Robust tests in generalized linear models with missing responses*

Ana M. Bianco, Graciela Boente and Isabel M. Rodrigues

Abstract

In many situations, data follow a generalized linear model in which the mean of the responses is modelled, through a link function, linearly on the covariates. In this paper, robust estimators for the regression parameter are considered in order to build test statistics for this parameter when missing data occur in the responses. We derive the asymptotic behaviour of the robust estimators for the regression parameter under the null hypothesis and under contiguous alternatives in order to obtain that of the robust Wald test statistics. Their influence function is also studied. A simulation study allows to compare the behaviour of the classical and robust tests, under different contamination schemes. The procedure is also illustrated analysing a real data set.

Key Words: Fisher–consistency, Generalized Linear Models, Influence Function, Missing Data, Outliers, Robust Estimation, Robust Testing

AMS Subject Classification: MSC 62F35 MSC 62F05

*Ana M. Bianco (Email: abianco@dm.uba.ar) and Graciela Boente (Email: gboente@dm.uba.ar) are Professors at the Instituto de Cálculo and the Departamento de Matemáticas, respectively, from the Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, Buenos Aires 1428, Argentina (Tel./FAX: 54-11-45763375). They have also a researcher position at the CONICET. Isabel M. Rodrigues (Email: irodrig@math.ist.utl.pt) is Professor at the Departamento de Matemática, Instituto Superior Técnico, Technical University of Lisbon (TULisbon), Avda. Rovisco Pais 1, 1049-001 Lisboa, Portugal, she is also a researcher at the CEMAT (Tel./FAX: 351-21-8417045/48).

1 Introduction

The generalized linear model (McCullagh and Nelder, 1989) is a popular technique for modelling a wide variety of data and assumes that the observations $(y_i, \mathbf{x}_i^{\mathrm{T}})$, $1 \leq i \leq n$, $\mathbf{x}_i \in \mathbb{R}^k$, are independent with the same distribution as $(y, \mathbf{x}^{\mathrm{T}}) \in \mathbb{R}^{k+1}$ such that the conditional distribution of $y|\mathbf{x}$ belongs to the canonical exponential family

$$\exp\left\{\left[y\theta(\mathbf{x}) - B\left(\theta(\mathbf{x})\right)\right]/A(\tau) + C(y,\tau)\right\} ,$$

for known functions A, B and C. In this situation, if we denote by B' the derivative of B, the mean $\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) = B'\left(\theta(\mathbf{x})\right)$ is modelled linearly through a known link function, g, i.e., $g(\mu\left(\mathbf{x}\right)) = \theta(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$. Robust procedures for generalized linear models have been considered, among others, by Stefanski *et al.* (1986), Künsch *et al.* (1989), Bianco and Yohai (1996), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2002) and Bianco *et al.* (2005), see also, Maronna *et al.* (2006). Recently, robust tests for the regression parameter under a logistic model were considered by Bianco and Martínez (2009).

In practice, some response variables may be missing by design (as in two-stage studies) or by happenstance. As it is well known, the methods proposed by the above mentioned authors are designed for complete data sets and problems arise when missing observations are present. Even if there are many situations in which both the response and the explanatory variables are missing, we will focus our attention on those cases in which missing data occur only in the responses. Actually, missingness of responses is very common in opinion polls, market research surveys, mail enquiries, social-economic investigations, medical studies and other scientific experiments, when the explanatory variables can be controlled. This pattern appears, for example, in the scheme of double sampling proposed by Neyman (1938), where first a complete sample is obtained and then some additional covariate values are computed since perhaps this is less expensive than to obtain more response values. Hence, we will focus our attention on robust inference when the response variable has missing observations but the covariate $\mathbf{x}$ is totally observed.

In this paper, we consider the robust estimators for the regression parameter $\boldsymbol{\beta}$ introduced by Bianco *et al.* (2010), under a GLM model. When there are no missing data, these estimators include the family of estimators previously studied by several authors such as Bianco and Yohai (1996), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2002) and Bianco *et al.* (2005). It is shown that the robust estimates of $\boldsymbol{\beta}$ are asymptotically normally distributed which allows to construct a robust procedure to test the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. The paper is organized as follows. The robust proposal is given in Section 2, the asymptotic distribution of the regression estimators and a robust Wald test for the regression parameter are provided in Section 3, while an expression for the influence function of the test is obtained in Section 4. The results of a Monte Carlo study are summarized in Section 5. A real data example is given in Section 6. Proofs are relegated to the Appendix.

2 Robust inference

2.1 Preliminaries: The robust estimators

Suppose we obtain a random sample of incomplete data $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i\right)$, $1 \leq i \leq n$, of a generalized linear model where $\delta_i = 1$ if y_i is observed, $\delta_i = 0$ if y_i is missing and $(y_i, \mathbf{x}_i^{\mathrm{T}}) \in \mathbb{R}^{k+1}$ are such that $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ with $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ and $\textsc{Var}(y_i|\mathbf{x}_i) = A^2(\tau)V^2(\mu_i) = A^2(\tau)B''\left(\theta(\mathbf{x}_i)\right)$ with B'' the second derivative of B. Let $(\boldsymbol{\beta}, \tau)$ denote the true parameter values and $\mathbb{E}_F$ the expectation under the true model, thus $\mathbb{E}_F(y|\mathbf{x}) = H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$. In a more general situation, we will think of τ as a nuisance parameter such as the tuning constant for the score function to be considered.

Let $(y, \mathbf{x}^{\mathrm{T}}, \delta)$ be a random vector with the same distribution as $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i\right)$. Bianco *et al.* (2010) defined robust estimators of the regression parameter when missing responses occur under an ignorable missing mechanism. To be more precise, they assumed that y is missing at random (MAR), that is, δ and y are conditionally independent given $\mathbf{x}$, i.e.,

$$P\left(\delta = 1|(y, \mathbf{x})\right) = P\left(\delta = 1|\mathbf{x}\right) = p\left(\mathbf{x}\right) \ . \tag{1}$$

A common assumption in the literature states that $\inf_{\mathbf{x}} p\left(\mathbf{x}\right) > 0$, meaning that at any value of the covariate response variables are observed. By introducing a weight function with compact support, this assumption will be relaxed.

For the sake of completeness, we remind the definition of the two families of estimators considered in Bianco *et al.* (2010). Let $w_1 : \mathbb{R}^k \to \mathbb{R}$ be a weight function to control leverage points on the carriers $\mathbf{x}$ and $\rho : \mathbb{R}^3 \to \mathbb{R}$ a loss function. For any $\mathbf{b} \in \mathbb{R}^k$, $t \in \mathbb{R}$ and any function $q : \mathbb{R}^k \to \mathbb{R}$, let us define

$$\begin{aligned}
S_n(\mathbf{b}, t) &= \frac{1}{n}\sum_{i=1}^{n} \delta_i \rho\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x}_i) \ , && (2)\\
S(\mathbf{b}, t) &= \mathbb{E}_F\left[\delta \rho\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x})\right] = \mathbb{E}_F\left[p(\mathbf{x})\rho\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x})\right] \ , && (3)\\
S_{\mathrm{P},n}(\mathbf{b}, t, q) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{q(\mathbf{x}_i)}\rho\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x}_i) \ , && (4)\\
S_{\mathrm{P}}(\mathbf{b}, t, q) &= \mathbb{E}_F\left[\frac{\delta}{q(\mathbf{x})}\rho\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x})\right] = \mathbb{E}_F\left[\frac{p(\mathbf{x})}{q(\mathbf{x})}\rho\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x})\right] \ . && (5)
\end{aligned}$$

For both $S(\mathbf{b}, \tau)$ and $S_{\mathrm{P}}(\mathbf{b}, \tau, q)$, τ plays the role of a nuisance parameter. Besides, for $S_{\mathrm{P}}(\mathbf{b}, \tau, q)$ q plays also the role of a nuisance parameter. Moreover, it is worth noticing that $S_{\mathrm{P}}(\mathbf{b}, t, p) = \mathbb{E}_F\left[\rho\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x})\right]$, i.e., it corresponds to the objective function when the sample contains no missing responses.

In order to define Fisher–consistent estimators, Bianco *et al.* (2010) assumed that $w_1(\cdot)$ and $\rho(\cdot)$ are such that, $S(\boldsymbol{\beta}, \tau) = \min_{\mathbf{b}} S(\mathbf{b}, \tau)$. Moreover, they also assumed that $\boldsymbol{\beta}$ is the unique minimum of $S_{\mathrm{P}}(\mathbf{b}, \tau, p)$.

Two families of estimators were defined therein as follows.

1. **The robust simplified estimators.** Let $\widehat{\tau} = \widehat{\tau}_n$ be robust consistent estimators of τ, the *robust simplified estimator* $\widehat{\boldsymbol{\beta}}$ of the regression parameter is defined as

$$\widehat{\boldsymbol{\beta}} \quad = \quad \operatorname*{argmin}_{\mathbf{b}} S_n(\mathbf{b}, \widehat{\tau}) \ . \tag{6}$$

   When ρ is continuously differentiable, if we denote by $\Psi\,(y,u,t) = \partial\rho(y,u,t)/\partial u$, $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$ satisfy the differentiated equations $S^{(1)}(\boldsymbol{\beta}, \tau) = \mathbf{0}_k$ and $S_n^{(1)}(\mathbf{b}, \widehat{\tau}) = \mathbf{0}_k$, respectively, where

$$\begin{aligned} S^{(1)}(\mathbf{b}, t) &= \mathbb{E}_F \left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x}) p(\mathbf{x}) \mathbf{x} \right) \ , \\ S_n^{(1)}(\mathbf{b}, t) &= \frac{1}{n} \sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x}_i) \mathbf{x}_i \ . \end{aligned}$$

2. **The robust propensity score estimator.** To improve the bias caused in the estimation by the missing mechanism, robust propensity score estimators may be considered using an estimator of the missingness probability. Denote by $\widehat{p}(\mathbf{x})$ any consistent estimator of $p(\mathbf{x})$. For instance, if we assume that the missingness probability is given by the logistic model, i.e., that $p(\mathbf{x}) = G_{\mathrm{L}}(\mathbf{x}^{\mathrm{T}}\boldsymbol{\lambda}_0)$, where $G_{\mathrm{L}}(s) = (1 + e^{-s})^{-1}$ is the logistic distribution function, we only need to estimate the parameter $\boldsymbol{\lambda}_0$ to define the estimator $\widehat{p}(\mathbf{x})$. The *robust propensity score estimator* $\widehat{\boldsymbol{\beta}}_{\mathrm{P}}$ is defined as

$$\widehat{\boldsymbol{\beta}}_{\mathrm{P}} \quad = \quad \operatorname*{argmin}_{\mathbf{b}} S_{\mathrm{P},n}(\mathbf{b}, \widehat{\tau}_{\mathrm{P}}, \widehat{p}) \ , \tag{7}$$

   where $\widehat{\tau}_{\mathrm{P}}$ is a robust consistent estimator of τ, possible different than the one previously considered. As above, when ρ is continuously differentiable, if we denote by $\Psi\,(y,u,t) = \partial\rho(y,u,t)/\partial u$, $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$ satisfy the differentiated equations $S_{\mathrm{P}}^{(1)}(\boldsymbol{\beta}, \tau, p) = \mathbf{0}_k$ and $S_{\mathrm{P},n}^{(1)}(\mathbf{b}, \widehat{\tau}, \widehat{p}) = \mathbf{0}_k$, respectively, where

$$\begin{aligned} S_{\mathrm{P}}^{(1)}(\mathbf{b}, t, q) &= \mathbb{E}_F \left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} \mathbf{x} \right) \ , \\ S_{\mathrm{P},n}^{(1)}(\mathbf{b}, t, q) &= \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{q(\mathbf{x}_i)} \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right) w_1(\mathbf{x}_i) \mathbf{x}_i \ . \end{aligned}$$

It is worth noticing that even if the *robust propensity score estimators* adapt to the situation when no missing responses arise, they may be more sensitive than the *simplified estimators* to values of the covariates $\mathbf{x}$ leading to small values of the missingness probability, i.e., to values where less responses arise. If for such a value of the covariate, an outlier occur in the response, its effect will be enlarged due to the estimator $\widehat{p}(\mathbf{x})$. The weight function w_1 may try to overcome this problem by controlling not only high leverage points but also values of $\mathbf{x}$ leading to small values of $\widehat{p}(\mathbf{x})$.

Remark 2.1.1. Two classes of loss functions ρ have been considered in the literature. The first one aims to bound the deviances, while the second one introduced by Cantoni and Ronchetti (2001) bounds the Pearson residuals. For a complete description see Bianco *et al.* (2010). In particular, the Poisson and log–Gamma model were considered therein. We refer to Bianco *et al.* (2005) for a description on the robust estimators based on deviances for complete data sets and to Heritier *et al.* (2009) for a description on $M-$type estimators for the log–Gamma model. For the sake of completeness, we will remind how to adapt the estimators based on deviances to the situation with missing responses since this will be the model used in our simulation study.

Denote by $d_i(\boldsymbol{\beta},\tau)$ the deviance component of the i-th observation, i.e., $d_i(\boldsymbol{\beta},\tau) = 2\tau\, d^*(y_i,\mathbf{x}_i,\boldsymbol{\beta})$ where

$$d^*(y,\mathbf{x},\boldsymbol{\beta}) = -1 - (\log(y) - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) + y\exp(-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}).$$

Let us now assume that we are dealing with the situation in which some of the responses y_i, and so the transformed responses $z_i = \log(y_i)$ may be missing. Indeed, we have that $\delta_i = 1$ if z_i is observed, while $\delta_i = 0$ if z_i is missing and $(z_i, \mathbf{x}_i^{\mathrm{T}}) \in \mathbb{R}^{k+1}$ are such that $z_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + u_i$, where $u_i \sim \log(\Gamma(\tau,1))$ and u_i and $\mathbf{x}_i$ are independent. Moreover, δ and z are conditionally independent given $\mathbf{x}$ and so δ and u are independent. Besides, the density of u is $g(u,\tau)$, where

$$g(u,\tau) = \frac{\tau^{\tau}}{\Gamma(\tau)}\ \exp(\tau(u - \exp(u)))\ , \tag{8}$$

is asymmetric and unimodal with maximum at $u_0 = 0$. Note that $d^*(y,\mathbf{x},\mathbf{b}) = -1-u-\mathbf{x}^{\mathrm{T}}(\boldsymbol{\beta}-\mathbf{b}) + \exp(u)\exp(\mathbf{x}^{\mathrm{T}}(\boldsymbol{\beta}-\mathbf{b})) = \widetilde{d}(u,\mathbf{x},\boldsymbol{\beta}-\mathbf{b})$. The maximum likelihood (ML) estimator of $\boldsymbol{\beta}$ is, thus, obtained as

$$\widehat{\boldsymbol{\beta}}_{\mathrm{ML}} = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{i=1}^{n} \delta_i d^*(y_i,\mathbf{x}_i,\mathbf{b}).$$

As described in Bianco *et al.* (2010) a three step procedure can be considered to compute a sort of generalized $MM-$estimators when missing responses are present.

- **Step 1.** We first compute an initial $S-$estimate $\widetilde{\boldsymbol{\beta}}_n$ of the regression parameter and the corresponding scale estimate $\widehat{\sigma}_n$, taking $b = \frac{1}{2}\sup\rho$ with the complete data set. To be more precise, for each value of $\mathbf{b}$ let $\sigma_n(\mathbf{b})$ be the $M-$scale estimate of $\sqrt{d^*(y_i,\mathbf{x}_i,\mathbf{b})}$ given by

$$\frac{1}{\sum_{i=1}^{n}\delta_i}\sum_{i=1}^{n}\delta_i\phi\left(\frac{\sqrt{d^*(y_i,\mathbf{x}_i,\mathbf{b})}}{\sigma_n(\mathbf{b})}\right) = b,$$

  where ϕ is Tukey's bisquare function. The $S-$estimate of $\boldsymbol{\beta}$ for the considered model is defined as $\widetilde{\boldsymbol{\beta}}_n = \operatorname{argmin}_{\mathbf{b}}\ \sigma_n(\mathbf{b})$ and the corresponding scale estimate as $\widehat{\sigma}_n = \min_{\mathbf{b}}\ \sigma_n(\mathbf{b})$.

  Let u be a random variable with density (8) and write $\sigma^*(\tau)$ the solution of

$$\mathbb{E}_G\left[\rho\left(\frac{\sqrt{h(u)}}{\sigma^*(\tau)}\right)\right] = b,$$

where $h(u) = 1 - u - \exp(u)$. Under mild conditions, $\widetilde{\boldsymbol\beta}_n \xrightarrow{a.s.} \boldsymbol\beta$ and $\widehat{\sigma}_n \xrightarrow{a.s.} \sigma^*(\tau)$, besides $\sigma^*(\tau)$ is a continuous and strictly decreasing function and so, an estimator of τ can be defined as $\widehat{\tau}_n = \sigma^{*-1}(\widehat{\sigma}_n)$ leading to a a strongly consistent estimator for τ.

- **Step 2.** In the second step, we compute $\widehat{\tau}_n = \sigma^{*-1}(\widehat{\sigma}_n)$ and

$$\widehat{c}_n = \max(\widehat{\sigma}_n, C_e(\widehat{\tau}_n)) = \max(\widehat{\sigma}_n, C_e(\sigma^{*-1}(\widehat{\sigma}_n)) .$$

We then have that $\widehat{c}_n \xrightarrow{p} c_0 = \max\{\sigma^*(\tau), C_e(\tau)\}$.

- **Step 3.** Let $\widehat{\boldsymbol\beta}_n$ be the adaptive estimator of $\boldsymbol\beta$ defined by

$$\widehat{\boldsymbol\beta}_n = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{i=1}^{n} \delta_i \phi\left(\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}/\widehat{c}_n\right) w_1(\mathbf{x}_i) . \tag{9}$$

The results stated in Section 3 allow to show that since $\widehat{c}_n \xrightarrow{p} c_0$

$$\sqrt{n}(\widehat{\boldsymbol\beta}_n - \boldsymbol\beta) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_k, \frac{B(\psi, \tau, c_0)}{A^2(\phi, \tau, c_0)}\mathbf{C}\right) ,$$

where $\mathbf{C} = \mathbf{A}^{-1}\mathbb{E}\left(p(\mathbf{x})w_1^2(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\mathbf{A}^{-1}$ with $\mathbf{A} = \mathbb{E}\left(p(\mathbf{x})w_1(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)$. Note that when a $0-1$ weight function is considered the asymptotic matrix $\mathbf{C}$ reduces to $\mathbf{C} = \mathbb{E}\left(p(\mathbf{x})w_1(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)^{-1}$.

For missing responses, the asymptotic relative efficiency of $\widehat{\boldsymbol\beta}_n$ depends on the asymptotic efficiency for the complete data set $\mathrm{ARE}(\phi, \tau, c_0)$ and on the matrix $\mathbf{C}$. For the sake of simplicity, in the simulation study we have calibrated the estimators for $p \equiv 1$ and $w_1 \equiv 1$ and so, an extra loss of efficiency should be expected.

