Electronic Journal of Statistics
Vol. 5 (2011) 1876–1934
ISSN: 1935-7524
DOI: 10.1214/11-EJS659

Distributional results for thresholding estimators in high-dimensional Gaussian regression models*

Benedikt M. Pötscher

Department of Statistics, University of Vienna
Universitätsstraße 5, 1010 Vienna, Austria
e-mail: benedikt.poetscher@univie.ac.at

Ulrike Schneider

Institute for Mathematical Methods in Economics, Vienna University of Technology
Argentinierstraße 8, 1040 Vienna, Austria
e-mail: ulrike.schneider@tuwien.ac.at

Abstract: We study the distribution of hard-, soft-, and adaptive soft-thresholding estimators within a linear regression model where the number of parameters k can depend on sample size n and may diverge with n. In addition to the case of known error-variance, we define and study versions of the estimators when the error-variance is unknown. We derive the finite-sample distribution of each estimator and study its behavior in the large-sample limit, also investigating the effects of having to estimate the variance when the degrees of freedom $n-k$ does not tend to infinity or tends to infinity very slowly. Our analysis encompasses both the case where the estimators are tuned to perform consistent variable selection and the case where the estimators are tuned to perform conservative variable selection. Furthermore, we discuss consistency, uniform consistency and derive the uniform convergence rate under either type of tuning.

AMS 2000 subject classifications: Primary 62F11, 62F12; secondary 62J05, 62J07, 62E15, 62E20.
Keywords and phrases: Thresholding, Lasso, adaptive Lasso, penalized maximum likelihood, finite-sample distribution, asymptotic distribution, variance estimation, uniform convergence rate, high-dimensional model, oracle property, variable selection.

Received June 2011.

1. Introduction

We study the distribution of thresholding estimators such as hard-thresholding, soft-thresholding, and adaptive soft-thresholding in a linear regression model when the number of regressors can be large. These estimators can be viewed as penalized least-squares estimators in the case of an orthogonal design matrix, with soft-thresholding then coinciding with the Lasso (introduced by Frank and

*We would like to thank Hannes Leeb, a referee, and an associate editor for comments on a previous version of the paper.

Friedman, 1993, Alliney and Ruzinsky, 1994, and Tibshirani, 1996) and with adaptive soft-thresholding coinciding with the adaptive Lasso (introduced by Zou, 2006). Thresholding estimators have of course been discussed earlier in the context of model selection (see Bauer, Pötscher and Hackl, 1988) and in the context of wavelets (see, e.g., Donoho et al., 1995). Contributions concerning distributional properties of thresholding and penalized least-squares estimators are as follows: Knight and Fu (2000) study the asymptotic distribution of the Lasso estimator when it is tuned to act as a conservative variable selection procedure, whereas Zou (2006) studies the asymptotic distribution of the Lasso and the adaptive Lasso estimators when they are tuned to act as consistent variable selection procedures. Fan and Li (2001) and Fan and Peng (2004) study the asymptotic distribution of the so-called smoothly clipped absolute deviation (SCAD) estimator when it is tuned to act as a consistent variable selection procedure. In the wake of Fan and Li (2001) and Fan and Peng (2004) a large number of papers have been published that derive the asymptotic distribution of various penalized maximum likelihood estimators under consistent tuning; see the introduction in Pötscher and Schneider (2009) for a partial list. Except for Knight and Fu (2000), all these papers derive the asymptotic distribution in a fixed-parameter framework. As pointed out in Leeb and Pötscher (2005), such a fixed-parameter framework is often highly misleading in the context of variable selection procedures and penalized maximum likelihood estimators. For that reason, Pötscher and Leeb (2009) and Pötscher and Schneider (2009) have conducted a detailed study of the finite-sample as well as large-sample distribution of various penalized least-squares estimators, adopting a moving-parameter framework for the asymptotic results. [Related results for so-called post-model-selection estimators can be found in Leeb and Pötscher (2003, 2005) and for model averaging estimators in Pötscher (2006); see also Sen (1979) and Pötscher (1991).] The papers by Pötscher and Leeb (2009) and Pötscher and Schneider (2009) are set in the framework of an orthogonal linear regression model with a fixed number of parameters and with the error-variance being known.

In the present paper we build on the just mentioned papers Pötscher and Leeb (2009) and Pötscher and Schneider (2009). In contrast to these papers, we do not assume the number of regressors k to be fixed, but let it depend on sample size – thus allowing for high-dimensional models. We also consider the case where the error-variance is unknown, which in case of a high-dimensional model creates non-trivial complications as then estimators for the error-variance will typically not be consistent. Considering thresholding estimators from the outset in the present paper allows us also to cover non-orthogonal design. While the asymptotic distributional results in the known-variance case do not differ in substance from the results in Pötscher and Leeb (2009) and Pötscher and Schneider (2009), not unexpectedly we observe different asymptotic behavior in the unknown-variance case if the number of degrees of freedom $n-k$ is constant, the difference resulting from the non-vanishing variability of the error-variance estimator in the limit. Less expected is the result that – under consistent tuning – for the variable selection probabilities (implied by all the estimators considered) as well as for the distribution of the hard-thresholding estimator, estimation of the error-variance still has an effect asymptotically even if $n-k$ diverges, but does so only slowly.

To give some idea of the theoretical results obtained in the paper we next present a rough summary of some of these results. For simplicity of exposition assume for the moment that the $n\times k$ design matrix X is such that the diagonal elements of $(X'X/n)^{-1}$ are equal to 1, and that the error-variance σ^2 is equal to 1. Let $\tilde{\theta}_{H,i}$ denote the hard-thresholding estimator for the i-th component θ_i of the regression parameter, the threshold being given by $\hat{\sigma}\eta_{i,n}$, with $\hat{\sigma}^2$ denoting the usual error-variance estimator and with $\eta_{i,n}$ denoting a tuning parameter. An infeasible version of the estimator, denoted by $\hat{\theta}_{H,i}$, which uses σ instead of $\hat{\sigma}$, is also considered (known-variance case). We then show that the uniform rate of convergence of the hard-thresholding estimator is $n^{-1/2}$ if the threshold satisfies $\eta_{i,n}\to 0$ and $n^{1/2}\eta_{i,n}\to e_i<\infty$ ("conservative tuning"), but that the uniform rate is only $\eta_{i,n}$ if the threshold satisfies $\eta_{i,n}\to 0$ and $n^{1/2}\eta_{i,n}\to\infty$ ("consistent tuning"). The same result also holds for the soft-thresholding estimator $\tilde{\theta}_{S,i}$ and the adaptive soft-thresholding estimator $\tilde{\theta}_{AS,i}$, as well as for infeasible variants of the estimators that use knowledge of σ (known-variance case). Furthermore, all possible limits of the centered and scaled distribution of the hard-thresholding estimator $\tilde{\theta}_{H,i}$ (as well as of the soft- and the adaptive soft-thresholding estimators $\tilde{\theta}_{S,i}$ and $\tilde{\theta}_{AS,i}$) under a moving parameter framework are obtained. Consider first the case of conservative tuning: then all possible limiting forms of the distribution of $n^{1/2}(\tilde{\theta}_{H,i}-\theta_{i,n})$ as well as of $n^{1/2}(\hat{\theta}_{H,i}-\theta_{i,n})$ for arbitrary parameter sequences $\theta_{i,n}$ are determined. It turns out that – in the known-variance case – these limits are of the same functional form as the finite-sample distribution, i.e., they are a convex combination of a pointmass and an absolutely continuous distribution that is an excised version of a normal distribution. In the unknown-variance case, when the number of degrees of freedom $n-k$ goes to infinity, exactly the same limits arise. However, if $n-k$ is constant, the limits are "averaged" versions of the limits in the known-variance case, the averaging being with respect to the distribution of the variance estimator $\hat{\sigma}^2$. Again these limits have the same functional form as the corresponding finite-sample distributions. Consider next the case of consistent tuning: Here the possible limits of $\eta_{i,n}^{-1}(\tilde{\theta}_{H,i}-\theta_{i,n})$ as well as of $\eta_{i,n}^{-1}(\hat{\theta}_{H,i}-\theta_{i,n})$ have to be considered, as $\eta_{i,n}$ is the uniform convergence rate. In the known-variance case the limits are convex combinations of (at most) two pointmasses, the location of the pointmasses as well as the weights depending on $\theta_{i,n}$ and $\eta_{i,n}$. In the unknown-variance case exactly the same limits arise if $n-k$ diverges to infinity sufficiently fast; however, if $n-k$ is constant or diverges to infinity sufficiently slowly, the limits are again convex combinations of the same pointmasses, but with weights that are typically different. The picture for soft-thresholding and adaptive soft-thresholding is somewhat different: in the known-variance case, as well as in the unknown-variance case when $n-k$ diverges to infinity, the limits are (single) pointmasses. However, in the unknown-variance case and if $n-k$ is constant, the limit distribution can have an absolutely continuous component. It is furthermore useful to point out that in case of consistent tuning the sequence of distributions of $n^{1/2}(\tilde{\theta}_{H,i} - \theta_{i,n})$ is not stochastically bounded in general (since $\eta_{i,n}$ is the uniform convergence rate), and the same is true for soft-thresholding $\tilde{\theta}_{S,i}$ and adaptive soft-thresholding $\tilde{\theta}_{AS,i}$. This throws a light on the fragility of the oracle-property, see Section 6.4 for more discussion.

While our theoretical results for the thresholding estimators immediately apply to Lasso and adaptive Lasso in case of orthogonal design, this is not so in the non-orthogonal case. In order to get some insight into the finite-sample distribution of the latter estimators also in the non-orthogonal case, we numerically compare the distribution of Lasso and adaptive Lasso with their thresholding counterparts in a simulation study.

The main take-away messages of the paper can be summarized as follows:

- The finite-sample distributions of the various thresholding estimators considered are highly non-normal, the distributions being in each case a convex combination of pointmass and an absolutely continuous (non-normal) component.
- The non-normality persists asymptotically in a moving parameter framework.
- Results in the unknown-variance case are obtained from the corresponding results in the known-variance case by smoothing with respect to the distribution of $\hat{\sigma}$. In line with this, one would expect the limiting behavior in the unknown-variance case to coincide with the limiting behavior in the known-variance whenever the degrees of freedom $n-k$ diverge to infinity. This indeed turns out to be so for some of the results, but not for others where we see that the speed of divergence of $n-k$ matters.
- In case of conservative tuning the estimators have the expected uniform convergence rate, which is $n^{-1/2}$ under the simplified assumptions of the above discussion, whereas under consistent tuning the uniform rate is slower, namely $\eta_{i,n}$ under the simplified assumptions of the above discussion. This is intimately connected with the fact that the so-called 'oracle property' paints a misleading picture of the performance of the estimators.
- The numerical study suggests that the results for the thresholding estimators $\tilde{\theta}_{S,i}$ and $\tilde{\theta}_{AS,i}$ qualitatively apply also to the (components of) the Lasso and the adaptive Lasso as long as the design matrix is not too ill-conditioned.

The paper is organized as follows. We introduce the model and define the estimators in Section 2. Section 3 treats the variable selection probabilities implied by the estimators. Consistency, uniform consistency, and uniform convergence rates are discussed in Section 4. We derive the finite-sample distribution of each estimator in Section 5 and study the large-sample behavior of these in Section 6. A numerical study of the finite-sample distribution of Lasso and adaptive Lasso can be found in Section 7. All proofs are relegated to Section 8.

2. The model and the estimators

Consider the linear regression model

$$Y = X\theta + u$$

with Y an $n \times 1$ vector, X a nonstochastic $n \times k$ matrix of rank $k \geq 1$, and $u \sim N(0, \sigma^2 I_n)$, $0 < \sigma < \infty$. We allow k, the number of columns of X, as well as the entries of Y, X, and u to depend on sample size n (in fact, also the probability spaces supporting Y and u may depend on n), although we shall almost always suppress this dependence on n in the notation. Note that this framework allows for high-dimensional regression models, where the number of regressors k is large compared to sample size n, as well as for the more classical situation where k is much smaller than n. Furthermore, let $\xi_{i,n}$ denote the nonnegative square root of $((X'X/n)^{-1})_{ii}$, the i-th diagonal element of $(X'X/n)^{-1}$. Now let

$$\hat{\theta}_{LS} = (X'X)^{-1} X'Y$$
$$\hat{\sigma}^2 = (n-k)^{-1}(Y - X\hat{\theta}_{LS})'(Y - X\hat{\theta}_{LS})$$

denote the least-squares estimator for θ and the associated estimator for σ^2, the latter being defined only if $n > k$. The hard-thresholding estimator $\tilde{\theta}_H$ is defined via its components as follows

$$\tilde{\theta}_{H,i} = \tilde{\theta}_{H,i}(\eta_{i,n}) = \hat{\theta}_{LS,i}\mathbf{1}\left(\left|\hat{\theta}_{LS,i}\right| > \hat{\sigma}\xi_{i,n}\eta_{i,n}\right),$$

where the tuning parameters $\eta_{i,n}$ are positive real numbers and $\hat{\theta}_{LS,i}$ denotes the i-th component of the least-squares estimator. We shall also need to consider its infeasible counterpart $\hat{\theta}_H$ given by

$$\hat{\theta}_{H,i} = \hat{\theta}_{H,i}(\eta_{i,n}) = \hat{\theta}_{LS,i}\mathbf{1}\left(\left|\hat{\theta}_{LS,i}\right| > \sigma\xi_{i,n}\eta_{i,n}\right).$$

The soft-thresholding estimator $\tilde{\theta}_S$ and its infeasible counterpart $\hat{\theta}_S$ are given by

$$\tilde{\theta}_{S,i} = \tilde{\theta}_{S,i}(\eta_{i,n}) = \operatorname{sign}(\hat{\theta}_{LS,i})\left(\left|\hat{\theta}_{LS,i}\right| - \hat{\sigma}\xi_{i,n}\eta_{i,n}\right)_+$$

and

$$\hat{\theta}_{S,i} = \hat{\theta}_{S,i}(\eta_{i,n}) = \operatorname{sign}(\hat{\theta}_{LS,i})\left(\left|\hat{\theta}_{LS,i}\right| - \sigma\xi_{i,n}\eta_{i,n}\right)_+,$$

where $(\cdot)_+ = \max(\cdot, 0)$. Finally, the adaptive soft-thresholding estimator $\tilde{\theta}_{AS}$ and its infeasible counterpart $\hat{\theta}_{AS}$ are defined via

$$\begin{aligned}\tilde{\theta}_{AS,i} &= \tilde{\theta}_{AS,i}(\eta_{i,n}) = \hat{\theta}_{LS,i}\left(1 - \hat{\sigma}^2\xi_{i,n}^2\eta_{i,n}^2/\hat{\theta}_{LS,i}^2\right)_+ \\ &= \begin{cases} 0 & \text{if } \left|\hat{\theta}_{LS,i}\right| \leq \hat{\sigma}\xi_{i,n}\eta_{i,n} \\ \hat{\theta}_{LS,i} - \hat{\sigma}^2\xi_{i,n}^2\eta_{i,n}^2/\hat{\theta}_{LS,i} & \text{if } \left|\hat{\theta}_{LS,i}\right| > \hat{\sigma}\xi_{i,n}\eta_{i,n} \end{cases}\end{aligned}$$

and

$$\begin{aligned}\hat{\theta}_{AS,i} &= \hat{\theta}_{AS,i}(\eta_{i,n}) = \hat{\theta}_{LS,i}\left(1-\sigma^2\xi_{i,n}^2\eta_{i,n}^2/\hat{\theta}_{LS,i}^2\right)_+ \\ &= \left\{\begin{array}{cl} 0 & \text{if } \left|\hat{\theta}_{LS,i}\right| \leq \sigma\xi_{i,n}\eta_{i,n} \\ \hat{\theta}_{LS,i} - \sigma^2\xi_{i,n}^2\eta_{i,n}^2/\hat{\theta}_{LS,i} & \text{if } \left|\hat{\theta}_{LS,i}\right| > \sigma\xi_{i,n}\eta_{i,n} \end{array}\right. .\end{aligned}$$

Note that $\tilde{\theta}_H$, $\tilde{\theta}_S$, and $\tilde{\theta}_{AS}$ as well as their infeasible counterparts are equivariant under scaling of the columns of $(Y : X)$ by non-zero column-specific scale factors. We have chosen to let the thresholds $\hat{\sigma}\xi_{i,n}\eta_{i,n}$ ($\sigma\xi_{i,n}\eta_{i,n}$, respectively) depend explicitly on $\hat{\sigma}$ (σ, respectively) and $\xi_{i,n}$ in order to give $\eta_{i,n}$ an interpretation independent of the values of σ and X. Furthermore, often $\eta_{i,n}$ will be chosen independently of i, i.e., $\eta_{i,n} = \eta_n$ where η_n is a positive real number. Clearly, for the feasible versions we always need to assume $n > k$, whereas for the infeasible versions $n \geq k$ suffices.

We note the simple fact that

$$0 \leq \tilde{\theta}_{S,i} \leq \tilde{\theta}_{AS,i} \leq \tilde{\theta}_{H,i} \leq \hat{\theta}_{LS,i} \tag{1}$$

holds on the event that $\hat{\theta}_{LS,i} \geq 0$, and that

$$\hat{\theta}_{LS,i} \leq \tilde{\theta}_{H,i} \leq \tilde{\theta}_{AS,i} \leq \tilde{\theta}_{S,i} \leq 0 \tag{2}$$

holds on the event that $\hat{\theta}_{LS,i} \leq 0$. Analogous inequalities hold for the infeasible versions of the estimators.

Remark 1 (*Lasso*)**.** (i) Consider the objective function

$$(Y - X\theta)'(Y - X\theta) + 2n\hat{\sigma}\sum_{i=1}^{k}\eta'_{i,n}\left|\theta_i\right|,$$

where $\eta'_{i,n}$ are positive real numbers. It is well-known that a unique minimizer $\tilde{\theta}_L$ of this objective function exists, the Lasso-estimator. It is easy to see that in case $X'X$ is diagonal we have

$$\tilde{\theta}_{L,i} = \operatorname{sign}(\hat{\theta}_{LS,i})\left(\left|\hat{\theta}_{LS,i}\right| - \hat{\sigma}\eta'_{i,n}\xi_{i,n}^2\right)_+ .$$

Hence, in the case of diagonal $X'X$, the components $\tilde{\theta}_{L,i}$ of the Lasso reduce to soft-thresholding estimators with appropriate thresholds; in particular, $\tilde{\theta}_{L,i}$ coincides with $\tilde{\theta}_{S,i}$ for the choice $\eta'_{i,n} = \eta_{i,n}\xi_{i,n}^{-1}$. Therefore all results derived below for soft-thresholding immediately give corresponding results for the Lasso as well as for the Dantzig-selector in the diagonal case. We shall abstain from spelling out further details.

(ii) Sometimes $\eta'_{i,n}$ in the definition of the Lasso is chosen independently of i; more reasonable choices seem to be (a) $\eta'_{i,n} = \eta_{i,n}\psi_{i,n}$ (where $\psi_{i,n}$ denotes the nonnegative square root of the i-th diagonal element of $(X'X/n)$), and (b) $\eta'_{i,n} = \eta_{i,n}\xi_{i,n}^{-1}$ where $\eta_{i,n}$ are positive real numbers (not depending on the design matrix and often not on i) as then $\eta_{i,n}$ again has an interpretation independent of the values of σ and X. Note that in case (a) or (b) the solution of the optimization problem is equivariant under scaling of the columns of $(Y : X)$ by non-zero column-specific scale factors.

(iii) Similar results obviously hold for the infeasible versions of the estimators.

Remark 2 (*Adaptive Lasso*)**.** Consider the objective function

$$(Y - X\theta)'(Y - X\theta) + 2n\hat{\sigma}^2 \sum_{i=1}^{k} (\eta'_{i,n})^2 \, |\theta_i| \, / \left|\hat{\theta}_{LS,i}\right|,$$

where $\eta'_{i,n}$ are positive real numbers. This is the objective function of the adaptive Lasso (where often $\eta'_{i,n} = \eta'_n$ is chosen independent of i). Again the minimizer $\tilde{\theta}_{AL}$ exists and is unique (at least on the event where $\hat{\theta}_{LS,i} \neq 0$ for all i). Clearly, $\tilde{\theta}_{AL}$ is equivariant under scaling of the columns of $(Y : X)$ by non-zero column-specific scale factors provided $\eta'_{i,n}$ does not depend on the design matrix. It is easy to see that in case $X'X$ is diagonal we have

$$\tilde{\theta}_{AL,i} = \hat{\theta}_{LS,i} \left(1 - \hat{\sigma}^2 \xi_{i,n}^2 \left(\eta'_{i,n}\right)^2 / \hat{\theta}_{LS,i}^2\right)_+ .$$

Hence, in the case of diagonal $X'X$, the components $\tilde{\theta}_{AL,i}$ of the adaptive Lasso reduce to the adaptive soft-thresholding estimators $\tilde{\theta}_{AS,i}$ (for $\eta'_{i,n} = \eta_{i,n}$). Therefore all results derived below for adaptive soft-thresholding immediately give corresponding results for the adaptive Lasso in the diagonal case. We shall again abstain from spelling out further details. Similar results obviously hold for the infeasible versions of the estimators.

Remark 3 (*Other estimators*)**.** (i) The adaptive Lasso as defined in Zou (2006) has an additional tuning parameter γ. We consider adaptive soft-thresholding only for the case $\gamma = 1$, since otherwise the estimator is not equivariant in the sense described above. Nonetheless an analysis for the case $\gamma \neq 1$, similar to the analysis in this paper, is possible in principle.

(ii) An analysis of a SCAD-based thresholding estimator is given in Pötscher and Leeb (2009) in the known-variance case. [These results are given in the orthogonal design case, but easily generalize to the non-orthogonal case.] The results obtained there for SCAD-based thresholding are similar in spirit to the results for the other thresholding estimators considered here. The unknown-variance case could also be analyzed in principle, but we refrain from doing so for the sake of brevity.

(iii) Zhang (2010) introduced the so-called minimax concave penalty (MCP) to be used for penalized least-squares estimation. Apart from the usual tuning parameter, MCP also depends on a shape parameter γ. It turns out that the thresholding estimator based on MCP coincides with hard-thresholding in case $\gamma \leq 1$, and thus is covered by the analysis of the present paper. In case $\gamma > 1$, the MCP-based thresholding estimator could similarly be analyzed, especially since the functional form of the MCP-based thresholding estimator is relatively simple (namely, a piecewise linear function of the least-squares estimator). We do not provide such an analysis for brevity.

For all asymptotic considerations in this paper we shall always assume without further mentioning that $\xi_{i,n}^2/n = ((X'X)^{-1})_{ii}$ *satisfies*

$$\sup_n \xi_{i,n}^2/n < \infty \tag{3}$$

for every fixed $i \geq 1$ *satisfying* $i \leq k(n)$ *for large enough* n. The case excluded by assumption (3) seems to be rather uninteresting as unboundedness of $\xi_{i,n}^2/n$ means that the information contained in the regressors gets weaker with increasing sample size (at least along a subsequence); in particular, this implies (coordinate-wise) inconsistency of the least-squares estimator. [In fact, if k as well as the elements of X do not depend on n, this case is actually impossible as $\xi_{i,n}^2/n$ is then necessarily monotonically nonincreasing.]

The following notation will be used in the paper: Let $\bar{\mathbb{R}}$ denote the extended real line $\mathbb{R} \cup \{-\infty, \infty\}$ endowed with the usual topology. On $\mathbb{N} \cup \{\infty\}$ we shall consider the topology it inherits from $\bar{\mathbb{R}}$. Furthermore, Φ and ϕ denote the cumulative distribution function (cdf) and the probability density function (pdf) of a standard normal distribution, respectively. By $T_{m,c}$ we denote the cdf of a non-central T-distribution with $m \in \mathbb{N}$ degrees of freedom and non-centrality parameter $c \in \mathbb{R}$. In the central case, i.e., $c = 0$, we simply write T_m. We use the convention $\Phi(\infty) = 1$, $\Phi(-\infty) = 0$ with a similar convention for $T_{m,c}$.

3. Variable selection probabilities

The estimators $\tilde{\theta}_H$, $\tilde{\theta}_S$, and $\tilde{\theta}_{AS}$ can be viewed as performing variable selection in the sense that these estimators set components of θ exactly equal to zero with positive probability. In this section we study the variable selection probability $P_{n,\theta,\sigma}(\tilde{\theta}_i \neq 0)$, where $\tilde{\theta}_i$ stands for any of the estimators $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, and $\tilde{\theta}_{AS,i}$. Since these probabilities are the same for any of the three estimators considered we shall drop the subscripts H, S, and AS in this section. We use the same convention also for the variable selection probabilities of the infeasible versions.

3.1. Known-variance case

Since $P_{n,\theta,\sigma}(\hat{\theta}_i \neq 0) = 1 - P_{n,\theta,\sigma}(\hat{\theta}_i = 0)$ it suffices to study the variable deletion probability

$$P_{n,\theta,\sigma}\left(\hat{\theta}_i = 0\right) = \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}\right)\right) - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}\right)\right). \tag{4}$$

As can be seen from the above formula, $P_{n,\theta,\sigma}(\hat{\theta}_i = 0)$ depends on θ only via θ_i. We first study the variable selection/deletion probabilities under a "fixed-parameter" asymptotic framework.

Proposition 4. *Let $0 < \sigma < \infty$ be given. For every $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have:*

(a) A necessary and sufficient condition for $P_{n,\theta,\sigma}(\hat{\theta}_i = 0) \to 0$ as $n \to \infty$ for all θ satisfying $\theta_i \neq 0$ (θ_i not depending on n) is $\xi_{i,n}\eta_{i,n} \to 0$.

(b) A necessary and sufficient condition for $P_{n,\theta,\sigma}(\hat{\theta}_i = 0) \to 1$ as $n \to \infty$ for all θ satisfying $\theta_i = 0$ is $n^{1/2}\eta_{i,n} \to \infty$.

(c) A necessary and sufficient condition for $P_{n,\theta,\sigma}(\hat{\theta}_i = 0) \to c_i < 1$ as $n \to \infty$ for all θ satisfying $\theta_i = 0$ is $n^{1/2}\eta_{i,n} \to e_i$, $0 \leq e_i < \infty$. The constant c_i is then given by $c_i = \Phi(e_i) - \Phi(-e_i)$.

Part (a) of the above proposition gives a necessary and sufficient condition for the procedure to correctly detect nonzero coefficients with probability converging to 1. Part (b) gives a necessary and sufficient condition for correctly detecting zero coefficients with probability converging to 1.

Remark 5. If $\xi_{i,n}/n^{1/2}$ does not converge to zero, the conditions on $\eta_{i,n}$ in Parts (a) and (b) are incompatible; also the conditions in Parts (a) and (c) are then incompatible (except when $e_i = 0$). However, the case where $\xi_{i,n}/n^{1/2}$ does not converge to zero is of little interest as the least-squares estimator $\hat{\theta}_{LS,i}$ is then not consistent.

