shorter period of time. The total number of interactions for the six intersections between a right turning car and a cyclist can be seen in Table 1, along with the number of observations of each of the three surrogate measures as well as the number of hours for which the data has

been collected. As evidenced, there are few observations for $\text{TTC}_{min}$ for most of the intersections, one being as low as 4 (Denmark 3). The number of observations for $\text{T}_{2min}$ and PET are much higher, with the lowest being 71 and 72 respectively, both belonging to Denmark 1. Due to the low number of observations for $\text{TTC}_{min}$ for Denmark 1 and 3, they will be removed from further analysis. Denmark 2 and the Netherlands 1 - 3 has had a period of 24 hours for the collection of data, Denmark 3 has had 48 hours and Denmark 1 has the longest time of data collection: 96 hours.

Table 1: Number of observations for the different data sets and surrogate measures and number of hours for the collection of data

| Intersection | $N$ | $n_{TTC}$ | $n_{T2}$ | $n_{PET}$ | $h$ |
|---|---|---|---|---|---|
| Denmark 1 | 76 | 10 | 71 | 72 | 96 |
| Denmark 2 | 136 | 39 | 136 | 136 | 24 |
| Denmark 3 | 114 | 4 | 104 | 104 | 48 |
| Netherlands 1 | 269 | 189 | 266 | 265 | 24 |
| Netherlands 2 | 137 | 67 | 136 | 136 | 24 |
| Netherlands 3 | 109 | 41 | 106 | 106 | 24 |

The histograms for all measures for Denmark 2 and the Netherlands 2 can be seen in Figure 2 - 4. The distributions for $\text{TTC}_{min}$ (Figure 2) look like they may be similar for the Denmark 2 and the Netherlands 2, however, the number of observations for Denmark 2 are fairly low (39) compared to the Netherlands 2 (67). The histograms for $\text{T2}_{min}$ for Denmark 2 and the Netherlands 2 (Figure 3) look very similar, with the same number of observations but a shifted mean. The histogram for PET for Denmark 2 have much heaver tails than the histogram for the Netherlands 2 for the same measure, and the distributions do not appear to be the same (Figure 4).

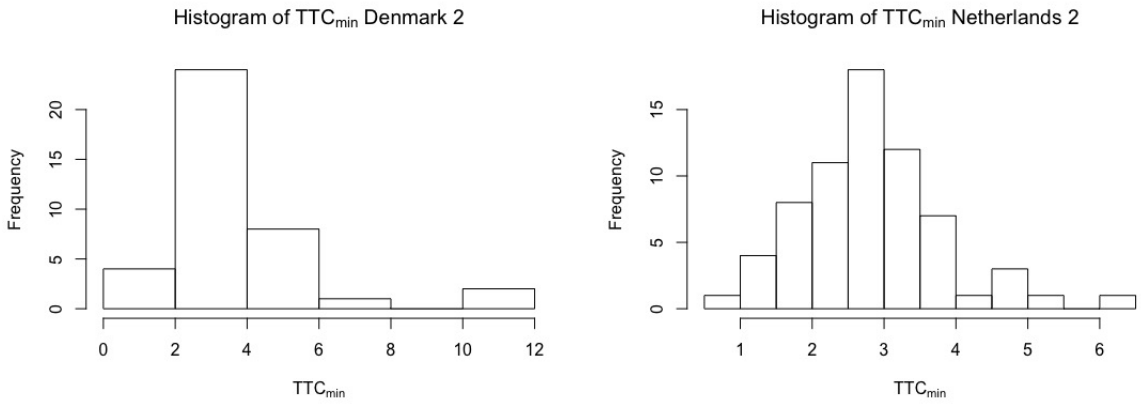


Figure 2: Histogram of $\text{TTC}_{min}$ for Denmark 2 and the Netherlands 2

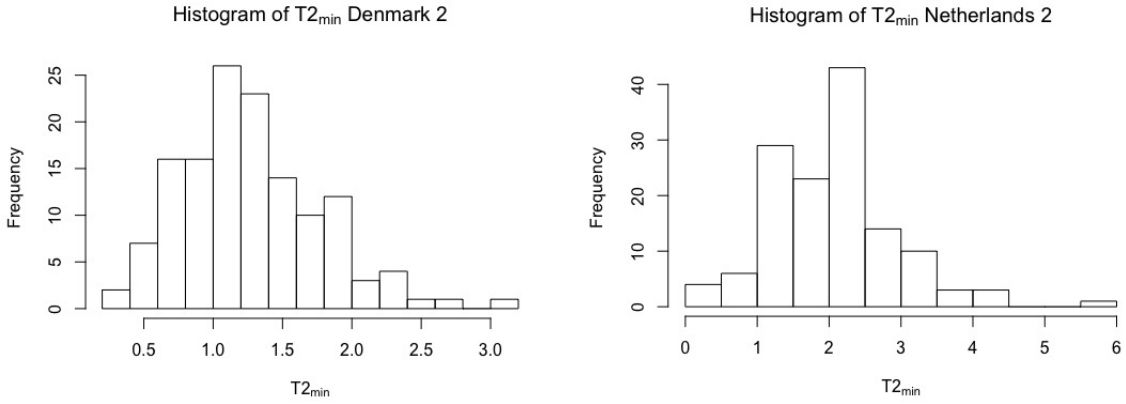


Figure 3: Histogram of $T2_{min}$ for Denmark 2 and the Netherlands 2

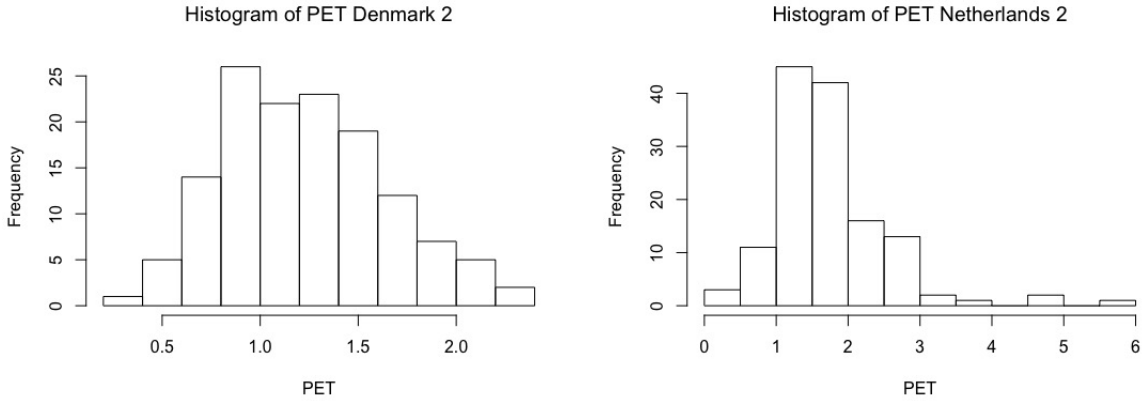


Figure 4: Histogram of $T2_{min}$ and PET for the Netherlands 2

Boxplots of the measures for intersection 2 in Denmark and the Netherlands are seen in Figure 5 - 7. The boxplot for $\text{TTC}_{min}$ look similar for the two countries, both with two outliers (Figure 5). The boxplots for $\text{T2}_{min}$ is much narrower for Denmark than the Netherlands (Figure 6). For PET, the box and whiskers are much narrower for Denmark than the Netherlands, with the later having 5 outliers compared to Denmark, which has none (Figure 7).

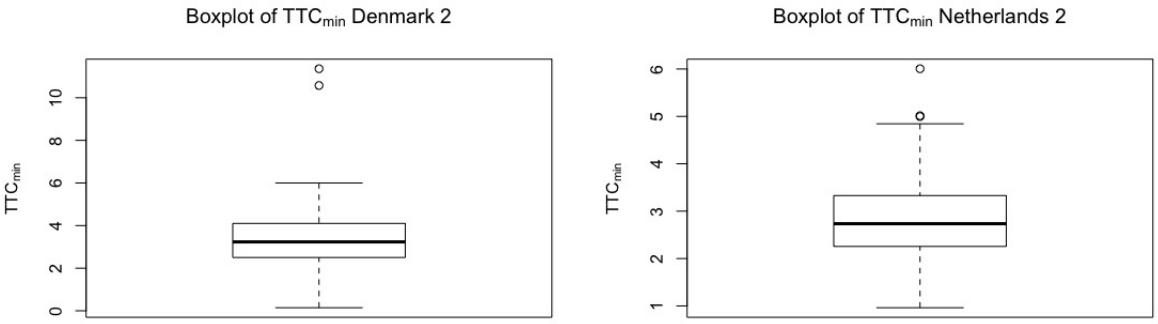


Figure 5: Boxplot of $\text{TTC}_{min}$ for Denmark 2 and the Netherlands 2

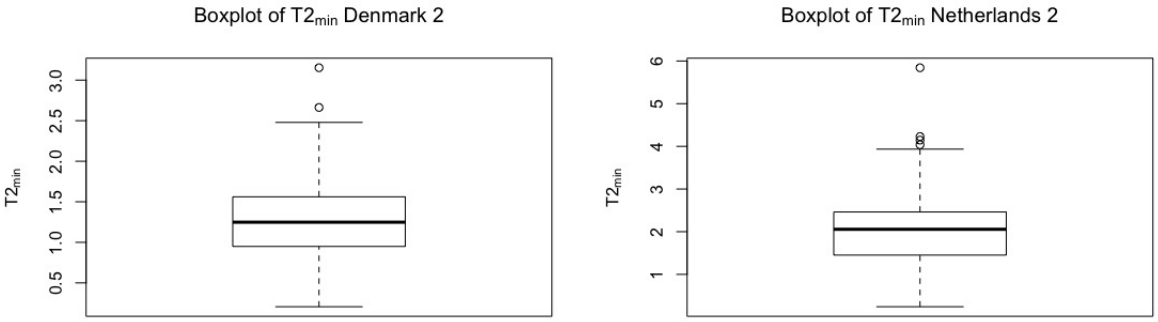


Figure 6: Boxplot of $\text{T2}_{min}$ for Denmark 2 and the Netherlands 2

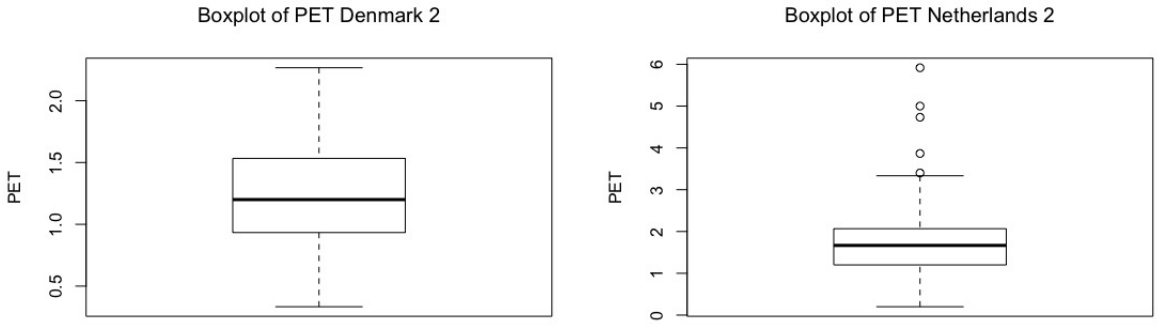


Figure 7: Boxplot of PET for Denmark 2 and the Netherlands 2

Descriptive statistics of the data can be seen in Table 2. The lowest standard error is for Denmark 2, PET, with a value of 0.4146, while the highest is for the Netherlands 3, $T2_{min}$, with a value of 2.1746. The standard error varies somewhat for the same measures across the intersections within one country. For instance, the lowest standard error in Denmark for $T2_{min}$ is 0.5096 and the highest is 1.8069 while the same measure for the Netherlands range from 0.8498 to 2.1746. The closest difference in the highest and lowest standard errors for a measure in one country are PET in Denmark and $TTC_{min}$ in the Netherlands, both with a difference close to 0.4. The standard errors for $TTC_{min}$ for Denmark 1 and 3 are 0.8588 and 1.9409 respectively.

All minimum values but two are below 0.54, with $TTC_{min}$ for the Netherlands 2 and 3 being the exceptions with a minimum of 0.9621 and 2.076 respectively. Mean values are fairly consistent with values ranging from 1.2392 to 3.5759 for all measures and intersections.

The maximum values for all measures and intersections range from 2.2667 for PET for Denmark 2, to 21.8707 for $T2_{min}$ for the Netherlands, while most values range from 2 to 8.

Table 2: Minimum, mean, maximum value and standard errors for the three measures in the different intersections and countries

| Data | Measure | Min | Mean | Max | Standard Error |
|---|---|---|---|---|---|
| Denmark 1 | $T2_{min}$ | 0.0094 | 2.0171 | 10.6758 | 1.8069 |
| | PET | 0.200 | 1.764 | 5.533 | 0.8433 |
| Denmark 2 | $TTC_{min}$ | 0.1437 | 3.5759 | 11.3627 | 2.1453 |
| | $T2_{min}$ | 0.2061 | 1.2751 | 3.1528 | 0.5096 |
| | PET | 0.3333 | 1.2392 | 2.2667 | 0.4146 |
| Denmark 3 | $T2_{min}$ | 0.1231 | 1.5120 | 6.3527 | 0.7342 |
| | PET | 0.5333 | 1.3660 | 2.9333 | 0.4484 |
| Netherlands 1 | $TTC_{min}$ | 0.000 | 3.074 | 7.687 | 1.1096 |
| | $T2_{min}$ | 0.2281 | 2.1927 | 21.1210 | 1.5581 |
| | PET | 0.2667 | 2.0340 | 12.4667 | 1.7660 |
| Netherlands 2 | $TTC_{min}$ | 0.9621 | 2.8289 | 6.0053 | 0.9756 |
| | $T2_{min}$ | 0.2417 | 2.0706 | 5.8423 | 0.8498 |
| | PET | 0.200 | 1.748 | 5.918 | 0.8152 |
| Netherlands 3 | $TTC_{min}$ | 2.076 | 3.149 | 5.364 | 0.7397 |
| | $T2_{min}$ | 0.4842 | 2.5846 | 21.8707 | 2.1746 |
| | PET | 0.400 | 1.874 | 8.267 | 0.9980 |

# 2 Theory

In this section we provide some necessary theory that justifies the method and results found in the course of this thesis.

## 2.1 Probability Distributions

In what follows, we will list some basic probability distributions and stochastic processes that will be used in the course of this thesis. Most of the material covered in this section is taken from Gut (2009) [15]. Probability distributions are usually characterized by their cumulative distribution function defined as

$$F(x) = P(X \leq x_0) = \int_{x<x_0} f(x)dx$$

for continuous distributions, and

$$F(x) = P(X \leq x_0) = \sum_{x \leq x_0} p(x),$$

for discrete distributions. Here, $f(x)$ and $p(x)$ refer to the probability density function (PDF) and the probability mass function (PMF) respectively.

Let $X$ be a Bernoulli variable with discrete outcomes $\{0, 1\}$ and PMF given by

$$p(x) = \begin{cases} \pi, \text{ for } x = 1 \\ 1 - \pi, \text{ for } x = 0. \end{cases}$$

If we have a sequence of identical and independent Bernoulli variables $\{X_1, ..., X_n\}$ and consider the sum $Y = \sum X_i$, then $Y \in Bin(n, \pi)$ with PMF

$$p(y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad \text{for } y = 0, ..., n.$$

A closely related distribution is the Poisson distribution with PMF

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \text{for } x = 0, 1, 2, ... \text{ and } \lambda > 0$$

Finally, let $X \in \exp(\lambda)$, meaning that it follows the exponential distribution with PDF

$$f(x) = \lambda \exp(-\lambda x), \text{ for } x > 0, \ \lambda > 0.$$

Now suppose we observe a sequence of integer-valued random variables $\{X(t), \ t \geq 0\}$ with $X(0) = 0$ and $X(t_2) \geq X(t_1)$ for $t_2 > t_1$. This sequence will constitute a Poisson process if it satisfies the following properties [15]:

1. The increments $X(t_{i+1}) - X(t_i)$ are independent random variables for all $i$.
2. $X(t) - X(s) \in Po(\lambda(t - s))$ for $\lambda > 0$.

The parameter $\lambda$ is usually called the intensity of the Poisson process. Suppose that the times at which an increment in the process occurs during a period T are given by $t_1 < t_2 < ... < t_n$, so that $X(t_n) = n$. These time points are usually referred to as jumping times, and the difference between two contiguous jumping times $t_i - t_{i-1}$ follow an exponential distribution, i.e., $t_i - t_{i-1} \in \exp(\lambda)$ [15].

## 2.2 Extreme Value Theory

Extreme value theory is concerned with the distribution of extremal events. An appropriate starting point is to consider the distribution of sample maxima.

**Definition 2.1.** Suppose we have the random variables $\{X_1,\ ...,\ X_n\}$, then the maxima is defined as $M_n = \max\{X_1,\ ...,\ X_n\}$.

Now, if we let $\{X_1,\ ...,\ X_n\}$ be independent and identically distributed with common distribution function $F(x)$, it follows that $M_n \in F^n(x)$. In many cases $F^n$ will converge to a point mass, but as we are interested in the distribution of the maxima, the question is whether there exists some distribution function $G$ such that $P\left(a_n\left(M_n - b_n\right) \leq x\right) \to G(x)$.

**Theorem 2.1.** A non-degenerate distribution function $G$ is max-stable if and only if there is a sequence $F_n$ and constants $a_n > 0$ and $b_n$ such that

$$F_n\left(a_{nk}^{-1}x + b_{nk}\right) \to G^{\frac{1}{k}}(x)$$

for $k = 1, 2, ...$

For a max-stable distribution, it follows that $G^n\left(a_n x + b_n\right) = G(x)$. Conveniently, it can be shown that every max-stable distribution belongs to three types [16].

**Theorem 2.2.** Suppose $G(x)$ is a max-stable distribution. Then $G$ belongs to one of the following three types

$$\begin{aligned}
&\text{Type I}: \ G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty \\
&\text{Type II}: \ G(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ e^{-x^{-\alpha}}, & \text{otherwise, } \alpha > 0 \end{cases} \\
&\text{Type III}: \ G(x) = \begin{cases} e^{-(-x)^{\alpha}}, & x \leq 0,\ \alpha > 0 \\ 1, & \text{otherwise.} \end{cases}
\end{aligned} \tag{1}$$

Furthermore, the three cases can be captured by a single generalized extreme value distribution given by

$$G(z) = \exp\left(-\left[1 + \gamma\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right), \tag{2}$$

for $\{z : 1 + \gamma\left(\frac{z-\mu}{\sigma}\right) > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \gamma < \infty$ [17].

The implication of Theorem 2.1 - 2.2 is that if we have a sequence of i.i.d random variables $\{X_1,\ ..., X_n\}$ with distribution function $F$ such that $F^n\left(a_n^{-1}\left(x + b_n\right)\right) \to G(x)$, then we obtain the distribution of the maxima $M_n$ as $P\left(M_n \leq z\right) \to G(z)$ where $G(z)$ is given as in Equation 2.

We may proceed by modelling $M_n$ immediately using the theorems above, however this approach is wasteful in terms of data. Instead, we may consider the distribution of exceedences above some threshold $u$ i.e.

$$P\left(X - u > y | X > u\right) = \frac{1 - F\left(u + y\right)}{1 - F(u)},\ y > 0. \tag{3}$$

The following theorem provides a way to approximate the distribution of exceedences.

**Theorem 2.3.** Suppose that $M_n = \max(X_1, ..., X_n)$ and $P(M_n \leq z) \to G(z)$ as in Equation 2, then the distribution of $(X - u)$ for sufficiently large $u$ is given by

$$P(X - u \leq y | X > u) = 1 - \left(1 + \gamma \frac{y}{\sigma}\right)^{-\frac{1}{\gamma}}$$

for $\{y : y > 0,\ \left(1 + \gamma \frac{y}{\sigma} > 0\right)\},\ \sigma > 0,\ \infty < \gamma < -\infty$ [17].

The distribution function in Theorem 2.3 is called the Generalized Pareto Distribution (GPD), and we see that the approximation to the GPD is contingent on the convergence of maxima to a non-degenerate distribution. For the case when $\gamma = 0$, we take the limit which is the exponential distribution given by

$$P(X - u \leq y | X > u) = 1 - \exp\left(\frac{y}{\sigma}\right),\ \text{for } y > 0. \tag{4}$$

We can use the theorem to obtain an estimate of the upper tail of the distribution of $F$ simply by applying Bayes' theorem i.e.

$$P(X \leq x) = P(X > u)\left[1 - \left(1 + \gamma \frac{x - u}{\sigma}\right)^{-\frac{1}{\gamma}}\right].$$

Theorem 2.3 motivates the modelling of points over high threshold using the GPD in what is sometimes called Points over Threshold model (POT). Suppose we have a sample of i.i.d. random variables $\{x_1, ..., x_n\}$. To approximate the upper tail of the distribution using the POT model, we identify an appropriate threshold $u$ and fit the GPD to the filtered sample $\{x_1^*,\ x_2^*, ...,\ x_{n^*}^*\}$ consisting of all $x$ that exceed $u$.

In this application we are interested in the lower tail of the distribution of a given SMoS. If we let $m_n$ denote the minima of a sequence, then it follows that $m_n = -\max(-X_1, ..., -X_n) = -\max\left(\tilde{X}_1, ..., \tilde{X}_n\right)$. Consequently, if there exists some sequence of constants $a_n$ and $b_n$ such that $\frac{-m_n - b_n}{a_n}$ converges to the generalized extreme value distribution, then by theorem 2.3 there exists some $\tilde{u}$ such that $\tilde{X} - \tilde{u} \mid \tilde{X} > \tilde{u}$ converges to the GPD. From a modelling perspective we simply have to apply POT to the negated data.

## 2.3 Statistical Inference

### 2.3.1 Estimation

Parametric inference in concerned with estimating some parameter vector $\boldsymbol{\theta}$ of a distribution $F(x; \boldsymbol{\theta})$ from a finite sample $\boldsymbol{x} = \{x_1, ..., x_n\}$. It is common in the field of extreme value theory to make use of the likelihood function given by

$$L(\boldsymbol{x}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}),$$

where $f$ is the density function associated with $F$. This assumes that $\boldsymbol{x}$ is an i.i.d. sample. Then, the maximum likelihood estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\{L(\boldsymbol{x};\boldsymbol{\theta})\} = \arg\max_{\boldsymbol{\theta}}\{\ell(\boldsymbol{x};\boldsymbol{\theta})\},$$

where $\ell(\boldsymbol{x};\boldsymbol{\theta}) = \log[L(\boldsymbol{x};\boldsymbol{\theta})]$. In general, we want any estimator of $\boldsymbol{\theta}$ to be efficient and consistent. It is known that the maximum likelihood estimator is consistent under certain regularity conditions, which implies that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\ell(\boldsymbol{x};\boldsymbol{\theta}) = 0$$

has roots $\hat{\boldsymbol{\theta}}$ which converges to the true value with probability 1. Moreover, under the same regularity conditions the maximum likelihood estimator achieves the Cramér-Rao lower bound, which implies that they are efficient [18].

It should be noted that the likelihood function is invariant to one-to-one transformations. Suppose we have a likelihood function for $\boldsymbol{\theta}$ but we are interested in $\boldsymbol{\psi} = g(\boldsymbol{\theta}) = [g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), ..., g_k(\boldsymbol{\theta})]$ where the mapping is one-to-one. Then it follows that

$$\sup_{\boldsymbol{\theta}} L(\boldsymbol{x};\boldsymbol{\theta}) = \sup_{\boldsymbol{\psi}} L(\boldsymbol{x}; g^{-1}(\boldsymbol{\psi})),$$

which implies that we can re-parametrize the likelihood function to obtain estimates of interest. However, if we are interested in estimating $g(\boldsymbol{\theta})$ we can obtain the estimate immediately from $\hat{\boldsymbol{\theta}}$, i.e. $\hat{g}(\boldsymbol{\theta}) = g(\hat{\boldsymbol{\theta}})$ even if $g$ is not a one-to-one bijection. If the transformation is one-to-one then certain asymptotic properties are preserved. For example, if $\hat{\boldsymbol{\theta}}$ is asymptotically normal, then $\hat{\boldsymbol{\psi}}$ will also be asymptotically normal [19].

Now suppose we are only interested in some subset of the parameter vector, i.e., let $\boldsymbol{\theta} = [\psi, \boldsymbol{\phi}]$, where $\psi$ is the parameter of interest and $\boldsymbol{\phi}$ are nuisance parameters. Then the profile log-likelihood function of $\psi$ is defined as

$$\ell_p(\psi) = \sup_{\boldsymbol{\phi}} \ell(\psi, \boldsymbol{\phi}) = \ell(\psi, \hat{\boldsymbol{\phi}}_\psi). \tag{5}$$

This implies that for a specific value of $\psi_0$ we compute $\hat{\boldsymbol{\phi}}$ that maximizes $\ell(\psi_0, \boldsymbol{\phi})$, and evaluating $\ell$ at $\left(\psi_0, \hat{\boldsymbol{\phi}}_0\right)$ yields the profile likelihood function at $\psi_0$. In many cases, inference based on the profile-likelihood performs better for the parameter $\psi$ compared to the standard likelihood function; particularly for small samples [17]. The profile likelihood estimate is given by

$$\hat{\psi}_p = \arg\max_{\psi}\{\ell_p(\psi)\}.$$

We can think of $\hat{\boldsymbol{\phi}}_{\boldsymbol{\psi}}$ as an estimator of the nuisance parameters, but these will generally be biased estimators except when $\psi = \hat{\psi}$. In order to adjust for the bias, we must multiply the profile likelihood with some function $M(\psi)$ so that the modified profile likelihood is given by

$$L_M(\psi) = M(\psi)L_p(\psi).$$

In general such a function exists and is given by

$$M(\psi) = |j_{\boldsymbol{\phi}\boldsymbol{\phi}}(\psi, \hat{\boldsymbol{\phi}}_\psi)|^{-\frac{1}{2}} \left| \frac{\partial \hat{\boldsymbol{\phi}}}{\partial \boldsymbol{\phi}} \right|, \tag{6}$$

where $j_{\boldsymbol{\phi}\boldsymbol{\phi}}$ is the corner of the observed information matrix corresponding to the nuisance parameters, and $\frac{\partial \hat{\boldsymbol{\phi}}}{\partial \boldsymbol{\phi}}$ is a sample space derivative which is usually difficult to compute [20]. One solution to this difficulty is to transform the parameter-space according to $(\psi, \boldsymbol{\phi}) \mapsto (\psi, \boldsymbol{\phi}')$ such that the $i^{-1}(\psi, \boldsymbol{\phi}')$ is a diagonal matrix. However, this involves solving a system of partial differential equations which have a solution in general, but may be difficult to find for particular problems [21]. As an alternative, we consider an approximation to the modified profile likelihood derived by [22]. Let us denote the approximate modified likelihood by $L_M$, then

$$L_M(\psi, \hat{\boldsymbol{\phi}}_\psi) = \frac{\left| j_{\boldsymbol{\phi}\boldsymbol{\phi}}(\psi, \hat{\boldsymbol{\phi}}_\psi) \right|}{\left| i(\psi, \hat{\boldsymbol{\phi}}_\psi; \hat{\psi}, \hat{\boldsymbol{\phi}}) \right|} L_p(\psi), \tag{7}$$

where

$$i(\psi, \hat{\boldsymbol{\phi}}_\psi; \hat{\psi}, \hat{\boldsymbol{\phi}}) = E_{\hat{\psi}, \hat{\boldsymbol{\phi}}} \left[ \left( \frac{\partial}{\partial \boldsymbol{\phi}} \ell(\psi, \boldsymbol{\phi}) \Big|_{\psi, \hat{\boldsymbol{\phi}}_\psi} \right) \left( \frac{\partial}{\partial \boldsymbol{\phi}} \ell(\psi, \boldsymbol{\phi}) \Big|_{\hat{\psi}, \hat{\boldsymbol{\phi}}} \right)^T \right].$$

Here, $E_{\hat{\psi}, \hat{\boldsymbol{\phi}}}$ is defined as the expectation under $f(\boldsymbol{x}; \hat{\psi}, \hat{\boldsymbol{\phi}})$. It turns out that the quantity $i(\psi, \hat{\boldsymbol{\phi}}_\psi; \hat{\psi}, \hat{\boldsymbol{\phi}})$ is quite cumbersome to compute for likelihoods involving the GPD. Fortunately, it can be approximated by

$$\tilde{i}(\psi, \hat{\boldsymbol{\phi}}_\psi; \hat{\psi}, \hat{\boldsymbol{\phi}}) = \sum_i^n \left( \frac{\partial}{\partial \boldsymbol{\phi}} \ell^{(i)}(\psi, \boldsymbol{\phi}) \Big|_{\psi, \hat{\boldsymbol{\phi}}_\psi} \right) \left( \frac{\partial}{\partial \boldsymbol{\phi}} \ell^{(i)}(\psi, \boldsymbol{\phi}) \Big|_{\hat{\psi}, \hat{\boldsymbol{\phi}}} \right)^T, \tag{8}$$

where $\ell^{(i)}$ denotes the log-likelihood for a single observation $i$. Meaning that that for $n$ observations $\ell(\theta) = \sum_{i=1}^n \ell^{(i)}(\theta)$ [23]. We will make use of this approximation to find MPL estimators $\hat{\psi}$ using the following likelihood function

$$L_M(\psi, \hat{\boldsymbol{\phi}}_\psi) = \frac{\left| j_{\boldsymbol{\phi}\boldsymbol{\phi}}(\psi, \hat{\boldsymbol{\phi}}_\psi) \right|}{\left| \tilde{i}(\psi, \hat{\boldsymbol{\phi}}_\psi; \hat{\psi}, \hat{\boldsymbol{\phi}}) \right|} L_p(\psi). \tag{9}$$

The motivation for this modification is that in practice one may obtain better approximations of the confidence bounds for small samples compared to the profile likelihood (see below) [24].

### 2.3.2 Confidence Intervals

Apart from estimating the parameters, we can also use certain properties of the likelihood functions to construct statistical tests from which we can derive confidence intervals for certain parameters

of interest. Ideally, we would like to derive some uniformly most powerful test based on Neymann-Pearson's lemma that would hold for both large and small samples. However, the nature of the likelihood function for the Generalized Pareto distribution makes this approach very difficult. Instead, large-sample theory is often used for inference with respect to the GPD. In what follows, we will list some fundamental theorems with respect to large-sample likelihood theory that will be used to construct confidence intervals in this thesis. These theorems require the following regularity conditions for $f(\boldsymbol{x};\boldsymbol{\theta}) = L(\boldsymbol{x};\boldsymbol{\theta})$:

1. The paremeter space $\Omega$ has finite dimension, and the true parameter $\boldsymbol{\theta}_0 \in \Omega$.
2. $f(\boldsymbol{x};\boldsymbol{\theta})$ is distinct for distinct values of $\boldsymbol{\theta}$.
3. The density $f(\boldsymbol{x};\boldsymbol{\theta})$ is three times differentiable with respect to $\boldsymbol{\theta}$, and $\mathrm{E}\left[|\frac{\partial^3}{\partial\boldsymbol{\theta}^3} f(\boldsymbol{x};\boldsymbol{\theta})|\right] < \infty$
4. Finally, we must have that

$$E\left[U(\boldsymbol{\theta})U(\boldsymbol{\theta})^T\right] = E\left[-\frac{\partial}{\partial\theta}U(\boldsymbol{\theta})^T\right] = i(\boldsymbol{\theta})$$

,

where the score function is $U(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}} f(\boldsymbol{x};\boldsymbol{\theta})$ and the Fisher information is $i(\boldsymbol{\theta}) = E\left[-\frac{\partial^2}{\boldsymbol{\theta}^2}\log\left(f(\boldsymbol{x};\boldsymbol{\theta})\right)\right]$. Note that if the dimension of $\boldsymbol{\theta}$ is greater than one, $\frac{\partial}{\partial\boldsymbol{\theta}}$ denotes the gradient operator and $\frac{\partial^2}{\partial\boldsymbol{\theta}^2}$ is the hessian. These conditions are unnecessarily restrictive, however, they are often easy to verify. Under these same regularity conditions one can show that the maximum likelihood estimate is asymptotically normally distributed [25].

