The Finite-Sample Distribution of Post-Model-Selection Estimators, and Uniform Versus Non-Uniform Approximations

Hannes Leeb and Benedikt M. Pötscher

March 2000

Abstract

In Pötscher (1991) the asymptotic distribution of a post-model-selection estimator, both unconditional and conditional on selecting a correct model, has been derived. Limitations of these results are (i) that they do not provide information on the distribution of the post-model-selection estimator conditional on selecting an incorrect model, and (ii) that the quality of this asymptotic approximation to the finite-sample distribution is not uniform w.r.t. the underlying parameters. In the present paper we first obtain the unconditional as well as the conditional finite-sample distribution of the post-model-selection estimator which turns out to be complicated and difficult to interpret. Second, we obtain approximations to the finite-sample distributions that are as simple and easy to interpret as the asymptotic distributions obtained in Pötscher (1991), but at the same time are close to the finite-sample distributions uniformly w.r.t. the underlying parameters. As a by-product, we also obtain the asymptotic distribution conditional on selecting an incorrect model.

2000 Mathematics Subject Classification: 62E15, 62F10, 62F12, 62J05
JEL-Classification: C20

1 Introduction

The traditional theory of parametric statistical inference is primarily concerned with statistical properties of estimators and inference procedures, like tests or confidence sets, under the central assumption of an a priori given model. That is, it is assumed that the model is known to the researcher prior to the statistical analysis, except for the value of the true parameter vector. In practice, however, the specification of the model (choice of functional form, choice of regressors, number of lags, etc.) is often also determined only after the data have been observed, violating the central assumption of an a priori given parametric model. As a consequence, the actual statistical properties of estimators or inference procedures following such a data-driven model selection step are not described by the traditional theory assuming an a priori given model; in fact, they may differ substantially from the properties predicted by the traditional theory, cf. Pötscher (1991, Section 3.3). Ignoring the additional uncertainty originating from the data-based model selection step and (inappropriately) applying traditional theory can hence result in very misleading conclusions.

Sen (1979) obtained the unconditional asymptotic distribution of a post-model-selection estimator in an iid maximum likelihood framework when there are two competing models. In Pötscher (1991) the asymptotic properties of a class of post-model-selection estimators (based on a sequence of tests) were studied in a rather general setting covering non-linear models, dependent processes, and more than two competing models.[1] In particular, the asymptotic distribution of the post-model-selection estimator both unconditional as well as conditional on having chosen a correct model (minimal or not) was derived. While constituting an important step towards understanding the distributional properties of post-model-selection estimators, those results appear to be limited in at least two ways:

- They do not provide information on the distribution of the post-model-selection estimator conditional on selecting an incorrect model. (Although asymptotically such models are never selected by any reasonable model selection procedure, including the one considered in Pötscher (1991), the finite-sample probability of selecting an incorrect model can be substantial, see, e.g., Pötscher and Novak (1998, Table II).)

- The approximation to the finite-sample distribution provided by the asymptotic distribution obtained in Pötscher (1991), while of reasonably good quality in many cases, can be very poor in others, cf. the Monte Carlo results in Pötscher and Novak (1998). As argued in Pötscher (1991) and Pötscher and Novak (1998), this is related to certain aspects of non-uniformity in the convergence of the finite-sample distributions.

In the present paper we take a closer look at the distribution of post-model-selection estimators within the more restricted framework of a linear regression model $Y = X\theta + \epsilon$. We obtain inter alia the finite-sample density $f_{n,\theta,\sigma}(\cdot\,|\,p)$ of the post-model-selection estimator $\tilde{\theta}$ conditional on selecting the first p regressors. It turns out that this density is quite complicated and hence difficult to interpret, cf. (23)-(24). The density of the corresponding asymptotic conditional distribution can be obtained from Pötscher (1991) in case the selected model is correct (i.e., all relevant regressors have been selected for inclusion), cf. (28)-(29). It is much simpler, but at the expense of not providing an approximation that is uniformly close to the finite-sample distribution when the regression parameter varies. In particular, it sometimes misses essential features of the finite-sample distribution. It turns out that a quite simple and easy to interpret uniform approximation capturing all the essential features of the conditional finite-sample distribution can be found, see Theorem 4.2. Because of its simplicity and the uniformity, this approximation allows one to easily get insight into the behaviour of the complicated finite-sample distribution of the post-model-selection estimator. It furthermore turns out that this approximating density comes in the form of the finite-sample density $f^*_{n,\theta,\sigma}(\cdot\,|\,p)$ of an idealized version of the post-model-selection estimator $\tilde{\theta}$.

While the asymptotic distribution of the post-model-selection estimator conditional on selecting a correct model is relatively easy to obtain (even in more general frameworks, cf. Sen (1979) and Pötscher (1991)), the asymptotic distribution conditional on selecting an incorrect model (i.e., conditional on not selecting all relevant regressors for inclusion) is much harder to derive, due to the fact that the conditioning event has limiting probability zero.[2] The uniform approximation result given in

[1] For further results in this direction see Kabaila (1995), Pötscher (1995), and Pötscher and Novak (1998).

[2] Sen (1979) studies the unconditional asymptotic distribution also when the true parameter is local to the smaller of the two competing models. From his derivation the asymptotic distribution

Theorem 4.2, however, allows one to obtain the asymptotic conditional distribution also in this case, thus completing the analysis in Sen (1979) and Pötscher (1991) (at least in the linear regression framework).

The plan of the paper is as follows: After introducing the model and the post-model-selection estimator in Section 2, we obtain the conditional as well as the unconditional finite-sample distribution of the post-model-selection estimator (and of its idealized version) in Section 3. The results in this section provide, in particular, a substantial extension of the results in Giles and Srivastava (1993) who consider a model with a constant and only one regressor; cf. Remark 6.5(ii). Section 4 studies the uniform approximation to the conditional finite-sample distribution of the post-model-selection estimator $\tilde{\theta}$ mentioned above (Theorem 4.2). In Corollary 4.5 the asymptotic conditional distribution is obtained, including the case where one conditions on selecting an incorrect model. Furthermore, regions over which convergence to the asymptotic distribution is uniform/non-uniform are also identified (Corollary 4.6 and its discussion). Analogous results for the unconditional distribution and the selection probabilities are given in Section 5. Section 6 contains technical remarks and extensions. Conclusions are drawn in Section 7. Proofs are relegated to the appendices.

2 The Model and Estimators

Consider the linear regression model

$$Y \;=\; X\theta \,+\, \epsilon \tag{1}$$

where X is a non-stochastic $n \times P$ matrix with $rank(X) = P$ and $\epsilon \sim N(0, \sigma^2 I_n)$, $\sigma^2 > 0$. Here n denotes the sample size and we assume $n > P \geq 1$. For the asymptotic results in Sections 4 and 5 we shall in addition assume that $Q = \lim_{n\to\infty} X'X/n$ exists and is non-singular; cf. Remark 6.4(ii). Similar as in Pötscher (1991), we consider model selection from a collection of nested models $M_0 \subseteq M_1 \subseteq \cdots \subseteq M_P$ which are given by

$$M_p \;=\; \Big\{(\theta_1, \ldots, \theta_P)' \in \mathbf{R}^P : \theta_{p+1} = \cdots = \theta_P = 0\Big\} \qquad (0 \leq p \leq P).$$

Hence, model M_p corresponds to the situation where only the first p regressors in (1) are included. Note that $M_0 = \{0\}$ and $M_P = \mathbf{R}^P$. We call M_p the regression model of order p.

The following notation will prove useful. For matrices A and B of the same row-dimension, the column-wise concatenation is denoted by $(A : B)$. If C is an $n \times P$ matrix, let $C[p]$ denote the $n \times p$ matrix consisting of the first p columns of C. Similarly, let $C[\neg p]$ denote the $n \times (P-p)$ matrix consisting of the last $P - p$ columns of C. If x is a $P \times 1$ vector, we write in abuse of notation $x[p]$ and $x[\neg p]$ for $(x'[p])'$ and $(x'[\neg p])'$, respectively. (We shall use these definitions also in the "boundary" cases $p = 0$ and $p = P$. It will always be clear from the context how expressions containing the symbols $C[0]$, $C[\neg P]$, $x[0]$, and $x[\neg P]$ are to be interpreted.) As usual, if x is a vector, the j-th component of x will be denoted by x_j. Using the above notation, the restricted least-squares estimator using only the first p regressors is given by

$$\hat{\theta}(p) \;=\; (X[p]'X[p])^{-1}\, X[p]'Y \tag{2}$$

conditional on selecting the smaller model could be easily obtained. However, in this local alternatives framework the limiting probabilities of selecting either one of the two competing models are positive.

for $p \geq 1$, and by $\hat{\theta}(p) = 0$ for $p = 0$.

Given a parameter value θ, its order is defined as

$$p_0(\theta) \;=\; \min\{p : 0 \leq p \leq P, \theta \in M_p\}.$$

Hence, if θ is the true parameter value, only models M_p with $p \geq p_0(\theta)$ are correct models. We stress that $p_0(\theta)$ is a property of a *single parameter value*, and hence needs to be distinguished from the notion of the order of the model M_p introduced earlier, which is a property of the *set* M_p.

A model selection procedure in general is now nothing else than a data-driven (measurable) rule $\hat{p}$ that selects a value from $\{0, \ldots, P\}$ and thus selects a model from the list of candidate models $M_0, \ldots, M_P$. In this paper, we shall consider a model selection procedure based on a sequence of hypothesis tests, which is given as follows: The sequence of hypotheses $H_0^p : \theta \in M_{p-1}$, is tested against the alternative $H_1^p : \theta \in M_p \setminus M_{p-1}$ in decreasing order starting at $p = P$ until the first rejection occurs. If H_0^p is the first hypothesis in this process that is rejected, we set $\hat{p} = p$. If no hypothesis is rejected we set $\hat{p} = 0$. (The case where a subvector of θ is assumed to be common to any submodel and hence is not subject to test is discussed in Remark 6.5(i).) Each hypothesis in this sequence is tested by a kind of t-test where the error variance estimator is always taken from the overall model. More formally, we have

$$\hat{p} \;=\; \max\{p : |T_p| \geq c_p, 0 \leq p \leq P\}$$

where the test-statistics are given by

$$T_p \;=\; \frac{n^{1/2}\hat{\theta}_p(p)}{\hat{\sigma}\xi_{n,p}} \qquad (1 \leq p \leq P) \tag{3}$$

with

$$\xi_{n,p} \;=\; \left(\left[(X[p]'X[p]/n)^{-1}\right]_{p,p}\right)^{\frac{1}{2}} \qquad (1 \leq p \leq P)$$

being the square root of the p-th diagonal element of the matrix indicated and with

$$\hat{\sigma}^2 \;=\; (n-P)^{-1}(Y - X\hat{\theta}(P))'(Y - X\hat{\theta}(P)). \tag{4}$$

The critical values c_p are independent of sample size n (cf., however, Remark 6.4(i)), and satisfy $0 < c_p < \infty$ for $1 \leq p \leq P$. We also set $T_0 = c_0 = 0$ in order to ensure a well-defined $\hat{p}$. Note that under the hypothesis H_0^p the statistic T_p for $1 \leq p \leq P$ is t-distributed with $n - P$ degrees of freedom.

The post-model-selection estimator $\tilde{\theta}$ is now defined as follows: The first $\hat{p}$ coordinates of $\tilde{\theta}$ are given by the restricted least-squares formula based on the first $\hat{p}$ regressors and the remaining $P - \hat{p}$ coordinates are set equal to zero. I.e., $\tilde{\theta}$ is the $P \times 1$ vector given by

$$\tilde{\theta} \;=\; (0, \ldots, 0)'\,\mathbf{1}\{\hat{p} = 0\} \;+\; \sum_{p=1}^{P}(\hat{\theta}(p)' : 0)'\,\mathbf{1}\{\hat{p} = p\}.$$

For theoretical reasons we shall also be interested in the idealized model selection procedure which assumes knowledge of σ^2 and hence uses T_p^* instead of T_p, where

$$T_p^* \;=\; \frac{n^{1/2}\hat{\theta}_p(p)}{\sigma\xi_{n,p}} \qquad (1 \leq p \leq P) \tag{5}$$

and $T_0^* = 0$. The corresponding model selector is denoted by $\hat{p}^*$ and the resulting idealized post-model-selection estimator by $\tilde{\theta}^*$.

We conclude this section by introducing some further notation. For $p \geq 1$, the expected value of the restricted least-squares estimator $\hat{\theta}(p)$ given by (2) will be denoted by $\mu_n(p)$ and is given by

$$\mu_n(p) \;=\; \theta[p] \;+\; (X[p]'X[p])^{-1}X[p]'X[\neg p]\theta[\neg p]. \tag{6}$$

In case $p = P$, (6) is to be interpreted as $\mu_n(P) = \theta[P] = \theta$. As usual, the j-th component of $\mu_n(p)$ will be denoted by $\mu_{n,j}(p)$. We note that $\mu_n(p)$ depends on θ, although this dependence is not shown explicitly in the notation. The distribution of $n^{1/2}(\hat{\theta}(p) - \mu_n(p))$ is normal with mean zero and variance-covariance matrix $\sigma^2(X[p]'X[p]/n)^{-1}$. The cdf and pdf of this distribution, respectively, will be denoted by $\Phi_{n,p}$ and $\phi_{n,p}$. Their corresponding large sample limits, i.e., the cdf and pdf of a normal distribution with mean zero and variance-covariance matrix equal to $\sigma^2 \lim_n(X[p]'X[p]/n)^{-1}$, will be denoted by $\Phi_{\infty,p}$ and $\phi_{\infty,p}$, respectively. As usual, the cdf and pdf of the standard normal distribution on the real line will be denoted by Φ and ϕ, respectively. We shall also use $\Delta(a, b)$ with $a \in \mathbf{R} \cup \{-\infty, \infty\}$ and $b \in \mathbf{R}$ as shorthand notation for $\Phi(a + b) - \Phi(a - b)$ with the usual convention that $\Phi(\infty) = 1$ and $\Phi(-\infty) = 0$.

