Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n1p1

Fixed size confidence regions for the parameters of the mixed effects logistic regression model
By Zoubeidi, El-Bassiouni

Published: 26 April 2019

Electronic Journal of Applied Statistical Analysis
Vol. 12, Issue 01, April 2019, 1-13
DOI: 10.1285/i20705948v12n1p1

Fixed size confidence regions for the parameters of the mixed effects logistic regression model

Taoufik Zoubeidi* and Mohamed Y. El-Bassiouni

Department of Statistics, UAE University, Al-Ain, UAE

Published: 26 April 2019

We develop fixed size confidence regions for estimating the fixed and random effects parameters of the mixed effects logistic regression model. This model applies to, among others, the study of the effects of covariates on a dichotomous response variable when subjects are sampled in clusters. Two sequential procedures are developed to estimate with a prescribed accuracy (confidence level) and fixed precision the set of fixed and random effects parameters and linear transformations of these parameters, respectively. We show that the two procedures are asymptotically consistent (i.e., the coverage probability converges to the nominal confidence level) and asymptotically efficient (i.e., the ratio of the expected random sample size to the unknown best fixed sample size converges to 1) as the width of the confidence region converges to 0. Suggestions to improve the performance of the procedures are provided based on Monte Carlo simulation and illustrated through a longitudinal clinical trial data.

keywords: Mixed effects logistic regression model, sequential estimation, fixed width confidence estimation.

1 Introduction

Studies of the effects of covariates on a dichotomous response variable based on clustered data are very common in many fields including medicine, education, and social sciences. A data cluster may consist of repeated measures on a single subject (longitudinal data)

*Corresponding author: taoufikz@uaeu.ac.ae

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

or cross-sectional responses from a group of subjects sampled as a cluster (e.g., sampled from the same site or organization). This latter type of clustering encompasses multilevel and hierarchical data (Tuerlinckx et al., 2006). For example, Davis (1991) reports a multi-center randomized clinical trial comparing two treatments for a respiratory illness. The primary outcome, respiratory status (0 = poor, 1 = good), was assessed at 4 visits. The covariates were site, sex, age, and baseline respiratory status. In this trial, correlations between responses of the same subject as well as responses from subjects treated at the same site cannot be overruled a priori. A standard model for analyzing this type of data is the mixed effects logistic regression where the random effects components of the model are used to fit the dependency structure within clusters. Another example of application of this model can be found in Boccuzzo and Luca (2012).

In the context of experiments where subjects are sampled sequentially as is common in clinical trials and other fields, we develop in this manuscript fixed size confidence regions for the set of fixed and random effects parameters of the mixed effects logistic regression model, and for linear transformations of these parameters as well. These regions have a controlled maximum width and contain all the target parameters with a prescribed coverage probability. This dual control of the precision of estimation (width of region) and coverage probability can only be achieved through sequential or adaptive experimental designs.

Specifically, let $\mathbf{X}_{ij}$ be a vector of covariates and Y_{ij} be a dichotomous response observed on the jth subject from cluster $i = 1, \ldots, n$, where $1 \leq j \leq k_i$. Let $\mathbb{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{ik_i})$ be the matrix of covariates and $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ik_i})'$ be the vector of responses collected from cluster i. Assume that, conditional on $\mathbb{X}_i$ and a cluster specific random variable $u_i \in \mathbb{R}$, $Y_{i1}, \ldots, Y_{ik_i}$ are independent Bernoulli with $\mathrm{P}[Y_{ij} = 1 \,|\, \mathbf{X}_{ij}, u_i] = \mu_{ij}$. Assume further that $(\boldsymbol{Y}_1, u_1, \mathbb{X}_1), \ldots, (\boldsymbol{Y}_n, u_n, \mathbb{X}_n)$ are independent and identically distributed (i.i.d.), and the unobservable random effects $u_1, \ldots, u_n$ are i.i.d. $N(0, \sigma^2)$, $\sigma^2 \geq 0$.

Suppose that $\mu_{ij} = \mathrm{P}[Y_{ij} = 1 | \mathbf{X}_{ij}, u_i]$ follows the logistic mixed effects model

$$\ln\left[\frac{\mu_{ij}}{1-\mu_{ij}}\right] = \mathbf{X}'_{ij}\boldsymbol{\beta} + u_i, \qquad i \geq 1, \;\; j = 1, \ldots, k_i \tag{1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ are unknown fixed effects parameters.

<u>*Aim:*</u> We develop fixed size confidence regions with a prescribed accuracy (confidence level) for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ in Model (1) and its linear transform $\boldsymbol{\gamma} = A\boldsymbol{\theta}$ where is A is a $q \times (p+1)$ matrix of rank q ($q \leq p+1$).

There is a rich literature about the fixed precision interval estimation of the fixed effects of a Generalized Linear Model (GLM) which inherently assumes independent responses, i.e., without random effects (Chang, 1999; Chang and Martinsek, 1992). However, to the best of our knowledge, the problem undertaken in this manuscript, i.e, the fixed precision estimation of the parameters of the mixed effects logistic regression model has not been addressed. This latter model has a wider scope including clustered data such as longitudinal, multi-level and hierarchical data. More advanced relevant work in the literature concern the estimation of the parameters of the logistic regression model (GLM) with fixed precision under a Covariate-Adjusted Response-Adaptive

design where subjects are allocated to treatments based on values of their covariates and information from previous subjects (Chang and Park, 2013; Chambaz et al., 2015; Zhang and Hu, 2009). In other applications, Grabovsky and Chang (2003) derived a stopping rule for sequential adaptive tests with application to the computerized CAT, GRE and GMAT standard tests. Chang (2011) considered the sequential estimation of GLM when there are measurement errors under both adaptive and fixed designs. Chien et al. (2011) developed a two-stage sequential method to estimate with fixed precision the fixed effects of a retrospective (case-control) logistic regression model. Spiessens et al. (2002) developed a group sequential method for testing an ordinal logistic random-effects model when the random-effects distribution is misspecified.