**Theorem 2.4.** Let $X_1, ..., X_n$ be i.i.d. and suppose the distribution satisfies the regularity conditions 1 - 4 above. Then the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ satisfies

$$\hat{\boldsymbol{\theta}} \to N\left(\boldsymbol{\theta}, i^{-1}(\boldsymbol{\theta})\right).$$

A similar result holds for the ML estimate of $g(\boldsymbol{\theta})$, which is often referred to as the delta method. Assuming that $\boldsymbol{\theta} = [\theta_1, ..., \theta_k]$, then we introduce the following notation

$$\nabla g(\boldsymbol{\theta}) = \left[\frac{\partial}{\partial\theta_1} g(\boldsymbol{\theta}), ..., \frac{\partial}{\partial\theta_k} g(\boldsymbol{\theta})\right].$$

**Theorem 2.5.** Let $X_1, ..., X_n$ be i.i.d. and suppose the distribution satisfies the regularity conditions 1 - 4 above. Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimate of $\boldsymbol{\theta}$ and $g(\boldsymbol{\theta})$ is a differentiable function. Then it follows that

$$g(\hat{\boldsymbol{\theta}}) \to N\left(g(\boldsymbol{\theta}), \sigma_g^2\right),$$

where,

$$\sigma_g^2 = \nabla g(\boldsymbol{\theta})\, i^{-1}(\boldsymbol{\theta})\, \nabla g(\boldsymbol{\theta})^T.$$

Theorem 2.4 constitutes the basis for constructing a series of hypothesis tests regarding $\boldsymbol{\theta}$. In this thesis, we are particularly interested in constructing confidence intervals for our estimates based on two-sided hypothesis testing of the form given in equations (10 - 11).

$$\mathrm{H}_0 : \ \theta = 0 \tag{10}$$

$$\mathrm{H}_1 : \ \theta \neq 0 \tag{11}$$

Let $\alpha$ be the desired significance level of the test, and let $z_p$ denote the $\mathrm{p}^{\mathrm{th}}$ quantile of the standard normal distribution, then a consequence of Theorem 2.4 is that we can we can reject the null hypothesis $\mathrm{H}_0$ if

$$|\hat{\theta}| \geq z_{1-\alpha} i^{-\frac{1}{2}} \left(\hat{\theta}\right),$$

which is called the Wald statistic. From this test we can construct the following confidence interval for $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} - z_\alpha i^{-\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}\right) \leq \boldsymbol{\theta} \leq \hat{\boldsymbol{\theta}} + z_{1-\alpha} i^{-\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}\right). \tag{12}$$

Moreover, the corresponding confidence interval for Theorem 2.5 is given by

$$g\left(\hat{\boldsymbol{\theta}}\right) - z_{1-\alpha}\hat{\sigma}_g \leq g\left(\boldsymbol{\theta}\right) \leq g\left(\hat{\boldsymbol{\theta}}\right) + z_{1-\alpha}\hat{\sigma}_g, \tag{13}$$

where $\hat{\sigma}_g$ is $\sigma_g$ evaluated at $\hat{\boldsymbol{\theta}}$. Notice that the confidence intervals given in Equations 12 - 13 are necessarily symmetric about the maximum-likelihood estimate, and the estimated variance is related to the likelihood function only through the Fisher information [26]. In practice, and in extreme value theory in particular, convergence to the normal distribution is poor for finite samples. To remedy this, it is often beneficial to confidence intervals based on the asymptotic distribution of the log-likelihood ratio. These confidence intervals are based on the following theorems [27].

**Theorem 2.6.** Suppose $X_1, ..., X_n$ are i.i.d. with distribution function F satisfying the regularity conditions 1 - 4 above. And suppose we're interested in testing the simple hypothesis

$$\mathrm{H}_0 : \ \boldsymbol{\theta} = \boldsymbol{\theta}_0$$

$$\mathrm{H}_1 : \ \boldsymbol{\theta} \neq \boldsymbol{\theta_0}.$$

The likelihood ratio is defined as $R = \frac{L(\hat{\boldsymbol{\theta}})}{L(\boldsymbol{\theta}_0)}$. Then, under $\mathrm{H}_0$ it follows that

$$2\log\left(\frac{L\left(\hat{\boldsymbol{\theta}}\right)}{L\left(\boldsymbol{\theta}_0\right)}\right) = 2\left(\ell\left(\hat{\boldsymbol{\theta}}\right) - \ell\left(\boldsymbol{\theta}_0\right)\right) \to \chi^2_k,$$

where k is the number of free parameters under $\mathrm{H}_1$.

The corresponding confidence region is then give by

$$\left\{\boldsymbol{\theta}:\ \ell\left(\boldsymbol{\theta}\right) \geq \ell\left(\hat{\boldsymbol{\theta}}\right) - \frac{1}{2}q_{1-\alpha}\right\}, \tag{14}$$

where $q_{1-\alpha}$ is the $1-\alpha$ quantile of the $\chi^2_k$ distribution.

Similarly, we can construct confidence intervals for $g\left(\boldsymbol{\theta}\right)$ based on the asymptotic likelihood ratio test using the following theorem.

**Theorem 2.7.** Let $X_1, ..., X_n$ be i.i.d. with distribution function F satisfying regularity conditions 1 - 4 above. Let $g\left(\boldsymbol{\theta}\right)$ be differentiable and suppose we're interested in testing the following hypothesis

$$\begin{aligned} \mathrm{H}_0:&\ \ g\left(\boldsymbol{\theta}\right) = 0 \\ \mathrm{H}_1:&\ \ g\left(\boldsymbol{\theta}\right) \neq 0. \end{aligned}$$

Then,

$$2\log\left(R\right) = 2\left(\ell(\hat{\boldsymbol{\theta}}) - \ell\left(\boldsymbol{\theta}_0\right)\right) \rightarrow \chi^2_1.$$

From Theorem 2.7 we can construct the following confidence region for $g\left(\boldsymbol{\theta}\right)$:

$$I^g_{1-\alpha} = \left\{g\left(\boldsymbol{\theta}\right):\ \ell\left(\boldsymbol{\theta}\right) \geq \ell(\hat{\boldsymbol{\theta}}) - \frac{q_{1-\alpha}}{2}\right\}. \tag{15}$$

The lower and upper bounds of the confidence interval with usual interpretation, is then given by $\min\left\{I^g_{1-\alpha}\right\}$ and $\max\left\{I^g_{1-\alpha}\right\}$ respectively. Computing these bounds essentially amounts to a constrained maximization problem and the following Lagrangian can be used to find said bounds immediately

$$\mathcal{L}\left(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\phi},\varepsilon\right) = \ell\left(\boldsymbol{\theta}\right) + \varepsilon\left(\phi - g(\boldsymbol{\theta})\right), \tag{16}$$

where $\varepsilon$ is the the multiplier. Let $(\hat{\boldsymbol{\theta}}, \hat{\phi}, \hat{\varepsilon})$ be the parameters that maximizes the $\mathcal{L}$, and let $(\hat{\boldsymbol{\theta}}_\varepsilon, \hat{\phi}_\varepsilon)$ be the parameters that maximizes $\mathcal{L}$ for a given $\varepsilon$. Then by the properties of the Lagrangian it follows that the confidence region in Equation 14 can be re-written in terms of $\varepsilon$ as

$$\left\{\varepsilon : \mathcal{L}(\hat{\boldsymbol{\theta}}_\varepsilon, \hat{\phi}_\varepsilon, \varepsilon) \geq \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\phi}, \hat{\varepsilon}) - \frac{q_{1-\alpha}}{2}\right\},$$

and the lower confidence bound is given by finding $\varepsilon_-$ and $\varepsilon_+$ such that $\mathcal{L}(\hat{\boldsymbol{\theta}}_\varepsilon, \hat{\phi}_\varepsilon, \varepsilon) = \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\phi}, \hat{\varepsilon}) - \frac{q_{1-\alpha}}{2}$, and so the upper and lower bounds are $g(\hat{\boldsymbol{\theta}}_{\varepsilon_-})$ and $g(\hat{\boldsymbol{\theta}}_{\varepsilon_+})$ [28].

It can be shown that the likelihood ratio for the profile likelihood and the modified profile likelihood function also converge to $\chi^2$ distribution, which consequently implies that we can construct the following confidence regions for the profile likelihood [26]

$$\left\{\boldsymbol{\psi} : \ell_p(\psi) \geq \ell_p(\hat{\psi}) - \frac{q_{1-\alpha}}{2}\right\}, \tag{17}$$

and for the modified profile likelihood we have [29]

$$\left\{\boldsymbol{\psi} : \ell_p(\psi) \geq \ell_{mp}(\hat{\psi}) - \frac{q_{1-\alpha}}{2}\right\}, \tag{18}$$

where $q_{1-\alpha}$ is $1-\alpha$ quantile of the the $\chi^2_k$ distribution, with $k$ equal to the dimension of $\boldsymbol{\psi}$.

# 3 Method

In this section we delineate how the theory from Section 2 can be applied to construct confidence intervals for the near-crash intensity and to assess the relative safety between two intersection using SMoS.

## 3.1 Probabilistic Model of Near-Crash Events

Let us denote the value of an SMoS by $S$, and treat it as a random variable. Then, in order to model the lower tail distribution of $S$ using the generalized Pareto distribution, we must assume that $S$ is continuous. Further, suppose we observe a sequence $\{S_1, ..., S_n\}$, and let $F$ be the distribution function for $\tilde{S} = -S$, then from Theorem 2.1 there must exist a sequence of constants $a_n$ and $b_n$ such that $F_n$ converges to a max-stable distribution which will be one of the extremal types. Consequently, if the above assumptions hold, and for a sufficiently large threshold $\tilde{u}$, it follows that

$$P(\tilde{S} - \tilde{u} \leq z | \tilde{S} > \tilde{u}) = 1 - \left(1 + \gamma\frac{z}{\sigma}\right)^{-\frac{1}{\gamma}},$$

from Theorem 2.3. For concision, we set $Z = \tilde{S} - \tilde{u}$, and as we are interested in $P(S < s)$, then the relation of the the probability $P(Z > z | Z > 0)$ to $P(S < s | S < u)$ is easily seen from the following identities

$$\begin{aligned}
&\left(1 + \gamma\frac{z}{\sigma}\right)^{-\frac{1}{\gamma}} = P(Z > z | Z > 0) = P(\tilde{S} - \tilde{u} > z | \tilde{S} > \tilde{u}) \\
&= P(-(\tilde{S} - \tilde{u}) < -z | -\tilde{S} < -\tilde{u}) = P(S - u < -z | S < u) \\
&= P(S - u < s - u | S < u) = P(S < s | S < u).
\end{aligned}$$

Meaning that $P(Z > z | Z > 0)$ is interpreted as the probability that $S$ is less than $s$ conditional on $S < u$. Finally, the lower tail probability the $P(S < s)$ is then given by multiplication with $P(S < s)$ i.e.

$$P(S < s) = P(S < s | S < u)P(S < u) = \left(1 + \gamma\frac{z}{\sigma}\right)^{-\frac{1}{\gamma}} P(Z > 0).$$

When assessing the safety of some type of traffic infrastructure, one is generally interested in the number of times $S < s$ during a period, where is $s$ is some level of interest. This suggests that we

are interested in some counting process where the intensity is a function of the GPD parameters. Let us denote this counting process by $C_f(s,t)$, since it is both a function of the sampling period $t$ and the quantile of interest $s$. We want $C_f(s,t)$ to be the count of the number of interactions during a period $t$ with an SMoS less than $s$. Now we make the following simplifying assumptions regarding the interactions between right turning vehicles and cyclists:

1. The number of interactions within any subinterval of $t$ are independent of the number in any other subinterval.
2. Interactions occur with constant intensity over the time period $t$.
3. The value of an individual SMoS is independent of the number of interactions.

Let $N(t)$ be the number of interactions between right turning vehicles and cyclists during a period $t$; $n$ is the number of interactions with $S < u$; and let $X$ be a Bernoulli variable defined as

$$p(x) = \begin{cases} \left(1 + \gamma \frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}, & \text{for } x = 1 \\ 1 - \left(1 + \gamma \frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}, & \text{for } x = 0, \end{cases}$$

where $q(s) = \tilde{s} - \tilde{u} = -s + u$ is the desired level of the SMoS for which we want to know the counting distribution. We will call all observations of $s < u$ "shortfalls below $u$". Together with the additional assumption that $N(t)$ follows a Poisson distribution, the assumptions above correspond to the following distributions

$$\begin{aligned} &N(t) \in Po(\lambda t) \\ &n \mid N(t) \in Bin(N(t), \pi) \\ &X \mid n, N(t) \in Bernoulli\left(\left\{1 + \gamma \frac{q(s)}{\sigma}\right\}^{-\frac{1}{\gamma}}\right). \end{aligned}$$

We use $\mid$ to denote conditioning so that $X \mid n, N(t)$ is defined as the distribution of $X$ conditional on $n$ and $N(t)$. Also, note that $\pi = P(S < u)$. The random variable $C_f(s,t)$ can then be seen as the sum of all $X$ over $n$, i.e.

$$C_f(s,t) = \sum_{i=1}^{n} X_i \tag{19}$$

We briefly set $C_f(s,t) = C$ and $\left(1 + \gamma \frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}} = p$. Note that $C \mid n, N(t) \in Bin(n,p)$, so the

distribution of $C_f(s,t)$ can now be derived using the law of total probability as follows

$$\begin{aligned}
P(C=c) &= \sum_{N=0}^{\infty}\sum_{n=0}^{N} P(N(t)=N)P(n=n \mid N(t)=N)P(C=c \mid n, N(t)) \\
&= \sum_{N=0}^{\infty}\sum_{n=0}^{N} e^{-\lambda t}\frac{(\lambda t)^N}{N!}\binom{N}{n}\pi^n(1-\pi)^{N-n}\binom{n}{c}p^c(1-p)^{n-c} \\
&= \sum_{N=0}^{\infty}\sum_{n=0}^{N} e^{-\lambda t}\frac{(\lambda t)^N}{N!}\binom{N}{c}\binom{N-c}{n-c}(\pi p)^c(\pi(1-p))^{n-c}(1-\pi)^{N-n} \\
&= \sum_{N=0}^{\infty}\binom{N}{c}e^{-\lambda t}\frac{(\lambda t)^N}{N!}(\pi p)^c\sum_{k=0}^{N-c}\binom{N-c}{k}(\pi(1-p))^k(1-\pi)^{N-c-k} \\
&= \sum_{N=0}^{\infty}\binom{N}{c}e^{-\lambda t}\frac{(\lambda t)^N}{N!}(\pi p)^c\left(\pi(1-p)+1-\pi\right)^{N-c} \\
&= e^{-\lambda t}\frac{(\lambda t\pi p)^c}{c!}\sum_{N=0}^{\infty}\frac{(\lambda t(1-\pi p))^{N-c}}{(N-c)!} = e^{-\lambda t\pi p}\frac{(\lambda t\pi p)^c}{c!},
\end{aligned} \tag{20}$$

where we have used the binomial theorem and some combinatoric identities. Now the PMF in Equation 20 can be recognized as the Poisson distribution with intensity $\lambda_c = \lambda t\pi\left(1+\gamma\frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}$. The implication of this model is that if we know the parameters $(\lambda, \pi, \gamma, \sigma)$, then we can compute the expected number of interactions with $S < s$ for all $s < u$. In other words, once we estimate the parameters, then we also obtain estimates of the counting distribution $C_f(s,t)$ for all $s < u$ and all $t$. As is standard in the literature, we will let $S < 0$ correspond to the event that an interaction results in a crash.

The model in Equation 20 also allows us to distinguish between the number interactions as a function of the traffic intensity corresponding to the parameter $\lambda$, and the probability that any one interaction generates a SMoS less than $s$, which corresponds to the factor $\pi\left(1+\gamma\frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}$.

## 3.2 Parameter Estimation

In the previous section, we concluded that the number of interactions with $S < s$ during a time period $t$ follows a Poisson distribution under certain simplifying assumptions, i.e.

$$C_f(s,t) \in Po\left(\lambda t\pi\left(1+\gamma\frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}\right).$$

What remains to be shown is how the parameter vector $\boldsymbol{\theta} = (\lambda, \pi, \gamma, \sigma)$ can be estimated. In extreme-value theory, parameter estimation is usually performed through likelihood- and moment-based estimators. In this thesis, we restrict ourselves to the class of likelihood-based methods, since the maximum likelihood estimates are invariant to one-to-one transformations of the parameter space; which will prove to be quite useful indeed.

From the data description, we know that we observe the random vector $\boldsymbol{X} = (N(t), n, \boldsymbol{z})$, where $N(t)$ is the total number of interactions, $n$ is the number of events with an SMoS $S < u$ and $z_i = \tilde{s}_i - \tilde{u} = -s_i + u$, for some reasonably chosen threshold $u$. The likelihood is then simply the density function $f(\boldsymbol{X})$ viewed as function of the parameter vector $\boldsymbol{\theta}$. Then, under the same assumptions as in Section 3.1, the density function follows immediately from Bayes' Theorem

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{X}) = f(\boldsymbol{X}) = f(\boldsymbol{z} \mid N(t), n) P(n = n \mid N(t)) P(N(t) = N) \\ = e^{-\lambda t} \frac{(\lambda t)^N}{N!} \binom{N}{n} \pi^n (1-\pi)^{N-n} \prod_{i=1}^{n} \frac{1}{\sigma} \left(1 + \gamma \frac{z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}. \end{aligned} \tag{21}$$

Then, the maximum likelihood estimator of $\boldsymbol{\theta}$ is given by the solution to the system of equations $\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}; \boldsymbol{X}) = 0$. Since the parameters $\lambda$ and $\pi$ are orthogonal to one another and to the GPD parameters, these can effectively be estimated separately. There is no analytical solution for the GPD parameters, so the solution will have to be computed numerically, which can be done using the extRemes package in R [30].

The question remains whether or not the likelihood in Equations 21 will converge to the normal distribution, and whether or not the estimators are consistent and efficient, as is desired. It was shown by Smith that for $\gamma > -0.5$ , the maximum likelihood estimator for the GPD parameters behaved regularly. Note that the number of interactions $N(t)$ and, as a consequence, the number of exceedences $n$ are generated by a Poisson process. Therefore, we are interested in the distribution of the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ as $t \to \infty$, and since the textbook proof of asymptotic normality proceeds by taking the limit as the number of observations goes to infinity, this requires an explicit proof for asymptotic normality. This turns out to be quite difficult. Of course, the model in Equation 21 is equivalent to the model proposed by Davison and Smith [31] if we only consider the marginal distribution of $(n, \boldsymbol{z})$ with the following density function

$$\sum_{N=0}^{\infty} f(\boldsymbol{X}) = e^{-\lambda \pi t} \frac{(\lambda \pi t)^N}{N!} \prod_{i=1}^{n} \frac{1}{\sigma} \left(1 + \gamma \frac{z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}.$$

In this case the exceedences follow a Poisson distribution. In other applications, the exceedences have been treated as Binomial, for example in Coles 2001 [17], which in our case can be obtained by conditioning on $N(t)$. The point is, that in both cases the asymptotic normality of the estimates are not proven, and they simply add the corresponding components to the Fisher information matrix (see below). We will follow the same approach.

In order to compute the confidence interval based on the Wald test as is given in Equation 13, we need to compute the Fisher information matrix $i(\boldsymbol{\theta}) = E\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell(\boldsymbol{\theta})\right]$. From [32] we know that

the Fisher information for the generalized Pareto distribution is given by

$$i_{\mathrm{GPD}}\left(\boldsymbol{\theta}\right) = \begin{pmatrix} \frac{2}{(1+\gamma)(1+2\gamma)} & \frac{-1}{\sigma(1+\gamma)(1+2\gamma)} \\ \frac{-1}{\sigma(1+\gamma)(1+2\gamma)} & \frac{1}{\sigma^2(1+2\gamma)} \end{pmatrix} = \boldsymbol{M}, \tag{22}$$

and we note that the following the following partial derivatives hold for the log-likelihood corresponding to Equation 21:

$$\begin{array}{cccc} \frac{\partial^2}{\partial\lambda^2}\ell = \frac{-N}{\lambda^2} & \frac{\partial^2}{\partial\lambda\partial\pi}\ell = 0 & \frac{\partial^2}{\partial\lambda\partial\gamma} = 0 & \frac{\partial^2}{\partial\lambda\partial\sigma}\ell = 0 \\ \frac{\partial^2}{\partial\pi^2}\ell = -\frac{n}{\pi^2} - \frac{N-n}{(1-\pi)^2} & \frac{\partial^2}{\partial\pi\lambda}\ell = 0 & \frac{\partial^2}{\partial\pi\partial\gamma}\ell = 0 & \frac{\partial^2}{\partial\pi\partial\sigma}\ell = 0. \end{array}$$

Let $j_{\gamma\sigma}(\boldsymbol{\theta})$ be the corner of the observed Fisher information corresponding to the GPD parameters for the likelihood in Equation 21. Since we treat the number of shortfalls as random, we have that

$$\begin{aligned} i_{\gamma\sigma}(\boldsymbol{\theta}) &= \sum_{N=0}^{\infty}\sum_{n=0}^{N} E\left[j_{\gamma\sigma}(\boldsymbol{\theta}) \mid N(t)=N, n=n\right] P(n \mid N)P(N) \\ &= \sum_{N=0}^{\infty}\sum_{n=0}^{N} n\boldsymbol{M}P(n \mid N)P(N) = \lambda t\pi\boldsymbol{M}. \end{aligned}$$

Using the fact that the marginal distribution of $n \in Po(\lambda t\pi)$, and $N-n \in Po(\lambda t(1-\pi))$, we have that the Fisher information matrix given by

$$i\left(\boldsymbol{\theta}\right) = \begin{pmatrix} \frac{t}{\lambda} & 0 & 0 & 0 \\ 0 & \frac{\lambda t}{\pi(1-\pi)} & 0 & 0 \\ 0 & 0 & \frac{2\lambda t\pi}{(1+\gamma)(1+2\gamma)} & \frac{-\lambda t\pi}{\sigma(1+\gamma)(1+2\gamma)} \\ 0 & 0 & \frac{-\lambda t\pi}{\sigma(1+\gamma)(1+2\gamma)} & \frac{\lambda t\pi}{\sigma^2(1+2\gamma)} \end{pmatrix}. \tag{23}$$

From Theorem 2.4, we know that $\hat{\boldsymbol{\theta}} \to N(\boldsymbol{\theta}, i^{-1}(\boldsymbol{\theta}))$, and so the asymptotic covariance matrix of the ML estimators is given by

$$i^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\lambda}{t} & 0 & 0 & 0 \\ 0 & \frac{\pi(1-\pi)}{\lambda t} & 0 & 0 \\ 0 & 0 & \frac{(1+\gamma)^2}{\lambda \pi t} & \frac{\sigma(1+\gamma)}{\lambda \pi t} \\ 0 & 0 & \frac{\sigma(1+\gamma)}{\lambda \pi t} & \frac{2(1+\gamma)\sigma^2}{\lambda \pi t} \end{pmatrix}. \tag{24}$$

We recall that $\lambda_c = \lambda t \pi \left(1 + \gamma \frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}$, meaning that $\lambda_c$ is a function of the parameter vector $\boldsymbol{\theta}$. By the properties of the maximum likelihood estimator, we have that $\hat{\lambda}_c = \lambda_c(\hat{\boldsymbol{\theta}})$. As a consequence of Theorem 2.5, we can then form the following confidence interval for $\lambda_c$ with significance level $\alpha$

$$\lambda_c(\hat{\boldsymbol{\theta}}) - z_{\frac{1-\alpha}{2}} \hat{\sigma}_c \leq \lambda_c \leq \lambda_c(\hat{\boldsymbol{\theta}}) + z_{1-\frac{\alpha}{2}} \hat{\sigma}_c, \tag{25}$$

where,

$$\hat{\sigma}_c^2 = \nabla \lambda_c(\hat{\boldsymbol{\theta}}) i^{-1}(\hat{\boldsymbol{\theta}}) \nabla \lambda_c(\hat{\boldsymbol{\theta}}),$$

and,

$$\nabla \hat{\lambda}_c = \left[ \frac{\partial}{\partial \lambda} \lambda_c,\ \frac{\partial}{\partial \pi} \lambda_c,\ \frac{\partial}{\partial \gamma} \lambda_c,\ \frac{\partial}{\partial \sigma} \lambda_c \right] \Big|_{(\hat{\lambda},\ \hat{\pi},\ \hat{\gamma},\ \hat{\sigma})}.$$

As was noted in Section 2.4, we may obtain estimates with better small sample properties using the profile likelihood function. To this end, the likelihood function in Equation 21 must be re-parametrized in terms of $\lambda_c$. Since we have that

$$\lambda = \frac{\lambda_c}{\pi} \left(1 + \gamma \frac{q(s)}{\sigma}\right)^{\frac{1}{\gamma}},$$

and since $\lambda_c$ is a monotone function of $\lambda$, the transformation $(\lambda, \pi, \gamma, \sigma) \mapsto (\lambda_c, \pi, \gamma, \sigma)$ is one-to-one. Consequently, the likelihood function is invariant to the transformation and is given by

$$L(\boldsymbol{\theta}; \boldsymbol{X}) = e^{-\frac{\lambda_c t}{\pi}\left(1+\gamma \frac{q(s)}{\sigma}\right)^{\frac{1}{\gamma}}} \frac{\left(\frac{\lambda_c t}{\pi}\left(1 + \gamma \frac{q(s)}{\sigma}\right)^{\frac{1}{\gamma}}\right)^N}{N!} \binom{N}{n} \pi^n (1-\pi)^{N-n} \prod_{i=1}^{n} \frac{1}{\sigma} \left(1 + \gamma \frac{z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}. \tag{26}$$

Since we are interested in the near-crash intensity $\lambda_c$, all other parameters can be considered nuisance parameters. Let $\boldsymbol{\phi} = [\pi, \gamma, \sigma]$, then the profile log-likelihood function with respect to $\lambda_c$

is given by

$$\ell_p(\lambda_c) = \ell(\lambda_c, \hat{\boldsymbol{\phi}}_{\lambda_c}) = \sup_{\boldsymbol{\phi}} \ell(\lambda_c, \boldsymbol{\phi}; \boldsymbol{X}). \tag{27}$$

The profile likelihood estimate of $\lambda_c$ is then given by

$$\hat{\lambda}_c^p = \arg\max_{\lambda_c} \{\ell_p(\lambda_c)\}.$$

There is no analytical expression for $\ell_p(\lambda_c)$, and so obtaining $\hat{\lambda}_c^p$ amounts to a numerical optimization problem in two parts. Firstly, for a given value of $\lambda_c^0$, we have to find $\hat{\boldsymbol{\phi}}_{\lambda_c^0}$ that maximizes $\ell(\lambda_c^0, \boldsymbol{\phi})$. Then we need to find $\hat{\lambda}_c^p$ that maximizes $\ell_p(\lambda_c)$. The problem of obtaining $\hat{\boldsymbol{\phi}}_{\lambda_c^0}$ corresponds to the following non-linear optimization problem:

$$\begin{gathered} \text{Maximize } \ell(\lambda_c^0, \boldsymbol{\phi}; \boldsymbol{X}) \text{ w.r.t. } \boldsymbol{\phi} \\ \text{Subject to } (1 + \gamma \tfrac{z}{\sigma}) > 0, \quad (1 + \gamma \tfrac{q(s)}{\sigma}) > 0. \end{gathered} \tag{28}$$

There are broad class of algorithms for optimization with non-linear inequality constraints, and following [33] we use the SLSQP algorithm which can easily be implemented in R using the NLOPTR package [34]. The second step is to maximize $\ell_p(\lambda_c)$, which is carried out using R's standard *optim* function.

Let $q_{1-\alpha}$ be the $1-\alpha$ quantile of the $\chi_1^2$ distribution, then the confidence region with confidence level $1-\alpha$ for $\lambda_c$ is given by

$$I_{1-\alpha} = \left\{ \lambda_c : \ell(\lambda_c) \geq \ell(\hat{\lambda}_c^p) - \frac{q_{1-\alpha}}{2} \right\},$$

so that the corresponding confidence interval is

$$\min(I_{1-\alpha}) \leq \lambda_c \leq \max(I_{1-\alpha}). \tag{29}$$

This amounts to finding the roots of the function $2\left(\ell_p(\hat{\lambda}_c^p) - \ell_p(\lambda_c)\right) = q_{1-\alpha}$. This argument for the profile likelihood based confidence for the crash intensity $\lambda_c$ is very similar to the profile likelihood based confidence for the return levels in Coles (2001) [17], and Davison and Smith (1990) [31].

Finally, we consider a modification to the profile likelihood according to Equation 7 with the appropriate approximation for $i(\psi, \hat{\boldsymbol{\phi}}_\psi; \hat{\psi}, \hat{\boldsymbol{\phi}})$ as given in Equation 8. This approximation requires a log-likelihood function that can be re-written as the sum of log-likelihood functions for single observations, i.e. $\ell(\theta) = \sum_{i=1}^N \ell^{(i)}(\theta)$. With the likelihood given in Equation 21, this is not possible. However, since we have the times at which an interaction between vehicle and cyclist was recorded, it means that we also observe a Poisson process for the number of interactions. Furthermore, the observance of the number of shortfalls $n$ is equivalent to observing $n$ Bernoulli

variables. So, in order to observe an SMoS at time $t_i$, there must occur an interaction and a shortfall below the threshold. Since we also observe interactions without shortfalls, this should be included in the likelihood function. So the proposed likelihood function for a single observation is

$$L^{(i)}(\boldsymbol{\theta}; \boldsymbol{X}) = \begin{cases} \lambda\pi(t_i - t_{i-1})e^{-\lambda(t_i - t_{i-1})} \frac{1}{\sigma}\left(1 + \gamma\frac{z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}, & \text{for } s_i < u \\ \lambda(1-\pi)(t_i - t_{i-1})e^{-\lambda(t_i - t_{i-1})}, & \text{for } s_i \geq u, \end{cases} \tag{30}$$

and the full likelihood is then simply given by $L(\boldsymbol{\theta}; \boldsymbol{X}) = \prod_{i=1}^{N} L^{(i)}(\boldsymbol{\theta}; \boldsymbol{X})$. Again, we can reparametrize the likelihood in terms of $(\lambda_c, \boldsymbol{\phi})$, and the profile likelihood function can be obtained using the same optimization scheme as above. Also, the approximation in Equation 8 is now well-defined - albeit tedious to compute. Fortunately, the components $j_{\boldsymbol{\phi}\boldsymbol{\phi}}(\lambda_c, \hat{\boldsymbol{\phi}}_{\lambda_c})$ and $\tilde{i}(\lambda_c, \hat{\boldsymbol{\phi}}_{\lambda_c}; \hat{\lambda}_c, \hat{\boldsymbol{\phi}})$ are straightforward computations in any symbolic computing software such as Maple, which can be translated to an R matrix. Then, to find the estimate of $\lambda_c$, we simply maximize the modified profile likelihood function. However, the observed information $j_{\phi\phi}$ is not necessarily positive definite, and we found that the MPL only existed in a neighbourhood of the maximum-likelihood estimate. Therefore, it is important to plot the MPL before optimizing over it. We refer to Appendix B for how to implement the profile likelihood and modified profile likelihood in R.

