MPRA
Munich Personal RePEc Archive

Confidence sets based on penalized maximum likelihood estimators

Pötscher, Benedikt M. and Schneider, Ulrike

Department of Statistics, University of Vienna

June 2008

Online at https://mpra.ub.uni-muenchen.de/9062/
MPRA Paper No. 9062, posted 11 Jun 2008 07:29 UTC

Confidence Sets Based on Penalized Maximum Likelihood Estimators

Benedikt M. Pötscher and Ulrike Schneider*
Department of Statistics, University of Vienna

Preliminary version: February 2008
This version: June 2008

Abstract

The finite-sample coverage properties of confidence intervals based on penalized maximum likelihood estimators like the LASSO, adaptive LASSO, and hard-thresholding are analyzed. It is shown that symmetric intervals are the shortest. The length of the shortest intervals based on the hard-thresholding estimator is larger than the length of the shortest interval based on the adaptive LASSO, which is larger than the length of the shortest interval based on the LASSO, which in turn is larger than the standard interval based on the maximum likelihood estimator. In the case where the penalized estimators are tuned to possess the 'sparsity property', the intervals based on these estimators are larger than the standard interval by an order of magnitude. A simple asymptotic confidence interval construction in the 'sparse' case, that also applies to the smoothly clipped absolute deviation estimator, is also discussed.

MSC Subject Classifications: Primary 62F25; secondary 62C25, 62J07.

Keywords: penalized maximum likelihood, Lasso, adaptive Lasso, hard-thresholding, confidence set, coverage probability, sparsity, model selection.

1 Introduction

Recent years have seen an increased interest in penalized maximum likelihood (least squares) estimators. Prominent examples of such estimators are the LASSO estimator (Tibshirani (1996)) and its variants like the adaptive LASSO (Zou (2006)), the Bridge estimators (Frank and Friedman (1993)), or the smoothly clipped absolute deviation (SCAD) estimator (Fan and Li (2001)). In linear regression models with orthogonal regressors, the hard- and soft-thresholding

*Department of Statistics, University of Vienna, Universitätsstrasse 5, A-1010 Vienna. Phone: +431 427738640. E-mail: benedikt.poetscher@univie.ac.at

estimators can also be reformulated as penalized least squares estimators, with the soft-thresholding estimator then coinciding with the LASSO estimator.

The asymptotic distributional properties of penalized maximum likelihood (least squares) estimators have been studied in the literature, mostly in the context of a finite-dimensional linear regression model; see Knight and Fu (2000), Fan and Li (2001), and Zou (2006). Knight and Fu (2000) study the asymptotic distribution of Bridge estimators and, in particular, of the LASSO estimator. Their analysis concentrates on the case where the estimators are tuned in such a way as to perform conservative model selection, and their asymptotic framework allows for dependence of parameters on sample size. In contrast, Fan and Li (2001) for the SCAD estimator and Zou (2006) for the adaptive LASSO estimator concentrate on the case where the estimators are tuned to possess the 'sparsity' property. They show that, with such tuning, these estimators possess what has come to be known as the 'oracle property'. However, their results are based on a fixed-parameter asymptotic framework only. Pötscher and Leeb (2007) and Pötscher and Schneider (2007) study the finite-sample distribution of the hard-thresholding, the soft-thresholding (LASSO), the SCAD, and the adaptive LASSO estimator under normal errors; they also obtain the asymptotic distributions of these estimators in a general 'moving parameter' asymptotic framework. The results obtained in these two papers clearly show that the distributions of the estimators studied are often highly non-normal and that the so-called 'oracle property' typically paints a misleading picture of the actual performance of the estimator.[1]

A natural question now is what all these distributional results mean for confidence intervals that are based on penalized maximum likelihood (least squares) estimators. This is the question we address in the present paper in the context of a normal linear regression model with orthogonal regressors. We obtain formulae for the finite-sample infimal coverage probabilities of fixed-width confidence intervals based on one of the following estimators: hard-thresholding, LASSO (soft-thresholding), and adaptive LASSO. We show that among those intervals the symmetric ones are the shortest, and we show that hard-thresholding leads to longer intervals than the adaptive LASSO, which in turn leads to longer intervals than the LASSO. All these intervals are longer than the standard confidence interval based on the maximum likelihood estimator, which is in line with Joshi (1969). In case the estimators are tuned to possess the 'sparsity' property, explicit asymptotic formulae for the length of the confidence intervals are furthermore obtained, showing that in this case the intervals based on the penalized maximum likelihood estimators are larger by an order of magnitude than the standard maximum likelihood based interval. This refines, for the particular estimators considered, a general result for 'sparse' estimators (Pötscher (2007)). Additionally, in the 'sparsely' tuned case a simple asymptotic construction of confidence intervals is provided that also applies to other penalized

[1] In the wake of Fan and Li (2001) a considerable literature has sprung up establishing the so-called 'oracle property' for a variety of estimators. All these results are fixed-parameter asymptotic results only and can be very misleading. See Leeb and Pötscher (2008) and Pötscher (2007) for more discussion.

maximum likelihood estimators such as the SCAD estimator.

2 The Model and Estimators

For a normal linear regression model with orthogonal regressors and known error variance, distributional properties of penalized maximum likelihood (least squares) estimators with a separable penalty can be reduced without loss of generality to the case of a Gaussian location problem; for details see, e.g., Pötscher and Schneider (2007). Hence, we may suppose that the data $y_1, \ldots, y_n$ are independent identically distributed as $N(\theta, 1)$, $\theta \in \mathbb{R}$, and $n \geq 1$. We shall be concerned with confidence sets based on penalized maximum likelihood estimators such as the hard-thresholding estimator, the LASSO (reducing to soft-thresholding), and the adaptive LASSO estimator. The hard-thresholding estimator $\hat{\theta}_H$ is given by

$$\hat{\theta}_H = \bar{y}\mathbf{1}(|\bar{y}| > \eta_n)$$

where the threshold η_n is a positive real number and $\bar{y}$ denotes the maximum likelihood estimator, i.e., the arithmetic mean of the data. The LASSO (or soft-thresholding) estimator $\hat{\theta}_S$ is given by

$$\hat{\theta}_S = \operatorname{sign}(\bar{y})(|\bar{y}| - \eta_n)_+.$$

Here $\operatorname{sign}(x)$ is defined as -1, 0, and 1 in case $x < 0$, $x = 0$, and $x > 0$, respectively, and z_+ is shorthand for $\max\{z, 0\}$. The adaptive LASSO estimator $\hat{\theta}_A$ is here given by

$$\hat{\theta}_A = \bar{y}(1 - \eta_n^2/\bar{y}^2)_+ = \begin{cases} 0 & \text{if } |\bar{y}| \leq \eta_n \\ \bar{y} - \eta_n^2/\bar{y} & \text{if } |\bar{y}| > \eta_n. \end{cases}$$

It coincides with the nonnegative Garotte in this simple model.