3 Finite-Sample Distributions

In a more general framework allowing for nonlinear models as well as dependent data, Pötscher (1991) obtained the asymptotic distribution of the corresponding post-model-selection estimator conditional on the event of choosing a correct model M_p (i.e., $p \geq p_0(\theta)$) as well as the unconditional asymptotic distribution[3]. In this section we obtain, within the more restrictive normal linear model framework, finite-sample results allowing also for the case where one conditions on the event of choosing an incorrect model, M_p (i.e., $p < p_0(\theta)$). A key ingredient in the development leading to the asymptotic distribution in Pötscher (1991) is an asymptotic independence result, cf. Lemma 3 in that paper. A finite-sample version of this result, but which also allows for $p < p_0(\theta)$, will play a similarly important role in the present context.

Proposition 3.1 *For every p, $0 \leq p \leq P$, the random quantities $\hat{\theta}(p)$, $\hat{\theta}_{p+1}(p+1), \ldots, \hat{\theta}_P(P)$ and $\hat{\sigma}^2$ are mutually independent. In particular, $\hat{\theta}(p)$ is independent of $\{T_{p+1}, \ldots, T_P\}$ as well as of $\{T_{p+1}^*, \ldots, T_P^*\}$. Furthermore, $T_1^*, T_2^*, \ldots, T_P^*$ are mutually independent.*

Inspection of the proof in Appendix A shows that in fact for every p, $0 \leq p \leq P$, the collection $\{\hat{\theta}(r) : 0 \leq r \leq p\}$ is independent of $\{\hat{\theta}_j(r) : p+1 \leq r \leq P, j \geq p+1\}$, and that the collection $\{\hat{\theta}(r) : 0 \leq r \leq P\}$ is independent of $\hat{\sigma}^2$.

3.1 The Known-Variance Case

In this subsection we assume that σ^2 is known and we are interested in the finite-sample distribution of the idealized version $\tilde{\theta}^*$ of the post-model-selection estimator.

[3] See also Appendix A of Pötscher and Novak (1998) for a relaxation of an unnecessarily restrictive assumption in Pötscher (1991).

The conditional distribution function of $\tilde{\theta}^*$ conditional on selecting model order p, i.e., conditional on $\hat{p}^* = p$, is considered first and satisfies

$$P_{n,\theta,\sigma}\left(\tilde{\theta}^* \le (t_1,\ldots,t_P)' \,\middle|\, \hat{p}^* = p\right) = P_{n,\theta,\sigma}\left(\hat{\theta}(p) \le (t_1,\ldots,t_p)' \,\middle|\, \hat{p}^* = p\right) \mathbf{1}_{\mathbf{R}_+^{P-p}}(t_{p+1},\ldots,t_P) \tag{7}$$

for $1 \le p \le P$, and is equal to point-mass at zero if $p = 0$. Here $P_{n,\theta,\sigma}$ denotes the probability measure under the true parameters θ and σ. Furthermore, $\mathbf{R}_+$ is the set of non-negative real numbers and we use the convention that the indicator function in (7) is identically equal to one if $p = P$. It suffices now to concentrate only on the case $p \ge 1$, and in this case we will consider a centered and scaled version of (7), namely

$$P_{n,\theta,\sigma}\left(\sqrt{n}(\tilde{\theta}^* - (\mu_n(p)' : 0)') \le (t_1,\ldots,t_P)' \,\middle|\, \hat{p}^* = p\right) = P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \le (t_1,\ldots,t_p)' \,\middle|\, \hat{p}^* = p\right) \mathbf{1}_{\mathbf{R}_+^{P-p}}(t_{p+1},\ldots,t_P) \tag{8}$$

where $\mu_n(p)$ is given by (6). We recall that $\mu_n(p)$ is the unconditional mean of the restricted least-squares estimator $\hat{\theta}(p)$ and satisfies $\mu_n(p) = \theta[p]$ whenever $p \ge p_0(\theta)$. The centering and scaling applied in (8) is of course only of trivial consequence for the finite-sample results, but will prove useful for the asymptotic analysis (cf. also Remark 6.1 on centering constants). For the sake of simplicity we shall henceforth concentrate on

$$F^*_{n,\theta,\sigma}(t\,|\,p) = P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \le t \,\middle|\, \hat{p}^* = p\right) \qquad (1 \le p \le P) \tag{9}$$

with $t \in \mathbf{R}^p$. For convenience, we shall refer also to (9) as the conditional distribution of the post-model-selection estimator $\tilde{\theta}^*$. Now, in light of Proposition 3.1,

$$\begin{aligned} F^*_{n,\theta,\sigma}(t\,|\,p) &= \frac{P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \le t,\, |T_p^*| \ge c_p, |T_{p+1}^*| < c_{p+1},\ldots,|T_P^*| < c_P\right)}{P_{n,\theta,\sigma}\left(|T_p^*| \ge c_p, |T_{p+1}^*| < c_{p+1},\ldots,|T_P^*| < c_P\right)} \\ &= \frac{P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \le t,\, |T_p^*| \ge c_p\right)}{P_{n,\theta,\sigma}\left(|T_p^*| \ge c_p\right)}. \end{aligned} \tag{10}$$

As a point of interest we note that (10), and hence (12) below, are not affected by the tests based on $T_{p+1},\ldots,T_P$ at all. Recalling that Φ denotes the cdf of a standard normal random variable and $\Delta(a,b) = \Phi(a+b) - \Phi(a-b)$, the denominator on the r.h.s. of (10) is easily seen to be $1 - \Delta(-n^{1/2}\mu_{n,p}(p)\sigma^{-1}\xi_{n,p}^{-1}, c_p)$, since T_p^* is normally distributed with mean $n^{1/2}\mu_{n,p}(p)\sigma^{-1}\xi_{n,p}^{-1}$ and variance 1. The numerator can be written as

$$\int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_{p-1}} \int_{-\infty}^{t_p} \phi_{n,p}(r)\mathbf{1}_{U(p,c_p)}(r_p)dr \tag{11}$$

where

$$U(p,c_p) = \mathbf{R} \setminus (-n^{1/2}\mu_{n,p}(p) - \sigma\xi_{n,p}c_p, -n^{1/2}\mu_{n,p}(p) + \sigma\xi_{n,p}c_p)$$

and $\phi_{n,p}$ is the normal density defined at the end of Section 2. From (10) and (11) the density $f^*_{n,\theta,\sigma}(t\,|\,p)$ of $F^*_{n,\theta,\sigma}(t\,|\,p)$ is now easily obtained as

$$f^*_{n,\theta,\sigma}(t\,|\,p) = \phi_{n,p}(t)\,\frac{\mathbf{1}_{U(p,c_p)}(t_p)}{1 - \Delta(-n^{1/2}\mu_{n,p}(p)\sigma^{-1}\xi_{n,p}^{-1}, c_p)}. \tag{12}$$

The conditional density of $\tilde{\theta}^*$ is thus seen to be an excised normal distribution, the denominator in (12) just being a normalizing constant. The excision interval is centered at $-n^{1/2}\mu_{n,p}(p)$. In case $\mu_{n,p}(p) = 0$, which is in particular the case if $p > p_0(\theta)$, the conditional density in (12) simplifies to

$$f^*_{n,\theta,\sigma}(t \,|\, p) \;=\; \phi_{n,p}(t)\, \frac{\mathbf{1}_{\mathbf{R}\setminus(-\sigma\xi_{n,p}c_p,\sigma\xi_{n,p}c_p)}(t_p)}{1-\Delta(0,c_p)} \tag{13}$$

which is symmetric about $t = 0$. It should also be noted that (13) depends on sample size only through the matrix $(X[p]'X[p]/n)^{-1}$ and through the square root of its last diagonal element $\xi_{n,p}$, both of which stabilize at their limits as the sample size increases (given $X'X/n \to Q$). However, in case $\mu_{n,p}(p)$ does not vanish, which is typical for the case $p \leq p_0(\theta)$, (12) shows a further and more important sample-size-dependence of $f^*_{n,\theta,\sigma}(\cdot \,|\, p)$ in that the center of the excision interval, i.e., $-n^{1/2}\mu_{n,p}(p)$, will typically diverge to $\pm\infty$. Furthermore, note that (13) is independent of θ, while in general (12) depends on θ through $\mu_{n,p}(p)$. See Section 4 below for more discussion.

Not too surprisingly, (13) coincides with the asymptotic distribution of the post-model-selection estimator obtained in Pötscher (1991, eq. (2)) in case $p > p_0(\theta)$, except that the matrix $\sigma^2(X[p]'X[p]/n)^{-1}$ replaces the asymptotic variance-covariance matrix of the restricted estimator in the latter reference. As a consequence, all further derivations in Pötscher (1991) based only on this distribution apply also to the conditional finite-sample distribution of the post-model-selection estimator $\tilde{\theta}^*$ as long as $p > p_0(\theta)$. In particular, the numerical results in Section 3.3 of Pötscher (1991) demonstrating the considerable discrepancy between the asymptotic distribution of the post-model-selection estimator conditional on $\hat{p} = p$ and the distribution of the restricted estimator treating p as given, directly apply here to the comparison of the conditional finite-sample distribution of $\tilde{\theta}^*$ with the finite-sample distribution of the restricted least-squares estimator $\hat{\theta}(p)$.

The unconditional finite-sample distribution of $\tilde{\theta}^*$ can now easily be expressed in terms of the conditional distributions $F^*_{n,\theta,\sigma}(\cdot|p)$ and the selection probabilities $\pi^*_{n,\theta,\sigma}(p) = P_{n,\theta,\sigma}(\hat{p}^* = p)$. Let

$$F^*_{n,\theta,\sigma}(u) \;=\; P_{n,\theta,\sigma}\left(\sqrt{n}(\tilde{\theta}^* - \theta) \;\leq\; u\right)$$

with $u \in \mathbf{R}^P$ denote the unconditional cdf. (Again, centering and scaling is of no consequence for the finite-sample results but will prove useful in later sections; cf. also Remark 6.2(ii).) Using (7) and (9) we obtain

$$F^*_{n,\theta,\sigma}(u) \;=\; \sum_{p=0}^{P} F^*_{n,\theta,\sigma}(u[p] + \sqrt{n}(\theta[p] - \mu_n(p))|p)\mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \sqrt{n}\theta[\neg p])\pi^*_{n,\theta,\sigma}(p) \tag{14}$$

where we use the convention that $F^*_{n,\theta,\sigma}(\cdot|0) \equiv 1$. Observe that the expressions $\theta[p] - \mu_n(p)$ and $\theta[\neg p]$ in (14) vanish for $p \geq p_0(\theta)$. In view of Proposition 3.1, the selection probabilities $\pi^*_{n,\theta,\sigma}(p)$ can be expressed as

$$\pi^*_{n,\theta,\sigma}(p) \;=\; P_{n,\theta,\sigma}(|T^*_p| \geq c_p) \prod_{j=p+1}^{P} P_{n,\theta,\sigma}(|T^*_j| < c_j) \tag{15}$$

with

$$P_{n,\theta,\sigma}(|T^*_j| < c_j) \;=\; \Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi^{-1}_{n,j}, c_j) \qquad (1 \leq j \leq P) \tag{16}$$

and $P_{n,\theta,\sigma}(|T_0^*| \geq c_0) = 1$. Note that for $j > p_0(\theta)$ the probability in (16) simplifies to $\Delta(0, c_j)$, which is nothing else than 1 minus the significance level of the test based on T_j^*. Plugging (10), (15) and (16) into (14) gives the alternative expression

$$\begin{aligned} F^*_{n,\theta,\sigma}(u) &= \mathbf{1}_{\mathbf{R}_+^P}(u + \sqrt{n}\theta) \prod_{j=1}^{P} \Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, c_j) + \\ &\quad \sum_{p=1}^{P} P_{n,\theta,\sigma}(\sqrt{n}(\hat{\theta}(p) - \theta[p]) \leq u[p], |T_p^*| \geq c_p) \cdot \\ &\qquad \mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \sqrt{n}\theta[\neg p]) \prod_{j=p+1}^{P} \Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, c_j) \\ &= \sum_{p=0}^{P} \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_p} \phi_{n,p}(r) 1_{U(p,c_p)}(r_p) dr \cdot \\ &\qquad \mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \sqrt{n}\theta[\neg p]) \prod_{j=p+1}^{P} \Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, c_j), \end{aligned} \tag{17}$$

with the convention that the integral in (17) is to be interpreted as 1 when $p = 0$.