Remark 6 (*Speed of convergence in Proposition 4*)**.** (i) The speed of convergence in (a) is $\xi_{i,n}\eta_{i,n}$ in case $n^{1/2}\xi_{i,n}^{-1}$ is bounded (an uninteresting case as noted above); if $n^{1/2}\xi_{i,n}^{-1} \to \infty$, the speed of convergence in (a) is not slower than $\exp\left(-cn\xi_{i,n}^{-2}\right) / \left(n^{1/2}\xi_{i,n}^{-1}\right)$ for some suitable $c > 0$ depending on θ_i/σ.

(ii) The speed of convergence in (b) is $\exp\left(-0.5n\eta_{i,n}^2\right) / \left(n^{1/2}\eta_{i,n}\right)$. In (c) the speed of convergence is given by the rate at which $n^{1/2}\eta_{i,n}$ approaches e_i.

[For the above results we have made use of Lemma VII.1.2 in Feller (1957).]

Remark 7. For $\theta \in \mathbb{R}^{k(n)}$ let $A_n(\theta) = \{i : 1 \leq i \leq k(n), \theta_i \neq 0\}$. Then (i) for every $i \in A_n(\theta)$

$$P_{n,\theta,\sigma}\left(\hat{\theta}_i = 0\right) \leq P_{n,\theta,\sigma}\left(\bigcup_{j \in A_n(\theta)} \left\{\hat{\theta}_j = 0\right\}\right) \leq \sum_{j \in A_n(\theta)} P_{n,\theta,\sigma}\left(\hat{\theta}_j = 0\right).$$

Suppose now that the entries of θ do not change with n (although the dimension of θ may depend on n).[1] Then, given that $\mathrm{card}(A_n(\theta))$ is bounded (this being in particular the case if $k(n)$ is bounded), the probability of incorrect non-detection of at least one nonzero coefficient converges to 0 if and only if $\xi_{i,n}\eta_{i,n} \to 0$ as $n \to \infty$ for every $i \in A_n(\theta)$. [If $\mathrm{card}(A_n(\theta))$ is unbounded then this probability converges to 0, e.g., if $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\xi_{i,n}^{-1} \to \infty$ as $n \to \infty$ for every $i \in A_n(\theta)$ and $\inf_{i \in A_n(\theta)} |\theta_i| > 0$ and $\sum_{i \in A_n(\theta)} \exp\left(-cn\xi_{i,n}^{-2}\right) / \left(n^{1/2}\xi_{i,n}^{-1}\right) \to 0$ as $n \to \infty$ for a suitable c that is determined by $\inf_{i \in A_n(\theta)} |\theta_i| / \sigma$.]

[1] More precisely, this means that θ is made up of the initial $k(n)$ elements of a fixed element of $\mathbb{R}^\infty$.

(ii) For every $i \notin A_n(\theta)$ we have

$$\begin{aligned} P_{n,\theta,\sigma}\left(\hat{\theta}_i = 0\right) &\geq P_{n,\theta,\sigma}\left(\bigcap_{j \notin A_n(\theta)} \left\{\hat{\theta}_j = 0\right\}\right) \\ &= 1 - P_{n,\theta,\sigma}\left(\bigcup_{j \notin A_n(\theta)} \left\{\hat{\theta}_j \neq 0\right\}\right) \\ &\geq 1 - \sum_{j \notin A_n(\theta)} \left[1 - P_{n,\theta,\sigma}\left(\hat{\theta}_j = 0\right)\right]. \end{aligned}$$

Suppose again that the entries of θ do not change with n. Then, given that $\operatorname{card}(A_n^c(\theta))$ is bounded (this being in particular the case if $k(n)$ is bounded), the probability of incorrectly classifying at least one zero parameter as a non-zero one converges to 0 as $n \to \infty$ if and only if $n^{1/2}\eta_{i,n} \to \infty$ for every $i \in A_n(\theta)$. [If $\operatorname{card}(A_n^c(\theta))$ is unbounded then this probability converges to 0, e.g., if $\sum_{i \notin A_n(\theta)} \exp\left(-0.5 n \eta_{i,n}^2\right) / \left(n^{1/2}\eta_{i,n}\right) \to 0$ as $n \to \infty$.]

(iii) In the case where $X'X$ is a diagonal matrix, the relevant probabilities $P_{n,\theta,\sigma}(\bigcup_{i \in A_n(\theta)}\{\hat{\theta}_i = 0\})$ as well as $P_{n,\theta,\sigma}(\bigcap_{i \notin A_n(\theta)}\{\hat{\theta}_i = 0\})$ can be directly expressed in terms of products of $P_{n,\theta,\sigma}(\hat{\theta}_i = 0)$ or $1 - P_{n,\theta,\sigma}(\hat{\theta}_i = 0)$, and Proposition 4 can then be applied.

Since the fixed-parameter asymptotic framework often gives a misleading impression of the actual behavior of a variable selection procedure (cf. Leeb and Pötscher, 2005, Pötscher and Leeb, 2009) we turn to a "moving-parameter" framework next, i.e., we allow the elements of θ as well as σ to depend on sample size n. In the proposition to follow (and all subsequent large-sample results) we shall concentrate only on the case where $\xi_{i,n}\eta_{i,n} \to 0$ as $n \to \infty$, since otherwise the estimators $\hat{\theta}_i$ are not even consistent for θ_i as a consequence of Proposition 4, cf. also Theorem 16 below. Given the condition $\xi_{i,n}\eta_{i,n} \to 0$, we shall then distinguish between the case $n^{1/2}\eta_{i,n} \to e_i$, $0 \leq e_i < \infty$, and the case $n^{1/2}\eta_{i,n} \to \infty$, which in light of Proposition 4 we shall call the case of "conservative tuning" and the case of "consistent tuning", respectively.[2]

Proposition 8. *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to e_i$ where $0 \leq e_i \leq \infty$.*

(a) Assume $e_i < \infty$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$. Then

$$\lim_{n\to\infty} P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i = 0\right) = \Phi\left(-\nu_i + e_i\right) - \Phi\left(-\nu_i - e_i\right).$$

(b) Assume $e_i = \infty$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $\theta_{i,n}/(\sigma_n \xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$. Then

[2] There is no loss of generality here in assuming convergence of $n^{1/2}\eta_{i,n}$ to a (finite or infinite) limit, in the sense that this convergence can, for any given sequence $n^{1/2}\eta_{i,n}$, be achieved along suitable subsequences in light of compactness of the extended real line.

1. $|\zeta_i| < 1$ *implies* $\lim_{n\to\infty} P_{n,\theta^{(n)},\sigma_n}(\hat{\theta}_i = 0) = 1$.
2. $|\zeta_i| > 1$ *implies* $\lim_{n\to\infty} P_{n,\theta^{(n)},\sigma_n}(\hat{\theta}_i = 0) = 0$.
3. $|\zeta_i| = 1$ *and* $r_{i,n} := n^{1/2}\left(\eta_{i,n} - \zeta_i\theta_{i,n}/(\sigma_n\xi_{i,n})\right) \to r_i$, *for some* $r_i \in \bar{\mathbb{R}}$,

imply

$$\lim_{n\to\infty} P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i = 0\right) = \Phi(r_i).$$

In a fixed-parameter asymptotic analysis, which in Proposition 8 corresponds to the case $\theta_{i,n} \equiv \theta_i$ and $\sigma_n \equiv \sigma$, the limit of the probabilities $P_{n,\theta,\sigma}(\hat{\theta}_i = 0)$ is always 0 in case $\theta_i \neq 0$, and is 1 in case $\theta_i = 0$ and consistent tuning (it is $\Phi(e_i) - \Phi(-e_i)$ in case $\theta_i = 0$ and conservative tuning); this does clearly not properly capture the finite-sample behavior of these probabilities. The moving-parameter asymptotic analysis underlying Proposition 8 better captures the finite-sample behavior and, e.g., allows for limits other than 0 and 1 even in the case of consistent tuning. In particular, Proposition 8 shows that the convergence of the variable selection/deletion probabilities to their limits in a fixed-parameter asymptotic framework is not uniform in θ_i, and this non-uniformity is local in the sense that it occurs in an arbitrarily small neighborhood of $\theta_i = 0$ (holding the value of $\sigma > 0$ fixed).[3] Furthermore, the above proposition entails that under consistent tuning deviations from $\theta_i = 0$ of larger order than under conservative tuning go unnoticed asymptotically with probability 1 by the variable selection procedure corresponding to $\hat{\theta}_i$. For more discussion in a special case (which in its essence also applies here) see Pötscher and Leeb (2009).

Remark 9 (*Speed of convergence in Proposition 8*)**.** (i) The speed of convergence in (a) is given by the slower of the rate at which $n^{1/2}\eta_{i,n}$ approaches e_i and $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})$ approaches ν_i provided that $|\nu_i| < \infty$; if $|\nu_i| = \infty$, the speed of convergence is not slower than

$$\exp\left(-cn\theta_{i,n}^2/(\sigma_n^2\xi_{i,n}^2)\right) / \left|n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})\right|$$

for any $c < 1/2$.

(ii) The speed of convergence in (b1) is not slower than $\exp(-cn\eta_{i,n}^2)/(n^{1/2}\eta_{i,n})$ where c depends on ζ_i. The same is true in case (b2) provided $|\zeta_i| < \infty$; if $|\zeta_i| = \infty$, the speed of convergence is not slower than $\exp(-cn\theta_{i,n}^2/(\sigma_n^2\xi_{i,n}^2))/|n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})|$ for every $c < 1/2$. In case (b3) the speed of convergence is not slower than the speed of convergence of

$$\max\left(\exp\left(-cn\eta_{i,n}^2\right) / \left(n^{1/2}\eta_{i,n}\right), |r_{i,n} - r_i|\right)$$

for any $c < 2$ in case $|r_i| < \infty$; in case $|r_i| = \infty$ it is not slower than

$$\max\left(\exp\left(-cn\eta_{i,n}^2\right) / \left(n^{1/2}\eta_{i,n}\right), \exp\left(-0.5r_{i,n}^2\right) / |r_{i,n}|\right)$$

for any $c < 2$.

The preceding remark corrects and clarifies the remarks at the end of Section 3 in Pötscher and Leeb (2009) and Section 3.1 in Pötscher and Schneider (2009).

[3] More generally, the non-uniformity arises for θ_i/σ in a neighborhood of zero.

3.2. Unknown-variance case

In the unknown-variance case the finite-sample variable selection/deletion probabilities can be obtained as follows:

$$\begin{aligned}
P_{n,\theta,\sigma}\left(\tilde{\theta}_i = 0\right) &= P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right| \leq \hat{\sigma}\xi_{i,n}\eta_{i,n}\right) \\
&= \int_0^\infty P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right| \leq \hat{\sigma}\xi_{i,n}\eta_{i,n} \mid \hat{\sigma} = s\sigma\right)\rho_{n-k}(s)ds \\
&= \int_0^\infty P_{n,\theta,\sigma}\left(\hat{\theta}_i(s\eta_{i,n}) = 0\right)\rho_{n-k}(s)ds \\
&= \int_0^\infty \Big[\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + s\eta_{i,n}\right)\right) \\
&\qquad -\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - s\eta_{i,n}\right)\right)\Big]\rho_{n-k}(s)ds \\
&= T_{n-k,n^{1/2}\theta_i/(\sigma\xi_{i,n})}\left(n^{1/2}\eta_{i,n}\right) - T_{n-k,n^{1/2}\theta_i/(\sigma\xi_{i,n})}\left(-n^{1/2}\eta_{i,n}\right). \qquad (5)
\end{aligned}$$

Here we have used (4), and independence of $\hat{\sigma}$ and $\hat{\theta}_{LS,i}$ allowed us to replace $\hat{\sigma}$ by $s\sigma$ in the relevant formulae, cf. Leeb and Pötscher (2003, p. 110). In the above ρ_{n-k} denotes the density of $(n-k)^{-1/2}$ times the square root of a chi-square distributed random variable with $n-k$ degrees of freedom. It will turn out to be convenient to set $\rho_{n-k}(s) = 0$ for $s < 0$, making ρ_{n-k} a bounded continuous function on $\mathbb{R}$.

We now have the following fixed-parameter asymptotic result for the variable selection/deletion probabilities in the unknown-variance case that perfectly parallels the corresponding result in the known-variance case, i.e., Proposition 4:

Proposition 10. *Let $0 < \sigma < \infty$ be given. For every $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have:*

(a) A necessary and sufficient condition for $P_{n,\theta,\sigma}(\tilde{\theta}_i = 0) \to 0$ as $n \to \infty$ for all θ satisfying $\theta_i \neq 0$ (θ_i not depending on n) is $\xi_{i,n}\eta_{i,n} \to 0$.

(b) A necessary and sufficient condition for $P_{n,\theta,\sigma}(\tilde{\theta}_i = 0) \to 1$ as $n \to \infty$ for all θ satisfying $\theta_i = 0$ is $n^{1/2}\eta_{i,n} \to \infty$.

(c) A necessary and sufficient condition for $P_{n,\theta,\sigma}(\tilde{\theta}_i = 0) - c_{i,n} \to 0$ as $n \to \infty$ for all θ satisfying $\theta_i = 0$ and with $c_{i,n} = T_{n-k}(e_i) - T_{n-k}(-e_i)$ satisfying $\limsup_{n\to\infty} c_{i,n} < 1$ *is $n^{1/2}\eta_{i,n} \to e_i$, $0 \leq e_i < \infty$.*

Proposition 10 shows that the dichotomy regarding conservative tuning and consistent tuning is expressed by the same conditions in the unknown-variance case as in the known-variance case. Furthermore, note that $c_{i,n}$ appearing in Part (c) of the above proposition converges to $c_i = \Phi(e_i) - \Phi(-e_i)$ in the case where $n-k \to \infty$, the limit thus being the same as in the known-variance case. This is different in case $n-k$ is constant equal to m, say, eventually, the sequence $c_{i,n}$ then being constant equal to $T_m(e_i) - T_m(-e_i)$ eventually. We finally note that Remark 5 also applies to Proposition 10 above.

For the same reasons as in the known-variance case we next investigate the asymptotic behavior of the variable selection/deletion probabilities under a moving-parameter asymptotic framework. We consider the case where $n-k$ is (eventually) constant and the case where $n-k\to\infty$. There is no essential loss in generality in considering these two cases only, since by compactness of $\mathbb{N}\cup\{\infty\}$ we can always assume (possibly after passing to subsequences) that $n-k$ converges in $\mathbb{N}\cup\{\infty\}$.

Theorem 11. *Suppose that for given $i\geq 1$ satisfying $i\leq k=k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n}\to 0$ and $n^{1/2}\eta_{i,n}\to e_i$ where $0\leq e_i\leq\infty$.*

(a) Assume $e_i<\infty$. Suppose that the true parameters $\theta^{(n)}=(\theta_{1,n},\ldots,\theta_{k_n,n})\in\mathbb{R}^{k_n}$ and $\sigma_n\in(0,\infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})\to\nu_i\in\bar{\mathbb{R}}$.

(a1) If $n-k$ is eventually constant equal to m, say, then

$$\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)=\int_0^\infty\left(\Phi\left(-\nu_i+se_i\right)-\Phi\left(-\nu_i-se_i\right)\right)\rho_m(s)ds.$$

(a2) If $n-k\to\infty$ holds, then

$$\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)=\Phi\left(-\nu_i+e_i\right)-\Phi\left(-\nu_i-e_i\right).$$

(b) Assume $e_i=\infty$. Suppose that the true parameters $\theta^{(n)}=(\theta_{1,n},\ldots,\theta_{k_n,n})\in\mathbb{R}^{k_n}$ and $\sigma_n\in(0,\infty)$ satisfy $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\to\zeta_i\in\bar{\mathbb{R}}$.

(b1) If $n-k$ is eventually constant equal to m, say, then

$$\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)=\int_{|\zeta_i|}^\infty\rho_m(s)ds=\Pr(\chi^2_m>m\zeta_i^2).$$

(b2) If $n-k\to\infty$ holds, then

1. $|\zeta_i|<1$ implies $\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}(\tilde{\theta}_i=0)=1$.

2. $|\zeta_i|>1$ implies $\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}(\tilde{\theta}_i=0)=0$.

3. $|\zeta_i|=1$ and $n^{1/2}\eta_{i,n}/\left(n-k\right)^{1/2}\to 0$ imply

$$\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)=\Phi(r_i)$$

provided $r_{i,n}:=n^{1/2}\left(\eta_{i,n}-\zeta_i\theta_{i,n}/(\sigma_n\xi_{i,n})\right)\to r_i$ for some $r_i\in\bar{\mathbb{R}}$.

4. $|\zeta_i|=1$ and $n^{1/2}\eta_{i,n}/\left(n-k\right)^{1/2}\to 2^{1/2}d_i$ with $0<d_i<\infty$ imply

$$\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)=\int_{-\infty}^\infty\Phi(d_it+r_i)\phi(t)dt$$

provided $r_{i,n}\to r_i$ for some $r_i\in\bar{\mathbb{R}}$. [Note that the integral in the above display reduces to 1 if $r_i=\infty$, and to 0 if $r_i=-\infty$.]

5. $|\zeta_i|=1$ and $n^{1/2}\eta_{i,n}/\left(n-k\right)^{1/2}\to\infty$ imply

$$\lim_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)=\Phi(r_i')$$

provided $(n^{1/2}\eta_{i,n}/\left(n-k\right)^{1/2})^{-1}r_{i,n}\to 2^{-1/2}r_i'$ for some $r_i'\in\bar{\mathbb{R}}$.

Theorem 11 shows, in particular, that also in the unknown-variance case the convergence of the variable selection/deletion probabilities to their limits in a fixed-parameter asymptotic framework is not locally uniform in θ_i. In the case of conservative tuning the theorem furthermore shows that the limit of the variable selection/deletion probabilities in the unknown-variance case is the same as in the known-variance case if the degrees of freedom $n-k$ go to infinity (entailing that the distribution of $\hat{\sigma}/\sigma$ concentrates more and more around 1); if $n-k$ is eventually constant, the limit turns out to be a mixture of the known-variance case limits (with σ replaced by $s\sigma$), the mixture being with respect to the distribution of $\hat{\sigma}/\sigma$. [We note that in the somewhat uninteresting case $e_i = 0$ this mixture also reduces to the same limit as in the known-variance case.] While this result is as one would expect, the situation is different and more subtle in the case of consistent tuning: If $n-k \to \infty$ the limits are the same as in the known-variance case if $|\zeta_i| < 1$ or $|\zeta_i| > 1$ holds, namely 1 and 0, respectively. However, in the "boundary" case $|\zeta_i| = 1$ the rate at which $n-k$ diverges to infinity becomes relevant. If the divergence is fast enough in the sense that $n^{1/2}\eta_{i,n}/\left(n-k\right)^{1/2} \to 0$, again the same limit as in the known-variance case, namely $\Phi(r_i)$, is obtained; but if $n-k$ diverges to infinity more slowly, a different limit arises (which, e.g., in case 4 of Part (b2) is obtained by averaging $\Phi(r_i + \cdot)$ with respect to a suitable distribution). The case where the degrees of freedom $n-k$ is eventually constant looks very much different from the known-variance case and again some averaging with respect to the distribution of $\hat{\sigma}/\sigma$ takes place. Note that in this case the limiting variable deletion probabilities are 1 and 0, respectively, only if $\zeta_i = 0$ and $|\zeta_i| = \infty$, respectively, which is in contrast to the known-variance case (and the unknown-variance case with $n-k \to \infty$).

Remark 12. (i) For later use we note that Proposition 8 and Theorem 11 also hold when applied to subsequences, as is easily seen.

(ii) The convergence conditions in Proposition 8 on the various quantities involving $\theta_{i,n}$ and σ_n are essentially cost-free in the sense that given any sequence $(\theta_{i,n}, \sigma_n)$ we can, due to compactness of $\bar{\mathbb{R}}$, select from any subsequence n_j a further subsubsequence $n_{j(l)}$ such that along this subsubsequence all relevant quantities such as $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n})$ (or $\theta_{i,n}/(\sigma_n \xi_{i,n} \eta_{i,n})$ and $r_{i,n}$) converge in $\bar{\mathbb{R}}$. Since Proposition 8 also holds when applied to subsequences as just noted, an application of this proposition to the subsubsequence $n_{j(l)}$ then results in a characterization of all possible accumulation points of the variable selection/deletion probabilities in the known-variance case.

(iii) In a similar manner, the convergence conditions in Theorem 11 (including the ones on $n-k$) are essentially cost-free, and thus this theorem provides a full characterization of all possible accumulation points of the variable selection/deletion probabilities in the unknown-variance case.

As just discussed, in the case of conservative tuning we get the same limiting behavior under moving-parameter asymptotics in the known-variance and in the unknown-variance case along any sequence of parameters if $n-k \to \infty$ or $e_i = 0$ (which in the conservatively tuned case can equivalently be stated as

$n^{1/2}\eta_{i,n}/(n-k)^{1/2} \to 0$). In the case of consistent tuning the same coincidence of limits occurs if $n-k \to \infty$ fast enough such that $n^{1/2}\eta_{i,n}/(n-k)^{1/2} \to 0$. This is not accidental but a consequence of the following fact:

Proposition 13. *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ as $n \to \infty$. Then*

$$\sup_{\theta\in\mathbb{R}^k, 0<\sigma<\infty} \left| P_{n,\theta,\sigma}\left(\hat{\theta}_i = 0\right) - P_{n,\theta,\sigma}\left(\tilde{\theta}_i = 0\right)\right| \to 0 \qquad \text{for } n \to \infty.$$

Remark 14. Suppose that $\xi_{i,n}\eta_{i,n} \to 0$ holds as $n \to \infty$, the other case being of little interest as noted earlier. If $n^{1/2}\eta_{i,n}(n-k)^{-1/2}$ does not converge to zero as $n \to \infty$, it can be shown from Proposition 8 and Theorem 11 that the limits of the variable deletion probabilities (along appropriate (sub)sequences $(\theta^{(n_j)}, \sigma_{n_j})$) for the known-variance and the unknown-variance case do not coincide. This shows that the condition $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ in the above proposition cannot be weakened (at least in case $\xi_{i,n}\eta_{i,n} \to 0$ holds).

4. Consistency, uniform consistency, and uniform convergence rate

For purposes of comparison we start with the following obvious proposition, which immediately follows from the observation that $\hat{\theta}_{LS,i}$ is $N(\theta_i, \sigma^2\xi^2_{i,n}/n)$-distributed.

Proposition 15. *For every $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have the following:*

(a) $\xi_{i,n}/n^{1/2} \to 0$ is a necessary and sufficient condition for $\hat{\theta}_{LS,i}$ to be consistent for θ_i, the convergence rate being $\xi_{i,n}/n^{1/2}$.

(b) Suppose $\xi_{i,n}/n^{1/2} \to 0$. Then $\hat{\theta}_{LS,i}$ is uniformly consistent for θ_i in the sense that for every $\varepsilon > 0$

$$\lim_{n\to\infty} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i} - \theta_i\right| > \sigma\varepsilon\right) = 0.$$

In fact, $\hat{\theta}_{LS,i}$ is uniformly $n^{1/2}/\xi_{i,n}$-consistent for θ_i in the sense that for every $\varepsilon > 0$ there exists a real number $M > 0$ such that

$$\sup_{n\in\mathbb{N}} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(\left(n^{1/2}/\xi_{i,n}\right)\left|\hat{\theta}_{LS,i} - \theta_i\right| > \sigma M\right) < \varepsilon.$$

[Note that the probabilities in the displays above in fact neither depend on θ nor σ. In particular, the l.h.s. of the above displays equal $2\Phi(-\varepsilon n^{1/2}/\xi_{i,n})$ and $2\Phi(-M)$, respectively.]

The corresponding result for the estimators $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, or $\tilde{\theta}_{AS,i}$ and their infeasible counterparts $\hat{\theta}_{H,i}$, $\hat{\theta}_{S,i}$, or $\hat{\theta}_{AS,i}$ is now as follows.

Theorem 16. *Let $\tilde{\theta}_i$ stand for any of the estimators $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, or $\tilde{\theta}_{AS,i}$. Then for every $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have the following:*

(a) $\tilde{\theta}_i$ is consistent for θ_i if and only if $\xi_{i,n}\eta_{i,n} \to 0$ and $\xi_{i,n}/n^{1/2} \to 0$.

(b) Suppose $\xi_{i,n}\eta_{i,n} \to 0$ and $\xi_{i,n}/n^{1/2} \to 0$. Then $\tilde{\theta}_i$ is uniformly consistent in the sense that for every $\varepsilon > 0$

$$\lim_{n\to\infty} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(\left|\tilde{\theta}_i - \theta_i\right| > \sigma\varepsilon\right) = 0.$$

Furthermore, $\tilde{\theta}_i$ is uniformly $a_{i,n}$-consistent with $a_{i,n} = \min(n^{1/2}/\xi_{i,n}, (\xi_{i,n}\eta_{i,n})^{-1})$ in the sense that for every $\varepsilon > 0$ there exists a real number $M > 0$ such that

$$\sup_{n\in\mathbb{N}} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(a_{i,n}\left|\tilde{\theta}_i - \theta_i\right| > \sigma M\right) < \varepsilon.$$

(c) Suppose $\xi_{i,n}\eta_{i,n} \to 0$ and $\xi_{i,n}/n^{1/2} \to 0$ and $b_{i,n} \geq 0$. If for every $\varepsilon > 0$ there exists a real number $M > 0$ such that

$$\limsup_{n\to\infty} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(b_{i,n}\left|\tilde{\theta}_i - \theta_i\right| > \sigma M\right) < \varepsilon \tag{6}$$

holds, then $b_{i,n} = O(a_{i,n})$ necessarily holds.

(d) Let $\hat{\theta}_i$ stand for any of the estimators $\hat{\theta}_{H,i}$, $\hat{\theta}_{S,i}$, or $\hat{\theta}_{AS,i}$. Then the results in (a)-(c) also hold for $\hat{\theta}_i$.

The preceding theorem shows that the thresholding estimators $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, and $\tilde{\theta}_{AS,i}$ (as well as their infeasible versions) are uniformly $a_{i,n}$-consistent and that this rate is sharp and cannot be improved. In particular, if the tuning is conservative these estimators are uniformly $n^{1/2}/\xi_{i,n}$-consistent, which is the usual rate one expects to find in a linear regression model as considered here. However, if consistent tuning is employed, the preceding theorem shows that these thresholding estimators are then only uniformly $(\xi_{i,n}\eta_{i,n})^{-1}$-consistent, i.e., have a slower uniform convergence rate than the least-squares (maximum likelihood) estimator (or the conservatively tuned thresholding estimators for that matter). For a discussion of the pointwise convergence rate see Section 6.4.