Finally, let $q_{1-\alpha}$ be the $1-\alpha$ quantile of $\chi^2_1$ distributione, then the the confidence region for $\lambda_c$ based on the MPL is given by

$$\left\{\lambda_c : \ell_p(\lambda_c) \geq \ell_{mp}(\hat{\lambda}_c^{mp}) - \frac{q_{1-\alpha}}{2}\right\}. \tag{31}$$

### 3.3 Threshold Selection and Model Validation

Provided that the variation of crash or near-crash events $C_f(s,t)$ is appropriately described by the Poisson distribution, we are left with two distinct problems; on the one hand, we require a reasonably high threshold $\tilde{u}$ to ensure that $\tilde{S} - \tilde{u} \mid \tilde{S} > \tilde{u}$ converges to the GPD, and on the other hand, we need a large enough sample to ensure accurate parameter estimates with the desired properties. Naturally, there is a trade-off between these two, since a higher threshold (for the negated observations) implies fewer observations. In this thesis, we restrict ourselves to two graphical methods for threshold selection.

Earlier we noted that there is stability in the GPD parameters for increasing thresholds, with an appropriate transformation of the scale parameter. To make this point more clear; suppose that $\tilde{S} - \tilde{u}_0 \mid \tilde{S} > \tilde{u}_0$ converges to the GPD, then it will also converge for any $\tilde{u} > \tilde{u}_0$ with the same shape parameter $\gamma$ but different scale parameter $\sigma$. Let $\sigma_{\tilde{u}_0}$ correspond to the scale parameter of the GPD when the threshold is $\tilde{u}_0$, and $\sigma_{\tilde{u}}$ when the threshold is $\tilde{u}$, then it can be shown that

$$\sigma_{\tilde{u}} = \sigma_{\tilde{u}_0} + \gamma\left(\tilde{u} - \tilde{u}_0\right). \tag{32}$$

If we set $\sigma^* = \sigma_{\tilde{u}} - \gamma\tilde{u}$, then by plugging into Equation 32, we have

$$\sigma^* = \sigma_{\tilde{u}} - \gamma\tilde{u} = \sigma_{\tilde{u}_0} + \gamma\left(\tilde{u} - \tilde{u}_0\right) - \gamma\tilde{u} = \sigma_{\tilde{u}_0} - \gamma\tilde{u}_0.$$

Consequently, $\sigma^*$ is constant for all $\tilde{u} > \tilde{u}_0$. Moreover, a confidence interval can be computed for $\sigma^*$ using the delta method. So, one strategy for threshold is to plot the following tuples along with the corresponding confidence intervals:

$$\left\{(\tilde{u},\ \hat{\sigma}^*) :\ \tilde{u} < max(\tilde{s})\right\},$$

and,

$$\left\{(\tilde{u},\ \hat{\gamma}) :\ \tilde{u} < max(\tilde{s})\right\}.$$

A suitable threshold is where the line connecting the points of the plots starts to become horizontal for both the $\gamma$ and $\sigma^*$ parameters when you account for the confidence intervals. To summarize; we fit the GPD to the data over a range of threshold values and compute the estimates and confidence intervals for the parameters $\gamma$ and $\sigma^*$, and then find the lowest threshold above which the parameters appear stable. These plots are usually called threshold range plots [17].

The second method that will inform our choice of threshold is the mean excess plot (also referred to as mean residual life plot). It is based on the mean excess function defined as

$$M(u) = \mathrm{E}[X - u \mid X > u],$$

and for the GPD in particular, it can be shown that $M$ is linear in u,

$$M(u) = \frac{\sigma}{1-\gamma} + \frac{\gamma}{1-\gamma}u,$$

provided that $\gamma < 1$. Now suppose that the $\tilde{S} - \tilde{u_0} \mid \tilde{S} > \tilde{u}_0$ converges to the GPD, then so will $\tilde{S} - \tilde{u} \mid \tilde{S} > \tilde{u}$ for any $\tilde{u} > \tilde{u}_0$. Furthermore, an estimate of the mean excess function is given by $\hat{M}(\tilde{u}) = \frac{1}{n_{\tilde{u}}}\sum_{i=1}^{n_{\tilde{u}}} \tilde{s}_{(i)} - \tilde{u}$, where $\tilde{s}_{(1)}, ..., \tilde{s}_{(n_{\tilde{u}})}$ are the negated values of $S$ that exceed $\tilde{u}$. Consequently, we would expect the mean excess plot given by

$$\left\{\left(\tilde{u},\ \frac{1}{n_{\tilde{u}}}\sum_{i=1}^{n_{\tilde{u}}} \tilde{s}_{(i)} - \tilde{u}\right) :\ \tilde{u} < \max(\tilde{s})\right\},$$

to be linear above the minimum threshold required to ensure convergence [17].

After estimating the parameters, we should also check that the GPD in fact provides a good fit to the data. We make use of three common plots, the first being the probability plot defined as

$$\left\{\left(\frac{i}{n+1},\ 1-\left(1+\hat{\gamma}\frac{\tilde{s}_{(i)}-\tilde{u}}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}}\right):\ \text{for } i=1,...,n\right\}, \tag{33}$$

where, $\tilde{s}_{(i)},...,\tilde{s}_{(n)}$ is the ordered sample. This amounts to plotting the empirical probabilities against the corresponding estimated probabilities according to the model. We may also plot the observed quantiles against the estimated quantiles in a quantile plot given by

$$\left\{\left(\tilde{u}+\frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{i}{n+1}\right)^{\hat{\gamma}}-1\right],\ \tilde{s}_{(i)}\right),\ \text{for } i=1,...,n\right\}. \tag{34}$$

In principle, the quantile and probability plot should be linear if the GPD is a good fit [17]. Finally, it is a good idea to check whether the Poisson distribution for $C_f(s,t)$ given in Equation 20 is a good fit to the data. One approach to this is to order the observed SMoS values $s_{(1)},...s_{(n)}$ and then plot the number of observations with an SMoS lower than $s_{(i)}$ for $i=1,...,n$. At the same time, we can plot the expected number of interactions with an SMoS lower than $s_{(i)}$ as according to $\hat{\lambda}_c(s_{(i)})$, along with the corresponding confidence intervals. This amounts to plotting the following two sets of tuples simultaneously:

$$\begin{gathered}\{(\,s_{(i)},\ i\,),\ \text{for } i=1,...,n\}\\ \{(s_{(i)},\ \hat{\lambda}_c(s_{(i)}),\ \text{for } i=1,...,n\}.\end{gathered} \tag{35}$$

If we also plot the corresponding confidence intervals for the $\lambda_c(s_{(i)})$, then we get a good idea of whether or not $C_f(s,t)$ is a good model for the observed data.

## 3.4 Relative Safety

The major purpose of this thesis is to compare the safety of two different road intersection designs with respect to cars and cyclists. Even though it may be difficult to obtain accurate estimates of the crash intensity, or probability of a crash using statistical models, we may nevertheless be able to draw correct conclusions as to whether or not a particular design is safer than the other.

To assess the relative safety of an intersection compared to another, we're interested in comparing the probability that an interaction between a car and a cyclist has an SMoS below a certain value for two intersection. As such, we're interested in the quantity $P(S<s)$, and for $s<u$ we know from Section 2.3 that

$$P(S<s)\approx P(S<u)\left(1+\gamma\frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}=\pi\left(1+\gamma\frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}.$$

Let us denote this probability by $\pi_c(s)$, so that $\pi_c(s)=\pi\left(1+\gamma\frac{q(s)}{\sigma}\right)^{-\frac{1}{\gamma}}$. We will sometimes drop the fact that the probability is a function of $s$ for convenience, so that $\pi_c(s)=\pi_c$. We're interested

in comparing $\pi_c$ for intersections in the Netherlands and intersections in Denmark, meaning that we're interested in the following hypothesis

$$\begin{aligned} \mathrm{H}_0 &: \pi_c^{NE} - \pi_c^{DK} = 0 \\ \mathrm{H}_1 &: \pi_c^{NE} - \pi_c^{DK} \neq 0. \end{aligned} \tag{36}$$

Note that our actual hypothesis is that the intersections in Denmark are generally safer than the intersections in the Netherlands, which would imply the one sided alternative hypothesis $\pi_c^{NE} > \pi_c^{DK}$. However, the asymptotic distribution of the likelihood ratio under the null hypothesis implies a two-sided hypothesis, but since the two-sided hypothesis is more difficult to reject, then provided that $\hat{\pi}_c^{NE} > \hat{\pi}_c^{DK}$, we can also reject the null hypothesis for the one-sided test.

We will use the confidence interval for $\pi_c^{NE} - \pi_c^{DK}$ as a test statistic in order to test the hypothesis; this means that the null hypothesis is accepted if zero is an element of the interval, and rejected if zero is not an element of the interval.

Now, let $\boldsymbol{\theta}^{NE} = [\lambda^{NE}, \pi^{NE}, \gamma^{NE}, \sigma^{NE}]$ and $\boldsymbol{\theta}^{DK}$ be the corresponding parameter vector for the Danish intersection, and let $\boldsymbol{\theta} = [\,\boldsymbol{\theta}^{NE},\ \boldsymbol{\theta}^{DK}]$. Further, we will assume that the observations from the Netherlands and Denmark are independent, so that the joint log-likelihood is given by

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}^{NE}) + \ell(\boldsymbol{\theta}^{DK}), \tag{37}$$

and therefore, by assumption, $\boldsymbol{\theta}^{NE}$ and $\boldsymbol{\theta}^{DK}$ are orthogonal. As a consequence, the corresponding Fisher information is given by

$$i(\boldsymbol{\theta}) = \begin{pmatrix} i(\boldsymbol{\theta}^{NE}) & \mathbf{0} \\ \\ \mathbf{0} & i(\boldsymbol{\theta}^{DK}) \end{pmatrix}, \tag{38}$$

where $i(\boldsymbol{\theta}^{NE})$ and $i(\boldsymbol{\theta}^{DK})$ are given as in Equation 23. Computing the confidence intervals based on the Wald statistic according to Equation 13 is now straightforward by setting $g(\boldsymbol{\theta}) = \pi_c^{NE} - \pi_c^{DK}$ and computing the corresponding gradient with respect to $\boldsymbol{\theta}$.

Again, the likelihood function in Equation 37 can be re-parametrized by noting that

$$\pi = \pi_c \left(1 + \gamma \frac{q(s)}{\sigma}\right)^{\frac{1}{\gamma}},$$

and if we set $\boldsymbol{\phi}^{NE} = [\lambda^{NE}, \gamma^{NE}, \sigma^{NE}]$, then the transformation $\boldsymbol{\theta} \mapsto (\pi_c^{NE}, \pi_c^{DK}, \boldsymbol{\phi}^{NE}, \boldsymbol{\phi}^{DK})$ is one-to-one. Since we're only interested in the parameters $\pi_c^{DK}$ and $\pi_c^{NE}$, we can consider the other elements of the vector $\boldsymbol{\theta}$ as nuisance parameters denoted by $\boldsymbol{\phi} = [\boldsymbol{\phi}^{NE},\ \boldsymbol{\phi}^{DK}]$. Due to orthogonality, we have the the profile likelihood for $(\pi_c^{NE},\ \pi_c^{DK})$ is given by

$$\ell_p(\pi_c^{NE}, \pi_c^{DK}) = \sup_{\boldsymbol{\phi}} \ell(\pi_c^{NE}, \pi_c^{DK}, \boldsymbol{\phi}) = \ell_p(\pi_c^{NE}) + \ell_p(\pi_c^{DK}). \tag{39}$$

As a consequence of Equation 39, the simultaneous profile-likelihood function simply amounts to finding the profile likelihood for observations in Denmark and the profile likelihood function for the observations in the Netherlands. So the optimization scheme in Equation 28 can be applied, and then the joint likelihood is simply the sum of those two functions.

By Theorem 2.7, it follows that the corresponding $1-\alpha$ confidence region for the difference $\pi_c^{NE} - \pi_c^{DK}$ is given by

$$I_{1-\alpha}^p = \left\{ \pi_c^{NE} - \pi_c^{DK} : \ \ell_p(\pi_c^{NE}, \pi_c^{DK}) \geq \ell_p(\hat{\pi}_c^{NE}, \hat{\pi}_c^{DK}) - \frac{q_{1-\alpha}}{2} \right\}, \tag{40}$$

where $q_{1-\alpha}$ is the $1-\alpha$ quantile of the $\chi_1^2$ distribution. Then, the corresponding confidence interval is given by

$$\min(I_{1-\alpha}^p) \leq \pi_c^{NE} - \pi_c^{DK} \leq \max(I_{1-\alpha}^p), \tag{41}$$

and computation of the upper bound in Equation 41 can be done through the following optimization

$$\begin{aligned} &\text{Maximize } \pi_c^{NE} - \pi_c^{DK} \\ &\text{Subject to } \ 2\left(\ell_p(\hat{\pi}_c^{NE}, \hat{\pi}_c^{DK}) - \ell_p(\pi_c^{NE}, \pi_c^{DK})\right) \leq q_{1-\alpha}. \end{aligned} \tag{42}$$

For the lower bound we simply minimize the difference under the same constraint.

Computation of the modified profile likelihood is also possible based on the likelihood in Equation 30 under the re-parametrization $\boldsymbol{\theta} \mapsto (\pi_c^{NE}, \pi_c^{DK}, \boldsymbol{\phi}^{NE}, \boldsymbol{\phi}^{DK})$. For the sake of clarity, the likelihood for a single observation from one intersection is given by

$$L^{(i)}(\pi_c, \boldsymbol{\phi}) = \begin{cases} \lambda \pi_c \left(1 + \gamma \frac{q(s)}{\sigma}\right)^{\frac{1}{\gamma}} (t_i - t_{i-1}) e^{-\lambda(t_i - t_{i-1})} \frac{1}{\sigma} \left(1 + \gamma \frac{z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}, & \text{for } s_i < u \\ \lambda(1 - \pi_c \left(1 + \gamma \frac{q(s)}{\sigma}\right)^{\frac{1}{\gamma}})(t_i - t_{i-1}) e^{-\lambda(t_i - t_{i-1})}, & \text{for } s_i \geq u. \end{cases} \tag{43}$$

We note that there will be a different number of observations from Denmark and the Netherlands. This is not an issue if we let $L^{(i)}(\pi_c^{DK}, \boldsymbol{\phi}^{DK}) = 1$ in the case where we observe no interaction. Then, the full joint likelihood is given by letting $N = \max(N_{DK}, N_{NE})$:

$$L(\pi_c^{NE}, \pi_c^{DK}, \boldsymbol{\phi}^{NE}, \boldsymbol{\phi}^{DK}) = \prod_{i=1}^{N} L^{(i)}(\pi_c^{NE}, \boldsymbol{\phi}^{NE}) L^{(i)}(\pi_c^{DK}, \boldsymbol{\phi}^{DK}). \tag{44}$$

Now, computation of the profile likelihood is straightforward as well as computation of the modified profile likelihood in Equation 9. Furthermore, the $1-\alpha$ confidence region is then given by

$$I_{1-\alpha}^{mp} = \left\{ \pi_c^{NE} - \pi_c^{DK} : \ \ell_p(\pi_c^{NE}, \pi_c^{DK}) \geq \ell_{mp}(\hat{\pi}_c^{NE}, \pi_c^{DK}) - \frac{q_{1-\alpha}}{2} \right\}. \tag{45}$$

Computation of the upper bound of the confidence interval amounts to the following optimization problem

$$\begin{aligned} &\text{Maximize } \pi_c^{NE} - \pi_c^{DK} \\ &\text{Subject to } \ 2\left(\ell_p(\hat{\pi}_c^{NE}, \hat{\pi}_c^{DK}) - \ell_p(\pi_c^{NE}, \pi_c^{DK})\right) \leq q_{1-\alpha}. \end{aligned} \tag{46}$$

Again, we use the SLSQP algorithm to optimize.

# 4 Simulation Study

The profile likelihood and the modified profile likelihood for $\lambda_c$ and $\pi_c$ introduces some additional complexity in the computation of estimates and confidence intervals, compared to the standard maximum likelihood. In order to motivate this complexity, we conduct a small simulation study to evaluate the power and confidence level of the confidence intervals described in Section 3.2 - 3.3.

We want the simulations to reflect the sampling problem of using video recordings to identify interactions between cars and cyclists, and compute the corresponding SMoS for the interaction. Since the number of interactions observed is a random process, it follows that the number of interactions is a function of the sampling time T. As before, we assume that the number of interactions is generated by a Poisson process, and we're interested in how the different confidence intervals perform for different sampling times and intensities.

The simulations are done in two steps, first we simulate the number of interactions from a Poisson process, and then from each interaction we simulate from some underlying distribution, representing the distribution of an SMoS. This means that we simulate a dataset equivalent to the datasets for Denmark and the Netherlands from the distribution defined in Equation 21. Then for each simulated dataset, we can estimate the crash intensity $\lambda_c$ and the probability of a crash $\pi_c$, as well as confidence intervals using the methods described in the previous section. Finally, we can use these simulated confidence intervals to estimate the power and the confidence level. The parameters used are given below.

In order to find the modified profile likelihood estimator, we need to know the times at which an interaction occurs $\{t_i : \ i = 1, ..., N\}$. Provided that $N(T)$ is generated by a Poisson process with intensity $\lambda$, the time between events $j_i = t_i - t_{i-1}$ is exponentially distributed with rate

$\lambda$. So, a simple algorithm to simulate from a Poisson process for a time period T, is to simulate $N+1$ exponentially distributed random variables until $\sum_{i=1}^{N+1} j_i > T$. Then set $N(T) = N$, and the jumping times is given by $t_i = j_i + t_{i-1}$, with $t_0 = 0$ [35]. We note that this is essentially equivalent to simulating from $Po(\lambda T)$. Then, we simulate $N$ random variables from some underlying distribution which represents the distribution of $S$. In this way, we have simulated a random number of interactions (or conflicts) and the outcome of those conflicts in terms of some SMoS.

We simulate $S$ from the gamma and beta distribution, shifted so that $P(S < 0) > 0$, and rescaled to reflect the actual scale in the dataset. We also vary the intensity for $N(T)$. We simulate from the following distributions:

1. $N(T) \in Po(3T),\ S = X_1 - 0.1,$ where $X_1 \in$ Gamma(3,2)
2. $N(T) \in Po(T),\ S = X_2 - 0.1,$ where $X_2 \in$ Gamma(2,2)
3. $N(T) \in Po(3T),\ S = 10(Y_1 - 0.01),$ where $Y_1 \in$ Beta$(6, 15)$
4. $N(T) \in Po(2T),\ S = 10(Y_2 - 0.01),$ where $Y_2 \in$ Beta$(2, 5)$.

By a similar argument as for the distribution of $C_f(s,t)$ in Equation 20, the corresponding crash intensity per hour is given by $\lambda_c(0) = \lambda P(S < 0)$. This implies that the true crash intensities (per hour) for the sampling distributions above are given by

1. $\lambda_c = 3P(S < 0) = 3P(X_1 < 0.1) = 6.020248 \times 10^{-5}$
2. $\lambda_c = P(S < 0) = P(X_2 < 0.1) = 1.209104 \times 10^{-3}$
3. $\lambda_c = 3P(S < 0) = 3P(Y_1 < 0.01) = 1.030925 \times 10^{-7}$
4. $\lambda_c = 2P(S < 0) = 2P(Y_2 < 0.01) = 2.920895 \times 10^{-3}$.

After simulating $N(T)$ and a random sample of $S$, we can compute confidence intervals for $\lambda_c$ and $\pi_c^{NE} - \pi_c^{DK}$ using the simulated datasets. Let $I_\lambda$ denote the confidence region for $\lambda_c$, and let $I_\pi$ denote the confidence region for $\pi_c^{NE} - \pi_c^{DK}$. Then, the confidence level $1 - \alpha$ for the confidence interval for $\lambda_c$ and $\pi_c^{NE} - \pi_c^{DK}$ is given by

$$1 - \alpha = P(\lambda_c \in I_\lambda)$$
$$1 - \alpha = P(\pi_c^{NE} - \pi_c^{DK} \in I_\pi).$$

When comparing the relative safety, we are also interested in the power of the test, and if we use the confidence intervals as our test statistic, the power is defined as

$$1 - \beta = P(0 \notin I_\pi).$$

If we simulate $n$ datasets from the some distribution 1 - 4 above, and for each dataset we compute the confidence region $I$ according to the methods in Section 3.2 - 3.3, then by the Monte Carlo

principle, we have that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left\{\pi_c^{NE}-\pi_c^{DK}\in I_\pi^i\right\}\to 1-\alpha.$$

Here $\mathbb{1}$ denotes the indicator function, and $I_\pi^i$ is the confidence interval for the $i^{th}$ simulated data set. The power can also be computed in the same way by changing the condition for which the indicator function is equal to one i.e.

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left\{0\notin I_\pi^i\right\}\to 1-\beta.$$

The results are shown in the tables below. Note that $T$ is the sampling time, and $E[N]$ is the expected number of interactions for that sampling time. For each simulation, we constructed confidence intervals using the nominal confidence level of 95%, so ideally, the actual confidence level should be the approximately 95%. Each estimate of the confidence level is based on $n = 100$ simulated datasets.

Table 3: Confidence level for interval estimates of $\lambda_c$ based on the maximum-,profile- and modified profile likelihood. The nominal confidence level is 95%.

| | E[N] | T | ML | PL | MPL |
|---|---|---|---|---|---|
| Gamma(3,2) | 72 | 24 | 0.12 | 0.68 | 0.31 |
| | 144 | 48 | 0.16 | 0.68 | 0.51 |
| | 216 | 72 | 0.20 | 0.74 | 0.62 |
| | 288 | 96 | 0.22 | 0.79 | 0.71 |
| Gamma(2,2) | 24 | 24 | 0.08 | 0.91 | 0.72 |
| | 48 | 48 | 0.12 | 0.94 | 0.88 |
| | 72 | 72 | 0.20 | 0.94 | 0.89 |
| | 96 | 96 | 0.19 | 0.95 | 0.93 |
| Beta(6,15) | 72 | 24 | 0.14 | 0.30 | 0.17 |
| | 144 | 48 | 0.15 | 0.33 | 0.19 |
| | 216 | 72 | 0.18 | 0.26 | 0.20 |
| | 288 | 96 | 0.22 | 0.30 | 0.22 |
| Beta(2,5) | 48 | 24 | 0.15 | 0.69 | 0.98 |
| | 96 | 48 | 0.27 | 0.74 | 0.94 |
| | 144 | 72 | 0.34 | 0.78 | 0.94 |
| | 192 | 96 | 0.46 | 0.80 | 0.92 |

Table 3 shows the confidence level estimates for $n = 100$ simulated datasets. In general, the significance level increases with increased sampling time, as we would expect. The confidence intervals based on the modified profile likelihood for Beta(2,5) is an exception in this regard, but

if we assume that a confidence level of 95% is already achieved for $T = 24$, then the results are not unreasonable given that they are generated by a Binomial distribution.

We also see that the profile- and modified profile likelihood achieves significantly higher confidence level than the maximum likelihood method. Also, profile likelihood method achieves a higher significance level for all distributions except for Beta(2,5), which suggests that the relative merit between the profile- and modified profile likelihood may be sensitive to the underlying distribution generating the data. Finally, confidence level tends to be higher for distributions with higher values of $\lambda_c$, like Beta(2,5) and Gamma(2,3).

Next, we evaluate the confidence level for the difference in crash probability $\pi_c^{NE} - \pi_c^{DK}$ and the power of the testing the difference as in Equation 36. We conduct the test for the difference in crash probability for the two gamma distributions, and the two beta distributions. For the gamma distributions we set

$$\pi_c^{NE} - \pi_c^{DK} = P(X_2 < 0.1) - P(X_1 < 0.1) = 0.001189037$$

And for the beta distributions we set

$$\pi_c^{NE} - \pi_c^{DK} = P(Y_2 < 0.01) - P(Y_1 < 0.01) = 0.001460413$$

The results are shown in Table 4 and 5. Again, we see that the confidence intervals based on the maximum likelihood perform very poorly both in terms of confidence level and statistical power. In terms of confidence level, both the profile- and modified profile likelihood constitute big improvements compared to the maximum likelihood method, but the profile likelihood clearly has a higher confidence level while the modified profile likelihood has a higher power. Consequently, the profile likelihood based confidence intervals have a high probability of containing the true difference in crash probability, but it has a hard time rejecting the null hypothesis that the crash probabilities are equal. In contrast, the modified profile likelihood has a relatively low probability of covering the true difference, but it is able to identify reject the false null hypothesis in more cases.

Table 4: Estimates of the statistical confidence level of the hypothesis test in Equation 36 using confidence intervals based on ML, PL and MPL as test statistics. The nominal confidence level is 95%.

| | T | ML | PL | MPL |
|---|---|---|---|---|
| Gamma | 24 | 0.01 | 0.98 | 0.74 |
| | 48 | 0.03 | 0.94 | 0.53 |
| | 72 | 0.03 | 0.92 | 0.64 |
| | 96 | 0.09 | 0.96 | 0.68 |
| Beta | 24 | 0.03 | 0.98 | 0.51 |
| | 48 | 0.06 | 0.99 | 0.64 |
| | 72 | 0.08 | 0.99 | 0.71 |
| | 96 | 0.05 | 0.98 | 0.82 |

Table 5: Estimates of the statistical power of the hypothesis test in Equation 36 using the confidence intervals based on ML, PL and MPL as test statistics. The nominal confidence level is 95%.

| | T | ML | PL | MPL |
|---|---|---|---|---|
| Gamma | 24 | 0 | 0.02 | 0.23 |
| | 48 | 0 | 0.05 | 0.33 |
| | 72 | 0 | 0.11 | 0.37 |
| | 96 | 0.01 | 0.06 | 0.47 |
| Beta | 24 | 0 | 0.01 | 0.18 |
| | 48 | 0 | 0.02 | 0.31 |
| | 72 | 0 | 0.02 | 0.37 |
| | 96 | 0 | 0.03 | 0.47 |

Based on these results, it may be prudent to use the profile likelihood method when computing the confidence estimate of $\lambda_c$, since it has a much higher probability of covering the true value. In terms of testing the difference in crash probability, the modified profile likelihood will be useful since rejecting the null hypothesis is easier.

# 5 Results

In this section, we apply the methods from Section 3 to the six datasets for all measures. Since Denmark 1 (Dk1) and Denmark 3 (Dk3) $\text{TTC}_{min}$ had very few observations, these were discarded.

Moreover, in order to make the results presentable we only plotted the results for Dk2 and Ne2. The corresponding plots for the other intersections can be found in Appendix C.

## 5.1 Threshold Selection

A combination of approaches has been used in order to find suitable thresholds for the six data sets. Threshold range plots and mean residual life plots has been used to find where $\tilde{S}$ has converged to GPD, and where the parameters $\sigma$ and $\gamma$ are stable. We also considered choosing a threshold such that the endpoint for the distribution of the SMoS $S \leq 0$ such that 0 is in the support, as we are interested in the probability of a crash (when SMoS=0). Finally, the model diagnostics for a chosen threshold $u$ must suggest that the model of choice is a good fit.

The plots of the threshold ranges are seen in Figure 8 - 13. The thresholds are given in terms of the negated data, which corresponds to $\tilde{u}$ from Section 3. For Denmark 2, the highest threshold we can select, where the confidence intervals are not too large, for $\text{TTC}_{min}$ seems to be around $-3$. For $\text{T2}_{min}$, it is $-0.75$, and for PET it seems to be just below $-1$. For the Netherlands 2, these thresholds seem to be around $-2$ for $\text{TTC}_{min}$, and around $-1.5$ for both $\text{T2}_{min}$ and PET.