3 Confidence Intervals: Finite Sample Results

We are interested in the finite-sample coverage properties of intervals of the form $[\hat{\theta} - a_n, \hat{\theta} + b_n]$ where a_n and b_n are nonnegative real numbers and $\hat{\theta}$ stands for one of the estimators $\hat{\theta}_H$, $\hat{\theta}_S$, or $\hat{\theta}_A$. We also consider one-sided intervals $(-\infty, \hat{\theta} + c_n]$ and $[\hat{\theta} - c_n, \infty)$ with $0 \leq c_n < \infty$. In the following let $P_{n,\theta}$ denote the distribution of the sample when θ is the true parameter. Furthermore, let Φ denote the standard normal cumulative distribution function.

3.1 Soft-thresholding

Let $C_{S,n}$ denote the interval $[\hat{\theta}_S - a_n, \hat{\theta}_S + b_n]$. We first determine the infimum of the covering probability $p_{S,n}(\theta) = P_{n,\theta}(\theta \in C_{S,n})$ of this interval.

Proposition 1 *The infimal coverage probability of the interval* $C_{S,n} = [\hat{\theta}_S - a_n, \hat{\theta}_S + b_n]$ *is given by*

$$\inf_{\theta\in\mathbb{R}} p_{S,n}(\theta) = \begin{cases} \Phi(n^{1/2}(a_n - \eta_n)) - \Phi(n^{1/2}(-b_n - \eta_n)) & \text{if } a_n \le b_n \\ \Phi(n^{1/2}(b_n - \eta_n)) - \Phi(n^{1/2}(-a_n - \eta_n)) & \text{if } a_n > b_n. \end{cases} \quad (1)$$

Proof. Using the expression for the finite sample distribution of $n^{1/2}(\hat{\theta}_S - \theta)$ given in Pötscher and Leeb (2007) and noting that this distribution function has a jump at $-n^{1/2}\theta$ we obtain

$$\begin{aligned} p_{S,n}(\theta) &= [\Phi(n^{1/2}(a_n - \eta_n)) - \Phi(n^{1/2}(-b_n - \eta_n))]\mathbf{1}(\theta < -a_n) \\ &+ [\Phi(n^{1/2}(a_n + \eta_n)) - \Phi(n^{1/2}(-b_n - \eta_n))]\mathbf{1}(-a_n \le \theta \le b_n) \\ &+ [\Phi(n^{1/2}(a_n + \eta_n)) - \Phi(n^{1/2}(-b_n + \eta_n))]\mathbf{1}(b_n < \theta). \end{aligned}$$

It follows that $\inf_{\theta\in\mathbb{R}} p_{S,n}(\theta)$ is as given in the proposition. ■

As a point of interest we note that $p_{S,n}(\theta)$ is a piecewise constant function with jumps at $\theta = -a_n$ and $\theta = b_n$.

Remark 2 *(i) If we consider the open interval* $C^o_{S,n} = (\hat{\theta}_S - a_n, \hat{\theta}_S + b_n)$ *the formula for the coverage probability becomes*

$$\begin{aligned} P_{n,\theta}\left(\theta \in C^o_{S,n}\right) &= [\Phi(n^{1/2}(a_n - \eta_n)) - \Phi(n^{1/2}(-b_n - \eta_n))]\mathbf{1}(\theta \le -a_n) \\ &+ [\Phi(n^{1/2}(a_n + \eta_n)) - \Phi(n^{1/2}(-b_n - \eta_n))]\mathbf{1}(-a_n < \theta < b_n) \\ &+ [\Phi(n^{1/2}(a_n + \eta_n)) - \Phi(n^{1/2}(-b_n + \eta_n))]\mathbf{1}(b_n \le \theta). \end{aligned}$$

As a consequence, the infimal coverage probability of $C^o_{S,n}$ *is again given by (1). A fortiori, the half-open intervals* $(\hat{\theta}_n - a_n, \hat{\theta}_n + b_n]$ *and* $[\hat{\theta}_n - a_n, \hat{\theta}_n + b_n)$ *then also have infimal coverage probability given by (1).*

(ii) It is not difficult to see that the one-sided intervals $(-\infty, \hat{\theta}_S + c_n]$, $(-\infty, \hat{\theta}_S + c_n)$, $[\hat{\theta}_S - c_n, \infty)$, *and* $(\hat{\theta}_S - c_n, \infty)$, *with* c_n *a nonnegative real number, have infimal coverage probability* $\Phi(n^{1/2}(c_n - \eta_n))$.

3.2 Hard-thresholding

Let $C_{H,n}$ denote the interval $[\hat{\theta}_H - a_n, \hat{\theta}_H + b_n]$. The infimum of the covering probability $p_{H,n}(\theta) = P_{n,\theta}(\theta \in C_{H,n})$ of this interval has been obtained in Proposition 9 in Pötscher (2007), which is summarized in the following proposition.

Proposition 3 *The infimal coverage probability of the interval* $C_{H,n} = [\hat{\theta}_H - a_n, \hat{\theta}_H + b_n]$ *is given by*

$$\inf_{\theta\in\mathbb{R}} p_{H,n}(\theta) = \begin{cases} \Phi(n^{1/2}(a_n - \eta_n)) - \Phi(-n^{1/2}b_n) & \text{if } \eta_n \le a_n + b_n \text{ and } a_n \le b_n \\ \Phi(n^{1/2}(b_n - \eta_n)) - \Phi(-n^{1/2}a_n) & \text{if } \eta_n \le a_n + b_n \text{ and } a_n > b_n \\ 0 & \text{if } \eta_n > a_n + b_n \end{cases} \quad (2)$$

For later use we observe that the interval $C_{H,n}$ has positive infimal coverage probability if and only if the length of the interval $a_n + b_n$ is larger than η_n. As a point of interest we also note that the coverage probability $p_{H,n}(\theta)$ is discontinuous (discontinuity points at $\theta = -a_n$ and $\theta = b_n$). Furthermore, as discussed in Pötscher (2007), the infimum in (2) is attained if $\eta_n > a_n + b_n$, but not in case $\eta_n \leq a_n + b_n$.

Remark 4 *(i) If we consider the open interval $C^o_{H,n} = (\hat{\theta}_H - a_n, \hat{\theta}_H + b_n)$ the coverage probability satisfies*

$$P_{n,\theta}\left(\theta \in C^o_{H,n}\right) = P_{n,\theta}\left(\theta \in C_{H,n}\right) - [\mathbf{1}(\theta = b_n) + \mathbf{1}(\theta = -a_n)][\Phi(n^{1/2}(-\theta + \eta_n)) - \Phi(n^{1/2}(-\theta - \eta_n))].$$

Inspection of the proof of Proposition 9 in Pötscher (2007) then shows that $C^o_{H,n}$ has the same infimal coverage probability as $C_{H,n}$. However, now the infimum is always a minimum. Furthermore, the half-open intervals $(\hat{\theta}_H - a_n, \hat{\theta}_H + b_n]$ and $[\hat{\theta}_H - a_n, \hat{\theta}_H + b_n)$ then a fortiori have infimal coverage probability given by (2); for these intervals the infimum is attained if $\eta_n > a_n + b_n$, but not necessarily if $\eta_n \leq a_n + b_n$.