This paper is structured as follows. In the next section we derive the fixed size confidence region for the vector of fixed and random effects $\boldsymbol{\theta}$ and describe its asymptotic properties. In Section 3, we deduce a fixed size confidence region for linear transforms of $\boldsymbol{\theta}$. Simulation results are presented in Section 4. In Section 5 we discuss some design considerations to improve the flexibility of the procedure. The proposed procedure is illustrated on data from a longitudinal clinical trial in Section 6. Conclusions are provided in Section 8 and the proofs of results are presented in the Appendix. In the sequel all calculations are done conditional on $\mathbf{X}_{ij}$, unless otherwise specified.

2 The fixed size confidence region for $\boldsymbol{\theta}$

Consider the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ based on a sample of n clusters using the Gauss-Hermite or Laplace quadratures. The derivation of the score function and information matrix is shown in the Appendix. Moreover, $\hat{\boldsymbol{\theta}}_n$ and its standard error can be computed using the glmmML package in the R statistical software (http://CRAN.R-project.org/package=glmmML) developed by Brostrom and Holmberg (2011). Sinha (2004) defined the regularity conditions for the consistency and asymptotic normality of the MLE in the Generalized Linear Mixed Model (GLMM). Under these regularity conditions, as $n \to +\infty$,

$$\Sigma_n^{1/2} \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right) \xrightarrow{\mathcal{L}} N \left(\mathbf{0}, \mathbb{I}_{p+1} \right), \tag{2}$$

where $\mathbb{I}_{p+1}$ is the identity matrix with $p+1$ rows, and $\Sigma_n^{1/2}$ is the Cholesky decomposition of the conditional Fisher Information matrix given $\mathbb{X}_1, \ldots, \mathbb{X}_n$.

As shown in (12), Σ_n is the sum of n cluster specific independent random matrices depending on the corresponding $\mathbb{X}_i$, namely, $\mathrm{E}_{\boldsymbol{Y}_i}\left[W_i(\boldsymbol{\theta}, \boldsymbol{Y}_i) \middle| \mathbb{X}_i \right]$ where $\mathrm{E}_{\boldsymbol{Y}_i}$ represents the expectation with respect to $\boldsymbol{Y}_i$. Then, by the i.i.d property of $\mathbb{X}_1, \mathbb{X}_2, \ldots$ and the Strong Law of Large Numbers, $\lim_{n \to +\infty} n^{-1} \Sigma_n = \Sigma$ where $\Sigma = \mathrm{E}_{\mathbb{X}_i} \left\{ \mathrm{E}_{\boldsymbol{Y}_i}\left[W_i(\boldsymbol{\theta}, \boldsymbol{Y}_i) \middle| \, \mathbb{X}_i \right] \right\}$.

Let $\hat{\Sigma}_n$ be Σ_n with $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\theta}}_n$. By (2) and the consistency of MLE,

$$\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right)' \hat{\Sigma}_n \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right) \xrightarrow{\mathcal{L}} \chi^2(p+1),$$

where $\chi^2(p+1)$ is the chi-square distribution with $p+1$ degrees of freedom. Let $\mathbf{R}_n$ be the ellipsoid defined by

$$\mathbf{R}_n = \left\{ \mathbf{Z} \in \mathbf{R}^{p+1} \,:\, \left(\mathbf{Z} - \hat{\boldsymbol{\theta}}_n\right)' \hat{\Sigma}_n \left(\mathbf{Z} - \hat{\boldsymbol{\theta}}_n\right) \leq d^2 \hat{\lambda}_n \right\}, \tag{3}$$

where d is a positive constant and $\hat{\lambda}_n$ is the smallest eigenvalue of $\hat{\Sigma}_n$. $\mathbf{R}_n$ has a maximum axis equal to $2d$ and $\lim_{n\to+\infty} n^{-1}\hat{\lambda}_n = \lambda$, the smallest eigenvalue of Σ.

Observe that if $d^2\hat{\lambda}_n \approx \chi^2_{1-\alpha}(p+1)$, then $\mathrm{P}[\boldsymbol{\theta} \in \mathbf{R}_n] \simeq 1-\alpha$ where $\chi^2_{1-\alpha}(p+1)$ is the $(1-\alpha)\times 100$th percentile of the $\chi^2(p+1)$ distribution. This suggests the stopping time

$$T_{\boldsymbol{\theta}} = \inf\left\{ n \geq 2 : d^2\hat{\lambda}_n \geq c_n \right\}, \tag{4}$$

where $c_n = \left(1 + cn^{-1}\right)\chi^2_{1-\alpha}(p+1)$ and $c > 0$ is a known moderating constant to avoid premature stopping. The proposed fixed size confidence region for $\boldsymbol{\theta}$ with a maximum axis of $2d$ and coverage probability converging to $1-\alpha$ when $d \to 0$ is

$$\mathbf{R}_{T_{\boldsymbol{\theta}}} = \left\{ \mathbf{Z} \in \mathbf{R}^{p+1} \,:\, \left(\mathbf{Z} - \hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}}\right)' \hat{\Sigma}_{T_{\boldsymbol{\theta}}} \left(\mathbf{Z} - \hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}}\right) \leq d^2 \hat{\lambda}_{T_{\boldsymbol{\theta}}} \right\}. \tag{5}$$

The asymptotic properties of the stopping time $T_{\boldsymbol{\theta}}$ and the confidence region $\mathbf{R}_{T_{\boldsymbol{\theta}}}$ are presented in the following two theorems whose proofs are postponed to the Appendix.

Theorem 1 *Under the regularity conditions for the consistency and asymptotic normality of MLE, see Sinha (2004), as* $d \to 0$,

i. $P\left[T_{\boldsymbol{\theta}} < +\infty\right] \to 1$ *a.s.*

ii. $T_{\boldsymbol{\theta}} \to +\infty$ *a.s.*

iii. $E\left[T_{\boldsymbol{\theta}}\right] \to +\infty$ *a.s.*

iv. $T_{\boldsymbol{\theta}}\lambda d^2 / c_{T_{\boldsymbol{\theta}}} \to 1$ *a.s.*

Theorem 2 *Under the regularity conditions for the consistency and asymptotic normality of MLE, see Sinha (2004), as* $d \to 0$,

i. $\hat{\Sigma}^{1/2}_{T_{\boldsymbol{\theta}}}\left(\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbb{I}_{p+1}\right)$,

ii. $\left(\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\right)'\hat{\Sigma}_{T_{\boldsymbol{\theta}}}\left(\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \chi^2(p+1)$,

iii. $P\left[\boldsymbol{\theta} \in \mathbf{R}_{T_{\boldsymbol{\theta}}}\right] \longrightarrow 1-\alpha$.