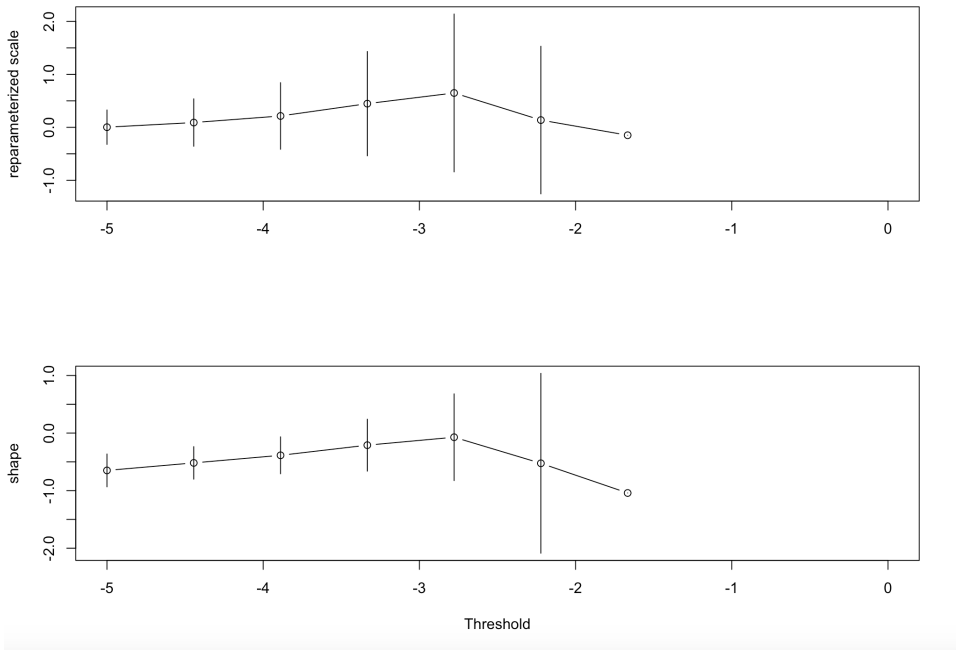


Figure 8: Threshold range plot of $\text{TTC}_{min}$ of Denmark 2, where a General Pareto model has been fit to a sequence of thresholds with 95% confidence intervals

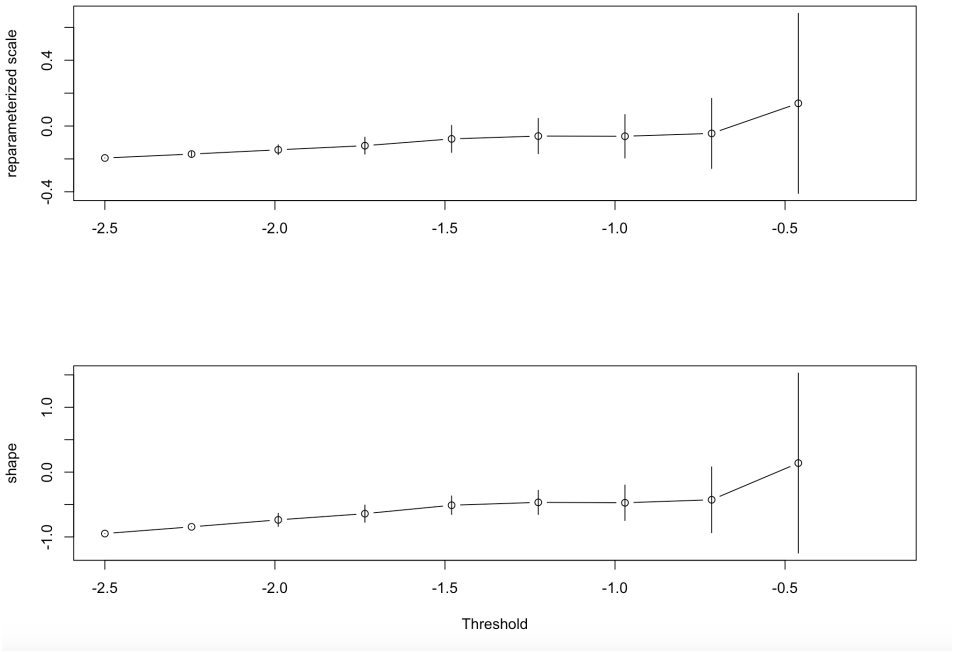


Figure 9: Threshold range plot of $T2_{min}$ of Denmark 2, where a General Pareto model has been fit to a sequence of thresholds with 95% confidence intervals

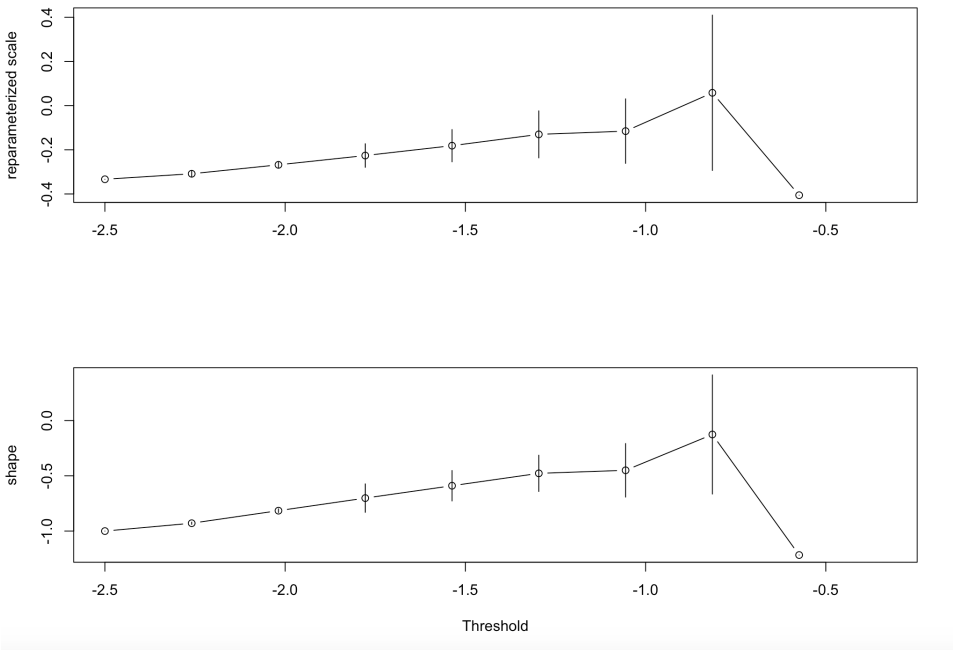


Figure 10: Threshold range plot of PET of Denmark 2, where a General Pareto model has been fit to a sequence of thresholds with 95% confidence intervals

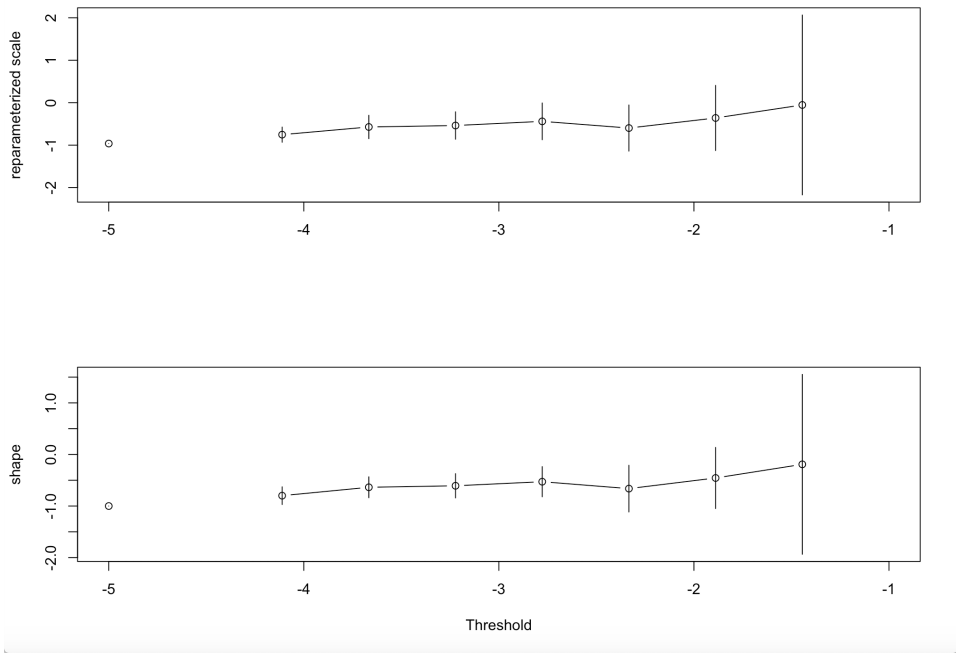


Figure 11: Threshold range plot of $\text{TTC}_{min}$ of the Netherlands 2, where a General Pareto model has been fit to a sequence of thresholds with 95% confidence intervals

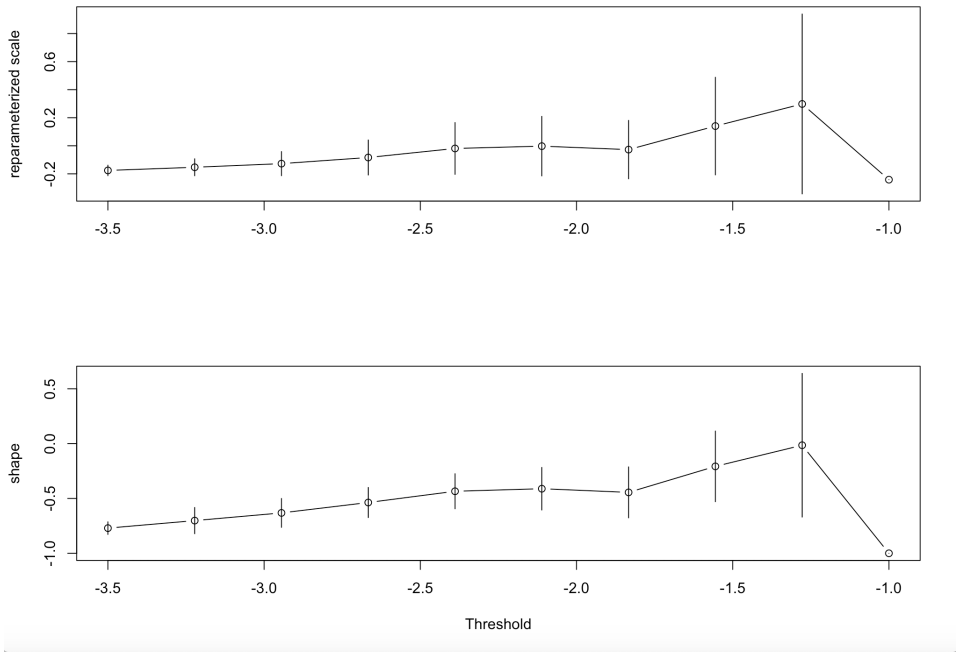


Figure 12: Threshold range plot of $\text{T2}_{min}$ of the Netherlands 2, where a General Pareto model has been fit to a sequence of thresholds with 95% confidence intervals

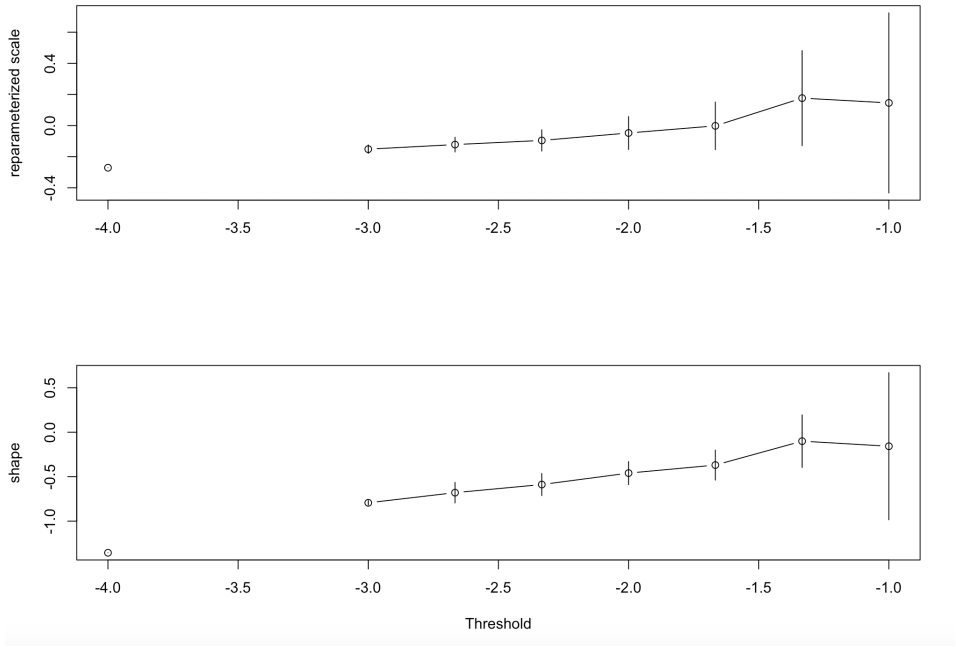


Figure 13: Threshold range plot of PET of the Netherlands 2, where a General Pareto model has been fit to a sequence of thresholds with 95% confidence intervals

The mean residual life plots are seen in Figure 14 - 16, in which we want to find where linearity occurs. For $\text{TTC}_{min}$, Denmark 2, the graph has a curvature at around $-4$, suggesting that $\tilde{S}$ has not converged to GPD before this value, thus, the threshold must be at least $-4$. There is some linearity between $-4$ and just before $-2$, with a spike at $-2$, suggesting this value might be an appropriate threshold.

For the Netherlands 2, there is a curvature at $-3$ with a plateau between the thresholds $-3$ and $-2.5$, after which there is a slight negative linear trend, suggesting $-2.5$ is an appropriate threshold. For $\text{T2}_{min}$, $-1.4$ and $-2$ seem to be appropriate thresholds for Denmark 2 and the Netherlands 2 respectively. A suitable threshold for Denmark 2, PET, is difficult to find. There seems to be a linear trend for all thresholds, but as the observed value closest to 0 for this data set and measure is 0.3 (Table 2), a threshold of around $-1.4$ might be suitable. For the same measure for the Netherlands 2, a threshold of $-1.8$ seems appropriate.

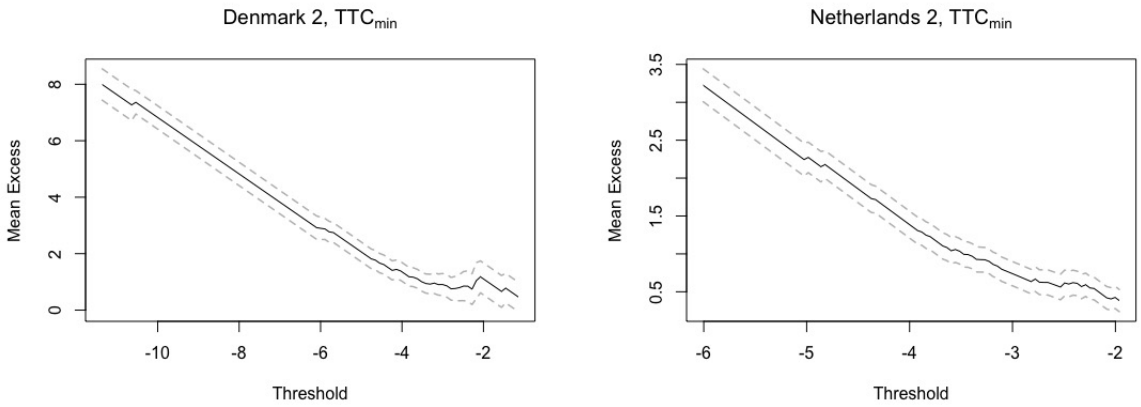


Figure 14: Mean residual life plot for Denmark 2 and the Netherlands 2 for $\text{TTC}_{min}$ with 95% confidence intervals

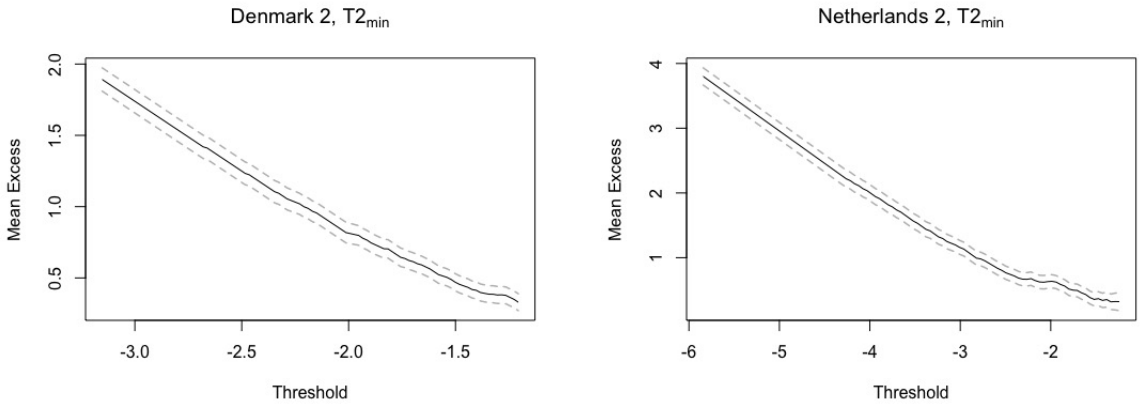


Figure 15: Mean residual life plot for Denmark 2 and the Netherlands 2 for $\text{T2}_{min}$ with 95% confidence intervals

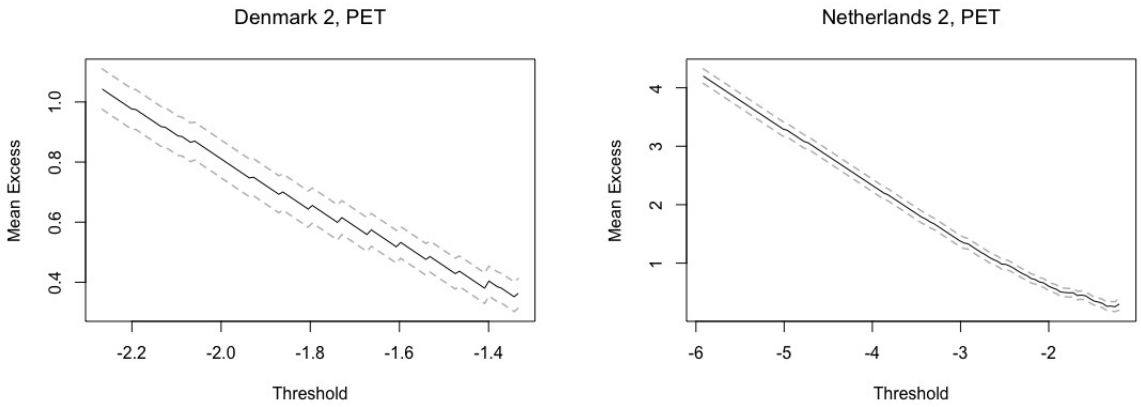


Figure 16: Mean residual life plot for Denmark 2 and the Netherlands 2 for PET with 95% confidence intervals

As we are interested in estimating the crash frequency (corresponding to $S \leq 0$), we also want to consider choosing a $u$ such that the lower endpoint for the estimated tail distribution of $S$ includes 0, as well as a low enough threshold $\tilde{u}$ in order to have an appropriate number of observations for the analysis. After some testing, and considering the threshold plots and mean residual life plots, the resulting thresholds and maximum likelihood estimates with corresponding standard errors for the GPD parameters can be seen in Table 6.

For many of the estimates, the shape parameter $\gamma$ is not significantly different from zero. In many applications this would indicate that we need to consider the GPD corresponding to $\gamma = 0$ with distribution function given in Equation 4. However, applying such a model to would imply that $S$ does not have a lower bound. In that sense, the distribution in Equation 4 would be a poor representation $S$.

Table 6: Thresholds, number of exceedances above thresholds, GPD parameter estimates and their standard errors for all intersections and measures

| Data | Measure | $u$ | $n$ | Scale ($\sigma$) | Shape ($\gamma$) | s.e. Scale | s.e. Shape |
|---|---|---|---|---|---|---|---|
| Denmark 1 | $T2_{min}$ | 2 | 52 | 0.9937 | −0.4613 | 0.1615 | 0.1062 |
| | PET | 1.5 | 30 | 0.5172 | −0.3151 | 0.1175 | 0.1469 |
| Denmark 2 | $TTC_{min}$ | 3 | 16 | 1.1347 | −0.2456 | 0.4267 | 0.2876 |
| | $T2_{min}$ | 0.75 | 22 | 0.2173 | −0.2677 | 0.0707 | 0.2522 |
| | PET | 0.88 | 30 | 0.1975 | −0.2101 | 0.0544 | 0.2099 |
| Denmark 3 | $T2_{min}$ | 1.5 | 63 | 0.4817 | −0.2933 | 0.0700 | 0.0814 |
| | PET | 0.88 | 12 | 0.6710 | −0.6155 | 0.0827 | 0.0857 |
| Netherlands 1 | $TTC_{min}$ | 2.5 | 66 | 0.5025 | −0.0735 | 0.0776 | 0.0937 |
| | $T2_{min}$ | 0.74 | 18 | 0.2057 | −0.2237 | 0.0974 | 0.4137 |
| | PET | 1.07 | 67 | 0.2730 | −0.2425 | 0.0448 | 0.1134 |
| Netherlands 2 | $TTC_{min}$ | 1.693 | 9 | 0.2734 | −0.1574 | 0.1502 | 0.4398 |
| | $T2_{min}$ | 2.3 | 92 | 0.9402 | −0.4085 | 0.1217 | 0.0878 |
| | PET | 1.4 | 50 | 0.3850 | −0.2047 | 0.0713 | 0.1228 |
| Netherlands 3 | $TTC_{min}$ | 2.597 | 11 | 0.1592 | −0.0512 | 0.0733 | 0.3484 |
| | $T2_{min}$ | 0.96 | 10 | 0.9402 | −0.4085 | 0.1217 | 0.0878 |
| | PET | 0.88 | 10 | 0.1694 | −0.1532 | 0.0776 | 0.3348 |

Model diagnostics of the generalized Pareto distributions for Denmark 2 and the Netherlands 2 for all measures are seen in Figure 17 - 19. For all models, the empirical and model density seem to be fairly consistent, suggesting the models, thus the thresholds, are appropriate. In the probability- and quantile plots, there are some curvatures for all measures, however, the points follow the line fairly well. Overall, the diagnostics seem to suggest the chosen models are appropriate.

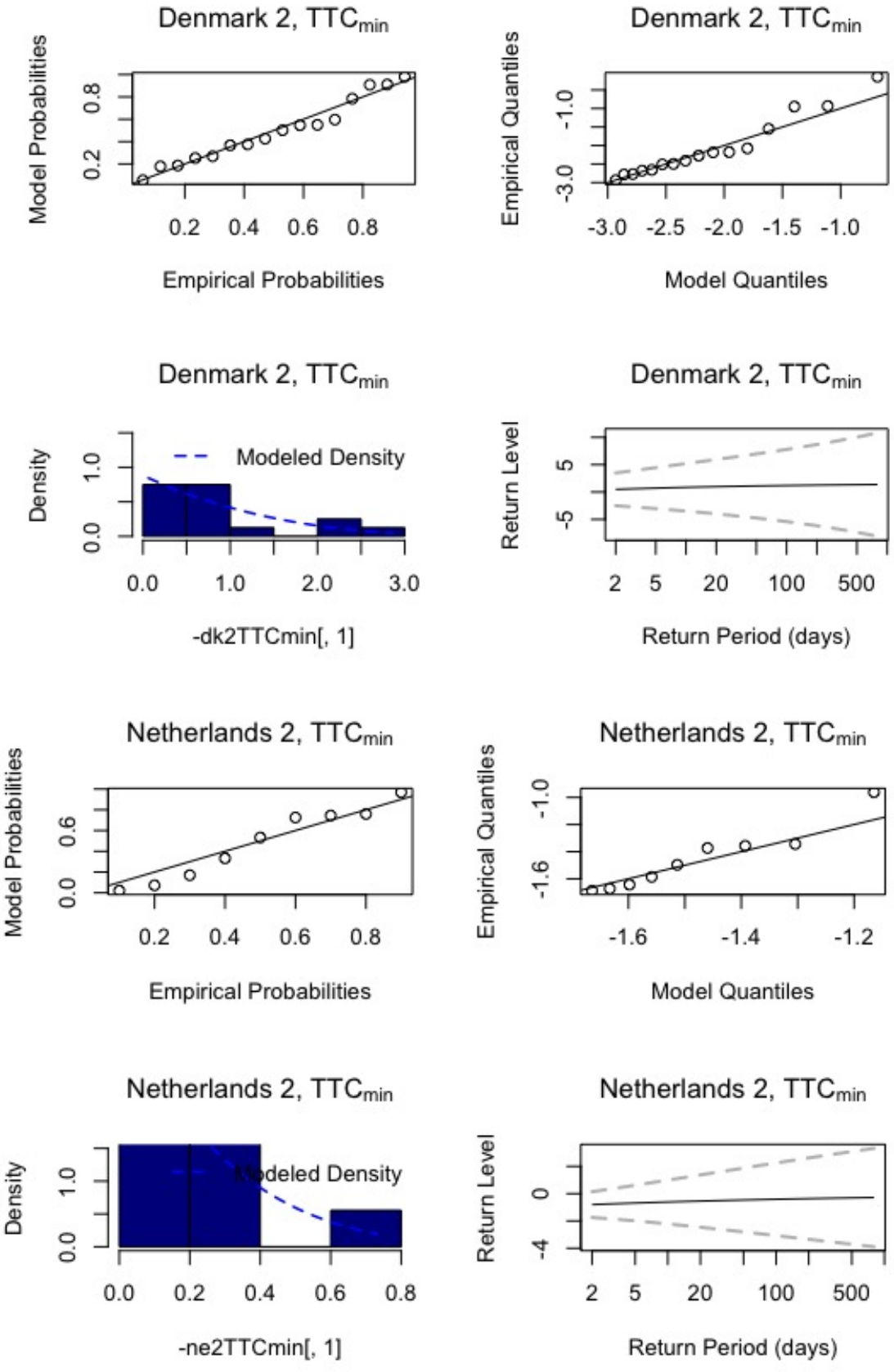


Figure 17: Model diagnostics for the estimated GP distributions for Denmark 2 and the Netherlands 2 for $\text{TTC}_{min}$

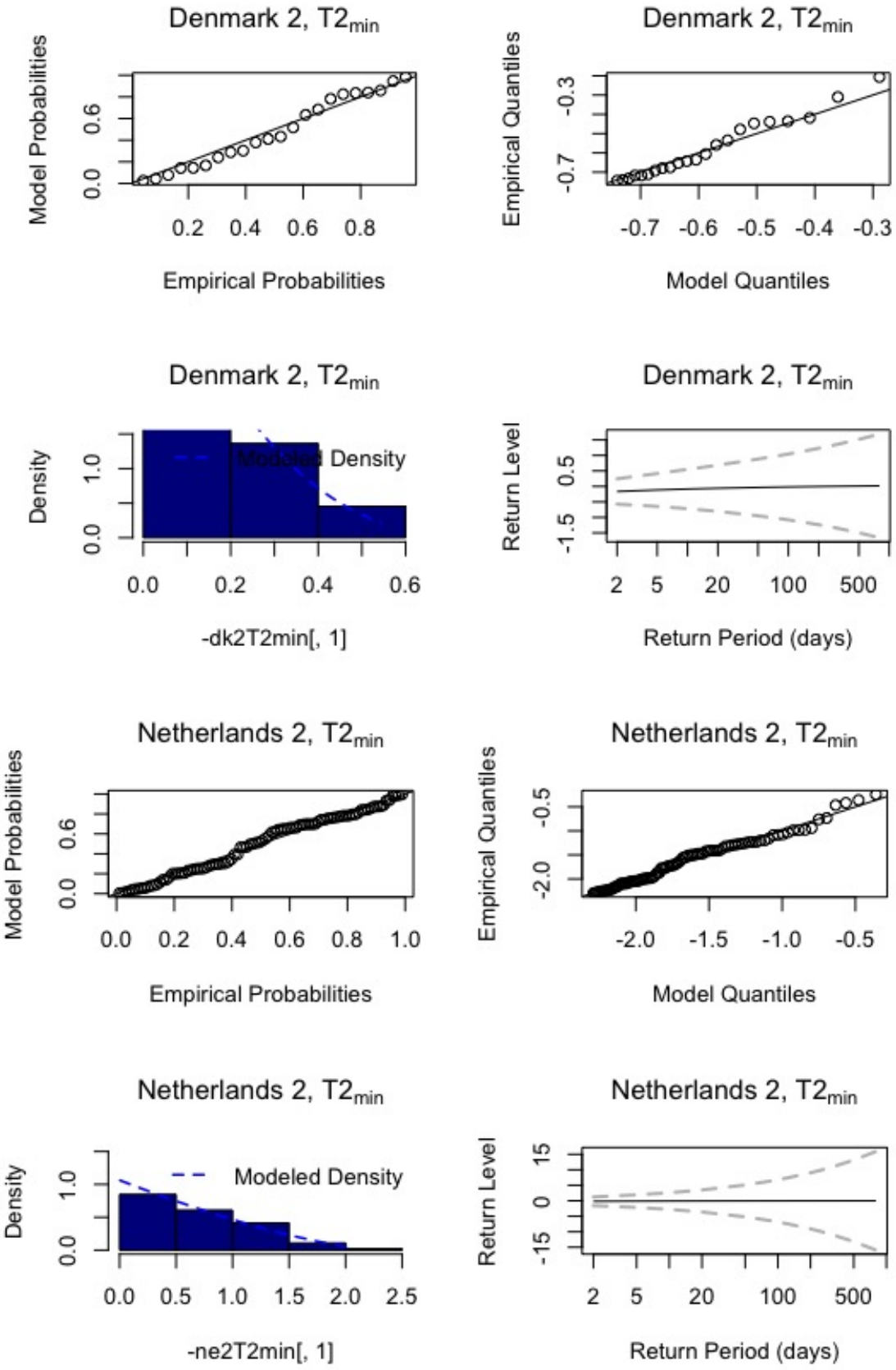


Figure 18: Model diagnostics for the estimated GP distributions for Denmark 2 and the Netherlands 2 for $T2_{min}$

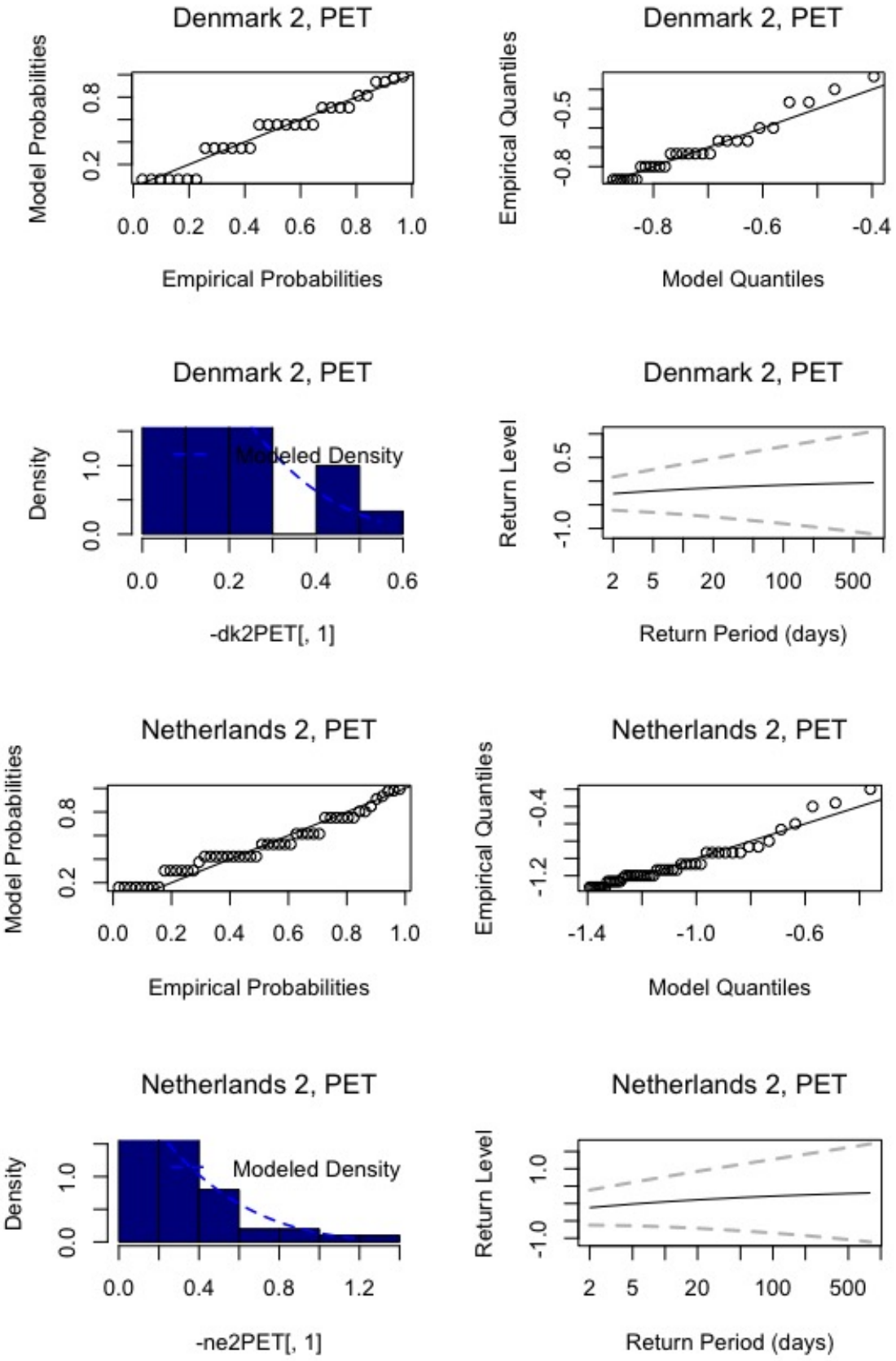


Figure 19: Model diagnostics for the estimated GP distributions for Denmark 2 and the Netherlands 2 for PET

## 5.2 Estimates of Near-Crash Intensities

In this section, we present the estimates of $\lambda_c(s)$ using the maximum- , profile- and modified profile likelihood method. We also plot the corresponding confidence intervals and the empirical counts for reference. The empirical counts are defined as the number of observations with a measure (SMoS) lower than $s$.