Propensity score estimators $\widehat{\boldsymbol\beta}_{\mathrm{P},n}$ can be defined similarly. In this case, results stated in Section 3 allow to show that

$$\sqrt{n}(\widehat{\boldsymbol\beta}_{\mathrm{P},n} - \boldsymbol\beta) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_k, \frac{B(\psi, \tau, c_0)}{A^2(\phi, \tau, c_0)}\mathbf{C}_{\mathrm{P}}\right)$$

where

$$\mathbf{C}_{\mathrm{P}} = \mathbf{A}_{\mathrm{P}}^{-1}\mathbb{E}\left(\frac{w_1^2(\mathbf{x})}{p(\mathbf{x})}\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\mathbf{A}_{\mathrm{P}}^{-1}$$

with $\mathbf{A}_{\mathrm{P}} = \mathbb{E}\left(w_1(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)$.

2.2 Test statistics

In Section 3 it will be shown that, if $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i\right)$, $1 \le i \le n$ are as described above, under mild conditions,

$$\sqrt{n}\left(\widehat{\boldsymbol\beta} - \boldsymbol\beta\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_k, \boldsymbol\Sigma\right) \quad \text{and} \quad \sqrt{n}\left(\widehat{\boldsymbol\beta}_{\mathrm{P}} - \boldsymbol\beta\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_k, \boldsymbol\Sigma_{\mathrm{P}}\right) ,$$

where $\mathbf{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$, $\mathbf{\Sigma}_{\mathrm{P}} = \mathbf{A}_{\mathrm{P}}^{-1}\mathbf{B}_{\mathrm{P}}\mathbf{A}_{\mathrm{P}}^{-1}$ and the symmetric matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}_{\mathrm{P}}$ and $\mathbf{B}_{\mathrm{P}}$ are defined as

$$\begin{aligned}
\mathbf{A} &= \mathbb{E}_F\left(\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x}) p(\mathbf{x}) \mathbf{x}\mathbf{x}^{\mathrm{T}}\right) &(10)\\
\mathbf{B} &= \mathbb{E}_F\left(\Psi^2\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1^2(\mathbf{x}) p(\mathbf{x}) \mathbf{x}\mathbf{x}^{\mathrm{T}}\right) &(11)\\
\mathbf{A}_{\mathrm{P}} &= \mathbb{E}_F\left(\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x}) \mathbf{x}\mathbf{x}^{\mathrm{T}}\right) &(12)\\
\mathbf{B}_{\mathrm{P}} &= \mathbb{E}_F\left(\Psi^2\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1^2(\mathbf{x}) p(\mathbf{x})^{-1} \mathbf{x}\mathbf{x}^{\mathrm{T}}\right), &(13)
\end{aligned}$$

with $\chi(y, u, \tau) = \partial\Psi(y, u, \tau)/\partial u$. Estimators of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}_{\mathrm{P}}$ and $\mathbf{B}_{\mathrm{P}}$ can be obtained through their sample versions. So, let us define $\widehat{\mathbf{A}} = \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widehat{\tau})$, $\widehat{\mathbf{B}} = \widehat{\mathbf{B}}(\widehat{\boldsymbol{\beta}}, \widehat{\tau})$, $\widehat{\mathbf{A}}_{\mathrm{P}} = \widehat{\mathbf{A}}_{\mathrm{P}}(\widehat{\boldsymbol{\beta}}, \widehat{\tau}, \widehat{p})$ and $\widehat{\mathbf{B}}_{\mathrm{P}} = \widehat{\mathbf{B}}_{\mathrm{P}}(\widehat{\boldsymbol{\beta}}, \widehat{\tau}, \widehat{p})$ where

$$\begin{aligned}
\widehat{\mathbf{A}}(\mathbf{b}, t) &= \frac{1}{n}\sum_{i=1}^{n} \delta_i\, \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right)\, w_1(\mathbf{x}_i)\, \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\\
\widehat{\mathbf{B}}(\mathbf{b}, t) &= \frac{1}{n}\sum_{i=1}^{n} \delta_i\, \Psi^2\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right)\, w_1^2(\mathbf{x}_i)\, \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\\
\widehat{\mathbf{A}}_{\mathrm{P}}(\mathbf{b}, t, q) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{q(\mathbf{x}_i)}\, \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right)\, w_1(\mathbf{x}_i)\, \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\\
\widehat{\mathbf{B}}_{\mathrm{P}}(\mathbf{b}, t, q) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{q^2(\mathbf{x}_i)}\, \Psi^2\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, t\right)\, w_1^2(\mathbf{x}_i)\, \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}.
\end{aligned}$$

Two Wald–type test statistics to test the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ can thus be defined as $\widehat{\mathcal{W}}_n = n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}\widehat{\mathbf{\Sigma}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $\widehat{\mathcal{W}}_{\mathrm{P},n} = n(\widehat{\boldsymbol{\beta}}_{\mathrm{P}} - \boldsymbol{\beta}_0)^{\mathrm{T}}\widehat{\mathbf{\Sigma}}_{\mathrm{P}}^{-1}(\widehat{\boldsymbol{\beta}}_{\mathrm{P}} - \boldsymbol{\beta}_0)$. These test statistics will be asymptotically χ^2_k distributed under the null hypothesis. Their asymptotic behaviour under contiguous alternatives is derived in Section 3. Also, a score type test as defined in Heritier and Ronchetti (1994) when there are no missing responses in the sample, can be adapted to the present situation.

3 Asymptotic behaviour of the test statistics

In this section, we will derive the asymptotic distribution of the test statistics under the null hypothesis and under contiguous alternatives. We will consider the following set of assumptions.

N1. The functions $w_1(\mathbf{x})$ and $w_1(\mathbf{x})\|\mathbf{x}\|$ are bounded.

N2. $\mathbb{E}_F\left(p(\mathbf{x})w_1(\mathbf{x})\|\mathbf{x}\|^2\right) < \infty$.

N3. $\Psi(y, u, v)$ and $\chi(y, u, v) = \partial\Psi(y, u, v)/\partial u$ are bounded continuous functions.

N4. The matrix $\mathbf{A}$ defined in (10) is non–singular.

N5. The class of functions $\mathcal{F} = \{f_\tau(y, \mathbf{x}, \delta) = \delta\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x})\mathbf{x}, \tau \in \mathcal{K}\}$ where $\mathcal{K}$ is a compact neighbourhood of τ, has finite entropy.

N6. $\mathbb{E}_F\left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right)|\mathbf{x}\right) = \mathbf{0}_k$ for any fixed $\tau \in \mathcal{K}$.

N7. $\inf_{\mathbf{x}\in\mathcal{S}_{w_1}\cap\mathcal{S}_{\mathbf{x}}} p(\mathbf{x}) = A > 0$, where $\mathcal{S}_{w_1}$ and $\mathcal{S}_{\mathbf{x}}$ stand for the support of w_1 and $\mathbf{x}$, respectively.

N8. $p(\mathbf{x}) = G(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x})$ for some continuous function $G : \mathbb{R} \to (0, 1)$ with bounded variation.

N9. The matrix $\mathbf{A}_{\mathrm{P}}$ defined in (12) is non–singular.

Remark 3.1. Assumptions **N1** and **N3** are standard requirements since they state that the weight function control large values of the covariates and that the score function bound large residuals, respectively. **N2** is fulfilled for instance, for a $0-1$ weight function and more generally, if $w_1(\mathbf{x})\|\mathbf{x}\|^2$ is bounded. Note that **N6** holds for the usual functions considered in robustness, it is the conditional Fisher–consistency defined by Künsch *et al.* (1989). Moreover, **N5** is fulfilled for the family of functions studied in Bianco *et al.* (2005), when τ plays the role of the tuning constant, and $\rho(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau) = \rho(\sqrt{d^*(y, \mathbf{x}, \boldsymbol{\beta})}/\tau)$ if ρ is twice continuously differentiable and there exists M such that $|u| > M$ implies that $\rho(u) = \sup_v \rho(v)$. Assumptions **N4** and **N9** are standard conditions in the robustness literature to guarantee that the regression estimators will be root$-n$ consistent. Note that if w_1 has compact support, as it is the case for the Tukey weight function used in the simulation study, **N7** holds for any continuous missingness probability such that $p(\mathbf{x}) > 0$. This includes, for instance, a logistic model for $p(\mathbf{x})$. On the other hand, if $\mathcal{S}_{\mathbf{x}} = \mathbb{R}^k$ and $w_1 \equiv 1$, i.e., if high leverage points are not downweighted, **N7** restricts the family of missing probabilities to be considered.

The following Lemma will be useful when deriving the asymptotic distribution of the robust simplified and robust propensity estimators defined in Section 2.1. We omit its proof since it follows using analogous arguments to those considered in Lemma 1 of Bianco and Boente (2002).

Lemma 3.1. *Assume that $(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i)$ satisfy (1) and are such that $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ where $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)$. Assume that **N1** and **N2** hold. Let $\varphi(y, u, v)$ be a continuous function and assume that $\widetilde{\tau} \xrightarrow{p} \tau$ and $\widetilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. Define*

$$\mathbf{V} = \mathbb{E}_F\left(\varphi(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau)w_1(\mathbf{x})p(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) \qquad \widehat{\mathbf{V}} = \frac{1}{n}\sum_{i=1}^{n} \delta_i\varphi(y_i, \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\tau})w_1(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} \,,$$

$$\mathbf{V}_{\mathrm{P}} = \mathbb{E}_F\left(\varphi(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau)w_1(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) \qquad \widehat{\mathbf{V}}_{\mathrm{P}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\widehat{p}(\mathbf{x}_i)}\varphi(y_i, \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}, \widetilde{\tau})w_1(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} \,.$$

Then, we have that, $\widehat{\mathbf{V}} \xrightarrow{p} \mathbf{V}$. Moreover, assume that either

i) ***N7*** *holds and* $\sup_{\mathbf{x}\in\mathcal{S}_{w_1}\cap\mathcal{S}_{\mathbf{x}}} |\widehat{p}(\mathbf{x}) - p(\mathbf{x})| \xrightarrow{p} 0$
or

ii) ***N8*** *holds,* $\widehat{p}(\mathbf{x}) = G(\widetilde{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{x})$ *where* $\widetilde{\boldsymbol{\lambda}} \xrightarrow{p} \boldsymbol{\lambda}$*,*

then, $\widehat{\mathbf{V}}_{\mathrm{P}} \xrightarrow{p} \mathbf{V}$ *holds.*

Theorem 3.1. *Assume that* $(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i)$ *satisfy (1) and are such that* $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ *where* $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n)$*, with* $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$*. Assume that* ***N1*** *to* ***N6*** *hold and that* $\widehat{\tau}_n \xrightarrow{p} \tau$*. Let* $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ *where the symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are defined in (10) and (11), respectively. Then, we have that*

a) Under $H_0 : \mathbf{c} = \mathbf{0}_k$*, i.e., under* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$*,* $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_k, \boldsymbol{\Sigma}\right)$*.*

b) Under $H_{1,n} : \mathbf{c} \neq \mathbf{0}_k$*, i.e., under* $H_{1,n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$*, if* $\mathbb{E}\left(|H'(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)|\,\|\mathbf{x}\|^2\right) < \infty$*,* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N\left(-\mathbf{c}, \boldsymbol{\Sigma}\right)$*. Moreover,* $\widehat{\mathbf{B}}^{-1/2}\widehat{\mathbf{A}}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N\left(-\mathbf{B}^{-1/2}\mathbf{A}\mathbf{c}, \boldsymbol{I}_k\right)$ *where* $\widehat{\mathbf{A}}$ *and* $\widehat{\mathbf{B}}$ *are defined in Section 2.2.*

We omit the proof of the following result since it follows using analogous arguments to those considered in the proof of Theorem 3.1 using that **N8** implies that the class of functions $\mathcal{P} = \{p(\mathbf{x}) = p_{\boldsymbol{\lambda}}(\mathbf{x}) = G(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}), \boldsymbol{\lambda} \in \mathbb{R}^k\}$ has finite entropy.

Theorem 3.2. *Assume that* $(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i)$ *satisfy (1) and are such that* $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ *where* $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n)$*, with* $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$*. Consider* $\widehat{\boldsymbol{\beta}}_{\mathrm{P}}$ *the propensity estimator defined through (7) with* $\widehat{p}(\mathbf{x}) = G(\widehat{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{x})$ *such that* $\widehat{\boldsymbol{\lambda}} \xrightarrow{p} \boldsymbol{\lambda}$*. Assume that* ***N1*** *to* ***N3*** *and* ***N5*** *to* ***N9*** *hold and that* $\widehat{\tau}_n \xrightarrow{p} \tau$*. Let* $\boldsymbol{\Sigma}_{\mathrm{P}} = \mathbf{A}_{\mathrm{P}}^{-1}\mathbf{B}_{\mathrm{P}}\mathbf{A}_{\mathrm{P}}^{-1}$ *where the symmetric matrices* $\mathbf{A}_{\mathrm{P}}$ *and* $\mathbf{B}_{\mathrm{P}}$ *are defined in (12) and (13), respectively. Then, we have that*

a) Under $H_0 : \mathbf{c} = \mathbf{0}_k$*, i.e., under* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$*,* $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{P}} - \boldsymbol{\beta}_0\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_k, \boldsymbol{\Sigma}_{\mathrm{P}}\right)$*.*

b) Under $H_{1,n} : \mathbf{c} \neq \mathbf{0}_k$*, i.e., under* $H_{1,n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$*, if* $\mathbb{E}\left(|H'(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)|\,\|\mathbf{x}\|^2\right) < \infty$*,* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\mathrm{P}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N\left(-\mathbf{c}, \boldsymbol{\Sigma}_{\mathrm{P}}\right)$*. Moreover,* $\widehat{\mathbf{B}}_{\mathrm{P}}^{-1/2}\widehat{\mathbf{A}}_{\mathrm{P}}\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\mathrm{P}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N\left(-\mathbf{B}_{\mathrm{P}}^{-1/2}\mathbf{A}_{\mathrm{P}}\mathbf{c}, \boldsymbol{I}_k\right)$ *where* $\widehat{\mathbf{A}}_{\mathrm{P}}$ *and* $\widehat{\mathbf{B}}_{\mathrm{P}}$ *are defined in Section 2.2.*

The following Theorems state the asymptotic behaviour of the proposed Wald–type test statistics and their proofs follow easily applying Theorem 3.1, 3.2 and Lemma 3.1.

Theorem 3.3. *Assume that* $(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i)$ *satisfy (1) and are such that* $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ *where* $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n)$*, with* $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$*. Let* $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ *where the symmetric matrices*

$\mathbf{A}$ *and* $\mathbf{B}$ *are defined in (10) and (11), respectively. Assume that* ***N1*** *to* ***N6*** *hold and that* $\widehat{\tau}_n \xrightarrow{p} \tau$*, then we have that*

a) *Under* $H_0 : \mathbf{c} = \mathbf{0}_k$*, i.e., under* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\widehat{\mathcal{W}}_n \xrightarrow{\mathcal{D}} \chi^2_k$.

b) *Under* $H_{1,n} : \mathbf{c} \neq \mathbf{0}_k$*, i.e., under* $H_{1,n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$*, if* $\mathbb{E}\left(|H'(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)|\,\|\mathbf{x}\|^2\right) < \infty$, $\widehat{\mathcal{W}}_n \xrightarrow{\mathcal{D}} \chi^2_k(\theta)$*, where* $\theta = \mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{c}$.

Theorem 3.4. *Assume that* $(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i)$ *satisfy (1) and are such that* $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ *where* $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n)$*, with* $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$*. Consider* $\widehat{\boldsymbol{\beta}}_{\mathrm{P}}$ *be the propensity estimator defined through (7) with* $\widehat{p}(\mathbf{x}) = G(\widehat{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{x})$ *such that* $\widehat{\boldsymbol{\lambda}} \xrightarrow{p} \boldsymbol{\lambda}$*. Assume that* ***N1*** *to* ***N3*** *and* ***N5*** *to* ***N9*** *hold and that* $\widehat{\tau}_n \xrightarrow{p} \tau$*. Let* $\boldsymbol{\Sigma}_{\mathrm{P}} = \mathbf{A}_{\mathrm{P}}^{-1}\mathbf{B}_{\mathrm{P}}\mathbf{A}_{\mathrm{P}}^{-1}$ *where the symmetric matrices* $\mathbf{A}_{\mathrm{P}}$ *and* $\mathbf{B}_{\mathrm{P}}$ *are defined in (12) and (13), respectively. Then, we have that*

a) *Under* $H_0 : \mathbf{c} = \mathbf{0}_k$*, i.e., under* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\widehat{\mathcal{W}}_{\mathrm{P},n} \xrightarrow{\mathcal{D}} \chi^2_k$.

b) *Under* $H_{1,n} : \mathbf{c} \neq \mathbf{0}_k$*, i.e., under* $H_{1,n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$*, if* $\mathbb{E}\left(|H'(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)|\,\|\mathbf{x}\|^2\right) < \infty$, $\widehat{\mathcal{W}}_{\mathrm{P},n} \xrightarrow{\mathcal{D}} \chi^2_k(\theta_{\mathrm{P}})$*, where* $\theta_{\mathrm{P}} = \mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{P}}^{-1}\mathbf{c}$.

From Theorems 3.3 and 3.4, to test the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ at a given asymptotic level α, the following consistent tests can be used

- Reject H_0 if $\widehat{\mathcal{W}}_n > \chi^2_{k,\alpha}$

  or

- Reject H_0 if $\widehat{\mathcal{W}}_{\mathrm{P},n} > \chi^2_{k,\alpha}$.

In regression, one of the most frequent hypothesis testing problems involves only a subset of the regression parameter. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^{\mathrm{T}}, \boldsymbol{\beta}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$, $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_{(1)}^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{x} = (\mathbf{x}_{(1)}^{\mathrm{T}}, \mathbf{x}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\beta}_{(1)} \in \mathbb{R}^{k_1}$ with $k_1 < k$. In order to test $H_{0\,\boldsymbol{\beta}_{(1)}} : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$, $\boldsymbol{\beta}_{(2)}$ unspecified, one may use the statistic

$$\widehat{\mathcal{W}}_{1,n} = n(\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})^{\mathrm{T}}\, \widehat{\boldsymbol{\Sigma}}_{11}^{-1} (\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})\,, \tag{14}$$

where $\widehat{\boldsymbol{\Sigma}}_{11}$ denotes the $k_1 \times k_1$ submatrix of $\widehat{\boldsymbol{\Sigma}}$, corresponding to the coordinates of $\boldsymbol{\beta}_{(1)}$.