3.2 The Unknown-Variance Case

In this subsection we consider the finite-sample distribution of the post-model-selection estimator $\tilde{\theta}$. Again

$$\begin{aligned} &P_{n,\theta,\sigma}\left(\tilde{\theta} \leq (t_1, \ldots, t_P)' \,\middle|\, \hat{p} = p\right) \\ &\quad = \quad P_{n,\theta,\sigma}\left(\hat{\theta}(p) \leq (t_1, \ldots, t_p)' \,\middle|\, \hat{p} = p\right) \mathbf{1}_{\mathbf{R}_+^{P-p}}(t_{p+1}, \ldots, t_P) \end{aligned} \tag{18}$$

for $1 \leq p \leq P$ holds, and the distribution of $\tilde{\theta}$ conditional on $\hat{p} = 0$ is point mass at zero. Similarly as before, in case $p \geq 1$ we may instead consider

$$\begin{aligned} &P_{n,\theta,\sigma}\left(\sqrt{n}(\tilde{\theta} - (\mu_n(p)' : 0)') \leq (t_1, \ldots, t_P)' \,\middle|\, \hat{p} = p\right) \\ &\quad = \quad P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \leq (t_1, \ldots, t_p)' \,\middle|\, \hat{p} = p\right) \mathbf{1}_{\mathbf{R}_+^{P-p}}(t_{p+1}, \ldots, t_P) \end{aligned} \tag{19}$$

and may concentrate only on the first factor

$$F_{n,\theta,\sigma}(t \,|\, p) \;=\; P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \leq t \,\middle|\, \hat{p} = p\right) \qquad (1 \leq p \leq P) \tag{20}$$

where $t \in \mathbf{R}^p$. For convenience we shall again refer also to (20) as the conditional distribution of $\tilde{\theta}$. Now

$$\begin{aligned} &F_{n,\theta,\sigma}(t \,|\, p) \\ &\quad = \frac{P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \leq t, |T_p| \geq c_p, |T_{p+1}| < c_{p+1}, \ldots, |T_P| < c_P\right)}{P_{n,\theta,\sigma}(|T_p| \geq c_p, |T_{p+1}| < c_{p+1}, \ldots, |T_P| < c_P)} \\ &\quad = \frac{\int_0^\infty P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \leq t, |T_p| \geq c_p, |T_{p+1}| < c_{p+1}, \ldots, |T_P| < c_P \,\middle|\, \hat{\sigma}/\sigma = s\right) h(s)\, ds}{\int_0^\infty P_{n,\theta,\sigma}(|T_p| \geq c_p, |T_{p+1}| < c_{p+1}, \ldots, |T_P| < c_P \mid \hat{\sigma}/\sigma = s) h(s)\, ds} \end{aligned} \tag{21}$$

where h denotes the density of $\hat{\sigma}/\sigma$, i.e., h is the density of $(n - P)^{-1/2}$ times the square-root of a chi-square distributed random variable with $n - P$ degrees of freedom. In view of Proposition 3.1 and Theorem 5.3.22 in Gaenssler and Stute (1977), we may evaluate the conditional probabilities in (21) by fixing $\hat{\sigma}/\sigma$ at the value s and then by computing the unconditional probabilities. This gives

$$\begin{aligned} &F_{n,\theta,\sigma}(t \,|\, p) \\ &\quad = \frac{\int_0^\infty P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \leq t, |T_p^*| \geq sc_p, |T_{p+1}^*| < sc_{p+1}, \ldots, |T_P^*| < sc_P\right) h(s)\, ds}{\int_0^\infty P_{n,\theta,\sigma}\left(|T_p^*| \geq sc_p, |T_{p+1}^*| < sc_{p+1}, \ldots, |T_P^*| < sc_P\right) h(s)\, ds} \\ &\quad = \frac{\int_0^\infty P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \leq t, |T_p^*| \geq sc_p\right) \prod_{j=p+1}^{P} P_{n,\theta,\sigma}(|T_j^*| < sc_j) h(s)\, ds}{\int_0^\infty P_{n,\theta,\sigma}\left(|T_p^*| \geq sc_p\right) \prod_{j=p+1}^{P} P_{n,\theta,\sigma}(|T_j^*| < sc_j) h(s)\, ds} \end{aligned} \tag{22}$$

where we have made use of independence of $\hat{\theta}(p)$ from $(T^*_{p+1},\ldots,T^*_P)$ as well as of independence of T^*_j, cf. Proposition 3.1. Observe that the first factor of the integrand in the numerator is identical to the numerator in the expression for $F^*_{n,\theta,\sigma}(t\,|\,p)$ given in (10), except for the fact that the critical value c_p has been replaced by sc_p. Replacing this factor in (22) by the equivalent expression given in (11) with sc_p taking now the place of c_p, using Fubini's theorem and making use of (16) one can immediately read off the density of $F_{n,\theta,\sigma}(t\,|\,p)$:

$$f_{n,\theta,\sigma}(t\,|\,p) \;=\; \phi_{n,p}(t)\, m_{n,p,\theta,\sigma}(t_p) \tag{23}$$

where

$$\begin{aligned}
&m_{n,p,\theta,\sigma}(t_p)\\
&\quad= \left[\int_0^\infty \mathbf{1}_{U(p,sc_p)}(t_p)\prod_{j=p+1}^{P}\Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, sc_j)h(s)ds\right] \Big/\\
&\qquad\left[\int_0^\infty(1-\Delta(-n^{1/2}\mu_{n,p}(p)\sigma^{-1}\xi_{n,p}^{-1}, sc_p))\prod_{j=p+1}^{P}\Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, sc_j)h(s)ds\right]\\
&\quad= \left[\int_0^{|t_p+n^{1/2}\mu_{n,p}(p)|/(\sigma\xi_{n,p}c_p)}\prod_{j=p+1}^{P}\Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, sc_j)h(s)ds\right] \Big/\\
&\qquad\left[\int_0^\infty(1-\Delta(-n^{1/2}\mu_{n,p}(p)\sigma^{-1}\xi_{n,p}^{-1}, sc_p))\prod_{j=p+1}^{P}\Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, sc_j)h(s)ds\right].
\end{aligned}\tag{24}$$

Formula (23) shows that $f_{n,\theta,\sigma}(t|p)$ is the multivariate normal density $\phi_{n,p}(t)$ times a "modification factor" $m_{n,p,\theta,\sigma}$ that is only a function of t_p. While the corresponding modification factor for $f^*_{n,\theta,\sigma}(t|p)$ was seen in (12) to be an indicator function (divided by a normalizing constant) leading to an excised normal distribution, the modification factor (24) is of a more complicated nature; in particular, the indicator function (and thus the excision interval) are "smoothed" through integration w.r.t. the distribution of $\hat{\sigma}/\sigma$. The modification factor $m_{n,p,\theta,\sigma}(t_p)$ is zero at $t_p = -n^{1/2}\mu_{n,p}(p)$, is symmetric about that point and strictly increases to unity as the argument t_p moves away from $-n^{1/2}\mu_{n,p}(p)$. Also note that $f_{n,\theta,\sigma}(\cdot|p)$ is symmetric about zero if $\mu_{n,p}(p)=0$, which arises, e.g., if $p > p_0(\theta)$. Furthermore, if $\mu_{n,j}(j)=0$ for $p\le j\le P$, which is the case if $p>p_0(\theta)$ holds, the conditional density $f_{n,\theta,\sigma}(\cdot|p)$ is independent of θ.

As a point of interest we note that, contrary to the known-variance case, (22) and (23)-(24) show a dependence on the tests based on $T_{p+1},\ldots,T_P$. However, this dependence will be seen to fade out with increasing sample size, cf. Remark 6.3(ii).

The unconditional finite-sample distribution of the post-model-selection estimator $\tilde{\theta}$ can now easily be expressed in terms of the conditional distributions $F_{n,\theta,\sigma}(\cdot|p)$ and the selection probabilities $\pi_{n,\theta,\sigma}(p) = P_{n,\theta,\sigma}(\hat{p}=p)$. Let

$$F_{n,\theta,\sigma}(u) \;=\; P_{n,\theta,\sigma}(\sqrt{n}(\tilde{\theta}-\theta)\;\le\; u) \tag{25}$$

with $u\in\mathbf{R}^P$ denote the unconditional cdf. Using (18) and (20) we obtain

$$F_{n,\theta,\sigma}(u) \;=\; \sum_{p=0}^{P} F_{n,\theta,\sigma}(u[p]+\sqrt{n}(\theta[p]-\mu_n(p))|p)\mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p]+\sqrt{n}\theta[\neg p])\pi_{n,\theta,\sigma}(p) \tag{26}$$

where we use the convention that $F_{n,\theta,\sigma}(\cdot|0)\equiv 1$. Note that the selection probabilities $\pi_{n,\theta,\sigma}(p)$ for $p\ge 1$ are precisely given by the denominator in (24) (or (22)); for

$p = 0$ this formula also gives $\pi_{n,\theta,\sigma}(0)$ if we use the convention that the first factor in these integrals is set equal to 1. Consequently, (26) can be written as

$$\begin{aligned}
F_{n,\theta,\sigma}(u) &= \mathbf{1}_{\mathbf{R}_+^P}(u+\sqrt{n}\theta)\int_0^\infty \prod_{j=1}^P \Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, sc_j)h(s)ds + \\
&\quad \sum_{p=1}^P \mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p]+\sqrt{n}\theta[\neg p])\cdot \\
&\qquad \int_0^\infty P_{n,\theta,\sigma}(\sqrt{n}(\hat{\theta}(p)-\theta[p]) \le u[p],\, |T_p^*| \ge sc_p)\cdot \\
&\qquad\quad \prod_{j=p+1}^P \Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, sc_j)h(s)ds \\
&= \sum_{p=0}^P \mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p]+\sqrt{n}\theta[\neg p])\cdot \\
&\qquad \int_0^\infty \int_{-\infty}^{u_1}\cdots\int_{-\infty}^{u_p} \phi_{n,p}(r)1_{U(p,sc_p)}(r_p)dr\cdot \\
&\qquad\quad \prod_{j=p+1}^P \Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}sc_j)h(s)ds,
\end{aligned} \tag{27}$$

with the convention that the integral w.r.t. r in (27) is to be interpreted as 1 when $p = 0$.

4 Asymptotic Results for Conditional Distributions

In the previous section the finite-sample distributions of the post-model-selection estimator $\tilde{\theta}$ and of its idealized version $\tilde{\theta}^*$ conditional on selecting a model of order p have been obtained. In the known-variance case the resulting formulas are quite simple and easy to interpret, but this is not so in the unknown-variance case; compare (12) with (23)-(24). In the latter case one could thus think of achieving simplification by passing to the large-sample limit. Assuming that $X'X/n$ converges to a positive definite matrix Q, the results in Pötscher (1991) imply that in case $p \ge p_0(\theta)$ and $p > 0$ the conditional finite-sample distribution of $\tilde{\theta}$ (and also of $\tilde{\theta}^*$) defined in (20) (and (9)) converges to a limiting distribution $F_{\infty,\theta,\sigma}(\cdot\,|\,p)$ with a density $f_{\infty,\theta,\sigma}(\cdot\,|\,p)$ given by

$$f_{\infty,\theta,\sigma}(t\,|\,p) \;=\; \phi_{\infty,p}(t)\,\frac{\mathbf{1}_{\mathbf{R}\setminus(-\sigma\xi_{\infty,p}c_p,\sigma\xi_{\infty,p}c_p)}(t_p)}{1-\Delta(0,c_p)} \tag{28}$$

if $p > p_0(\theta)$, i.e., if $\theta \in M_{p-1}$, and by

$$f_{\infty,\theta,\sigma}(t\,|\,p) \;=\; \phi_{\infty,p}(t) \tag{29}$$

if $p = p_0(\theta)$, i.e., if $\theta \in M_p \setminus M_{p-1}$. Here $\xi_{\infty,p}$ is the limit of $\xi_{n,p}$ as $n \to \infty$. The asymptotic distribution given by $f_{\infty,\theta,\sigma}(\cdot\,|\,p)$ in (28)-(29) is certainly simpler than the conditional finite-sample distribution of $\tilde{\theta}$ given by $f_{n,\theta,\sigma}(\cdot\,|\,p)$ in (23)-(24). To gain insight into the accuracy of $f_{\infty,\theta,\sigma}(\cdot\,|\,p)$ as an approximation to $f_{n,\theta,\sigma}(\cdot\,|\,p)$, we first compare $f_{\infty,\theta,\sigma}(\cdot\,|\,p)$ with the conditional finite-sample density $f^*_{n,\theta,\sigma}(\cdot\,|\,p)$ of the idealized version $\tilde{\theta}^*$: As discussed after (13), in case $p > p_0(\theta)$, i.e., $\theta \in M_{p-1}$, the asymptotic distribution (28) is identical to its finite-sample counterpart given by (13), except that $X[p]'X[p]/n$ and $\xi_{n,p}$ have been replaced by their respective limits; furthermore, (13) is independent of θ. As a consequence, the accuracy of (28) as an approximation to (13) is only governed by the speed of convergence of $X'X/n$ to its limit Q and, in particular, is uniform over all $\theta \in M_{p-1}$. In case $p = p_0(\theta)$, i.e., $\theta \in M_p \setminus M_{p-1}$, the asymptotic expression in (29) is somewhat simpler than (12), but at the expense of missing an essential feature of the finite-sample distribution, namely the interval of excision in (12). The reason for this is that in case $p = p_0(\theta)$ the center of the excision interval $-\sqrt{n}\mu_{n,p}(p)$ equals

$-\sqrt{n}\theta_p$, which diverges to $\pm\infty$ (regardless of how small θ_p is) while the interval's width stays bounded. Thus the excision interval "disappears" at $\pm\infty$ in the limit. (Note that in contrast the excision interval is centered at zero in case $p > p_0(\theta)$.) As a consequence, the asymptotic distribution (29) will poorly approximate the corresponding conditional finite-sample distribution of $\tilde{\theta}^*$ in case θ_p is of order $1/\sqrt{n}$. In particular, convergence of the corresponding cdfs is not uniform for all $\theta \in M_p \setminus M_{p-1}$, cf. Corollary 4.6. (See also Remark 4(iii) in Pötscher (1991) and the related discussion in Kabaila (1995) and Pötscher (1995).) Both phenomena just discussed, namely uniformity over M_{p-1} in the convergence to the asymptotic distribution and non-uniformity over $M_p \setminus M_{p-1}$, can also be expected to occur for $\tilde{\theta}$ instead of $\tilde{\theta}^*$ since the conditional finite-sample distribution of $\tilde{\theta}$ is essentially a "smoothed" version of the corresponding distribution of $\tilde{\theta}^*$ (cf. (23)-(24) and the attending discussion). This expectation is confirmed by the theoretical results obtained below (cf. (34)-(35) and Corollary 4.6). It is also in agreement with the results of a simulation study reported in Pötscher and Novak (1998), which showed that the conditional finite-sample distribution $F_{n,\theta,\sigma}(\cdot \mid p)$ of $\tilde{\theta}$ is reasonably well approximated by its asymptotic counterpart in case $p > p_0(\theta)$, i.e., $\theta \in M_{p-1}$, while this approximation can be quite poor if $p = p_0(\theta)$, i.e., if $\theta \in M_p \setminus M_{p-1}$.