Comparing the estimates with the empirical counts is a good way to evaluate the model fit $C_f(s,t)$, however, it says very little about the accuracy of the model when extrapolating outside the range of the dataset. With the estimates of the crash intensity $\lambda_c(0)$ we can still apply some common sense judgement, in particular when the estimate is very high it is obvious that the estimate is nonsensical.

### 5.2.1 Maximum Likelihood

The maximum likelihood estimates of $\hat{\lambda}_c(s)\cdot t$ with 95% confidence intervals for the three measures for Denmark 2 and the Netherlands 2 are seen in Figure 20 - 22. The empirical counts are also plotted as points. We see that the the estimates fit the data well in the sense that empirical counts appear to be randomly dispersed about the estimate. However, we also see that the lower confidence bound are less than zero for certain values of $s$, and since $\lambda_c(s) > 0$ this is unreasonable. For this reason, we need to consider other methods for computing interval estimates of $\lambda_c$.

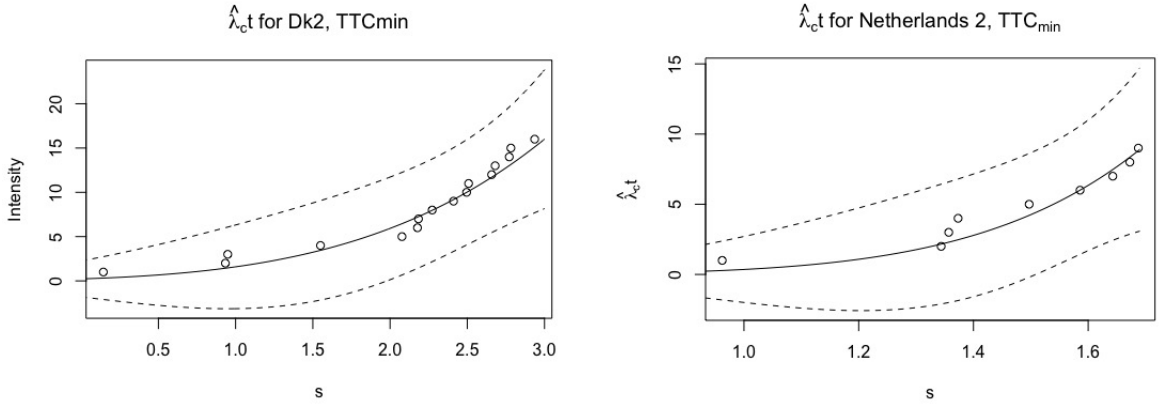


Figure 20: Maximum likelihood estimates of crash intensity for $\text{TTC}_{min}$ for Denmark 2 and the Netherlands 2 with 95% confidence intervals

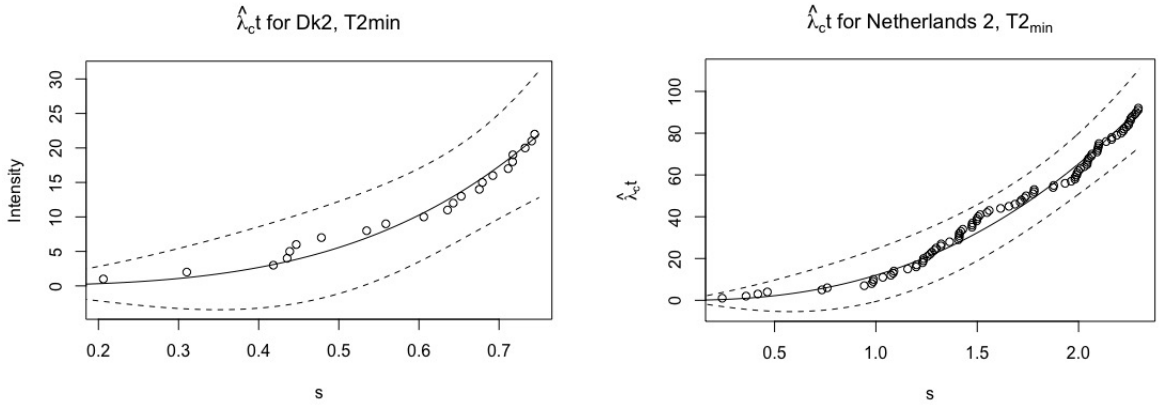


Figure 21: Maximum likelihood estimates of crash intensity for $T2_{min}$ for Denmark 2 and the Netherlands 2 with 95% confidence intervals

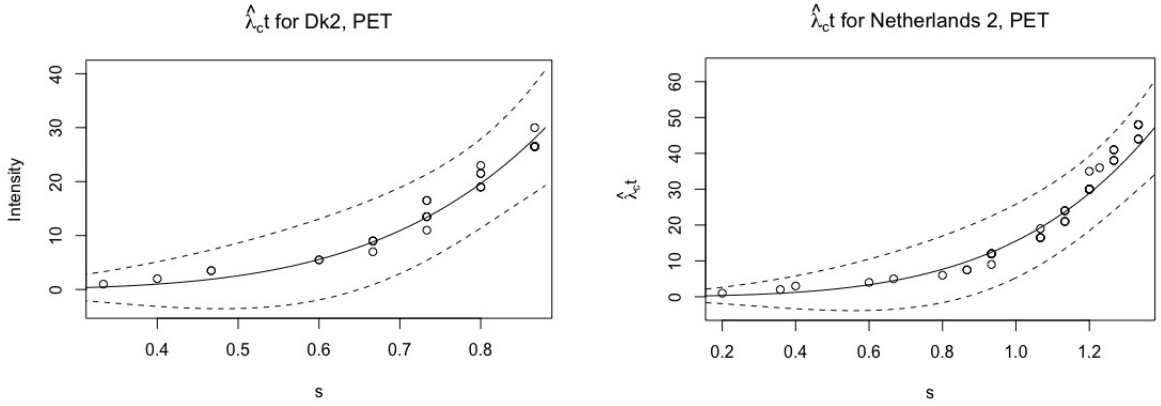


Figure 22: Maximum likelihood estimates of crash intensity for PET for Denmark 2 and the Netherlands 2 with 95% confidence intervals

In Table 7 are the maximum likelihood estimates of the near-crash intensities per hour with 95% confidence intervals for selected values of the measures (0, 0.3 and 0.6), for all data sets and measures. As can be seen from the table, the (near) crash intensity $\hat{\lambda}_c$ can vary significantly for different measures. We also see that the crash intensity (i.e. for $s < 0$) $\hat{\lambda}_c(0)$ can be quite high. For example, for Dk1 the hourly crash intensity estimated from $T2_{min}$ is $\hat{\lambda}_c(0) = 0.0018$. Consequently, the yearly crash intensity would therefore be $\hat{\lambda}_c(0) \times 24 \times 365 = 15.768$. We do not have access to any crash data, but this seems to be a serious overestimate of the number of collisions per year. However, we also get very low estimates for e.g. Ne1 PET, we have a yearly intensity of $\hat{\lambda}_c(0) \times 24 \times 365 = 0.09636$. This implies that we would expect about one crash every 10 years between right turning vehicles and cyclist.

We also note that the confidence intervals based on the delta method are quite wide, meaning

that even if we get a reasonable point estimate of $\lambda_c$, we may still obtain unreasonable interval estimates. So, for Ne1 PET, the point estimate of the crash intensity seems reasonable, but the upper confidence bound is equal to 0.0008, which would imply an expected number of 7 collisions per year, which is indeed quite high.

### 5.2.2 Profile Likelihood

The profile likelihood estimates of $\lambda_c(s)t$ with a 95% confidence interval for the three measures for Denmark 2 and the Netherlands 2 are seen in Figure 23 - 25. In this case, the lower confidence bounds are all greater than zero. The point estimates of the near-crash intensity are identical with maximum likelihood estimates, but obviously the confidence bounds are different.

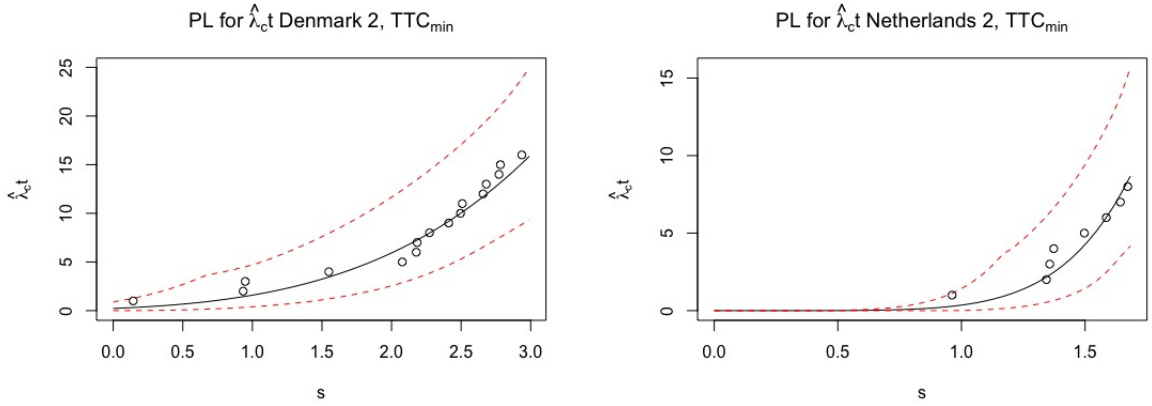


Figure 23: Profile likelihood estimates of crash intensity for $TTC_{min}$ for Denmark 2 and the Netherlands 2 with 95% confidence intervals

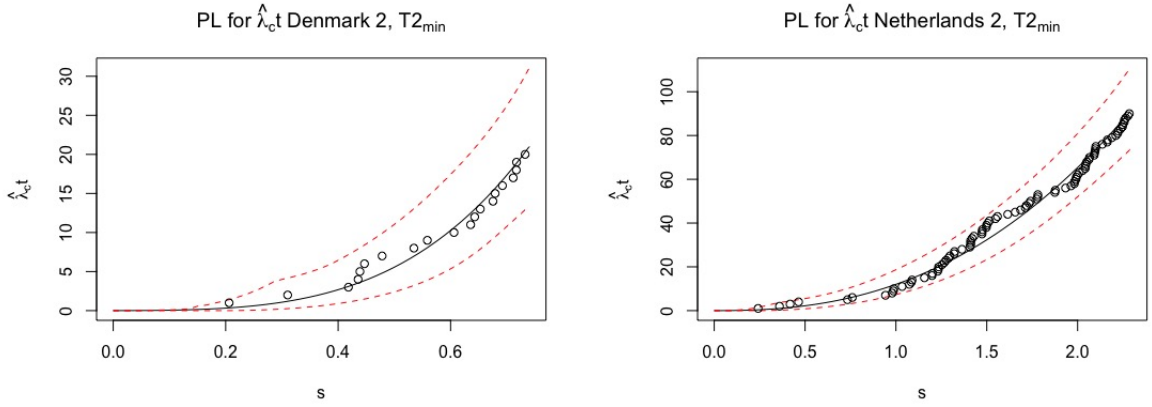


Figure 24: Profile likelihood estimates of crash intensity for $T2_{min}$ for Denmark 2 and the Netherlands 2 with 95% confidence intervals

Table 7: Maximum likelihood estimates of the crash intensities per hour with 95% confidence intervals for measures below 0, 0.3 and 0.6 for all data sets and measures

| | | Measure $\leq 0$ | | | Measure $\leq 0.3$ | | | Measure $\leq 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Data | Measure | lower | $\hat{\lambda}_c$ | upper | lower | $\hat{\lambda}_c$ | upper | lower | $\hat{\lambda}_c$ | upper |
| Dk 1 | $T2_{min}$ | $-0.02479$ | 0.0018 | 0.0284 | $-0.0514$ | 0.0185 | 0.0884 | $-0.0389$ | 0.0557 | 0.1502 |
| | PET | $-0.0042$ | 0.0001 | 0.0044 | $-0.0282$ | 0.0048 | 0.0379 | $-0.0396$ | 0.0251 | 0.0898 |
| Dk 2 | $TTC_{min}$ | $-0.0751$ | 0.0094 | 0.0938 | $-0.1006$ | 0.0187 | 0.1379 | $-0.1209$ | 0.0337 | 0.1884 |
| | $T2_{min}$ | $-0.0037$ | $6.1 \cdot 10^{-5}$ | 0.0038 | $-1.1364$ | 0.0448 | 0.2260 | 0.1456 | 0.4273 | 0.7091 |
| | PET | $-0.0003$ | $2.5 \cdot 10^{-6}$ | 0.0003 | $-0.0808$ | 0.0130 | 0.1067 | $-0.0803$ | 0.2322 | 0.5447 |
| Dk 3 | $T2_{min}$ | $-0.0078$ | 0.0003 | 0.0085 | $-0.0650$ | 0.0150 | 0.0950 | $-0.0873$ | 0.0876 | 0.2624 |
| | PET | $-0.0001$ | $1.6 \cdot 10^{-6}$ | 0.0001 | $-0.0084$ | 0.0005 | 0.0094 | $-0.0546$ | 0.0199 | 0.0943 |
| Ne1 | $TTC_{min}$ | $-0.0346$ | 0.0056 | 0.0457 | $-0.0604$ | 0.0140 | 0.0883 | $-0.0951$ | 0.0328 | 0.1606 |
| | $T2_{min}$ | $-0.0132$ | 0.0005 | 0.0142 | $-0.1260$ | 0.0409 | 0.2077 | 0.1014 | 0.3585 | 0.6155 |
| | PET | $-0.0008$ | $1.1 \cdot 10^{-5}$ | 0.0008 | $-0.1082$ | 0.0241 | 0.1564 | $-0.1118$ | 0.3005 | 0.7128 |
| Ne2 | $TTC_{min}$ | $-2.6 \cdot 10^{-8}$ | $2.8 \cdot 10^{-11}$ | $2.6 \cdot 10^{-8}$ | $-0.0009$ | $1.3 \cdot 10^{-5}$ | 0.0009 | $-0.0154$ | 0.0007 | 0.0168 |
| | $T2_{min}$ | $-0.0001$ | $4.9 \cdot 10^{-8}$ | 0.0001 | $-0.1536$ | 0.0264 | 0.2065 | $-0.2253$ | 0.1435 | 0.5123 |
| | PET | $-0.0305$ | 0.0027 | 0.0358 | $-0.1116$ | 0.0284 | 0.1684 | $-0.1592$ | 0.1391 | 0.4375 |
| Ne3 | $TTC_{min}$ | $-1.8 \cdot 10^{-13}$ | $2.4 \cdot 10^{-16}$ | $1.8 \cdot 10^{-13}$ | $-6.8 \cdot 10^{-10}$ | $1.9 \cdot 10^{-12}$ | $6.8 \cdot 10^{-10}$ | $-1.6 \cdot 10^{-7}$ | $8.9 \cdot 10^{-10}$ | $1.7 \cdot 10^{-7}$ |
| | $T2_{min}$ | $-0.0002$ | $1.6 \cdot 10^{-6}$ | 0.0002 | $-0.0178$ | 0.0009 | 0.0197 | $-0.1011$ | 0.0279 | 0.1570 |
| | PET | $-0.0009$ | $1.3 \cdot 10^{-5}$ | 0.0009 | $-0.0392$ | 0.0033 | 0.0458 | $-0.1085$ | 0.0620 | 0.2324 |

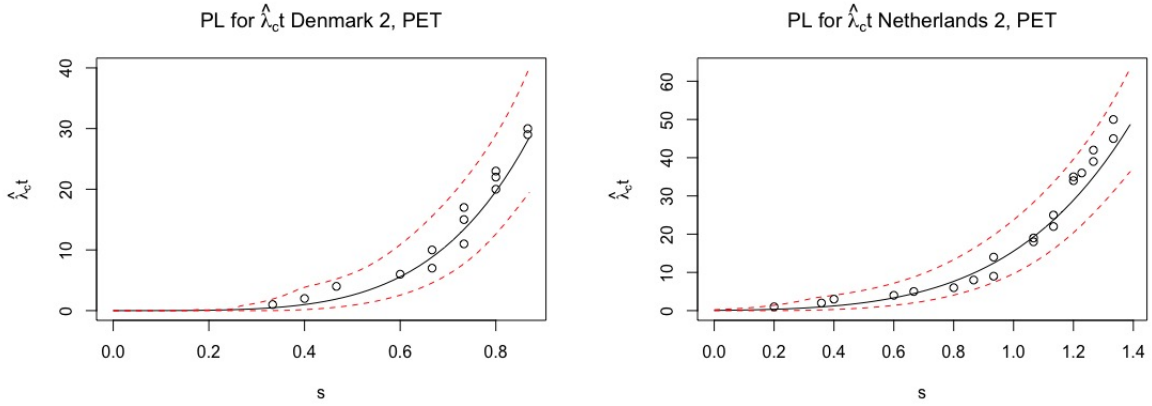


Figure 25: Profile likelihood estimates of crash intensity for PET for Denmark 2 and the Netherlands 2 with 95% confidence intervals

Table 8 shows the profile likelihood estimates of $\lambda_c(s)$ for selected values of the measure $s$, along with 95% confidence intervals. We see that these confidence intervals are generally thinner than those based on the maximum likelihood method. This means that we obtain more reasonable interval estimates for $\lambda_c$. We noted earlier that the upper bound of the confidence interval for $\lambda_c(0)$ based on the delta method for for Ne1 PET would imply a yearly crash intensity of 7 collisions per year. From Table 8, we see that the corresponding upper bound for the hourly based crash intensity on the profile likelihood is $1.3 \times 10^{-5}$, which implies a yearly crash intensity of 0.11388. So, we would expect one crash every 10 years, which seems more realistic than the corresponding upper bound using the delta method. It is interesting also, that in general the confidence intervals for $\lambda_c(0)$ based on the PET tends to have quite reasonable upper bounds, and in many cases they produce the lowest upper bounds of all measures. One exception to this rule is the crash frequency for Ne2.

### 5.2.3 Modified Profile Likelihood

The modified profile likelihood estimates of $\lambda_c(s)t$ with a 95% confidence interval for the three measures for Denmark 2 and the Netherlands 2 are seen in Figure 26 - 28. Furthermore, the modified profile likelihood for $s < 0$, $s < 0.3$ and $s < 0.6$ are given in Table 9.

Looking at the point estimates, we see that in general, for $\hat{\lambda}_c(0)$, the MPL estimates seems to be systematically higher than the maximum likelihood estimates (and therefore also the profile likelihood estimates).

The confidence intervals for the modified profile likelihood are generally slim compared to the delta method, and compared the profile likelihood the relative width are mixed. However, if we look at the lower bounds for the intervals, we can see that the modified profile likelihood creates

Table 8: Profile likelihood estimates of the crash intensities with 95% confidence intervals for measures below 0, 0.3 and 0.6 for all data sets and measures

| | | Measure $\leq 0$ | | | Measure $\leq 0.3$ | | | Measure $\leq 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Data | Measure | lower | $\hat{\lambda}_c$ | upper | lower | $\hat{\lambda}_c$ | upper | lower | $\hat{\lambda}_c$ | upper |
| Dk1 | $T2_{min}$ | $1.91 \cdot 10^{-7}$ | 0.0018 | 0.0071 | 0.0063 | 0.0185 | 0.0490 | 0.0269 | 0.0557 | 0.1011 |
| | PET | $6.0 \cdot 10^{-9}$ | 0.0001 | 0.0005 | 0.0005 | 0.0048 | 0.0194 | 0.0086 | 0.0251 | 0.0609 |
| Dk2 | $TTC_{min}$ | $1.4 \cdot 10^{-7}$ | 0.0094 | 0.0374 | 0.0010 | 0.0187 | 0.0747 | 0.0049 | 0.0337 | 0.1349 |
| | $T2_{min}$ | $1.4 \cdot 10^{-8}$ | $6.1 \cdot 10^{-5}$ | $7.2 \cdot 10^{-5}$ | 0.0087 | 0.0448 | 0.1667 | 0.2227 | 0.4273 | 0.7263 |
| | PET | $1.2 \cdot 10^{-9}$ | $2.5 \cdot 10^{-6}$ | $3.0 \cdot 10^{-6}$ | $2.0 \cdot 10^{-6}$ | 0.0130 | 0.0363 | 0.1062 | 0.2322 | 0.4525 |
| Dk3 | $T2_{min}$ | $1.5 \cdot 10^{-8}$ | 0.0003 | 0.0005 | 0.0032 | 0.0150 | 0.0600 | 0.0405 | 0.0876 | 0.1718 |
| | PET | $1.2 \cdot 10^{-12}$ | $1.6 \cdot 10^{-6}$ | $2.1 \cdot 10^{-6}$ | $7.5 \cdot 10^{-9}$ | 0.0005 | 0.0013 | 0.0029 | 0.0199 | 0.0795 |
| Ne1 | $TTC_{min}$ | $7.4 \cdot 10^{-5}$ | 0.0056 | 0.0225 | 0.0011 | 0.0140 | 0.0558 | 0.0064 | 0.0328 | 0.1250 |
| | $T2_{min}$ | $1.6 \cdot 10^{-7}$ | 0.0005 | 0.0006 | 0.0063 | 0.0409 | 0.1634 | 0.1593 | 0.3585 | 0.6481 |
| | PET | $1.1 \cdot 10^{-6}$ | $1.1 \cdot 10^{-5}$ | $1.3 \cdot 10^{-5}$ | 0.0018 | 0.0241 | 0.0947 | 0.1611 | 0.3005 | 0.5204 |
| Ne2 | $TTC_{min}$ | $2.7 \cdot 10^{-11}$ | $2.8 \cdot 10^{-11}$ | $5.8 \cdot 10^{-11}$ | $8.2 \cdot 10^{-11}$ | $1.3 \cdot 10^{-5}$ | $3.3 \cdot 10^{-5}$ | $6.5 \cdot 10^{-11}$ | 0.0007 | 0.0027 |
| | $T2_{min}$ | $4.1 \cdot 10^{-8}$ | $4.8 \cdot 10^{-8}$ | $5.8 \cdot 10^{-8}$ | 0.0029 | 0.0264 | 0.1058 | 0.0648 | 0.1435 | 0.2984 |
| | PET | $4.6 \cdot 10^{-8}$ | 0.0027 | 0.0094 | 0.0035 | 0.0284 | 0.1137 | 0.0574 | 0.1391 | 0.2986 |
| Ne3 | $TTC_{min}$ | $2.3 \cdot 10^{-16}$ | $2.4 \cdot 10^{-16}$ | $9.6 \cdot 10^{-16}$ | $1.3 \cdot 10^{-13}$ | $1.9 \cdot 10^{-12}$ | $5.3 \cdot 10^{-12}$ | $3.5 \cdot 10^{-14}$ | $8.8 \cdot 10^{-10}$ | $2.6 \cdot 10^{-9}$ |
| | $T2_{min}$ | $1.4 \cdot 10^{-11}$ | $1.6 \cdot 10^{-6}$ | $2.0 \cdot 10^{-6}$ | $4.9 \cdot 10^{-9}$ | 0.0009 | 0.0024 | 0.0032 | 0.0279 | 0.1118 |
| | PET | $4.2 \cdot 10^{-22}$ | $1.3 \cdot 10^{-5}$ | $1.6 \cdot 10^{-5}$ | $7.0 \cdot 10^{-8}$ | 0.0033 | 0.0101 | 0.0130 | 0.0620 | 0.1973 |

some problems that were not present for the profile likelihood. For example, the lower bound for $\lambda_c(0)$ based on MPL for Dk1 $T2_{min}$ is 0.0009, which would imply a yearly crash frequency of $0.0009 \times 24 \times 365 = 7.884$. For comparison, the corresponding bound for using the profile likelihood is $1.91 \times 10^{-7} \times 24 \times 365 = 1.67 \times 10^{-3}$. We can see the same problem with the lower bounds of MPL confidence intervals for other intersections and measures like Dk2 $\text{TTC}_{min}$, Dk3 $T2_{min}$ and Ne2 PET.

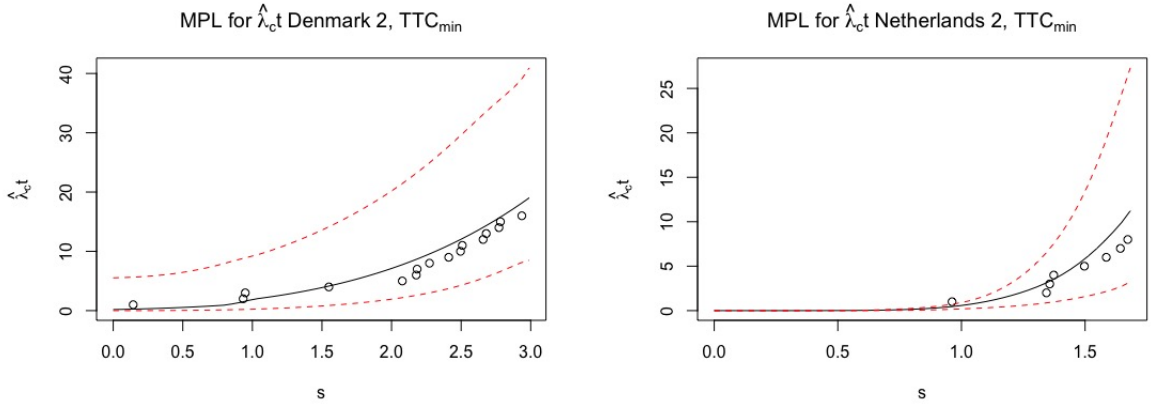


Figure 26: Modified profile likelihood estimates of crash intensity for $\text{TTC}_{min}$ for Denmark 2 and the Netherlands 2 with 95% confidence intervals

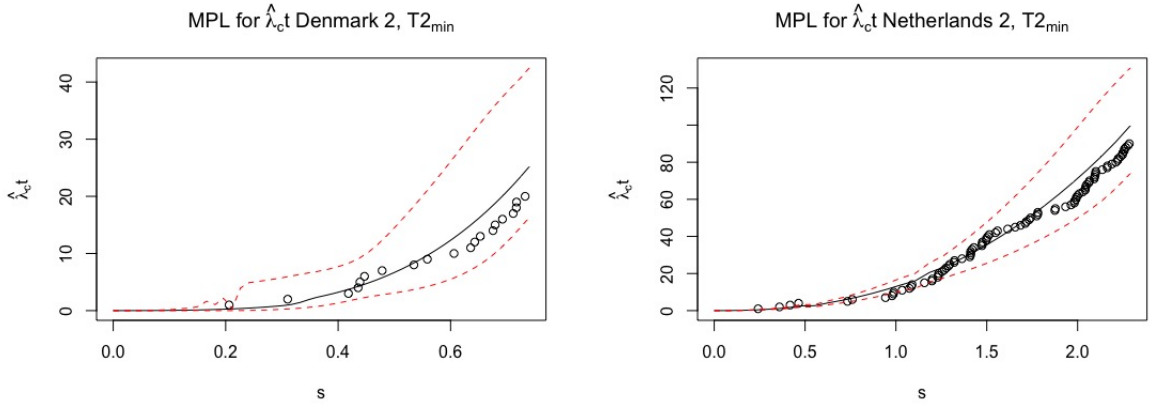


Figure 27: Modified profile likelihood estimates of crash intensity for $T2_{min}$ for Denmark 2 and the Netherlands 2 with 95% confidence intervals

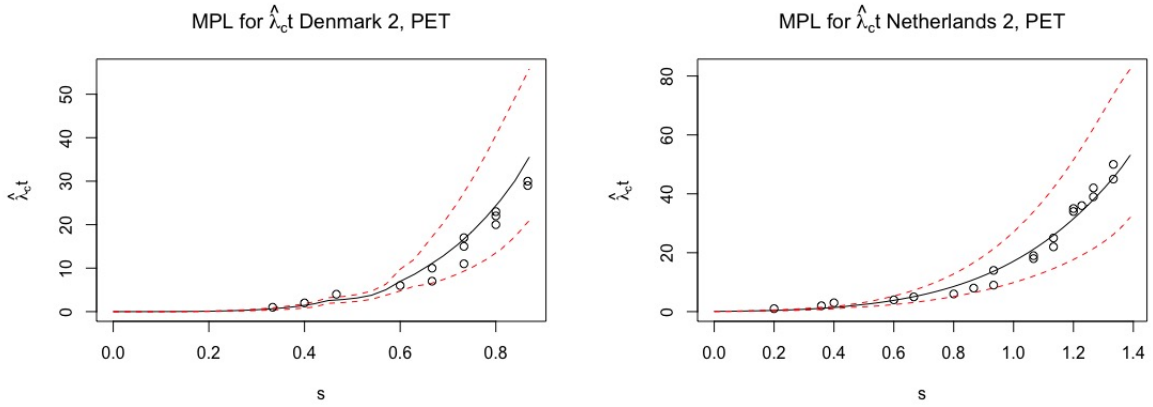


Figure 28: Modified profile likelihood estimates of crash intensity for PET for Denmark 2 and the Netherlands 2 with 95% confidence intervals

## 5.3 Assessing Relative Safety

In this section, we present the results of assessing the relative safety between intersections in Denmark and the Netherlands. We construct confidence intervals based on the maximum-, profile- and modified profile likelihood for $\pi_c^{NE} - \pi_c^{DK}$ in order to test the following hypotheses

$$\begin{aligned} \mathrm{H}_0 : \ & \pi_c^{NE}(s) - \pi_c^{DK}(s) = 0 \\ \mathrm{H}_1 : \ & \pi_c^{NE}(s) - \pi_c^{DK}(s) \neq 0. \end{aligned}$$

Recall that the null hypothesis is acecepted in the event that the confidence interval spans over zero, and if this is not the case then the null hypothesis is rejected. Note that if the hypothesis is rejected, and the $\pi_c^{NE} - \pi_c^{DK} > 0$, then we have good reason to draw the conclusion that risk of a crash is greater in the Netherlands compared to Denmark.

The confidence intervals for $\pi_c^{NE} - \pi_c^{DK}$ based on the maximum likelihood as described in Section 3.4 are given in Figures 29 - 30. The confidence interval for $\pi_c^{NE} - \pi_c^{DK}$ based on the profile likelihood are given in Figures 31 - 32. Finally the corresponding confidence intervals using the MPL are given in Figures 33 - 34.

By inspecting these confidence intervals graphically, it appears that the confidence intervals based on the maximum likelihood method are generally very wide. Furthermore, in general, the width of the confidence interval becomes wider as the $s$ increases, which is quite natural as the number of events with an SMoS $S < s$ follows a binomial distribution if we condition on the interactions, and the variance of a binomial random variable increases as the parameter $\pi \to 0.5$. So, the general results are congruent with what we expect in theory.