Remark 17. If $n^{1/2}\eta_{i,n} \to e_i = 0$, then $\tilde{\theta}_i$ is asymptotically equivalent to $\hat{\theta}_{LS,i}$ in the sense that for every $\varepsilon > 0$

$$\lim_{n\to\infty} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(\left(n^{1/2}/\xi_{i,n}\right)|\tilde{\theta}_i - \hat{\theta}_{LS,i}| > \sigma\varepsilon\right) = 0.$$

A similar statement holds for $\hat{\theta}_i$. For $\tilde{\theta}_i$ this follows immediately from (27) in Section 8 and the fact that the family of distributions corresponding to ρ_{n-k} is tight; for $\hat{\theta}_i$ this follows from the relation $|\hat{\theta}_i - \hat{\theta}_{LS,i}| \leq \sigma\xi_{i,n}\eta_{i,n}$.

Remark 18. (i) A variation of the proof of Theorem 16 shows that in case of consistent tuning for the infeasible estimators additionally also

$$\lim_{n\to\infty} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(a_{i,n}\left|\hat{\theta}_i - \theta_i\right| > \sigma M\right) = 0$$

holds for every $M > 1$, and that for the feasible estimators

$$\lim_{n\to\infty} \sup_{\theta\in\mathbb{R}^k} \sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(a_{i,n}\left|\tilde{\theta}_i - \theta_i\right| > \sigma M\right) = 0$$

holds for every $M > 1$ provided that $n - k \to \infty$.

(ii) Inspection of the proof shows that the conclusion of Theorem 16(c) continues to hold if the supremum over $\mathbb{R}^k$ is replaced by the supremum over an arbitrarily small neighborhood of 0 and σ is held fixed at an arbitrary positive value.

(iii) If $\sigma\varepsilon$ and σM are replaced by ε and M, respectively, in the displays in Proposition 15 and Theorem 16 as well as in Remark 17, the resulting statements remain true provided the suprema over $0 < \sigma < \infty$ are replaced by suprema over $0 < \sigma \le c$, where $c > 0$ is an arbitrary real number.

5. Finite-sample distributions

5.1. Known-variance case

We next present the finite-sample distributions of the infeasible thresholding estimators. It will turn out to be convenient to give the results for scaled versions, where the scaling factor $\alpha_{i,n}$ is a positive real number, but is otherwise arbitrary. *Note that below we suppress the dependence of the distribution functions of the thresholding estimators on the scaling sequence $\alpha_{i,n}$ in the notation.* Furthermore, observe that the finite-sample distributions depend on θ only through θ_i.

Proposition 19. *The cdf $H^i_{H,n,\theta,\sigma} := H^i_{H,\eta_{i,n},n,\theta,\sigma}$ of $\sigma^{-1}\alpha_{i,n}(\hat{\theta}_{H,i} - \theta_i)$ is given by*

$$\begin{aligned} H^i_{H,n,\theta,\sigma}(x) &= \Phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})\right)\mathbf{1}\left(\left|\alpha_{i,n}^{-1}x + \theta_i/\sigma\right| > \xi_{i,n}\eta_{i,n}\right) \\ &\quad + \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}\right)\right)\mathbf{1}\left(0 \le \alpha_{i,n}^{-1}x + \theta_i/\sigma \le \xi_{i,n}\eta_{i,n}\right) \\ &\quad + \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}\right)\right)\mathbf{1}\left(-\xi_{i,n}\eta_{i,n} \le \alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right), \end{aligned} \tag{7}$$

or, equivalently,

$$\begin{aligned} &dH^i_{H,n,\theta,\sigma}(x) \\ &\quad = \left\{\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}\right)\right) - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}\right)\right)\right\} \\ &\quad \times d\delta_{-\alpha_{i,n}\theta_i/\sigma}(x) \\ &\quad + \left(n^{1/2}/(\alpha_{i,n}\xi_{i,n})\right)\phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})\right)\mathbf{1}\left(\left|\alpha_{i,n}^{-1}x + \theta_i/\sigma\right| > \xi_{i,n}\eta_{i,n}\right)dx \end{aligned} \tag{8}$$

where δ_z denotes pointmass at z.

Proposition 20. *The cdf* $H^i_{S,n,\theta,\sigma} := H^i_{S,\eta_{i,n},n,\theta,\sigma}$ *of* $\sigma^{-1}\alpha_{i,n}(\hat{\theta}_{S,i}-\theta_i)$ *is given by*

$$H^i_{S,n,\theta,\sigma}(x) = \Phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n}) + n^{1/2}\eta_{i,n}\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma \geq 0\right) + \Phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n}) - n^{1/2}\eta_{i,n}\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right), \quad (9)$$

or, equivalently,

$$\begin{aligned} &dH^i_{S,n,\theta,\sigma}(x) \\ &= \left\{\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}\right)\right) - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}\right)\right)\right\} \\ &\times d\delta_{-\alpha_{i,n}\theta_i/\sigma}(x) \\ &+ \left(n^{1/2}/(\alpha_{i,n}\xi_{i,n})\right)\left\{\phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n}) + n^{1/2}\eta_{i,n}\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma > 0\right)\right. \\ &\left.+\phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n}) - n^{1/2}\eta_{i,n}\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right)\right\} dx. \end{aligned} \quad (10)$$

Proposition 21. *The cdf* $H^i_{AS,n,\theta,\sigma} := H^i_{AS,\eta_{i,n},n,\theta,\sigma}$ *of* $\sigma^{-1}\alpha_{i,n}(\hat{\theta}_{AS,i}-\theta_i)$ *is given by*

$$\begin{aligned} H^i_{AS,n,\theta,\sigma}(x) &= \Phi\left(z^{(2)}_{n,\theta,\sigma}(x,\eta_{i,n})\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma \geq 0\right) \\ &+ \Phi\left(z^{(1)}_{n,\theta,\sigma}(x,\eta_{i,n})\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right), \end{aligned} \quad (11)$$

where $z^{(1)}_{n,\theta,\sigma}(x,y) \leq z^{(2)}_{n,\theta,\sigma}(x,y)$ *are defined by*

$$0.5n^{1/2}\xi_{i,n}^{-1}(\alpha_{i,n}^{-1}x - \theta_i/\sigma) \pm n^{1/2}\sqrt{\left(0.5\xi_{i,n}^{-1}(\alpha_{i,n}^{-1}x + \theta_i/\sigma)\right)^2 + y^2}.$$

Or, equivalently,

$$\begin{aligned} &dH^i_{AS,n,\theta,\sigma}(x) \\ &= \left\{\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}\right)\right) - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}\right)\right)\right\} \\ &\times d\delta_{-\alpha_{i,n}\theta_i/\sigma}(x) \; + \; (0.5n^{1/2}/(\alpha_{i,n}\xi_{i,n})) \\ &\times \left\{\phi\left(z^{(2)}_{n,\theta,\sigma}(x,\eta_{i,n})\right)(1 + t_{n,\theta,\sigma}(x,\eta_{i,n}))\,\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma > 0\right)\right. \\ &\left.+ \; \phi\left(z^{(1)}_{n,\theta,\sigma}(x,\eta_{i,n})\right)(1 - t_{n,\theta,\sigma}(x,\eta_{i,n}))\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right)\right\}, \end{aligned}$$

where $t_{n,\theta,\sigma}(x,y) = 0.5\xi_{i,n}^{-1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma\right) / \left((0.5\xi_{i,n}^{-1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma\right))^2 + y^2\right)^{1/2}$.

The finite-sample distributions of $\hat{\theta}_{H,i}$, $\hat{\theta}_{S,i}$, and $\hat{\theta}_{AS,i}$ are seen to be non-normal. They are made up of two components, one being a multiple of pointmass at $-\alpha_{i,n}\theta_i/\sigma$ and the other one being absolutely continuous with a density that is generally bimodal. For more discussion and some graphical illustrations in a special case see Pötscher and Leeb (2009) and Pötscher and Schneider (2009).

Remark 22. In the case where $X'X$ is diagonal, the estimators of the components θ_i and θ_j for $i \neq j$ are independent and hence the above results immediately allow one to determine the finite-sample distributions of the entire vectors $\hat{\theta}_H$, $\hat{\theta}_S$, and $\hat{\theta}_{AS}$. In particular, this provides the finite-sample distribution of the Lasso $\hat{\theta}_L$ and the adaptive Lasso $\hat{\theta}_{AS}$ in the diagonal case (cf. Remarks 1 and 2).

5.2. Unknown-variance case

The finite-sample distributions of $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, $\tilde{\theta}_{AS,i}$ are obtained next. The same remark on the scaling as in the previous section applies here.

Proposition 23. *The cdf* $H^{i\maltese}_{H,n,\theta,\sigma} := H^{i\maltese}_{H,\eta_{i,n},n,\theta,\sigma}$ *of* $\sigma^{-1}\alpha_{i,n}(\tilde{\theta}_{H,i}-\theta_i)$ *is given by*

$$\begin{aligned}
H^{i\maltese}_{H,n,\theta,\sigma}(x) &= \Phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})\right)\int_0^\infty \mathbf{1}\left(\left|\alpha_{i,n}^{-1}x+\theta_i/\sigma\right| > \xi_{i,n}s\eta_{i,n}\right)\rho_{n-k}(s)ds \\
&+ \int_0^\infty \Phi\Big(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n})+s\eta_{i,n}\right)\Big)\mathbf{1}\left(0 \le \alpha_{i,n}^{-1}x+\theta_i/\sigma \le \xi_{i,n}s\eta_{i,n}\right)\rho_{n-k}(s)ds \\
&+ \int_0^\infty \Phi\Big(n^{1/2}(-\theta_i/(\sigma\xi_{i,n})-s\eta_{i,n})\Big)\mathbf{1}(-\xi_{i,n}s\eta_{i,n} \le \alpha_{i,n}^{-1}x+\theta_i/\sigma < 0)\rho_{n-k}(s)ds.
\end{aligned} \tag{12}$$

Or, equivalently,

$$\begin{aligned}
dH^{i\maltese}_{H,n,\theta,\sigma}(x) \;=\; &\int_0^\infty \Big\{\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n})+s\eta_{i,n}\right)\right) \\
&-\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n})-s\eta_{i,n}\right)\right)\Big\}\rho_{n-k}(s)dsd\delta_{-\alpha_{i,n}\theta_i/\sigma}(x) + n^{1/2}\alpha_{i,n}^{-1}\xi_{i,n}^{-1} \\
&\times \phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})\right)\int_0^\infty \mathbf{1}\left(\left|\alpha_{i,n}^{-1}x+\theta_i/\sigma\right| > \xi_{i,n}s\eta_{i,n}\right)\rho_{n-k}(s)dsdx.
\end{aligned} \tag{13}$$

Proposition 24. *The cdf* $H^{i\maltese}_{S,n,\theta,\sigma} := H^{i\maltese}_{S,\eta_{i,n},n,\theta,\sigma}$ *of* $\sigma^{-1}\alpha_{i,n}(\tilde{\theta}_{S,i}-\theta_i)$ *is given by*

$$\begin{aligned}
&H^{i\maltese}_{S,n,\theta,\sigma}(x) \\
&\quad= \int_0^\infty \Phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})+n^{1/2}s\eta_{i,n}\right)\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x+\theta_i/\sigma \ge 0\right) \\
&\qquad+ \int_0^\infty \Phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})-n^{1/2}s\eta_{i,n}\right)\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x+\theta_i/\sigma < 0\right) \\
&\quad= T_{n-k,-n^{1/2}x/(\alpha_{i,n}\xi_{i,n})}\left(n^{1/2}\eta_{i,n}\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x+\theta_i/\sigma \ge 0\right) \\
&\qquad+ T_{n-k,-n^{1/2}x/(\alpha_{i,n}\xi_{i,n})}\left(-n^{1/2}\eta_{i,n}\right)\mathbf{1}\left(\alpha_{i,n}^{-1}x+\theta_i/\sigma < 0\right).
\end{aligned} \tag{14}$$

Or, equivalently,

$$
\begin{aligned}
dH^{i\maltese}_{S,n,\theta,\sigma}(x) \;=\; & \int_0^\infty \left\{\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + s\eta_{i,n}\right)\right)\right. \\
& \left. -\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - s\eta_{i,n}\right)\right)\right\} \rho_{n-k}(s)dsd\delta_{-\alpha_{i,n}\theta_i/\sigma}(x) + n^{1/2}\alpha_{i,n}^{-1}\xi_{i,n}^{-1} \\
& \times \left\{\int_0^\infty \phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n}) + n^{1/2}s\eta_{i,n}\right)\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma > 0\right)\right. \\
& \left. + \int_0^\infty \phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n}) - n^{1/2}s\eta_{i,n}\right)\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right)\right\} dx.
\end{aligned}
\tag{15}
$$

Proposition 25. *The cdf* $H^{i\maltese}_{AS,n,\theta,\sigma} := H^{i\maltese}_{AS,\eta_{i,n},n,\theta,\sigma}$ *of* $\sigma^{-1}\alpha_{i,n}(\tilde{\theta}_{AS,i} - \theta_i)$ *is given by*

$$
\begin{aligned}
H^{i\maltese}_{AS,n,\theta,\sigma}(x) = & \int_0^\infty \Phi\left(z^{(2)}_{n,\theta,\sigma}(x, s\eta_{i,n})\right)\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma \geq 0\right) \\
& + \int_0^\infty \Phi\left(z^{(1)}_{n,\theta,\sigma}(x, s\eta_{i,n})\right)\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right).
\end{aligned}
\tag{16}
$$

Or, equivalently,

$$
\begin{aligned}
dH^{i\maltese}_{AS,n,\theta,\sigma}(x) = & \int_0^\infty \left\{\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + s\eta_{i,n}\right)\right)\right. \\
& \left. - \Phi\left(n^{1/2}(-\theta_i/(\sigma\xi_{i,n}) - s\eta_{i,n})\right)\right\}\rho_{n-k}(s)dsd\delta_{-\alpha_{i,n}\theta_i/\sigma}(x) + (0.5n^{1/2}/(\alpha_{i,n}\xi_{i,n})) \\
& \times \left\{\int_0^\infty \phi\left(z^{(2)}_{n,\theta,\sigma}(x, s\eta_{i,n})\right)(1 + t_{n,\theta,\sigma}(x, s\eta_{i,n}))\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma > 0\right)\right. \\
& \left. + \int_0^\infty \phi\left(z^{(1)}_{n,\theta,\sigma}(x, s\eta_{i,n})\right)(1 - t_{n,\theta,\sigma}(x, s\eta_{i,n}))\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x + \theta_i/\sigma < 0\right)\right\} dx.
\end{aligned}
\tag{17}
$$

As in the known-variance case the distributions are a convex combination of pointmass and an absolutely continuous part. In case of hard-thresholding, the averaging with respect to the density ρ_{n-k} smoothes the indicator functions leading to a continuous density function for the absolutely continuous part (while in the known-variance case the density function is only piece-wise continuous, cf. Figure 1 in Pötscher and Leeb (2009)). This is not so for soft-thresholding and adaptive soft-thresholding, where the averaging with respect to the density ρ_{n-k} does not affect the indicator functions involved; here the shape of the distribution is qualitatively the same as in the known-variance case (Figure 2 in Pötscher and Leeb, 2009 and Figure 1 in Pötscher and Schneider, 2009).

Remark 26. In the case where $X'X$ is diagonal, the finite-sample distributions of the entire vectors $\tilde{\theta}_H$, $\tilde{\theta}_S$, and $\tilde{\theta}_{AS}$ can be found from the distributions of $\hat{\theta}_H$,

$\hat{\theta}_S$, and $\hat{\theta}_{AS}$ (see Remark 22) by conditioning on $\hat{\sigma} = s\sigma$ and integrating with respect to $\rho_{n-k}(s)$. In particular, this provides the finite-sample distributions of the Lasso $\tilde{\theta}_L$ and the adaptive Lasso $\tilde{\theta}_{AS}$ in the diagonal case (cf. Remarks 1 and 2).

6. Large-sample distributions

We next derive the asymptotic distributions of the thresholding estimators under a moving-parameter (and not only under a fixed-parameter) framework since it is well-known that asymptotics based only on a fixed-parameter framework often lead to misleading conclusions regarding the performance of the estimators (cf. also the discussion in Section 6.4).

6.1. The known-variance case

We first consider the infeasible versions of the thresholding estimators.

Proposition 27. *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to e_i$ where $0 \leq e_i \leq \infty$.*

(a) Assume $e_i < \infty$. Set the scaling factor $\alpha_{i,n} = n^{1/2}/\xi_{i,n}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$. Then $H^i_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution with cdf

$$\begin{aligned} \Phi(x)\mathbf{1}(|x+\nu_i| > e_i) \quad &+ \quad \Phi(-\nu_i + e_i)\mathbf{1}(0 \leq x + \nu_i \leq e_i) \\ &+ \quad \Phi(-\nu_i - e_i)\mathbf{1}(-e_i \leq x + \nu_i < 0), \end{aligned}$$

the corresponding measure being

$$\{\Phi(-\nu_i + e_i) - \Phi(-\nu_i - e_i)\}\, d\delta_{-\nu_i}(x) + \phi(x)\mathbf{1}(|x+\nu_i| > e_i)\, dx. \tag{18}$$

[This distribution reduces to a standard normal distribution in case $|\nu_i| = \infty$ or $e_i = 0$.]

(b) Assume $e_i = \infty$. Set the scaling factor $\alpha_{i,n} = (\xi_{i,n}\eta_{i,n})^{-1}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$.

1. If $|\zeta_i| < 1$, then $H^i_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to $\delta_{-\zeta_i}$.

2. If $|\zeta_i| > 1$, then $H^i_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to δ_0.

3. If $|\zeta_i| = 1$ and $n^{1/2}(\eta_{i,n} - \zeta_i\theta_{i,n}/(\sigma_n\xi_{i,n})) \to r_i$, for some $r_i \in \bar{\mathbb{R}}$, then $H^i_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to

$$\Phi(r_i)\delta_{-\zeta_i} + (1 - \Phi(r_i))\delta_0.$$

Proposition 28. *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to e_i$ where $0 \leq e_i \leq \infty$.*

(a) Assume $e_i < \infty$. Set the scaling factor $\alpha_{i,n} = n^{1/2}/\xi_{i,n}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0, \infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$. Then $H^i_{S,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution with cdf

$$\Phi\left(x + e_i\right) \mathbf{1}\left(x + \nu_i \geq 0\right) + \Phi\left(x - e_i\right) \mathbf{1}\left(x + \nu_i < 0\right),$$

the corresponding measure being

$$\begin{aligned} &\left\{\Phi\left(-\nu_i + e_i\right) - \Phi\left(-\nu_i - e_i\right)\right\} d\delta_{-\nu_i}(x) \\ &\qquad + \left\{\phi\left(x + e_i\right) \mathbf{1}\left(x + \nu_i > 0\right) + \phi\left(x - e_i\right) \mathbf{1}\left(x + \nu_i < 0\right)\right\} dx. \end{aligned} \tag{19}$$

[This distribution reduces to a $N(-\operatorname{sign}(\nu_i)e_i, 1)$-distribution in case $|\nu_i| = \infty$ or $e_i = 0$.]

(b) Assume $e_i = \infty$. Set the scaling factor $\alpha_{i,n} = (\xi_{i,n}\eta_{i,n})^{-1}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0, \infty)$ satisfy $\theta_{i,n}/(\sigma_n \xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$. Then $H^i_{S,n,\theta^{(n)},\sigma_n}$ converges weakly to $\delta_{-\operatorname{sign}(\zeta_i)\min(1,|\zeta_i|)}$.

Proposition 29. *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to e_i$ where $0 \leq e_i \leq \infty$.*

(a) Assume $e_i < \infty$. Set the scaling factor $\alpha_{i,n} = n^{1/2}/\xi_{i,n}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0, \infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$. Then $H^i_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution with cdf

$$\begin{aligned} &\Phi\left(0.5(x - \nu_i) + \sqrt{\left(0.5(x + \nu_i)\right)^2 + e_i^2}\right) \mathbf{1}\left(x + \nu_i \geq 0\right) \\ &+\Phi\left(0.5(x - \nu_i) - \sqrt{\left(0.5(x + \nu_i)\right)^2 + e_i^2}\right) \mathbf{1}\left(x + \nu_i < 0\right) \end{aligned} \tag{20}$$

in case $|\nu_i| < \infty$, the corresponding measure being

$$\begin{aligned} &\left\{\Phi\left(-\nu_i + e_i\right) - \Phi\left(-\nu_i - e_i\right)\right\} d\delta_{-\nu_i}(x) \\ &+0.5\left\{\phi\left(0.5(x - \nu_i) + \sqrt{\left(0.5(x + \nu_i)\right)^2 + e_i^2}\right)(1 + t(x))\, \mathbf{1}\left(x + \nu_i > 0\right)\right. \\ &\left. + \phi\left(0.5(x - \nu_i) - \sqrt{\left(0.5(x + \nu_i)\right)^2 + e_i^2}\right)(1 - t(x))\, \mathbf{1}\left(x + \nu_i < 0\right)\right\} dx, \end{aligned}$$

where $t(x) = \left(x + \nu_i\right)/\sqrt{((x + \nu_i)^2 + 4e_i^2)}$. In the case where $|\nu_i| = \infty$, the cdf $H^i_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to Φ, i.e., to a standard normal distribution. [In case $e_i = 0$ the limit always reduces to a standard normal distribution.]

(b) Assume $e_i = \infty$. Set the scaling factor $\alpha_{i,n} = (\xi_{i,n}\eta_{i,n})^{-1}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0, \infty)$ satisfy $\theta_{i,n}/(\sigma_n \xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$.

1. *If $|\zeta_i| < 1$, then $H^i_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to $\delta_{-\zeta_i}$.*
2. *If $1 \leq |\zeta_i| < \infty$, then $H^i_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to δ_{-1/ζ_i}.*
3. *If $|\zeta_i| = \infty$, then $H^i_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to δ_0.*

Observe that the scaling factors $\alpha_{i,n}$ used in the above propositions are exactly of the same order as $a_{i,n}$ in the case of conservative as well as in the case of consistent tuning and thus correspond to the uniform rate of convergence in both cases. In the case of conservative tuning the limiting distributions have essentially the same form as the finite-sample distributions, demonstrating that the moving-parameter asymptotic framework captures the finite-sample behavior of the estimators in a satisfactory way. In contrast, a fixed-parameter asymptotic framework, which corresponds to setting $\theta_{i,n} \equiv \theta_i$ and $\sigma_n \equiv \sigma$ in the above propositions, misrepresents the finite-sample properties of the thresholding estimators whenever $\theta_i \neq 0$ but small, as the fixed-parameter limiting distribution is – in case of hard-thresholding and adaptive soft-thresholding – then always $N(0,1)$, regardless of the size of θ_i. For soft-thresholding we also observe a strong discrepancy between the finite-sample distribution and the fixed-parameter limit for $\theta_i \neq 0$ which is given by $N(-\operatorname{sign}(\theta_i)e_i, 1)$. In particular, the above propositions demonstrate non-uniformity in the convergence of finite-sample distributions to their limit in a fixed-parameter framework.

In the case of consistent tuning we observe an interesting phenomenon, namely that the limiting distributions now correspond to pointmasses (but not always located at zero!), or are convex combinations of two pointmasses in some cases when considering the hard-thresholding estimator. This essentially means that consistently tuned thresholding estimators are plagued by a bias-problem in that the "bias-component" is the dominant component and is of larger order than the "stochastic variability" of the estimator.[4] In a fixed-parameter framework we get the trivial limits δ_0 for every value of θ_i in case of hard-thresholding and adaptive soft-thresholding. At first glance this seems to suggest that we have used a scaling sequence that does not increase fast enough with n, but recall that the scaling used here corresponds to the uniform convergence rate. We shall take this issue further up in Section 6.4. The situation is different for the soft-thresholding estimator where the fixed-parameter limit is $\delta_{-\operatorname{sign}(\theta_i)}$, which reduces to δ_0 only for $\theta_i = 0$; this is a reflection of the well-known fact that soft-thresholding is plagued by bias problems to a higher degree than are hard-thresholding and adaptive soft-thresholding.

6.2. Uniform closeness of distributions in the known- and unknown-variance case

We next show that the finite-sample cdfs of $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, and $\tilde{\theta}_{AS,i}$ and of their infeasible counterparts $\hat{\theta}_{H,i}$, $\hat{\theta}_{S,i}$, and $\hat{\theta}_{AS,i}$, respectively, are uniformly (with respect to the parameters) close in the total variation distance (or the supremum norm) provided the number of degrees of freedom $n-k$ diverges to infinity fast enough. Apart from being of interest in their own right, these results will be

[4] For the hard-thresholding estimator some randomness survives in the limit in the case $|\zeta_i| = 1$, where we can achieve a limiting probability for $\hat{\theta}_{H,i} = 0$ that is strictly between 0 and 1. That this randomness does not survive for the other two estimators in the limit seems to be connected to the fact that these estimators are continuous functions of the data, whereas $\hat{\theta}_{H,i}$ is not.

instrumental in the subsequent section. We note that the results in Theorem 30 below hold for any choice of the scaling factors $\alpha_{i,n}$.

Theorem 30. *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ as $n \to \infty$. Then*

$$\sup_{\theta \in \mathbb{R}^k, 0<\sigma<\infty} \left\| H^{i}_{H,n,\theta,\sigma} - H^{i\maltese}_{H,n,\theta,\sigma} \right\|_{TV} \to 0 \qquad \text{for } n \to \infty,$$

$$\sup_{\theta \in \mathbb{R}^k, 0<\sigma<\infty} \left\| H^{i}_{S,n,\theta,\sigma} - H^{i\maltese}_{S,n,\theta,\sigma} \right\|_{TV} \to 0 \qquad \text{for } n \to \infty,$$

and

$$\sup_{\theta \in \mathbb{R}^k, 0<\sigma<\infty} \left\| H^{i}_{AS,n,\theta,\sigma} - H^{i\maltese}_{AS,n,\theta,\sigma} \right\|_{\infty} \to 0 \qquad \text{for } n \to \infty$$

hold.[5]

Remark 31. In case of conservative tuning, the condition $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ is always satisfied if $n-k \to \infty$. [In fact it is then equivalent to $n-k \to \infty$ or $e_i = 0$.] In case of consistent tuning $n-k \to \infty$ is clearly a weaker condition than $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$. However, in general, a sufficient condition for $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ is that $\eta_{i,n} \to 0$ and $\limsup_{n\to\infty} k/n < 1$.

Remark 32. Suppose that $\xi_{i,n}\eta_{i,n} \to 0$ holds as $n \to \infty$. If $n^{1/2}\eta_{i,n}(n-k)^{-1/2}$ does not converge to zero as $n \to \infty$, Remark 14 shows that none of the convergence results in Theorem 30 holds. [To see this note that the variable deletion probabilities constitute the weight of the pointmass in the respective distribution functions.] This shows that the condition $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ in the above theorem cannot be weakened (at least in case $\xi_{i,n}\eta_{i,n} \to 0$ holds).