The just discussed "non-uniformity" defect of the asymptotic distribution as an approximation to the conditional finite-sample distribution of $\tilde{\theta}$ may prompt one to look for alternative (asymptotic) approximations that retain the same level of simplicity as the asymptotic distribution without suffering from its defects. It turns out that such asymptotic approximations can indeed be found if we allow these approximating distributions to depend – contrary to (28) and (29) – on sample size. In fact, the distributions given by $f^*_{n,\theta,\sigma}(\cdot \mid p)$ in (12) turn out to be the appropriate approximations. They are as simple and easy to interpret as $f_{\infty,\theta,\sigma}(\cdot \mid p)$ given in (28) and (29), but they will be shown in Theorem 4.2 to approximate the finite-sample distributions given by $f_{n,\theta,\sigma}(\cdot \mid p)$ in (23)-(24) uniformly over (suitable large subsets of) the parameter space. In particular, the unpleasant restriction $p \geq p_0(\theta)$, i.e., $\theta \in M_p$, underlying the asymptotic results in Pötscher (1991) can be removed entirely in the linear model framework considered in the present paper. As a by-product, we then also obtain in Corollary 4.5 the limiting conditional distribution even for the case $p < p_0(\theta)$, i.e., $\theta \notin M_p$.

4.1 Uniform Asymptotic Closeness of the Conditional Distributions $F_{n,\theta,\sigma}(\cdot \mid p)$ and $F^*_{n,\theta,\sigma}(\cdot \mid p)$

We first recall a few facts about distances between distribution functions F and G on $\mathbf{R}^p$. The total variation distance is given by $||F - G||_{TV} = \sup_A |\int_A dF - \int_A dG|$ where the sup is taken over all Borel sets A in $\mathbf{R}^p$. Certainly, $\sup_{t \in \mathbf{R}^p} |F(t) - G(t)| \leq ||F - G||_{TV}$ holds, and hence convergence in total variation distance implies weak convergence in particular. If λ is a dominating (σ-finite) measure for F and G (which always exists), then $||F - G||_{TV} = \frac{1}{2} \int |f - g| d\lambda$, where f and g are the densities of F and G w.r.t. λ. As a consequence, convergence in total variation distance is equivalent to convergence of the corresponding density functions in the L_1-sense. In the following we shall only be concerned with the case where the dominating measure λ is Lebesgue-measure. The L_1-distance between f and g w.r.t. Lebesgue-measure will be denoted by $||f - g||_1$.

The following preliminary result is instrumental in allowing the transfer of results (like the uniformity/non-uniformity phenomena discussed above) from the known-variance case to the unknown-variance case.

Proposition 4.1 *For every p, $1 \leq p \leq P$, we have*

$$\sup_{\theta \in M_p} \sup_{\sigma > 0} ||f_{n,\theta,\sigma}(\cdot \mid p) - f^*_{n,\theta,\sigma}(\cdot \mid p)||_1 \longrightarrow 0 \tag{30}$$

for $n \to \infty$. As a consequence,

$$\sup_{\theta \in M_p} \sup_{\sigma > 0} \sup_{t \in \mathbf{R}^p} |F_{n,\theta,\sigma}(t \mid p) - F^*_{n,\theta,\sigma}(t \mid p)| \longrightarrow 0 \tag{31}$$

for $n \to \infty$.

As argued in the introduction to this section, the conditional distribution of $\tilde{\theta}^*$ is close to its asymptotic counterpart uniformly for $\theta \in M_{p-1}$. In fact, it is easy to see that

$$\sup_{\theta \in M_{p-1}} \sup_{\sigma > 0} \left|\left| f^*_{n,\theta,\sigma}(\cdot|p) - f_{\infty,\theta,\sigma}(\cdot|p) \right|\right|_1 \longrightarrow 0, \tag{32}$$

and hence

$$\sup_{\theta \in M_{p-1}} \sup_{\sigma > 0} \sup_{t \in \mathbf{R}^p} \left| F^*_{n,\theta,\sigma}(t|p) - F_{\infty,\theta,\sigma}(t|p) \right| \longrightarrow 0 \tag{33}$$

for $n \to \infty$.[4] Proposition 4.1 now immediately allows one to transfer this result from the known-variance case to the unknown-variance case, yielding

$$\sup_{\theta \in M_{p-1}} \sup_{\sigma > 0} ||f_{n,\theta,\sigma}(\cdot|p) - f_{\infty,\theta,\sigma}(\cdot|p)||_1 \longrightarrow 0, \tag{34}$$

and hence

$$\sup_{\theta \in M_{p-1}} \sup_{\sigma > 0} \sup_{t \in \mathbf{R}^p} |F_{n,\theta,\sigma}(t|p) - F_{\infty,\theta,\sigma}(t|p)| \longrightarrow 0 \tag{35}$$

for $n \to \infty$. Similarly, the non-uniformity in the closeness of the conditional distribution of $\tilde{\theta}^*$ and its asymptotic counterpart when θ varies in $M_p \setminus M_{p-1}$, outlined in the introduction to this section, can be transfered to the unknown-variance case by means of Proposition 4.1. For a precise statement see Corollary 4.6 below.

A further important implication of Proposition 4.1 is that $f^*_{n,\theta,\sigma}(\cdot \mid p)$ provides a better approximation to $f_{n,\theta,\sigma}(\cdot \mid p)$ than the asymptotic distribution given by (28)-(29) in that (30) holds uniformly for $\theta \in M_p$, whereas (34) holds only uniformly for $\theta \in M_{p-1}$. Despite its usefulness, Proposition 4.1 suffers from the limitation that θ is confined to M_p, and hence this result does not tell us anything about closeness of $f^*_{n,\theta,\sigma}(\cdot|p)$ and $f_{n,\theta,\sigma}(\cdot|p)$ if the true value of θ falls outside of M_p, i.e., if $p_0(\theta) > p$. It would therefore be advantageous to have available a similar result but without any restriction on $p_0(\theta)$ and, in particular, without confining θ to M_p. This is indeed possible at the expense of a more difficult proof.

Theorem 4.2 *For every p, $1 \leq p \leq P$, any bounded subset C of $\mathbf{R}^P$ and any positive constant $\sigma_* > 0$ we have*

$$\sup_{\theta \in C} \sup_{\sigma \geq \sigma_*} \left|\left| f_{n,\theta,\sigma}(\cdot \mid p) - f^*_{n,\theta,\sigma}(\cdot \mid p) \right|\right|_1 \longrightarrow 0$$

for $n \to \infty$. As a consequence,

$$\sup_{\theta \in C} \sup_{\sigma \geq \sigma_*} \sup_{t \in \mathbf{R}^p} \left| F_{n,\theta,\sigma}(t \mid p) - F^*_{n,\theta,\sigma}(t \mid p) \right| \longrightarrow 0$$

for $n \to \infty$.

[4] Subject to $\theta \in M_{p-1}$, the densities $f^*_{n,\theta,\sigma}(\cdot|p)$ and $f_{\infty,\theta,\sigma}(\cdot|p)$ do in fact not depend on θ. Furthermore, it is easy to see that the L_1-distance in (32) is independent of σ. Result (32) now follows from λ-a.e. pointwise (in t) convergence and Scheffe's theorem. (Cf. also the proof of Proposition 4.3.)

The significance of the above theorem is to establish that the easily interpretable formula (12) for $f^*_{n,\theta,\sigma}(\cdot|p)$ indeed provides a *uniform* (w.r.t. θ, σ) approximation to (23)-(24), the conditional density of the post-model-selection estimator $\tilde{\theta}$, over large subsets of the entire parameter space.[5] The uniformity over M_p in the approximation of $f_{n,\theta,\sigma}(\cdot|p)$ by $f^*_{n,\theta,\sigma}(\cdot|p)$ deduced earlier from Proposition 4.1, which has led us to conclude that $f^*_{n,\theta,\sigma}(\cdot|p)$ is superior to $f_{\infty,\theta,\sigma}(\cdot|p)$ as an approximation, is thus extended by Theorem 4.2 to a range of θ's outside of M_p, a range for which the asymptotic distributional results of Pötscher (1991) do not even apply. In fact, as shown in the next subsection, this limitation of the results in Pötscher (1991) can be lifted in the linear model framework considered here, i.e., the limiting conditional distribution even in case $\theta \notin M_p$ can be obtained. Theorem 4.2 will also be instrumental in that development, as it will allow one to transfer the rather easily obtainable results for the known-variance case to the more complicated unknown-variance case.

4.2 Limiting Conditional Distributions

The limiting conditional distributions of the post-model-selection estimator $\tilde{\theta}$ and of its idealized version $\tilde{\theta}^*$ will be obtained next for *all* values of p and not only under the restriction $p \geq p_0(\theta)$ as in Pötscher (1991). Such a complete description of the limiting behaviour of the conditional distributions in the linear model framework also rests on Theorem 4.2. (We repeat, however, that the limiting conditional distributions do not provide a uniform approximation to the finite-sample distributions over the complement of M_{p-1}, cf. Corollary 4.6.) The stepping stone to the results in this subsection is the following proposition.

Proposition 4.3 *Let p satisfy $1 \leq p \leq P$, and let $\theta^{(n)} \in \mathbf{R}^P$ be a sequence such that $n^{1/2}\mu_{n,p}(p)$ converges to a limit $\nu_p \in \mathbf{R} \cup \{-\infty, \infty\}$, where $\mu_n(p)$ is defined by (6) with $\theta^{(n)}$ replacing θ. Furthermore, let $\sigma^{(n)}$ be a sequence of positive real numbers with limit $\sigma > 0$ and recall that $\xi_{\infty,p}$ is the large sample limit of $\xi_{n,p}$. Then:*

(a) The conditional density $f^*_{n,\theta^{(n)},\sigma^{(n)}}(t\,|\,p)$ converges to the density

$$\phi_{\infty,p}(t)\,\frac{\mathbf{1}_{\mathbf{R}\setminus(-\nu_p-\sigma\xi_{\infty,p}c_p,\,-\nu_p+\sigma\xi_{\infty,p}c_p)}(t_p)}{1-\Delta(-\nu_p\sigma^{-1}\xi^{-1}_{\infty,p},c_p)} \tag{36}$$

for λ-almost all $t \in \mathbf{R}^p$, and hence in the L_1-sense. (Note that (36) reduces to the normal density $\phi_{\infty,p}$ in (29) in case $\nu_p = \pm\infty$, and to (28) in case $\nu_p = 0$.) Consequently, the conditional cdf $F^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ converges to the cdf with density (36) in the total variation distance.

(b) If, additionally, $\theta^{(n)}$ is bounded, also the conditional density $f_{n,\theta^{(n)},\sigma^{(n)}}(t\,|\,p)$ converges to (36) in the L_1-sense, and hence the conditional cdf $F_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ converges to the cdf with density (36) in the total variation distance.

We note that (36) is an excised normal density, i.e., (36) is the density of $P(U \leq t,\, |U_p + \nu_p| \geq \sigma\xi_{\infty,p}c_p)/P(|U_p+\nu_p| \geq \sigma\xi_{\infty,p}c_p)$ where U is a $p \times 1$ random vector with cdf $\Phi_{\infty,p}$.

[5] The restriction of θ to a bounded set C is of no real significance and in particular implies local uniformity. Perhaps it could be removed by another method of proof. In this context we note that the boundedness assumption can indeed be avoided for the corresponding result concerning the unconditional distributions, cf. Theorem 5.2

Remark 4.4 (i) If $n^{1/2}\mu_{n,p}(p)$ does not converge, Proposition 4.3 can be applied to subsequences of $\theta^{(n)}$ along which $n^{1/2}\mu_{n,p}(p)$ converges. Since $\mathbf{R}\cup\{-\infty,\infty\}$ is compact, and hence every subsequence contains a convergent subsequence, Proposition 4.3 in fact shows for *any* sequence $\theta^{(n)}$ that every accumulation point of the sequence $f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ w.r.t. the L_1-distance necessarily is of the form (36) (i.e., any accumulation point of the sequence $F^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ w.r.t. the total variation distance has a density of the form (36)). The same holds for $f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $F_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ provided the sequence $\theta^{(n)}$ is bounded.

(ii) If $\theta^{(n)} \in M_p$ for all n, the boundedness assumption in part (b) of Proposition 4.3 can be dropped. The proof remains unchanged, except that Proposition 4.1 instead of Theorem 4.2 has to be used.

The limiting conditional distributions of the post-model-selection estimator $\tilde{\theta}$ and of its idealized version $\tilde{\theta}^*$ under fixed parameter as well as local alternatives asymptotics follow now easily. In the following $Q = \lim_{n\to\infty} X'X/n$ is partitioned as

$$Q = \begin{pmatrix} Q[p:p] & Q[p:\neg p] \\ Q[\neg p:p] & Q[\neg p:\neg p] \end{pmatrix}$$

where $Q[p:p]$ is of dimension $p\times p$.

Corollary 4.5 *Let p satisfy $1\le p\le P$. For $\theta\in\mathbf{R}^P$ consider the local alternatives given by $\theta^{(n)} = \theta+\gamma/\sqrt{n}$ where $\gamma\in\mathbf{R}^P$. Let $\sigma^{(n)}$ be a sequence of positive numbers converging to $\sigma>0$.*

(a) Suppose $\theta\in M_{p-1}$ is satisfied, i.e., $p>p_0(\theta)$ holds. Then $f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ both converge to (36) in the L_1-sense, where

$$\nu_p = \gamma_p + \left((Q[p:p])^{-1}Q[p:\neg p]\gamma[\neg p]\right)_p. \tag{37}$$

As a consequence, the conditional cdfs $F_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $F^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ converge to an excised normal distribution, namely to the cdf of (36), in the total variation distance, and hence uniformly in $t\in\mathbf{R}^p$.