Table 9: Modified profile likelihood estimates of the crash intensities with 95% confidence intervals for measures below 0, 0.3 and 0.6 for all data sets and measures

| | | Measure $\leq 0$ | | | Measure $\leq 0.3$ | | | Measure $\leq 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Data | Measure | lower | $\hat{\lambda}_c$ | upper | lower | $\hat{\lambda}_c$ | upper | lower | $\hat{\lambda}_c$ | upper |
| Dk1 | $T2_{min}$ | 0.0009 | 0.0020 | 0.0034 | 0.0101 | 0.0200 | 0.0336 | 0.0308 | 0.0588 | 0.0994 |
| | PET | $5.1 \cdot 10^{-5}$ | $1.5 \cdot 10^{-4}$ | $2.9 \cdot 10^{-4}$ | 0.0021 | 0.0055 | 0.0102 | 0.0116 | 0.0271 | 0.0509 |
| Dk2 | $TTC_{min}$ | 0.0039 | 0.0138 | 0.0279 | 0.0081 | 0.0273 | 0.0544 | 0.0154 | 0.0487 | 0.0955 |
| | $T2_{min}$ | $2.2 \cdot 10^{-5}$ | $8.9 \cdot 10^{-5}$ | $1.9 \cdot 10^{-4}$ | 0.0202 | 0.0634 | 0.1276 | 0.2260 | 0.5645 | 1.1121 |
| | PET | $1.0 \cdot 10^{-6}$ | $3.5 \cdot 10^{-6}$ | $7.2 \cdot 10^{-6}$ | 0.0061 | 0.0175 | 0.0339 | 0.1274 | 0.2936 | 0.5534 |
| Dk3 | $T2_{min}$ | 0.0002 | 0.0004 | 0.0007 | 0.0100 | 0.0195 | 0.0318 | 0.0600 | 0.1110 | 0.1827 |
| | PET | $3.9 \cdot 10^{-7}$ | $2.6 \cdot 10^{-6}$ | $6.8 \cdot 10^{-6}$ | 0.0001 | 0.0008 | 0.0021 | 0.0067 | 0.0326 | 0.0735 |
| Ne1 | $TTC_{min}$ | $4.0 \cdot 10^{-10}$ | 0.0069 | 0.0194 | $3.0 \cdot 10^{-5}$ | 0.0161 | 0.1691 | 0.0007 | 0.0373 | 0.4442 |
| | $T2_{min}$ | $9.4 \cdot 10^{-11}$ | $4.6 \cdot 10^{-4}$ | 0.0015 | 0.0005 | 0.0368 | 0.5494 | 0.0827 | 0.4099 | 1.1787 |
| | PET | $4.5 \cdot 10^{-10}$ | $1.4 \cdot 10^{-5}$ | $2.3 \cdot 10^{-5}$ | $4.1 \cdot 10^{-5}$ | 0.0266 | 0.07256 | 0.0731 | 0.3366 | 1.0599 |
| Ne2 | $TTC_{min}$ | $3.7 \cdot 10^{-12}$ | $3.7 \cdot 10^{-11}$ | $1.3 \cdot 10^{-10}$ | $2.4 \cdot 10^{-6}$ | $2.0 \cdot 10^{-5}$ | $5.4 \cdot 10^{-5}$ | 0.0002 | 0.0011 | 0.0027 |
| | $T2_{min}$ | $2.7 \cdot 10^{-8}$ | $5.7 \cdot 10^{-8}$ | $9.3 \cdot 10^{-8}$ | 0.0169 | 0.0301 | 0.0458 | 0.0949 | 0.1601 | 0.2422 |
| | PET | 0.0013 | 0.0031 | 0.0054 | 0.0151 | 0.0326 | 0.0557 | 0.0783 | 0.1555 | 0.2629 |
| Ne3 | $TTC_{min}$ | $4.2 \cdot 10^{-17}$ | $3.2 \cdot 10^{-16}$ | $10.0 \cdot 10^{-16}$ | $3.6 \cdot 10^{-13}$ | $2.6 \cdot 10^{-12}$ | $7.7 \cdot 10^{-12}$ | $1.7 \cdot 10^{-10}$ | $1.2 \cdot 10^{-9}$ | $3.5 \cdot 10^{-9}$ |
| | $T2_{min}$ | $2.9 \cdot 10^{-7}$ | $2.2 \cdot 10^{-6}$ | $5.9 \cdot 10^{-6}$ | 0.0002 | 0.0013 | 0.0032 | 0.0090 | 0.0362 | 0.0801 |
| | PET | $2.4 \cdot 10^{-6}$ | $1.8 \cdot 10^{-5}$ | $4.8 \cdot 10^{-5}$ | 0.0008 | 0.0045 | 0.0108 | 0.0209 | 0.0776 | 0.1738 |

We also see that the difference in $\pi_c^{NE}(s) - \pi_c^{DK}(s)$ can (and often do) change sign as $s$ varies. This is because the GP distributions used to estimate $\pi_c^{NE}$ and $\pi_c^{DK}$ differ in shape and kurtosis.

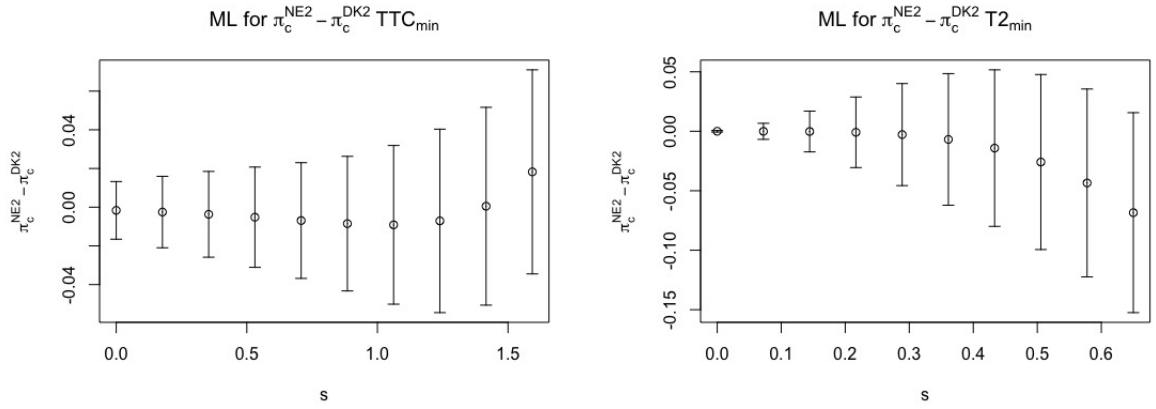


Figure 29: Maximum likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk2}}$ for TTC$_{min}$ and T2$_{min}$ with 95% confidence intervals

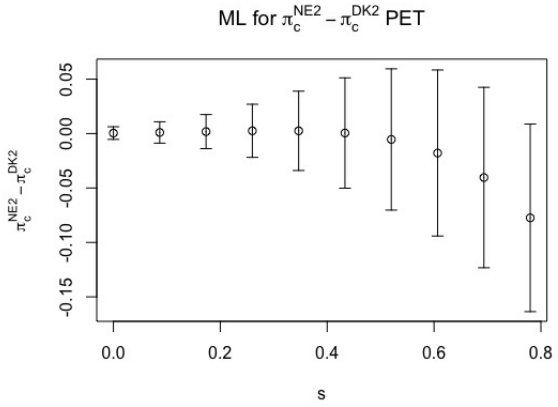


Figure 30: Maximum likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk2}}$ for PET with 95% confidence intervals

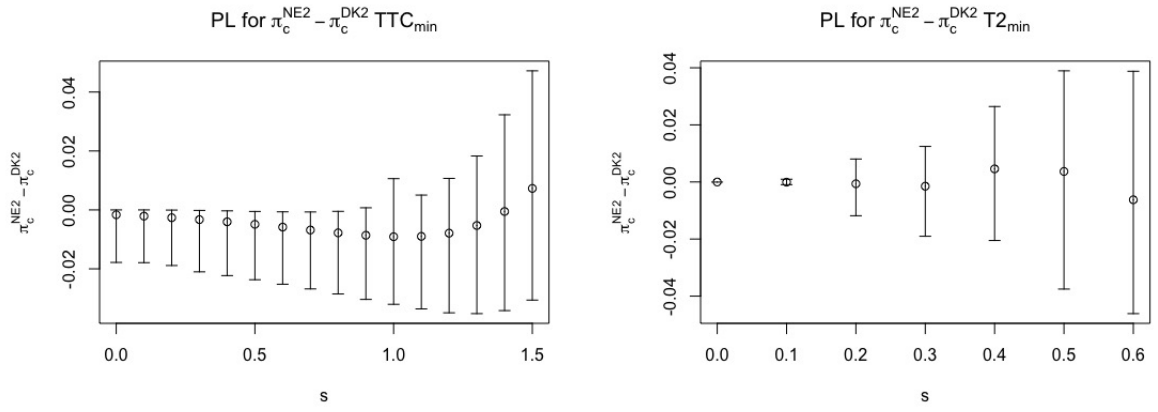


Figure 31: Profile likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk2}}$ for TTC$_{min}$ and T2$_{min}$ with 95% confidence intervals

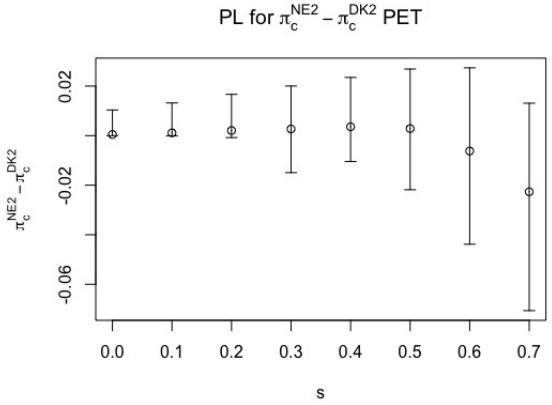


Figure 32: Profile likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk2}}$ for PET with 95% confidence intervals

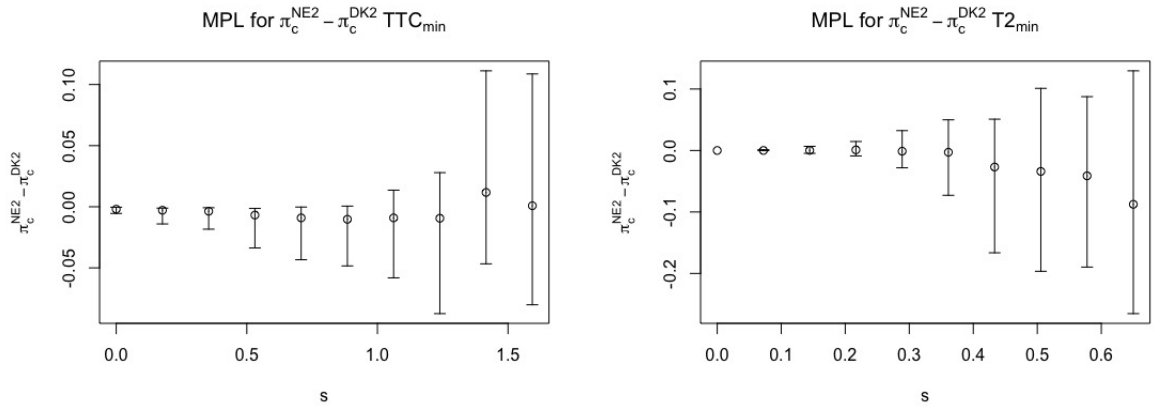


Figure 33: Modified profile likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk2}}$ for TTC$_{min}$ and T2$_{min}$ with 95% confidence intervals

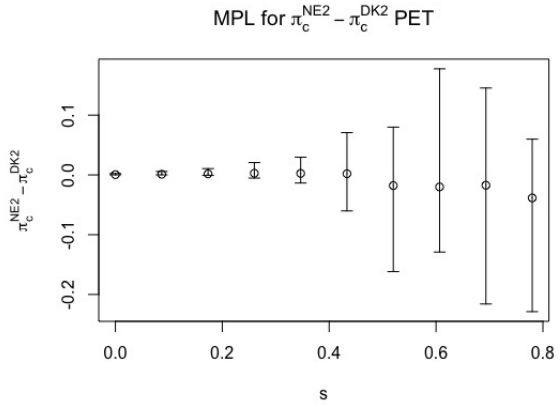


Figure 34: Modified profile likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk2}}$ for PET with 95% confidence intervals

We are primarily interested in the difference in crash probabilities between intersections in Denmark and the Netherlands, since this has the most bearing on the relative safety of the respective intersection design. The confidence intervals for the difference in $\pi_c^{Ne}(0) - \pi_c^{Dk}(0)$ are given in Table 10 for all measures and all combination of intersections in Denmark and the Netherlands.

As can be seen from the table, the profile likelihood based confidence intervals suggest that we can reject the null hypothesis when comparing Ne2−Dk2 using PET. We also reject the null hypothesis for Ne2−Dk1 using $\text{T2}_{min}$, but in this case the difference is negative, which suggests that the intersection in the Netherlands has a lower crash probability. In fact, the conclusion that an intersection is safer in the Netherlands compared to Denmark is made for 7 comparisons using the hypothesis test based on the profile likelihood.

The modified profile likelihood also rejects the null hypothesis for Ne2−Dk1 using $\text{T2}_{min}$ and Ne2−Dk2 $\text{TTC}_{min}$ with a negative sign, suggesting the intersections in the Netherlands have a lower crash probability. However, the null hypothesis is rejected for both Ne2−Dk3 $\text{T2}_{min}$ and Ne2−Dk3 PET with a positive sign, suggesting that the corresponding Danish intersections have a relatively lower crash probability compared to the respective intersections in the Netherlands.

One worrying sign for both MPL and PL is that we have a contradictory result for the comparison of Ne2-Dk2 when using different SMoS. For Ne2-Dk2 PET, we reject the null hypothesis with $\hat{\pi}_c^{Ne} > \hat{\pi}_c^{Dk}$, but for Ne2-Dk2 $\text{TTC}_{min}$ and $\text{T2}_{min}$, we reject the null hypothesis with $\hat{\pi}_c^{Ne} < \hat{\pi}_c^{Dk}$. Consequently, $\text{TTC}_{min}$ and $\text{T2}_{min}$ indicate that Ne2 is safer than Dk2, but PET suggests that Dk2 is safer than Ne2.

These results suggest that there is no real evidence for the general conclusion that the intersection design in any of the countries are safer than the other. However, we were able to reject the null hypothesis for particular intersections using particular measures.

Table 10: Estimates of $\pi_c^{NE}(0) - \pi_c^{DK}(0)$ with 95% confidence intervals for all data sets and measures

| | | Maximum likelihood | | | Profile likelihood | | | Modified profile likelihood | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Data | Measure | lower | $\hat{\pi}_{ne} - \hat{\pi}_{dk}$ | upper | lower | $\hat{\pi}_{ne} - \hat{\pi}_{dk}$ | upper | lower | $\hat{\pi}_{ne} - \hat{\pi}_{dk}$ | upper |
| Ne1–Dk1 | $T2_{min}$ | $-0.0358$ | $-0.0022$ | 0.0314 | $-0.0266$ | $-0.0022$ | 0.0001 | $-0.0113$ | $-0.0023$ | $-0.0002$ |
| | PET | $-0.0056$ | $-0.0002$ | 0.0053 | $-0.0057$ | $-0.0002$ | $4.1 \cdot 10^{-7}$ | $-0.0009$ | $-0.0002$ | $-3.6 \cdot 10^{-5}$ |
| Ne1–Dk2 | $TTC_{min}$ | $-0.0165$ | $-0.0011$ | 0.0142 | $-0.0165$ | $-0.0011$ | 0.0010 | $-0.0080$ | $-0.0013$ | 0.0007 |
| | $T2_{min}$ | $-0.0014$ | $3.4 \cdot 10^{-5}$ | 0.0014 | $-1.8 \cdot 10^{-5}$ | $3.4 \cdot 10^{-5}$ | 0.0002 | $-1.7 \cdot 10^{-5}$ | $4.1 \cdot 10^{-5}$ | 0.0002 |
| | PET | $-8.5 \cdot 10^{-5}$ | $5.7 \cdot 10^{-7}$ | $8.6 \cdot 10^{-5}$ | $-6.6 \cdot 10^{-7}$ | $5.6 \cdot 10^{-7}$ | $1.2 \cdot 10^{-6}$ | $-1.9 \cdot 10^{-6}$ | $6.3 \cdot 10^{-7}$ | $2.5 \cdot 10^{-6}$ |
| Ne1–Dk3 | $T2_{min}$ | $-0.0037$ | $-8.7 \cdot 10^{-5}$ | 0.0036 | $-0.0004$ | $-8.7 \cdot 10^{-5}$ | $4.0 \cdot 10^{-5}$ | $-0.0013$ | $-0.0002$ | $4.3 \cdot 10^{-5}$ |
| | PET | $-8.5 \cdot 10^{-5}$ | $3.5 \cdot 10^{-7}$ | $8.6 \cdot 10^{-5}$ | $-8.6 \cdot 10^{-7}$ | $3.3 \cdot 10^{-7}$ | $1.1 \cdot 10^{-6}$ | $-1.2 \cdot 10^{-6}$ | $2.9 \cdot 10^{-7}$ | $3.6 \cdot 10^{-6}$ |
| Ne2–Dk1 | $T2_{min}$ | $-0.0358$ | $-0.0023$ | 0.0313 | $-0.0184$ | $-0.0023$ | $-0.0004$ | $-0.0031$ | $-0.0025$ | $-0.0007$ |
| | PET | $-0.0077$ | 0.0003 | 0.0083 | $-0.0003$ | 0.0003 | 0.0031 | $-0.0007$ | 0.0003 | 0.0023 |
| Ne2–Dk2 | $TTC_{min}$ | $-0.0165$ | $-0.0017$ | 0.0132 | $-0.0178$ | $-0.0017$ | $-5.3 \cdot 10^{-8}$ | $-0.0079$ | $-0.0019$ | $-0.0002$ |
| | $T2_{min}$ | $-0.0007$ | $-1.1 \cdot 10^{-5}$ | 0.0006 | $-5.0 \cdot 10^{-5}$ | $-1.1 \cdot 10^{-5}$ | $-2.2 \cdot 10^{-6}$ | $-5.4 \cdot 10^{-5}$ | $-1.1 \cdot 10^{-5}$ | $-3.9 \cdot 10^{-6}$ |
| | PET | $-0.0053$ | 0.0005 | 0.0063 | $6.9 \cdot 10^{-6}$ | 0.0005 | 0.0104 | 0.0002 | 0.0005 | 0.0021 |
| Ne2–Dk3 | $T2_{min}$ | $-0.0036$ | $-0.0001$ | 0.0033 | $-0.0002$ | $-0.0001$ | $-0.0001$ | $-0.0002$ | $-0.0001$ | $-8.4 \cdot 10^{-5}$ |
| | PET | $-0.0053$ | 0.0005 | 0.0063 | $-8.2 \cdot 10^{-7}$ | 0.0005 | 0.0104 | 0.0002 | 0.0005 | 0.0024 |
| Ne3–Dk1 | $T2_{min}$ | $-0.0358$ | $-0.0023$ | 0.0313 | $-0.0266$ | $-0.0023$ | $-9.2 \cdot 10^{-6}$ | $-0.0112$ | $-0.0023$ | $-0.0003$ |
| | PET | $-0.0056$ | $-0.0002$ | 0.0053 | $-0.0176$ | $-0.0002$ | $2.5 \cdot 10^{-6}$ | $-0.0009$ | $-0.0002$ | $-7.1 \cdot 10^{-6}$ |
| Ne3–Dk2 | $TTC_{min}$ | $-0.0165$ | $-0.0017$ | 0.0132 | $-0.0178$ | $-0.0017$ | $-3.5 \cdot 10^{-6}$ | $-0.0076$ | $-0.0018$ | $-0.0003$ |
| | $T2_{min}$ | $-0.0007$ | $-1.0 \cdot 10^{-5}$ | 0.0006 | $-2.5 \cdot 10^{-5}$ | $-1.0 \cdot 10^{-5}$ | $-1.03 \cdot 10^{-5}$ | $-3.3 \cdot 10^{-5}$ | $-1.1 \cdot 10^{-5}$ | $-2.8 \cdot 10^{-6}$ |
| | PET | $-0.0002$ | $2.4 \cdot 10^{-6}$ | 0.0002 | $-4.0 \cdot 10^{-7}$ | $2.4 \cdot 10^{-6}$ | $1.1 \cdot 10^{-5}$ | $-6.5 \cdot 10^{-7}$ | $3.2 \cdot 10^{-6}$ | $1.3 \cdot 10^{-5}$ |
| Ne3–Dk3 | $T2_{min}$ | $-0.0036$ | $-0.0001$ | 0.0033 | $-0.0084$ | $-0.0001$ | $2.9 \cdot 10^{-7}$ | $-0.0007$ | $-0.0001$ | $-1.6 \cdot 10^{-5}$ |
| | PET | $-0.0002$ | $2.2 \cdot 10^{-6}$ | 0.0002 | $-1.2 \cdot 10^{-6}$ | $2.2 \cdot 10^{-6}$ | $3.0 \cdot 10^{-5}$ | $-1.2 \cdot 10^{-6}$ | $2.6 \cdot 10^{-6}$ | $1.2 \cdot 10^{-5}$ |

# 6 Discussion

In the course of this thesis we have proposed three methods for computing confidence intervals for the near-crash intensity $\lambda_c(s)$ as well as three methods for testing the relative safety of two intersection $\pi_c^{Ne} - \pi_c^{Dk}$.

For the confidence intervals we noted that we can obtain reasonable point estimates (in terms of not being too high), but that the corresponding interval can give unreasonable upper bounds. This motivates the computation of confidence intervals for $\lambda_c(s)$, since they indicate the uncertainty in our point estimates.

Furthermore, we saw that the confidence intervals based on the delta method were uninformative, since in many cases the lower bounds were less than zero and the intervals were wide. Moreover, the simulation study indicated that these intervals had very low confidence levels.

In contrast, the confidence intervals based on the profile likelihood remedied many of the problems with the delta method. None of the lower bounds were lower than zero, and in many cases the upper bounds of the interval were reasonable. If we couple this with the fact that the profile likelihood based confidence intervals performed very well in the simulation study, it suggests that the profile likelihood produce appropriate interval estimates for $\lambda_c(s)$. The modified profile likelihood was considered since it purportedly has better small sample properties compared to the profile likelihood. However, based on the simulation study, the confidence level for the confidence intervals based on the MPL generally performed worse than the profile likelihood, except in one case. Moreover, we saw that the MPL consistently produced higher point estimates of $\lambda_c(0)$ than the maximum- and profile likelihood, as a consequence we sometimes obtained very high lower bounds for the confidence interval. It should also be noted, that the implementation of the MPL was complicated and the optimization took a long time compared to the corresponding computations with the ML and PL.

In terms of the choice of SMoS we saw that the profile likelihood confidence intervals based on PET produced reasonable interval estimates of $\lambda_c(s)$ more times than any other surrogate measure. However, it is difficult to conclude that any one measure produce better estimates than another based on this data.

We also compared the relative safety of an intersection in Denmark with an intersection in the Netherlands, and we considered all pairwise combinations for all SMoS using the hypothesis test in Equation 36. The maximum likelihood based confidence intervals using the delta method were again very wide and we never rejected the null hypothesis. The profile likelihood rejected the null hypothesis for Ne2$-$Dk2 PET, which suggests that we have some evidence for the fact that Dk2 is relatively safer compared to Ne2. We were also able to reject the null hypothesis using MPL. However, the contrary conclusion was more prevalent since we rejected the null hypothesis with

$\hat{\pi}_c^{Ne} > \hat{\pi}_c^{Dk}$ more often. But in general, it was difficult to reject the null hypothesis regardless of the SMoS used, and regardless of the method used.

According to the literature comparing the two intersection designs, the bicycle lanes in Denmark (separated by painted lines) are generally safer than the intersections in the Netherlands with physically separated bicycle crossings. This is based on the literature review in Prati et al. [14]. Our findings do not indicate that one type infrastructure design is safer than the other based on the hypothesis tests in Table 10.

# 7 Proposals for Future Research

There are many aspects of this thesis that could be altered or pursued further, and so we limit ourselves to listing a few that we think will be most successful with respect to the purpose. One of the problems of finding good point and interval estimates is to have an accurate model of the data-generating process. In this thesis, we have made the simplifying assumption that the near-crash frequency $C_f(s,t)$ can be modelled as a stationary poisson process. However, we know that the traffic intensity is a function of time, and we saw that $C_f(s,t)$ is a function of the traffic intensity, so we should include the non-stationarity in our model.

A simple way to introduce non-stationarity is to let the parameter estimates be a function of time. In doing so, the intensity $\lambda_c(s)$ will also be a function of time. For example, we could allow the scale parameter to be periodic with a 24 hour period, so that

$$\sigma(t) = \beta_0 + \beta_1 \cos\left(\frac{2\pi}{24}t\right) + \beta_2 \sin\left(\frac{2\pi}{24}t\right).$$

Letting the parameters of the GPD vary with time is intuitive, since in principle the outcome of an interaction is affected by other traffic conditions such as traffic flow. Detailed descriptions of implementation of such models are given in Davison and Smith (1990) [31]. Profile likelihood estimates for $\lambda_c$ are still possible under this increased model complexity, the only difference being an increase in the amount of nuisance parameters.

Earlier we noted that when we conduct statistical inference based on the GPD, we are forced to rely on asymptotic properties of likelihood function. There are many different adjustments to the classical asymptotic results related to the likelihood which improve inference for small samples. In this thesis, we considered the modified profile likelihood, however, we found that it was cumbersome to compute and optimize over, and the results were not an improvement compared to the profile likelihood. In fact, a major problem with modified proifle likelihood is that the point estimates of $\lambda_c(s)$ were generally higher than the profile likelihood. It may be worth considering other types of adjustments, for example, modifications to likelihood ratio can be made by multiplication with some correction factor (see Lawley 1956). By doing so, we would only

change the confidence bounds. Tajvidi (2003) derived the corresponding correction factors for the likelihood ratio statistic for the shape and scale parameters of the Generalized Pareto Distribution [13]. Similar calculations could be made for $\lambda_c(s)$ and $\pi_c^{NE}(s) - \pi_c^{DK}(s)$ and the corresponding likelihood functions in Section 3.

Finally, a prescient point was made in a previous thesis that in principle, the lower endpoint of the SMoS $S$ is equal to zero; so that $S = 0$ corresponds to a crash. In order to include this piece of prior information, the truncated GPD for $\tilde{S} - \tilde{u} \mid \tilde{S} > \tilde{u}$ was proposed. It was used to compute crash probabilities after estimating the parameters of the GPD using the standard untruncated likelihood function [36]. In theory, one should use the truncated GPD to estimate the corresponding parameters. The profile- and modified profile likelihood could be derived using the truncated distribution, however, to our knowledge no one has derived the asymptotic properties of the likelihood based on the truncated GPD. This may be something worth investigating, since the truncated GPD is a more accurate model for the data.

# References


[1] "Towards a European road safety area: policy orientations on road safety 2011-2020". In: (July 20, 2010). European Commission. Brussels, 2010.

[2] Khabat Amin et al. *Analys av trafiksäkerhetsutvecklingen 2018. Målstyrning av trafiksäkerhetsarbetet mot etappmålen 2020.* Tech. rep. Trafikverket, 2019, pp. 6–7, 67.

[3] G. A. Davis et al. "Outline for a casual model of traffic conflicts and crashes". In: *Accident Analysis and Prevention* 43 (2011), pp. 1907–1919.

[4] Christer Hydén. "The development of a method for traffic safety evaluation: the swedish traffic conflict technique". Doctoral thesis. Lund University, 1987. Department of Traffic Planning and Engineering.

[5] A. Laureshyn, Å. Svensson, and C. Hydén. "Evaluation of traffic safety, based on micro–level behavioural data: Theoretical framework and first implementation". In: *Accident Analysis & Prevention* 42 (2010).

[6] J. L. Harris and S. R. Perkins. Traffic conflict characteristics – accident potential at intersections. In: Highway Research Record vol. 225 (1968).

[7] J.C Hayward. Near–miss determination through use of a scale of danger. In: Highway Research Record vol. 384 (1972).

[8] K. L. Campbell, H. C. Joksch, and P. E. Green. *A Bridging Analysis for Estimating the Benefits of Active Safety Technologies.* UMTRI-96-18 Final Report. Ann Arbor, MI: University of Michigan. Transportation Research Institute, Apr. 1996.

[9] P. Songchitruksa and A. Tarko. "The extreme value theory approach to safety estimation". In: *Accident Analysis and Prevention* 38 (2006).

[10] L. Zheng, K. Ismail, and X. Meng. "Freeway safety estimation using extreme value theory approaches: A comparative study". In: *Accident Analysis & Prevention* 62 (2014).

[11] A. P. Tarko. "Use of crash surrogates and exceedance statistics to estimate road safety". In: *Accident Analysis & Prevention* 45 (2012).

[12] Haneen Farah and Carlos Lima Azevedo. "Safety analysis of passing maneuvers using extreme value theory". In: *IATSS Research* 41 (2017).

[13] Nader Tajvidi. "Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions". In: *Extremes* 6 (2004), pp. 111–123.

[14] Gabriele Prati et al. "Factors contributing to bicycle–motorised vehicle collisions: a systematic literature review". In: *Transport Reviews* 38.2 (2018), pp. 184–208.

[15] Allan Gut. *An Intermediate Course in Probability.* 1st ed. Springer Science + Business Media, 2009.

[16] M.R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* 1st ed. Springer-Verlag, 1983.

[17] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. 1st ed. London: Springer-Verlag, 2001.

[18] E.L. Lehmann and George Casella. *Theory of Point Estimation*. 1st ed. New York: Springer-Verlag, 1998.

[19] Thomas A. Severini. *Likelihood Methods in Statistics*. 1st ed. New York: Oxford University Press, 2000.

[20] O.E. Barndorff-Nielsen. "On a formula for the distribution of the maximum likelihood estimator". In: *Biometrika* 70 (1983), pp. 343–365.

[21] D.R. Cox and N. Reid. "Parameter Orthogonality and Approximate Conditional Inference". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 49 (1987), pp. 1–39.

[22] Thomas A. Severini. "An Approximation to the Modified Likelihood Function". In: *Biometrika* 85 (1998), pp. 403–411.

[23] Thomas A. Severini. "An Empirical Adjustment to the Likelihood Ratio Statistic". In: *Biometrika* 86 (June 1999).

[24] N. Reid. "Summary of Some Statistical Issues". In: *Statistical Problems in Particle Physics Astrophysics and Cosmology*. Toronto: Imperial College Press, 2006, pp. 279–282.

[25] D.R. Cox and D.V. Hinkley. *Theoretical Statistics*. 1st ed. London: Chapman and Hall, 1974.

[26] A.C. Davison. *Statistical Models*. 1st ed. Cambridge: Cambridge University Press, 2003.

[27] E.L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. 3rd ed. New York: Springer Science + Business Media, 2005.

[28] David E. Mathews. "Likelihood-Based Confidence Intervals for Functions of Many Parameters". In: *Biometrika* 75 (Mar. 1988).

[29] Thomas J. Diciccio and Steven E. Stern. *An Adjustment to Profile Likelihood Based on Observed Information*. Tech. rep. Department of Statistics, University of Stanford, June 1993.

[30] Eric Gilleland and Richard W. Katz. "extRemes 2.0: An Extreme Value Analysis Package in R". In: *Journal of Statistical Software* 72.8 (2016), pp. 1–39. DOI: 10.18637/jss.v072.i08.