(ii) Using the reasoning in Pötscher (2007), the one-sided intervals $(-\infty, \hat{\theta}_H + c_n]$, $(-\infty, \hat{\theta}_H + c_n)$, $[\hat{\theta}_H - c_n, \infty)$, and $(\hat{\theta}_H - c_n, \infty)$, with c_n a nonnegative real number, can be shown to have infimal coverage probability $\Phi(n^{1/2}(c_n - \eta_n))$.

3.3 Adaptive LASSO

Let $C_{A,n}$ denote the interval $[\hat{\theta}_A - a_n, \hat{\theta}_A + b_n]$. The infimum of the covering probability $p_{A,n}(\theta) = P_{n,\theta}\left(\theta \in C_{A,n}\right)$ of this interval is given next.

Proposition 5 *The infimal coverage probability of the interval $C_{A,n} = [\hat{\theta}_A - a_n, \hat{\theta}_A + b_n]$ is given by*

$$\inf_{\theta \in \mathbb{R}} p_{A,n}(\theta) = \Phi(n^{1/2}(a_n - \eta_n)) - \Phi\left(n^{1/2}\left((a_n - b_n)/2 - \sqrt{((a_n + b_n)/2)^2 + \eta_n^2}\right)\right)$$

if $a_n \leq b_n$, and by

$$\inf_{\theta \in \mathbb{R}} p_{A,n}(\theta) = \Phi(n^{1/2}(b_n - \eta_n)) - \Phi\left(n^{1/2}\left((b_n - a_n)/2 - \sqrt{((a_n + b_n)/2)^2 + \eta_n^2}\right)\right)$$

if $a_n > b_n$.

Proof. The distribution function $F_{A,n,\theta} = P_{n,\theta}(n^{1/2}(\hat{\theta}_A - \theta) \leq x)$ of the adaptive LASSO estimator is given by

$$\begin{aligned} \mathbf{1}(x + n^{1/2}\theta &\geq 0)\Phi\left(-((n^{1/2}\theta - x)/2) + \sqrt{((n^{1/2}\theta + x)/2)^2 + n\eta_n^2}\right) + \\ \mathbf{1}(x + n^{1/2}\theta &< 0)\Phi\left(-((n^{1/2}\theta - x)/2) - \sqrt{((n^{1/2}\theta + x)/2)^2 + n\eta_n^2}\right) \end{aligned}$$

(see Pötscher and Schneider (2007)). Hence, the coverage probability $p_{A,n}(\theta) = F_{A,n,\theta}(n^{1/2}a_n) - \lim_{x\to(-n^{1/2}b_n)_-} F_{A,n,\theta}(x)$ is

$$p_{A,n}(\theta) = \begin{cases} \Phi\left(n^{1/2}\gamma^{(-)}(\theta,-a_n)\right) - \Phi\left(n^{1/2}\gamma^{(-)}(\theta,b_n)\right) & \text{if } \theta < -a_n \\ \Phi\left(n^{1/2}\gamma^{(+)}(\theta,-a_n)\right) - \Phi\left(n^{1/2}\gamma^{(-)}(\theta,b_n)\right) & \text{if } -a_n \le \theta \le b_n \\ \Phi\left(n^{1/2}\gamma^{(+)}(\theta,-a_n)\right) - \Phi\left(n^{1/2}\gamma^{(+)}(\theta,b_n)\right) & \text{if } \theta > b_n. \end{cases} \tag{3}$$

Here

$$\begin{aligned} \gamma^{(-)}(\theta,x) &= -((\theta+x)/2) - \sqrt{((\theta-x)/2)^2+\eta_n^2} \\ \gamma^{(+)}(\theta,x) &= -((\theta+x)/2) + \sqrt{((\theta-x)/2)^2+\eta_n^2}, \end{aligned}$$

which are clearly smooth functions of (θ,x). Observe that $\gamma^{(-)}$ and $\gamma^{(+)}$ are nonincreasing in $\theta\in\mathbb{R}$ (for every $x\in\mathbb{R}$). As a consequence, we obtain for $-a_n\le\theta\le b_n$ the lower bound

$$\begin{aligned} p_{A,n}(\theta) &\ge \Phi\left(n^{1/2}\gamma^{(+)}(b_n,-a_n)\right) - \Phi\left(n^{1/2}\gamma^{(-)}(-a_n,b_n)\right) \\ &= \Phi\left(n^{1/2}\left((a_n-b_n)/2+\sqrt{((a_n+b_n)/2)^2+\eta_n^2}\right)\right) \\ &\quad -\Phi\left(n^{1/2}\left((a_n-b_n)/2-\sqrt{((a_n+b_n)/2)^2+\eta_n^2}\right)\right). \end{aligned} \tag{4}$$

Consider first the case where $a_n\le b_n$. We then show that $p_{A,n}(\theta)$ is nonincreasing on $(-\infty,-a_n)$: The derivative $dp_{A,n}(\theta)/d\theta$ is given by

$$\begin{aligned} &dp_{A,n}(\theta)/d\theta = \\ &n^{1/2}[\phi(n^{1/2}\gamma^{(-)}(\theta,-a_n))\partial\gamma^{(-)}(\theta,-a_n)/\partial\theta - \phi(n^{1/2}\gamma^{(-)}(\theta,b_n))\partial\gamma^{(-)}(\theta,b_n)/\partial\theta] \end{aligned}$$

where ϕ denotes the standard normal density function. Using the relation $a_n\le b_n$, elementary calculations show that

$$\partial\gamma^{(-)}(\theta,-a_n)/\partial\theta \le \partial\gamma^{(-)}(\theta,b_n)/\partial\theta \quad \text{for } \theta\in(-\infty,-a_n).$$

Furthermore, given $a_n\le b_n$, it is not too difficult to see that $\left|\gamma^{(-)}(\theta,-a_n)\right| \le \left|\gamma^{(-)}(\theta,b_n)\right|$ for $\theta\in(-\infty,-a_n)$ (cf. Lemma 6 below), which implies that

$$\phi(n^{1/2}\gamma^{(-)}(\theta,-a_n)) \ge \phi(n^{1/2}\gamma^{(-)}(\theta,b_n)).$$

The last two displays together with the fact that $\partial\gamma^{(-)}(\theta,-a_n)/\partial\theta$ as well as $\partial\gamma^{(-)}(\theta,b_n)/\partial\theta$ are less than or equal to zero, imply that $dp_{A,n}(\theta)/d\theta\le 0$ on $(-\infty,-a_n)$. This proves that

$$\inf_{\theta<-a_n} p_{A,n}(\theta) = \lim_{\theta\to(-a_n)_-} p_{A,n}(\theta) = c$$

with

$$c = \Phi\left(n^{1/2}(a_n-\eta_n)\right) - \Phi\left(n^{1/2}\left((a_n-b_n)/2-\sqrt{((a_n+b_n)/2)^2+\eta_n^2}\right)\right). \tag{5}$$