Theorem 3.5. *Assume that* $(y_i, \mathbf{x}_i^{\mathrm{T}}, \delta_i)$ *satisfy (1) and are such that* $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ *where* $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n)$*, with* $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$ *with* $\mathbf{c} = (\mathbf{c}_{(1)}^{\mathrm{T}}, \mathbf{0}_{k-k_1}^{\mathrm{T}})^{\mathrm{T}}$*. Assume that* ***N1*** *to* ***N7*** *hold and that* $\widehat{\tau}_n \xrightarrow{p} \tau$*. Denote by* $\boldsymbol{\Sigma}_{11}$ *the* $k_1 \times k_1$ *submatrix of* $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$*, corresponding to the coordinates of* $\boldsymbol{\beta}_{(1)}$ *where the symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are defined in (10) and (11), respectively. Then, we have that* *a) Under* $H_{0\boldsymbol{\beta}_{(1)}} : \mathbf{c}_{(1)} = \mathbf{0}_{k_1}$*, i.e., under* $H_0 : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$*,* $\widehat{\mathcal{W}}_{1,n} \xrightarrow{\mathcal{D}} \chi^2_{k_1}$*.*

b) Under $H_{1\boldsymbol{\beta}_{(1)},n} : \mathbf{c}_{(1)} \neq \mathbf{0}_{k_1}$*, i.e., under* $H_{1\boldsymbol{\beta}_{(1)},n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$*, if* $\mathbb{E}\left(|H'(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)|\,\|\mathbf{x}\|^2\right) < \infty$*,* $\widehat{\mathcal{W}}_{1,n} \xrightarrow{\mathcal{D}} \chi^2_k(\theta)$*, where* $\theta = \mathbf{c}_{(1)}^{\mathrm{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{c}_{(1)}$*.*

An analogous result can be obtained for the propensity score test.

4 Influence functions of the test functionals

Influence functions are measures of robustness with respect to single outliers. The influence functions allows us to study the local robustness and the asymptotic efficiency of the estimators, providing a rationale for choosing appropriate weight functions and tuning parameters. It can be thought as the first derivative of the functional version of the estimator. The influence function of a functional $T(F)$ is defined as:

$$\mathrm{IF}(\mathbf{z}_0, T, F) = \lim_{\epsilon \to 0} \frac{T(F_{\mathbf{z}_0,\epsilon}) - T(F)}{\epsilon}\,, \qquad (15)$$

where $F_{\mathbf{z}_0,\epsilon} = (1-\epsilon)F + \epsilon\Delta_{\mathbf{z}_0}$ and $\Delta_{\mathbf{z}_0}$ denotes the probability measure which puts mass 1 at the point $\mathbf{z}_0 = (y_0, \mathbf{x}_0^{\mathrm{T}}, \delta_0)$ and represents the contaminated model.

4.1 Influence function of the test functionals based on the simplified estimators

For any distribution F_1, let $\mathbf{V}(F_1)$ be a Fisher–consistent scatter functional at F, i.e., such that $\mathbf{V}(F) = \boldsymbol{\Sigma}$. Denote by $\boldsymbol{\beta}(F_1)$ and $\tau(F_1)$ the functionals related to the estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\tau}$, respectively, and assume that $\boldsymbol{\beta}(F_1)$, the solution of

$$\begin{aligned} S^{(1)}(\boldsymbol{\beta}(F_1), \tau(F_1)) &= \mathbb{E}_{F_1}\left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F_1), \tau(F_1)\right) w_1(\mathbf{x}) p(\mathbf{x})\mathbf{x}\right) \\ &= \mathbb{E}_{F_1}\left(\delta\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F_1), \tau(F_1)\right) w_1(\mathbf{x})\mathbf{x}\right) = \mathbf{0}_k\,, \end{aligned}$$

is a Fisher–consistent functional at F, i.e., $\boldsymbol{\beta}(F) = \boldsymbol{\beta}$. Define the functionals

$$\begin{aligned} \mathbf{A}(F_1) &= \mathbb{E}_{F_1}\left(\delta\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F_1), \tau(F_1)\right) w_1(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) \\ \mathbf{B}(F_1) &= \mathbb{E}_{F_1}\left(\delta\Psi^2\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F_1), \tau(F_1)\right) w_1^2(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\,. \end{aligned}$$

The Wald–type test functional related to the statistic used to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is given by

$$\mathcal{W}(F_1) = (\boldsymbol{\beta}(F_1) - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{V}(F_1)^{-1}(\boldsymbol{\beta}(F_1) - \boldsymbol{\beta}_0)\,.$$

It is easy to see that, under H_0, $\mathrm{IF}(\mathbf{x}, \mathcal{W}, F) = 0$. In order to obtain a non-null influence function, we consider the square root of the test statistics (as in Hampel *et al.*, 1986, p. 348), $\mathcal{T}(F_1) = \mathcal{W}(F_1)^{1/2}$. As for the linear model, using that, under H_0 $\mathcal{W}(F) = 0$, we have that

$$\mathrm{IF}(\mathbf{x}, \mathcal{T}, F) = \left\{\frac{1}{2}\frac{\partial^2}{\partial\epsilon^2}\mathcal{W}(F_{\mathbf{x},\epsilon})\Big|_{\epsilon=0}\right\}^{\frac{1}{2}}\,.$$

The following Theorem gives the value of the influence function of the test functional $\mathcal{T}(F)$.

Theorem 4.1. *Let $\mathbf{V}(F_1)$ be a scatter functional such that $\mathbf{V}(F) = \mathbf{\Sigma}$ where $\mathbf{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ with symmetric matrices $\mathbf{A} = \mathbf{A}(F)$ and $\mathbf{B} = \mathbf{B}(F)$ defined in (10) and (11). Assume that the influence function $IF(\mathbf{z}_0, \boldsymbol{\beta}, F)$ and that $\partial^2\boldsymbol{\beta}(F_{\mathbf{x},\epsilon})/\partial\epsilon^2\big|_{\epsilon=0}$ exist. Then, the influence function at F of the functional $\mathcal{T}(F_1)$ to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is given by*

$$IF(\mathbf{z}_0, \mathcal{T}, F)^2 \quad = \quad IF(\mathbf{z}_0, \boldsymbol{\beta}, F)^{\mathrm{T}}\mathbf{\Sigma}^{-1}IF(\mathbf{z}_0, \boldsymbol{\beta}, F)\,. \tag{16}$$

*Besides, under **N3**, **N4** and **N6**, we have that $IF(\mathbf{z}_0, \boldsymbol{\beta}, F)$ exists and if $\tau(F) = \tau$, then*

$$\begin{aligned} IF(\mathbf{z}_0, \boldsymbol{\beta}, F) \quad &= \quad -\,\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x}_0)\delta_0\{\mathbb{E}_F\left(\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x})p(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\}^{-1}\mathbf{x}_0 \\ &= \quad -\,\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x}_0)\delta_0\mathbf{A}^{-1}\mathbf{x}_0\,. \end{aligned} \tag{17}$$

Replacing (17) in (16), it is easy to see that the influence function equals $\mathrm{IF}(\mathbf{z}_0, \mathcal{T}, F)^2 = \Psi^2\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}(F), \tau(F)\right) w_1^2(\mathbf{x}_0)\delta_0\, \mathbf{x}_0^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x}_0$.

Remark 4.1. It is worth noticing that the influence function depends on the indicator of the missingness response δ_0 and so, it will be 0 if no responses arise. A more reliable function to measure the sensitivity to outliers of a given functional $T(F_1)$ under a missing scheme may be to consider the *expected influence function* given an observed data $\mathbf{z}_0^* = (y_0, \mathbf{x}_0^{\mathrm{T}})^{\mathrm{T}}$, denoted $\mathrm{EIF}(\mathbf{z}_0^*, T, F)$, i.e., $\mathrm{EIF}(\mathbf{z}_0^*, T, F) = \mathbb{E}(\mathrm{IF}(\mathbf{z}_0, T, F)|(y_0, \mathbf{x}_0))$. For the functionals under study, we have that

$$\begin{aligned} \mathrm{EIF}(\mathbf{z}_0^*, \boldsymbol{\beta}, F) \quad &= \quad -\,\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x}_0)p(\mathbf{x}_0)\mathbf{A}^{-1}\mathbf{x}_0 \\ \mathrm{EIF}(\mathbf{z}_0^*, \mathcal{T}, F)^2 \quad &= \quad \Psi^2\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}(F), \tau(F)\right) w_1^2(\mathbf{x}_0)p^2(\mathbf{x}_0)\, \mathbf{x}_0^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x}_0\,. \end{aligned}$$

When considering a test, a different measure may be to consider the *expected squared influence function* $\mathrm{EIF}_2(\mathbf{z}_0^*, \mathcal{T}, F) = \mathbb{E}(\mathrm{IF}(\mathbf{z}_0, \mathcal{T}, F)^2|(y_0, \mathbf{x}_0))$. In our case, we obtain

$$\mathrm{EIF}_2(\mathbf{z}_0^*, \mathcal{T}, F) = \Psi^2\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}(F), \tau(F)\right) w_1^2(\mathbf{x}_0)p(\mathbf{x}_0)\, \mathbf{x}_0^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x}_0\,.$$

Note that $\mathrm{EIF}(\mathbf{z}_0^*, \mathcal{T}, F) = \mathrm{EIF}_2(\mathbf{z}_0^*, \mathcal{T}, F)p(\mathbf{x})$ and so, the difference between both measures is the importance given to the missingness probability.

4.2 Influence function of the test functionals based on the propensity score estimators

Let us assume a parametric model for the probability of missing, i.e., $p(\mathbf{x}) = G(\mathbf{x}^{\mathrm{T}}\boldsymbol{\lambda})$. As in the previous section, denote by $\boldsymbol{\beta}_{\mathrm{P}}(F_1)$, $\tau_{\mathrm{P}}(F_1)$ and $\boldsymbol{\lambda}(F_1)$ the functionals related to the estimators $\widehat{\boldsymbol{\beta}}_{\mathrm{P}}$, $\widehat{\tau}_{\mathrm{P}}$ and $\widehat{\boldsymbol{\lambda}}$, respectively, where $\widehat{\boldsymbol{\beta}}_{\mathrm{P}}$ is solution of $S_{\mathrm{P},n}^{(1)}(\mathbf{b}, \widehat{\tau}_{\mathrm{P}}, \widehat{\boldsymbol{\lambda}}) = \mathbf{0}_k$ for

$$S_{\mathrm{P},n}^{(1)}(\mathbf{b}, \widehat{\tau}_{\mathrm{P}}, \widehat{\boldsymbol{\lambda}}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{p}(\mathbf{x}_i)}\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b}, \widehat{\tau}_{\mathrm{P}}\right) w_1(\mathbf{x}_i)\mathbf{x}_i\,,$$

with $\widehat{p}(\mathbf{x}) = G(\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\lambda}})$ the propensity score estimate and $\widehat{\boldsymbol{\lambda}}$ is a consistent estimator of $\boldsymbol{\lambda}$. Assume that $\boldsymbol{\beta}_{\mathrm{P}}(F_1)$ is a Fisher–consistent functional at F, i.e., $\boldsymbol{\beta}_{\mathrm{P}}(F) = \boldsymbol{\beta}$. Note that $\boldsymbol{\beta}_{\mathrm{P}}(F_1)$ is the solution of $S_{\mathrm{P}}^{(1)}(\mathbf{b}, \tau_{\mathrm{P}}(F_1), \boldsymbol{\lambda}(F_1)) = \mathbf{0}_k$ with

$$S_{\mathrm{P}}^{(1)}(\mathbf{b}, \tau, \boldsymbol{\lambda}(F_1)) = \mathbb{E}_{F_1}\left(\delta\Psi\left(y, \mathbf{x}^{\mathrm{T}}\mathbf{b}, \tau\right) w_1^*(\mathbf{x}, F_1)\mathbf{x}\right) ,$$

and $w_1^*(\mathbf{x}, F_1) = w(\mathbf{x})/G(\mathbf{x}^{\mathrm{T}}\boldsymbol{\lambda}(F_1))$. Define the following functionals for any distribution F_1

$$\begin{aligned}
\mathbf{A}_{\mathrm{P}}(F_1) &= \mathbb{E}_F\left(\delta\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F_1), \tau_{\mathrm{P}}(F_1)\right) w_1^*(\mathbf{x}, F_1)\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) \\
\mathbf{B}_{\mathrm{P}}(F_1) &= \mathbb{E}_F\left(\delta\Psi^2\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F), \tau_{\mathrm{P}}(F_1)\right) w_1^{*2}(\mathbf{x}, F_1)\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) .
\end{aligned}$$

Note that, since $\boldsymbol{\beta}_{\mathrm{P}}(F) = \boldsymbol{\beta}$, $\tau_{\mathrm{P}}(F) = \tau$ and $\boldsymbol{\lambda}(F) = \boldsymbol{\lambda}$, we have that $\mathbf{A}_{\mathrm{P}}(F) = \mathbf{A}_{\mathrm{P}}$ and $\mathbf{B}_{\mathrm{P}}(F) = \mathbf{B}_{\mathrm{P}}$ where $\mathbf{A}_{\mathrm{P}}$ and $\mathbf{B}_{\mathrm{P}}$ are defined in (12) and (13), respectively.

The Wald–type test functional related to the statistic based in the estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{P}}$, used to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, is given by $\mathcal{W}_{\mathrm{P}}(F_1) = (\boldsymbol{\beta}_{\mathrm{P}}(F_1) - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{V}_{\mathrm{P}}(F_1)^{-1}(\boldsymbol{\beta}_{\mathrm{P}}(F_1) - \boldsymbol{\beta}_0)$.

The following Theorem gives the value of the influence function of the test functional $\mathcal{T}_{\mathrm{P}}(F_1) = \mathcal{W}_{\mathrm{P}}(F_1)^{1/2}$.

Theorem 4.2. *Let $\mathbf{V}_{\mathrm{P}}(F_1)$ be a scatter functional such that $\mathbf{V}_{\mathrm{P}}(F) = \boldsymbol{\Sigma}_{\mathrm{P}}$, where $\boldsymbol{\Sigma}_{\mathrm{P}} = \mathbf{A}_{\mathrm{P}}^{-1}\mathbf{B}_{\mathrm{P}}\mathbf{A}_{\mathrm{P}}^{-1}$ for $\mathbf{A}_{\mathrm{P}} = \mathbf{A}_{\mathrm{P}}(F)$ and $\mathbf{B}_{\mathrm{P}} = \mathbf{B}_{\mathrm{P}}(F)$ defined in (12) and (13), respectively. Assume that the influence function $IF(\mathbf{z}_0, \boldsymbol{\beta}_{\mathrm{P}}, F)$ and the $\partial^2\boldsymbol{\beta}_{\mathrm{P}}(F_{\mathbf{z}_0,\epsilon})/\partial\epsilon^2\big|_{\epsilon=0}$ exist. Then, the influence function at F_1 of the functional $\mathcal{T}_{\mathrm{P}}(F)$ to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is given by*

$$IF(\mathbf{z}_0, \mathcal{T}_{\mathrm{P}}, F)^2 = IF(\mathbf{z}_0, \boldsymbol{\beta}_{\mathrm{P}}, F)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{P}}^{-1}IF(\mathbf{z}_0, \boldsymbol{\beta}_{\mathrm{P}}, F) . \tag{18}$$

*Moreover, under **N3**, **N6** and **N8**, we have that $IF(\mathbf{z}_0, \boldsymbol{\beta}_{\mathrm{P}}, F)$ exists and*

$$\begin{aligned}
IF(\mathbf{z}_0, \boldsymbol{\beta}_{\mathrm{P}}, F) &= -\left\{\mathbb{E}_F\left(\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\right\}^{-1}\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1^*(\mathbf{x}_0, F)\delta_0\mathbf{x}_0 \\
&= -\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x}_0)\frac{\delta_0}{p(\mathbf{x}_0)}\mathbf{A}_{\mathrm{P}}^{-1}\mathbf{x}_0 .
\end{aligned} \tag{19}$$

Besides, we have that

$$\begin{aligned}
EIF(\mathbf{z}_0^*, \boldsymbol{\beta}_{\mathrm{P}}, F) &= -\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1(\mathbf{x}_0)\mathbf{A}_{\mathrm{P}}^{-1}\mathbf{x}_0 \\
EIF(\mathbf{z}_0^*, \mathcal{T}_{\mathrm{P}}, F)^2 &= \Psi^2\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1^2(\mathbf{x}_0)\mathbf{x}_0^{\mathrm{T}}\mathbf{B}_{\mathrm{P}}^{-1}\mathbf{x}_0 \\
EIF_2(\mathbf{z}_0^*, \mathcal{T}_{\mathrm{P}}, F) &= \Psi^2\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}, \tau\right) w_1^2(\mathbf{x}_0)\mathbf{x}_0^{\mathrm{T}}\mathbf{B}_{\mathrm{P}}^{-1}\mathbf{x}_0/p(\mathbf{x}_0) .
\end{aligned}$$

It is worth noticing that except for the matrix $\mathbf{B}_{\mathrm{P}}$, the expected square influence of the propensity test, $\mathrm{EIF}_2(\mathbf{z}_0^*, \mathcal{T}_{\mathrm{P}}, F)$ is that of the Wald–type test corresponding to the complete case. This is quite natural since $\boldsymbol{\beta}_{\mathrm{P}}(F_1) = \boldsymbol{\beta}_{\mathrm{C}}(F_1)$ where $\boldsymbol{\beta}_{\mathrm{C}}(F_1)$ is the weighted estimator when the sample contains no missing responses.

4.3 Expected influence functions for the Gamma model

Consider the generalized linear model where the distribution of the response y given the vector of covariates $\mathbf{x}$ is $\Gamma(\tau, \mu(\mathbf{x}))$ and the link function is $\log(\mu(\mathbf{x})) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$, where, for any $\tau > 0$ and $\mu > 0$, we denote by $\Gamma(\tau, \mu)$ the parametrization of the Gamma distribution given by the density

$$f(y, \tau, \mu) = \begin{cases} \dfrac{\tau^{\tau}}{\mu^{\tau}\,\Gamma(\tau)}\; y^{\tau-1}\; \exp(-(\tau/\mu)y) & \text{if} \quad y \geq 0 \\ \qquad\qquad 0 & \text{if} \quad y < 0 \end{cases} .$$

Note that, if $z \sim \Gamma(\tau, \mu)$, we have that $\mathbb{E}(z) = \mu$ and $\mathrm{VAR}(z) = \mu^2/\tau$, where τ is a shape parameter.

Figures 1 to 5 show the squared *expected influence functions* (EIF^2) and the *expected square influence functions* (EIF_2) at $y = \exp(1)$ corresponding to the maximum likelihood estimators, the estimators related to those introduced in Bianco *et al.* (2005) based on the deviance, i.e., with $w_1 \equiv 1$, and the weighted estimators computed with the Tukey bisquare weight function, denoted as $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$, $\widehat{\boldsymbol{\beta}}_{\mathrm{BGY}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{TUK}}$, respectively. The weights were computed over the Mahalanobis distances with tuning constant $c = \chi^2_{k,0.95}$. The gamma regression model considered was

$$y_i|\mathbf{x}_i \sim \Gamma(\tau, \mu(\mathbf{x}_i)) \quad \text{with} \quad \mu(\mathbf{x}_i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3,\ i = 1, ..., n, \tag{20}$$

with $\tau = 3$, $\beta_1 = \beta_2 = \beta_3 = 0$ and $(x_{1i}, x_{2i}) \sim N(0, \mathbf{I})$. Figure 1 corresponds to the test functionals related to the *simplified estimators* with $p(\mathbf{x}) = 1$, while Figure 2 and Figure 3 to the test functionals related to the simplified and propensity estimators with missing probabilities satisfying the logistic model $p(\mathbf{x}) = 1/(1 + \exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} - 2))$ with $\boldsymbol{\lambda} = (2, 2)^{\mathrm{T}}$ and $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2, 2)^{\mathrm{T}}$, respectively. To plot the influence functions a grid of values for each component x_j, $j = 1, 2$, was taken between -4 and 4 with step 0.1.