3 Fixed size confidence regions for linear transformations of $\boldsymbol{\theta}$

Experimenters may be interested only in some components of $\boldsymbol{\theta}$ or their linear combinations, e.g., the fixed effects parameters or their contrasts. When these parameters of interest are linear transformations of $\boldsymbol{\theta}$ of the form $\boldsymbol{\gamma} = A\boldsymbol{\theta}$ where A is a $q \times (p+1)$ matrix of rank q ($q \leq p+1$), fixed size confidence regions for $\boldsymbol{\gamma}$ are easily deduced from Section 2. The MLE, $\hat{\boldsymbol{\gamma}}_n$, of $\boldsymbol{\gamma}$ has a $N(\boldsymbol{\gamma},\, A\Sigma_n^{-1}A')$ distribution and $A\Sigma_n^{-1}A'$ is definite positive (d.f.) when Σ_n is d.f. Let $\hat{\delta}_n$ be the smallest eigenvalue of $[A\hat{\Sigma}_n^{-1}A']^{-1} = \hat{\Omega}_n$ (say). We define the $(1-\alpha)\times 100\%$ fixed size confidence region for $\boldsymbol{\gamma}$, with a maximum width of $2d$, by

$$\mathbf{Q}_{T_{\boldsymbol{\gamma}}} = \left\{ \mathbf{Z} \in \mathbf{R}^q \,:\, \left(\mathbf{Z} - \hat{\boldsymbol{\gamma}}_{T_{\boldsymbol{\gamma}}}\right)' \hat{\Omega}_{T_{\boldsymbol{\gamma}}} \left(\mathbf{Z} - \hat{\boldsymbol{\gamma}}_{T_{\boldsymbol{\gamma}}}\right) \leq d^2 \hat{\delta}_{T_{\boldsymbol{\gamma}}} \right\}, \tag{6}$$

where

$$T_{\boldsymbol{\gamma}} = \inf\left\{ n \geq 2 : d^2 \hat{\delta}_n \geq \tilde{c}_n \right\}, \tag{7}$$

where $\tilde{c}_n = \left(1 + cn^{-1}\right)\chi^2_{1-\alpha}(q)$. The results of theorems 1 and 2 apply to $T_{\boldsymbol{\gamma}}$ and $\mathbf{Q}_{T_{\boldsymbol{\gamma}}}$. The proof that Theorem 1 applies to $T_{\boldsymbol{\gamma}}$ is a straightforward application of (2) and Lemma 1 in the Appendix. The proof of the properties of $\mathbf{Q}_{T_{\boldsymbol{\gamma}}}$ follows along the lines of the proof of Theorem 2. The proofs of these results are omitted.

4 Simulation Study

To assess the performance of the proposed fixed size confidence regions for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma} = \boldsymbol{\beta}$ defined in (5) and (6), simulations were run in the statistical software R (version 3.1.2), using the glmmML package. The procedures were iterated 1000 times under a wide range of Model (1) parameter settings. Table 1 displays the results in the case of one covariate, i.e., $\mathbf{X}_{ij} = (1,\, X_{ij1})'$, with $X_{ij1} \sim N(-0.5\,,\, 0.05^2)$, balanced designs with equal cluster sizes $k_i = k = 10$ and 25, and unbalanced design where k_i is randomly selected from $\{10,\, 15,\, 20\}$, standard deviation of random effects $\sigma = 0.1$, fixed effects $\beta_0 = 2.5$, $\beta_1 = 0,\, 0.5,\, 1,\, 1.5,\, 3,\, 5$, and a maximum confidence region width of $2d = 10$. Table 2 displays the case of two correlated covariates, i.e., $\mathbf{X}_{ij} = (1,\, X_{ij1},\, X_{ij2})'$, where $(X_{ij1},\, X_{ij2})' \sim N\Big((-0.5, 0.5)',\, \left[(0.33^2,\, 0.1)', (0.1,\, 0.33^2)'\right]\Big)$, under unbalanced design $k_i \in \{10,\, 15,\, 20\}$, $\sigma = 0.2$, $\beta_0 = 2.2$, $(\beta_1,\, \beta_2) \in \{(0,0), (0.5,-1), (1,-1), (1,-2), (2,-2)\}$ and $d = 3$. The settings in tables 1 and 2 correspond to response probabilities $\mu_{ij} = \mathrm{P}[Y_{ij} = 1]$ varying from 0.5 to 0.95. We don't report the results for $\mu_{ij} < 0.5$ because simulation results are symmetric about $\mu_{ij} = 0.5$. The moderating constant, c, was set in proportion to the cluster size, that is, $c = 10$ when $k_i = 25$, $c = 20$ when $k_i \in \{10, 15, 20\}$, and $c = 40$ when $k_i = 10$. Higher values of c prevent premature stopping of the procedure which affects the coverage probability in small cluster sizes. The nominal coverage probability was set to 95%. In estimating $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, respectively, we calculated the average numbers of clusters $T_{\boldsymbol{\theta}}$ and $T_{\boldsymbol{\beta}}$, the average total sample sizes $N_{T_{\boldsymbol{\theta}}}$ and $N_{T_{\boldsymbol{\beta}}}$, the estimated coverage

probabilities of the confidence regions, $\hat{\mathrm{P}}\big[\boldsymbol{\theta} \in R_{T_{\boldsymbol{\theta}}}\big]$ and $\hat{\mathrm{P}}\big[\boldsymbol{\beta} \in \mathbf{Q}_{T_{\boldsymbol{\beta}}}\big]$, and the average Euclidean distance between the MLE at stopping and the corresponding parameter, $\|\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\|$ and $\|\hat{\boldsymbol{\beta}}_{T_{\boldsymbol{\beta}}} - \boldsymbol{\beta}\|$.