Since the lower bound given in (4) is not less than c, we have

$$\inf_{\theta \leq b_n} p_{A,n}(\theta) = \inf_{\theta < -a_n} p_{A,n}(\theta) = c.$$

It remains to show that $p_{A,n}(\theta) \geq c$ for $\theta > b_n$. From (3) and (5) after rearranging terms we obtain for $\theta > b_n$

$$\begin{aligned} p_{A,n}(\theta) - c &= \left[\Phi(n^{1/2}\gamma^{(+)}(\theta, -a_n)) - \Phi(n^{1/2}\gamma^{(-)}(-a_n, -a_n))\right] - \\ &\quad \left[\Phi(n^{1/2}\gamma^{(+)}(\theta, b_n)) - \Phi(n^{1/2}\gamma^{(-)}(-a_n, b_n))\right]. \end{aligned}$$

It is elementary to show that $\gamma^{(+)}(\theta, -a_n)) \geq \gamma^{(-)}(-a_n, -a_n) = a_n - \eta_n$ and $\gamma^{(+)}(\theta, b_n)) \geq \gamma^{(-)}(-a_n, b_n)$. We next show that

$$\gamma^{(+)}(\theta, -a_n) - \gamma^{(-)}(-a_n, -a_n)) \geq \gamma^{(+)}(\theta, b_n) - \gamma^{(-)}(-a_n, b_n). \tag{6}$$

To establish this note that (6) can equivalently be rewritten as

$$f(0) + f((\theta + a_n)/2) \geq f((\theta - b_n)/2) + f((a_n + b_n)/2) \tag{7}$$

where $f(x) = (x^2 + \eta_n^2)^{1/2}$. Observe that $0 \leq (\theta - b_n)/2 \leq (\theta + a_n)/2$ holds since $0 \leq a_n \leq b_n < \theta$. Writing $(\theta - b_n)/2$ as $\lambda(\theta + a_n)/2 + (1 - \lambda)0$ with $0 \leq \lambda \leq 1$ gives $(a_n + b_n)/2 = (1 - \lambda)(\theta + a_n)/2 + \lambda 0$. Because f is convex, the inequality (7) and hence (6) follows.

Next observe that in case $a_n \geq \eta_n$ we have (using monotonicity of $\gamma^{(+)}(\theta, b_n)$)

$$0 \leq \gamma^{(-)}(-a_n, -a_n)) = a_n - \eta_n \leq b_n - \eta_n = -\gamma^{(+)}(b_n, b_n) \leq -\gamma^{(+)}(\theta, b_n) \tag{8}$$

for $\theta > b_n$. In case $a_n < \eta_n$ we have (using symmetry and monotonicity of $\gamma^{(-)}$)

$$\gamma^{(-)}(-a_n, b_n) \leq \gamma^{(-)}(-a_n, -a_n) = a_n - \eta_n < 0, \tag{9}$$

and (using monotonicity of $\gamma^{(+)}$)

$$\gamma^{(-)}(-a_n, b_n) \leq -\gamma^{(+)}(b_n, -a_n) \leq -\gamma^{(+)}(\theta, -a_n) \tag{10}$$

for $\theta > b_n$. Applying Lemma 7 below with $\alpha = n^{1/2}\gamma^{(-)}(-a_n, -a_n)$, $\beta = n^{1/2}\gamma^{(+)}(\theta, -a_n)$, $\gamma = n^{1/2}\gamma^{(-)}(-a_n, b_n)$, and $\delta = n^{1/2}\gamma^{(+)}(\theta, b_n)$ and using (6)-(10), establishes $p_{A,n}(\theta) - c \geq 0$. This completes the proof in case $a_n \leq b_n$.

The case $a_n > b_n$ follows from the observation that (3) remains unchanged if a_n and b_n are interchanged and θ is replaced by $-\theta$. ∎

We note that $p_{A,n}$ is continuous except at $\theta = b_n$ and $\theta = -a_n$ and that the infimum of $p_{A,n}$ is not attained which can be seen from a simple refinement of the above proof.

Lemma 6 *Suppose* $a_n \leq b_n$. *Then* $\left|\gamma^{(-)}(\theta, -a_n)\right| \leq \left|\gamma^{(-)}(\theta, b_n)\right|$ *holds for* $\theta \in (-\infty, -a_n)$.

Proof. Squaring both sides of the claimed inequality shows that the claim is equivalent to

$$a_n^2/2 - (a_n - \theta)\sqrt{((a_n+\theta)/2)^2 + \eta^2} \leq b_n^2/2 + (b_n+\theta)\sqrt{((b_n-\theta)/2)^2+\eta^2}.$$

But, for $\theta < -a_n$, the left-hand side of the preceding display is not larger than

$$a_n^2/2 + (a_n+\theta)\sqrt{((a_n-\theta)/2)^2+\eta^2}.$$

Since $a_n^2/2 \leq b_n^2/2$, it hence suffices to show that

$$-(a_n+\theta)\sqrt{((a_n-\theta)/2)^2+\eta^2} \geq -(b_n+\theta)\sqrt{((b_n-\theta)/2)^2+\eta^2}$$

for $\theta < -a_n$. This is immediately seen by distinguishing the cases where $-b_n \leq \theta < -a_n$ and where $\theta < -b_n$, and observing that $a_n \leq b_n$. ■

The following lemma is elementary to prove.

Lemma 7 *Suppose α, β, γ, and δ are real numbers satisfying $\alpha \leq \beta$, $\gamma \leq \delta$, and $\beta - \alpha \geq \delta - \gamma$. If $0 \leq \alpha \leq -\delta$, or if $\gamma \leq \alpha \leq 0$ and $\gamma \leq -\beta$, then $\Phi(\beta) - \Phi(\alpha) \geq \Phi(\delta) - \Phi(\gamma)$.*

Remark 8 *(i) If $C_{A,n}^o$ denotes the open interval $(\hat{\theta}_A - a_n, \hat{\theta}_A + b_n)$, the formula for the coverage probability becomes*

$$P_{n,\theta}\left(\theta \in C_{A,n}^o\right) = \begin{cases} \Phi\left(n^{1/2}\gamma^{(-)}(\theta,-a_n)\right) - \Phi\left(n^{1/2}\gamma^{(-)}(\theta,b_n)\right) & \text{if } \theta \leq -a_n \\ \Phi\left(n^{1/2}\gamma^{(+)}(\theta,-a_n)\right) - \Phi\left(n^{1/2}\gamma^{(-)}(\theta,b_n)\right) & \text{if } -a_n < \theta < b_n \\ \Phi\left(n^{1/2}\gamma^{(+)}(\theta,-a_n)\right) - \Phi\left(n^{1/2}\gamma^{(+)}(\theta,b_n)\right) & \text{if } \theta \geq b_n. \end{cases}$$

Again this is continuous except at $\theta = b_n$ and $\theta = -a_n$ (except in the trivial case $a_n = b_n = 0$). It is now easy to see that the infimal coverage probability of $C_{A,n}^o$ coincides with the infimal coverage probability of the closed interval $C_{A,n}$, the infimum of the coverage probability of $C_{A,n}^o$ now always being a minimum. Furthermore, the half-open intervals $(\hat{\theta}_A - a_n, \hat{\theta}_A + b_n]$ and $[\hat{\theta}_A - a_n, \hat{\theta}_A + b_n)$ a fortiori have the same infimal coverage probability as $C_{A,n}$ and $C_{A,n}^o$.