(b) Suppose $\theta\in M_p\setminus M_{p-1}$ is satisfied, i.e., $p=p_0(\theta)$ holds. Then $f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ both converge to the normal density $\phi_{\infty,p}$ in the L_1-sense. As a consequence, the conditional cdfs $F_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $F^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ converge to the normal cdf $\Phi_{\infty,p}$ in the total variation distance, and hence uniformly in $t\in\mathbf{R}^p$.

(c) Suppose $\theta\in\mathbf{R}^P\setminus M_p$ is satisfied, i.e., $p<p_0(\theta)$ holds. Suppose

$$\theta_p + \left((Q[p:p])^{-1}Q[p:\neg p]\theta[\neg p])\right)_p \tag{38}$$

is non-zero (which is the case for all θ except for the elements of a hyperplane). Then $f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ both converge to the normal density $\phi_{\infty,p}$ in the L_1-sense. As a consequence, the conditional cdfs $F_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $F^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ converge to the normal cdf $\Phi_{\infty,p}$ in the total variation distance, and hence uniformly in $t\in\mathbf{R}^p$.

The limiting density function obtained in Corollary 4.5 will be denoted by $f_{\infty,\theta,\sigma,\gamma}(\cdot\,|\,p)$ and the corresponding cdf by $F_{\infty,\theta,\sigma,\gamma}(\cdot\,|\,p)$. In case $\gamma=0$ we shall write $f_{\infty,\theta,\sigma}(\cdot\,|\,p)$ and $F_{\infty,\theta,\sigma}(\cdot\,|\,p)$ as shorthand for $f_{\infty,\theta,\sigma,0}(\cdot\,|\,p)$ and

$F_{\infty,\theta,\sigma,0}(\cdot\,|\,p)$. Note that Corollary 4.5 only ensures well-definedness of $f_{\infty,\theta,\sigma,\gamma}(\cdot\,|\,p)$ and $F_{\infty,\theta,\sigma,\gamma}(\cdot\,|\,p)$ for all $\theta \in M_p$, as well as for those $\theta \in \mathbf{R}^P \setminus M_p$ making (38) non-zero.

Parts (a) and (b) of Corollary 4.5 in case $\gamma = 0$ (i.e., fixed parameter asymptotics) reproduce the asymptotic distributions given in (28) and (29) obtained from the results in Pötscher (1991).[6] Part (c) complements these results by considering the case $\theta \notin M_p$, a case excluded from the analysis in Pötscher (1991).

Corollary 4.5 (and Proposition 4.3) show that the limiting behaviour of the distribution of the post-model-selection estimator $\tilde{\theta}$ (and of $\tilde{\theta}^*$), conditional on having selected model order p, is mainly governed by the behaviour of $\sqrt{n}\mu_{n,p}(p)$ given by (6) with $\theta^{(n)}$ replacing θ. In discussing the implications of the above corollary, we consider first the case $\theta \in M_p$ corresponding to parts (a) and (b). For simplicity of presentation, assume also that $\gamma \in M_p$, i.e., $\gamma[\neg p] = 0$.[7] The crucial quantity $\sqrt{n}\mu_{n,p}(p)$ then reduces to $\sqrt{n}(\theta_p + \gamma_p/\sqrt{n})$ and one immediately sees that it is the p-th component of $\theta^{(n)} = \theta + \gamma/\sqrt{n}$ that determines whether the limiting distribution is a normal (if $\theta_p \neq 0$) or an excised normal distribution (if $\theta_p = 0$). This dichotomy is exactly the same as in the case of fixed parameter asymptotics, but now in case of an excised normal distribution the center of excision is not necessarily zero but is given by $-\gamma_p$. As discussed earlier, the shape of the finite-sample distribution of $\tilde{\theta}$ (or $\tilde{\theta}^*$), conditional on having selected model order p, will – in case of a true parameter value in M_p close to but outside of M_{p-1} – be badly captured by the shape of the limiting conditional distribution obtained from fixed parameter asymptotics with $\theta^{(n)} \equiv \theta \in M_p \setminus M_{p-1}$, regardless of how close θ is to M_{p-1}. However, if this case is modelled in a local alternatives framework by taking the true parameter value $\theta^{(n)}$ to satisfy $\theta^{(n)} \in M_p \setminus M_{p-1}$ but approaching M_{p-1} (i.e., $\theta^{(n)} = \theta + \gamma/\sqrt{n}$ with $\theta \in M_{p-1}$, $\gamma_p \neq 0$), then the limiting conditional distribution will better capture the essential features of the finite-sample distribution by predicting an excised normal distribution rather than a normal distribution.

We next turn to the case $\theta \notin M_p$ corresponding to part (c) or Corollary 4.5. Observe that $\sqrt{n}\mu_{n,p}(p)$ is given by

$$
\begin{aligned}
\sqrt{n}\mu_{n,p}(p) = \sqrt{n}\left(\theta_p + \left(\left(\frac{X[p]'X[p]}{n}\right)^{-1}\frac{X[p]'X[\neg p]}{n}\theta[\neg p]\right)_p\right) \\
+\ \gamma_p\ + \left(\left(\frac{X[p]'X[p]}{n}\right)^{-1}\frac{X[p]'X[\neg p]}{n}\gamma[\neg p]\right)_p .
\end{aligned}
\tag{39}
$$

If the expression in (38) is non-zero, $\sqrt{n}\mu_{n,p}(p)$ tends to $\pm\infty$ resulting in a limiting normal distribution. Otherwise (without additional assumptions on the regressors), $\sqrt{n}\mu_{n,p}(p)$ can show all kinds of limiting behaviour, including oscillatory behaviour, in which case no limiting distribution for $F_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $F^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ exists. Of course, along convergent subsequences of $\sqrt{n}\mu_{n,p}(p)$ the limiting distribution follows from Proposition 4.3, and hence will be an excised normal distribution or a normal distribution, depending on whether the limit of the subsequence of $\sqrt{n}\mu_{n,p}(p)$ is finite or $\pm\infty$, cf. Remark 4.4.

[6] The convergence results for the conditional cdfs $F_{n,\theta,\sigma}(\cdot\,|\,p)$ in Pötscher (1991) are formulated in terms of pointwise convergence, i.e., for each fixed $t \in \mathbf{R}^p$. Since the limiting cdfs are continuous (when viewed on the appropriate subspace), uniformity w.r.t. $t \in \mathbf{R}^p$, however, follows automatically from Polya's theorem, cf. Billingsley and Topsoe (1967, Ex. 6) and Chandra (1989).

[7] This assumption is rather innocuous, as the general case $\gamma \in \mathbf{R}^P$ is completely similar.

We conclude this section by a formal statement of the non-uniformity phenomenon in the convergence of the conditional finite-sample distributions to their limits. As a starting point, recall that the conditional finite-sample densities $f^*_{n,\theta,\sigma}(\cdot\,|\,p)$ as well as $f_{n,\theta,\sigma}(\cdot\,|\,p)$ converge to their limit $f_{\infty,\theta,\sigma}(\cdot\,|\,p)$ in the L_1-sense uniformly in θ and σ as long as θ varies in M_{p-1} only and $\sigma > 0$, cf. (32)–(35). However, in contrast to the result given in Theorem 4.2, uniformity in the convergence to the limiting density $f_{\infty,\theta,\sigma}(\cdot\,|\,p)$ (or to the cdf $F_{\infty,\theta,\sigma}(\cdot\,|\,p)$) does not hold when θ varies in (bounded subsets of) the entire parameter space (or even only in M_p). In fact, uniformity already breaks down locally near M_{p-1}.

Corollary 4.6 *Let p satisfy $1 \leq p \leq P$. For every $t \in \mathbf{R}^p$ and every $\sigma > 0$ there exists a positive $\rho \in \mathbf{R}$ such that for every $\theta \in M_{p-1}$*

$$\liminf_{n\to\infty} \sup_{\overline{\theta}\in B_p(\theta,\rho/\sqrt{n})} \left|F_{n,\overline{\theta},\sigma}(t\,|\,p) - F_{\infty,\overline{\theta},\sigma}(t\,|\,p)\right| \;>\; 0, \tag{40}$$

where $B_p(\theta,\rho/\sqrt{n})$ denotes the ball $\{\overline{\theta} \in M_p : ((\overline{\theta}-\theta)'(\overline{\theta}-\theta))^{1/2} < \rho/\sqrt{n}\}$. Consequently,

$$\liminf_{n\to\infty} \sup_{\overline{\theta}\in B_p(\theta,\rho/\sqrt{n})} ||f_{n,\overline{\theta},\sigma}(\cdot\,|\,p) - f_{\infty,\overline{\theta},\sigma}(\cdot\,|\,p)||_1 \;>\; 0.$$

The same is also true if $F_{n,\overline{\theta},\sigma}(\cdot\,|\,p)$ and $f_{n,\overline{\theta},\sigma}(\cdot\,|\,p)$ are replaced by $F^*_{n,\overline{\theta},\sigma}(\cdot\,|\,p)$ and $f^*_{n,\overline{\theta},\sigma}(\cdot\,|\,p)$, respectively.

Corollary 4.6 shows in particular that convergence to the limiting distribution is not uniform over tubes in M_p that surround M_{p-1}. It is easy to show that uniformity in fact holds in the complement (relative to M_p) of any such tube (of fixed positive radius). Hence, within M_p the non-uniformity arises precisely near M_{p-1}.[8]

5 Asymptotic Results for the Unconditional Distribution and for the Selection Probabilities

In this section we study the asymptotic properties of the unconditional distribution functions of the post-model-selection estimator $\tilde{\theta}$ and of its idealized version $\tilde{\theta}^*$, and obtain results paralleling those in Section 4. In particular, we find again that the limiting distribution does not provide a uniform approximation to the unconditional finite-sample distributions $F_{n,\theta,\sigma}(u)$ and $F^*_{n,\theta,\sigma}(u)$; cf. Corollary 5.5. Also, uniform closeness of $F_{n,\theta,\sigma}(u)$ and $F^*_{n,\theta,\sigma}(u)$ can again be established; see Theorem 5.2. Analogous results for the selection probabilities $\pi_{n,\theta,\sigma}(p)$ and $\pi^*_{n,\theta,\sigma}(p)$ defined in Section 3 are also obtained.

Since the cdfs $F_{n,\theta,\sigma}$ and $F^*_{n,\theta,\sigma}$, respectively, can be linearly expressed in terms of $F_{n,\theta,\sigma}(t|p)\pi_{n,\theta,\sigma}(p)$ and $F^*_{n,\theta,\sigma}(t|p)\pi^*_{n,\theta,\sigma}(p)$, cf. (26) and (14), a key step for the results in this section is the following lemma.

[8] It is easy to see that the non-uniformity near M_{p-1} typically also arises when approaching M_{p-1} from within $\mathbf{R}^P \setminus M_p$. Furthermore, outside of M_p additional non-uniformity can arise near the hyperplane mentioned in Corollary 4.5(c). Again, outside a tube of fixed positive radius surrounding this hyperplane convergence is uniform.

Lemma 5.1 *For every p, $1 \leq p \leq P$, we have*

$$\sup_{\theta \in \mathbf{R}^P} \sup_{\sigma > 0} \left|\left| f_{n,\theta,\sigma}(\cdot|p)\pi_{n,\theta,\sigma}(p) \; - \; f^*_{n,\theta,\sigma}(\cdot|p)\pi^*_{n,\theta,\sigma}(p) \right|\right|_1 \; \longrightarrow \; 0 \tag{41}$$

for $n \to \infty$. As a consequence,

$$\sup_{\theta \in \mathbf{R}^P} \sup_{\sigma > 0} \sup_{t \in \mathbf{R}^p} \left| F_{n,\theta,\sigma}(t|p)\pi_{n,\theta,\sigma}(p) \; - \; F^*_{n,\theta,\sigma}(t|p)\pi^*_{n,\theta,\sigma}(p) \right| \; \longrightarrow \; 0$$

for $n \to \infty$.

The following results parallel Theorem 4.2.

Theorem 5.2 *(a) For the unconditional distribution functions we have*

$$\sup_{\theta \in \mathbf{R}^P} \sup_{\sigma > 0} \left|\left| F_{n,\theta,\sigma} \; - \; F^*_{n,\theta,\sigma} \right|\right|_{TV} \; \longrightarrow \; 0 \tag{42}$$

and hence

$$\sup_{\theta \in \mathbf{R}^P} \sup_{\sigma > 0} \sup_{u \in \mathbf{R}^P} \left| F_{n,\theta,\sigma}(u) \; - \; F^*_{n,\theta,\sigma}(u) \right| \; \longrightarrow \; 0$$

for $n \to \infty$.

(b) Let p satisfy $0 \leq p \leq P$. The selection probabilities satisfy

$$\sup_{\theta \in \mathbf{R}^P} \sup_{\sigma > 0} \left| \pi_{n,\theta,\sigma}(p) - \pi^*_{n,\theta,\sigma}(p) \right| \; \longrightarrow \; 0$$

for $n \to \infty$.

In contrast to Theorem 4.2, we do not need to restrict θ to a bounded set or to require σ to satisfy $\sigma \geq \sigma_* > 0$. These restrictions arose in the proof of Theorem 4.2 from the need to control the behaviour of ratios of probabilities which vanish asymptotically. In the unconditional case considered in Theorem 5.2 this difficulty does not arise, allowing us to avoid these restrictions. We also note that the cdfs $F_{n,\theta,\sigma}$ and $F^*_{n,\theta,\sigma}$ do not have densities w.r.t. Lebesgue measure on $\mathbf{R}^P$, since both distributions put positive mass on every subspace $M_0, \ldots, M_P$. For this reason, (42) is given in terms of total variation distance rather than in terms of the L_1-distance between density functions.[9]

We next consider the limit of the unconditional distributions of the post-model-selection estimator $\tilde{\theta}$ and of its idealized version $\tilde{\theta}^*$. The following proposition contains the main technical step.