[31] A.C. Davison and R.L. Smith. "Models for Exceedences Over High Thresholds". In: *Journal of The Royal Statistical Society. Series B* 52.3 (1990), pp. 393–442.

[32] R.L. Smith. "Threshold Methods for Sample Extremes". In: *Statistical Extremes and Applications*. Ed. by J. Tiago de Oliveira. Dordrecht: Springer-Science + Business Media, 1984.

[33] Juliana F. Pires, Audrey H. M. A. Cysneiros, and Francisco Cribari-Neto. "Improved inference for the generalized Pareto distribution". In: *Brazilian Journal of Probability and Statistics* 32.1 (2018).

[34] Steven G. Johnson. *The NLopt non-linear optimization package*. URL: http://github.com/stevengj/nlopt.

[35] J.R. Norris. *Markov Chains*. Cambridge: Cambridge University Press, 1997.

[36] Johanna Lägnert. “An Extreme Value Approach to Road Safety Analysis”. MA thesis. Lund: Lund University, 2019.

# Appendices

## A Scripts

The R functions that we developed to compute the confidence intervals can be found and downloaded at the following URL: github.com/koba96/masters-thesis .

## B Results For All Datasets

### B.1 Estimates of Near-Crash Intensities

#### B.1.1 Maximum Likelihood

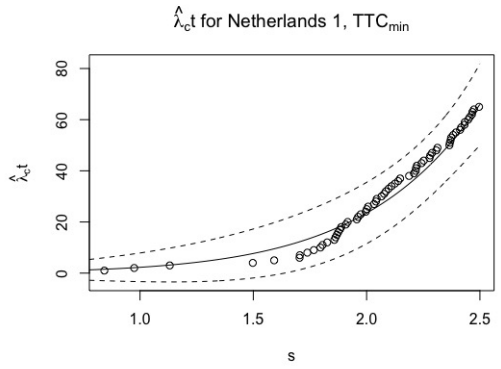


Figure 35: Maximum likelihood estimates of crash intensity for $\text{TTC}_{min}$ for the Netherlands 1 with a 95% confidence interval

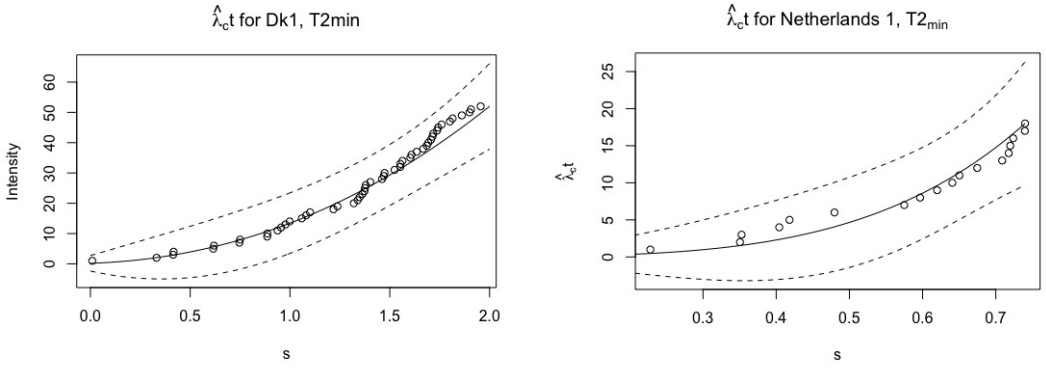


Figure 36: Maximum likelihood estimates of crash intensity for $\text{T2}_{min}$ for Denmark 1 and the Netherlands 1 with 95% confidence intervals

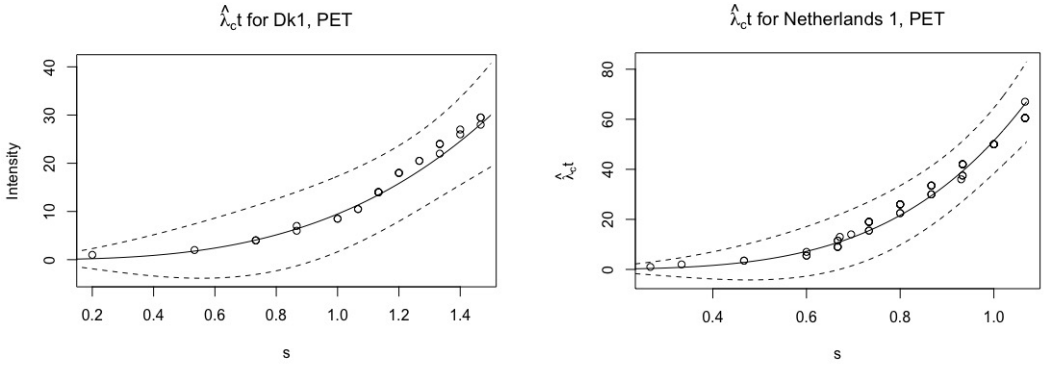


Figure 37: Maximum likelihood estimates of crash intensity for PET for Denmark 1 and the Netherlands 1 with 95% confidence intervals

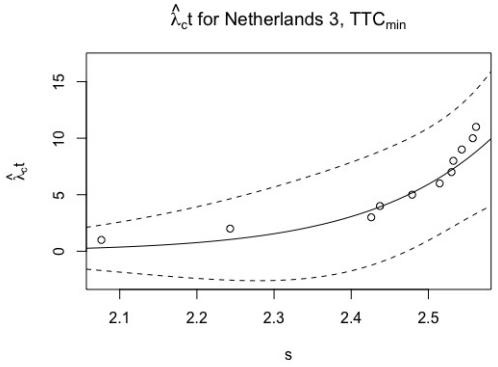


Figure 38: Maximum likelihood estimates of crash intensity for $\text{TTC}_{min}$ for the Netherlands 3 with a 95% confidence interval

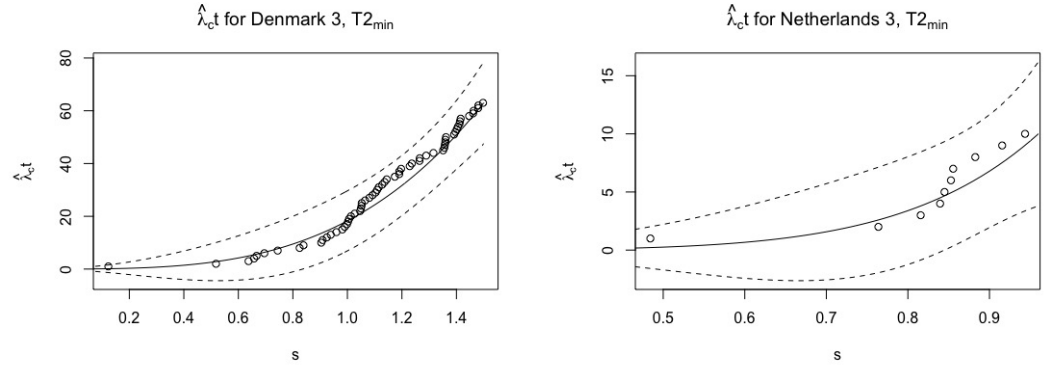


Figure 39: Maximum likelihood estimates of crash intensity for $\text{T2}_{min}$ for Denmark 3 and the Netherlands 3 with 95% confidence intervals

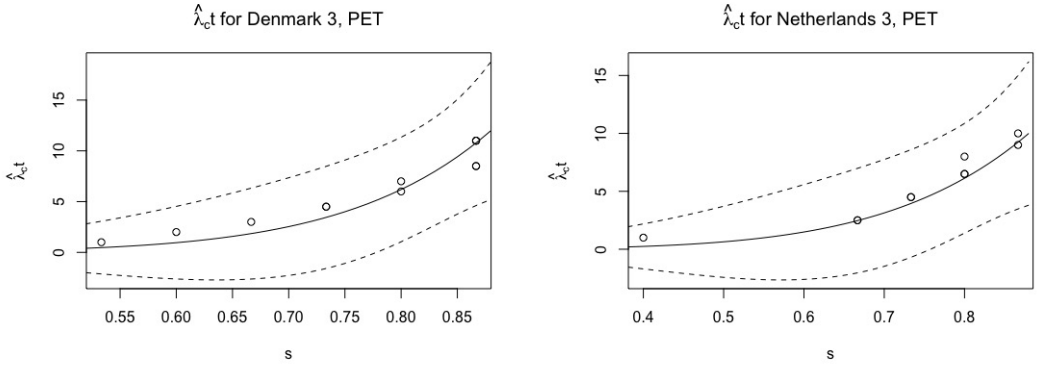


Figure 40: Maximum likelihood estimates of crash intensity for PET for Denmark 3 and the Netherlands 3 with 95% confidence intervals

### B.1.2 Profile Likelihood

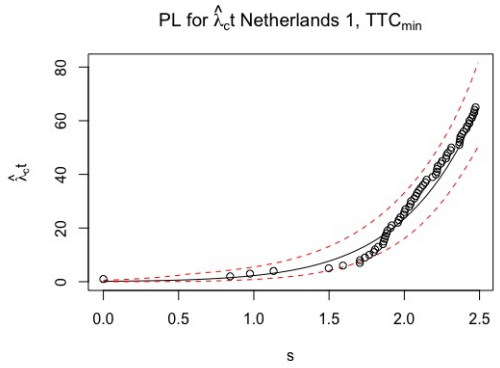


Figure 41: Profile likelihood estimates of crash intensity for $\text{TTC}_{min}$ for the Netherlands 1 with a 95% confidence interval

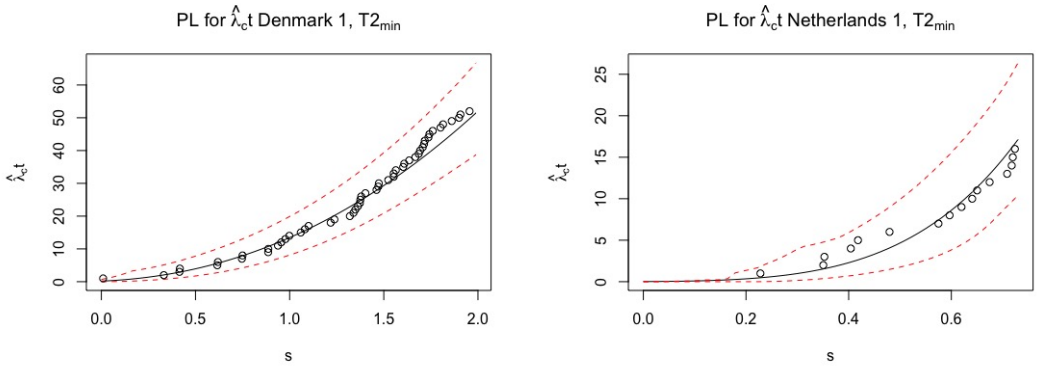


Figure 42: Profile likelihood estimates of crash intensity for $\text{T2}_{min}$ for Denmark 1 and the Netherlands 1 with 95% confidence intervals

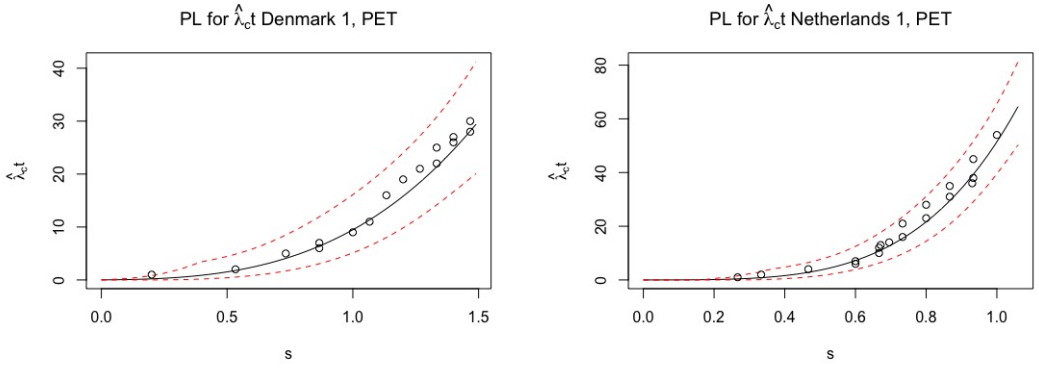


Figure 43: Profile likelihood estimates of crash intensity for PET for Denmark 1 and the Netherlands 1 with 95% confidence intervals

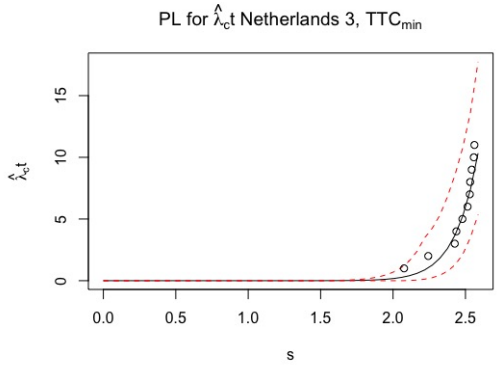


Figure 44: Profile likelihood estimates of crash intensity for $TTC_{min}$ for the Netherlands 3 with a 95% confidence interval

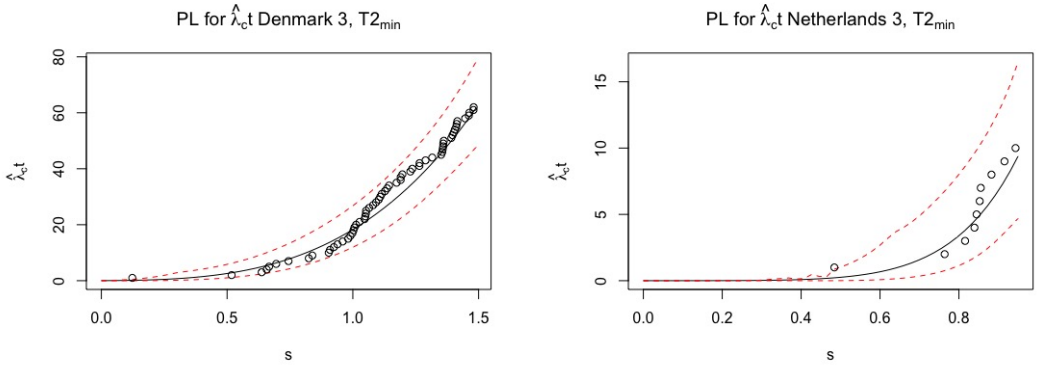


Figure 45: Profile likelihood estimates of crash intensity for $T2_{min}$ for Denmark 3 and the Netherlands 3 with 95% confidence intervals

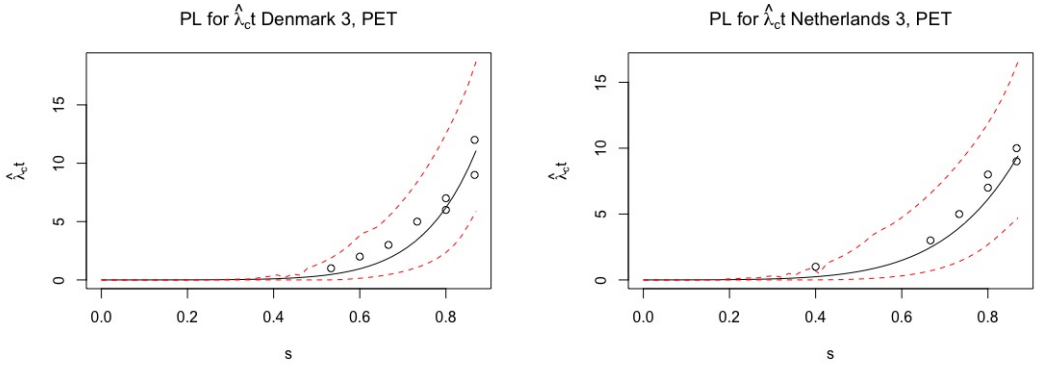


Figure 46: Profile likelihood estimates of crash intensity for PET for Denmark 3 and the Netherlands 3 with 95% confidence intervals

### B.1.3 Modified Profile Likelihood

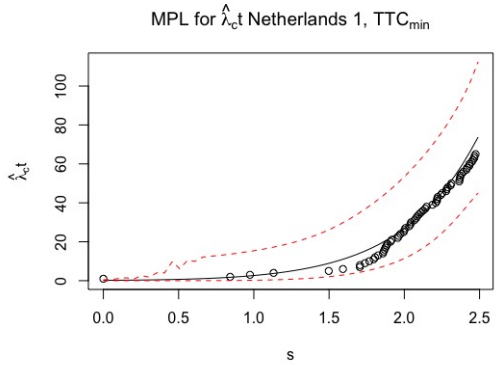


Figure 47: Modified profile likelihood estimates of crash intensity for $TTC_{min}$ for the Netherlands 1 with 95% confidence intervals

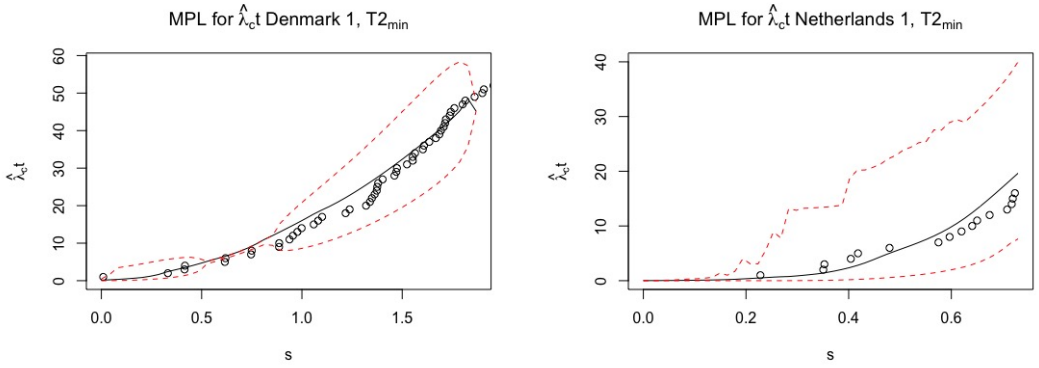


Figure 48: Modified profile likelihood estimates of crash intensity for $T2_{min}$ for Denmark 1 and the Netherlands 1 with 95% confidence intervals

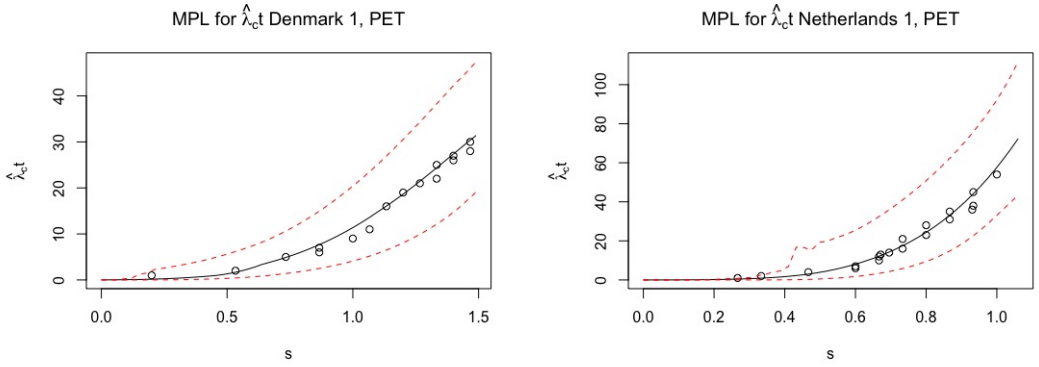


Figure 49: Modified profile likelihood estimates of crash intensity for PET for Denmark 1 and the Netherlands 1 with 95% confidence intervals

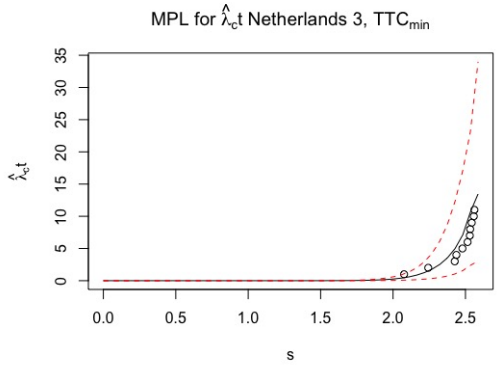


Figure 50: Modified profile likelihood estimates of crash intensity for $\text{TTC}_{min}$ and for the Netherlands 3 with a 95% confidence interval

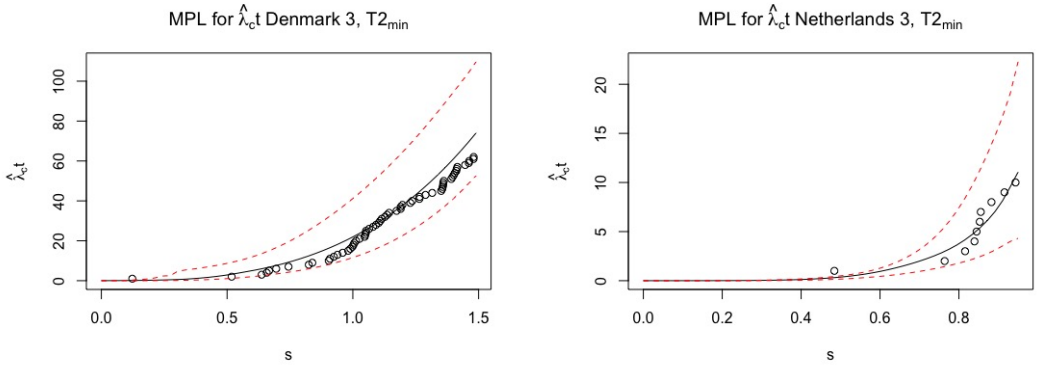


Figure 51: Modified profile likelihood estimates of crash intensity for $\text{T2}_{min}$ for Denmark 3 and the Netherlands 3 with 95% confidence intervals

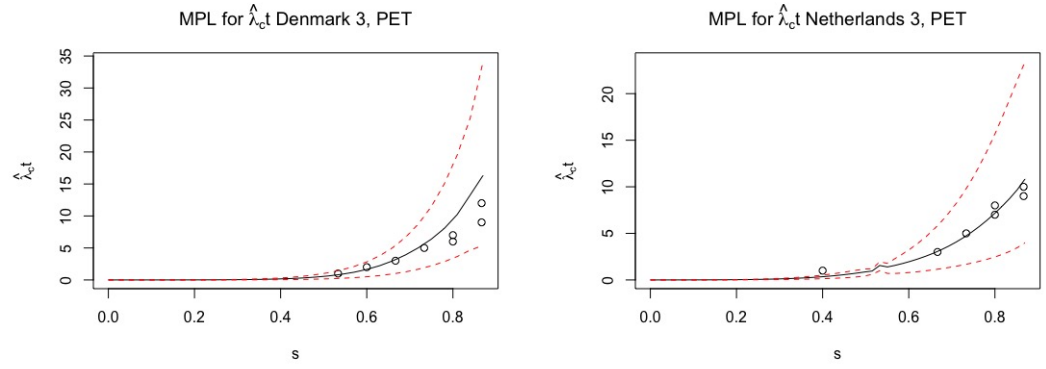


Figure 52: Modified profile likelihood estimates of crash intensity for PET for Denmark 3 and the Netherlands 3 with 95% confidence intervals

## B.2 Assessing Relative Safety

### B.2.1 Maximum Likelihood

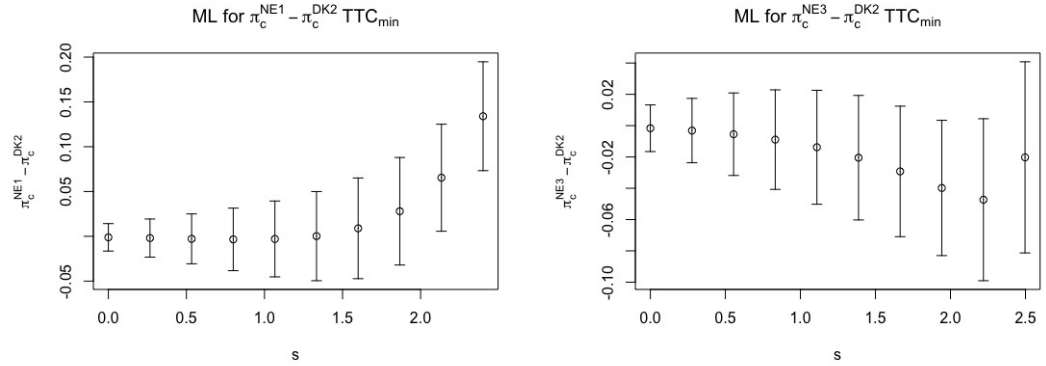


Figure 53: Maximum likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ for $\text{TTC}_{min}$ with 95% confidence intervals

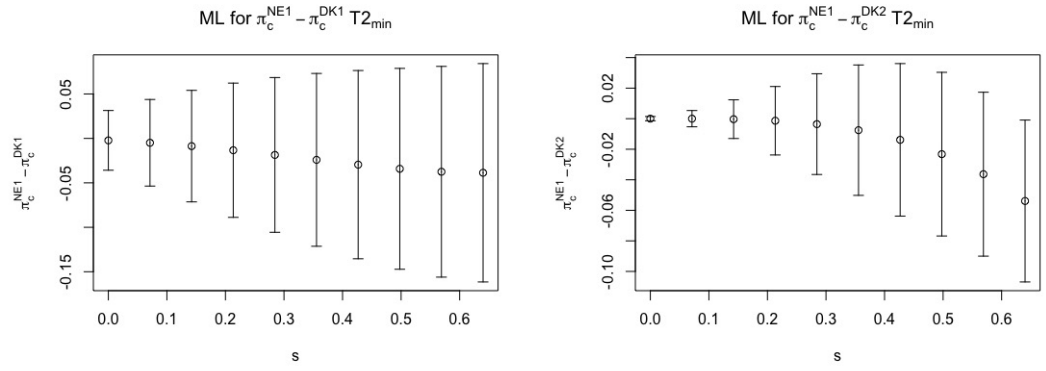


Figure 54: Maximum likelihood estimates of $\pi_{\mathrm{Ne1}} - \pi_{\mathrm{Dk1}}$ and $\pi_{\mathrm{Ne1}} - \pi_{\mathrm{Dk2}}$ for $\mathrm{T2}_{min}$ with 95% confidence intervals

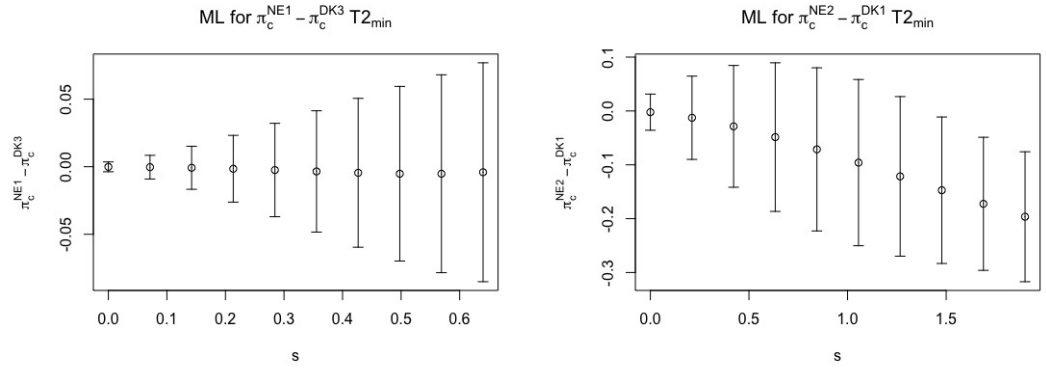


Figure 55: Maximum likelihood estimates of $\pi_{\mathrm{Ne1}} - \pi_{\mathrm{Dk3}}$ and $\pi_{\mathrm{Ne2}} - \pi_{\mathrm{Dk1}}$ for $\mathrm{T2}_{min}$ with 95% confidence intervals

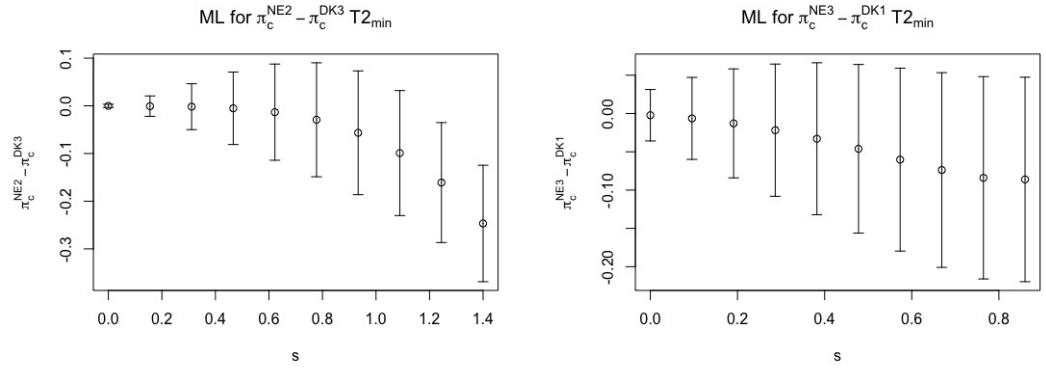


Figure 56: Maximum likelihood estimates of $\pi_{\mathrm{Ne3}} - \pi_{\mathrm{Dk3}}$ and $\pi_{\mathrm{Ne3}} - \pi_{\mathrm{Dk1}}$ for $\mathrm{T2}_{min}$ with 95% confidence intervals

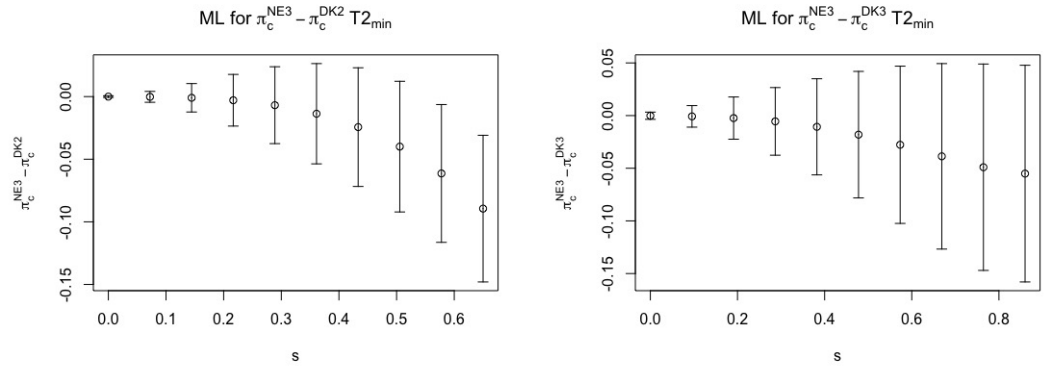


Figure 57: Maximum likelihood estimates of $\pi_{\mathrm{Ne3}} - \pi_{\mathrm{Dk2}}$ and $\pi_{\mathrm{Ne3}} - \pi_{\mathrm{Dk3}}$ for $\mathrm{T2}_{min}$ with 95% confidence intervals