Results indicate that the coverage probabilities of $\mathbf{R}_{T_{\boldsymbol{\theta}}}$ and $\mathbf{Q}_{T_{\boldsymbol{\beta}}}$ are controlled for all values of β_1 and β_2, except when they are closer to 0 which corresponds to $\mu_{ij} = \mathrm{P}[Y_{ij} = 1]$ close to 1. This reflects the known deficiencies of MLEs when the estimated parameter is close to the boundary of the parameter space. However, adjusting the value of the moderating constant, c, in proportion to the cluster sample size k_i will improve the coverage probability when μ_{ij} is close to the boundary. Higher values of c prevent premature stopping, thus reducing instances of estimation errors. The number of sampled clusters and the total number of sampled subjects increase as μ_{ij} moves away from 0.5 to the boundary values of 0 and 1. The average distance between the MLE at stopping and the parameter is well under the maximum allowed distance of $2d$.

Table 1: Simulations of 95% confidence regions for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ under Model (1) where $\mathbf{X}_{ij} = (1,\, X_{ij1})'$, $X_{ij1} \sim N(-0.5\,,\, 0.05^2)$, $\sigma = 0.1$, $\beta_0 = 2.5$, $d = 5$

β_1	$T_{\boldsymbol{\theta}}$	$N_{T_{\boldsymbol{\theta}}}$	$\mathrm{P}\big[\boldsymbol{\theta} \in \mathbf{R}_{T_{\boldsymbol{\theta}}}\big]$	$\|\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\|$	$T_{\boldsymbol{\beta}}$	$N_{T_{\boldsymbol{\beta}}}$	$\mathrm{P}\big[\boldsymbol{\beta} \in \mathbf{Q}_{T_{\boldsymbol{\beta}}}\big]$	$\|\hat{\boldsymbol{\beta}}_{T_{\boldsymbol{\beta}}} - \boldsymbol{\beta}\|$
Constant $k_i = k = 25$ and $c = 10$								
0.0	81.1	2026.8	0.931	1.48	64.1	1603.0	0.941	1.65
0.5	67.6	1689.4	0.946	1.47	53.9	1346.5	0.942	1.67
1.0	57.1	1427.1	0.953	1.51	45.6	1141.0	0.949	1.68
3.0	34.4	858.8	0.976	1.33	27.9	697.4	0.966	1.48
5.0	29.2	730.8	0.971	1.35	23.9	597.8	0.967	1.48
Variable $k_i \in \{10, 15, 25\}$ and $c = 20$								
0.0	137.9	2070.9	0.914	1.53	109.6	1645.0	0.933	1.69
0.5	115.3	1728.8	0.920	3.03	92.0	1380.2	0.936	3.19
1.0	97.2	1461.2	0.960	1.44	77.8	1169.1	0.949	2.16
3.0	58.9	884.8	0.962	1.43	48.1	722.3	0.954	1.58
5.0	50.5	758.8	0.974	1.35	41.4	621.0	0.967	1.45
Constant $k_i = k = 10$ and $c = 40$								
0.0	215.1	2150.7	0.903	1.49	171.0	1709.5	0.947	2.55
0.5	179.9	1799.3	0.927	2.07	144.4	1443.5	0.958	2.18
1.0	153.2	1531.7	0.939	1.42	123.5	1234.6	0.956	1.57
3.0	93.9	939.3	0.970	1.35	77.0	769.5	0.964	1.47
5.0	81.0	810.2	0.971	1.32	66.8	668.1	0.962	1.42

5 Design Considerations

When designing experiments, it is important to carefully select the values of d and c because of their impact on the total sample size. Large values of c should be used for small cluster sizes as illustrated in the previous section. The choice of d is determined

by the desired precision of estimation and the maximum sample size, N, allowed by the time and budgetary constraints. Usually, the desired precision corresponds to a range of values of d, rather than a single value. In order to introduce more flexibility in choosing d and reduce the chance of sampling beyond N, a range of values of d could be pre-specified before sampling starts ranging from an acceptable, but not ideal, precision of estimation represented by a lager value of d to an ideal precision corresponding to a smaller value d. As sampling progresses, the stopping time will be evaluated at all the pre-specified values of d and the value of the stopping time that is smaller and closest to N, if any, will be used as the sample size for computing the fixed size confidence regions. In case the stopping time is greater than N for all values of d, then sampling will continue to the first stopping time instance. This latter scenario is an undesirable situation that may happen in open-ended sequential sampling.

Table 2: Simulations of 95% confidence regions for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ under Model (1) where $\mathbf{X}_{ij} = (1,\, X_{ij1},\, X_{ij2})'$, $(X_{ij1},\, X_{ij2})' \sim N\Big((-0.5, 0.5)',\, [(0.33^2,\, 0.1)', (0.1,\, 0.33^2)']\Big)$, $\sigma = 0.2$, $\beta_0 = 2.2$, $d = 3$

β_1	β_2	$T_{\boldsymbol{\theta}}$	$N_{T_{\boldsymbol{\theta}}}$	$\mathrm{P}\big[\boldsymbol{\theta} \in \mathbf{R}_{T_{\boldsymbol{\theta}}}\big]$	$\|\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\|$	$T_{\boldsymbol{\beta}}$	$N_{T_{\boldsymbol{\beta}}}$	$\mathrm{P}\big[\boldsymbol{\beta} \in \mathbf{Q}_{T_{\boldsymbol{\beta}}}\big]$	$\|\hat{\boldsymbol{\beta}}_{T_{\boldsymbol{\beta}}} - \boldsymbol{\beta}\|$
Variable $k_i \in \{10,\, 15,\, 25\}$ and $c = 20$									
0.0	0.0	105.9	1587.9	0.964	0.92	89.9	1348.0	0.956	0.98
0.5	-1.0	77.1	1157.0	0.976	0.91	66.0	989.9	0.968	0.93
1.0	-1.0	61.4	920.9	0.983	0.86	52.7	791.7	0.984	0.90
1.0	-2.0	53.3	799.8	0.984	0.86	46.0	690.3	0.977	0.88
2.0	-2.0	49.4	742.0	0.989	0.85	42.7	641.8	0.978	0.89