(ii) The one-sided intervals $(-\infty, \hat{\theta}_A + c_n]$, $(-\infty, \hat{\theta}_A + c_n)$, $(\hat{\theta}_A - c_n, \infty)$, and $[\hat{\theta}_A - c_n, \infty)$, with c_n a nonnegative real number, have infimal coverage probability given by $\Phi(n^{1/2}(c_n - \eta_n))$. This follows by similar, but simpler, reasoning as in the proof of Proposition 5.

3.4 Symmetric intervals are shortest

For the two-sided confidence sets considered above, we show first that given a prescribed infimal coverage probability the symmetric intervals are shortest. We then show that these shortest intervals are longer than the standard interval based on the maximum likelihood estimator and quantify the excess length of these intervals.

Theorem 9 *Let δ satisfy $0 < \delta < 1$.*

(a) Among all intervals $C_{S,n}$ with infimal coverage probability not less than δ there is a unique shortest interval $C^*_{S,n} = [\hat{\theta}_S - a^*_{n,S}, \hat{\theta}_S + b^*_{n,S}]$ characterized by $a^*_{n,S} = b^*_{n,S}$ with $a^*_{n,S}$ being the unique solution to

$$\Phi(n^{1/2}(a_n - \eta_n)) - \Phi(n^{1/2}(-a_n - \eta_n)) = \delta. \tag{11}$$

The interval $C^*_{S,n}$ has infimal coverage probability equal to δ and $a^*_{n,S}$ is positive.

(b) Among all intervals $C_{H,n}$ with infimal coverage probability not less than δ there is a unique shortest interval $C^*_{H,n} = [\hat{\theta}_H - a^*_{n,H}, \hat{\theta}_H + b^*_{n,H}]$ characterized by $a^*_{n,H} = b^*_{n,H}$ with $a^*_{n,H}$ being the unique solution to

$$\Phi(n^{1/2}(a_n - \eta_n)) - \Phi(-n^{1/2}a_n) = \delta. \tag{12}$$

The interval $C^*_{H,n}$ has infimal coverage probability equal to δ and $a^*_{n,H}$ satisfies $a^*_{n,H} > \eta_n/2$.

(c) Among all intervals $C_{A,n}$ with infimal coverage probability not less than δ there is a unique shortest interval $C^*_{A,n} = [\hat{\theta}_A - a^*_{n,A}, \hat{\theta}_A + b^*_{n,A}]$ characterized by $a^*_{n,A} = b^*_{n,A}$ with $a^*_{n,A}$ being the unique solution of

$$\Phi(n^{1/2}(a_n - \eta_n)) - \Phi\left(-n^{1/2}\sqrt{a_n^2 + \eta_n^2}\right) = \delta. \tag{13}$$

The interval $C^*_{A,n}$ has infimal coverage probability equal to δ and $a^*_{n,A}$ is positive.

Proof. (a) Since δ is positive, any solution to (11) has to be positive. Now the equation (11) has a unique solution $a^*_{n,S}$, since (11) as a function of $a_n \in [0, \infty)$ is easily seen to be strictly increasing with range $[0, 1)$. Furthermore, the infimal coverage probability (1) is a continuous function of the pair (a_n, b_n) on $[0, \infty) \times [0, \infty)$. Let $K \subseteq [0, \infty) \times [0, \infty)$ consist of all pairs (a_n, b_n) such that (i) the corresponding interval $[\hat{\theta}_S - a_n, \hat{\theta}_S + b_n]$ has infimal coverage probability not less than δ, and (ii) the length $a_n + b_n$ is less than or equal $2a^*_{n,S}$. Then K is compact. It is also nonempty as the pair $(a^*_{n,S}, a^*_{n,S})$ belongs to K. Since the length $a_n + b_n$ is obviously continuous, it follows that there is a pair $(a^o_n, b^o_n) \in K$ having minimal length within K. Since confidence sets corresponding to pairs not belonging to K always have length larger than $2a^*_{n,S}$, the pair (a^o_n, b^o_n) gives rise to an interval with shortest length within the set of all intervals with infimal coverage probability not less than δ. We next show that $a^o_n = b^o_n$ must hold: Suppose not, then we may assume without loss of generality that $a^o_n < b^o_n$, since (1) remains invariant under permutation of a^o_n and b^o_n. But now increasing a^o_n by $\varepsilon > 0$ and decreasing b^o_n by the same amount such that $a^o_n + \varepsilon < b^o_n - \varepsilon$ holds, will result in an interval of the same length with infimal coverage probability

$$\Phi(n^{1/2}(a^o_n + \varepsilon - \eta_n)) - \Phi(n^{1/2}(-(b^o_n - \varepsilon) - \eta_n)).$$

This infimal coverage probability will be strictly larger than

$$\Phi(n^{1/2}(a^o_n - \eta_n)) - \Phi(n^{1/2}(-b^o_n - \eta_n)) \geq \delta$$

provided ε is chosen sufficiently small. But then, by continuity of the infimal coverage probability as a function of a_n and b_n, the interval $[\hat{\theta}_S - a_n^o - \varepsilon, \hat{\theta}_S + b_n' - \varepsilon]$ with $\varepsilon < b_n' < b_n^o$ will still have infimal coverage probability not less than δ as long as b_n' is sufficiently close to b_n^o; at the same time this interval will be shorter than the interval $[\hat{\theta}_S - a_n^o, \hat{\theta}_S + b_n^o]$. This leads to a contradiction and establishes $a_n^o = b_n^o$. By what was said at the beginning of the proof, it is now obvious that $a_n^o = b_n^o = a_{n,S}^*$ must hold, thus also establishing uniqueness. The last claim is obvious in view of the construction of $a_{n,S}^*$.