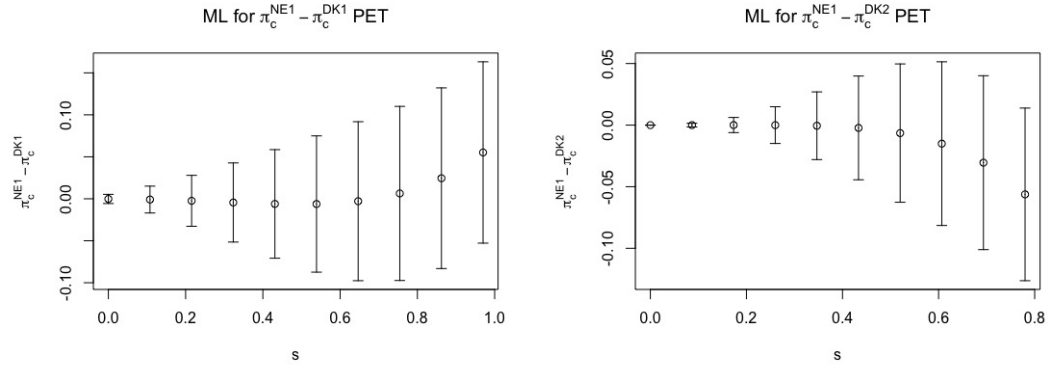


Figure 58: Maximum likelihood estimates of $\pi_{\mathrm{Ne1}} - \pi_{\mathrm{Dk1}}$ and $\pi_{\mathrm{Ne1}} - \pi_{\mathrm{Dk2}}$ for PET with 95% confidence intervals

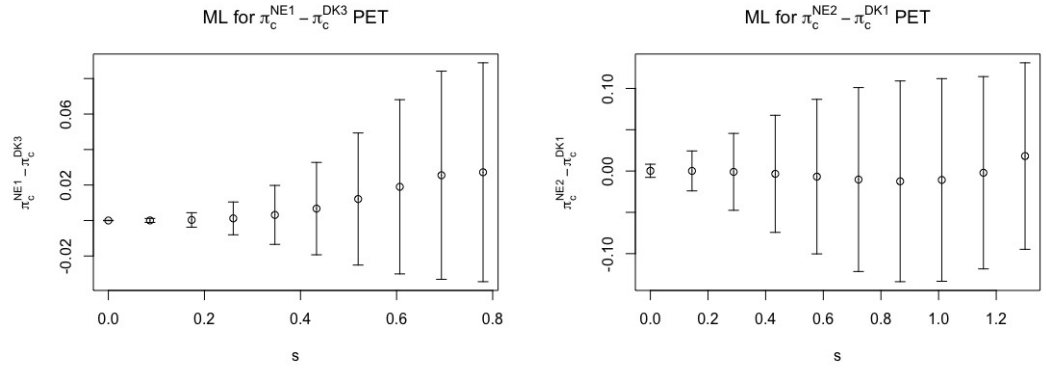


Figure 59: Maximum likelihood estimates of $\pi_{\mathrm{Ne1}} - \pi_{\mathrm{Dk3}}$ and $\pi_{\mathrm{Ne2}} - \pi_{\mathrm{Dk1}}$ for PET with 95% confidence intervals

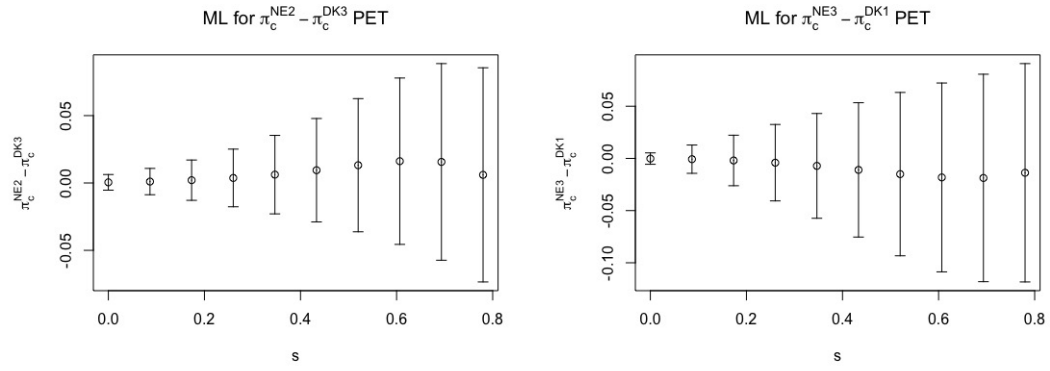


Figure 60: Maximum likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk1}}$ for PET with 95% confidence intervals

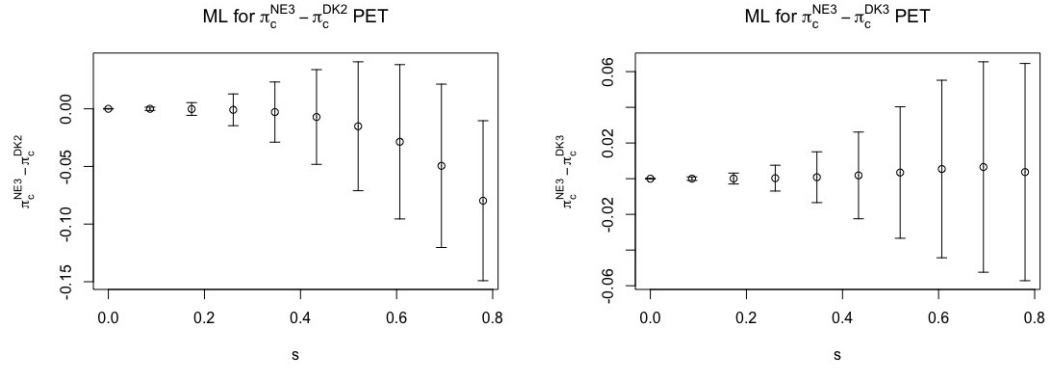


Figure 61: Maximum likelihood estimates of $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk3}}$ for PET with 95% confidence intervals

### B.2.2 Profile Likelihood

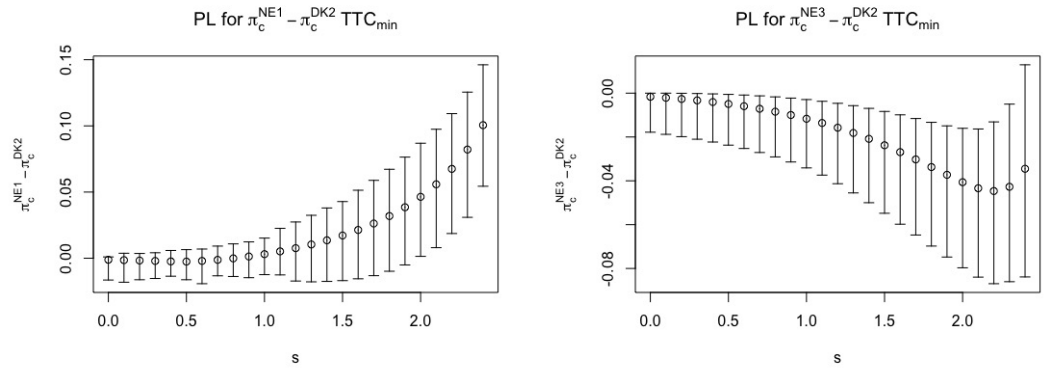


Figure 62: Profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ for $\text{TTC}_{min}$ with 95% confidence intervals

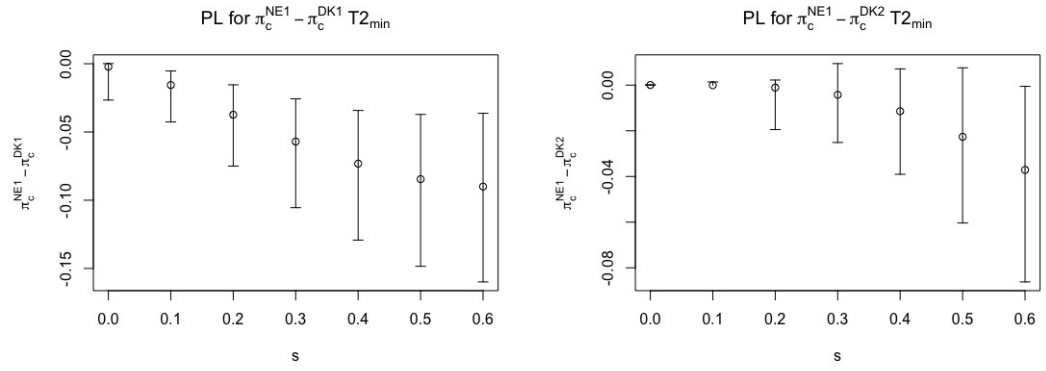


Figure 63: Profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk1}}$ and $\pi_{\text{Ne1}} - \pi_{\text{Dk2}}$ for $\text{T2}_{min}$ with 95% confidence intervals

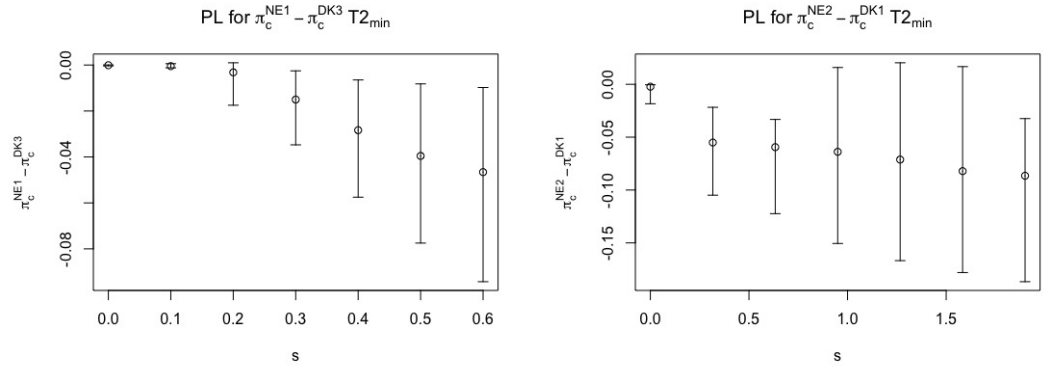


Figure 64: Profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne2}} - \pi_{\text{Dk1}}$ for $\text{T2}_{min}$ with 95% confidence intervals

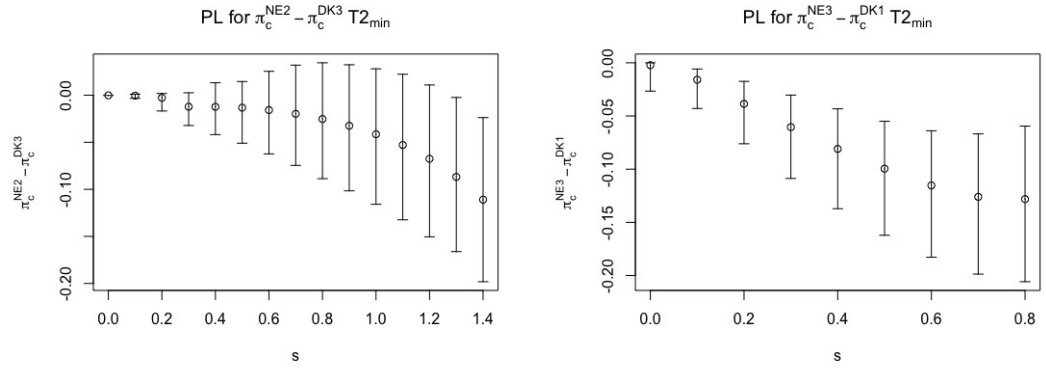


Figure 65: Profile likelihood estimates of $\pi_{\text{Ne3}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk1}}$ for $\text{T2}_{min}$ with 95% confidence intervals

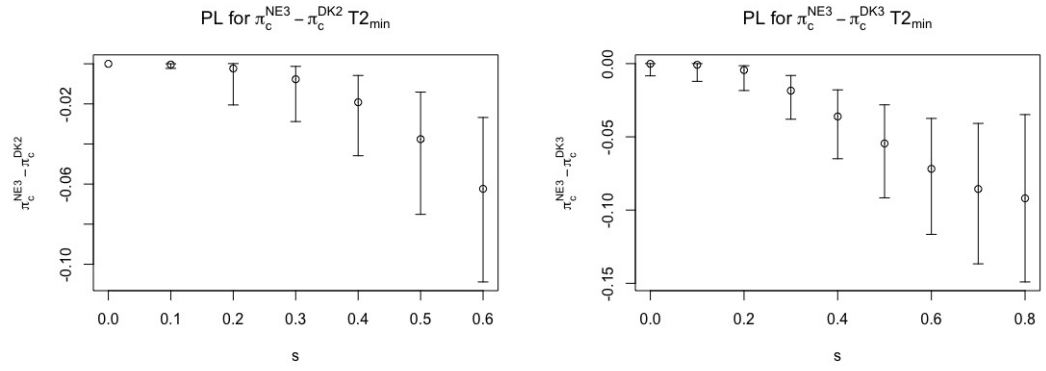


Figure 66: Profile likelihood estimates of $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk3}}$ for $\text{T2}_{min}$ with 95% confidence intervals

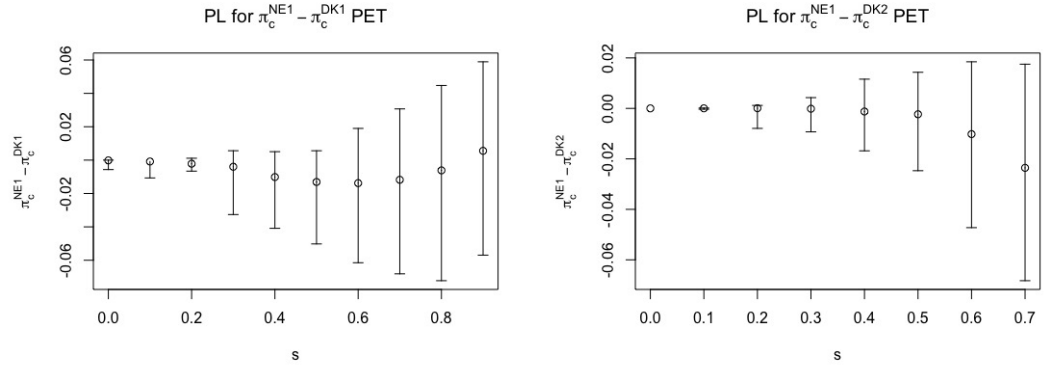


Figure 67: Profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk1}}$ and $\pi_{\text{Ne1}} - \pi_{\text{Dk2}}$ for PET with 95% confidence intervals

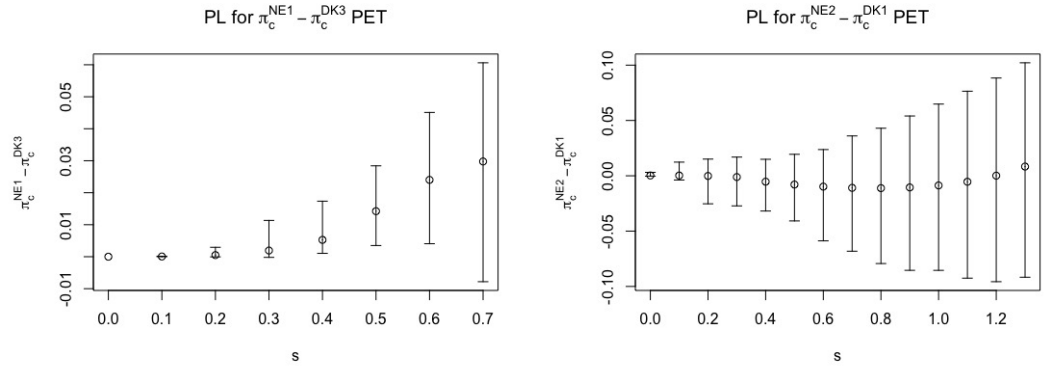


Figure 68: Profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne2}} - \pi_{\text{Dk1}}$ for PET with 95% confidence intervals

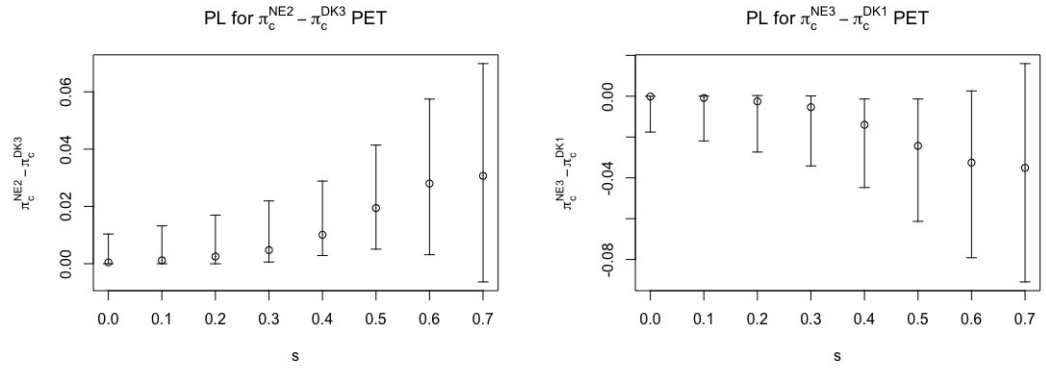


Figure 69: Profile likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk1}}$ for PET with 95% confidence intervals

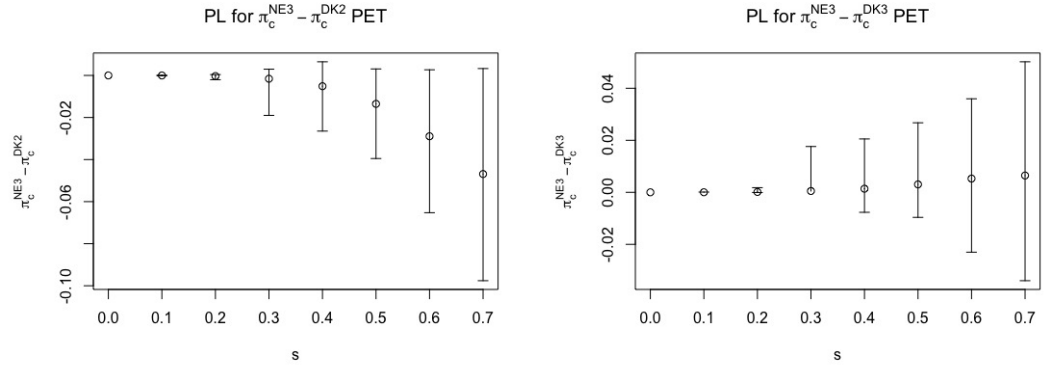


Figure 70: Profile likelihood estimates of $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk3}}$ for PET with 95% confidence intervals

### B.2.3 Modified profile Likelihood

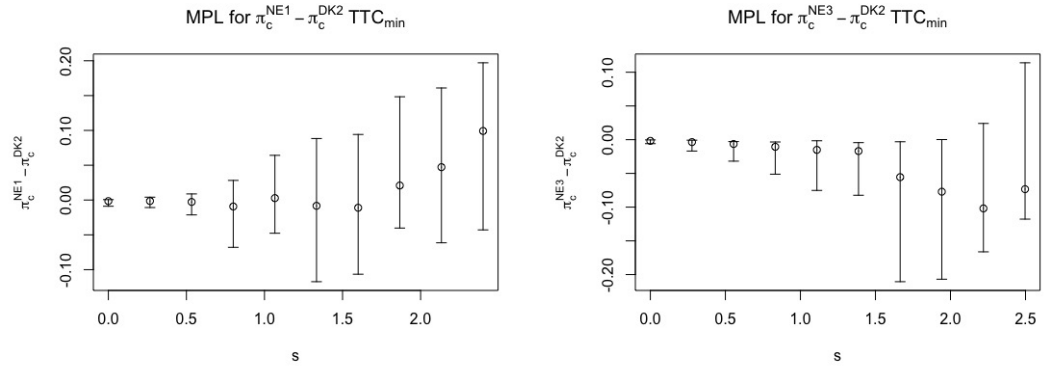


Figure 71: Modified profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ for $\text{TTC}_{min}$ with 95% confidence intervals

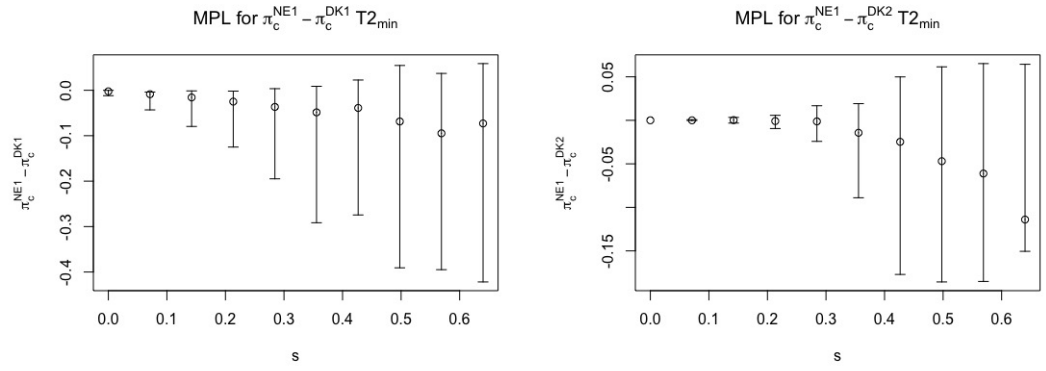


Figure 72: Modified profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk1}}$ and $\pi_{\text{Ne1}} - \pi_{\text{Dk2}}$ for $\text{T2}_{min}$ with 95% confidence intervals

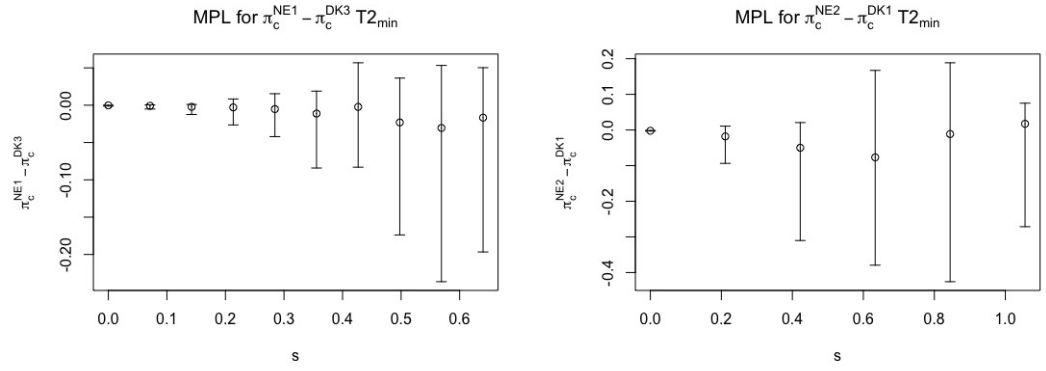


Figure 73: Modified profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne2}} - \pi_{\text{Dk1}}$ for $\text{T2}_{min}$ with 95% confidence intervals

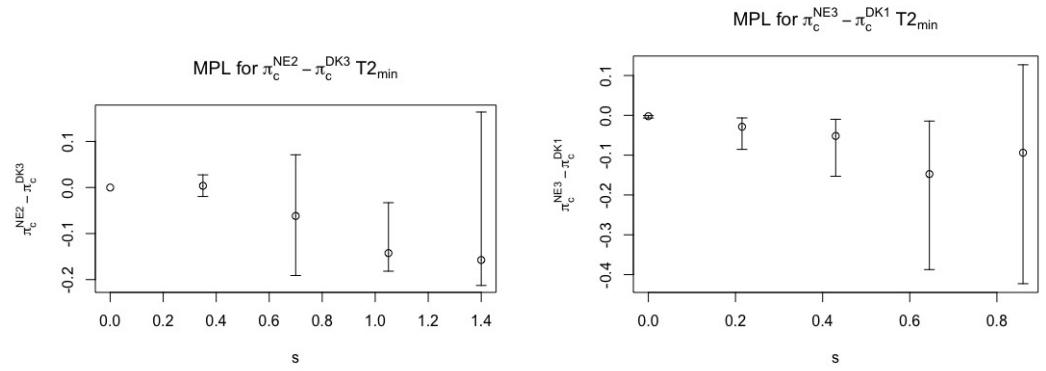


Figure 74: Modified profile likelihood estimates of $\pi_{\text{Ne3}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk1}}$ for $\text{T2}_{min}$ with 95% confidence intervals

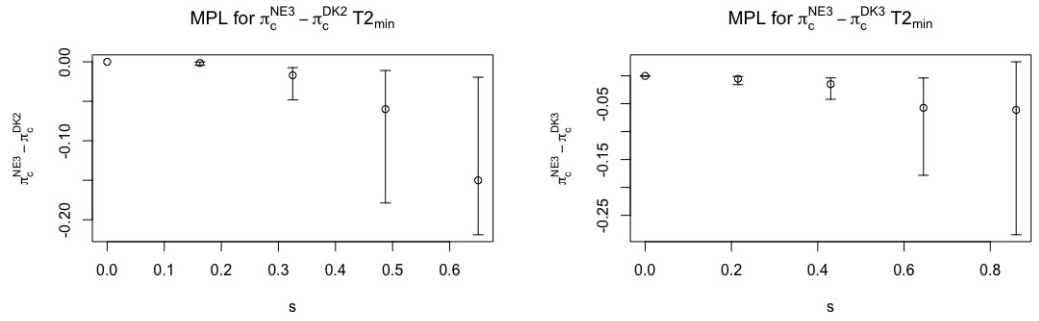


Figure 75: Modified profile likelihood estimates of $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk3}}$ for $\text{T2}_{min}$ with 95% confidence intervals

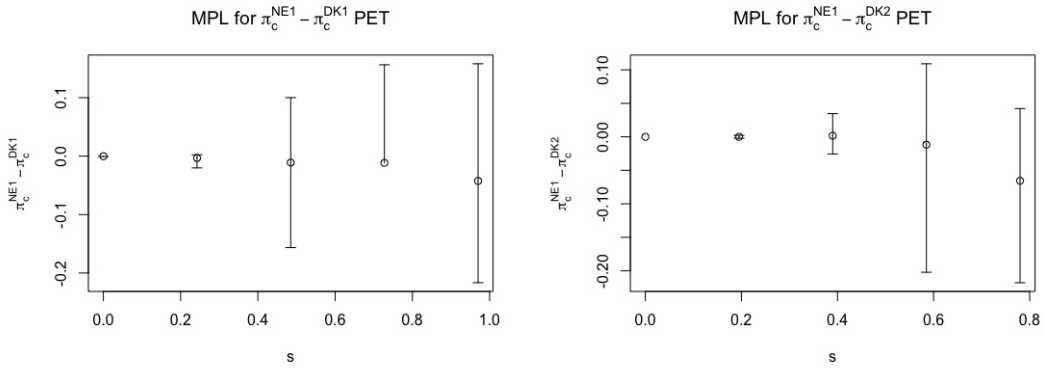


Figure 76: Modified profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk1}}$ and $\pi_{\text{Ne1}} - \pi_{\text{Dk2}}$ for PET with 95% confidence intervals

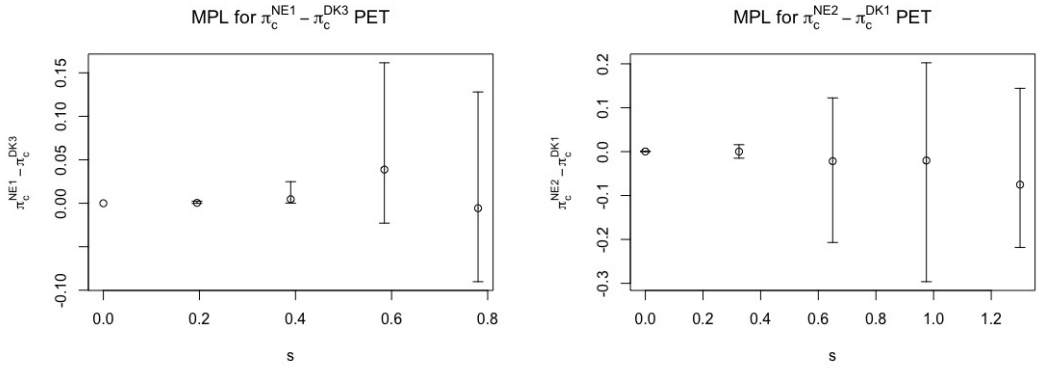


Figure 77: Modified profile likelihood estimates of $\pi_{\text{Ne1}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne2}} - \pi_{\text{Dk1}}$ for PET with 95% confidence intervals

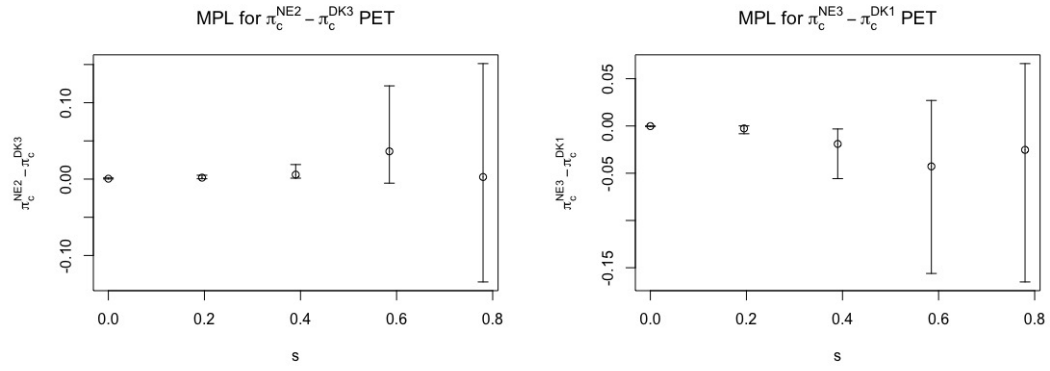


Figure 78: Modified profile likelihood estimates of $\pi_{\text{Ne2}} - \pi_{\text{Dk3}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk1}}$ for PET with 95% confidence intervals

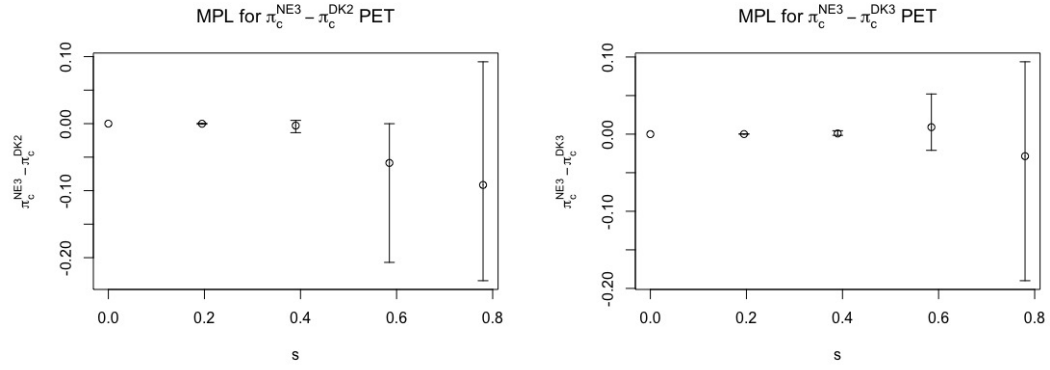


Figure 79: Modified profile likelihood estimates of $\pi_{\text{Ne3}} - \pi_{\text{Dk2}}$ and $\pi_{\text{Ne3}} - \pi_{\text{Dk3}}$ for PET with 95% confidence intervals