Let $d_1 > \cdots > d_K$ represent the pre-specified values of d for a fixed $K \geq 1$ such that $d_i \to 0$ and $d_i/d_{i'} \to 1$ for all $i,\, i' = 1, \ldots, K$. For example, set $d_i = d + a_i(o(d))$ where $a_i,\, i \geq 1$, are decreasing scalars. Let $T_{\boldsymbol{\theta},\, i} = T_{\boldsymbol{\theta}}$ and $T_{\boldsymbol{\gamma},\, i} = T_{\boldsymbol{\gamma}}$ when $d = d_i$ where $T_{\boldsymbol{\theta}}$ and $T_{\boldsymbol{\gamma}}$ are defined in (4) and (7). Observe that $T_{\boldsymbol{\theta},1} \leq \cdots \leq T_{\boldsymbol{\theta},K}$ and $T_{\boldsymbol{\gamma},1} \leq \cdots \leq T_{\boldsymbol{\gamma},K}$. When estimating $\boldsymbol{\gamma}$, sample clusters until the following stopping time

$$\tilde{T}_{\boldsymbol{\gamma}} = \max\Big\{T_{\boldsymbol{\gamma},\, i} I_{\{T_{\boldsymbol{\gamma},\, i} \leq N\}},\ i = 1, \ldots, K\Big\} + T_{\boldsymbol{\gamma},1} I_{\{T_{\boldsymbol{\gamma},\, 1} > N\}}, \tag{8}$$

and compute the fixed size region $\mathbf{R}_{\tilde{T}_{\boldsymbol{\theta}}}$ defined in (5). Similarly, when estimating $\boldsymbol{\theta}$, sample until $\tilde{T}_{\boldsymbol{\theta}} = \max\{T_{\boldsymbol{\theta},\, i} I_{\{T_{\boldsymbol{\theta},\, i} \leq N\}},\ i = 1, \ldots, K\} + T_{\boldsymbol{\theta},1} I_{\{T_{\boldsymbol{\theta},\, 1} > N\}}$, and compute the fixed size region $Q_{\tilde{T}_{\boldsymbol{\gamma}}}$ defined in (6). Then, Theorem 1 applies to $\tilde{T}_{\boldsymbol{\theta}}$ and $\tilde{T}_{\boldsymbol{\gamma}}$ and Theorem 2 applies to $\mathbf{R}_{\tilde{T}_{\boldsymbol{\theta}}}$ and $Q_{\tilde{T}_{\boldsymbol{\gamma}}}$, that is, both $\mathbf{R}_{\tilde{T}_{\boldsymbol{\theta}}}$ and $Q_{\tilde{T}_{\boldsymbol{\gamma}}}$ have asymptotic coverage probabilities equal to $1 - \alpha$, and maximum width d_i corresponding to $\tilde{T}_{\boldsymbol{\theta}} = T_{\boldsymbol{\theta},\, i}$ or $\tilde{T}_{\boldsymbol{\gamma}} = T_{\boldsymbol{\gamma},\, i}$, respectively. The proof of these results is postponed to the Appendix.

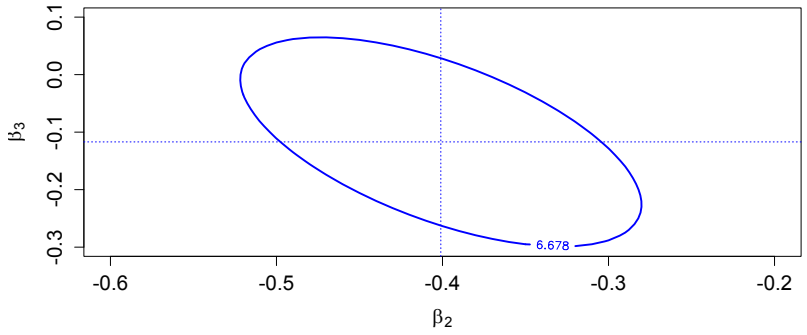
6 Illustrative Example

In this section we illustrate the application of the confidence regions using data from a longitudinal clinical trial comparing two oral treatments for toenail infection (der-

matophyte onychomycosis), see Backer et al. (1998). Patients were evaluated for the degree of onycholysis (the degree of separation of the nail plate from the nail-bed) at baseline (week 0) and at follow-up visits in weeks 4, 8, 12, 24, 36, and 48. The binary outcome variable, onycholysis (none or mild versus moderate or severe), was recorded for 146 and 148 patients assigned to treatments A and B, respectively, (data source http://www.blackwellpublishing.com/rss/ datasets/C4827r.txt). A question of interest was whether the percentage of severe infections decreased over time, and whether this evolution was different for the two treatment groups. Model (1) is a standard model for analyzing the presence of mild/severe onycholysis (represented by $Y_{ij} = 1$). The vector of covariates $\mathbf{X}_{ij}$ consists of Treatment (coded 1 for A and 0 for B), Time (exact time of the visit in months) and Treatment×Time. The correlations among repeated measures on the same patient are modeled by the random effects u_i. The parameters of interest are $\boldsymbol{\gamma} = (\beta_2, \beta_3)'$ corresponding to Time and Treatment×Time.