6.3. The unknown-variance case

6.3.1. Conservative tuning

We next obtain the limiting distributions of $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, and $\tilde{\theta}_{AS,i}$ in a moving-parameter framework under conservative tuning.

Theorem 33 (Hard-thresholding with conservative tuning)**.** *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to e_i$ where $0 \leq e_i < \infty$. Set the scaling factor $\alpha_{i,n} = n^{1/2}/\xi_{i,n}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$.*

(a) If $n-k$ is eventually constant equal to m, say, then $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution with cdf

$$\int_0^\infty \left\{ \Phi(x)\,\mathbf{1}(|x+\nu_i| > se_i) + \Phi(-\nu_i + se_i)\,\mathbf{1}(0 \leq x+\nu_i \leq se_i) \right.$$
$$\left. + \,\Phi(-\nu_i - se_i)\,\mathbf{1}(-se_i \leq x+\nu_i < 0) \right\} \rho_m(s)\,ds,$$

[5]Uniform closeness of the respective cdfs of the adaptive soft-thresholding estimators in the total variation distance, and not only in the supremum norm, could probably be obtained at the expense of a more cumbersome proof. We do not pursue this.

the corresponding measure being

$$\int_0^\infty \left\{\Phi\left(-\nu_i + se_i\right) - \Phi\left(-\nu_i - se_i\right)\right\} \rho_m(s) ds d\delta_{-\nu_i}(x) + \phi\left(x\right) \int_0^\infty \mathbf{1}\left(\left|x+\nu_i\right| > se_i\right) \rho_m(s) ds dx. \tag{21}$$

[The distribution reduces to a standard normal distribution in case $|\nu_i| = \infty$ or $e_i = 0$.]

(b) If $n-k \to \infty$ holds, then $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution given in Proposition 27(a).

Theorem 34 (Soft-thresholding with conservative tuning)**.** *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to e_i$ where $0 \leq e_i < \infty$. Set the scaling factor $\alpha_{i,n} = n^{1/2}/\xi_{i,n}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$.*

(a) If $n-k$ is eventually constant equal to m, say, then $H^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution with cdf

$$\int_0^\infty \left\{\Phi\left(x+se_i\right)\mathbf{1}\left(x+\nu_i \geq 0\right) + \Phi\left(x-se_i\right)\mathbf{1}\left(x+\nu_i < 0\right)\right\} \rho_m(s) ds,$$

the corresponding measure being

$$\int_0^\infty \left\{\Phi\left(-\nu_i + se_i\right) - \Phi\left(-\nu_i - se_i\right)\right\} \rho_m(s) ds d\delta_{-\nu_i}(x) + \int_0^\infty \left\{\phi(x+se_i)\mathbf{1}(x+\nu_i > 0) + \phi(x-se_i)\mathbf{1}(x+\nu_i < 0)\right\} \rho_m(s) ds dx. \tag{22}$$

[The atomic part in the above expression is absent in case $|\nu_i| = \infty$. Furthermore, the distribution reduces to a standard normal distribution if $e_i = 0$.]

(b) If $n-k \to \infty$ holds, then $H^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution given in Proposition 28(a).

Theorem 35 (Adaptive soft-thresholding with conservative tuning)**.** *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to e_i$ where $0 \leq e_i < \infty$. Set the scaling factor $\alpha_{i,n} = n^{1/2}/\xi_{i,n}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$.*

(a) Suppose $n-k$ is eventually constant equal to m, say. Then $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution with cdf

$$\int_0^\infty \Phi\left(0.5(x-\nu_i) + \sqrt{\left(0.5(x+\nu_i)\right)^2 + s^2e_i^2}\right) \rho_m(s) ds \mathbf{1}\left(x+\nu_i \geq 0\right) + \int_0^\infty \Phi\left(0.5(x-\nu_i) - \sqrt{\left(0.5(x+\nu_i)\right)^2 + s^2e_i^2}\right) \rho_m(s) ds \mathbf{1}\left(x+\nu_i < 0\right) \tag{23}$$

in case $|\nu_i| < \infty$, the corresponding measure being given by

$$\int_0^\infty \{\Phi(-\nu_i + se_i) - \Phi(-\nu_i - se_i)\} \rho_m(s) ds d\delta_{-\nu_i}(x) \;+\; 0.5$$
$$\times \int_0^\infty \left\{ \phi\left(0.5(x-\nu_i) + \sqrt{(0.5(x+\nu_i))^2 + s^2 e_i^2}\right)(1 + t(x,s))\,\mathbf{1}\,(x+\nu_i > 0)\right.$$
$$\left. + \,\phi\left(0.5(x-\nu_i) - \sqrt{(0.5(x+\nu_i))^2 + s^2 e_i^2}\right)(1 - t(x,s))\,\mathbf{1}\,(x+\nu_i < 0)\right\}$$
$$\times\, \rho_m(s) ds dx,$$

where $t(x,s) = (x+\nu_i)/\sqrt{((x+\nu_i)^2 + 4s^2e_i^2)}$. In case $|\nu_i| = \infty$, the cdf $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to Φ, i.e., a standard normal distribution. [If $e_i = 0$, the limit always reduces to a standard normal distribution.]

(b) If $n - k \to \infty$, then $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution given in Proposition 29(a).

It transpires that in case of conservative tuning and $n - k \to \infty$ we obtain exactly the same limiting distributions as in the known-variance case and hence the relevant discussion given at the end of Section 6.1 applies also here. [That one obtains the same limits does not come as a surprise given the results in Section 6.2 and the observation made in Remark 31.] In the case, where $n - k$ is eventually constant, the limits are obtained from the limits in the known-variance case (with σ replaced by σs) by averaging with respect to the distribution of $\hat{\sigma}/\sigma$. Again the limiting distributions essentially have the same structure as the corresponding finite-sample distributions. The fixed-parameter limiting distributions (corresponding to setting $\theta_{i,n} \equiv \theta_i$ and $\sigma_n \equiv \sigma$ in the above theorems) again misrepresent the finite-sample properties of the thresholding estimators whenever $\theta_i \neq 0$ but small, as the fixed-parameter limiting distribution is – in case of hard-thresholding and adaptive soft-thresholding – then always $N(0,1)$, regardless of the size of θ_i. For soft-thresholding we also observe a strong discrepancy between the finite-sample distribution and the fixed-parameter limit especially for $\theta_i \neq 0$ but small, which is given by the distribution with pdf $\int_0^\infty \phi\left(x + s\,\mathrm{sign}(\theta_i)e_i\right)\rho_m(s)ds$ regardless of the size of θ_i. As a consequence, we again observe non-uniformity in the convergence of finite-sample distributions to their limit in a fixed-parameter framework also in the case where the number of degrees of freedom is (eventually) constant.

6.3.2. *Consistent tuning*

We next derive the limiting distributions of $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, and $\tilde{\theta}_{AS,i}$ in a moving-parameter framework under consistent tuning.

Theorem 36 (Hard-thresholding with consistent tuning)**.** *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to \infty$. Set the scaling factor $\alpha_{i,n} = (\xi_{i,n}\eta_{i,n})^{-1}$. Suppose that* *the true parameters* $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ *and* $\sigma_n \in (0,\infty)$ *satisfy* $\theta_{i,n}/(\sigma_n \xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$.

(a) If $n-k$ *is eventually constant equal to* m*, say, then* $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ *converges weakly to*

$$\left(\int_{|\zeta_i|}^{\infty} \rho_m(s)ds\right)\delta_{-\zeta_i} + \left(1 - \int_{|\zeta_i|}^{\infty} \rho_m(s)ds\right)\delta_0$$
$$= \Pr(\chi^2_m > m\zeta_i^2)\delta_{-\zeta_i} + \Pr(\chi^2_m \leq m\zeta_i^2)\delta_0.$$

[The above display reduces to δ_0 *for* $|\zeta_i| = \infty$*.]*

(b) If $n-k \to \infty$ *holds, then*

1. $|\zeta_i| < 1$ *implies that* $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ *converges weakly to* $\delta_{-\zeta_i}$.
2. $|\zeta_i| > 1$ *implies that* $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ *converges weakly to* δ_0.
3. $|\zeta_i| = 1$ *and* $n^{1/2}\eta_{i,n}/(n-k)^{1/2} \to 0$ *imply that* $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ *converges weakly to*

$$\Phi(r_i)\delta_{-\zeta_i} + (1-\Phi(r_i))\,\delta_0$$

provided $r_{i,n} = n^{1/2}(\eta_{i,n} - \zeta_i\theta_{i,n}/(\sigma_n\xi_{i,n})) \to r_i$ *for some* $r_i \in \bar{\mathbb{R}}$.

4. $|\zeta_i| = 1$ *and* $n^{1/2}\eta_{i,n}/(n-k)^{1/2} \to 2^{1/2}d_i$ *with* $0 < d_i < \infty$ *imply that* $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ *converges weakly to*

$$\left(\int_{-\infty}^{\infty} \Phi(d_i t + r_i)\phi(t)dt\right)\delta_{-\zeta_i} + \left(1 - \int_{-\infty}^{\infty} \Phi(d_i t + r_i)\phi(t)dt\right)\delta_0$$

provided $r_{i,n} \to r_i$ *for some* $r_i \in \bar{\mathbb{R}}$. *[Note that the above display reduces to* $\delta_{-\zeta_i}$ *if* $r_i = \infty$*, and to* δ_0 *if* $r_i = -\infty$*.]*

5. $|\zeta_i| = 1$ *and* $n^{1/2}\eta_{i,n}/(n-k)^{1/2} \to \infty$ *imply that* $H^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ *converges weakly to*

$$\Phi(r_i')\delta_{-\zeta_i} + (1-\Phi(r_i'))\,\delta_0$$

provided $(n^{1/2}\eta_{i,n}/(n-k)^{1/2})^{-1} r_{i,n} \to 2^{-1/2}r_i'$ *for some* $r_i' \in \bar{\mathbb{R}}$.

Theorem 37 (Soft-thresholding with consistent tuning)**.** *Suppose that for given* $i \geq 1$ *satisfying* $i \leq k = k(n)$ *for large enough* n *we have* $\xi_{i,n}\eta_{i,n} \to 0$ *and* $n^{1/2}\eta_{i,n} \to \infty$*. Set the scaling factor* $\alpha_{i,n} = (\xi_{i,n}\eta_{i,n})^{-1}$*. Suppose that the true parameters* $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ *and* $\sigma_n \in (0,\infty)$ *satisfy* $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$.

(a) If $n-k$ *is eventually constant equal to* m*, say, then* $H^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ *converges weakly to the distribution given by*

$$\begin{aligned}&\int_{|\zeta_i|}^{\infty} \rho_m(s)ds d\delta_{-\zeta_i}(x) + \{\rho_m(x)\mathbf{1}(x+\zeta_i<0) + \rho_m(-x)\mathbf{1}(x+\zeta_i>0)\}\,dx\\ &= \Pr(\chi^2_m > m\zeta_i^2)d\delta_{-\zeta_i}(x) + \{\rho_m(x)\mathbf{1}(x+\zeta_i<0) + \rho_m(-x)\mathbf{1}(x+\zeta_i>0)\}\,dx,\end{aligned} \tag{24}$$

where we recall the convention that $\rho_m(x) = 0$ *for* $x < 0$*. [In case* $|\zeta_i| = \infty$*, the atomic part in (24) is absent and (24) reduces to* $\rho_m(-\operatorname{sign}(\zeta_i)x)dx$*.]*

(b) If $n-k \to \infty$ *holds, then* $H^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ *converges weakly to* $\delta_{-\operatorname{sign}(\zeta_i)\min(1,|\zeta_i|)}$.

Theorem 38 (Adaptive soft-thresholding with consistent tuning)**.** *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to \infty$. Set the scaling factor $\alpha_{i,n} = (\xi_{i,n}\eta_{i,n})^{-1}$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$.*

(a) Suppose $n-k$ is eventually constant equal to m, say. Then $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to the distribution with cdf

$$\begin{aligned} &\int_{\sqrt{|x\zeta_i|}}^{\infty} \rho_m(s)ds\mathbf{1}\left(-\zeta_i \leq x < 0\right) + \mathbf{1}\left(x \geq 0\right) \\ = \quad & \Pr(\chi^2_m > m\,|x\zeta_i|)\mathbf{1}\left(-\zeta_i \leq x < 0\right) + \mathbf{1}\left(x \geq 0\right) \end{aligned}$$

in case $0 \leq \zeta_i < \infty$, and to the distribution with cdf

$$\begin{aligned} &\int_{0}^{\sqrt{|x\zeta_i|}} \rho_m(s)ds\mathbf{1}\left(0 \leq x < -\zeta_i\right) + \mathbf{1}\left(x \geq -\zeta_i\right) \\ = \quad & \Pr(\chi^2_m \leq m\,|x\zeta_i|)\mathbf{1}\left(0 \leq x < -\zeta_i\right) + \mathbf{1}\left(x \geq -\zeta_i\right) \end{aligned}$$

in case $-\infty < \zeta_i < 0$. Furthermore, $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to δ_0 if $|\zeta_i| = \infty$. [In case $|\zeta_i| < \infty$, the distribution has a jump of height $\int_{|\zeta_i|}^{\infty}\rho_m(s) = \Pr(\chi^2_m > m\zeta_i^2)$ at $x = -\zeta_i$ and is otherwise absolutely continuous. In particular, it reduces to δ_0 in case $\zeta_i = 0$.]

(b) If $n-k \to \infty$ holds, then

1. *$|\zeta_i| \leq 1$ implies that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to $\delta_{-\zeta_i}$,*
2. *$1 < |\zeta_i| < \infty$ implies that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to δ_{-1/ζ_i},*
3. *$|\zeta_i| = \infty$ implies that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to δ_0.*

We know from Theorem 30 that we obtain the same limiting distributions for $\tilde{\theta}_{H,i}$, $\tilde{\theta}_{S,i}$, and $\tilde{\theta}_{AS,i}$ as for $\hat{\theta}_{H,i}$, $\hat{\theta}_{S,i}$, and $\hat{\theta}_{AS,i}$, respectively, provided $n-k$ diverges to infinity sufficiently fast in the sense that $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$. The theorems in this section now show that for the soft-thresholding as well as for the adaptive soft-thresholding estimator we actually get the same limiting distribution as in the unknown-variance case whenever $n-k$ diverges even if $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ is violated. However, for the hard-thresholding estimator the picture is different, and in case $n-k$ diverges but $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ is violated, limit distributions different from the known-variance case arise (these limiting distributions still being convex combinations of two pointmasses, but with weights different from the known-variance case). It seems that this is a reflection of the fact that the hard-thresholding estimator is a discontinuous function of the data, whereas the other two estimators considered depend continuously on the data. The fixed-parameter limiting distributions for all three estimators are again the same as in the known-variance case.

In the case where the degrees of freedom $n-k$ are eventually constant, the limiting distribution of the hard-thresholding estimator is again a convex combination of two pointmasses, with weights that are in general different from the known-variance case. However, for the soft-thresholding as well as for the adaptive soft-thresholding estimator the limiting distributions can also contain an absolutely continuous component. This component seems to stem from an interaction of the more pronounced "bias-component" (as compared to hard-thresholding) with the nonvanishing randomness in the estimated variance. The fixed-parameter limiting distributions for hard-thresholding and adaptive soft-thresholding are again given by δ_0 for all values of θ_i as in the known-variance case, whereas for soft-thresholding the fixed-parameter limiting distribution is δ_0 only for $\theta_i = 0$ and otherwise has a pdf given by $\rho_m(-\operatorname{sign}(\theta_i)x)$ (as compared to a limit of $\delta_{-\operatorname{sign}(\theta_i)}$ in the known-variance case).

6.4. Consistent tuning: Some comments on fixed-parameter large-sample distributions and the "oracle-property"

6.4.1. Hard-thresholding and adaptive soft-thresholding

As already mentioned at the end of Sections 6.1 and 6.3.2, under consistent tuning the *fixed-parameter* limiting distributions of the hard-thresholding and of the adaptive soft-thresholding estimator – in the known-variance as well as in the unknown-variance case – always degenerate to pointmass at zero. Recall that in these results the estimators (after centering at θ_i) are scaled by $\sigma^{-1}(\xi_{i,n}\eta_{i,n})^{-1}$, which corresponds to the uniform convergence rate. We next show that if the estimators are scaled by $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}$ instead, a limit distribution under *fixed-parameter* asymptotics arises that is not degenerate in general (under an additional condition on the tuning parameter in case of adaptive soft-thresholding). In fact, we show that the hard-thresholding as well as the adaptive soft-thresholding estimators then satisfy what has been called the "oracle-property". However, it should be kept in mind that – with this faster scaling sequence $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}$ – the centered estimators are no longer stochastically bounded in a moving-parameter framework (for certain sequences of parameters), cf. Theorem 16. This shows the fragility of the "oracle-property", which is a fixed-parameter concept, and calls into question the statistical significance of this notion. For a more extensive discussion of the "oracle-property" and its consequences see Leeb and Pötscher (2008), Pötscher and Leeb (2009), and Pötscher and Schneider (2009).

Proposition 39. *Let $0 < \sigma < \infty$ be given. Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to \infty$.*

(a) $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}(\tilde{\theta}_{H,i} - \theta_i)$ as well as $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}(\hat{\theta}_{H,i} - \theta_i)$ converge in distribution to $N(0,1)$ when $\theta_i \neq 0$, and to $\delta_0 = N(0,0)$ when $\theta_i = 0$.

(b) $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}(\tilde{\theta}_{AS,i} - \theta_i)$ as well as $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}(\hat{\theta}_{AS,i} - \theta_i)$ converge in distribution to $N(0,1)$ when $\theta_i \neq 0$, and to $\delta_0 = N(0,0)$ when $\theta_i = 0$, provided the tuning parameter additionally satisfies $n^{1/4}\xi_{i,n}^{1/2}\eta_{i,n} \to 0$ for $n \to \infty$.

Remark 40. Inspection of the proof of Part (b) given in Section 8.4 shows that the condition $n^{1/4}\xi_{i,n}^{1/2}\eta_{i,n} \to 0$ is used for the result only in case $\theta_i \neq 0$. If now

$n^{1/4}\xi_{i,n}^{1/2}\eta_{i,n} \to \omega$ with $0 < \omega < \infty$, inspection of the proof shows that then in case $\theta_i \neq 0$ we have that $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}(\tilde{\theta}_{AS,i} - \theta_i) = Z_n - \sigma\omega^2\theta_i^{-1}(\hat{\sigma}/\sigma)^2 + o_p(1)$, where Z_n is standard normal and is independent of $\hat{\sigma}/\sigma$. Hence, we see that the distribution of $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}(\tilde{\theta}_{AS,i} - \theta_i)$ asymptotically behaves like the convolution of an $N(0,1)$-distribution and the distribution of $-\sigma\omega^2\theta_i^{-1}(n-k)^{-1}$ times a chi-square distributed random variable with $n-k$ degrees of freedom (if $n-k \to \infty$ this reduces to an $N(-\sigma\omega^2\theta_i^{-1}, 1)$-distribution). If $n^{1/4}\xi_{i,n}^{1/2}\eta_{i,n} \to \infty$, then $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}(\tilde{\theta}_{AS,i} - \theta_i)$ is stochastically unbounded. Note that this shows that the consistently tuned adaptive soft-thresholding estimator – even in a fixed-parameter setting – has a convergence rate slower than $n^{1/2}\xi_{i,n}^{-1}$ if $\theta_i \neq 0$ and if the tuning parameter is "too large" in the sense that $n^{1/4}\xi_{i,n}^{1/2}\eta_{i,n} \to \infty$. The same conclusion applies to the infeasible estimator $\hat{\theta}_{AS,i}$ (with the simplification that one always obtains an $N(-\sigma\omega^2\theta_i^{-1}, 1)$-distribution in case $n^{1/4}\xi_{i,n}^{1/2}\eta_{i,n} \to \omega$ with $0 < \omega < \infty$).

We further illustrate the fragility of the fixed-parameter asymptotic results under a $\sigma^{-1}n^{1/2}\xi_{i,n}^{-1}$-scaling obtained above by providing the moving-parameter limits under this scaling. Let $F^i_{H,n,\theta,\sigma} := F^i_{H,\eta_{i,n},n,\theta,\sigma}$ denote the cdf of $\sigma^{-1}\times n^{1/2}\xi_{i,n}^{-1}(\hat{\theta}_{H,i} - \theta_i)$, and define $F^i_{S,n,\theta,\sigma}$ and $F^i_{AS,n,\theta,\sigma}$ analogously. The proofs of the subsequent propositions are completely analogous to the proofs of Theorem 9 in Pötscher and Leeb (2009) and Theorem 5 in Pötscher and Schneider (2009), respectively.

Proposition 41 (Hard-thresholding)**.** *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to \infty$. Suppose that the true parameters $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ and $\sigma_n \in (0,\infty)$ satisfy $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$ and $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$. [Note that in case $\zeta_i \neq 0$ the convergence of $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})$ already follows from that of $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})$, and ν_i is then given by* $\mathrm{sign}(\zeta_i)\infty$*.]*

1. Suppose $|\zeta_i| < 1$. Then $F^i_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to $\delta_{-\nu_i}$ if $|\nu_i| < \infty$; if $|\nu_i| = \infty$ the total mass of $F^i_{H,n,\theta^{(n)},\sigma_n}$ escapes to $-\nu_i$, in the sense that $F^i_{H,n,\theta^{(n)},\sigma_n}(x) \to 0$ for every $x \in \mathbb{R}$ if $\nu_i = -\infty$, and that $F^i_{H,n,\theta^{(n)},\sigma_n}(x) \to 1$ for every $x \in \mathbb{R}$ if $\nu_i = \infty$.

2. Suppose $|\zeta_i| > 1$. Then $F^i_{H,n,\theta^{(n)},\sigma_n}$ converges weakly to Φ.

3. Suppose $|\zeta_i| = 1$ and $n^{1/2}(\eta_{i,n} - \zeta_i\theta_{i,n}/(\sigma_n\xi_{i,n})) \to r_i$ for some $r_i \in \bar{\mathbb{R}}$. Then $F^i_{H,n,\theta^{(n)},\sigma_n}(x)$ converges to

$$\Phi(r_i)\mathbf{1}(\zeta_i = 1) + \int_{-\infty}^{x} \phi(t)\mathbf{1}(\zeta_i t > r_i)\, dt$$

for every $x \in \mathbb{R}$. [In case $r_i = -\infty$ the limit reduces to a standard normal distribution.]

Proposition 42 (Adaptive soft-thresholding)**.** *Suppose that for given $i \geq 1$ satisfying $i \leq k = k(n)$ for large enough n we have $\xi_{i,n}\eta_{i,n} \to 0$ and $n^{1/2}\eta_{i,n} \to$*

∞. *Suppose that the true parameters* $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ *and* $\sigma_n \in (0,\infty)$ *satisfy* $\theta_{i,n}/(\sigma_n \xi_{i,n} \eta_{i,n}) \to \zeta_i \in \bar{\mathbb{R}}$.

1. If $\zeta_i = 0$ *and* $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to \nu_i \in \mathbb{R}$, *then* $F^i_{AS,n,\theta^{(n)},\sigma_n}$ *converges weakly to* $\delta_{-\nu_i}$.

2. The total mass of $F^i_{AS,n,\theta^{(n)},\sigma_n}$ *escapes to* ∞ *or* $-\infty$ *in the following cases: If* $-\infty < \zeta_i < 0$, *or if* $\zeta_i = 0$ *and* $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to -\infty$, *or if* $\zeta_i = -\infty$ *and* $n^{1/2}\eta^2_{i,n}\xi_{i,n}\theta^{-1}_{i,n}\sigma_n \to -\infty$, *then* $F^i_{AS,n,\theta^{(n)},\sigma_n}(x) \to 0$ *for every* $x \in \mathbb{R}$. *If* $0 < \zeta_i < \infty$, *or if* $\zeta_i = 0$ *and* $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to \infty$, *or if* $\zeta_i = \infty$ *and* $n^{1/2}\eta^2_{i,n}\xi_{i,n}\theta^{-1}_{i,n}\sigma_n \to \infty$, *then* $F^i_{AS,n,\theta^{(n)},\sigma_n}(x) \to 1$ *for every* $x \in \mathbb{R}$.

3. If $|\zeta_i| = \infty$ *and* $n^{1/2}\eta^2_{i,n}\xi_{i,n}\theta^{-1}_{i,n}\sigma_n \to w_i \in \mathbb{R}$, *then* $F^i_{AS,n,\theta^{(n)},\sigma_n}$ *converges weakly to* $\Phi(\cdot + w_i)$.

It is easy to see that setting $\theta_{i,n} \equiv \theta_i$ and $\sigma_n \equiv \sigma$ in Proposition 41 immediately recovers the "oracle-property" for $\hat{\theta}_{H,i}$. Similarly, we recover the "oracle property" for $\hat{\theta}_{AS,i}$ from Proposition 42 provided $n^{1/4}\xi^{1/2}_{i,n}\eta_{i,n} \to 0$. The propositions also characterize the sequences of parameters along which the mass of the distributions of the hard-thresholding and the adaptive soft-thresholding estimator escapes to infinity; loosely speaking these are sequences along which the bias of the estimators exceeds all bounds.

The theorems in Section 6.2 also show that the last two propositions above carry over immediately to the unknown-variance case whenever $n - k \to \infty$ sufficiently fast such that $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ holds. To save space, we do not extend these two propositions to the case where the latter condition fails to hold.

6.4.2. Soft-thresholding

The situation is somewhat different for the soft-thresholding estimator. It follows from Theorem 37 that the distribution of $\sigma^{-1}(\xi_{i,n}\eta_{i,n})^{-1}(\tilde{\theta}_{S,i} - \theta_i)$ does not degenerate to pointmass at zero (in fact, has no mass at zero) if $\theta_i \neq 0$ and is held fixed. Consequently, $(\xi_{i,n}\eta_{i,n})^{-1}$ is also the fixed-parameter convergence rate of $\tilde{\theta}_{S,i}$, in the sense that scaling with a faster rate (e.g., $n^{1/2}\xi^{-1}_{i,n}$) leads to the escape of the total mass of the finite-sample distribution of the so-scaled (and centered) estimator to $-\operatorname{sign}(\theta_i)\infty$. For $\theta_i = 0$ we get with the same argument as for hard-thresholding that $\sigma^{-1}n^{1/2}\xi^{-1}_{i,n}(\tilde{\theta}_{S,i} - \theta_i)$ converges to δ_0. For the infeasible version $\hat{\theta}_{S,i}$ the situation is identical. We conclude by a result analogous to Propositions 41 and 42. The proof of this result is completely analogous to the proof of Theorem 10 in Pötscher and Leeb (2009).