(b) Since δ is positive, any solution to (12) is has to be larger than $\eta_n/2$. Now equation (12) has a unique solution $a_{n,H}^*$, since (12) as a function of $a_n \in [\eta_n/2, \infty)$ is easily seen to be strictly increasing with range $[0, 1)$. Furthermore, define K similarly as in the proof of part (a). Then by the same reasoning as in (a), the set K is compact and non-empty, leading to a pair (a_n^o, b_n^o) that gives rise to an interval with shortest length within the set of all intervals with infimal coverage probability not less than δ. We next show that $a_n^o = b_n^o$ must hold: Suppose not, then we may again assume without loss of generality that $a_n^o < b_n^o$. Note that $a_n^o + b_n^o > \eta_n$ must hold, since the infimal coverage probability of the corresponding interval is positive by construction. Since all this entails $|a_n^o - \eta_n| < b_n^o$, increasing a_n^o by $\varepsilon > 0$ and decreasing b_n^o by the same amount such that $a_n^o + \varepsilon < b_n^o - \varepsilon$ holds, will result in an interval of the same length with infimal coverage probability

$$\begin{aligned}\Phi(n^{1/2}(a_n^o + \varepsilon - \eta_n)) - \Phi(-n^{1/2}(b_n^o - \varepsilon)) &> \\ \Phi(n^{1/2}(a_n^o - \eta_n)) - \Phi(-n^{1/2} b_n^o) &\geq \delta\end{aligned}$$

provided ε is chosen sufficiently small. By continuity of the infimal coverage probability as a function of a_n and b_n, the interval $[\hat{\theta}_S - a_n^o - \varepsilon, \hat{\theta}_S + b_n' - \varepsilon]$ with $\varepsilon < b_n' < b_n^o$ will still have infimal coverage probability not less than δ as long as b_n' is sufficiently close to b_n^o; at the same time this interval will be shorter than the interval $[\hat{\theta}_S - a_n^o, \hat{\theta}_S + b_n^o]$, leading to a contradiction thus establishing $a_n^o = b_n^o$. As in (a) it now follows that $a_n^o = b_n^o = a_{n,H}^*$ must hold, thus also establishing uniqueness. The last claim is then obvious in view of the construction of $a_{n,H}^*$.

(c) Since δ is positive, it is easy to see that any solution to (13) has to be positive. Now equation (13) has a unique solution $a_{n,A}^*$, since (13) as a function of $a_n \in [0, \infty)$ is strictly increasing with range $[0, 1)$. Furthermore, the infimal coverage probability as given in Proposition 5 is a continuous function of the pair (a_n, b_n) on $[0, \infty) \times [0, \infty)$. Define K similarly as in the proof of part (a). Then by the same reasoning as in (a), the set K is compact and non-empty, leading to a pair (a_n^o, b_n^o) that gives rise to an interval with shortest length within the set of all intervals with infimal coverage probability not less than δ. We next show that $a_n^o = b_n^o$ must hold: Suppose not, then we may again assume without loss of generality that $a_n^o < b_n^o$. But now increasing a_n^o by $\varepsilon > 0$ and decreasing b_n^o by the same amount such that $a_n^o + \varepsilon < b_n^o - \varepsilon$ holds, will result in an interval of the same length with infimal coverage probability

$$\Phi(n^{1/2}(a_n^o+\varepsilon-\eta_n))-\Phi\left(n^{1/2}\left(\varepsilon+(a_n^o-b_n^o)/2-\sqrt{((a_n^o+b_n^o)/2)^2+\eta_n^2}\right)\right)>$$
$$\Phi(n^{1/2}(a_n^o-\eta_n))-\Phi\left(n^{1/2}\left((a_n^o-b_n^o)/2-\sqrt{((a_n^o+b_n^o)/2)^2+\eta_n^2}\right)\right)\geq\delta,$$

provided ε is chosen sufficiently small. This is so since $a_n^o < b_n^o$ implies

$$|a_n^o-\eta_n|<\left|(a_n^o-b_n^o)/2-\sqrt{((a_n^o+b_n^o)/2)^2+\eta_n^2}\right|$$

as is easily seen. But then, by continuity of the infimal coverage probability as a function of a_n and b_n, the interval $[\hat{\theta}_S-a_n^o-\varepsilon, \hat{\theta}_S+b_n'-\varepsilon]$ with $\varepsilon < b_n' < b_n^o$ will still have infimal coverage probability not less than δ as long as b_n' is sufficiently close to b_n^o; at the same time this interval will be shorter than the interval $[\hat{\theta}_S-a_n^o, \hat{\theta}_S+b_n^o]$. This leads to a contradiction and establishes $a_n^o=b_n^o$. As in (a) it now follows that $a_n^o=b_n^o=a_{n,A}^*$ must hold, thus also establishing uniqueness. The last claim is obvious in view of the construction of $a_{n,A}^*$. ∎

In the statistically uninteresting case $\delta=0$ the interval with $a_n=b_n=0$ is the unique shortest interval in all three cases. However, for the case of the hard-thresholding estimator also any interval with $a_n=b_n$ and $a_n\leq\eta_n/2$ has infimal covering probability equal to zero.

The above proposition shows that given a prespecified δ $(0<\delta<1)$, the shortest confidence set with infimal coverage probability equal to δ based on the soft-thresholding (LASSO) estimator is shorter than the corresponding interval based on the adaptive LASSO estimator, which in turn is shorter than the corresponding interval based on the hard-thresholding estimator. All three intervals are longer than the corresponding standard confidence interval based on the maximum likelihood estimator. That is,

$$a_{n,H}^*>a_{n,A}^*>a_{n,S}^*>n^{-1/2}\Phi^{-1}((1+\delta)/2).$$

Figure 1 below shows $n^{1/2}$ times the half-length of the shortest δ-level confidence intervals based on hard-thresholding, adaptive LASSO, soft-thresholding, and the maximum likelihood estimator, respectively, as a function of $n^{1/2}\eta_n$ for various values of δ. The graphs illustrate that the intervals based on hard-thresholding, adaptive LASSO, and soft-thresholding substantially exceed the length of the maximum likelihood based interval except if $n^{1/2}\eta_n$ is very small. For large values of $n^{1/2}\eta_n$ the graphs suggest a linear increase in the length of the intervals based on the penalized estimators. This is formally confirmed in Section 3.4.1 below.