We set $c = 30$ and adopted the approach of Section 5 with $d_1 = 0.5$, $d_2 = 0.4$, $d_3 = 0.2$ and $d_4 = 0.1$. Using the glmmML package, we sequentially sampled $\tilde{T}_{\boldsymbol{\gamma}} = T_{\boldsymbol{\gamma},3} = 276$ subjects corresponding to $d_3 = 0.2$, as defined in (8). While $T_{\boldsymbol{\gamma},i}$, $i = 1, 2, 3$ stopped before $N = 294$, $T_{\boldsymbol{\gamma},4}$ did not. The MLE of $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \sigma^2)$ is $\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\gamma},3}} = (-2.580, -0.258, -0.401, -0.117, 4.590)$. The elements of the matrix $\hat{\Omega}_{T_{\boldsymbol{\gamma},3}}$ defined in (6) are $\hat{\Omega}_{T_{\boldsymbol{\gamma},3}}(1,1) = 713.639$, $\hat{\Omega}_{T_{\boldsymbol{\gamma},3}}(1,2) = 284.414$, and $\hat{\Omega}_{T_{\boldsymbol{\gamma},3}}(2,2) = 314.904$, and its smallest eigenvalue $\hat{\delta}_{T_{\boldsymbol{\gamma},3}} = 166.940$. The resulting fixed size confidence region for $\boldsymbol{\gamma} = (\beta_2, \beta_3)$ with a 95% confidence level and maximum width $2d_3 = 0.4$ is given by $Q_{T_{\boldsymbol{\gamma},3}}$ in (6) and it is displayed in Figure 1. Observe that the MLEs of β_2 and β_3 are negatively correlated. Moreover, $Q_{T_{\boldsymbol{\gamma},3}}$ contains $\beta_3 = 0$ and only negative values of β_2. We therefore conclude at a 5% experimentwise significance level that the percentage of severe infections declines with time and that the rate of decrease over time is the same for both treatments.

Figure 1: 95% confidence region with maximum width 0.4 for β_2 & β_3



7 Conclusion

We introduced fixed size confidence regions for the set of parameters of the mixed effects logistic regression model. The regions, which can be readily computed using the glmmML package in R, allow the experimenter to control both the precision and the familywise confidence level in estimating the fixed and random effects parameters of the model

or their linear combinations. These results extend existing methods to the analysis of clustered data such as longitudinal and hierarchical data. Suggestions about the selection of the stopping time moderating constant, c, and the maximum width of the confidence region, d, are provided to improve the performance of the procedure.

References

Backer, M. D., Vroey, C. D., Lesaffre, E., Scheys, I., and Keyser, P. D. (1998). Twelve weeks of continuous oral therapy for toenail onychomycosis caused by dermatophytes: A double-blind comparative trial of terbinafine 250 mg/day versus itraconazole 200 mg/day. *Journal of the American Academy of Dermatology*, 38:57–63.

Boccuzzo, G. and Luca, F. D. (2012). Sudden infant death syndrome: knowledge of its risk factors among italian healthcare professionals. *Electronic Journal of Applied Statistical Analysis*, 5(3):374–380.

Brostrom, G. and Holmberg, H. (2011). Generalized linear models with clustered data: Fixed and random effects models. *Computational Statistics and Data Analysis*, 55.

Chambaz, A., Laan, M. V. D., and Zheng, W. (2015). Targeted covariate-adjusted response-adaptive lasso-based randomized controlled trials. In Sverdlov, O., editor, *Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects*, pages 345–368. CRC.

Chang, Y. C. and Martinsek, A. T. (1992). Fixed size confidence regions for parameters of a logistic regression model. *Annals of Statistics*, 20:175–191.

Chang, Y. I. (1999). Strong consistency of maximum quasi-likelihood estimate in generalized linear models via a last time. *Statistics & Probability Letters*, 45:237–246.

Chang, Y. I. (2011). Sequential estimation in generalized linear models when covariates are subject to errors. *Metrika*, 73:93–120.

Chang, Y. I. and Park, E. (2013). Sequential estimation for covariate-adjusted response-adaptive designs. *Journal of the Korean Statistical Society*, 42:105–116.

Chien, C., Chang, Y. I., and Hsueh, H. (2011). Optimal sampling in retrospective logistic regression via two-stage method. *Biometrical Journal*, 53:5–18.

Davis, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine*, 10:1959–1980.

Govindarajulu, Z. (2004). *Sequential Statististics*. World Scientific Publishing Co., Singapore.

Grabovsky, I. and Chang, H. H. (2003). Deriving a stopping rule for sequential adaptive tests. http://www.iacat.org/sites/default/files/biblio/gr01-01.pdf. Accessed on 17-11-2011.

Sinha, S. K. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*, 99:451–460.

Spiessens, B., Lesaffre, E., Verbeke, G., and Kim, K. (2002). Group sequential meth-

ods for an ordinal logistic random-effects model under misspecification. *Biometrics*, 58:569–575.

Tuerlinckx, F., Rijmen, F., Verbeke, G., and Boeck, P. D. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statististical Psychology*, 59:225–255.

Zhang, L. X. and Hu, F. F. A. (2009). New family of covariate-adjusted response-adaptive designs and their properties. *Applied Mathematics Journal of Chinese Universities*, 24(1):1–13.

8 Appendix

8.1 *Computation of the score function and the Fisher information matrix*

The full likelihood function of $(\boldsymbol{Y}_1, u_1), \ldots, (\boldsymbol{Y}_n, u_n)$, given $\mathbb{X}_1, \ldots, \mathbb{X}_n$, is

$$\begin{aligned}
&\prod_{i=1}^{n} L_i(\boldsymbol{\theta}, u_i \boldsymbol{Y}_i) \\
&= \exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{k_i}\left[Y_{ij}\ln\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) + \ln(1-\mu_{ij})\right]\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i^2\right\} \frac{\sigma^{-n}}{(2\pi)^{n/2}} \\
&= \exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{k_i}\left[Y_{ij}\left(\mathbf{X}_{ij}'\boldsymbol{\beta} + u_i\right) - \ln\left(1+\exp\left(\mathbf{X}_{ij}'\boldsymbol{\beta} + u_i\right)\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i^2\right\} \frac{\sigma^{-n}}{(2\pi)^{n/2}} \qquad (9)
\end{aligned}$$