Proposition 5.3 *Let p satisfy $0 \leq p \leq P$, and let $\theta^{(n)} \in \mathbf{R}^P$ be a sequence such that $n^{1/2}\mu_{n,j}(j)$ converges to a limit $\nu_j \in \mathbf{R} \cup \{-\infty, \infty\}$ for $j = \max(1,p), \ldots, P$, where $\mu_n(j)$ is defined by (6) with $\theta^{(n)}$ replacing θ. Furthermore, let $\sigma^{(n)}$ be a sequence of positive real numbers with limit $\sigma > 0$. Then:*

(a) If $p \geq 1$, the quantity $f^*_{n,\theta^{(n)},\sigma^{(n)}}(t|p)\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p)$ converges to

$$\phi_{\infty,p}(t)\mathbf{1}_{\mathbf{R} \setminus (-\nu_p - \sigma\xi_{\infty,p}c_p, -\nu_p + \sigma\xi_{\infty,p}c_p)}(t_p) \prod_{j=p+1}^{P} \Delta(-\nu_j \sigma^{-1}\xi^{-1}_{\infty,j}, c_j) \tag{43}$$

[9] This certainly could be done upon an appropriate choice of dominating measure, which, of course, now cannot be Lebesgue measure.

for λ-almost all $t \in \mathbf{R}^p$, and hence in the L_1-sense. Furthermore, $f_{n,\theta^{(n)},\sigma^{(n)}}(t|p)\pi_{n,\theta^{(n)},\sigma^{(n)}}(p)$ also converges to (43) in the L_1-sense. (In case $\nu_p = \pm\infty$ the indicator function in (43) is to interpreted as the constant 1.) As a consequence, $F^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot|p)\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p)$ and $F_{n,\theta^{(n)},\sigma^{(n)}}(\cdot|p)\pi_{n,\theta^{(n)},\sigma^{(n)}}(p)$ converge to the distribution with density (43) in the total variation distance, and hence uniformly in $t \in \mathbf{R}^p$.[10]

(b) *The selection probabilities $\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p)$ and $\pi_{n,\theta^{(n)},\sigma^{(n)}}(p)$ converge to*

$$(1 - \Delta(-\nu_p \sigma^{-1}\xi^{-1}_{\infty,p}, c_p)) \prod_{j=p+1}^{P} \Delta(-\nu_j \sigma^{-1}\xi^{-1}_{\infty,j}, c_j) \tag{44}$$

where we use the convention that in case $p = 0$ the first factor in (44) is set equal to one. (Note that (43) and (44) are zero, whenever $|\nu_j| = \infty$ for some j satisfying $p + 1 \le j \le P$.)

(c) *If $|\nu_j| = \infty$ for some $p \le j \le P$, then $\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(r)$ and $\pi_{n,\theta^{(n)},\sigma^{(n)}}(r)$ converge to zero for $0 \le r < p$.*

We note that Remark 4.4 following Proposition 4.3 applies to Proposition 5.3 as well. With the help of Proposition 5.3 we can now obtain the unconditional asymptotic distribution under fixed as well as local alternatives asymptotics.

Corollary 5.4 *For $\theta \in \mathbf{R}^P$ consider the local alternatives given by $\theta^{(n)} = \theta + \gamma/\sqrt{n}$ where $\gamma \in \mathbf{R}^P$, and let $\sigma^{(n)}$ be a sequence of positive numbers with limit $\sigma > 0$.*

(a) *For $0 \le p \le P$, the selection probabilities $\pi_{n,\theta^{(n)},\sigma^{(n)}}(p)$ and $\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p)$ converge to $\pi_{\infty,\theta,\sigma,\gamma}(p)$ which is given by*

$$\begin{aligned} &\pi_{\infty,\theta,\sigma,\gamma}(p) \\ &= \begin{cases} 0 & \text{if } p < p_0(\theta), \\ \prod_{j=p+1}^{P} \Delta(-\upsilon_j \sigma^{-1}\xi^{-1}_{\infty,j}, c_j) & \text{if } p = p_0(\theta), \\ (1 - \Delta(-\upsilon_p \sigma^{-1}\xi^{-1}_{\infty,p}, c_p)) \prod_{j=p+1}^{P} \Delta(-\upsilon_j \sigma^{-1}\xi^{-1}_{\infty,j}, c_j) & \text{if } p > p_0(\theta) \end{cases} \end{aligned} \tag{45}$$

with

$$\upsilon_j \;=\; \gamma_j \;+\; \left((Q[j:j])^{-1}Q[j:\neg j]\gamma[\neg j]\right)_j . \tag{46}$$

(b) *The unconditional finite-sample distributions $F_{n,\theta^{(n)},\sigma^{(n)}}$ and $F^*_{n,\theta^{(n)},\sigma^{(n)}}$ converge to $F_{\infty,\theta,\sigma,\gamma}$ in the total variation distance (and hence uniformly in $u \in \mathbf{R}^P$). The limit $F_{\infty,\theta,\sigma,\gamma}$ satisfies*

$$\begin{aligned} F_{\infty,\theta,\sigma,\gamma}(u) = \sum_{p=p_0(\theta)}^{P} & F_{\infty,\theta,\sigma,\gamma}(u[p] - (Q[p:p])^{-1}Q[p:\neg p]\gamma[\neg p]|p) \cdot \\ & \pi_{\infty,\theta,\sigma,\gamma}(p)\mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \gamma[\neg p]), \end{aligned} \tag{47}$$

where $F_{\infty,\theta,\sigma,\gamma}(\cdot|p)$ was introduced after Corollary 4.5 for $p \ge 1$ and is set equal to 1 for $p = 0$.

[10] The total variation distance for distribution functions with finite mass (not necessarily equal to one) is defined exactly as in the case where total mass is one. In this more general context the inequality $\sup_t |F(t) - G(t)| \le ||F - G||_{TV} \le \int |f - g| d\lambda$ still holds, where f and g, respectively, denote densities of F and G w.r.t. a dominating (σ-finite) measure λ.

The limiting distribution (47) is a convex combination of excised and non-excised normal distributions concentrated on subspaces of different dimensions. In contrast to the unconditional finite-sample distribution, cf. (17) and (27), however, only terms with $p \geq p_0(\theta)$ enter into the limiting distribution due to the fact that the selection probabilities for models with $p < p_0(\theta)$ converge to zero.[11]

In case $\gamma = 0$ (i.e., fixed parameter asymptotics), the limiting distribution in Corollary 5.4 reduces to the limiting distribution one obtains by applying the results in Pötscher (1991) to the linear model considered here. We shall write $F_{\infty,\theta,\sigma}$ and $\pi_{\infty,\theta,\sigma}(p)$, respectively, for $F_{\infty,\theta,\sigma,0}$ and $\pi_{\infty,\theta,\sigma,0}(p)$. In case $\gamma = 0$ the limiting distribution in Corollary 5.4 reduces to

$$
\begin{aligned}
F_{\infty,\theta,\sigma}(u) &= \sum_{p=p_0(\theta)}^{P} F_{\infty,\theta,\sigma}(u[p] \mid p)\pi_{\infty,\theta,\sigma}(p)\,\mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p]) \\
&= \int_{-\infty}^{u_1}\cdots\int_{-\infty}^{u_{p_0(\theta)}} \phi_{\infty,p_0(\theta)}(r)\,dr \prod_{j=p_0(\theta)+1}^{P} \Delta(0,c_j)\,\mathbf{1}_{\mathbf{R}_+^{P-p_0(\theta)}}(u[\neg p_0(\theta)]) + \\
&\quad \sum_{p=p_0(\theta)+1}^{P} \int_{-\infty}^{u_1}\cdots\int_{-\infty}^{u_p} \phi_{\infty,p}(r)\mathbf{1}_{\mathbf{R}\setminus(-\sigma\xi_{\infty,p}c_p,\sigma\xi_{\infty,p}c_p)}(r_p)\,dr \cdot \\
&\quad \prod_{j=p+1}^{P} \Delta(0,c_j)\,\mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p])
\end{aligned} \tag{48}
$$

with the convention that the first integral on the r.h.s of (48) is set equal to 1 if $p_0(\theta) = 0$. The asymptotic selection probabilities also simplify to $\pi_{\infty,\theta,\sigma}(p) = \prod_{j=p+1}^{P} \Delta(0,c_j)$ if $p = p_0(\theta)$ and to $\pi_{\infty,\theta,\sigma}(p) = (1-\Delta(0,c_p))\prod_{j=p+1}^{P} \Delta(0,c_j)$ if $p > p_0(\theta)$. Note that the (excised) normal distributions in (48) are now all symmetric about zero and that the asymptotic selection probabilities depend only on the critical values c_j; cf. Lemma 4 in Pötscher (1991).

As in the conditional case discussed in Section 4, the convergence of the unconditional finite-sample distribution to its limit is not uniform w.r.t. θ. Heuristically speaking, there are two sources for this non-uniformity. First, the non-uniformity in the convergence of the conditional distributions (Corollary 4.6), which spills over into the unconditional distribution. Second, the non-uniformity in the convergence of the selection probabilities (established below in Corollary 5.6) taken together with the fact that the limiting distribution (48) does not contain terms with $p < p_0(\theta)$.

Corollary 5.5 *For every $u \in \mathbf{R}^P$ and every $\sigma > 0$ there exists a positive $\rho \in \mathbf{R}$ such that for every $\theta \in M_{P-1}$*

$$
\liminf_{n\to\infty} \sup_{\overline{\theta}\in B_P(\theta,\rho/\sqrt{n})} \left| F_{n,\overline{\theta},\sigma}(u) - F_{\infty,\overline{\theta},\sigma}(u) \right| > 0, \tag{49}
$$

where $B_P(\theta,\rho/\sqrt{n})$ denotes the ball $\{\overline{\theta} \in \mathbf{R}^P : ((\overline{\theta}-\theta)'(\overline{\theta}-\theta))^{1/2} < \rho/\sqrt{n}\}$. The same is also true if $F_{n,\overline{\theta},\sigma}$ is replaced by $F^*_{n,\overline{\theta},\sigma}$.

Corollary 5.5 shows in particular that convergence to the limiting distribution is not uniform over tubes surrounding M_{P-1}. It is easy to show that uniformity in fact holds in the complement of any such tube (of fixed positive radius). Hence, the non-uniformity arises precisely near M_{P-1}. Recall from Corollary 4.6 that the non-uniformity in the convergence of the conditional finite-sample cdf $F_{n,\theta,\sigma}(\cdot|p)$ arises near M_{p-1}. Since the unconditional finite-sample distribution is a convex

[11] For this reason the limit $F_{\infty,\theta,\sigma,\gamma}$ is well-defined for all $\theta \in \mathbf{R}^P$, since $F_{\infty,\theta,\sigma,\gamma}(\cdot|p)$ is well-defined on a set containing all θ satisfying $p \geq p_0(\theta)$.

combination of the conditional distributions, it should not come as a surprise that now non-uniformity arises near the union of all sets M_{p-1} which is precisely M_{P-1}.[12]

We finally turn to a discussion of uniformity and non-uniformity in the convergence of the selection probabilities to their limits. If $\theta \in M_{p-1}$, $1 \leq p \leq P$, we have $\mu_{n,j}(j) = 0$ for $j \geq p$ and hence $\pi^{*}_{n,\theta,\sigma}(p)$ given by (15)–(16) coincides with $\pi_{\infty,\theta,\sigma}(p)$ given by (45) with $\gamma = 0$. Consequently,

$$\sup_{\theta \in M_{p-1}} \sup_{\sigma > 0} \left| \pi^{*}_{n,\theta,\sigma}(p) - \pi_{\infty,\theta,\sigma}(p) \right| \; = \; 0$$

holds for $1 \leq p \leq P$, and an application of Theorem 5.2 gives

$$\sup_{\theta \in M_{p-1}} \sup_{\sigma > 0} \left| \pi_{n,\theta,\sigma}(p) - \pi_{\infty,\theta,\sigma}(p) \right| \; \longrightarrow \; 0$$

as $n \to \infty$ for $1 \leq p \leq P$. In the next result we describe regions of non-uniformity in the convergence of the selection probabilities to their limits.

Corollary 5.6 *(a) Let p satisfy $1 \leq p \leq P$. Then for any positive $\rho \in \mathbf{R}$, any $\sigma > 0$, and any $\theta \in M_{p-1}$*

$$\liminf_{n \to \infty} \sup_{\overline{\theta} \in B_p(\theta, \rho/\sqrt{n})} \left| \pi_{n,\overline{\theta},\sigma}(p) \; - \; \pi_{\infty,\overline{\theta},\sigma}(p) \right| \; > \; 0$$

where $B_p(\theta, \rho/\sqrt{n})$ was defined in Corollary 4.6.

(b) Let p satisfy $0 \leq p < P$. Then for any positive $\rho \in \mathbf{R}$, any $\sigma > 0$, and any $\theta \in M_p$ we have

$$\liminf_{n \to \infty} \sup_{\overline{\theta} \in B_{p+1}(\theta, \rho/\sqrt{n})} \left| \pi_{n,\overline{\theta},\sigma}(p) \; - \; \pi_{\infty,\overline{\theta},\sigma}(p) \right| \; > \; 0.$$

Parts (a) and (b) also hold with $\pi_{n,\overline{\theta},\sigma}(p)$ replaced by $\pi^{*}_{n,\overline{\theta},\sigma}(p)$.

Corollary 5.6(a) shows, in particular, that convergence of the selection probabilities $\pi_{n,\theta,\sigma}(p)$ and $\pi^{*}_{n,\theta,\sigma}(p)$ to their limiting value is not uniform over tubes in M_p that surround M_{p-1}. It is easy to show that over the complement (w.r.t. M_p) of any such tube of fixed positive radius the convergence is in fact uniform. These results parallel the results for the conditional distributions given in Corollary 4.6 and in the attending discussion. In contrast to the results for the conditional distributions, however, Corollary 5.6(b) furthermore shows that non-uniformity in the convergence of $\pi_{n,\theta,\sigma}(p)$ and $\pi^{*}_{n,\theta,\sigma}(p)$ also arises in tubes (in $\mathbf{R}^P$) surrounding M_p. Again, outside any such tube of fixed radius convergence is uniform.