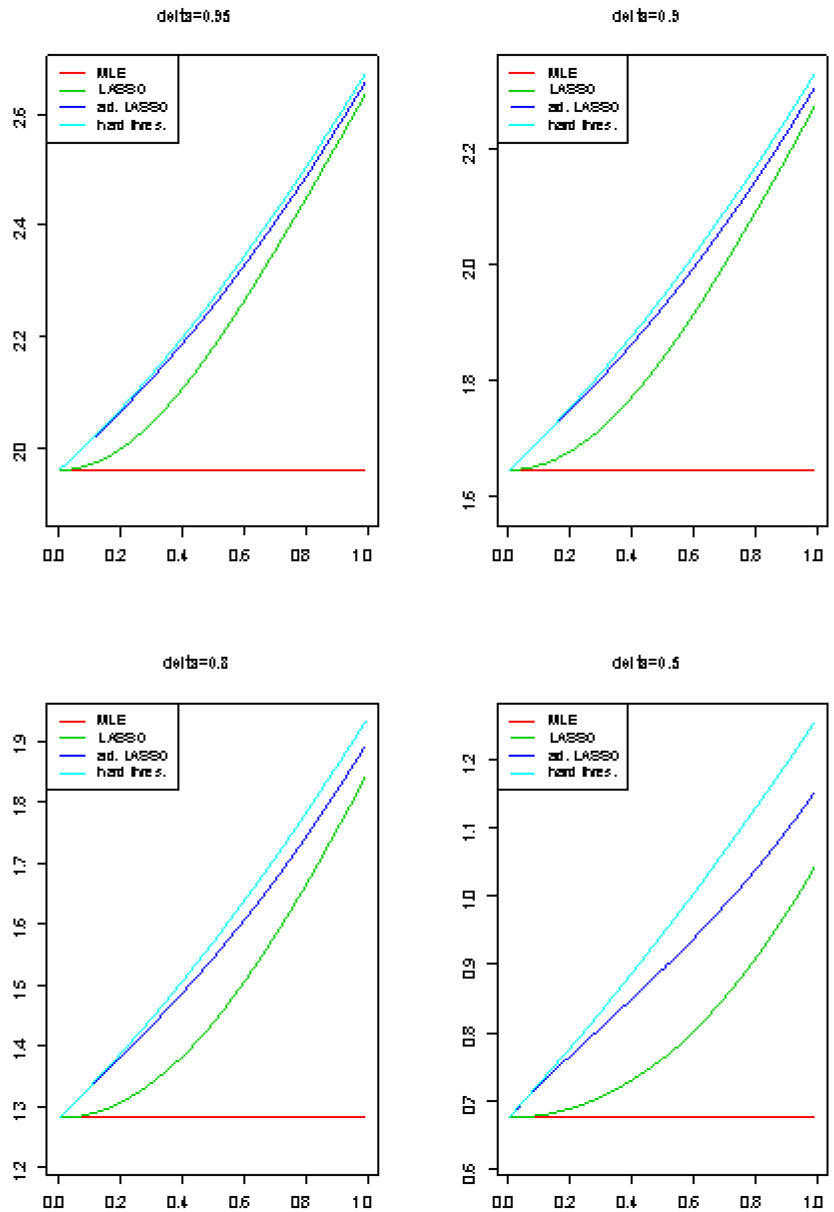

Figure 1: $n^{1/2}a^*_{n,H}$, $n^{1/2}a^*_{n,A}$, $n^{1/2}a^*_{n,S}$ as a function of $n^{1/2}\eta_n$ for coverage probabilities $\delta = 0.5$, 0.8, 0.9, 0.95. The horizontal line at height $\Phi^{-1}((1+\delta)/2)$ indicates $n^{1/2}$ times the half-length of the standard maximum likelihood based interval.

3.4.1 Asymptotic behavior of the length

It is well-known that as $n \to \infty$ two different regimes for the tuning parameter η_n can be distinguished. In the first regime $\eta_n \to 0$ and $n^{1/2}\eta_n \to e$, $0 < e < \infty$. This choice of tuning parameter leads to estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_A$ that perform conservative model selection. In the second regime $\eta_n \to 0$ and $n^{1/2}\eta_n \to \infty$, leading to estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_A$ that perform consistent model selection (also known as the 'sparsity property'); that is, with probability approaching 1, the estimators are exactly zero if the true value $\theta = 0$, and they are different from zero if $\theta \neq 0$. See Pötscher and Leeb (2007) and Pötscher and Schneider (2007) for a detailed discussion. We now discuss the asymptotic behavior, under the two regimes, of the half-length $a^*_{n,S}$, $a^*_{n,H}$, and $a^*_{n,A}$ of the shortest intervals $C^*_{S,n}$, $C^*_{H,n}$, and $C^*_{A,n}$ with a fixed infimal coverage probability δ, $0 < \delta < 1$.

If $\eta_n \to 0$ and $n^{1/2}\eta_n \to e$, $0 < e < \infty$, then it follows immediately from Proposition 9 that $n^{1/2}a^*_{n,S}$, $n^{1/2}a^*_{n,H}$, and $n^{1/2}a^*_{n,A}$ converge to the unique solutions of

$$\Phi(a - e) - \Phi(-a - e) = \delta,$$

$$\Phi(a - e) - \Phi(-a) = \delta,$$

and

$$\Phi\left(\sqrt{a^2 + e^2}\right) - \Phi(-a + e) = \delta,$$

respectively. Hence, while $a^*_{n,H}$, $a^*_{n,A}$, and $a^*_{n,S}$ are larger than the half-length $n^{-1/2}\Phi^{-1}((1+\delta)/2)$ of the standard interval, they are of the same order $n^{-1/2}$.

The situation is different, however, if $\eta_n \to 0$ but $n^{1/2}\eta_n \to \infty$. In this case Proposition 9 shows that

$$\Phi(n^{1/2}(a^*_{n,S} - \eta_n)) \to \delta$$

since $n^{1/2}(-a^*_{n,S} - \eta_n) \leq -n^{1/2}\eta_n \to -\infty$. In other words,

$$a^*_{n,S} = \eta_n + n^{-1/2}\Phi^{-1}(\delta) + o(n^{-1/2}). \tag{14}$$

Similarly, noting that $n^{1/2}a^*_{n,H} > n^{1/2}\eta_n/2 \to \infty$, we get

$$a^*_{n,H} = \eta_n + n^{-1/2}\Phi^{-1}(\delta) + o(n^{-1/2}); \tag{15}$$

and since $n^{1/2}\sqrt{a_n^2 + \eta_n^2} \geq n^{1/2}\eta_n \to \infty$ we obtain

$$a^*_{n,A} = \eta_n + n^{-1/2}\Phi^{-1}(\delta) + o(n^{-1/2}). \tag{16}$$

Hence, the intervals $C^*_{S,n}$, $C^*_{H,n}$, and $C^*_{A,n}$ are asymptotically of the same length. They are also longer than the standard interval by an order of magnitude: the ratio of each of $a^*_{n,S}$ ($a^*_{n,H}$, $a^*_{n,A}$, respectively) to the half-length of the standard interval is $n^{1/2}\eta_n$, which diverges to infinity. Hence, when the estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_A$ are tuned to possess the 'sparsity property', the corresponding confidence sets become very large. For the particular intervals considered here this is a refinement of a general result in Pötscher (2007) for confidence sets based on arbitrary estimators possessing the 'sparsity property'.

4 A simple asymptotic confidence interval

The finite-sample confidence intervals obtained in Section 3 required a detail case by case analysis based on the finite-sample distribution of the estimator on which the interval is based. If the estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_A$ are tuned to possess the 'sparsity property', i.e., if the tuning parameter satisfies $\eta_n \to 0$ and $n^{1/2}\eta_n \to \infty$, a simple asymptotic confidence interval construction relying on asymptotic results obtained in Pötscher and Leeb (2007) and Pötscher and Schneider (2007) is possible as shown below. An advantage of this construction is that it easily extends to other estimators like the smoothly clipped absolute deviation (SCAD) estimator when tuned to possess the 'sparsity property'.

As shown in Pötscher and Leeb (2007) and Pötscher and Schneider (2007), the uniform rate of consistency of the 'sparsely' tuned estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_A$ is not $n^{1/2}$, but only η_n^{-1}; furthermore, the limiting distributions of these estimators under the appropriate η_n^{-1}-scaling and a moving-parameter asymptotic framework is always concentrated in the interval $[-1, 1]$. These facts can be used to obtain the following result.