A reduced range was also considered to compare the behaviour near the origin by using grid of points between -1 and 1 with step 0.025. Besides, when plotting EIF^2 for the classical propensity test and that based on the Bianco *et al.* (2005) propensity estimators a larger range of values (between -10 and 4 with step 0.25) was considered, in order to have a better idea of the EIF shape. As expected, the shape of both the EIF^2 and EIF_2 of the test functionals based on the Bianco *et al.* (2005) estimators and on their weighted versions is comparable to that of their classical relatives at the center of the distribution of the covariates. Note that EIF^2 and EIF_2 are unbounded due to leverage points for the classical test and also for that based on the Bianco *et al.* (2005) estimators. This feature make us suspect that even when the latter are based on a robust procedure to estimate the regression parameter, the test statistic may be sensitive to outliers. On the other hand, when using a weight function to downweight carriers with large Mahalanobis distances, the expected influence at points further away is downweighted.

It is worth noticing that the logistic model chosen corresponds to a missing probability that does not fulfil $\inf_{\mathbf{x}} p(\mathbf{x}) = A > 0$ and so, a worst performance of the propensity estimators when $w_1 \equiv 1$ may be expected. Effectively, the EIF^2 for both the classical procedure and that based on the Bianco *et al.* (2005) propensity estimators attain very large values in this case for large negative values of the covariates. As noted above, the shape of the EIF_2 of all procedures is comparable at the center of the distribution of the covariates with that of the classical test (see Figure 4). A different pattern to that described above is observed in this situation due to the effect of the missing probability. The EIF^2 of the test related to the simplified estimators are downweighted for large negative values of the carriers for the logistic model. Thus, only leverage points with at least one large positive component will lead to large values of the influence function for the test based on the simplified classical estimators and also for that based on the simplified version of the Bianco *et al.* (2005) estimators. On the other hand, as above, when using the weighted estimators, the influence at points further away is downweighted.

When considering $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$, the defined influence measures of both the classical tests and those based on the Bianco et al. (2005) estimators are unbounded, while for the weighted version, observations with large Mahalanobis distances have null expected influence.

As mentioned above, the EIF^2 of the propensity estimators for both missing probabilities is similar to that obtained when $p \equiv 1$. The difference in the shape is only due to the effect of the missingness probability on the matrix $\mathbf{B}_{\mathrm{P}}$.

5 Monte Carlo Study

As in Section 4.3, the observations follow the log–gamma regression model (20), with $\beta_1 = \beta_2 = \beta_3 = 0$ and $(x_{1i}, x_{2i}) \sim N(0, \mathbf{I})$. We now consider two different values for the shape parameter: $\tau = 1$ and $\tau = 3$. The sample size was $n = 100$ and the number of Monte Carlo replications was $K = 1000$.

We study the behaviour of the test statistic for samples that do not contain outliers and samples contaminated with 5% outliers. In the contaminated samples, the outliers are all equal, say $(y_0, \mathbf{x}_0^{\mathrm{T}})$. Since the magnitude of the effect of these outliers depends on x_{10} and x_{20} only throughout $x_{10}^2 + x_{20}^2$, without loss of generality they were taken of the form $(y_0, \mathbf{x}_0^{\mathrm{T}})$ with $\mathbf{x}_0^{\mathrm{T}} = (x_0, 0, 1)$ and $y_0 = \exp(m\ x_0)$. The value m represents the slope of the outliers observations. We chose two values of x_0 : low leverage outliers with $x_0 = 1$, moderate leverage outliers with $x_0 = 3$ and high leverage outliers with $x_0 = 10$. As values for m we considered $m = 0.5$ and 2.5. These contaminations are denoted C_{m,x_0}. We have also considered an intermediate contamination C_1 by replacing 5 observations by $(y_0, \mathbf{x}_0^\star)$ where $\mathbf{x}_0^\star = (2.5, 2.6, 1)^{\mathrm{T}}$ and $y_0 = \exp(1)$.

The robust estimators were computed as described in Remark 2.1.1. For the weighted estimators, we used the Tukey's bisquare weight function with tuning constant $c = \chi^2_{k,0.95}$. The weights were computed over the robust Mahalanobis distances based on an $S-$estimator with breakdown point 0.25 using 500 subsamples. From now on, we denote by $\mathcal{W}_{\mathrm{ML}}$, $\mathcal{W}_{\mathrm{BGY}}$, $\mathcal{W}_{\mathrm{TUK}}$, the tests based on the maximum likelihood estimators, the estimators related to those defined in Bianco *et al.* (2005), i.e., with $w_1 \equiv 1$, and their weighted version with Tukey's weights, respectively. The propensity score tests will be denoted as $\mathcal{W}_{\mathrm{P,ML}}$, $\mathcal{W}_{\mathrm{P,BGY}}$, $\mathcal{W}_{\mathrm{P,TUK}}$, respectively. The nominal level was set equal to 0.05.

We considered four models for the missing probability

- $p \equiv 1$
- $p \equiv 0.8$, missing completely at random
- $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2, 2)^{\mathrm{T}}$.
- $p(\mathbf{x}) = 1/(1 + \exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} - 2))$ with $\boldsymbol{\lambda} = (2, 2)^{\mathrm{T}}$, i.e., a logistic model for the missing probability.

In order to study the convergence speed to the χ^2 distribution, Table 1 reports the observed frequencies of rejection in the non–contaminated case C_0, for different sample sizes $n = 50, 100, 250$ and 500. It is worth noticing that, as mentioned in Section 4.3, in the logistic case, **N8** does not hold and so, the propensity score tests perform worst than the test based on the simplified estimators. In all cases the convergence is quite slow and so, a bootstrap approach may be considered. However, providing bootstrap tests in this setting is beyond the scope of the paper. On the other hand, due to the strong asymmetry of the exponential distribution, the observed frequencies of rejection are further away to the nominal level than when $\tau = 3$.

Figures 6 to 17 allow to study the power performance of the tests when $n = 100$. Therein, we have plotted the observed frequencies of rejection under the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}_3$ and the alternatives $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \Delta n^{-1/2}(1, 0, 0)^{\mathrm{T}}$ with $\Delta = \pm 6$, ± 4.8, ± 3.6, ± 2.4, ± 1.2, ± 0.8, ± 0.4 and ± 0.2. The lines in black, green and red correspond, respectively, to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$ or $\boldsymbol{\beta}_{\mathrm{P,ML}}$, to the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ or $\boldsymbol{\beta}_{\mathrm{P,BGY}}$ and to their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$ and $\boldsymbol{\beta}_{\mathrm{P,TUK}}$.

As expected, under C_0 the test procedures based on classical or robust estimators perform quite similarly, under all the missing schemes. As noted in Section 4.3, the simplified methods perform better than the propensity ones when considering a logistic missing probability. For large values of m and/or x_0, the classical procedure is non–informative. On the other hand, the tests $\mathcal{W}_{\mathrm{BGY}}$ and $\mathcal{W}_{\mathrm{P,BGY}}$ show their sensitivity to moderate outliers ($m = 0.5$ and $x_0 = 3$) for all the missingness models and also to extreme outliers ($x_0 = 10$), when considering a logistic missingness model, probably due to the effect of high leverage points on the estimation of the asymptotic covariance matrix. Their weighted versions are more stable with respect to all the contaminations considered.

6 Example: Leukemia data

The data of Feigl and Zelen (1965) represent the survivorship of 33 patients of acute myelogenous leukemia divided in two groups, that correspond to a factor variable AG which classifies the patients as positive or negative depending on the presence or absence of a morphological characteristic in the white cells. The original data are time at death and also the white blood cells count WBC, which is a useful tool for diagnosing the initial condition of the patient,

		$\mathcal{W}_n$				$\mathcal{W}_{\mathrm{P},n}$			
	n	50	100	250	500	50	100	250	500
		$p \equiv 1$							
$\tau = 1$	$\mathcal{W}_{\mathrm{ML}}$	0.160	0.092	0.079	0.065				
	$\mathcal{W}_{\mathrm{BGY}}$	0.163	0.107	0.069	0.044				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.136	0.105	0.060	0.053				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.135	0.102	0.061	0.051				
$\tau = 3$	$\mathcal{W}_{\mathrm{ML}}$	0.114	0.097	0.062	0.051				
	$\mathcal{W}_{\mathrm{BGY}}$	0.136	0.101	0.066	0.061				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.121	0.084	0.055	0.057				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.116	0.085	0.056	0.056				
		$p \equiv 0.8$							
$\tau = 1$	$\mathcal{W}_{\mathrm{ML}}$	0.186	0.122	0.080	0.070				
	$\mathcal{W}_{\mathrm{BGY}}$	0.207	0.118	0.081	0.064				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.171	0.107	0.073	0.075				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.168	0.105	0.070	0.070				
$\tau = 3$	$\mathcal{W}_{\mathrm{ML}}$	0.159	0.090	0.065	0.051				
	$\mathcal{W}_{\mathrm{BGY}}$	0.195	0.108	0.072	0.059				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.167	0.087	0.067	0.067				
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.170	0.082	0.068	0.063				
		$p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$							
$\tau = 1$	$\mathcal{W}_{\mathrm{ML}}$	0.197	0.124	0.087	0.060	0.230	0.123	0.084	0.054
	$\mathcal{W}_{\mathrm{BGY}}$	0.279	0.135	0.080	0.059	0.294	0.133	0.085	0.061
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.213	0.114	0.084	0.053	0.242	0.117	0.08	0.052
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.212	0.106	0.085	0.054	0.234	0.112	0.081	0.054
$\tau = 3$	$\mathcal{W}_{\mathrm{ML}}$	0.174	0.114	0.081	0.061	0.189	0.120	0.080	0.061
	$\mathcal{W}_{\mathrm{BGY}}$	0.253	0.142	0.074	0.063	0.260	0.135	0.071	0.063
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.207	0.110	0.063	0.049	0.218	0.112	0.067	0.059
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.205	0.115	0.062	0.050	0.220	0.112	0.062	0.056
		$p(\mathbf{x}) = 1/(1 + \exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} - 2))$							
$\tau = 1$	$\mathcal{W}_{\mathrm{ML}}$	0.191	0.116	0.088	0.082	0.294	0.203	0.169	0.145
	$\mathcal{W}_{\mathrm{BGY}}$	0.264	0.106	0.088	0.060	0.377	0.223	0.191	0.163
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.224	0.098	0.086	0.071	0.331	0.167	0.133	0.087
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.212	0.105	0.088	0.067	0.334	0.179	0.137	0.101
$\tau = 3$	$\mathcal{W}_{\mathrm{ML}}$	0.163	0.111	0.068	0.047	0.246	0.201	0.145	0.107
	$\mathcal{W}_{\mathrm{BGY}}$	0.233	0.121	0.053	0.051	0.329	0.224	0.178	0.135
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.95}}$	0.192	0.108	0.062	0.046	0.282	0.160	0.091	0.068
	$\mathcal{W}_{\mathrm{TUK},c=\chi^2_{2,0.975}}$	0.194	0.101	0.062	0.046	0.297	0.166	0.113	0.072

Table 1: Observed frequencies of rejection in the non–contaminated case C_0 with $p \equiv 1$, $p \equiv 0.8$, $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ and $p(\mathbf{x}) = 1/(1 + \exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} - 2))$ with $\boldsymbol{\lambda} = (2, 2)^{\mathrm{T}}$.

indeed higher counts seem to be associated with more severe conditions. Bianco *et al.* (2005) fit, to the complete data set, the model

$$\log(y_i) = \beta_1 WBC_i + \beta_2 AG_i + \beta_3 + u_i,$$

where u_i has $\log \Gamma(\tau_0, 1)$ distribution through their BGY–estimator. The QQ-plot of the residuals of the BGY-estimate computed by Bianco *et al.* (2005) reveals four clear outliers corresponding to patients with very high values of WBC who survived more than expected.

Denote by $\widehat{\boldsymbol{\beta}}_{\text{ML}}$, $\widehat{\boldsymbol{\beta}}_{\text{BGY}}$ and $\widehat{\boldsymbol{\beta}}_{\text{TUK}}$ the ML–estimates and the two robust estimates BGY and TUK–estimates computed with all the data, respectively. Besides, $\widehat{\boldsymbol{\beta}}_{\text{ML}}^{-\{4\text{OUT}\}}$ stands for the ML-estimator applied to the sample without the four outliers. In this case, since AG is a factor variable, when computing the weighted estimators with the Tukey's bisquare function, $\widehat{\boldsymbol{\beta}}_{\text{TUK}}$, the weights $w_1(\mathbf{x})$ were based only on the variable WBC and the tuning constant was chosen as $c = \chi^2_{1,0.95}$. The robust Mahalanobis distance of WBC equals in this case $|WBC_i - \text{median}_i(WBC_i)|/\text{MAD}(WBC_i)$.

Table 2 reports the values of the above mentioned estimates together with the $p-$values of the related Wald–type statistics to check $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$. It is worth noticing that according to the results given by $\widehat{\boldsymbol{\beta}}_{\text{ML}}$, the coefficient of the variable WBC is non-significant at a 5% level. On the other hand, the ML-estimator applied to the sample without the four outliers and the robust estimators lead to the opposite conclusion.

To evaluate the performance of the proposed tests for incomplete data sets, we introduced artificially missing data to this example and we took the above analysis as a natural counterpart. Missing responses among the non–outlying points were introduced at random according to two missing schemes, a completely at random situation with $p(\mathbf{x}) = 0.9$ and a missing at random case with logistic probability of missing $p(\mathbf{x}) = 1/(1+\exp(0.2\,WBC-4))$. In this way, for the logistic case, 8 responses (almost 25% of the data) result in missing observations. The analysis was repeated for each of the obtained samples. In Table 3 we summarize the corresponding results. Different conclusions are derived depending on the missing scheme. As expected, when missing responses occur completely at random, analogous results to those obtained with the complete data set are obtained. In this sense, when $p(\mathbf{x}) = 0.9$, according to the computed robust tests the variable WBC is significant, while from the ML-estimator we conclude otherwise. On the other hand, for the incomplete sample obtained through a logistic missingness probability, the estimators $\widehat{\boldsymbol{\beta}}_{\text{ML}}$ and $\widehat{\boldsymbol{\beta}}_{\text{BGY}}$ take similar values. Moreover, according to the robust test statistic $\widehat{\mathcal{W}}_{\text{TUK}}$, based on Tukey's weights, the variable WBC is significant, while both the classical test and $\widehat{\mathcal{W}}_{\text{BGY}}$ lead to the opposite conclusion. Besides, if the 4 identified outlying observations were removed from this incomplete sample, the three tests would lead to the same conclusion obtained for the situation with no missing responses, indeed the p–value of the classical test and both robust tests would be 0. These results show the advantage of introducing weights as a useful tool to prevent from outlying points even under different missing schemes.

7 Concluding Remarks

We have introduced two resistant procedures to test hypotheses on the regression parameter under a generalized regression model, when there are missing observations in the responses and it can be suspected that anomalous observations are present in the sample. The estimators turn out to be asymptotically normally distributed. In particular, the asymptotic distribution of the robust estimators based on a propensity score approach is the same when the missingness

	Estimated Coefficients				$H_0 : \beta_1 = 0$ $p-$value			
	$\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}^{-\{4\mathrm{OUT}\}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{BGY}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{TUK}}$	$\widehat{\mathcal{W}}_{\mathrm{ML}}$	$\widehat{\mathcal{W}}_{\mathrm{ML}}^{-\{4\mathrm{OUT}\}}$	$\widehat{\mathcal{W}}_{\mathrm{BGY}}$	$\widehat{\mathcal{W}}_{\mathrm{TUK}}$
$\frac{WBC}{1000}$	-0.007	-0.051	-0.051	-0.089	0.102	0	0	0
AG	-1.101	-1.574	-1.802	-1.510				
Intercept	4.227	4.795	4.849	5.101				

Table 2: Analysis of Feigl and Zelen data. Complete data set.

	Estimated Coefficients			$H_0 : \beta_1 = 0$ $p-$value		
	$\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{BGY}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{TUK}}$	$\widehat{\mathcal{W}}_{\mathrm{ML}}$	$\widehat{\mathcal{W}}_{\mathrm{BGY}}$	$\widehat{\mathcal{W}}_{\mathrm{TUK}}$
	$p(\mathbf{x}) = 0.9$					
$\frac{WBC}{1000}$	-0.008	-0.050	-0.084	0.071	0	0
AG	-0.974	-1.469	-1.364			
Intercept	4.333	4.841	5.055			
	$p(\mathbf{x}) = 1/(1+\exp(0.2\,WBC - 4)$					
$\frac{WBC}{1000}$	-0.0012	-0.0004	-0.1210	0.743	0.912	0
AG	-1.3718	-1.4371	-1.4315			
Intercept	4.4432	4.5055	5.2617			

Table 3: Analysis of Feigl and Zelen data with two missingness probabilities.

probability is assumed to be known or when it is estimated parametrically using consistent estimators. Moreover, if the test statistic involves a bounded support function w_1, it allows to control the effect of high leverage points and also that of continuous missing probabilities such that $p(\mathbf{x}) > 0$ for any $\mathbf{x}$, but $\inf_{\mathbf{x}} p(\mathbf{x}) = 0$.

The test statistics are robust versions of the classical Wald–type statistic. Even when the tests statistics have a limiting χ^2-distribution under the null hypothesis and under contiguous alternatives, the simulation study illustrates the slow convergence to the asymptotic distribution. Bootstrapping techniques could be implemented in order to improve the convergence rate, but this task deserves further research that will be the subject of a forthcoming work. A measure of sensitivity based on the influence function and adapted to the missing situation was also defined.

The simulation study also confirms the expected inadequate behaviour of the classical Wald–type test and of the unweighted robust estimators in the presence of outliers. The proposed robust procedures for the regression parameter perform quite similarly both in level and power, under the central model or under the contaminations studied.

Finally, through a real data set, we confirm the stability, under different missingness probability patterns, of the decision rule induced by $\widehat{\mathcal{W}}_{\mathrm{TUK}}$, which is based on weights that control high leverage points.

Acknowledgement. This work began while Isabel Rodrigues was visiting the Instituto de Cálculo from the Universidad de Buenos Aires, partially supported by the *Fundação para a Ciência e a Tecnologia*

(FCT), the *Center for Mathematics and its Applications* (CEMAT) and *Fundação Calouste Gulbenkian* (FCG). This research was also partially supported by Grants PIP 112-200801-00216 from CONICET, PICT 0821 from ANPCYT and X-018 from the Universidad de Buenos Aires at Buenos Aires, Argentina and by the joint cooperation program ANPCYT–GRICES PO09/05.