The MLE of $\boldsymbol{\theta}$ is obtained by integrating out $u_1, \ldots, u_n$ in (9), then approximating the logarithm of the integral using a Laplace quadrature, i.e.,

$$\begin{aligned}
l\left(\boldsymbol{\theta}, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\right) &= \sum_{i=1}^{n} \ln \int_{-\infty}^{+\infty} L_i\left(\boldsymbol{\theta}, u_i, \boldsymbol{Y}_i\right) du_i \\
&\approx \sum_{i=1}^{n} \ln\left[\sqrt{2\pi}\hat{\omega}_i L_i\left(\boldsymbol{\theta}, \hat{u}_i(\boldsymbol{\theta}), \boldsymbol{Y}_i\right)\right] \qquad (10)
\end{aligned}$$

where $\hat{u}_i(\boldsymbol{\theta})$ is the maxima of $g_i\big(\boldsymbol{\theta}, u, \boldsymbol{Y}_i\big) = \ln L_i(\boldsymbol{\theta}, u, \boldsymbol{Y}_i\big)$ w.r.t. u for fixed $\boldsymbol{\theta}$, that is,

$$\frac{\partial}{\partial u} g_i\left(\boldsymbol{\theta}, \hat{u}_i(\boldsymbol{\theta}), \boldsymbol{Y}_i\right) = 0, \qquad \text{and} \qquad \hat{\omega}_i = \hat{\omega}_i\left(\boldsymbol{\theta}\right) = \left[-\frac{\partial^2}{\partial u^2} g_i\left(\boldsymbol{\theta}, \hat{u}_i\left(\boldsymbol{\theta}\right), Y_i\right)\right]^{-1/2}.$$

After straightforward calculations we find the following expressions of the score function and hessian of (10). The score function is

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}} l\left(\boldsymbol{\theta}, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\right) &= \sum_{i=1}^{n}\left[-\frac{\frac{\partial^3}{\partial u^2 \partial \boldsymbol{\theta}} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right)}{2\frac{\partial^2}{\partial u^2} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right)} + \frac{\partial}{\partial \boldsymbol{\theta}} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right)\right] \\ &= \sum_{i=1}^{n} \mathbf{Z}_i\left(\boldsymbol{\theta}, \boldsymbol{Y}_i\right)\end{aligned} \tag{11}$$

where $\mathbf{Z}_i\big(\boldsymbol{\theta}, \boldsymbol{Y}_i\big)$ is the summand term in the right hand side (r.h.s.) of (11). The hessian of (10) is

$$\begin{aligned}\Sigma_n &= \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \\ &\sum_{i=1}^{n} \frac{\nabla_{uu\boldsymbol{\theta}\boldsymbol{\theta}'} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right) \nabla_{uu} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right) - \nabla_{uu\boldsymbol{\theta}'} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right)\left[\nabla_{uu\boldsymbol{\theta}} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right)\right]'}{2\left[\nabla_{uu} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right)\right]^2} \\ &+ \sum_{i=1}^{n} \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} g_i\left(\boldsymbol{\theta}, \hat{u}_i, \boldsymbol{Y}_i\right) = \sum_{i=1}^{n} W_i\left(\boldsymbol{\theta}, \boldsymbol{Y}_i\right) \quad \text{(say)},\end{aligned} \tag{12}$$

where $\nabla_{\mathbf{v}_1 \ldots \mathbf{v}_k} f(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ is the differential operator consisting of the k^{th} partial derivatives with respect to the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

8.2 *Proof of theorems 1 and 2*

The proof of Theorem 1 relies on the following lemma of Chow and Robbins (Govindarajulu, 2004, Lemma 4.7.1, p. 196).

Lemma 1 *Let $U_1, U_2, \ldots$ be any sequence of random variables such that $U_n > 0$ a.s. and $\lim_{n\to+\infty} U_n = 1$. Also, let a_n be a sequence of integers such that $a_n > 0$, $\lim_{n\to+\infty} a_n = +\infty$ and $\lim_{n\to+\infty} a_n/a_{n-1} = 1$. Define $N = N(t) = \inf\big\{k \geq 1 : U_k \leq a_k/t\big\}$, then*

1. *N is well defined and non-decreasing as a function of t,*
2. *$\lim_{t\to+\infty} N(t) = +\infty$ a.s.,*
3. *$\lim_{t\to+\infty} E\big(N(t)\big) = +\infty$,*
4. *$\lim_{t\to+\infty} a_N/t = 1$ a.s.*

Proof of Theorem 1: The stopping time $T_{\boldsymbol{\theta}}$ may be written as

$$\begin{aligned}T_{\boldsymbol{\theta}} &= \inf\left\{n \geq 2 :\ d^2\hat{\lambda}_n \geq c_n\right\} = \inf\left\{n \geq 2 :\ \frac{n\lambda}{\hat{\lambda}_n} \leq \frac{n\lambda}{c_n} d^2\right\} \\ &= \inf\left\{n \geq 2 :\ U_n \leq \frac{a_n}{t}\right\},\end{aligned}$$

where $U_n = n\lambda/\hat{\lambda}_n$, $a_n = n\lambda/c_n$, and $t = d^{-2}$ which satisfy the conditions of Lemma 1.

Proof of Theorem 2: Note that the second and third results of the theorem are direct consequences of the first result. Thus, we will only prove the first result.

By expanding the score function (11) w.r.t. $\hat{\boldsymbol{\theta}}_n$ and, by the property of the MLE, setting the first term of the expansion to zero, we get

$$\sum_{i=1}^{n} \mathbf{Z}_i\left(\boldsymbol{\theta}, \boldsymbol{Y}_i\right) = \tilde{\boldsymbol{\Delta}}_n\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \quad \text{where} \quad \tilde{\boldsymbol{\Delta}}_n = \sum_{i=1}^{n} W_i\left(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{Y}_i\right), \tag{13}$$

and $\tilde{\boldsymbol{\theta}}_n$ is between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}_n$. Observe that $\mathbf{Z}_i\big(\boldsymbol{\theta}, \boldsymbol{Y}_i\big)$, $i \geq 1$, are i.i.d. and, by the property of score functions, $\mathrm{E}\big[\mathbf{Z}_i\big] = \mathrm{E}\big[\mathbf{Z}_i\big(\boldsymbol{\theta}, \boldsymbol{Y}_i\big)\big|\mathbb{X}_i\big] = \mathbf{0}$ and $\mathrm{Var}\big[\mathbf{Z}_i\big(\boldsymbol{\theta}, \boldsymbol{Y}_i\big)\big] < +\infty$. This implies that $n^{-1}\sum_{i=1}^{n} \mathbf{Z}_i\big(\boldsymbol{\theta}, \boldsymbol{Y}_i\big)$ are uniformly continuous in probability. Since by Theorem 1, $T_{\boldsymbol{\theta}} \to +\infty$, as $d \to 0$, then by Anscombe's Theorem,

$$T_{\boldsymbol{\theta}}^{-1/2}\Sigma^{-1/2}\sum_{i=1}^{T_{\boldsymbol{\theta}}} \mathbf{Z}_i \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbb{I}_{p+1}\right). \tag{14}$$