Proposition 43 (Soft-thresholding)**.** *Suppose that for given* $i \geq 1$ *satisfying* $i \leq k = k(n)$ *for large enough* n *we have* $\xi_{i,n}\eta_{i,n} \to 0$ *and* $n^{1/2}\eta_{i,n} \to \infty$. *Suppose that the true parameters* $\theta^{(n)} = (\theta_{1,n}, \ldots, \theta_{k_n,n}) \in \mathbb{R}^{k_n}$ *and* $\sigma_n \in (0,\infty)$ *satisfy* $n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) \to \nu_i \in \bar{\mathbb{R}}$. *Then* $F^i_{S,n,\theta^{(n)},\sigma_n}$ *converges weakly to* $\delta_{-\nu_i}$ *if* $|\nu_i| < \infty$; *and if* $|\nu_i| = \infty$, *the total mass of* $F^i_{S,n,\theta^{(n)},\sigma_n}$ *escapes to* $-\nu_i$, *in the sense that* $F^i_{S,n,\theta^{(n)},\sigma_n}(x) \to 0$ *for every* $x \in \mathbb{R}$ *if* $\nu_i = -\infty$, *and that* $F^i_{S,n,\theta^{(n)},\sigma_n}(x) \to 1$ *for every* $x \in \mathbb{R}$ *if* $\nu_i = \infty$.

Again, this proposition immediately extends to the unknown-variance case whenever $n-k\to\infty$ sufficiently fast such that $n^{1/2}\eta_{i,n}(n-k)^{-1/2}\to 0$ holds. We abstain from extending the result to the case where the latter condition fails to hold.

6.5. Remarks

Remark 44. (i) The convergence conditions on the various quantities involving $\theta_{i,n}$ and σ_n (and on $n-k$) in the propositions in Sections 6.1 and 6.4 as well as in the theorems in Section 6.3 are essentially cost-free for the same reason as explained in Remark 12.

(ii) We note that all possible forms of the moving-parameter limiting distributions in the results in this section already arise for sequences $\theta_{i,n}$ belonging to an arbitrarily small neighborhood of zero (and with $\sigma>0$ fixed). Consequently, the non-uniformity in the convergence to the fixed-parameter limits is of a local nature.

Remark 45. Pötscher and Leeb (2009) and Pötscher and Schneider (2009) present impossibility results for estimating the finite-sample distribution of the thresholding estimators considered in these papers. In the present context, corresponding impossibility results could be derived under appropriate assumptions. We abstain from presenting such results.

7. Numerical study

As has been discussed in Remarks 1 and 2 in Section 2, the soft-thresholding estimator coincides with the Lasso, and the adaptive soft-thresholding estimator coincides with the adaptive Lasso in case of orthogonal design. A natural question now is if the distributional results for the (adaptive) soft-thresholding estimator derived in this paper are in any way indicative for the distribution of the (adaptive) Lasso in case of non-orthogonal design. In order to gain some insight into this we provide a simulation study to compare the finite-sample distributions of the respective estimators.

We simulate the Lasso estimator as defined in Remark 1 (with $\eta_{i,n}'=\eta_{i,n}\xi_{i,n}^{-1}$ and $\eta_{i,n}=\eta_n$ not depending on i) and the adaptive Lasso estimator as defined in Remark 2 (with $\eta_{i,n}'=\eta_n$ not depending on i) and show histograms of $n^{1/2}\sigma^{-1}\xi_{i,n}^{-1}\left(\bar{\theta}_i-\theta_i\right)$ where $\bar{\theta}_i$ stands for the i-th component of Lasso or adaptive Lasso. [The scaling used here is chosen on the basis that with this scaling the i-th component of the least-squares estimator is standard normally distributed.]

We set $n=8$ and $k=4$, resulting in $n-k=4$ degrees of freedom. Two different types of designs are considered: for Design I we use $X'X=n\Omega(\rho)$ with $\Omega(\rho)_{i,j}=\rho^{|i-j|}$. More concretely, X is partitioned into $d=n/k=2$ blocks of size $k\times k$ and each of these blocks is set equal to $k^{1/2}L$ with $LL'=\Omega(\rho)$, the Cholesky factorization of $\Omega(\rho)$. The value of ρ is set equal to 0.3, 0.5, and 0.9, implying condition numbers for $X'X$ of 2.7, 5.6, and 57.0, respectively. Design

II is an "equicorrelated" design. Here we set the matrix comprised of the first k rows of X equal to $I_k + cE_k$, where E_k is the $k \times k$ matrix with all components equal to 1 and c is a real number greater than $-1/k = -0.25$. The remaining entries of X are all set equal to 0. We choose three values for c: first, $c = 0.2$ which implies a correlation of 0.36 between any two regressors and a condition number of 3.2 for $X'X$; second, $c = 2$ which implies a correlation of 0.952 and a condition number of 81; and $c = -0.2$ which implies a correlation of -0.32 and a condition number of 25. For either type of design we proceed as follows: For the given parameters $\theta = (3, 1.5, 0, 0)'$ and $\sigma = 1$, we simulate $10,000$ data vectors Y and compute the corresponding estimator, i.e., the Lasso and adaptive Lasso as specified above. We set $\eta_n = n^{-1/2}\Phi^{-1}(0.975)$, implying that the thresholding estimators delete a given irrelevant variable with probability 0.95.

For the non-zero outcomes of the estimators, we plot the histogram of $n^{1/2}\sigma^{-1}\xi_{i,n}^{-1}(\bar{\theta}_i - \theta_i)$ which is normalized such that its mass corresponds to the proportion of the non-zero values. The zero values are accounted for by plotting "pointmass" with height representing the proportion of zero values, i.e., the simulated variable selection probability. For the purpose of comparison the graph of the distribution of the corresponding (centered and scaled) thresholding estimator (using the same $\eta_{i,n} = \eta_n$) as derived analytically in Section 5 is then superimposed in red color. The results of the simulation study are presented in Figures 1–12 below.

In comparing the adaptive Lasso with the adaptive soft-thresholding estimator, we find remarkable agreement between the respective marginal distributions in all cases where the design matrix is not too multicollinear, see Figures 1, 2, and 4. For the cases where the design matrix is no longer well-conditioned a difference between the respective marginal distributions emerges but seems to be surprisingly moderate, see Figures 3, 5, and 6.

Turning to the Lasso and its thresholding counterpart, we find a similar situation with a somewhat stronger disagreement between the respective marginal distributions. Again in the cases where the design matrix is well-conditioned (Figures 7, 8, and 10) the difference is less pronounced than in the case of an ill-conditioned design matrix (Figures 9, 11, and 12).

We have also experimented with other values of n, k, θ, ρ, c, and η_n and have found the results to be qualitatively the same for these choices.

8. Proofs

8.1. Proofs for Section 3

Proof of Proposition 4. We first prove Part (a). Rewrite $P_{n,\theta,\sigma}(\hat{\theta}_i = 0)$ as

$$\Phi\left(n^{1/2}\xi_{i,n}^{-1}\left(-\theta_i/\sigma + \xi_{i,n}\eta_{i,n}\right)\right) - \Phi\left(n^{1/2}\xi_{i,n}^{-1}\left(-\theta_i/\sigma - \xi_{i,n}\eta_{i,n}\right)\right). \tag{25}$$

Assume first that $\xi_{i,n}\eta_{i,n} \to 0$ and fix $\theta_i \neq 0$. By a standard subsequence argument we may assume without loss of generality that $n^{1/2}\xi_{i,n}^{-1}$ converges

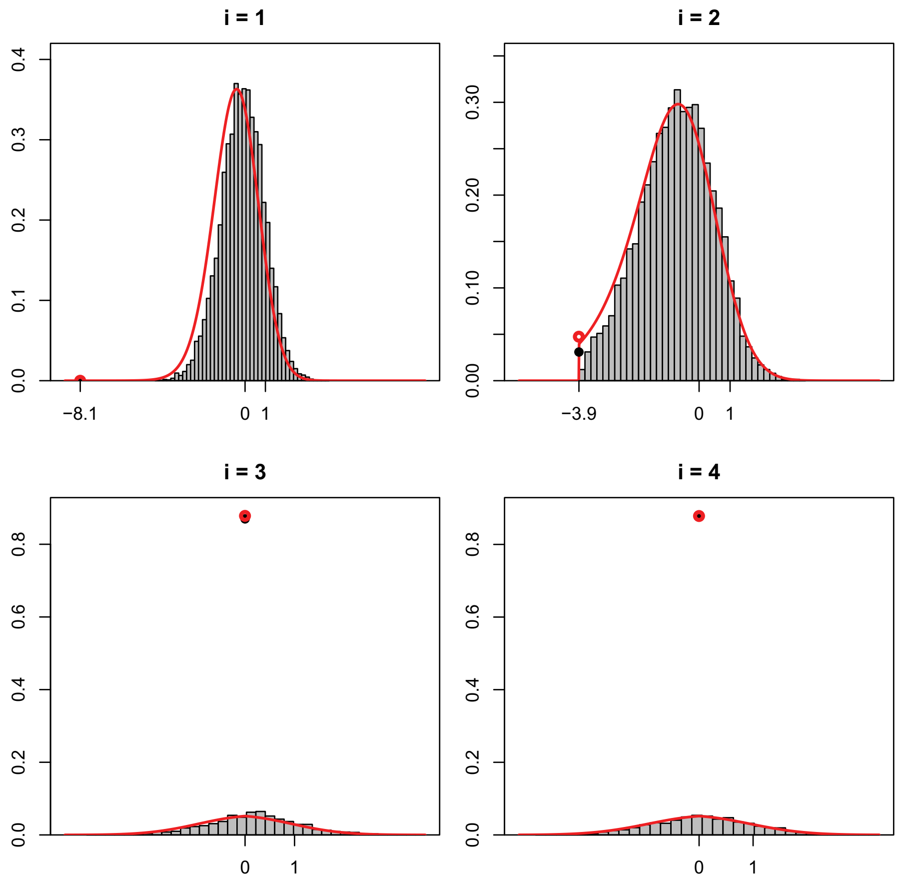

FIG 1. *Adaptive Lasso, Design I:* $\rho = 0.3$.

to a constant κ which by our maintained assumption (3) must satisfy $0 < \kappa \leq \infty$. Now $-\theta_i/\sigma \pm \xi_{i,n}\eta_{i,n}$ both converge to $-\theta_i/\sigma$, which is non-zero, and consequently both arguments in (25) converge to $-\kappa\theta_i/\sigma$. Since Φ is continuous on $\bar{\mathbb{R}}$, the expression (25) converges to zero. To prove the converse, now assume that (25) converges to zero for all $\theta_i \neq 0$. By a standard subsequence argument, we may assume without loss of generality that $\xi_{i,n}\eta_{i,n}$ converges to a constant $\varkappa$ satisfying $0 \leq \varkappa \leq \infty$. Suppose $\varkappa > 0$ holds. Choose θ_i such that $0 < -\theta_i/\sigma < \varkappa$ holds. It follows that $-\theta_i/\sigma + \xi_{i,n}\eta_{i,n}$ and $-\theta_i/\sigma - \xi_{i,n}\eta_{i,n}$ eventually have opposite signs and are bounded away from zero. By our maintained assumption (3), the same is then true for the arguments in (25) leading to a contradiction. Hence $\varkappa = 0$ must hold, completing the proof of Part (a). Parts (b) and (c) are obvious since $P_{n,\theta,\sigma}(\hat{\theta}_i = 0) = \Phi(n^{1/2}\eta_{i,n}) - \Phi(-n^{1/2}\eta_{i,n})$ whenever $\theta_i = 0$. $\square$

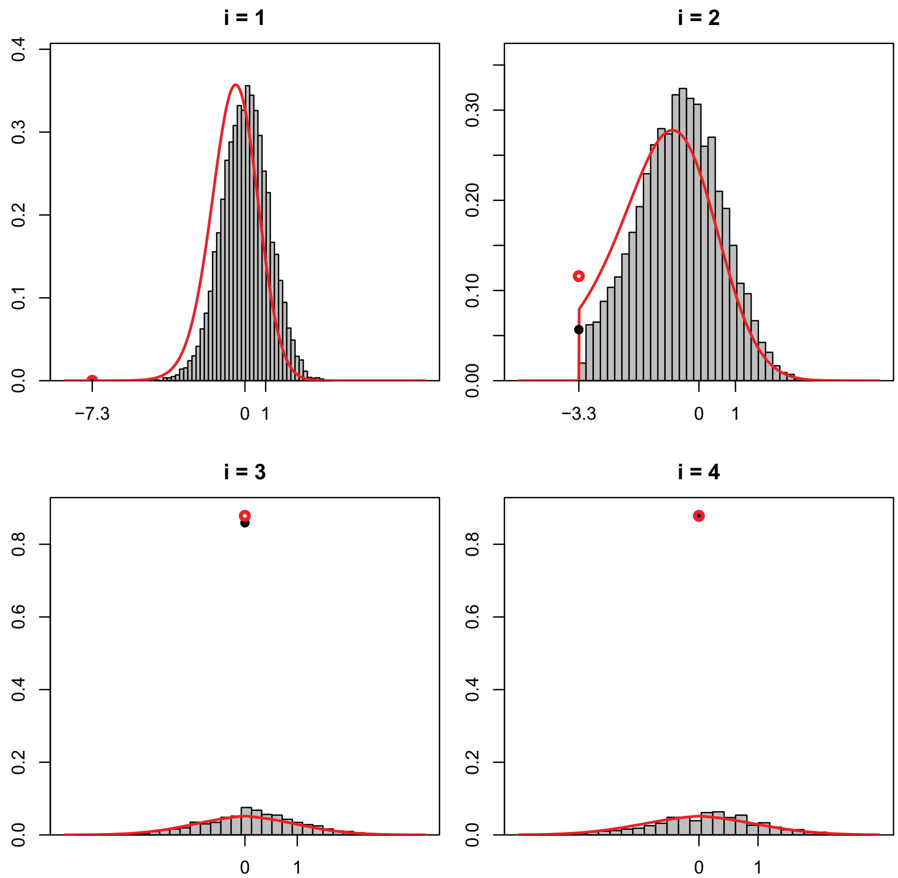

FIG 2. *Adaptive Lasso, Design I:* $\rho = 0.5$.

Proof of Proposition 8. Part (a) follows immediately from (4) and the assumptions. To prove Part (b) we use (4) to write

$$P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i = 0\right) = \Phi\left(n^{1/2}\eta_{i,n}\left(1 - \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\right)\right) - \Phi\left(n^{1/2}\eta_{i,n}\left(-1 - \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\right)\right).$$

The first and the second claim then follow immediately. For the third claim, assume first that $\zeta_i = 1$. Then

$$P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i = 0\right) = \Phi\left(n^{1/2}\left(\eta_{i,n} - \zeta_i\theta_{i,n}/(\sigma_n\xi_{i,n})\right)\right) - \Phi\left(n^{1/2}\eta_{i,n}\left(-1 - \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\right)\right) \to \Phi(r_i).$$

The case $\zeta_i = -1$ is handled analogously. $\square$

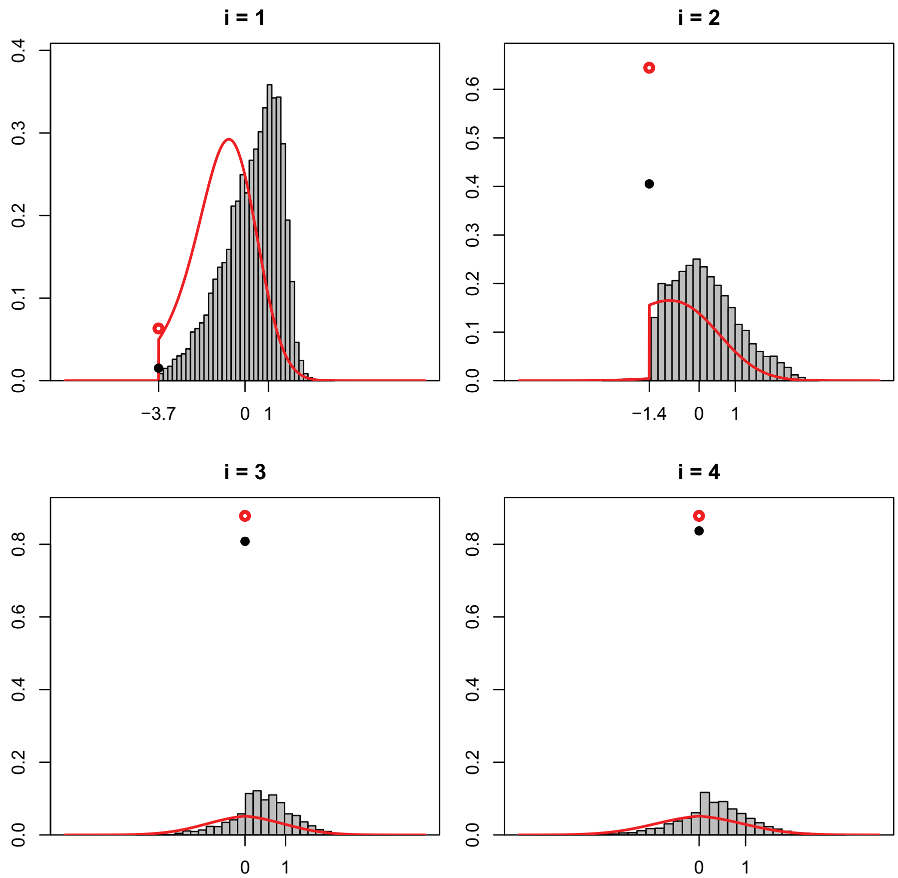

FIG 3. *Adaptive Lasso, Design I:* $\rho = 0.9$.

Proof of Proposition 10. We prove Part (b) first. Observe that

$$\begin{aligned} P_{n,\theta,\sigma}\left(\tilde{\theta}_i = 0\right) &= \int_0^\infty \left[\Phi\left(n^{1/2} s \eta_{i,n}\right) - \Phi\left(-n^{1/2} s \eta_{i,n}\right)\right] \rho_{n-k}(s) ds \\ &= T_{n-k}\left(n^{1/2}\eta_{i,n}\right) - T_{n-k}\left(-n^{1/2}\eta_{i,n}\right). \end{aligned}$$

By a subsequence argument it suffices to prove the result under the assumption that $n - k = n - k(n)$ converges in $\mathbb{N} \cup \{\infty\}$. If the limit is finite, then $n - k(n)$ is eventually constant and the result follows since every t-distribution has unbounded support. If $n - k \to \infty$ then

$$\begin{aligned} & \Phi\left(n^{1/2}\eta_{i,n}\right) - \Phi\left(-n^{1/2}\eta_{i,n}\right) - 2\left\|T_{n-k} - \Phi\right\|_\infty \\ \leq\ & P_{n,\theta,\sigma}\left(\tilde{\theta}_i = 0\right) \leq \Phi\left(n^{1/2}\eta_{i,n}\right) - \Phi\left(-n^{1/2}\eta_{i,n}\right) + 2\left\|T_{n-k} - \Phi\right\|_\infty, \end{aligned}$$

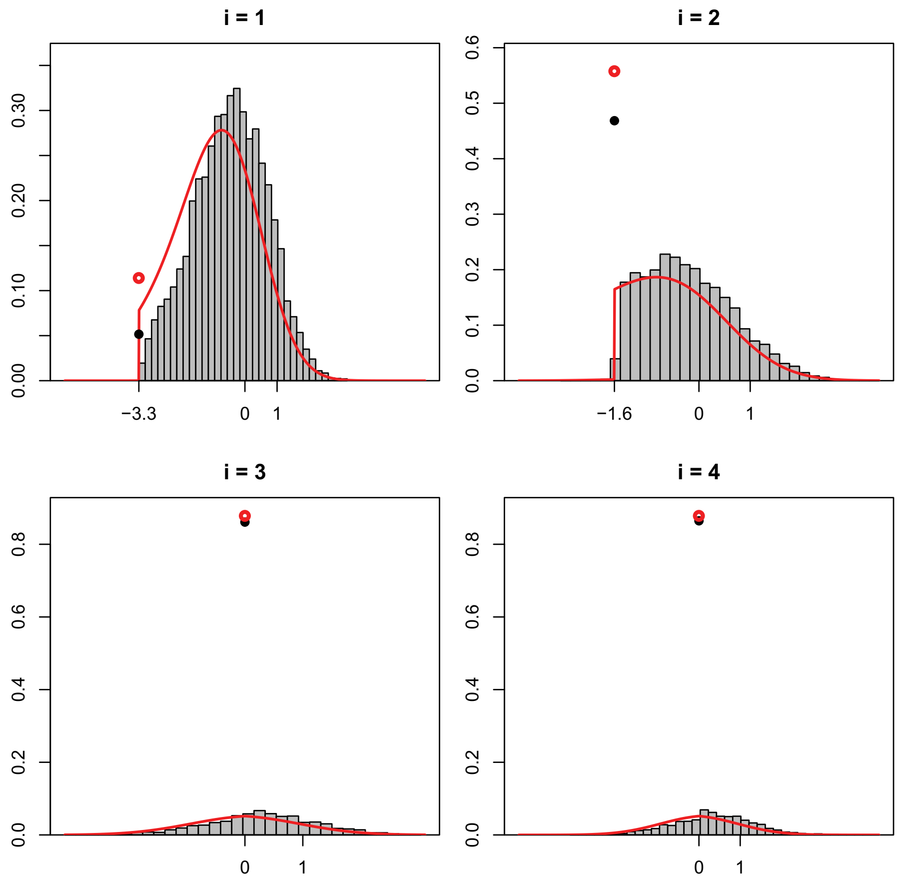

FIG 4. *Adaptive Lasso, Design II:* $c = 0.2$.

where $\|\cdot\|_\infty$ denotes the supremum norm. Since $\|T_{n-k} - \Phi\|_\infty \to 0$ if $n-k \to \infty$ by Polya's Theorem, the result follows. Part (c) is proved analogously.

We next prove Part (a). Observe that the collection of distributions corresponding to $\{\rho_m : m \in \mathbb{N}\}$ is tight on $(0, \infty)$, meaning that for every $0 < \delta < 1$ there exist $0 < c_*(\delta) < c^*(\delta) < \infty$ such that $\sup_{m\in\mathbb{N}} \int_0^{c_*(\delta)} \rho_m ds < \delta$ and $\sup_{m\in\mathbb{N}} \int_{c^*(\delta)}^\infty \rho_m ds < \delta$. Note that the map $s \mapsto P_{n,\theta,\sigma}(\hat{\theta}_i(s\eta_{i,n}) = 0)$ is monotonically nondecreasing. Hence,

$$
\begin{aligned}
(1-\delta)P_{n,\theta,\sigma}\left(\hat{\theta}_i(c_*(\delta)\eta_{i,n}) = 0\right) &\leq \int_{c_*(\delta)}^\infty P_{n,\theta,\sigma}\left(\hat{\theta}_i(s\eta_{i,n}) = 0\right)\rho_{n-k}(s)ds \\
&\leq P_{n,\theta,\sigma}\left(\tilde{\theta}_i = 0\right) = \int_0^\infty P_{n,\theta,\sigma}\left(\hat{\theta}_i(s\eta_{i,n}) = 0\right)\rho_{n-k}(s)ds \\
&\leq P_{n,\theta,\sigma}\left(\hat{\theta}_i(c^*(\delta)\eta_{i,n}) = 0\right) + \delta.
\end{aligned}
$$

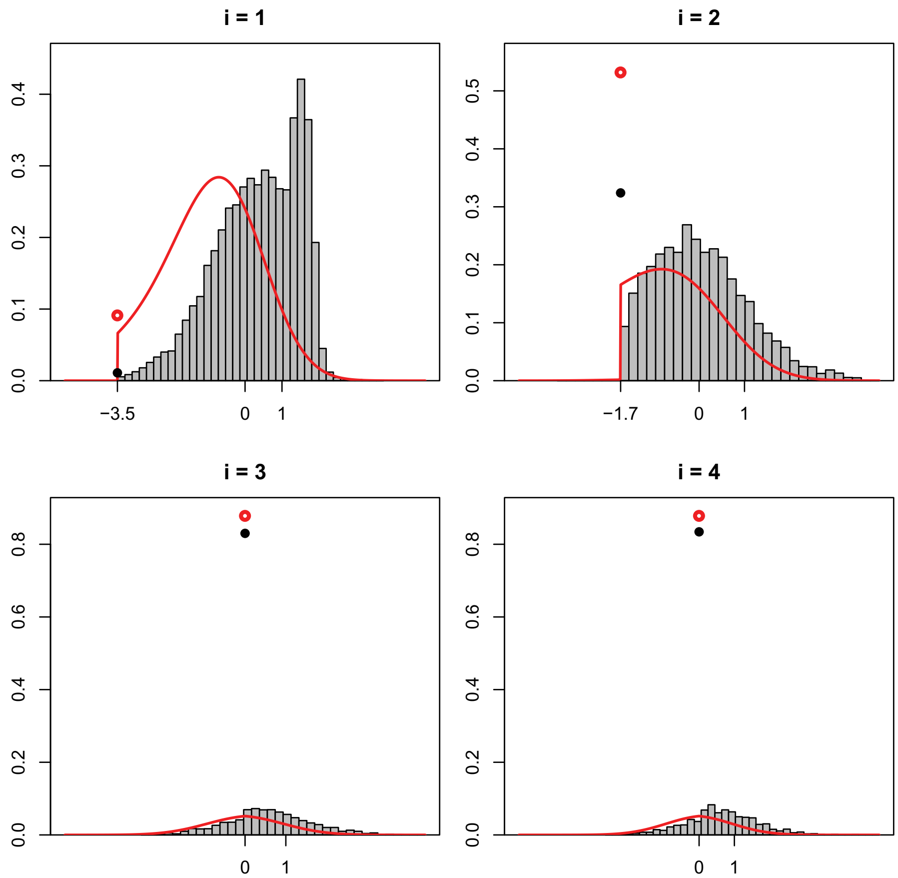

Fig 5. *Adaptive Lasso, Design II: $c = 2$.*

Since $\xi_{i,n} c_*(\delta)\eta_{i,n}$ ($\xi_{i,n} c^*(\delta)\eta_{i,n}$, respectively) converges to zero if and only if $\xi_{i,n}\eta_{i,n}$ does so, Part (a) follows from Proposition 4 applied to the estimators $\hat{\theta}_i(c_*(\delta)\eta_{i,n})$and $\hat{\theta}_i(c^*(\delta)\eta_{i,n})$. □

Proof of Theorem 11. (a) Set $P_n(s) = P_{n,\theta^{(n)},\sigma_n}(\hat{\theta}_i(s\eta_{i,n}) = 0)$ for $s > 0$. By Proposition 8 we have that $P_n(s)$ converges to $P(s)$ for all $s > 0$, where $P(s) = \Phi(-\nu_i + se_i) - \Phi(-\nu_i - se_i)$ for $s > 0$. Since $P_n(s)$ as well as $P(s)$ are continuous functions of s, are monotonically nondecreasing in s, and have the property that their limits for $s \to 0$ are 0 while the limits for $s \to \infty$ are 1, it follows from Polya's Theorem that the convergence is uniform in s. But then using (5) gives

$$
\begin{aligned}
& \left| P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i = 0\right) - \int_0^\infty \left(\Phi(-\nu_i + se_i) - \Phi(-\nu_i - se_i)\right) \rho_{n-k}(s) ds \right| \\
\leq\ & \sup_{s>0} |P_n(s) - P(s)| \int_0^\infty \rho_{n-k}(s) ds = \sup_{s>0} |P_n(s) - P(s)| \to 0
\end{aligned}
$$

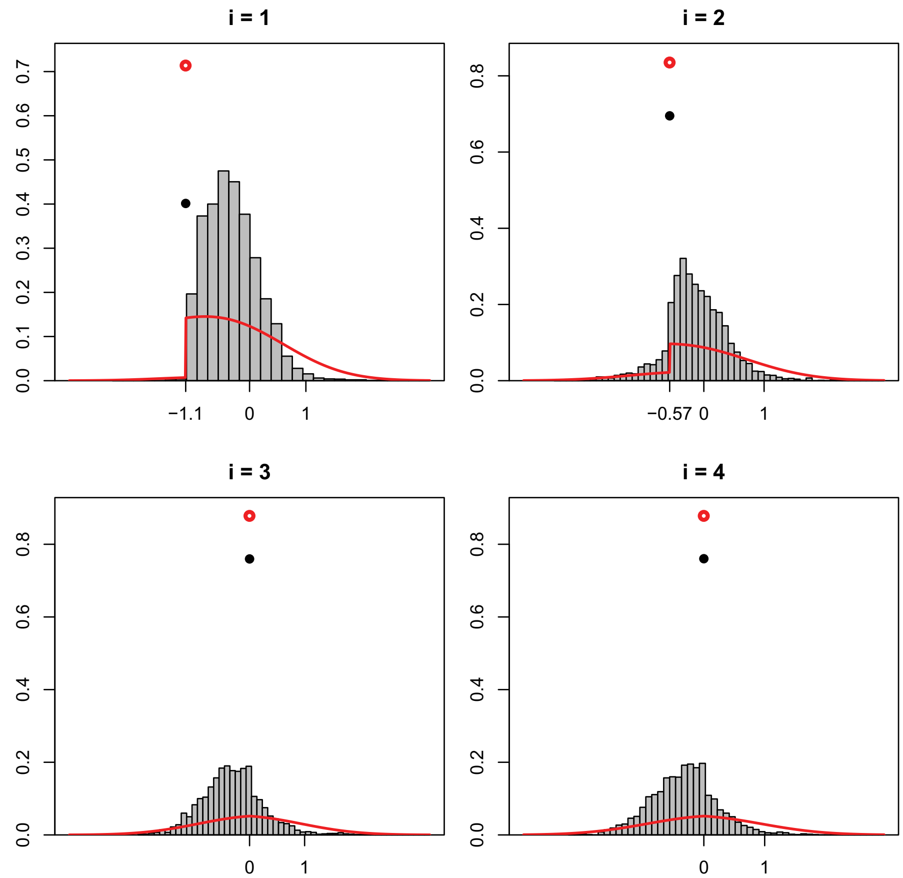

FIG 6. *Adaptive Lasso, Design II:* $c = -0.2$.

as $n \to \infty$. This completes the proof in case $n-k = m$ eventually; in case $n-k \to \infty$ observe that $\int_0^\infty (\Phi(-\nu_i + se_i) - \Phi(-\nu_i - se_i))\, \rho_{n-k}(s)ds$ then converges to $\Phi(-\nu_i + e_i) - \Phi(-\nu_i - e_i)$ as the distribution corresponding to ρ_{n-k} converges weakly to pointmass at $s = 1$ and the integrand is bounded and continuous.