8 Appendix

PROOF OF THEOREM 3.1. a) Using a Taylor's expansion of order one, we get that

$$\begin{aligned}
\mathbf{0}_k &= \frac{1}{n}\sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}, \widehat{\tau}\right) w_1(\mathbf{x}_i)\mathbf{x}_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0, \widehat{\tau}\right) w_1(\mathbf{x}_i)\mathbf{x}_i + \frac{1}{n}\sum_{i=1}^{n} \delta_i \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}, \widehat{\tau}\right) w_1(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\ ,
\end{aligned}$$

and so, we have that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{A}_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0, \widehat{\tau}\right) w_1(\mathbf{x}_i)\mathbf{x}_i\ .$$

Note that Lemma 3.1 entails that

$$\mathbf{A}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_i \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}, \widehat{\tau}\right) w_1(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} \xrightarrow{p} \mathbf{A}$$

Using that $\widehat{\tau} \xrightarrow{p} \tau$, **N5** and the maximal inequality, we get that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0, \widehat{\tau}\right) w_1(\mathbf{x}_i)\mathbf{x}_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0, \tau_0\right) w_1(\mathbf{x}_i)\mathbf{x}_i \xrightarrow{p} 0\ .$$

Hence, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0, \widehat{\tau}\right) w_1(\mathbf{x}_i)\mathbf{x}_i \xrightarrow{\mathcal{D}} N(\mathbf{0}_k, \mathbf{B})$ entailing that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{A}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0, \tau\right) w_1(\mathbf{x}_i)\mathbf{x}_i + o_p(1) \tag{21}$$

and so, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}_k, \boldsymbol{\Sigma})$.

b) Let T_n stand for $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ or $\widehat{\mathbf{B}}^{-1/2}\widehat{\mathbf{A}}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. We will use Le Cam's third Lemma (see van der Vaart, 2000, page 90). Therefore, we need to obtain the asymptotic distribution of $(T_n, \ln(q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})/p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})))$, where $p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})$ is the joint density under the null hypothesis and $q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})$ is the corresponding one under the alternative, $\mathbf{y} = (y_1, \dots, y_n)^{\mathrm{T}}$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$.

Let $\theta_n(\mathbf{x}_i) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \mathbf{x}_i^{\mathrm{T}}\mathbf{c}n^{-1/2} = \theta(\mathbf{x}_i) + \mathbf{x}_i^{\mathrm{T}}\mathbf{c}n^{-1/2}$, using (1) we get that

$$\begin{aligned}
\frac{q_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})}{p_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})} &= \prod_{i=1}^{n} \frac{\exp\left\{\left[y_i\theta(\mathbf{x}_i) + y_i\mathbf{x}_i^{\mathrm{T}}\mathbf{c}n^{-1/2} - B\left(\theta(\mathbf{x}_i) + \mathbf{x}_i^{\mathrm{T}}\mathbf{c}n^{-1/2}\right)\right]/A(\tau) + C(y_i,\tau)\right\}}{\exp\left\{\left[y_i\theta(\mathbf{x}_i) - B(\theta(\mathbf{x}_i))\right]/A(\tau) + C(y_i,\tau)\right\}} \\
&= \prod_{i=1}^{n} \exp\left\{y_i\mathbf{x}_i^{\mathrm{T}}\mathbf{c}n^{-1/2}/A(\tau) + \left[B(\theta(\mathbf{x}_i)) - B\left(\theta(\mathbf{x}_i) + \mathbf{x}_i^{\mathrm{T}}\mathbf{c}n^{-1/2}\right)\right]/A(\tau)\right\}\ .
\end{aligned}$$

Then, if $\xi_i = \xi_{i,n}$ denotes an intermediate point between $\theta(\mathbf{x}_i)$ and $\theta(\mathbf{x}_i) + \mathbf{x}_i^{\mathrm{T}}\mathbf{c}n^{-1/2}$, we have

$$\begin{aligned}
\ln\frac{q_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})}{p_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})} &= \frac{1}{A(\tau)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(y_i\mathbf{x}_i^{\mathrm{T}}\mathbf{c} - B'(\theta(\mathbf{x}_i))\mathbf{x}_i^{\mathrm{T}}\mathbf{c}\right) - \frac{1}{A(\tau)}\frac{1}{2n}\sum_{i=1}^{n}B''(\theta(\mathbf{x}_i))(\mathbf{x}_i^{\mathrm{T}}\mathbf{c})^2 \\
&- \frac{1}{A(\tau)}\frac{1}{2n}\sum_{i=1}^{n}[B''(\xi_i) - B''(\theta(\mathbf{x}_i))](\mathbf{x}_i^{\mathrm{T}}\mathbf{c})^2\ .
\end{aligned}$$

Since $B'(\theta(\mathbf{x}_i)) = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)$, we have that

$$\begin{aligned}
A(\tau)\ \ln\frac{q_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})}{p_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)]\mathbf{x}_i^{\mathrm{T}}\mathbf{c} - \frac{1}{2n}\sum_{i=1}^{n}H'(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)(\mathbf{x}_i^{\mathrm{T}}\mathbf{c})^2 \\
&- \frac{1}{2n}\sum_{i=1}^{n}[B''(\xi_i) - B''(\theta(\mathbf{x}_i))](\mathbf{x}_i^{\mathrm{T}}\mathbf{c})^2\ . \qquad (22)
\end{aligned}$$

Let $\sigma^2 = \mathbf{c}^{\mathrm{T}}\mathbb{E}(H'(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0))\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}})\mathbf{c}/A(\tau)$. Then, (22) entails that

$$A(\tau)\ \ln\frac{q_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})}{p_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)]\mathbf{x}_i^{\mathrm{T}}\mathbf{c} - \frac{1}{2n}\sum_{i=1}^{n}H'(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)(\mathbf{x}_i^{\mathrm{T}}\mathbf{c})^2 + o_p(1)\ ,$$

implying that $\ln\left(q_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})/p_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})\right) \xrightarrow{\mathcal{D}} N(-\sigma^2/2, \sigma^2)$.

On the other hand, in the proof of a), see (21), we obtained that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{A}^{-1}\ \mathbf{C}_n + o_p(1)$, where $\mathbf{A}$ is defined in (10) and

$$\mathbf{C}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0, \tau\right)w_1(\mathbf{x}_i)\mathbf{x}_i\ .$$

Moreover, Lemma 3.1 entails that $\widehat{\mathbf{A}} \xrightarrow{p} \mathbf{A}$ and $\widehat{\mathbf{B}} \xrightarrow{p} \mathbf{B}$, thus, using that $\mathbf{C}_n$ is asymptotically normally distributed, we get that $\widehat{\mathbf{B}}^{-1/2}\widehat{\mathbf{A}}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{B}^{-1/2}\ \mathbf{C}_n + o_p(1)$. Hence, to derive the joint asymptotic distribution of $(T_n, \ln(q_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})/p_n(\mathbf{y},\mathbf{X},\boldsymbol{\delta})))^{\mathrm{T}}$, it is enough to compute the covariance between $\mathbf{C}_n$ and

$$R_1 = \frac{1}{A(\tau)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)]\mathbf{x}_i^{\mathrm{T}}\mathbf{c}\ .$$

Using that $\mathbb{E}_F(y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)) = 0$, we get that

$$\begin{aligned}
\textsc{Cov}(\mathbf{C}_n, R_1) &= \frac{1}{A(\tau)}\textsc{Cov}(\delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0, \tau\right) w_1(\mathbf{x}_i)\mathbf{x}_i, [y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)]\mathbf{x}_i^{\mathrm{T}}\mathbf{c}) \\
&= \frac{1}{A(\tau)}\mathbb{E}_F[(y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0))\delta_i \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0, \tau\right) w_1(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}]\mathbf{c} \\
&= \frac{1}{A(\tau)}\mathbb{E}_F[(y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0))\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0, \tau\right) p(\mathbf{x}_i) w_1(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}]\mathbf{c} \,.
\end{aligned}$$

It is easy to see that **N5** entails that

$$\frac{1}{A(\tau)}\mathbb{E}_F[(y_1 - H(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0))\Psi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0, \tau\right)|\mathbf{x}_1] = -\mathbb{E}_F[\chi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0, \tau\right)|\mathbf{x}_1] \,,$$

and so, $\textsc{Cov}(\mathbf{C}_n, R_1) = -\mathbf{A}\,\mathbf{c}$, which implies that

$$(\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}, \ln(q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})/p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})))^{\mathrm{T}} \xrightarrow{\mathcal{D}} N\left(\begin{pmatrix} \mathbf{0}_k \\ -\frac{1}{2}\sigma^2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & -\mathbf{c} \\ -\mathbf{c} & \sigma^2 \end{pmatrix}\right)$$

and so, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N\left(-\mathbf{c}, \boldsymbol{\Sigma}\right)$ under $H_{1,n}$, concluding the proof. □

Proof of Theorem 4.1. We need to compute $\partial^2 \mathcal{W}(F_{\mathbf{z}_0,\epsilon})/\partial\epsilon^2\Big|_{\epsilon=0}$. Since under H_0 $\boldsymbol{\beta}(F) = \boldsymbol{\beta}_0$, we get

$$\frac{\partial^2}{\partial\epsilon^2}\mathcal{W}(F_{\mathbf{z}_0,\epsilon})\Big|_{\epsilon=0} = 2\left(\frac{\partial}{\partial\epsilon}(\boldsymbol{\beta}(F_{\mathbf{z}_0,\epsilon}) - \boldsymbol{\beta}_0)\right)^{\mathrm{T}} \mathbf{V}^{-1}(F_{\mathbf{z}_0,\epsilon})(\frac{\partial}{\partial\epsilon}(\boldsymbol{\beta}(F_{\mathbf{z}_0,\epsilon}) - \boldsymbol{\beta}_0))\Big|_{\epsilon=0}$$

and so,

$$\mathrm{IF}(\mathbf{z}_0, \mathcal{T}, F)^2 = \mathrm{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathrm{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F) \,. \tag{23}$$

To obtain the influence function of $\boldsymbol{\beta}(F)$, note that

$$\mathbb{E}_{F_{\mathbf{z}_0,\epsilon}}\left(\delta\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F_{\mathbf{z}_0,\epsilon}), \tau(F_{\mathbf{z}_0,\epsilon})\right) w_1(\mathbf{x})\mathbf{x}\right) = \mathbf{0}_k$$

implies

$$\begin{aligned}
\mathbf{0}_k &= (1-\epsilon)\mathbb{E}_F\left(\delta\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F_{\mathbf{z}_0,\epsilon}), \tau(F_{\mathbf{z}_0,\epsilon})\right) w_1(\mathbf{x})\mathbf{x}\right) \\
&+ \epsilon\,\delta_0\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}(F_{\mathbf{z}_0,\epsilon}), \tau(F_{\mathbf{z}_0,\epsilon})\right) w_1(\mathbf{x}_0)\mathbf{x}_0 \,.
\end{aligned} \tag{24}$$

Therefore, differentiating (24) with respect to ϵ and evaluating at $\epsilon = 0$, we obtain that

$$\begin{aligned}
\mathbf{0}_k &= \mathbb{E}_F\left(\delta\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F), \tau(F)\right) w_1(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\mathrm{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F) \\
&+ \frac{\partial}{\partial t}\mathbb{E}_F\left(\delta\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F), t\right) w_1(\mathbf{x})\mathbf{x}\right)\Big|_{t=\tau(F)}\mathrm{IF}(\mathbf{z}_0, \tau, F) \\
&+ \delta_0\Psi\left(y_0, \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}(F), \tau(F)\right) w_1(\mathbf{x}_0)\mathbf{x}_0 - \mathbb{E}_F\left(\delta\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F), \tau(F)\right) w_1(\mathbf{x})\mathbf{x}\right) \,.
\end{aligned}$$

Using the condition **N6** and the fact that

$$\mathbb{E}_F\left(\delta\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F),\tau(F)\right)w_1(\mathbf{x})\mathbf{x}\right)=\mathbb{E}_F\left(\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}(F),\tau(F)\right)w_1(\mathbf{x})p(\mathbf{x})\mathbf{x}\right)\ ,$$

we conclude the proof. □

Proof Theorem 4.2. Following the same steps as in the proof of Theorem 4.1, we need to compute influence function of $\boldsymbol{\beta}_{\mathrm{P}}(F)$. From

$$\mathbb{E}_{F_{\mathbf{z}_0,\epsilon}}\left(\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F_{\mathbf{z}_0,\epsilon}),\tau_{\mathrm{P}}(F_{\mathbf{z}_0,\epsilon})\right)w_1^*(\mathbf{x},F_{\mathbf{z}_0,\epsilon})\delta\mathbf{x}\right)=\mathbf{0}_k$$

we have that

$$\begin{aligned}\mathbf{0}_k &= (1-\epsilon)\mathbb{E}_F\left(\delta\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F_{\mathbf{z}_0,\epsilon}),\tau_{\mathrm{P}}\left(F_{\mathbf{z}_0,\epsilon}\right)\right)w_1^*(\mathbf{x},F_{\mathbf{z}_0,\epsilon})\mathbf{x}\right)\\ &+ \epsilon\,\delta_0\Psi\left(y_0,\mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F_{\mathbf{z}_0,\epsilon}),\tau_{\mathrm{P}}\left(F_{\mathbf{z}_0,\epsilon}\right)\right)w_1^*(\mathbf{x}_0,F_{\mathbf{z}_0,\epsilon})\mathbf{x}_0\ .\end{aligned}\tag{25}$$

Differentiating (25) with respect to ϵ and evaluating at $\epsilon=0$, we obtain

$$\begin{aligned}\mathbf{0}_k &= \mathbb{E}_F\left(\chi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F),\tau_{\mathrm{P}}(F)\right)w_1^*(\mathbf{x},F)p(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\mathrm{IF}(\mathbf{z}_0,\boldsymbol{\beta}_{\mathrm{P}},F)\\ &+ \mathbb{E}_F\left(\frac{\partial}{\partial t}\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F),t\right)\Big|_{t=\tau_{\mathrm{P}}(F)}w_1^*(\mathbf{x},F)p(\mathbf{x})\mathbf{x}\right)\mathrm{IF}(\mathbf{z}_0,\tau_{\mathrm{P}},F)\\ &+ \delta_0\Psi\left(y_0,\mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F),\tau_{\mathrm{P}}(F)\right)w_1^*(\mathbf{x}_0,F)\mathbf{x}_0-\mathbb{E}_F\left(\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F),\tau_{\mathrm{P}}(F)\right)w_1^*(\mathbf{x},F)p(\mathbf{x})\mathbf{x}\right)\\ &- \mathbb{E}_F\left(\frac{G'(\mathbf{x}^{\mathrm{T}}\boldsymbol{\lambda})}{G(\mathbf{x}^{\mathrm{T}}\boldsymbol{\lambda})}\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{P}}(F),\tau_{\mathrm{P}}(F)\right)w_1^*(\mathbf{x},F)p(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\mathrm{IF}(\mathbf{z}_0,\boldsymbol{\lambda},F)\end{aligned}$$

and the proof follows from **N6**. □

References

Bianco, A., Boente, G. and Rodrigues, I. M. (2010). Resistant estimators in Poisson and Gamma models with missing responses. Available on `http://www.ic.fcen.uba.ar/preprints/biancoboenterodrigues2010a.pdf`

Bianco, A., García Ben, M. and Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *Canad. J. Statist.* **33**, 511-528.

Bianco, A. and Martínez, E. (2009). Robust testing in the logistic regression model. *Comp. Statist. Data Anal.* **53**, 4095-4105.

Bianco, A. and Yohai, V. (1996). Robust estimation in the logistic regression model. *Lecture Notes in Statistics*, **109**, 17-34. Springer-Verlag, New York.

Boente G., González–Manteiga W. and Pérez–González A. (2009). Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference*, **139**: 571-592.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.* **96**, 1022-1030.

Croux, C. and Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Comp. Statist. Data Anal.*, **44**, 273-295.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics* **21**, 826838.

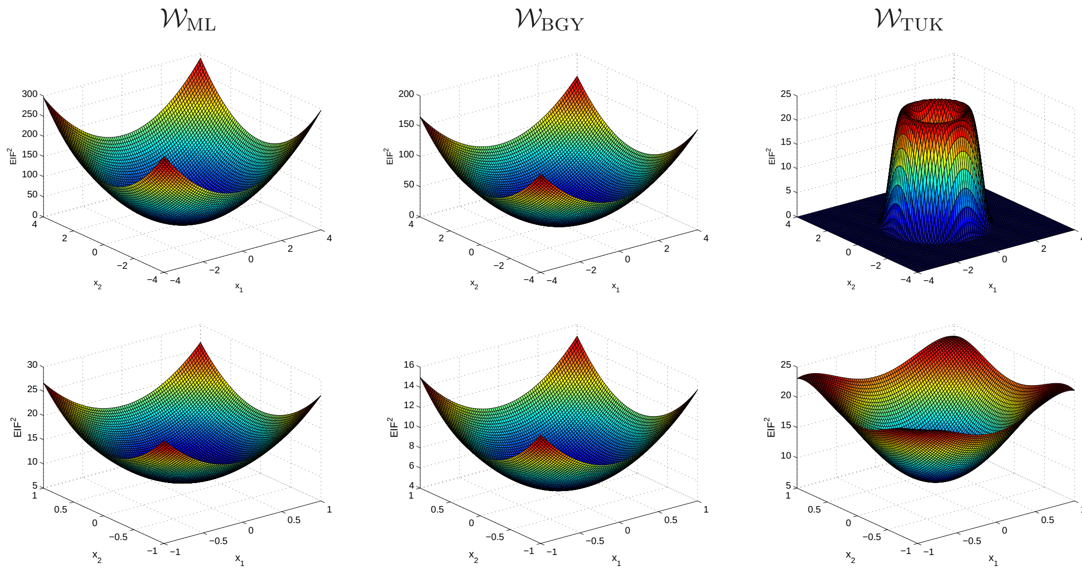

Figure 1: Square of the expected influence functions of the Wald–type statistics, EIF^2, under the Gamma model (20) when $p \equiv 1$. The upper figures correspond to a range of values for x_j between -4 and 4 and the lower ones to a range of values for x_j between -1 and 1.

Künsch, H., Stefanski, L. and Carroll, R. (1989). Conditionally unbiased bounded influence estimation in general regression models with applications to generalized linear models. *J. Amer. Assoc.* **84**, 460-466.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. (2nd ed.) London: Champman and Hall.

Maronna R., Martin D. and Yohai V. (2006). *Robust statistics: Theory and methods*. Wiley, New York.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.* **33**, 101-116.

Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.* **89**, 897-904.

Heritier, S., Cantoni, E., Copt, S. and Victoria–Feser, M.P. (2009). *Robust Methods in Biostatistics*. Wiley Series in Probability and Statistics. Wiley.

Stefanski, L., Carroll, R. and Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika* **73**, 413-424.