Observe that

$$\begin{aligned}
\hat{\Sigma}_{T_{\boldsymbol{\theta}}}^{1/2}\left(\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\right) &= \hat{\Sigma}_{T_{\boldsymbol{\theta}}}^{1/2}\tilde{\boldsymbol{\Delta}}_{T_{\boldsymbol{\theta}}}^{-1}\tilde{\boldsymbol{\Delta}}_{T_{\boldsymbol{\theta}}}\left(\hat{\boldsymbol{\theta}}_{T_{\boldsymbol{\theta}}} - \boldsymbol{\theta}\right) \\
&= \hat{\Sigma}_{T_{\boldsymbol{\theta}}}^{1/2}\tilde{\boldsymbol{\Delta}}_{T_{\boldsymbol{\theta}}}^{-1}\sum_{i=1}^{T_{\boldsymbol{\theta}}} \mathbf{Z}_i \\
&= \left[\left(T_{\boldsymbol{\theta}}^{-1}\hat{\Sigma}_{T_{\boldsymbol{\theta}}}\right)^{1/2}\left(T_{\boldsymbol{\theta}}^{-1}\tilde{\boldsymbol{\Delta}}_{T_{\boldsymbol{\theta}}}\right)^{-1} - \Sigma^{-1/2}\right] T_{\boldsymbol{\theta}}^{-1/2}\sum_{i=1}^{T_{\boldsymbol{\theta}}} \mathbf{Z}_i + T_{\boldsymbol{\theta}}^{-1/2}\Sigma^{-1/2}\sum_{i=1}^{T_{\boldsymbol{\theta}}} \mathbf{Z}_i.
\end{aligned} \tag{15}$$

The first term on the r.h.s. of (15) converges to 0 in probability, as $d \to 0$, because

$$\left(T_{\boldsymbol{\theta}}^{-1}\hat{\Sigma}_{T_{\boldsymbol{\theta}}}\right)^{1/2}\left(T_{\boldsymbol{\theta}}^{-1}\tilde{\boldsymbol{\Delta}}_{T_{\boldsymbol{\theta}}}\right)^{-1} - \Sigma^{-1/2} \to 0$$

in probability and $T_{\boldsymbol{\theta}}^{-1/2}\sum_{i=1}^{T_{\boldsymbol{\theta}}} \mathbf{Z}_i$ converges in distribution. The second term on the r.h.s. converges in distribution to $N\big(\mathbf{0}, \mathbb{I}_{p+1}\big)$ by (14).

8.3 *Proof of the results of Section 5*

Proof of the Results of Theorem 1 for $\tilde{T}_{\boldsymbol{\theta}}$ and $\tilde{T}_{\boldsymbol{\gamma}}$: The proofs for $\tilde{T}_{\boldsymbol{\theta}}$ and $\tilde{T}_{\boldsymbol{\gamma}}$ are identical, therefore we will only show the proof for $\tilde{T}_{\boldsymbol{\theta}}$. Observe that $T_{\boldsymbol{\theta},1} \leq \tilde{T}_{\boldsymbol{\theta}} \leq T_{\boldsymbol{\theta},K}$. Result (i) follows from $P[T_{\boldsymbol{\theta}} < +\infty] \geq P[T_{\boldsymbol{\theta},1} < +\infty] \to 1$ as $d_1 \to 0$. Results (ii) and three follow

from $\tilde{T}_{\boldsymbol{\theta}} \geq T_{\boldsymbol{\theta},1} \to +\infty$ *a.s.* and $\mathrm{E}[\tilde{T}_{\boldsymbol{\theta}}] \geq \mathrm{E}[T_{\boldsymbol{\theta},1}] \to +\infty$ as $d_1 \to 0$. Result (iv) follows from

$$\lim_{d_1 \to 0} \frac{\tilde{d}\, c_{T_{\boldsymbol{\theta},1}}}{d_1\, c_{\tilde{T}_{\boldsymbol{\theta}}}} \frac{T_{\boldsymbol{\theta},1} \lambda d_1}{c_{T_{\boldsymbol{\theta},1}}} \leq \lim_{d_1 \to 0} \frac{\tilde{T}_{\boldsymbol{\theta}} \lambda \tilde{d}}{c_{\tilde{T}_{\boldsymbol{\theta}}}} \leq \lim_{d_1 \to 0} \frac{\tilde{d}\, c_{T_{\boldsymbol{\theta},K}}}{d_K\, c_{\tilde{T}_{\boldsymbol{\theta}}}} \frac{T_{\boldsymbol{\theta},K} \lambda d_K}{c_{T_{\boldsymbol{\theta},K}}} \tag{16}$$

where $\tilde{d}$ is equal to d_i such that $\tilde{T}_{\boldsymbol{\theta}} = T_{\boldsymbol{\theta},i}$. Both the r.h.s and l.h.s of (16) converge to 1 a.s.

Proof of the Results of Theorem 2 for $\mathbf{R}_{\tilde{T}_{\boldsymbol{\theta}}}$ *and* $Q_{\tilde{T}_{\boldsymbol{\gamma}}}$: The proof of Theorem 2 remains unchanged as long as the result of Theorem 1 applies to $\tilde{T}_{\boldsymbol{\theta}}$ and $\tilde{T}_{\boldsymbol{\gamma}}$, and d_i is independent of the data $(\boldsymbol{Y}_j, u_j, \mathbb{X}_j)$, $j \geq 1$.