Proposition 10 *Suppose $\eta_n \to 0$ and $n^{1/2}\eta_n \to \infty$. Let $\hat{\theta}$ stand for any of the estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_A$. Let d be a real number, and define the interval $D_n = [\hat{\theta} - d\eta_n, \hat{\theta} + d\eta_n]$. If $d > 1$, the interval D_n has infimal coverage probability converging to 1, i.e.,*

$$\lim_{n\to\infty} \inf_{\theta\in\mathbb{R}} P_{n,\theta}(\theta \in D_n) = 1.$$

If $d < 1$,

$$\lim_{n\to\infty} \inf_{\theta\in\mathbb{R}} P_{n,\theta}(\theta \in D_n) = 0.$$

Proof. Let

$$c = \liminf_{n\to\infty} \inf_{\theta\in\mathbb{R}} P_{n,\theta}\left(-d \le \eta_n^{-1}(\hat{\theta} - \theta) \le d\right).$$

By definition of c, we can find a subsequence n_k and elements $\theta_{n_k} \in \mathbb{R}$ such that

$$P_{n_k,\theta_{n_k}}\left(-d \le \eta_{n_k}^{-1}(\hat{\theta} - \theta_{n_k}) \le d\right) \to c$$

for $k \to \infty$. Now, by Theorem 9 (for $\hat{\theta} = \hat{\theta}_H$), Theorem 10 (for $\hat{\theta} = \hat{\theta}_S$), and Remark 12 in Pötscher and Leeb (2007), and by Theorem 6 (for $\hat{\theta} = \hat{\theta}_A$) and Remark 7 in Pötscher and Schneider (2007), any accumulation point of the distribution of $\eta_{n_k}^{-1}(\hat{\theta} - \theta_{n_k})$ w.r.t. weak convergence is a probability measure concentrated on $[-1, 1]$. Since $d > 1$, it follows that $c = 1$ must hold, which proves the first claim. We next prove the second claim. In view of Theorem 9 (for $\hat{\theta} = \hat{\theta}_H$) and Theorem 10 (for $\hat{\theta} = \hat{\theta}_S$) in Pötscher and Leeb (2007), and in view of Theorem 6 (for $\hat{\theta} = \hat{\theta}_A$) in Pötscher and Schneider (2007) it is possible to choose a sequence $\theta_n \in \mathbb{R}$ such that the distribution of $\eta_n^{-1}(\hat{\theta} - \theta_n)$ converges to point mass located at one of the endpoints of the interval $[-1, 1]$. But then clearly

$$P_{n,\theta_n}\left(-d \le \eta_n^{-1}(\hat{\theta} - \theta_n) \le d\right) \to 0$$

for $d < 1$ which implies the second claim. ■

The asymptotic distributional results in the above proposition do not provide information on the case $d = 1$. However, from the finite-sample results in Section 3 we see that in this case the infimal coverage probability of D_n converges to 1/2.

Since the interval D_n for $d > 1$ has asymptotic infimal coverage probability equal to one, one may wonder how much cruder this interval is compared to the finite-sample intervals $C^*_{S,n}$, $C^*_{H,n}$, and $C^*_{A,n}$ constructed in Section 3, which have infimal coverage probability equal to a prespecified level δ, $0 < \delta < 1$: The ratio of the half-length of D_n to the half-length of the corresponding interval $C^*_{S,n}$, $C^*_{H,n}$, and $C^*_{A,n}$ is

$$d(1 + o(n^{-1/2}\eta_n^{-1})) = d(1 + o(1))$$

as can be seen from equations (14), (15), and (16). Since d can be chosen arbitrarily close to one, this ratio can be made arbitrarily small. This may sound somewhat strange, since we are comparing an interval with asymptotic infimal coverage probability 1 with the shortest finite-sample confidence intervals that have a fixed infimal coverage probability δ less than 1. The reason for this phenomenon is that, in the relevant moving-parameter asymptotic framework, the distribution of $\hat{\theta} - \theta$ is made up of a bias-component which is of the order η_n and a random component which is of the order $n^{-1/2}$. Since $\eta_n \to 0$ and $n^{1/2}\eta_n \to \infty$, the deterministic bias-component dominates the random component. This can also be gleaned from equations (14), (15), and (16), where the level δ enters the formula for the half-length only in the lower order term.[2]

We note that using Theorem 11 in Pötscher and Leeb (2007) the same proof immediately shows that Proposition 10 also holds for the smoothly clipped absolute deviation (SCAD) estimator when tuned to possess the 'sparsity property'. In fact, the argument in the proof of the above proposition can be applied to a large class of post-model-selection estimators based on a consistent model selection procedure.

Remark 11 *(i) Suppose $D'_n = [\hat{\theta} - d_1\eta_n, \hat{\theta} + d_2\eta_n]$ where $\hat{\theta}$ stand for any of the estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_A$. If $\min(d_1, d_2) > 1$, then the limit of the infimal coverage probability of D'_n is 1; if $\max(d_1, d_2) < 1$ then this limit is zero. This follows immediately from an inspection of the proof of Proposition 10.*

(ii) Proposition 10 also remains correct if D_n is replaced by the corresponding open interval. A similar comment applies to the open version of D'_n.

[2] The sparsely tuned hard-thresholding estimator or the sparsely tuned adaptive LASSO (under an additional condition on η_n) are known to possess the so-called 'oracle property'. In light of the 'oracle property' it is sometimes argued that valid confidence intervals based on these estimators with length proportional to $n^{-1/2}$ can be obtained. But note that intervals with length proportional to $n^{-1/2}$ have infimal coverage probability that converges to zero; this follows immediately from the discussion in Section 3.4.1. This once more shows that *fixed-parameter* asymptotic results like the 'oracle' property are dangerously misleading.

References

[1] Fan, J. & R. Li (2001): Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348-1360.

[2] Frank, I. E. & J. H. Friedman (1993): A statistical view of some chemometrics regression tools (with discussion). *Technometrics* 35, 109-148.

[3] Joshi, V. M. (1969): Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Annals of Mathematical Statistics* 40, 1042-1067.

[4] Knight, K. & W. Fu (2000): Asymptotics of lasso-type estimators. *Annals of Statistics* 28, 1356-1378.

[5] Leeb, H. & B. M. Pötscher (2008): Sparse estimators and the oracle property, or the return of Hodges' estimator. *Journal of Econometrics* 142, 201-211.

[6] Pötscher, B. M. (2007): Confidence sets based on sparse estimators are necessarily large. Working Paper, Department of Statistics, University of Vienna. ArXiv: 0711.1036.

[7] Pötscher, B. M. & H. Leeb (2007): On the distribution of penalized maximum likelihood estimators: the LASSO, SCAD, and thresholding. Working Paper, Department of Statistics, University of Vienna. ArXiv: 0711.0660.

[8] Pötscher, B. M. & U. Schneider (2007): On the distribution of the adaptive LASSO estimator. Working Paper, Department of Statistics, University of Vienna. ArXiv: 0801.4627.

[9] Tibshirani, R. (1996): Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* 58, 267-288.

[10] Zou, H. (2006): The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101, 1418-1429.