Van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press

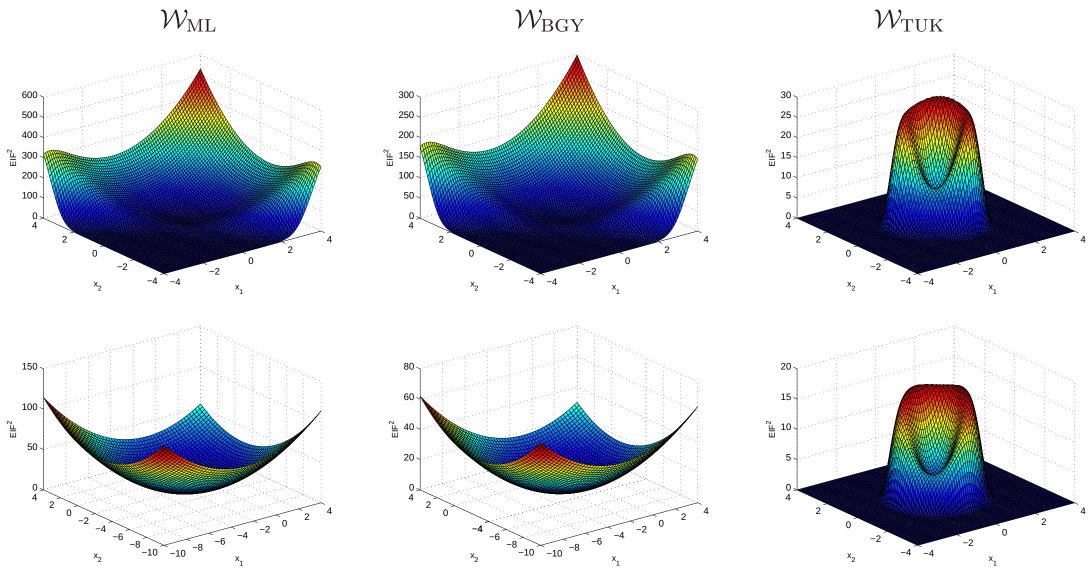

Figure 2: Square of the expected influence functions of the Wald–type statistics, EIF^2, under the Gamma model (20) when $p(\mathbf{x}) = 1/(1 + \exp(-\boldsymbol{\lambda}^{\text{T}}\mathbf{x} - 2))$ with $\boldsymbol{\lambda} = (2, 2)^{\text{T}}$ for the test functionals. The upper figures correspond to the simplified estimators and the lower ones to the propensity estimators.

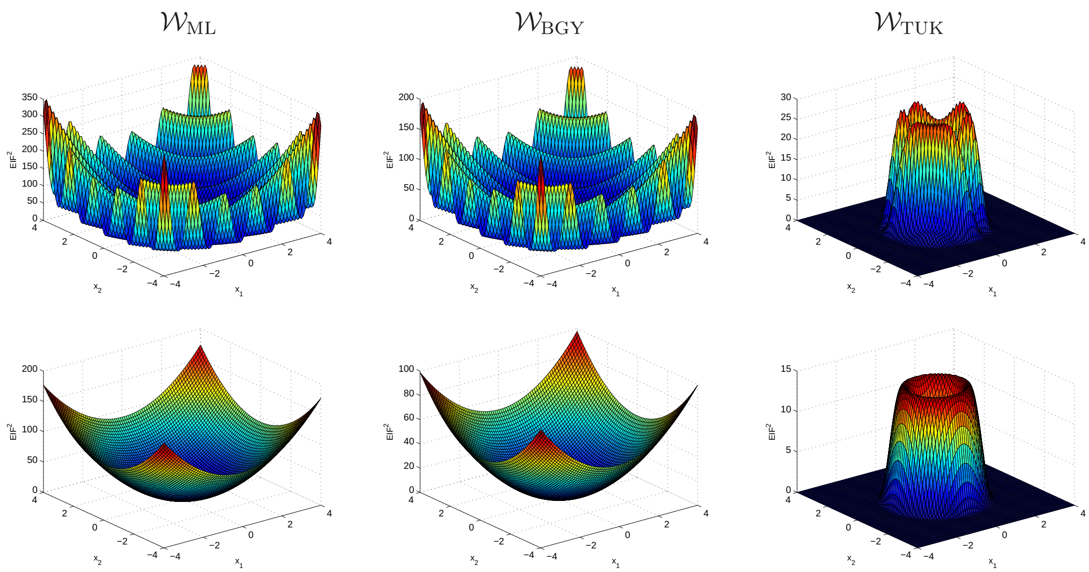

Figure 3: Square of the expected influence functions of the Wald–type statistics, EIF^2, under the Gamma model (20) when $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\text{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2, 2)^{\text{T}}$ for the test functionals. The upper figures correspond to the simplified estimators and the lower ones to the propensity estimators.

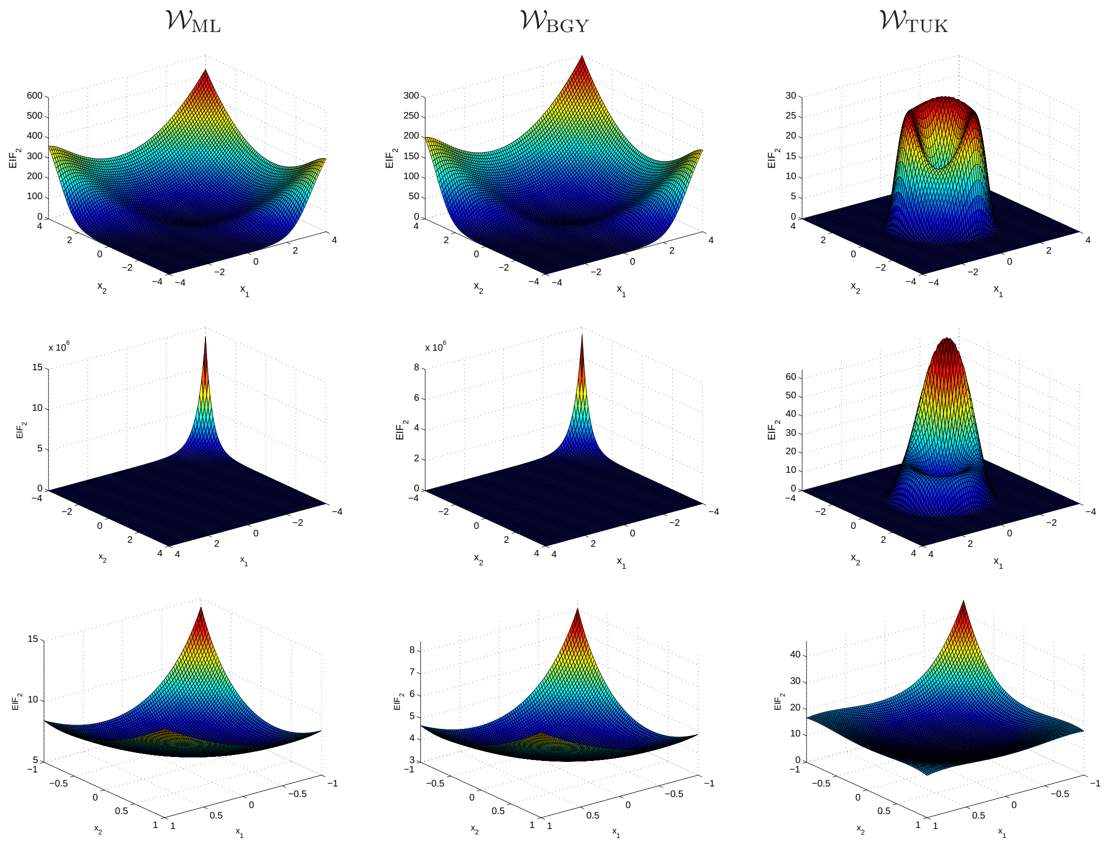

Figure 4: Expected squared influence function of the Wald–type statistics, EIF_2, under the Gamma model (20) when $p(\mathbf{x}) = 1/(1 + \exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} - 2))$ with $\boldsymbol{\lambda} = (2, 2)^{\mathrm{T}}$ for the test functionals. The upper figures correspond to the simplified estimators, the middle ones to the propensity estimators and the lower ones to the propensity estimators in a reduced range of values.

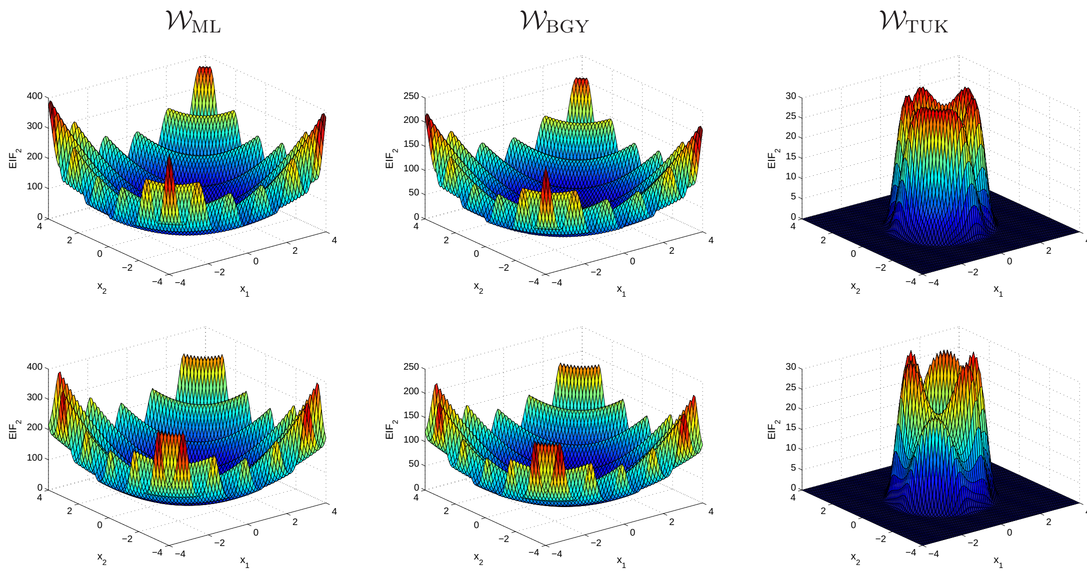

Figure 5: Expected squared influence function of the Wald–type statistics, EIF_2, under the Gamma model (20) when $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$ for the test functionals. The upper figures correspond to the simplified estimators and the lower ones to the propensity estimators.

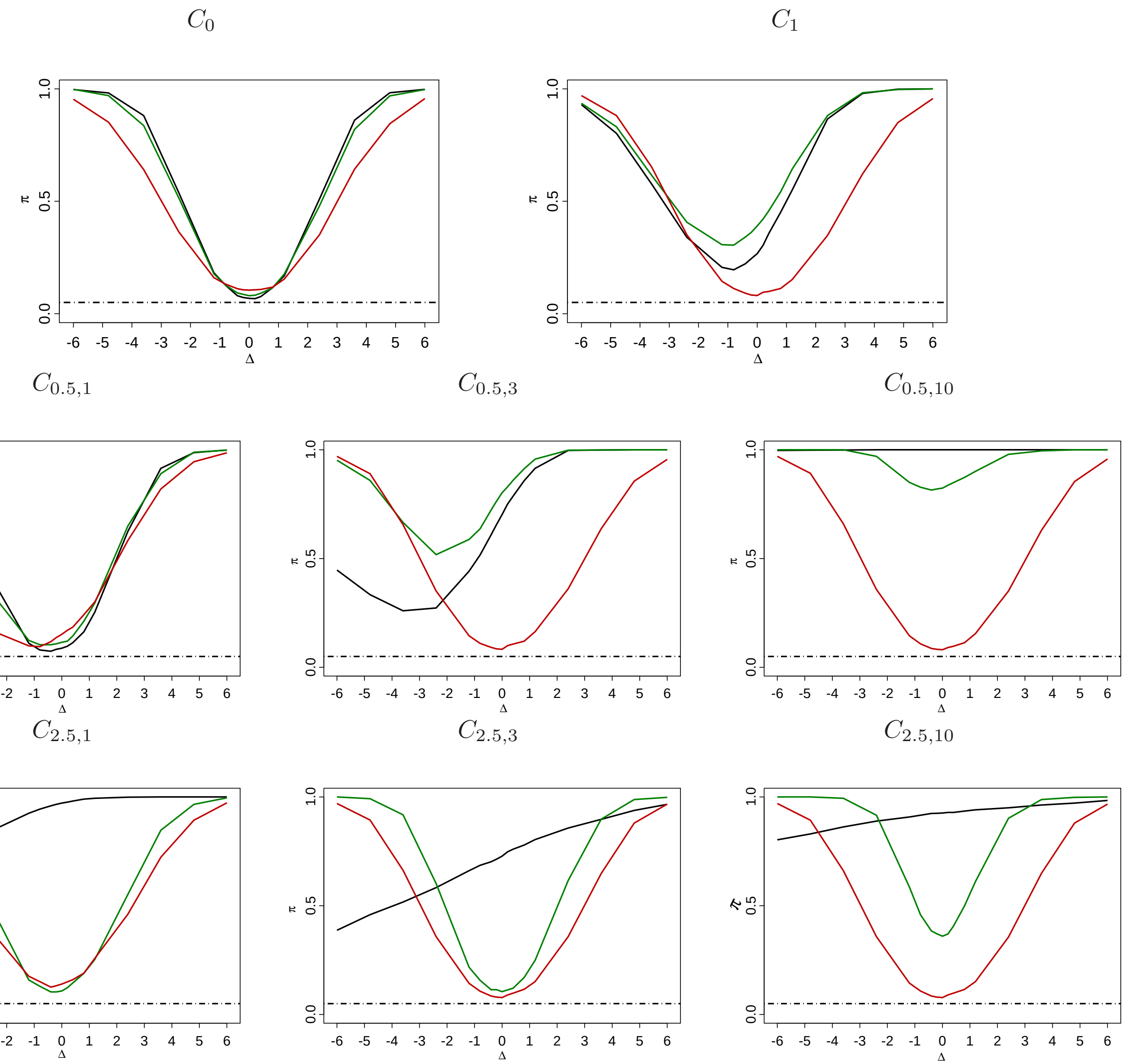

Figure 6: Observed frequencies of rejection under the Gamma model when $\tau = 1$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 1$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$.

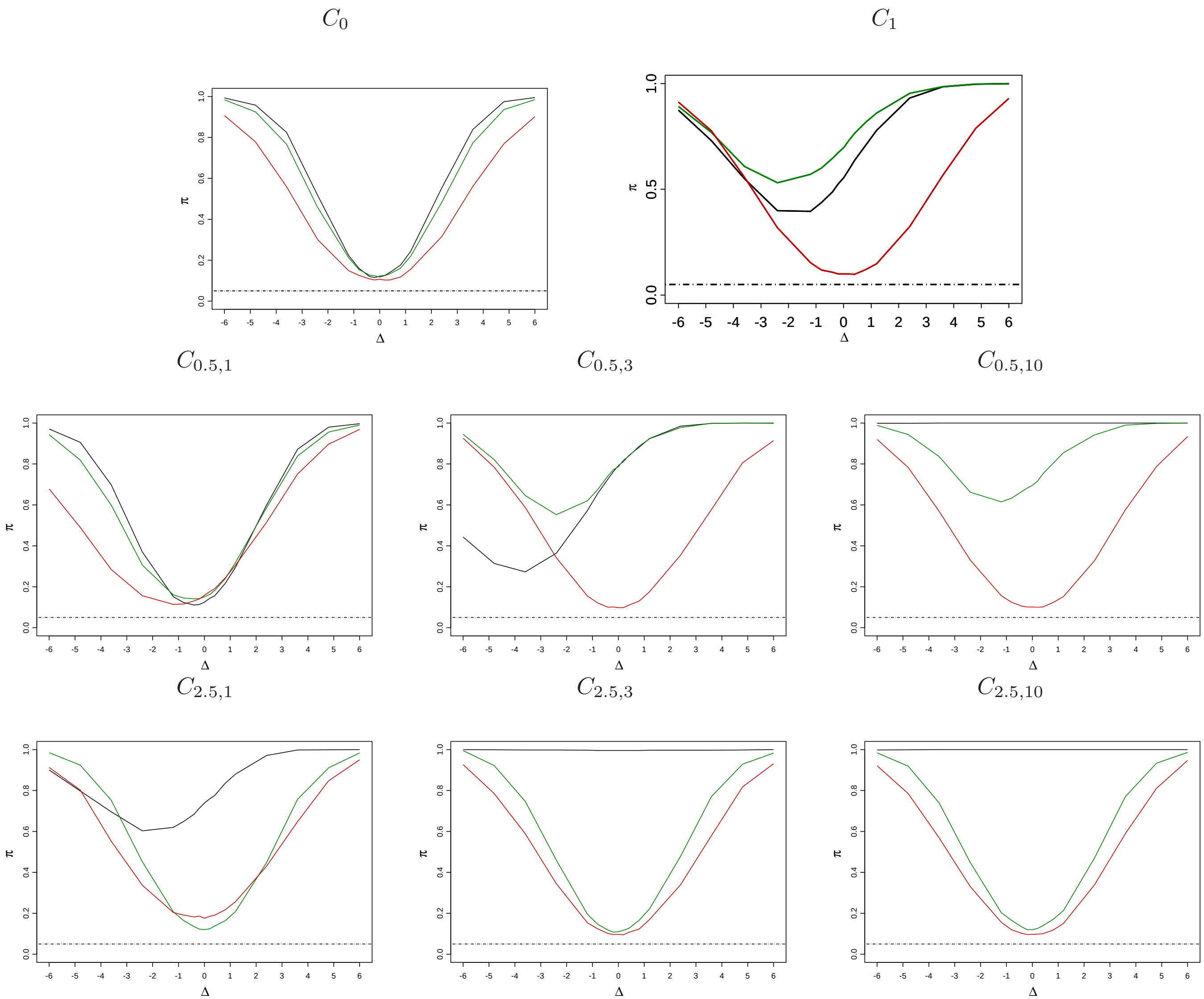

Figure 7: Observed frequencies of rejection under the Gamma model when $\tau = 1$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 0.8$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$.