(b) Observe that $P_{n,\theta^{(n)},\sigma_n}(\hat{\theta}_i(s\eta_{i,n}) = 0)$ converges to 1 for $s > |\zeta_i|$ and to 0 for $s < |\zeta_i|$ by Proposition 8 applied to the estimator $\hat{\theta}_i(s\eta_{i,n})$. Now (5) and dominated convergence deliver the result in (b1).

Next consider (b2): Suppose first that $|\zeta_i| < 1$. Choose $\varepsilon > 0$ small enough such that $|\zeta_i| + \varepsilon < 1$. Then, recalling that $P_{n,\theta^{(n)},\sigma_n}(\hat{\theta}_i(s\eta_{i,n}) = 0)$ is monotonically nondecreasing in s, eq. (5) gives

$$\begin{aligned} P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i = 0\right) &\geq \int_{|\zeta_i|+\varepsilon}^{\infty} P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i(s\eta_{i,n}) = 0\right) \rho_{n-k}(s)ds \\ &\geq P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i((|\zeta_i| + \varepsilon)\,\eta_{i,n}) = 0\right) \int_{|\zeta_i|+\varepsilon}^{\infty} \rho_{n-k}(s)ds. \end{aligned}$$

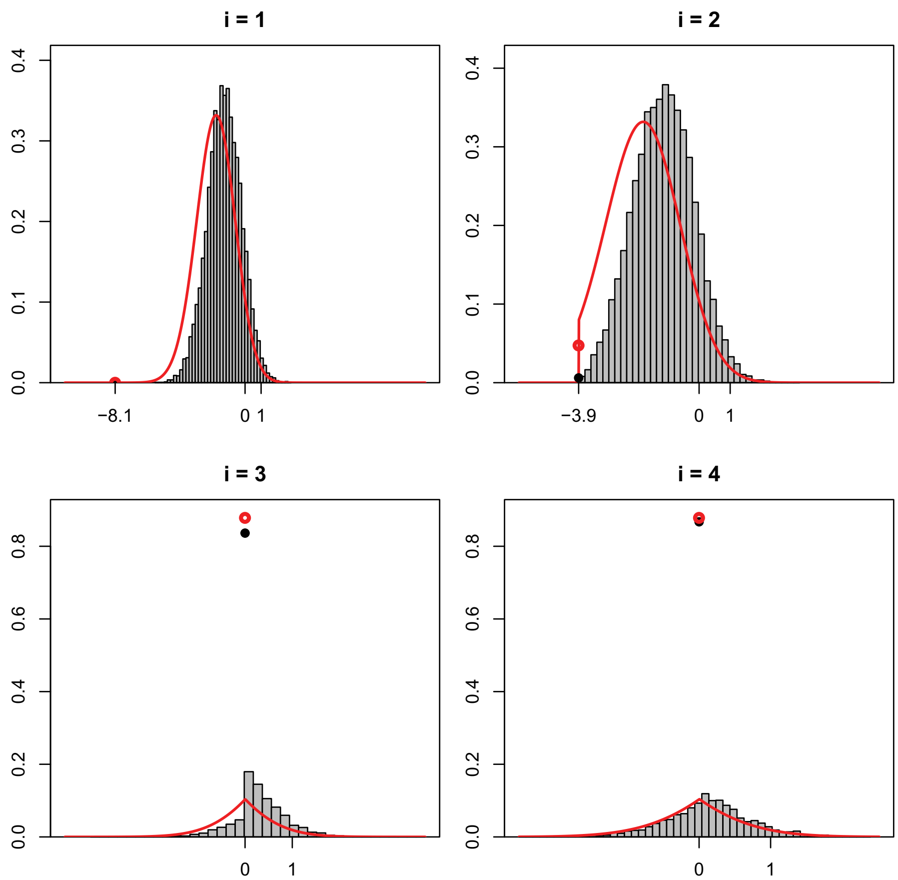

FIG 7. *Lasso, Design I:* $\rho = 0.3$.

Now the integral on the r.h.s. converges to 1 since $|\zeta_i| + \varepsilon < 1$, and the probability on the r.h.s. converges to 1 by Proposition 8 applied to the estimator $\hat{\theta}_i((|\zeta_i| + \varepsilon)\,\eta_{i,n})$. This completes the proof for the case $|\zeta_i| < 1$. Next assume that $|\zeta_i| > 1$. Choose $\varepsilon > 0$ small enough such that $|\zeta_i| - \varepsilon > 1$ holds. Then from (5) we have

$$\begin{aligned} P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i = 0\right) &\leq \int_0^{|\zeta_i|-\varepsilon} P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i(s\eta_{i,n}) = 0\right)\rho_{n-k}(s)ds \\ &\quad + \int_{|\zeta_i|-\varepsilon}^{\infty} \rho_{n-k}(s)ds \\ &\leq P_{n,\theta^{(n)},\sigma_n}\left(\hat{\theta}_i((|\zeta_i| - \varepsilon)\,\eta_{i,n}) = 0\right) + \int_{|\zeta_i|-\varepsilon}^{\infty} \rho_{n-k}(s)ds \end{aligned}$$

since $P_n(s)$ is monotonically nondecreasing in s and $\int_0^{|\zeta_i|-\varepsilon} \rho_{n-k}(s)ds$ is not larger than 1. Since $|\zeta_i| - \varepsilon > 1$ and $n - k \to \infty$ the second term on the r.h.s. goes to zero, while the first term goes to zero by Proposition 8 applied to the estimator $\hat{\theta}_i((|\zeta_i| - \varepsilon)\,\eta_{i,n})$.

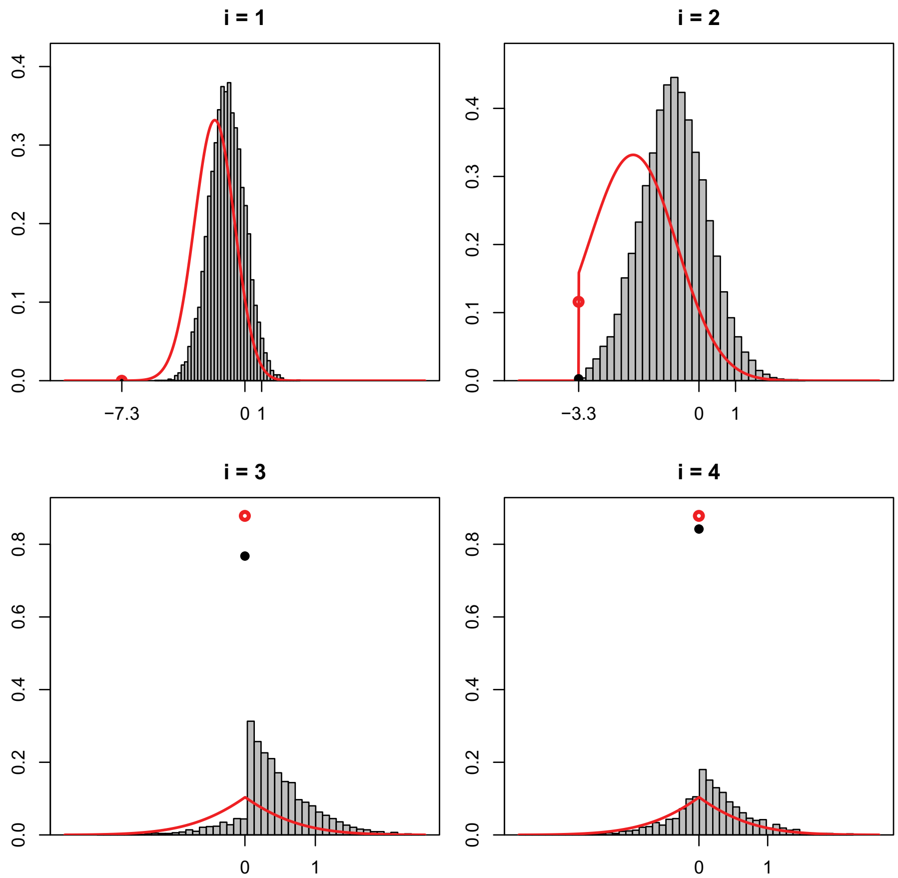

FIG 8. *Lasso, Design I:* $\rho = 0.5$.

Next we prove 3. and 4. We assume $\zeta_i = 1$ first. Then using eq. (5) and performing the substitution $s-1 = (2(n-k))^{-1/2} t$ we obtain the following expression (recalling that ρ_{n-k} is zero for negative arguments and using the abbreviations $r_{i,n} = n^{1/2}(\eta_{i,n} - \theta_{i,n}/(\sigma_n \xi_{i,n}))$ and $r^*_{i,n} = n^{1/2}(-\eta_{i,n} - \theta_{i,n}/(\sigma_n \xi_{i,n}))$).

$$\begin{aligned} P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i = 0\right) &= \int_{-\infty}^{\infty} \left[\Phi\left(r_{i,n} + n^{1/2}\eta_{i,n}\left(2(n-k)\right)^{-1/2} t\right)\right. \\ &\quad \left. - \Phi\left(r^*_{i,n} - n^{1/2}\eta_{i,n}\left(2(n-k)\right)^{-1/2} t\right)\right] \\ &\quad \times \left(2(n-k)\right)^{-1/2} \rho_{n-k}((2(n-k))^{-1/2} t + 1) dt \\ &= \int_{-\infty}^{\infty} \left[\Phi\left(r_{i,n} + n^{1/2}\eta_{i,n}\left(2(n-k)\right)^{-1/2} t\right)\right. \\ &\quad \left. -\Phi\left(r^*_{i,n} - n^{1/2}\eta_{i,n}\left(2(n-k)\right)^{-1/2} t\right)\right] \phi(t) dt + o(1). \end{aligned}$$

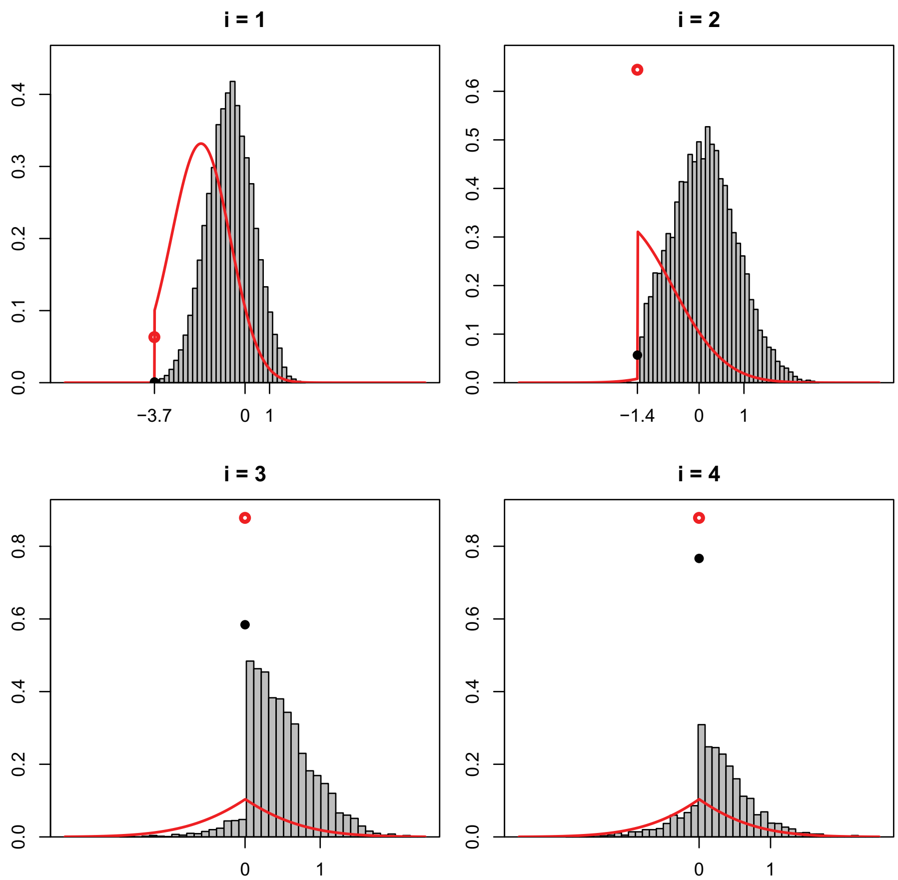

FIG 9. *Lasso, Design I:* $\rho = 0.9$.

The indicated term in the above display is $o(1)$ by the Lemma in the Appendix and because the expression in brackets inside the integral is bounded by 1. Since $r_{i,n} \to r_i$ and $r^*_{i,n} \to -\infty$, the integrand converges to $\Phi(r_i)$ under 3. and to $\Phi(r_i + d_i t)$ under 4. The dominated convergence theorem then completes the proof. The case $\zeta_i = -1$ is treated similarly.

It remains to prove 5. Again assume $\zeta_i = 1$ first. Define $r'_{i,n} = 2^{1/2} n^{-1/2} \eta^{-1}_{i,n} \times (n-k)^{1/2} r_{i,n}$ and $r''_{i,n} = 2^{1/2} n^{-1/2} \eta^{-1}_{i,n} (n-k)^{1/2} r^*_{i,n}$ and rewrite the above display as

$$P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i = 0\right) = \int_{-\infty}^{\infty} \left[\Phi\left(n^{1/2}\eta_{i,n}\left(2\left(n-k\right)\right)^{-1/2}\left(r'_{i,n} + t\right)\right) - \Phi\left(n^{1/2}\eta_{i,n}\left(2\left(n-k\right)\right)^{-1/2}\left(r''_{i,n} - t\right)\right)\right] \phi(t)dt + o(1).$$

Observe that $r'_{i,n} \to r'_i$ and $r''_{i,n} \to -\infty$. The expression in brackets inside the integral hence converges to 1 for $t > -r'_i$ and to 0 for $t < -r'_i$. By dominated

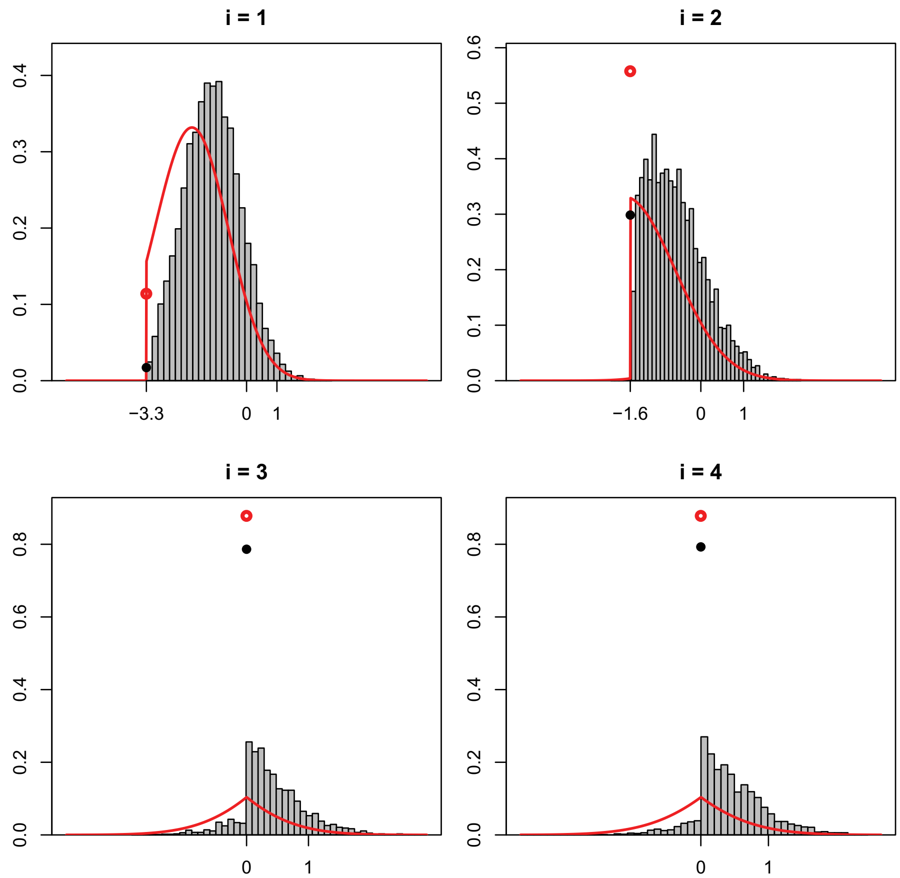

Fig 10. *Lasso, Design II: $c = 0.2$.*

convergence the integral converges to $\int_{-r_i'}^{\infty} \phi(t)dt = \Phi(r_i')$. The case $\zeta_i = -1$ is treated similarly. □

Proof of Proposition 13. Observe that

$$\begin{aligned} &\left| P_{n,\theta,\sigma}\left(\hat{\theta}_i = 0\right) - P_{n,\theta,\sigma}\left(\tilde{\theta}_i = 0\right) \right| \\ &\leq \int_0^\infty \left\{ \left| \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}\right)\right) - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + s\eta_{i,n}\right)\right) \right| \right. \\ &\left. + \left| \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}\right)\right) - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - s\eta_{i,n}\right)\right) \right| \right\} \rho_{n-k}(s)ds. \end{aligned} \tag{26}$$

By a trivial modification of Lemma 13 in Pötscher and Schneider (2010) we conclude that for every $\varepsilon > 0$ there exists a real number $c = c(\varepsilon) > 0$

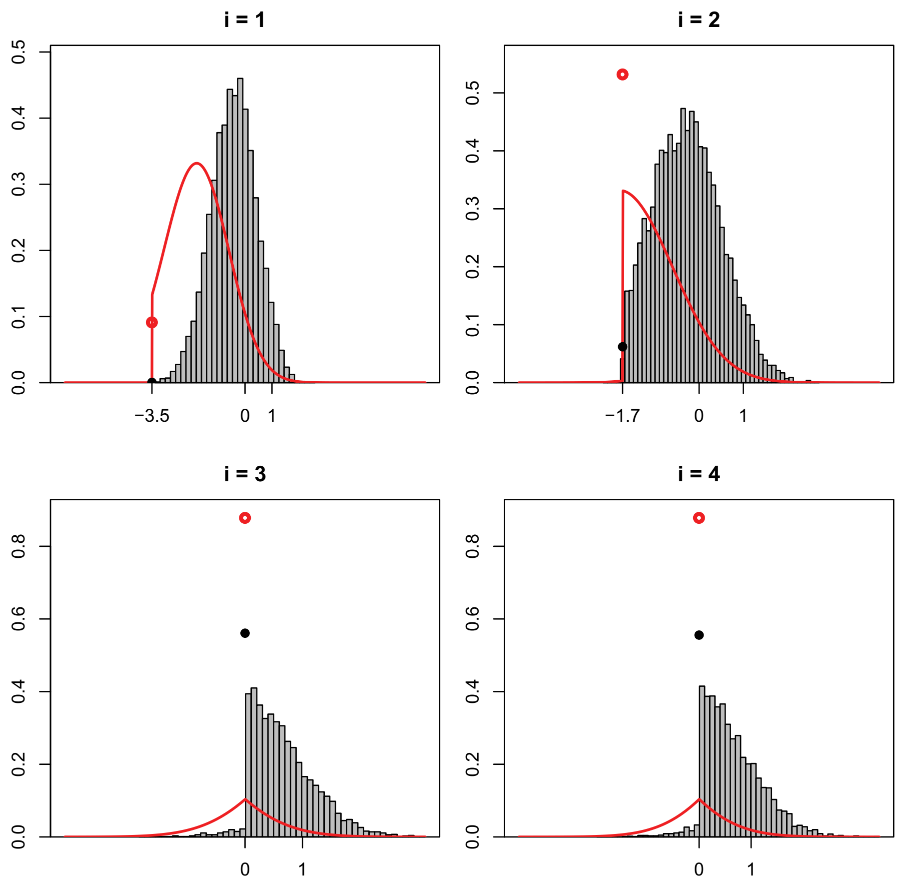

FIG 11. *Lasso, Design II:* $c = 2$.

such that

$$\int_{|s-1|>(n-k)^{-1/2}c} \rho_{n-k}(s)ds < \varepsilon$$

for every $n > k$. Using the fact, that Φ is globally Lipschitz with constant $(2\pi)^{-1/2}$, this gives

$$\begin{aligned}
&\sup_{\theta\in\mathbb{R}^k, 0<\sigma<\infty} \left| P_{n,\theta,\sigma}\left(\hat{\theta}_i = 0\right) - P_{n,\theta,\sigma}\left(\tilde{\theta}_i = 0\right)\right| \\
\leq\ & 2\int_{|s-1|>(n-k)^{-1/2}c} \rho_{n-k}(s)ds \\
&+2(2\pi)^{-1/2}n^{1/2}\eta_{i,n}\int_{|s-1|\leq(n-k)^{-1/2}c} |s-1|\,\rho_{n-k}(s)ds \\
\leq\ & 2\varepsilon + 2(2\pi)^{-1/2}n^{1/2}\eta_{i,n}(n-k)^{-1/2}c
\end{aligned}$$

which proves the result since ε can be made arbitrarily small. $\square$

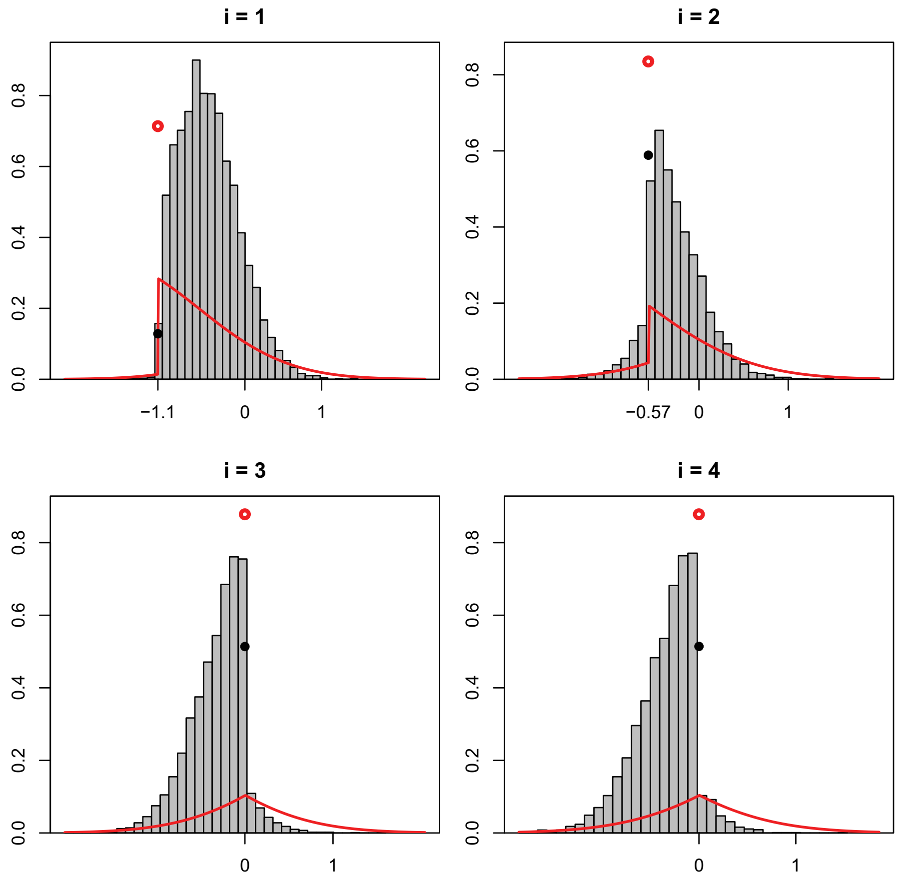

Fig 12. *Lasso, Design II: $c = -0.2$.*

8.2. Proofs for Section 4

Proof of Theorem 16. (a) Observe that

$$\left|\tilde{\theta}_i - \hat{\theta}_{LS,i}\right| \leq \hat{\sigma}\xi_{i,n}\eta_{i,n} \tag{27}$$

holds for any of the estimators. Hence, consistency of $\tilde{\theta}_i$ under $\xi_{i,n}\eta_{i,n} \to 0$ and $\xi_{i,n}/n^{1/2} \to 0$ follows immediately from Proposition 15(a) since the distributions of $\hat{\sigma}/\sigma$ are tight. Conversely, suppose $\tilde{\theta}_i$ is consistent. Then clearly $P_{n,\theta,\sigma}(\tilde{\theta}_i = 0) \to 0$ whenever $\theta_i \neq 0$ must hold, which implies $\xi_{i,n}\eta_{i,n} \to 0$ by Proposition 10(a). This then entails consistency of $\hat{\theta}_{LS,i}$ by (27) and tightness of the distributions of $\hat{\sigma}/\sigma$; this in turn implies $\xi_{i,n}/n^{1/2} \to 0$ by Proposition 15(a).