[12] One could furthermore ask if results similar to (32)-(35) hold, i.e., if $F_{n,\theta,\sigma}$ converges to $F_{\infty,\theta,\sigma}$ (either in total variation distance or for every $u \in \mathbf{R}^P$) uniformly over M_{P-1} (or over a smaller set M_r, $r > 0$). Since $F_{n,\theta,\sigma}$ is a convex combination of all the conditional distributions $F_{n,\theta,\sigma}(\cdot|p)$, it should again not come as a surprise that the answer is no. By an argument similar to the proof of Corollary 5.5, the following more general result can be established: Let $q(u) = \min\{i : 0 \leq i \leq P, u_{i+1} > 0, \ldots, u_P > 0\}$ and set $q(u) = P$ if the set over which the minimum is taken is empty. Then (49) holds with $B_P(\theta, \rho/\sqrt{n})$ replaced by any of the smaller sets $B_P(\theta, \rho/\sqrt{n}) \cap (M_r \setminus M_{r-1})$ for any r satisfying $\max\{p_0(\theta) + 1, q(u)\} \leq r \leq P$.

6 Remarks and Extensions

Remark 6.1 (i) In defining the conditional finite-sample distributions $F^*_{n,\theta,\sigma}(\cdot \mid p)$ and $F_{n,\theta,\sigma}(\cdot \mid p)$, cf. (9) and (20), the estimator has been centered at $\mu_n(p)$, the expected value of the restricted least-squares estimator given in (6). For the finite-sample results in Section 3 a different choice of centering constants of course only amounts to a translation of the conditional distributions and is hence inconsequential. Furthermore, in case $p \geq p_0(\theta)$ we have $\mu_n(p) = \theta[p]$ as remarked earlier and thus centering at $\mu_n(p)$ coincides with centering at the true value in this case, which is also the centering used in Pötscher (1991).

(ii) Proposition 4.1 and Theorem 4.2 are in fact totally independent of the particular centering that is being used, since the L_1-distance (w.r.t. Lebesgue-measure) between $f_{n,\theta,\sigma}(\cdot \mid p)$ and $f^*_{n,\theta,\sigma}(\cdot \mid p)$ is not affected by any shift in the argument t. Similarly, the sup-norm of the difference $F_{n,\theta,\sigma}(\cdot \mid p) - F^*_{n,\theta,\sigma}(\cdot \mid p)$ is unaffected by any shift in the argument t. Hence, Proposition 4.1 and Theorem 4.2 hold, in particular, without any centering of the estimator.

(iii) We are next concerned with the question to which extent the limiting results in Section 4.2 are affected by the choice of the centering constants. Let p satisfy $1 \leq p \leq P$ and let $d_{n,\theta}$ denote a $p \times 1$ vector. Then centering at $d_{n,\theta}$ leads to

$$P_{n,\theta,\sigma}\left(\sqrt{n}(\hat{\theta}(p) - d_{n,\theta}) \leq t \,\middle|\, \hat{p} = p\right) \;=\; F_{n,\theta,\sigma}\left(t + \sqrt{n}(d_{n,\theta} - \mu_n(p)) \,\middle|\, p\right). \quad (50)$$

As shown in Proposition 4.3 and Remark 4.4, every subsequence of $F_{n,\theta,\sigma}(\cdot \mid p)$ contains a subsequence that converges in total variation distance. If $\sqrt{n}(d_{n,\theta} - \mu_n(p))$ is bounded, it follows from Lemma D.1 that every accumulation point of (50) w.r.t. total variation distance is a shifted version of a cdf with density (36), the shift being given by the limit of $\sqrt{n}(d_{n,\theta} - \mu_n(p))$ along the appropriate subsequence. If $\sqrt{n}(d_{n,\theta} - \mu_n(p))$ is unbounded, then every subsequence of (50) contains a subsequence that converges to a shifted version of the cdf with density (36) for every t, where the shift is again given by the limit of $\sqrt{n}(d_{n,\theta} - \mu_n(p))$ along the appropriate subsequence. Note that now some of the components of this shift may be $\pm\infty$, implying that the limiting cdf will be degenerate in the sense that at least for one marginal distribution mass will have escaped to ∞ or $-\infty$. The above shows that choosing centering constants other that $\mu_n(p)$ either leads to essentially the same limiting distributions (if $\sqrt{n}(d_{n,\theta} - \mu_n(p))$ is bounded) or to loss of information about the distribution (if $\sqrt{n}(d_{n,\theta} - \mu_n(p))$ is unbounded). Identical remarks apply to $\tilde{\theta}^*$. (A similar discussion applies mutatis mutandis also to the case of asymptotics under a sequence of parameters $\theta^{(n)}$ and, in particular, to asymptotics under local alternatives.)

Remark 6.2 (i) The results concerning the conditional distributions in Sections 3 and 4 are given in terms of $F_{n,\theta,\sigma}(\cdot \mid p)$ and $F^*_{n,\theta,\sigma}(\cdot \mid p)$, which are actually only the conditional distributions of the first p components of the (scaled and centered) post-model-selection estimator and its idealized version, compare (8) with (9) and (19) with (20). Since (8) and (9) as well as (19) and (20) only differ by a multiplicative factor independent of sample size and θ, all the results obtained for $F_{n,\theta,\sigma}(\cdot \mid p)$ and $F^*_{n,\theta,\sigma}(\cdot \mid p)$, respectively, can be trivially translated into results for (19) and (8). We next discuss to which extent this is true if we use centering constants different from the ones used in (8) and (19). Since such a different centering has only a trivial effect on the finite-sample results, we concentrate on the asymptotic results in Section 4. Let $d_{n,\theta}$ now be a $P \times 1$ vector. Then

$$P_{n,\theta,\sigma}(\sqrt{n}(\tilde{\theta} - d_{n,\theta}) \le u \,|\, \hat{p} = p) = F_{n,\theta,\sigma}(u[p] + \sqrt{n}(d_{n,\theta}[p] - \mu_n(p)) \,|\, p) \, \mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \sqrt{n} d_{n,\theta}[\neg p]) \tag{51}$$

and

$$P_{n,\theta,\sigma}(\sqrt{n}(\tilde{\theta}^* - d_{n,\theta}) \le u \,|\, \hat{p}^* = p) = F^*_{n,\theta,\sigma}(u[p] + \sqrt{n}(d_{n,\theta}[p] - \mu_n(p)) \,|\, p) \, \mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \sqrt{n} d_{n,\theta}[\neg p]). \tag{52}$$

Note that the indicator function in (51) and (52) may now depend on n and θ. From Remark 6.1(iii) and from the fact that the indicator functions in (51) and (52) are identical, it immediately follows that Proposition 4.1 and Theorem 4.2 continue to hold upon replacing the L_1-distance between the densities by the total variation distance between (51) and (52). We next turn to the limiting behaviour of (51) (or (52)): The limiting behaviour of the first factor of (51) (or (52)) has been discussed in Remark 6.1(iii), where all accumulation points have been characterized. It now follows immediately that any subsequence of (51) (or (52)) has a further subsequence (along which we may assume that $\sqrt{n} d_{n,\theta}[\neg p]$ converges to a vector $\delta(\theta) \in (\mathbf{R} \cup \{-\infty, \infty\})^{P-p}$) that converges to the limit of the first factor described in Remark 6.1(iii) times the indicator function $\mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \delta(\theta))$ for all $u \in \mathbf{R}^P$, except possibly for those where $u[\neg p] + \delta(\theta)$ belongs to the boundary of $\mathbf{R}_+^{P-p}$. Note that this convergence will in general not hold in the total variation distance, except if $\sqrt{n} d_{n,\theta}[\neg p]$ does not depend on n. (A similar discussion applies mutatis mutandis also to the case of asymptotics under a sequence of parameters $\theta^{(n)}$ and, in particular, to asymptotics under local alternatives.)

(ii) Remarks similar to Remark 6.1 and (i) above also apply to the unconditional distributions of $\tilde{\theta}$ and $\tilde{\theta}^*$. In particular, a complete characterization of all accumulation points of $F_{n,\theta^{(n)},\sigma^{(n)}}(u)$ and $F^*_{n,\theta^{(n)},\sigma^{(n)}}(u)$ w.r.t. weak convergence for arbitrary sequences $\theta^{(n)}$, $\sigma^{(n)}$ is possible, but will not be included here for the sake of brevity.

Remark 6.3 (i) Let $F^{\circ}_{n,\theta,\sigma}(t \,|\, p) = P_{n,\theta,\sigma}(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)) \le t, |T_p| \ge c_p)/P_{n,\theta,\sigma}(|T_p| \ge c_p)$ and let $f^{\circ}_{n,\theta,\sigma}(\cdot \,|\, p)$ denote the corresponding density. Then it is not difficult to show that Proposition 4.1 and Theorem 4.2 continue to hold with $F^{\circ}_{n,\theta,\sigma}(\cdot \,|\, p)$ and $f^{\circ}_{n,\theta,\sigma}(\cdot \,|\, p)$ replacing $F^*_{n,\theta,\sigma}(\cdot \,|\, p)$ and $f^*_{n,\theta,\sigma}(\cdot \,|\, p)$ (or $F_{n,\theta,\sigma}(\cdot|p)$ and $f_{n,\theta,\sigma}(\cdot|p)$), respectively. Analogues to the remaining results in Section 4 can then be easily obtained.

(ii) In a more general setting, Pötscher (1991) obtained the limiting conditional distribution $F_{\infty,\theta,\sigma}(\cdot \,|\, p)$ of the post-model-selection estimator in case $p \ge p_0(\theta)$. The proof in that paper proceeds by first showing that the analogues of $(\sqrt{n}(\hat{\theta}(p) - \mu_n(p)), T_p)$ and $(T_{p+1}, \ldots, T_P)$ are asymptotically independent, i.e., by showing that the tests based on $T_{p+1}, \ldots, T_P$ have no effect asymptotically. Second, the limiting conditional distribution is obtained by computing the corresponding conditional distribution from the limiting distribution. This method of proof relies on interchanging the operations of taking limits and of conditioning. It is precisely the assumption $p \ge p_0(\theta)$ that implies that the conditioning event has a positive probability in the limit and thus allows for conditioning in the limiting distribution without difficulty. The results in the present paper allow for the case $p < p_0(\theta)$ and thus seem to require a different – and more complex – method of proof. We note that the tests based on $(T_{p+1}, \ldots, T_P)$ again have only a vanishing effect on the conditional distribution $F_{n,\theta,\sigma}(\cdot \,|\, p)$ of the post-model-selection estimator as shown by Theorem 4.2 and Remark 6.3(i) above.

Remark 6.4 (i) The model selection procedures considered in this paper are based on a sequence of tests which use critical values that do not depend on sample size and satisfy $0 < c_j < \infty$. If the critical values are allowed to depend on sample size, the finite-sample results trivially remain valid without any change. The asymptotic results all remain valid if the sample-size-dependent critical values $c_{n,j}$ satisfy $c_{n,j} \to c_{\infty,j}$ for $n \to \infty$ with $0 < c_{\infty,j} < \infty$ (and if $c_{\infty,j}$ replaces c_j in the formulas for the limiting pdfs, cdfs, and selection probabilities). In fact, the asymptotic results not involving the limiting cdf or pdf like Proposition 4.1, Theorem 4.2, Lemma 5.1, and Theorem 5.2 even hold as long as $0 < \liminf c_{n,j} \leq \limsup c_{n,j} < \infty$ is satisfied.

(ii) The assumption that $X'X/n \to Q > 0$, used for the asymptotic results in Sections 4 and 5, can be weakened to $0 < \liminf \lambda_{\min}(X'X/n) \leq \limsup \lambda_{\max}(X'X/n) < \infty$ for all asymptotic results that do not involve the limiting cdf or pdf. (The results involving the limiting distribution of course also hold under this weaker condition along subsequences for which $X'X/n$ converges.) Furthermore, Proposition 4.1, Lemma 5.1, and Theorem 5.2 in fact even hold without any condition on the asymptotic behaviour of $X'X/n$.

Remark 6.5 (i) The results of this paper readily adapt to situations where a subvector of θ is assumed to be common to all candidate models, i.e., where there is a $p_\bullet > 0$ such that only models from the list $M_{p_\bullet}, \ldots, M_P$ should be selected. First observe that setting $c_{p_\bullet} = 0$ forces the model selectors $\hat{p}$ and $\hat{p}^*$ to do precisely this, i.e., makes them satisfy $p_\bullet \leq \hat{p} \leq P$ and $p_\bullet \leq \hat{p}^* \leq P$. Second, redefining $p_0(\theta)$ now as $p_0(\theta) = \min\{p : p_\bullet \leq p \leq P, \theta \in M_p\}$, one immediately sees that the conditional finite-sample distributions of $p > p_\bullet$ are the same as in the case $p_\bullet = 0$ discussed in Section 3, and that now $f^*_{n,\theta,\sigma}(\cdot \mid p_\bullet) = f_{n,\theta,\sigma}(\cdot \mid p_\bullet) = \phi_{n,p_\bullet}(\cdot)$, a density that does not depend on θ at all. As a consequence, all results concerning conditional distributions continue to hold without change for $p > p_\bullet$ and become trivial in case $p = p_\bullet$. The results for the unconditional distributions and the selection probabilities also continue to hold with obvious, mainly notational, changes.

(ii) The finite-sample results in Section 3 adapted to the situation discussed in (i) substantially generalize the results in Giles and Srivastava (1993), who considered a regression model $y_t = \alpha + \beta x_t + \epsilon_t$ with only one regressor. Giles and Srivastava (1993) obtained the unconditional finite-sample distribution of the post-model-selection estimator for α, when model selection is based on a t-test for the hypothesis $\beta = 0$. Elsinger (1994) generalized their result to the model $y_t = x'_{1,t}\alpha + x'_{2,t}\beta + \epsilon_t$ and obtained the unconditional finite-sample distribution of the post-model-selection estimator for the vector α when model selection is based on an F-test for the hypothesis that the vector β satisfies $\beta = 0$.