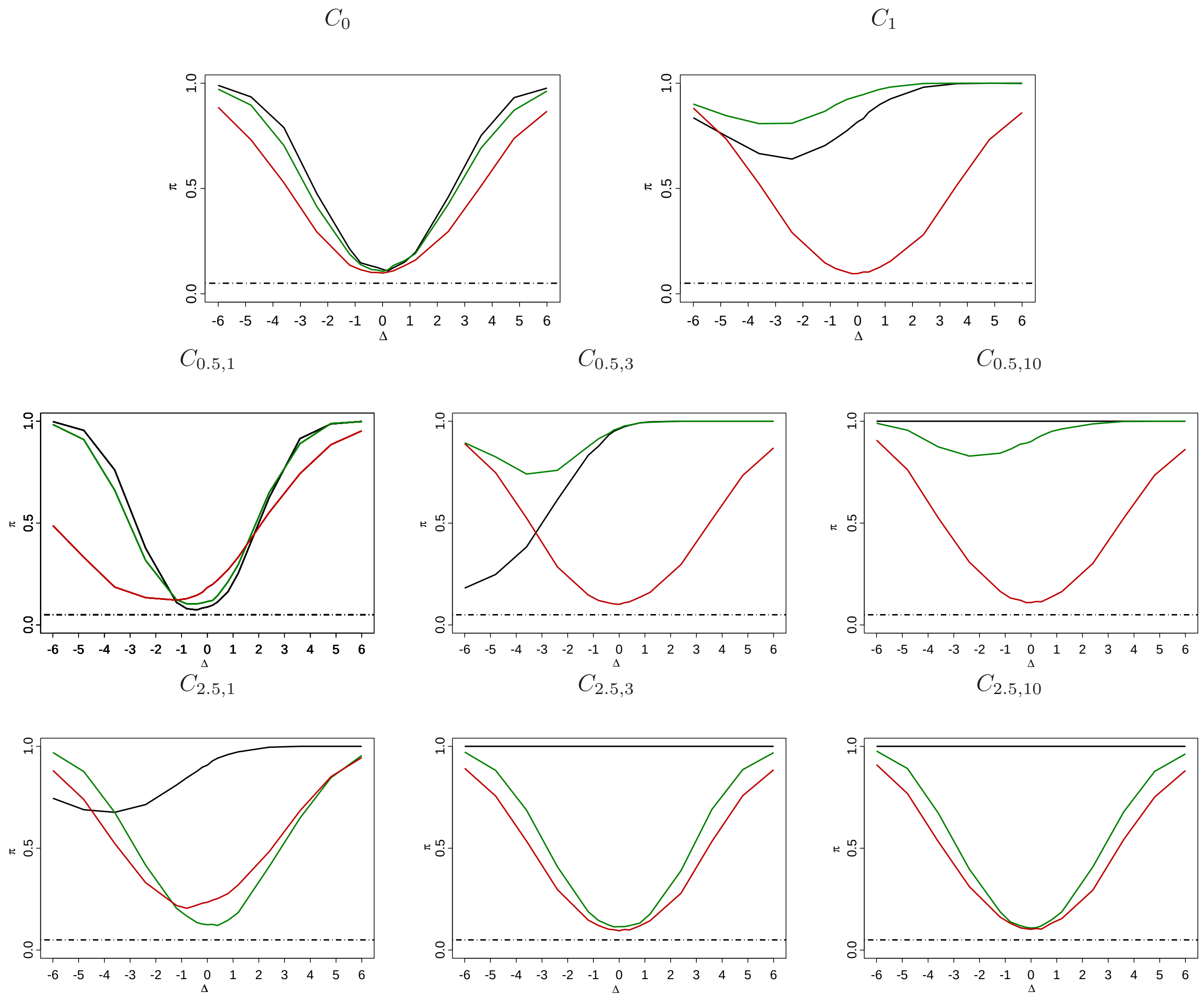

Figure 8: Observed frequencies of rejection under the Gamma model when $\tau = 1$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 1/(1+\exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}-2))$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$.

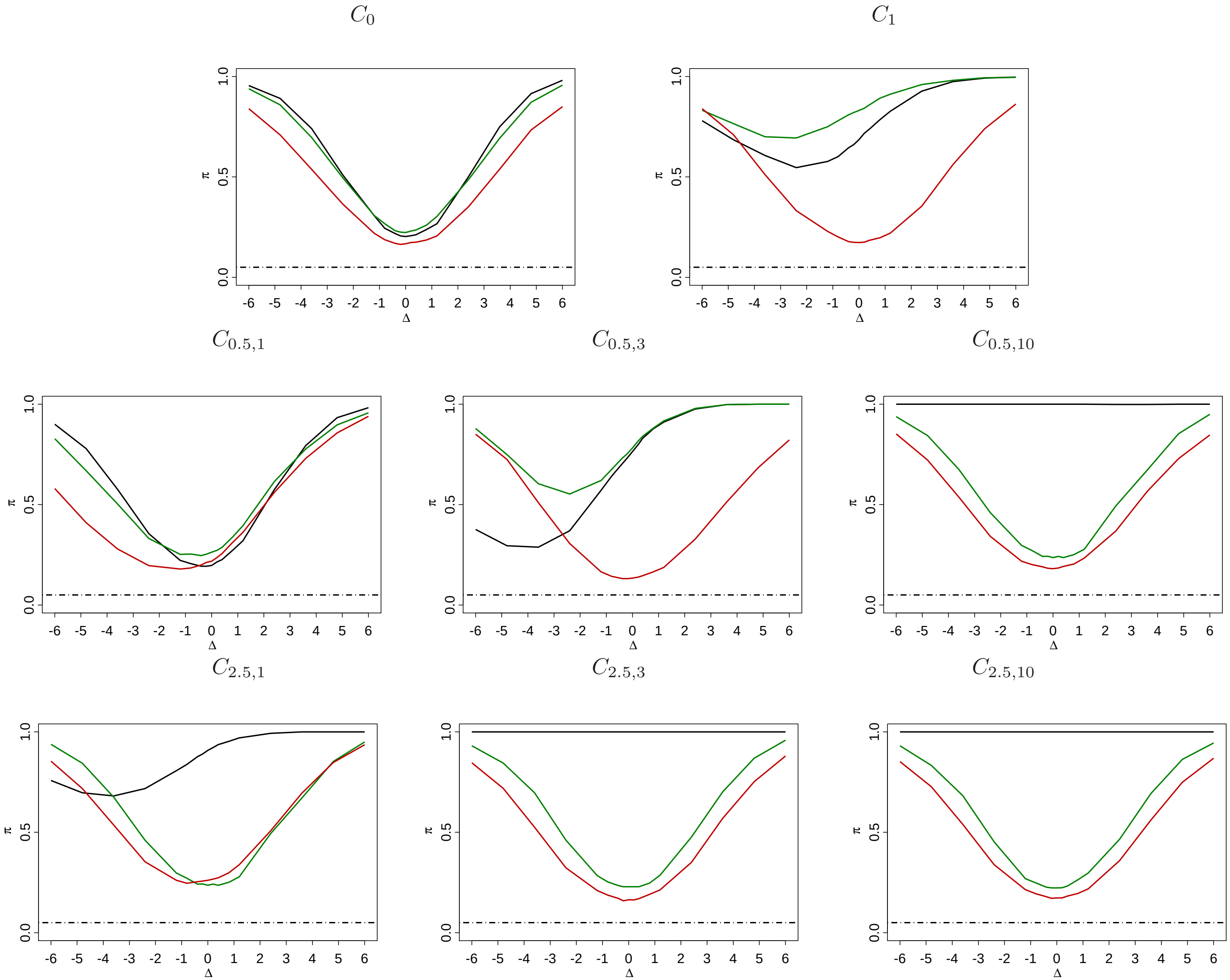

Figure 9: Observed frequencies of rejection under the Gamma model when $\tau = 1$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 1/(1+\exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} - 2))$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Propensity estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{P,ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{P,BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{P,TUK}}$.

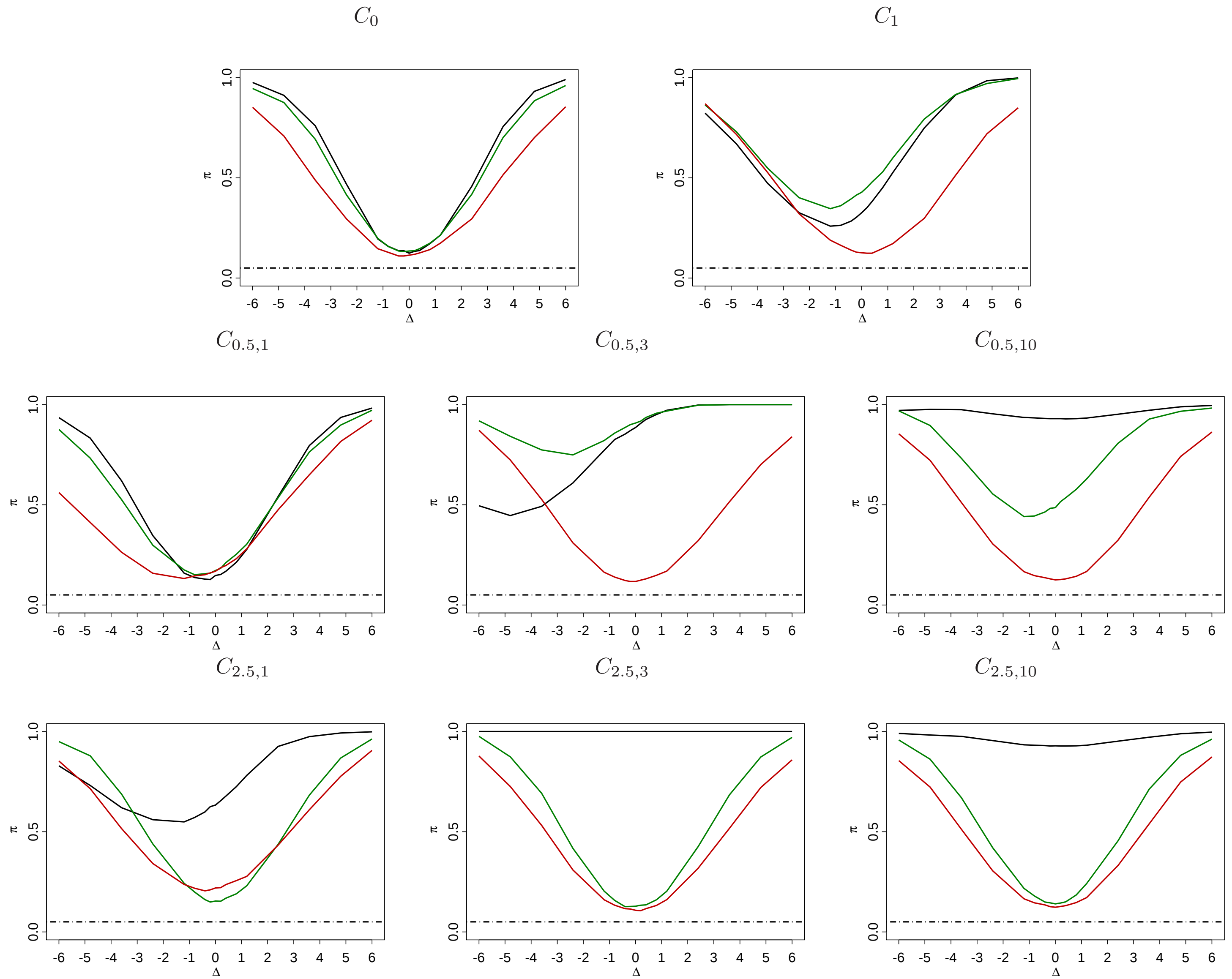

Figure 10: Observed frequencies of rejection under the Gamma model when $\tau = 1$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$.

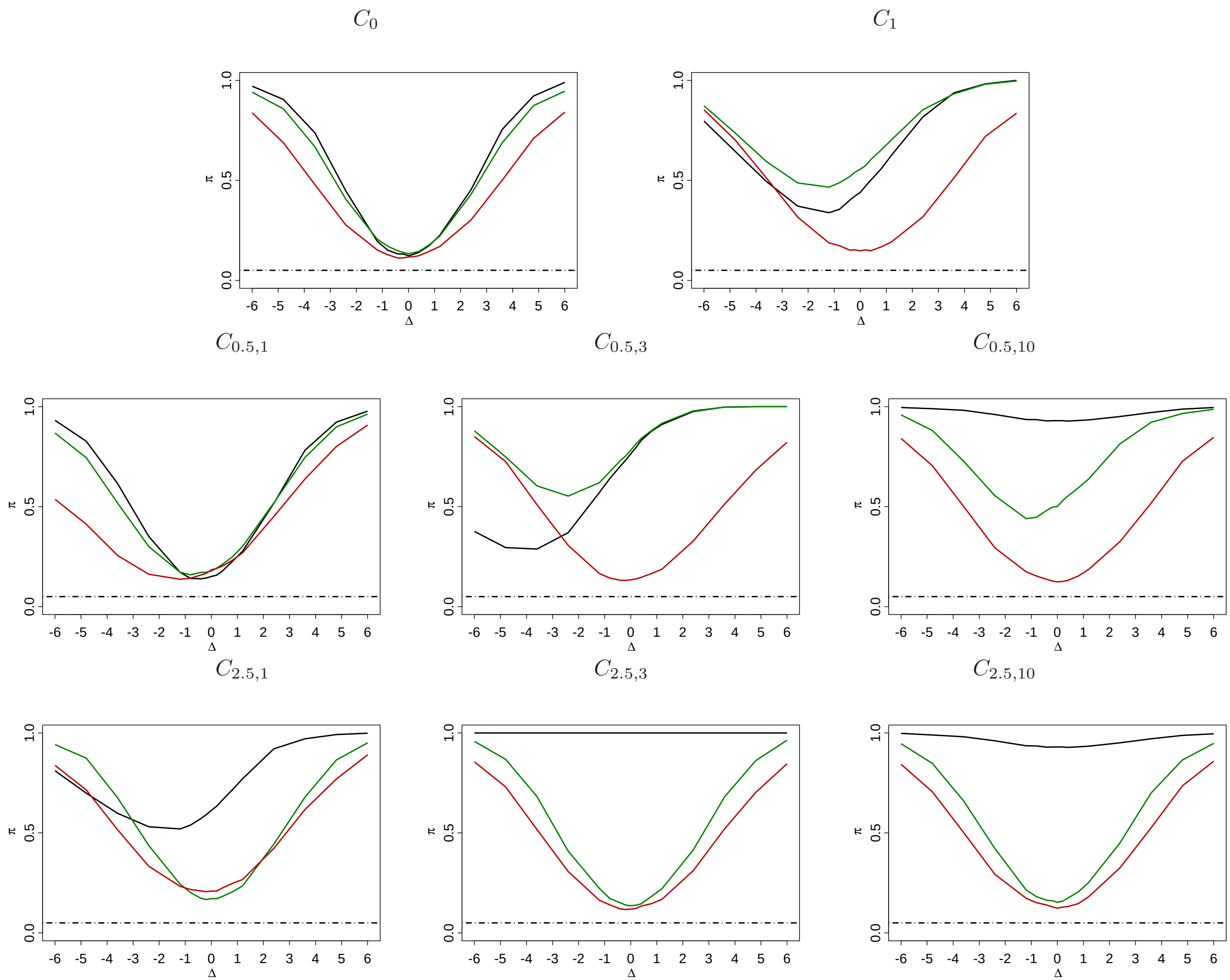

Figure 11: Observed frequencies of rejection under the Gamma model when $\tau = 1$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Propensity estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{P,ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{P,BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{P,TUK}}$.

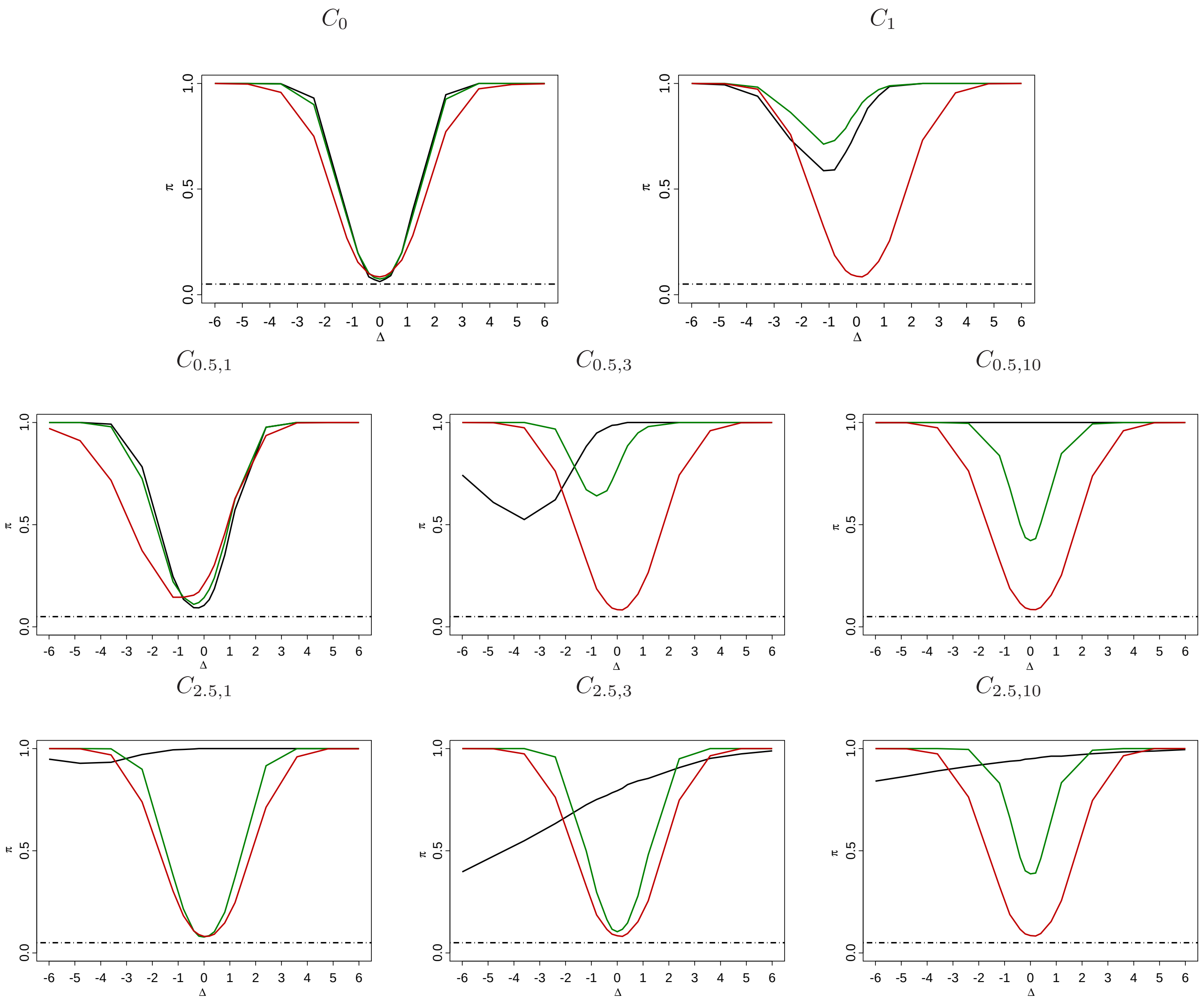

Figure 12: Observed frequencies of rejection under the Gamma model when $\tau = 3$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 1$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$.