(b) Since $a_{i,n} \to \infty$, it suffices to prove the second claim in (b). Now for every real $M > 0$ we have

$$
\begin{aligned}
& P_{n,\theta,\sigma}\left(a_{i,n}\left|\tilde{\theta}_{H,i}-\theta_i\right|>\sigma M\right) \\
&= P_{n,\theta,\sigma}\left(a_{i,n}\left|\hat{\theta}_{LS,i}-\theta_i\right|>\sigma M, \left|\hat{\theta}_{LS,i}\right|>\hat{\sigma}\xi_{i,n}\eta_{i,n}\right) \\
&\qquad + \mathbf{1}\left(a_{i,n}\left|\theta_i\right|>\sigma M\right) P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right|\leq\hat{\sigma}\xi_{i,n}\eta_{i,n}\right) \\
&\leq P_{n,\theta,\sigma}\left(a_{i,n}\left|\hat{\theta}_{LS,i}-\theta_i\right|>\sigma M\right) \\
&\qquad + \mathbf{1}\left(a_{i,n}\left|\theta_i\right|>\sigma M\right) P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right|\leq\hat{\sigma}\xi_{i,n}\eta_{i,n}\right) \\
&\leq P_{n,\theta,\sigma}\left(\left(n^{1/2}/\xi_{i,n}\right)\left|\hat{\theta}_{LS,i}-\theta_i\right|>\sigma M\right) \\
&\qquad + \mathbf{1}\left(a_{i,n}\left|\theta_i\right|>\sigma M\right) P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right|\leq\hat{\sigma}\xi_{i,n}\eta_{i,n}\right).
\end{aligned}
$$

This gives

$$
\begin{aligned}
& \sup_{n\in\mathbb{N}}\sup_{\theta\in\mathbb{R}^k}\sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(a_{i,n}\left|\tilde{\theta}_{H,i}-\theta_i\right|>\sigma M\right) \\
\leq\ & \sup_{n\in\mathbb{N}}\sup_{\theta\in\mathbb{R}^k}\sup_{0<\sigma<\infty} P_{n,\theta,\sigma}\left(\left(n^{1/2}/\xi_{i,n}\right)\left|\hat{\theta}_{LS,i}-\theta_i\right|>\sigma M\right) \\
& +\sup_{n\in\mathbb{N}}\sup_{0<\sigma<\infty}\sup_{\theta\in\mathbb{R}^k:|\theta_i|>\sigma M/a_{i,n}} P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right|\leq\hat{\sigma}\xi_{i,n}\eta_{i,n}\right)
\end{aligned}
$$

where the first term on the r.h.s. can be made arbitrarily small in view of Proposition 15(b) by choosing M large enough. The second term on the r.h.s. can be written as (cf. (5))

$$
\begin{aligned}
& \sup_{n\in\mathbb{N}}\sup_{0<\sigma<\infty}\sup_{\theta\in\mathbb{R}^k:|\theta_i|>\sigma M/a_{i,n}}\int_0^\infty P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right|\leq s\sigma\xi_{i,n}\eta_{i,n}\right)\rho_{n-k}(s)ds \\
\leq\ & \sup_{n\in\mathbb{N}}\sup_{0<\sigma<\infty}\int_0^\infty\sup_{\theta\in\mathbb{R}^k:|\theta_i|>\sigma M/a_{i,n}} P_{n,\theta,\sigma}\left(\left|\hat{\theta}_{LS,i}\right|\leq s\sigma\xi_{i,n}\eta_{i,n}\right)\rho_{n-k}(s)ds.
\end{aligned}
$$

For $\varepsilon>0$ choose $c^*(\varepsilon/2)$ as in the proof of Proposition 10. Using continuity of Φ and the fact that the probability appearing on the r.h.s. above is monotonically increasing as $|\theta_i|$ approaches $\sigma M/a_{i,n}$ from above, this can be further bounded by

$$
\begin{aligned}
& \sup_{n\in\mathbb{N}}\int_0^\infty\Phi\left(sn^{1/2}\eta_{i,n}-Ma_{i,n}^{-1}n^{1/2}/\xi_{i,n}\right)\rho_{n-k}(s)ds \\
\leq\ & \varepsilon/2+\sup_{n\in\mathbb{N}}\int_0^{c^*(\varepsilon/2)}\Phi\left(sn^{1/2}\eta_{i,n}-Ma_{i,n}^{-1}n^{1/2}/\xi_{i,n}\right)\rho_{n-k}(s)ds \\
\leq\ & \varepsilon/2+\sup_{n\in\mathbb{N}}\Phi\left(n^{1/2}\xi_{i,n}^{-1}a_{i,n}^{-1}\left(c^*(\varepsilon/2)\xi_{i,n}\eta_{i,n}a_{i,n}-M\right)\right) \\
\leq\ & \varepsilon/2+\Phi\left(c^*(\varepsilon/2)-M\right),
\end{aligned}
$$

the last inequality holding for $M>c^*(\varepsilon/2)$ and since $n^{1/2}\xi_{i,n}^{-1}a_{i,n}^{-1}\geq 1$ and $\xi_{i,n}\eta_{i,n}a_{i,n}\leq 1$. Choosing M sufficiently large (depending on ε) completes the proof for $\tilde{\theta}_{H,i}$. Next observe that

$$a_{i,n}\left|\tilde{\theta}_{H,i}-\tilde{\theta}_{S,i}\right|\leq\hat{\sigma}\min\left(n^{1/2}\eta_{i,n},1\right)\leq\hat{\sigma}$$

and similarly $a_{i,n}\left|\tilde{\theta}_{H,i}-\tilde{\theta}_{AS,i}\right|\leq\hat{\sigma}$ hold. Since the set of distributions of $\hat{\sigma}/\sigma$ (i.e., the set of distributions corresponding to ρ_{n-k}) is tight as already noted, this proves (b) then also for $\hat{\theta}_{S,i}$ and $\hat{\theta}_{AS,i}$.

(c) By a subsequence argument we can reduce the argument to the case where $n^{1/2}\eta_{i,n}\to e_i\in\bar{\mathbb{R}}$ and $n-k$ converges in $\mathbb{N}\cup\{\infty\}$. Suppose first that $e_i=\infty$: Observe that then $a_{i,n}=(\xi_{i,n}\eta_{i,n})^{-1}$ eventually. Choose $\theta_{i,n}$ and σ_n such that $\theta_{i,n}/\left(\sigma_n\xi_{i,n}\eta_{i,n}\right)=\zeta_i$, where ζ_i does not depend on n and $0<|\zeta_i|<1$ holds, and set the other coordinates of $\theta^{(n)}$ to arbitrary values (e.g., equal to zero). Observe that there exists a constant $\delta>0$ such that

$$\liminf_{n\to\infty}P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)>\delta \tag{28}$$

holds: If $n-k$ converges to a finite limit, i.e., is eventually constant, the claim follows from Theorem 11(b1); if $n-k\to\infty$, then use Theorem 11(b2). By (6) we have for $\varepsilon=\delta$ and a suitable M that

$$\begin{aligned}
\delta &> P_{n,\theta^{(n)},\sigma_n}\left(b_{i,n}\left|\tilde{\theta}_i-\theta_{i,n}\right|>\sigma_n M\right)\\
&\geq P_{n,\theta^{(n)},\sigma_n}\left(b_{i,n}\left|\tilde{\theta}_i-\theta_{i,n}\right|>\sigma_n M,\tilde{\theta}_i=0\right)\\
&= P_{n,\theta^{(n)},\sigma_n}\left(\left|b_{i,n}\theta_{i,n}\right|/\sigma_n>M,\tilde{\theta}_i=0\right)\\
&= \mathbf{1}\left(\left|b_{i,n}\theta_{i,n}\right|/\sigma_n>M\right)P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_i=0\right)>\delta\,\mathbf{1}\left(\left|b_{i,n}\theta_{i,n}\right|/\sigma_n>M\right)
\end{aligned}$$

for all n sufficiently large. But this is only possible if $b_{i,n}\xi_{i,n}\eta_{i,n}\leq M/\left|\zeta_i\right|<\infty$ holds eventually, implying that $b_{i,n}=O(a_{i,n})$. Next consider the case where $0<e_i<\infty$: Observe that then $a_{i,n}$ is of the same order as $n^{1/2}/\xi_{i,n}$. Then define $\theta_{i,n}$ and σ_n such that $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})=\nu_i$, where ν_i does not depend on n and $0<|\nu_i|<\infty$ holds, and set the other coordinates of $\theta^{(n)}$ to arbitrary values (e.g., equal to zero). Observe that then (28) also holds, in view of Theorem 11(a1) in case $n-k$ is eventually constant, and in view of Theorem 11(a2) in case $n-k\to\infty$. The rest of the proof is then similar as before. It remains to consider the case $e_i=0$: It follows from (27), the assumptions on $\xi_{i,n}$ and $\eta_{i,n}$, from $e_i=0$, and from the observation that $\hat{\theta}_{LS,i}$ is $N(\theta_i,\sigma^2\xi_{i,n}^2/n)$-distributed, that $n^{1/2}\xi_{i,n}^{-1}\sigma^{-1}(\tilde{\theta}_i-\theta_i)$ converges in distribution to a standard normal distribution for each fixed θ_i and σ. Hence, stochastic boundedness of $\sigma^{-1}b_{i,n}|\tilde{\theta}_i-\theta_i|$ for each θ_i (and a fortiori (6)) necessarily implies that $b_{i,n}=O(n^{1/2}\xi_{i,n}^{-1})=O(a_{i,n})$.

(d) The proof for $\hat{\theta}_i$ is similar and in fact simpler: note that now $|\hat{\theta}_i-\hat{\theta}_{LS,i}|\leq\sigma\xi_{i,n}\eta_{i,n}$ holds and that in the proof of (b) the integration over s can simply be replaced by evaluation at $s=1$. For (c) one uses Proposition 8 instead of Theorem 11. ∎

8.3. Proofs for Section 5

Proofs of Propositions 19, 20, and 21. Observe that

$$\hat{\theta}_{H,i}/(\sigma\xi_{i,n}) = \left(\hat{\theta}_{LS,i}/(\sigma\xi_{i,n})\right) \mathbf{1}\left(\left|\hat{\theta}_{LS,i}/(\sigma\xi_{i,n})\right| > \eta_{i,n}\right)$$

and that $\hat{\theta}_{LS,i}/(\sigma\xi_{i,n})$ is $N\left(\theta_i/(\sigma\xi_{i,n}), 1/n\right)$. Furthermore, we have

$$\begin{aligned} H^i_{H,n,\theta,\sigma}(x) &= P_{n,\theta,\sigma}\left(\sigma^{-1}\alpha_{i,n}(\hat{\theta}_{H,i} - \theta_i) \le x\right) \\ &= P_{n,\theta,\sigma}\left(n^{1/2}(\hat{\theta}_{H,i} - \theta_i)/(\sigma\xi_{i,n}) \le n^{1/2}\alpha_{i,n}^{-1}\xi_{i,n}^{-1}x\right). \end{aligned}$$

Identifying $\hat{\theta}_{LS,i}/(\sigma\xi_{i,n})$ and $\theta_i/(\sigma\xi_{i,n})$ with $\bar{y}$ and θ in Pötscher and Leeb (2009) and making use of eq. (4) in that reference immediately gives the result for $dH^i_{H,n,\theta,\sigma}$. The result for $H^i_{H,n,\theta,\sigma}$ then follows from elementary calculations.

The result for $dH^i_{S,n,\theta,\sigma}$ follows similarly by making use of eq. (5) instead of eq. (4) in Pötscher and Leeb (2009). The result for $H^i_{S,n,\theta,\sigma}$ then follows from elementary calculations.

The results for $dH^i_{AS,n,\theta,\sigma}$ and $H^i_{AS,n,\theta,\sigma}$ follow similarly by making use of eqs. (9)-(11) in Pötscher and Schneider (2009). □

Proofs of Propositions 23, 24, and 25. We have

$$\begin{aligned} H^{i\maltese}_{H,n,\theta,\sigma}(x) &= \int_0^\infty P_{n,\theta,\sigma}\left(\sigma^{-1}\alpha_{i,n}(\tilde{\theta}_{H,i} - \theta_i) \le x \mid \hat{\sigma} = s\sigma\right)\rho_{n-k}(s)ds \\ &= \int_0^\infty H^i_{H,s\eta_{i,n},n,\theta,\sigma}(x)\rho_{n-k}(s)ds, \end{aligned}$$

where we have used independence of $\hat{\sigma}$ and $\hat{\theta}_{LS,i}$ allowing us to replace $\hat{\sigma}$ by $s\sigma$ in the relevant formulae, cf. Leeb and Pötscher (2003, p. 110). Substituting (7), with $\eta_{i,n}$ replaced by $s\eta_{i,n}$, into the above equation gives (12). Representing $H^i_{H,s\eta_{i,n},n,\theta,\sigma}(x)$ as an integral of $dH^i_{H,s\eta_{i,n},n,\theta,\sigma}$ given in (8) and applying Fubini's theorem then gives (13).

Similarly, we have

$$H^{i\maltese}_{S,n,\theta,\sigma}(x) = \int_0^\infty H^i_{S,s\eta_{i,n},n,\theta,\sigma}(x)\rho_{n-k}(s)ds.$$

Substituting (9), with $\eta_{i,n}$ replaced by $s\eta_{i,n}$, into the above equation and noting that $\int_0^\infty \Phi(a+bs)\rho_\nu(s)\,ds = T_{\nu,-a}(b)$ gives (14). Elementary calculations then yield (15).

Finally, we have

$$H^{i\maltese}_{AS,n,\theta,\sigma}(x) = \int_0^\infty H^i_{AS,s\eta_{i,n},n,\theta,\sigma}(x)\rho_{n-k}(s)ds.$$

Substituting (11), with $\eta_{i,n}$ replaced by $s\eta_{i,n}$, into the above equation gives (16). Elementary calculations then yield (17). □

8.4. Proofs for Section 6

Proof of Proposition 27. The proof of (a) is completely analogous to the proof of Theorem 4 in Pötscher and Leeb (2009), whereas the proof of (b) is analogous to the proof of Theorem 17 in the same reference. □

Proof of Proposition 28. The proof of (a) is completely analogous to the proof of Theorem 5 in Pötscher and Leeb (2009), whereas the proof of (b) is analogous to the proof of Theorem 18 in the same reference. □

Proof of Proposition 29. The proof of (a) is completely analogous to the proof of Theorem 4 in Pötscher and Schneider (2009), whereas the proof of (b) is analogous to the proof of Theorem 6 in the same reference. □

Proof of Theorem 30. Observe that the total variation distance between two cdfs is bounded by the sum of the total variation distances between the corresponding discrete and continuous parts. Furthermore, recall that the total variation distance between the absolutely continuous parts is bounded from above by the L_1-distance of the corresponding densities. Hence, from (8) and (13) we obtain

$$\left\| H^{i}_{H,n,\theta,\sigma} - H^{i\maltese}_{H,n,\theta,\sigma} \right\|_{TV} \leq A + B$$

where

$$A = \left| P_{n,\theta,\sigma}\left(\hat{\theta}_{H,i} = 0\right) - P_{n,\theta,\sigma}\left(\tilde{\theta}_{H,i} = 0\right) \right|$$

and

$$\begin{aligned}
B &= \int_{-\infty}^{\infty} \int_{0}^{\infty} \left| \mathbf{1}\left(\left| \alpha_{i,n}^{-1} x + \theta_i/\sigma \right| > \xi_{i,n}\eta_{i,n} \right) \right. \\
&\quad \left. - \mathbf{1}\left(\left| \alpha_{i,n}^{-1} x + \theta_i/\sigma \right| > \xi_{i,n} s \eta_{i,n} \right) \right| \rho_{n-k}(s) ds \\
&\quad \times n^{1/2} \alpha_{i,n}^{-1} \xi_{i,n}^{-1} \phi\left(n^{1/2} x / (\alpha_{i,n}\xi_{i,n}) \right) dx \\
&= \int_{0}^{\infty} \int_{-\infty}^{\infty} \left| \mathbf{1}\left(\left| u + n^{1/2}\theta_i / \left(\sigma\xi_{i,n}\right) \right| > n^{1/2}\eta_{i,n} \right) \right. \\
&\quad \left. - \mathbf{1}\left(\left| u + n^{1/2}\theta_i / \left(\sigma\xi_{i,n}\right) \right| > s n^{1/2}\eta_{i,n} \right) \right| \phi(u) du \rho_{n-k}(s) ds \\
&= \int_{0}^{\infty} \int_{-\infty}^{\infty} \mathbf{1}\left(n^{1/2}\eta_{i,n}(s \wedge 1) < \left| u + n^{1/2}\theta_i/(\sigma\xi_{i,n}) \right| \leq n^{1/2}\eta_{i,n}(s \vee 1) \right) \\
&\quad \times \phi(u) du \rho_{n-k}(s) ds \\
&= \int_{0}^{\infty} \left\{ \left[\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}(s \vee 1) \right) \right) \right. \right. \\
&\qquad \left. - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) + \eta_{i,n}(s \wedge 1) \right) \right) \right] \\
&\quad + \left[\Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}(s \wedge 1) \right) \right) \right. \\
&\qquad \left. \left. - \Phi\left(n^{1/2}\left(-\theta_i/(\sigma\xi_{i,n}) - \eta_{i,n}(s \vee 1) \right) \right) \right] \right\} \rho_{n-k}(s) ds,
\end{aligned}$$

where we have made use of Fubini's theorem and performed an obvious substitution. By a trivial modification of Lemma 13 in Pötscher and Schneider (2010) we conclude that for every $\varepsilon > 0$ there exists a real number $c = c(\varepsilon) > 0$ such that

$$\int_{|s-1|>(n-k)^{-1/2}c} \rho_{n-k}(s)ds < \varepsilon \tag{29}$$

for every $n - k > 0$. Using the fact, that Φ is globally Lipschitz with constant $(2\pi)^{-1/2}$, this gives

$$\begin{aligned}\sup_{\theta\in\mathbb{R}^k,0<\sigma<\infty} B \;&\leq\; 2\int_{|s-1|>(n-k)^{-1/2}c} \rho_{n-k}(s)ds \\ &\quad + 2(2\pi)^{-1/2}n^{1/2}\eta_{i,n}\int_{|s-1|\leq(n-k)^{-1/2}c} |(s\vee 1)-(s\wedge 1)|\,\rho_{n-k}(s)ds \\ &\leq\; 2\varepsilon + 2(2\pi)^{-1/2}n^{1/2}\eta_{i,n}(n-k)^{-1/2}c.\end{aligned}$$

The r.h.s. now converges to 2ε because $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$. Since $\varepsilon > 0$ was arbitrary, this shows that $\sup_{\theta\in\mathbb{R}^k,0<\sigma<\infty} B$ converges to zero. Note also that $\sup_{\theta\in\mathbb{R}^k,0<\sigma<\infty} A$ has already been shown to converge to zero in Proposition 13. This completes the proof for the hard-thresholding estimator.

With the same argument as above we obtain

$$\left\|H^{i}_{S,n,\theta,\sigma} - H^{i\maltese}_{S,n,\theta,\sigma}\right\|_{TV} \leq A + B,$$

where

$$A = \left|P_{n,\theta,\sigma}\left(\hat{\theta}_{S,i} = 0\right) - P_{n,\theta,\sigma}\left(\tilde{\theta}_{S,i} = 0\right)\right|$$

and

$$\begin{aligned} B \;&=\; n^{1/2}\alpha_{i,n}^{-1}\xi_{i,n}^{-1}\int_{-\infty}^{\infty}\int_{0}^{\infty}\left|\phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})+n^{1/2}\eta_{i,n}\right)\right. \\ &\quad \left.-\phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})+n^{1/2}s\eta_{i,n}\right)\right|\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x+\theta_i/\sigma>0\right)dx \\ &\quad +n^{1/2}\alpha_{i,n}^{-1}\xi_{i,n}^{-1}\int_{-\infty}^{\infty}\int_{0}^{\infty}\left|\phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})-n^{1/2}\eta_{i,n}\right)\right. \\ &\quad \left.-\phi\left(n^{1/2}x/(\alpha_{i,n}\xi_{i,n})-n^{1/2}s\eta_{i,n}\right)\right|\rho_{n-k}(s)ds\mathbf{1}\left(\alpha_{i,n}^{-1}x+\theta_i/\sigma<0\right)dx\end{aligned}$$

where we have used (10) and (15). Now,

$$B \leq \int_0^{\infty}\left(B_1(s)+B_2(s)\right)\rho_{n-k}(s)ds$$

where

$$\begin{aligned}B_1(s) &= \int_{-\infty}^{\infty}\left|\phi\left(u+n^{1/2}\eta_{i,n}\right)-\phi\left(u+n^{1/2}s\eta_{i,n}\right)\right|du, \\ B_2(s) &= \int_{-\infty}^{\infty}\left|\phi\left(u-n^{1/2}\eta_{i,n}\right)-\phi\left(u-n^{1/2}s\eta_{i,n}\right)\right|du,\end{aligned}$$

and where we have used Fubini's theorem and an obvious substitution. It is elementary to verify that

$$B_1(s) = B_2(s) = 2\left|\Phi(n^{1/2}\eta_{i,n}(s-1)/2) - \Phi(-n^{1/2}\eta_{i,n}(s-1)/2)\right|,$$

and that $B_1(s) \leq 2$ holds. Consequently, using (29) we obtain that B is bounded above by

$$\begin{aligned}
&4\int_{|s-1|>(n-k)^{-1/2}c}\rho_{n-k}(s)ds + \int_{|s-1|\leq(n-k)^{-1/2}c}\left(B_1(s)+B_2(s)\right)\rho_{n-k}(s)ds \\
&\leq\; 4\varepsilon + 4(2\pi)^{-1/2}n^{1/2}\eta_{i,n}\int_{|s-1|\leq(n-k)^{-1/2}c}|s-1|\,\rho_{n-k}(s)ds \\
&\leq\; 4\varepsilon + 4(2\pi)^{-1/2}n^{1/2}\eta_{i,n}(n-k)^{-1/2}c,
\end{aligned}$$

where we have again used the fact that Φ is globally Lipschitz with constant $(2\pi)^{-1/2}$. Since $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ and $\varepsilon > 0$ was arbitrary, the proof for soft-thresholding is complete, because $\sup_{\theta\in\mathbb{R}^k, 0<\sigma<\infty} A$ goes to zero by Proposition 13.

Finally, from (11) and (16) we obtain

$$\begin{aligned}
&\| H^i_{AS,n,\theta,\sigma} - H^{i\maltese}_{AS,n,\theta,\sigma}\|_\infty \\
&\qquad\leq \int_0^\infty \sup_{x\in\mathbb{R}}\left|\Phi\left(z^{(2)}_{n,\theta,\sigma}(x,\eta_{i,n})\right) - \Phi\left(z^{(2)}_{n,\theta,\sigma}(x,s\eta_{i,n})\right)\right|\rho_{n-k}(s)ds \\
&\qquad\quad + \int_0^\infty \sup_{x\in\mathbb{R}}\left|\Phi\left(z^{(1)}_{n,\theta,\sigma}(x,\eta_{i,n})\right) - \Phi\left(z^{(1)}_{n,\theta,\sigma}(x,s\eta_{i,n})\right)\right|\rho_{n-k}(s)ds \\
&\qquad =: \int_0^\infty C_1(s)\rho_{n-k}(s)ds + \int_0^\infty C_2(s)\rho_{n-k}(s)ds.
\end{aligned}$$

Observe that on the one hand $C_1(s)$ and $C_2(s)$ are bounded by 1, and that on the other hand, using the Lipschitz-property of Φ and the mean-value theorem,

$$\begin{aligned}
|C_1(s)| &\leq (2\pi)^{-1/2}\sup_{x\in\mathbb{R}}\left|z^{(2)}_{n,\theta,\sigma}(x,\eta_{i,n}) - z^{(2)}_{n,\theta,\sigma}(x,s\eta_{i,n})\right| \\
&= (2\pi)^{-1/2}\sup_{x\in\mathbb{R}}\Bigg|n^{1/2}\sqrt{\left(0.5\xi^{-1}_{i,n}(\alpha^{-1}_{i,n}x+\theta_i/\sigma)\right)^2+\eta^2_{i,n}} \\
&\qquad - n^{1/2}\sqrt{\left(0.5\xi^{-1}_{i,n}(\alpha^{-1}_{i,n}x+\theta_i/\sigma)\right)^2+s^2\eta^2_{i,n}}\Bigg| \\
&\leq\; (2\pi)^{-1/2}n^{1/2}\eta^2_{i,n}\,|s-1|\sup_{x\in\mathbb{R}}\left|\left(\left(0.5\xi^{-1}_{i,n}(\alpha^{-1}_{i,n}x+\theta_i/\sigma)\right)^2\bar{s}^{-2}+\eta^2_{i,n}\right)^{-1/2}\right|,
\end{aligned}$$

where $\bar{s}$ is a mean-value between s and 1 which may depend on x. The supremum over x on the r.h.s. is now clearly assumed for $x = -\alpha_{i,n}\theta_i/\sigma$, resulting in the bound

$$|C_1(s)| \leq (2\pi)^{-1/2}n^{1/2}\eta_{i,n}\,|s-1|\,.$$

The same bound is obtained for C_2 in exactly the same way. Consequently, using (29) we obtain

$$\begin{aligned}\sup_{\theta\in\mathbb{R}^k,0<\sigma<\infty} \| H^{i}_{AS,n,\theta,\sigma} - H^{i\maltese}_{AS,n,\theta,\sigma} \|_\infty \;&\leq\; 2\int_{|s-1|>(n-k)^{-1/2}c} \rho_{n-k}(s)ds \\ &\quad + 2(2\pi)^{-1/2}n^{1/2}\eta_{i,n}\int_{|s-1|\leq(n-k)^{-1/2}c} |s-1|\,\rho_{n-k}(s)ds \\ &\leq 2\left[\varepsilon + (2\pi)^{-1/2}n^{1/2}\eta_{i,n}(n-k)^{-1/2}c\right].\end{aligned}$$

Since $n^{1/2}\eta_{i,n}(n-k)^{-1/2} \to 0$ and $\varepsilon > 0$ was arbitrary, the proof is complete. □

Proof of Theorem 33. (a) The atomic part of $dH^{i\maltese}_{H,n,\theta^{(n)},\sigma_n}$ as given in (13) clearly converges weakly to the atomic part of (21) in view of Theorem 11(a1) and the fact that $\alpha_{i,n}\theta_{i,n}/\sigma_n = n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) \to \nu_i$ by assumption; also note that the atomic part converges to the zero measure in case $|\nu_i| = \infty$ or $e_i = 0$ as then the total mass of the atomic part converges to zero. We turn to the absolutely continuous part next. For later use we note that what has been established so far also implies that the total mass of the absolutely continuous part converges to the total mass of the absolutely continuous part of the limit, since it is easy to see that the limiting distribution given in the theorem has total mass 1. The density of the absolutely continuous part of (13) takes the form

$$\phi\left(x\right)\int_0^\infty \mathbf{1}\left(\left|x+n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})\right| > sn^{1/2}\eta_{i,n}\right)\rho_{n-k}(s)ds.$$

Observe that for given $x\in\mathbb{R}$, the indicator function in the above display converges to $\mathbf{1}\left(|x+\nu_i| > se_i\right)$ for Lebesgue almost all s. [If $e_i = 0$, this is necessarily true only for $x \in \mathbb{R}$ with $x \neq -\nu_i$.] Since $n-k = m$ eventually, we get from the dominated convergence theorem that the above display converges to $\phi\left(x\right)\int_0^\infty \mathbf{1}\left(|x+\nu_i| > se_i\right)\rho_m(s)ds$ for every $x\in\mathbb{R}$ (for every $x\in\mathbb{R}$ with $x\neq -\nu_i$ in case $e_i = 0$), which is the density of the absolutely continuous part in (21). Since the total mass of the absolutely continuous part is preserved in the limit as shown above, the proof is completed by Scheffé's Lemma.