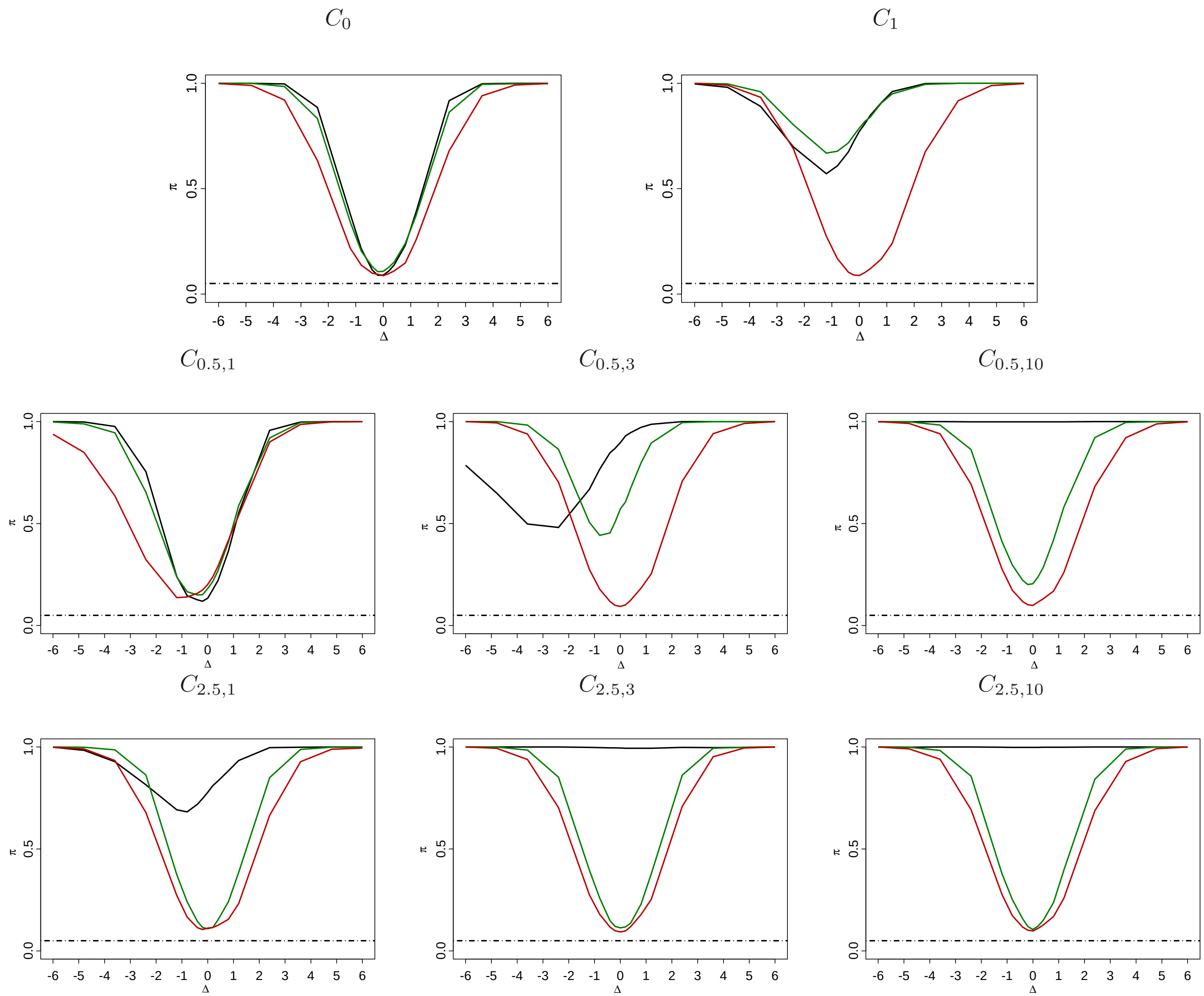

Figure 13: Observed frequencies of rejection under the Gamma model when $\tau = 3$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 0.8$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\rm ML}$, the robust estimators $\boldsymbol{\beta}_{\rm BGY}$ and their weighted version $\boldsymbol{\beta}_{\rm TUK}$.

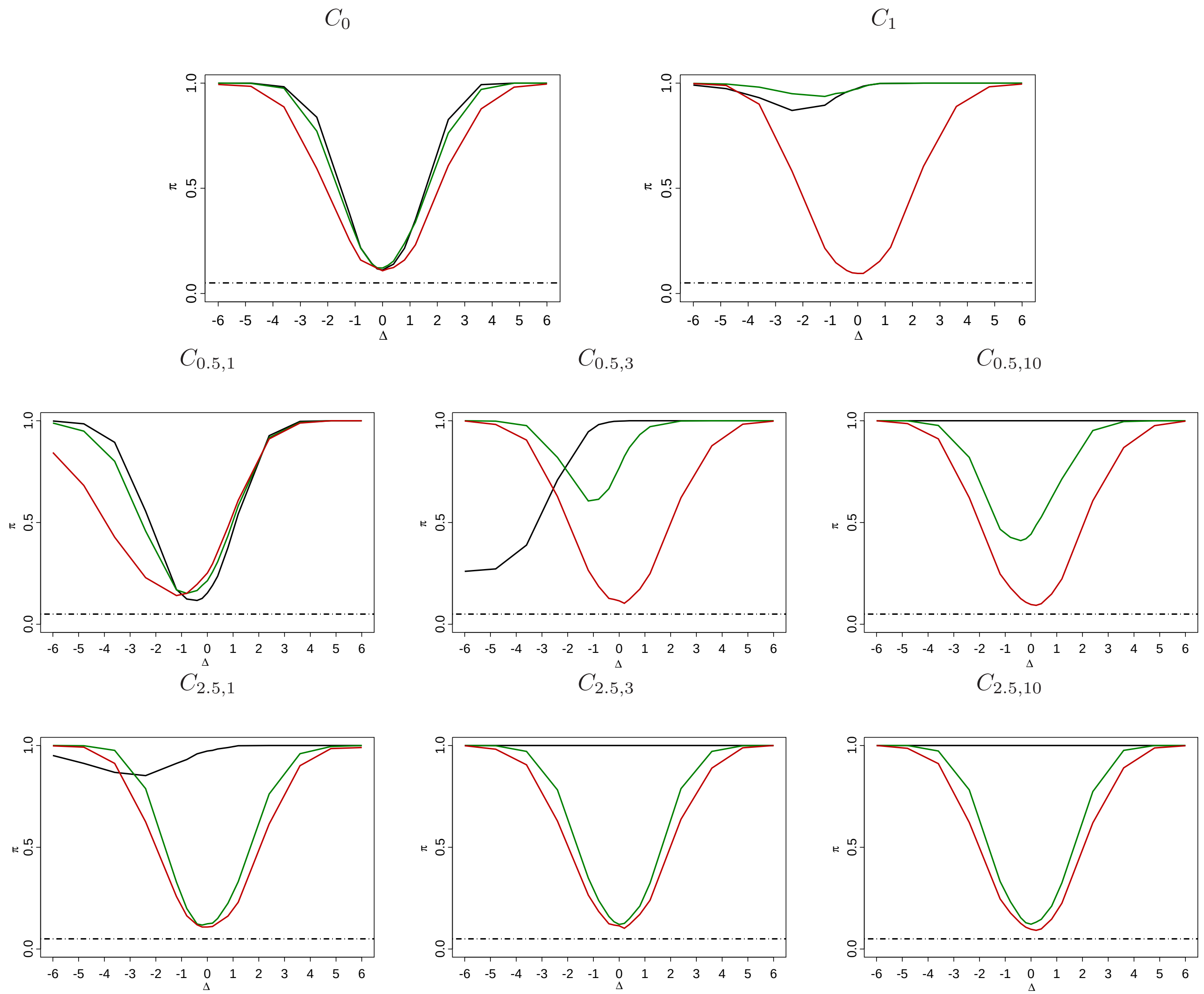

Figure 14: Observed frequencies of rejection under the Gamma model when $\tau = 3$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 1/(1+\exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}-2))$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$.

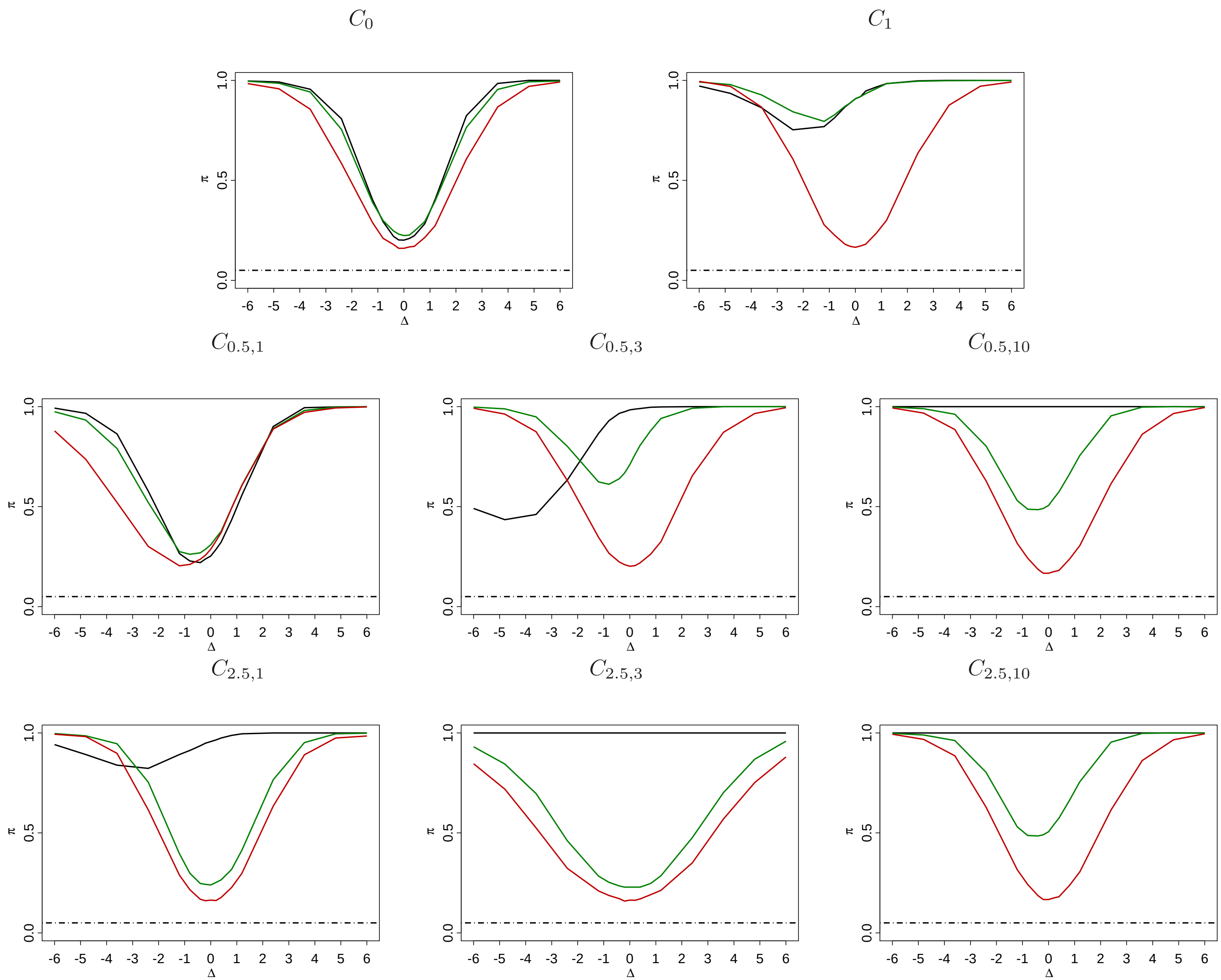

Figure 15: Observed frequencies of rejection under the Gamma model when $\tau = 1$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 1/(1+\exp(-\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}-2))$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Propensity estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{P,ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{P,BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{P,TUK}}$.

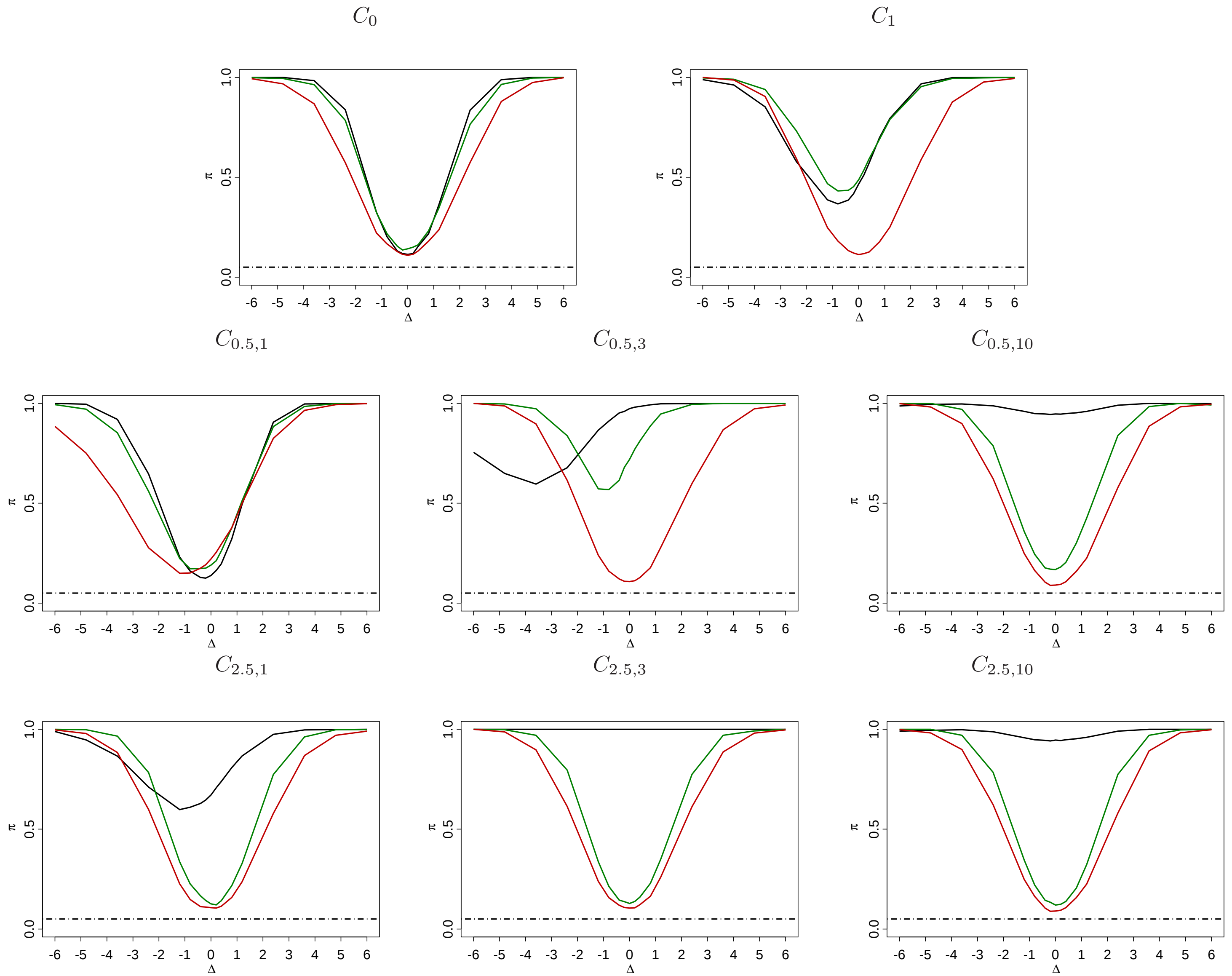

Figure 16: Observed frequencies of rejection under the Gamma model when $\tau = 3$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Simplified estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{TUK}}$.

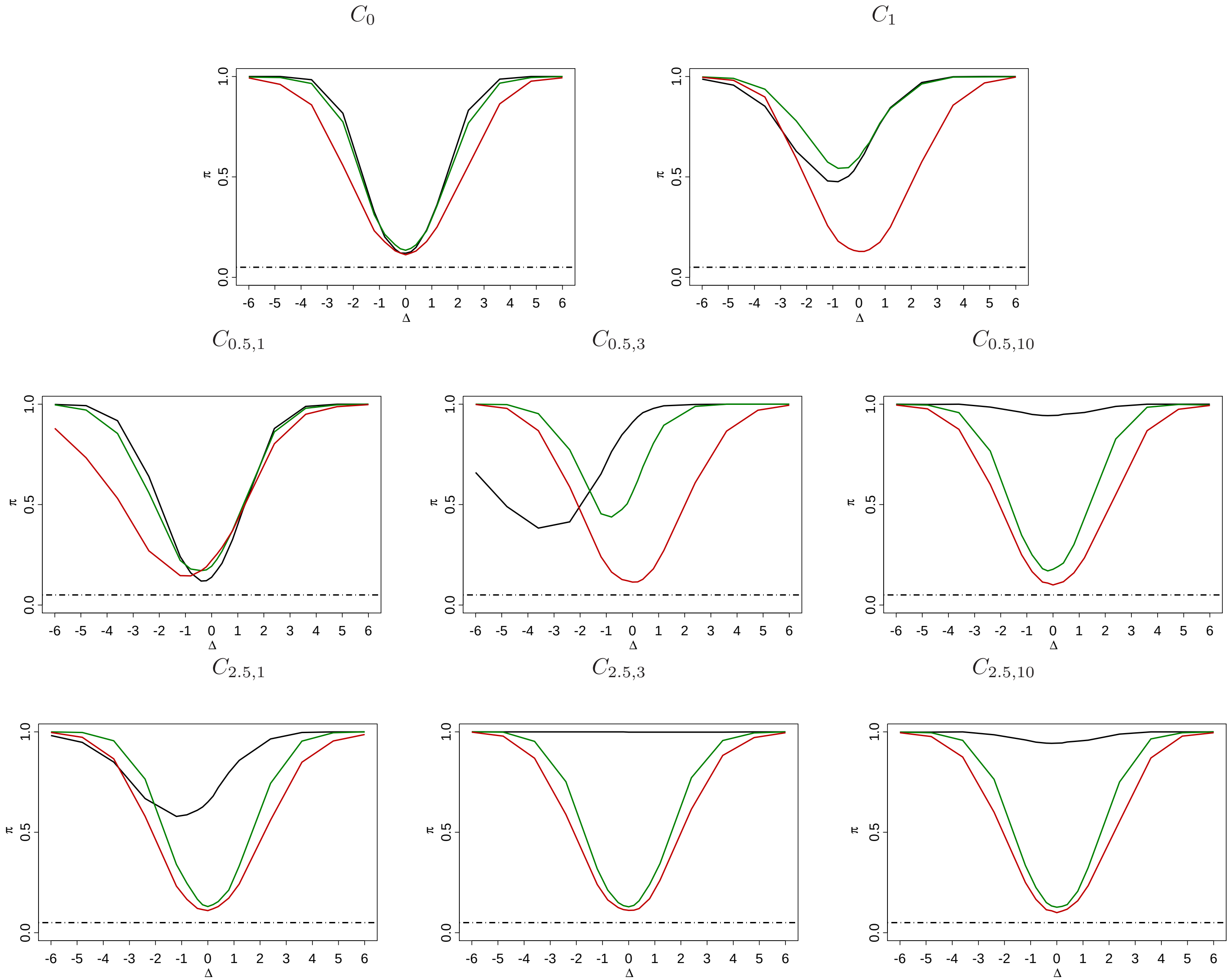

Figure 17: Observed frequencies of rejection under the Gamma model when $\tau = 3$, $c = \chi^2_{p,0.95}$, $p(\mathbf{x}) = 0.4 + 0.5(\cos(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + 0.4))^2$ with $\boldsymbol{\lambda} = (2,2)^{\mathrm{T}}$. Propensity estimators. The lines in black, green and red correspond to the test based on the classical estimators, $\boldsymbol{\beta}_{\mathrm{P,ML}}$, the robust estimators $\boldsymbol{\beta}_{\mathrm{P,BGY}}$ and their weighted version $\boldsymbol{\beta}_{\mathrm{P,TUK}}$.