(b) Follows immediately from Proposition 27 and Theorem 30. □

Proof of Theorem 34. (a) The atomic part of $dH^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ as given in (15) converges weakly to the atomic part of (22) in view of Theorem 11(a1) and the fact that $\alpha_{i,n}\theta_{i,n}/\sigma_n = n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) \to \nu_i$ by assumption; also note that the atomic part converges to the zero measure in case $|\nu_i| = \infty$ or $e_i = 0$ as then the total mass of the atomic part converges to zero. We turn to the absolutely continuous part next. For later use we note that what has been established so far also implies that the total mass of the absolutely continuous part converges to the total mass of the absolutely continuous part of the limit, since it is easy to see that the limiting distribution given in the theorem has total mass 1. The density of the absolutely continuous part of (15) takes the form

$$\int_0^\infty \phi\left(x + sn^{1/2}\eta_{i,n}\right)\rho_{n-k}(s)ds\mathbf{1}\left(x + n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) > 0\right) + \int_0^\infty \phi\left(x - sn^{1/2}\eta_{i,n}\right)\rho_{n-k}(s)ds\mathbf{1}\left(x + n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) < 0\right).$$

Observe that for given $x \in \mathbb{R}$, the functions $\phi\left(x \pm sn^{1/2}\eta_{i,n}\right)$ converge to $\phi(x \pm se_i)$, respectively, for all s. Since $n-k=m$ eventually, we then get from the dominated convergence theorem that the above display converges to

$$\int_0^\infty \phi(x + se_i)\rho_m(s)ds\mathbf{1}(x+\nu_i > 0) + \int_0^\infty \phi(x - se_i)\rho_m(s)ds\mathbf{1}(x+\nu_i < 0)$$

for every $x \neq -\nu_i$; the last display is precisely the density of the absolutely continuous part in (22). Since the total mass of the absolutely continuous part is preserved in the limit as shown above, the proof is completed by Scheffé's Lemma.

(b) Follows immediately from Proposition 28 and Theorem 30. □

Proof of Theorem 35. (a) Observe that

$$\begin{aligned} &H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}(x) \\ &= \int_0^\infty \Phi\left(z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})\right)\rho_{n-k}(s)ds\mathbf{1}\left(x + n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) \geq 0\right) \\ &\quad + \int_0^\infty \Phi\left(z^{(1)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})\right)\rho_{n-k}(s)ds\mathbf{1}\left(x + n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}) < 0\right) \end{aligned} \tag{30}$$

where $z^{(1)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})$ and $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})$ reduce to

$$0.5(x - n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})) \pm \sqrt{\left(0.5(x + n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n}))\right)^2 + s^2 n\eta_{i,n}^2}.$$

Clearly, $\Phi\left(z^{(1)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})\right)$ as well as $\Phi\left(z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})\right)$ converge for every $s \geq 0$ to

$$\Phi\left(0.5(x-\nu_i) - \sqrt{\left(0.5(x+\nu_i)\right)^2 + s^2e_i^2}\right)$$

and

$$\Phi\left(0.5(x-\nu_i) + \sqrt{\left(0.5(x+\nu_i)\right)^2 + s^2e_i^2}\right),$$

respectively, if $|\nu_i| < \infty$, and the dominated convergence theorem shows that the weights of the indicator functions in (30) converge to the corresponding weights in (23). Since $n^{1/2}\theta_{i,n}/(\sigma_n\xi_{i,n})$ converges to ν_i by assumption, it follows that for every $x \neq -\nu_i$ we have convergence of $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ to the cdf given in (23). This proves part (a) in case $|\nu_i| < \infty$. In case $\nu_i = \infty$, we have that $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})$

converges to x by an application of Proposition 15 in Pötscher and Schneider (2009). Consequently, the limit of $\Phi(z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n}))$ is now $\Phi(x)$. Again applying the dominated convergence theorem and observing that for each $x \in \mathbb{R}$ we have that $\mathbf{1}(x + n^{1/2}\theta_{i,n}/(\sigma_n \xi_{i,n}) < 0)$ is eventually zero, shows that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}(x)$ converges to $\Phi(x)$. The case $\nu_i = -\infty$ is proved analogously.

(b) Follows immediately from Proposition 29 and Theorem 30. ☐

Proof of Theorem 36. Observe that

$$\begin{aligned}
\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{H,i} - \theta_{i,n}) \;&=\; -\theta_{i,n}/(\sigma_n \xi_{i,n}\eta_{i,n})\mathbf{1}\left(\tilde{\theta}_{H,i} = 0\right) \\
&\quad + (\sigma_n \xi_{i,n}\eta_{i,n})^{-1}\left(\hat{\theta}_{LS,i} - \theta_{i,n}\right)\mathbf{1}\left(\tilde{\theta}_{H,i} \neq 0\right) \\
&= -\,\theta_{i,n}/(\sigma_n \xi_{i,n}\eta_{i,n})\mathbf{1}\left(\tilde{\theta}_{H,i} = 0\right) + n^{-1/2}\eta_{i,n}^{-1} Z_n \mathbf{1}\left(\tilde{\theta}_{H,i} \neq 0\right)
\end{aligned}$$

where Z_n is standard normally distributed. The expressions in front of the indicator functions now converge to $-\zeta_i$ and 0, respectively, in probability as $n \to \infty$. Inspection of the cdf of $\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{H,i} - \theta_{i,n})$ then shows that this cdf converges weakly to

$$\left(\lim_{n\to\infty} P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_{H,i} = 0\right)\right)\delta_{-\zeta_i} + \left(1 - \lim_{n\to\infty} P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_{H,i} = 0\right)\right)\delta_0$$

if $|\zeta_i| < \infty$. Part (b) of Theorem 11 completes the proof of both parts of the theorem in case $|\zeta_i| < \infty$. If $|\zeta_i| = \infty$ the same theorem shows that the weak limit is now δ_0. ☐

Proof of Theorem 37. (a) The atomic part of $dH^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ as given in (15) converges weakly to the atomic part given in (24) by Theorem 11(b1). The density of the absolutely continuous part of $dH^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ can be written as

$$\begin{aligned}
&n^{1/2}\eta_{i,n}\int_{-\infty}^{\infty}\phi\left(n^{1/2}\eta_{i,n}\left(x+s\right)\right)\rho_m(s)ds\mathbf{1}\left(x + \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}) > 0\right) \\
&+ n^{1/2}\eta_{i,n}\int_{-\infty}^{\infty}\phi\left(n^{1/2}\eta_{i,n}\left(x-s\right)\right)\rho_m(s)ds\mathbf{1}\left(x + \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}) < 0\right)
\end{aligned}$$

recalling the convention that $\rho_m(s) = 0$ for $s < 0$. Note that with this convention ρ_m is a bounded continuous function on the real line. Since the expressions $n^{1/2}\eta_{i,n}\phi\left(n^{1/2}\eta_{i,n}\left(x+\cdot\right)\right)$ and $n^{1/2}\eta_{i,n}\phi\left(n^{1/2}\eta_{i,n}\left(x-\cdot\right)\right)$ clearly converge weakly to δ_{-x} and δ_x, respectively, the density of the absolutely continuous part of $dH^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ is seen to converge to $\rho_m(-x)\mathbf{1}\left(x+\zeta_i > 0\right) + \rho_m(x)\mathbf{1}\left(x+\zeta_i < 0\right)$ for every $x \neq -\zeta_i$. An application of Scheffé's Lemma then completes the proof, noting that the total mass of the absolutely continuous part of $dH^{i\maltese}_{S,n,\theta^{(n)},\sigma_n}$ converges to the total mass of the absolutely continuous part of (24) as the same is true for the atomic part in view of Theorem 11(b1) (and since the distributions involved all have total mass 1).

(b) Rewrite $\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{S,i} - \theta_{i,n})$ as

$$-\theta_{i,n}/(\sigma_n \xi_{i,n}\eta_{i,n})\mathbf{1}\left(\tilde{\theta}_{S,i} = 0\right) + \left(W_n - (\hat{\sigma}/\sigma_n)\,\mathrm{sign}(W_n + \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}))\right)\mathbf{1}\left(\tilde{\theta}_{S,i} \neq 0\right),$$

where W_n is a sequence of $N(0, n^{-1}\eta_{i,n}^{-2})$-distributed random variables. Observe that $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})$ converges to ζ_i and that W_n converges to zero in $P_{n,\theta^{(n)},\sigma_n}$-probability. Now, if $|\zeta_i| < 1$, then $P_{n,\theta^{(n)},\sigma_n}(\tilde{\theta}_{S,i} = 0) \to 1$ by Theorem 11(b2), and hence $\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{S,i} - \theta_{i,n})$ converges to $-\zeta_i$ in $P_{n,\theta^{(n)},\sigma_n}$-probability. This proves the result in case $|\zeta_i| < 1$. In case $|\zeta_i| > 1$ we have that

$$P_{n,\theta^{(n)},\sigma_n}\left(\tilde{\theta}_{S,i} \neq 0\right) \to 1$$

and

$$P_{n,\theta^{(n)},\sigma_n}\left(\mathrm{sign}(W_n + \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})) = \mathrm{sign}(\zeta_i)\right) \to 1. \tag{31}$$

Clearly, also $\hat{\sigma}/\sigma_n$ converges to 1 in $P_{n,\theta^{(n)},\sigma_n}$-probability since $n-k \to \infty$. Consequently, $\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{S,i} - \theta_{i,n})$ converges to $-\mathrm{sign}(\zeta_i)$ in $P_{n,\theta^{(n)},\sigma_n}$-probability, which proves the case $|\zeta_i| > 1$. Finally, if $|\zeta_i| = 1$, then (31) continues to hold and we can write

$$\begin{aligned}
&\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{S,i} - \theta_{i,n}) \\
&\quad = (-\zeta_i + o(1))\,\mathbf{1}\left(\tilde{\theta}_{S,i} = 0\right) - (o_p(1) + (1 + o_p(1))\,\mathrm{sign}(\zeta_i))\,\mathbf{1}\left(\tilde{\theta}_{S,i} \neq 0\right) \\
&\quad = -\mathrm{sign}(\zeta_i) + o_p(1),
\end{aligned}$$

where $o_p(1)$ refers to a term that converges to zero in $P_{n,\theta^{(n)},\sigma_n}$-probability. This then completes the proof of part (b). □

Proof of Theorem 38. (a) Assume first that $0 \leq \zeta_i < \infty$ holds. Note that $z^{(1)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})$ and $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})$ now reduce to

$$n^{1/2}\eta_{i,n}\left[0.5(x - \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})) \pm \sqrt{\left(0.5(x + \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n}))\right)^2 + s^2}\right].$$

First, for $x > -\zeta_i$ we see that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}(x)$ eventually reduces to

$$\int_0^\infty \Phi\left(z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})\right)\rho_m(s)ds.$$

Furthermore, for $x \geq 0$ we see that $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n}) \to \infty$ for all $s > 0$ whereas for $-\zeta_i < x < 0$ we have that $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n}) \to \infty$ for $s > \sqrt{-x\zeta_i}$ and $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n}) \to -\infty$ for $s < \sqrt{-x\zeta_i}$. As a consequence, we obtain from the dominated convergence theorem that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}(x)$ converges to 1 for $x \geq 0$ and to $\int_{\sqrt{-x\zeta_i}}^{\infty} \rho_m(s)ds$ for $-\zeta_i < x < 0$. Second, for $x < -\zeta_i$ note that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}(x)$ eventually reduces to

$$\int_0^{\infty} \Phi\left(z^{(1)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})\right)\rho_m(s)ds$$

and that $z^{(1)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n}) \to -\infty$ for all $s > 0$ in this case. This shows that for $x < -\zeta_i$ we have that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}(x)$ converges to 0. But this proves the result for the case $0 \leq \zeta_i < \infty$. In case $\zeta_i = \infty$ the same reasoning shows that now $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}(x)$ eventually reduces to

$$\int_0^{\infty} \Phi\left(z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n})\right)\rho_m(s)ds$$

for all x, and that now for $x > 0$ we have $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n}) \to \infty$ for all $s > 0$ whereas for $x < 0$ we have that $z^{(2)}_{n,\theta^{(n)},\sigma_n}(x, s\eta_{i,n}) \to -\infty$ for all $s > 0$. This shows that $H^{i\maltese}_{AS,n,\theta^{(n)},\sigma_n}$ converges weakly to δ_0 in case $\zeta_i = \infty$. The proof for the case $\zeta_i < 0$ is completely analogous.

(b) Rewrite $\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{AS,i} - \theta_{i,n})$ as

$$\begin{aligned}
& -\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\mathbf{1}\left(\tilde{\theta}_{AS,i} = 0\right) \\
& + (\sigma_n\xi_{i,n}\eta_{i,n})^{-1}\left(\hat{\theta}_{LS,i} - \theta_{i,n} - \hat{\sigma}^2\xi_{i,n}^2\eta_{i,n}^2/\hat{\theta}_{LS,i}\right)\mathbf{1}\left(\tilde{\theta}_{AS,i} \neq 0\right) \\
= \quad & -\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\mathbf{1}\left(\tilde{\theta}_{AS,i} = 0\right) \\
& + \left(W_n - \left(\hat{\sigma}^2/\sigma_n\right)\xi_{i,n}\eta_{i,n}/\hat{\theta}_{LS,i}\right)\mathbf{1}\left(\tilde{\theta}_{AS,i} \neq 0\right) \\
= \quad & -\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\mathbf{1}\left(\tilde{\theta}_{AS,i} = 0\right) \\
& + \left(W_n - \left(\hat{\sigma}^2/\sigma_n^2\right)\left(W_n + \theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})\right)^{-1}\right)\mathbf{1}\left(\tilde{\theta}_{AS,i} \neq 0\right)
\end{aligned}$$

where W_n is a sequence of $N(0, n^{-1}\eta_{i,n}^{-2})$-distributed random variables. Note that $\theta_{i,n}/(\sigma_n\xi_{i,n}\eta_{i,n})$ converges to ζ_i by assumption. Now, if $|\zeta_i| < 1$, then $P_{n,\theta^{(n)},\sigma_n}(\tilde{\theta}_{AS,i} = 0) \to 1$ by Theorem 11(b2), hence $\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{AS,i} - \theta_{i,n})$ converges to $-\zeta_i$ in $P_{n,\theta^{(n)},\sigma_n}$-probability, establishing the result in this case. Furthermore, for $1 \leq |\zeta_i| \leq \infty$ rewrite the above display as

$$\begin{aligned}
(-\zeta_i + & \ o(1))\mathbf{1}\left(\tilde{\theta}_{AS,i} = 0\right) + \left(o_p(1) - (1 + o_p(1))\left(\zeta_i + o_p(1)\right)^{-1}\right)\mathbf{1}\left(\tilde{\theta}_{AS,i} \neq 0\right) \\
& = \ (-\zeta_i + o(1))\mathbf{1}\left(\tilde{\theta}_{AS,i} = 0\right) + \left(-\zeta_i^{-1} + o_p(1)\right)\mathbf{1}\left(\tilde{\theta}_{AS,i} \neq 0\right),
\end{aligned}$$

with the convention that $\zeta_i^{-1} = 0$ in case $|\zeta_i| = \infty$. If $|\zeta_i| > 1$ (including the case $|\zeta_i| = \infty$) then $P_{n,\theta^{(n)},\sigma_n}(\tilde{\theta}_{AS,i} \neq 0) \to 1$ by Theorem 11(b2), and hence the last display shows that $\sigma_n^{-1}\alpha_{i,n}(\tilde{\theta}_{AS,i} - \theta_{i,n})$ converges to $-\zeta_i^{-1}$ in $P_{n,\theta^{(n)},\sigma_n}$-probability, establishing the result in this case. Finally, if $|\zeta_i| = 1$ holds, then the last line in the above display reduces to $-\zeta_i + o_p(1)$, completing the proof of part (b). □

Proof of Proposition 39. (a) By a subsequence argument we may assume that $n-k$ converges in $\mathbb{N} \cup \{\infty\}$. Applying Theorem 11(b) we obtain that $P_{n,\theta,\sigma}(\tilde{\theta}_{H,i} = 0)$ converges to 1 in case $\theta_i = 0$, and to 0 in case $\theta_i \neq 0$. Observe that

$$\sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \left(\tilde{\theta}_{H,i} - \theta_i \right) = -\sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \theta_i$$

holds on the event $\tilde{\theta}_{H,i} = 0$, while

$$\sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \left(\tilde{\theta}_{H,i} - \theta_i \right) = \sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \left(\hat{\theta}_{LS,i} - \theta_i \right) =: Z_n$$

holds on the event $\tilde{\theta}_{H,i} \neq 0$. The result then follows in view of the fact that Z_n is standard normally distributed. The proof for $\hat{\theta}_{H,i}$ is similar using Proposition 8(b) instead of Theorem 11(b) (it is in fact simpler as the subsequence argument is not needed).

(b) Again we may assume that $n - k$ converges in $\mathbb{N} \cup \{\infty\}$. By the same reference as in the proof of (a) we obtain that $P_{n,\theta,\sigma}(\tilde{\theta}_{AS,i} = 0)$ converges to 1 in case $\theta_i = 0$, and to 0 in case $\theta_i \neq 0$. Now

$$\sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \left(\tilde{\theta}_{AS,i} - \theta_i \right) = -\sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \theta_i$$

holds on the event $\tilde{\theta}_{AS,i} = 0$ and the claim for $\theta_i = 0$ follows immediately. On the event $\tilde{\theta}_{AS,i} \neq 0$ we have from the definition of the estimator

$$\begin{aligned}\sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \left(\tilde{\theta}_{AS,i} - \theta_i \right) &= \sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \left(\hat{\theta}_{LS,i} - \theta_i - \hat{\sigma}^2 \xi_{i,n}^2 \eta_{i,n}^2 / \hat{\theta}_{LS,i} \right) \\ &= Z_n - (\hat{\sigma}/\sigma)^2 \left(\left(n\eta_{i,n}^2 \right)^{-1} Z_n + \sigma^{-1} \xi_{i,n}^{-1} n^{-1/2} \eta_{i,n}^{-2} \theta_i \right)^{-1}.\end{aligned}$$

Now, if $\theta_i \neq 0$, then the event $\tilde{\theta}_{AS,i} \neq 0$ has probability approaching 1 as shown above. Hence, we have on events that have probability tending to 1

$$\begin{aligned}\sigma^{-1} n^{1/2} \xi_{i,n}^{-1} \left(\tilde{\theta}_{AS,i} - \theta_i \right) &= Z_n - (\hat{\sigma}/\sigma)^2 \left(o_p(1) + \sigma^{-1} \xi_{i,n}^{-1} n^{-1/2} \eta_{i,n}^{-2} \theta_i \right)^{-1} \\ &= Z_n - o_p(1),\end{aligned}$$

since $n\eta_{i,n}^2 \to \infty$ and $\xi_{i,n}^{-1} n^{-1/2} \eta_{i,n}^{-2} \to \infty$ by the assumption and since $\theta_i \neq 0$; also note that $\hat{\sigma}/\sigma$ is stochastically bounded since the collection of distributions corresponding to ρ_m with $m \in \mathbb{N}$ is tight on $(0, \infty)$ as was noted earlier. The proof for $\hat{\theta}_{AS,i}$ is again similar (and simpler) by using Proposition 8(b) instead of Theorem 11(b). □

Appendix

Recall that $\rho_m(x) = 0$ for $x < 0$.

Lemma 46. $(2m)^{-1/2}\rho_m((2m)^{-1/2}t+1)$ *converges to* $\phi(t)$ *in the* L_1*-sense as* $m \to \infty$.

Proof. Observe that the expression $(2m)^{-1/2}\rho_m((2m)^{-1/2}t+1)$ is the density of $U_m = (2m)^{1/2}(\sqrt{\chi^2_m/m}-1)$ where χ^2_m denotes a chi-square distributed random variable with m degrees of freedom. By the central limit theorem and the delta-method U_m converges in distribution to a standard normal random variable. With

$$g_m(x) = 2^{-m/2}\left(\Gamma(m/2)\right)^{-1} x^{(m/2)-1}\exp(-x/2) \qquad \text{for } x > 0$$

being the density of χ^2_m we have for $x > 0$

$$\begin{aligned}
&\rho_m(x) \\
&\quad = 2mxg_m(mx^2) = 2^{1-m/2}\left(\Gamma(m/2)\right)^{-1} m^{1/2}\left(mx^2\right)^{(m/2)-1/2}\exp\left(-mx^2/2\right) \\
&\quad = (8m)^{1/2}\Gamma((m+1)/2)\left(\Gamma(m/2)\right)^{-1} g_{m+1}\left(mx^2\right)
\end{aligned}$$

and we have $\rho_m(x) = 0$ for $x \leq 0$. Since the cdf associated with g_{m+1} is unimodal, this shows that the same is true for the cdf associated with ρ_m. But then convergence in distribution of U_m implies convergence of $m^{-1/2}\rho_m(m^{-1/2}t+1)$ to $\phi(t)$ in the L_1-sense by a result of Ibragimov (1956), Scheffé's Lemma, and a standard subsequence argument. ☐

References

Alliney, S. and Ruzinsky, A. (1994). An Algorithm For the Minimization of Mixed l_1 and l_2 Norms With Applications to Bayesian Estimation. *IEEE Transactions on Signal Processing* **42** 618–627.

Bauer, P., Pötscher, B. M. and Hackl, P. (1988). Model Selection by Multiple Test Procedures. *Statistics* **19** 39–44. MR0921623

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet Shrinkage: Asymptopia? With discussion and a reply by the authors. *Journal of the Royal Statistical Society Series B* **57** 301–369. MR1323344

Fan, J. and Li, R. (2001). Variable Selection Via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association* **96** 1348–1360. MR1946581

Fan, J. and Peng, H. (2004). Nonconcave Penalized Likelihood With a Diverging Number of Parameters. *Annals of Statistics* **32** 928–961. MR2065194

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications* **1**. Wiley, New York. MR0088081

Frank, I. E. and Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools (with discussion). *Technometrics* **35** 109–148.

Ibragimov, I. A. (1956). On the Composition of Unimodal Distributions. *Theory of Probability and its Applications* **1** 255–260. MR0087249

Knight, K. and Fu, W. (2000). Asymptotics of Lasso-Type Estimators. *Annals of Statistics* **28** 1356–1378. MR1805787

Leeb, H. and Pötscher, B. M. (2003). The Finite-Sample Distribution of Post-Model-Selection Estimators and Uniform Versus Nonuniform Approximations. *Econometric Theory* **19** 100–142. MR1965844

Leeb, H. and Pötscher, B. M. (2005). Model Selection and Inference: Facts and Fiction. *Econometric Theory* **21** 21–59. MR2153856

Leeb, H. and Pötscher, B. M. (2008). Sparse Estimators and the Oracle Property, or the Return of Hodges' Estimator. *Journal of Econometrics* **142** 201–211. MR2394290

Pötscher, B. M. (1991). Effects of Model Selection on Inference. *Econometric Theory* **7** 163–185. MR1128410

Pötscher, B. M. (2006). The Distribution of Model Averaging Estimators and an Impossibility Result Regarding its Estimation. *IMS Lecture Notes - Monograph Series* **52** 113–129. MR2427842

Pötscher, B. M. and Leeb, H. (2009). On the Distribution of Penalized Maximum Likelihood Estimators: The LASSO, SCAD, and Thresholding. *Journal of Multivariate Analysis* **100** 2065–2082. MR2543087

Pötscher, B. M. and Schneider, U. (2009). On the Distribution of the Adaptive LASSO Estimator. *Journal of Statistical Planning and Inference* **139** 2775–2790. MR2523666

Pötscher, B. M. and Schneider, U. (2010). Confidence Sets Based on Penalized Maximum Likelihood Estimators in Gaussian Regression. *Electronic Journal of Statistics* **4** 334–360. MR2645488

Sen, P. K. (1979). Asymptotic Properties of Maximum Likelihood Estimators Based on Conditional Specification. *Annals of Statistics* **7** 1019–1033. MR0536504

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B* **58** 267–288. MR1379242

Zhang, C. (2010). Nearly Unbiased Variable Selection under Minimax Concave Penalty. *Annals of Statistics* **38** 893–942. MR2604701

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* **101** 1418–1429. MR2279469