Remark 6.6 (i) For the construction of the test-statistics T_p the estimator $\hat{\sigma}^2$ obtained from fitting the overall model M_P has been used rather than an estimator that is obtained from fitting model M_p. Such a choice is reasonable from a practical point of view, since it avoids overestimation of the error variance when model M_p is not true, which obviously has adverse effects on the test.

(ii) The finite-sample results in Section 3 are obviously tied very strongly to the assumptions of the paper (linear model, normality, choice of $\hat{\sigma}^2$). Also some of the asymptotic results like, e.g., Theorem 4.2, rely heavily on these assumptions and it is not obvious to which extent they can be generalized, whereas other asymptotic results should be more amenable to generalization. These and related questions are currently under investigation

7 Conclusion

The finite-sample distributions of the post-model-selection estimator $\tilde{\theta}$ derived in Section 3 are quite complicated and difficult to interpret. Insight into the behaviour of these distributions can, however, be gained through a simple and easy to interpret *uniform* approximation that is provided in the paper. This approximation is superior to the asymptotic distribution, which is known to provide an inadequate approximation to the finite-sample distribution under certain circumstances. As a by-product, the uniform approximation result also allows the completion of the analysis in Pötscher (1991), in that the asymptotic distribution of the post-model-selection estimator $\tilde{\theta}$ conditional on selecting an incorrect model can be obtained. We note that the uniform approximation as well as the asymptotic distribution both depend on the unknown parameter value θ, and hence cannot be used directly for purposes of inference. Consistent estimators for the finite-sample distribution of the post-model-selection estimator $\tilde{\theta}$ can be obtained based on the results of the present paper. Such estimators and their features are beyond the scope of this paper and will be discussed elsewhere.

A Proof of Proposition 3.1

The proof of Proposition 3.1 rests on the following lemma which establishes uncorrelatedness of certain random variables of key interest. The result is hardly new and variants of it are folklore. Note that we do *not* assume a true regression relationship to hold in this lemma.

Lemma A.1 *Let Z be an $n \times 1$ random vector whose components are uncorrelated and have constant variance. Let W be a non-stochastic $n \times k$ matrix of full column rank and let $\hat{\beta}$ be the ordinary least-squares estimator obtained from regressing Z on W. Let $W = [W_1 : W_2]$ and let $\hat{\beta} = [\hat{\beta}_1' : \hat{\beta}_2']'$ be partitioned conformably, with W_1 being of dimension $n \times k_1$, $0 < k_1 < k$. Let $\breve{\beta}$ be the ordinary least-squares estimator obtained from regressing Z on W_1 only. Then $\hat{\beta}$ as well as $\breve{\beta}$ are uncorrelated with $Z - W\hat{\beta}$. Furthermore, $\breve{\beta}$ and $\hat{\beta}_2$ are uncorrelated.*

Proof: The first claim is simply verified by calculating the appropriate covariances. To proof the second claim, let $P_{W_1}^{\perp} = (I_n - W_1(W_1'W_1)^{-1}W_1')$ be the projection on the orthogonal complement of the column space of W_1 and define $W^{\circ} = [W_1 : V]$, where $V = P_{W_1}^{\perp} W_2$. Let $\hat{\beta}^{\circ} = [\hat{\beta}_1^{\circ\prime} : \hat{\beta}_2^{\circ\prime}]'$ denote the ordinary least-squares estimator obtained from regressing Z on W°. Clearly, $\hat{\beta}_1^{\circ} = (W_1'W_1)^{-1}W_1'Z$ and $\hat{\beta}_2^{\circ} = (V'V)^{-1}V'Z$ and hence $\hat{\beta}_1^{\circ}$ and $\hat{\beta}_2^{\circ}$ are uncorrelated. Note that

$$W_1\hat{\beta}_1 + W_2\hat{\beta}_2 \;=\; W\hat{\beta} \;=\; W^{\circ}\hat{\beta}^{\circ} \;=\; W_1\hat{\beta}_1^{\circ} + V\hat{\beta}_2^{\circ}.$$

Multiplying this identity by $P_{W_1}^{\perp}$ from the left yields

$$V\hat{\beta}_2 \;=\; V\hat{\beta}_2^{\circ}.$$

Since V has full column rank, we obtain $\hat{\beta}_2^{\circ} = \hat{\beta}_2$. Observing that $\hat{\beta}_1^{\circ} = \breve{\beta}$ completes the proof. □

Proof of Proposition 3.1: Repeated application of Lemma A.1 implies uncorrelatedness of $\hat{\theta}(p), \hat{\theta}_{p+1}(p+1), \ldots, \hat{\theta}_P(P)$ and $Y - X\hat{\theta}(P)$. (In case $p = 0$, observe that $\hat{\theta}(p) = 0$.) In view of the Gaussianity assumption the above list of random variables is in fact independent. By definitions (3), (4) and (5), the proposition is then a simple consequence of this independence result. □

B Auxiliary Lemmata

Lemma B.1 *Let r_m, $m \geq 1$, and s_m, $m \geq 1$, be sequences of non-negative real numbers such that $r_m - s_m$ and $r_m s_m$ converge to ∞. Then*

$$(2\pi)^{1/2}\Delta(r_m, s_m)\,(r_m - s_m)e^{\frac{1}{2}(r_m-s_m)^2} \longrightarrow 1$$

for $m \to \infty$.

Proof: Without loss of generality we may assume that $r_m - s_m$ is positive for each m. Define $\psi(x) = \phi(x)/x$, where ϕ is the standard normal density. Then we may write

$$\frac{\Delta(r_m,s_m)}{\psi(r_m-s_m)} - 1 = \left(\frac{1-\Phi(r_m-s_m)}{\psi(r_m-s_m)} - 1\right) - \frac{\psi(r_m+s_m)}{\psi(r_m-s_m)}\left(\frac{1-\Phi(r_m+s_m)}{\psi(r_m+s_m)} - 1\right) - \frac{\psi(r_m+s_m)}{\psi(r_m-s_m)}.$$

From Gradstejn and Ryzik (1985, equation (8.254), p.931), we obtain for $x > 0$ that $1 - \Phi(x) = \psi(x)(1 - R(x))$ where $|R(x)| \leq x^{-2}$. This implies

$$\begin{aligned}\left|\frac{1-\Phi(r_m+s_m)}{\psi(r_m+s_m)} - 1\right| &\leq (r_m+s_m)^{-2} \to 0 \qquad \text{and}\\ \left|\frac{1-\Phi(r_m-s_m)}{\psi(r_m-s_m)} - 1\right| &\leq (r_m-s_m)^{-2} \to 0.\end{aligned}$$

Furthermore,

$$0 \leq \frac{\psi(r_m+s_m)}{\psi(r_m-s_m)} = \frac{r_m-s_m}{r_m+s_m}e^{-\frac{1}{2}(r_m+s_m)^2+\frac{1}{2}(r_m-s_m)^2} \leq e^{-2r_m s_m} \to 0,$$

which completes the proof. □

Lemma B.2 *Let W_m be a non-negative random variable such that mW_m^2 is chi-square distributed with m degrees of freedom, $m \geq 1$, and let $\epsilon \geq 0$. Then*

$$P\left(W_m \geq 1+\epsilon\right) < e^{-m\epsilon^2/2}.$$

Proof: The case $\epsilon = 0$ is trivial, hence assume $\epsilon > 0$. Using the moment generating function of the chi-square distribution and Markov's inequality we obtain for any $t < 1/2$

$$\begin{aligned}P(W_m \geq 1+\epsilon) &= P\left(e^{tmW_m^2} \geq e^{tm(1+\epsilon)^2}\right) \leq e^{-tm(1+\epsilon)^2}\cdot(1-2t)^{-m/2}\\ &= \exp\left(-m\left(t(1+\epsilon)^2 + \frac{1}{2}\log(1-2t)\right)\right) = \exp(-mh(t,\epsilon)).\end{aligned}$$

Setting $t(\epsilon) = \frac{1}{2}(2\epsilon+\epsilon^2)/(1+\epsilon)^2$ we see that $t(\epsilon) < 1/2$ and $h(t(\epsilon),\epsilon) = (\epsilon(2+\epsilon) - 2\log(1+\epsilon))/2$. Since $\log(1+\epsilon) < \epsilon$, we have $h(t(\epsilon),\epsilon) > (\epsilon(2+\epsilon) - 2\epsilon)/2 = \epsilon^2/2$. □

Lemma B.3 *Let W_m, $m \geq 1$, be as in the previous lemma and let $\tau_{i,m}$, $m \geq 1$, be a sequence of non-negative real numbers for $i = 1,\dots,l$. Assume that the sequence $\tau_{i,m}$ converges to some $\tau_i \in \mathbf{R}\cup\{\infty\}$ for $m \to \infty$ and that $\tau_{i,m} = O(m^{1/2})$ for each $i = 1,\dots,l$. Assume that $\tau_i = \infty$ for $i = 1,\dots,k$ where $1 \leq k \leq l$ and that $\tau_i \in \mathbf{R}$ for $i = k+1,\dots,l$. Define $\tau_{*,m} = \min\{\tau_{1,m},\dots,\tau_{k,m}\}$. Let d_i, $i = 1,\dots,l$,* *be positive real numbers and* $0 < \epsilon < 1$. *Then there exists a real number* K *such that for every* α *satisfying* $0 < \alpha < \min\{1/d_1, \ldots, 1/d_k\}$

$$0 \leq \frac{E\left(\prod_{i=1}^{l} \Delta(\tau_{i,m}, d_i W_m) \cdot \mathbf{1}\{1+\epsilon \leq W_m < \alpha\tau_{*,m}\}\right)}{E\left(\prod_{i=1}^{l} \Delta(\tau_{i,m}, d_i W_m) \cdot \mathbf{1}\{1-\epsilon \leq W_m < 1+\epsilon\}\right)} = O\left(\exp\left(-m\left(\epsilon^2/2 - K\alpha\right)\right)\right) \tag{53}$$

holds. The constant K *depends only on* k, d_i, $i = 1, \ldots, k$, *and the sequences* $\tau_{i,m}$, $i = 1, \ldots, k$, *while the constant implicit in the* $O(\cdot)$ *notation may depend on all sequences* $\tau_{i,m}$, *all* d_i, ϵ *and even on* α.

Proof: Since $\Delta(a, b)$ is strictly increasing in b, and since $\Delta(a, b) \leq 1$, the l.h.s. of (53) is bounded from above by

$$\left(\prod_{i=1}^{k} \frac{\Delta(\tau_{i,m}, d_i\alpha\tau_{*,m})}{\Delta(\tau_{i,m}, d_i(1-\epsilon))}\right) \frac{P(1+\epsilon \leq W_m < \alpha\tau_{*,m})}{E\left(\prod_{i=k+1}^{l} \Delta(\tau_{i,m}, d_i W_m) \cdot \mathbf{1}\{1-\epsilon \leq W_m < 1+\epsilon\}\right)}. \tag{54}$$

Observe that the expectation in the denominator of (54) converges to $\prod_{i=k+1}^{l} \Delta(\tau_i, d_i) > 0$ by dominated convergence since $W_m \to 1$ in probability. Hence, we can find a constant $K^{(1)}$ such that (54) is bounded from above by

$$K^{(1)} \cdot P(W_m \geq 1+\epsilon) \cdot \prod_{i=1}^{k} \frac{\Delta(\tau_{i,m}, d_i\alpha\tau_{*,m})}{\Delta(\tau_{i,m}, d_i(1-\epsilon))}. \tag{55}$$

Observe that $\tau_{i,m} - d_i\alpha\tau_{*,m} \to \infty$ for $i = 1, \ldots, k$, since $d_i\alpha < 1$. Hence, Lemma B.1 implies that there exists a positive constant $K^{(2)}$ such that (55) is bounded from above by

$$\begin{aligned} & K^{(2)} P(W_m \geq 1+\epsilon) \prod_{i=1}^{k} \frac{\psi(\tau_{i,m} - d_i\alpha\tau_{*,m})}{\psi(\tau_{i,m} - d_i(1-\epsilon))} \qquad (56) \\ = \quad & K^{(2)} P(W_m \geq 1+\epsilon) \prod_{i=1}^{k} \frac{\tau_{i,m} - d_i(1-\epsilon)}{\tau_{i,m} - d_i\alpha\tau_{*,m}} e^{-\frac{1}{2}(\tau_{i,m} - d_i\alpha\tau_{*,m})^2 + \frac{1}{2}(\tau_{i,m} - d_i(1-\epsilon))^2} \\ \leq \quad & K^{(3)} P(W_m \geq 1+\epsilon) \prod_{i=1}^{k} \frac{1 - d_i(1-\epsilon)/\tau_{i,m}}{1 - d_i\alpha\tau_{*,m}/\tau_{i,m}} e^{d_i\alpha\tau_{i,m}\tau_{*,m}} \end{aligned}$$

where $K^{(3)} = K^{(2)} \exp((\sum_{i=1}^{k} d_i^2)(1-\epsilon)^2/2)$ and ψ is as in the proof of Lemma B.1. Observe that $1 - d_i(1-\epsilon)/\tau_{i,m} \to 1$ and that $|1 - d_i\alpha\tau_{*,m}/\tau_{i,m}| \geq 1 - d_i\alpha > 0$ for $i = 1, \ldots, k$. Hence we can find a positive constant $K^{(4)}$ such that (56) is bounded from above by

$$K^{(4)} P(W_m \geq 1+\epsilon) \prod_{i=1}^{k} e^{\alpha d_i \tau_{i,m}\tau_{*,m}}. \tag{57}$$

Since $\tau_{i,m} = O(m^{1/2})$ by assumption, we can find a positive constant L depending only on the d_i's and the sequences $\tau_{i,m}$, $i = 1, \ldots, k$, such that $0 \leq d_i\tau_{i,m}\tau_{*,m}/m \leq L$ holds. Setting $K = kL$, (57) is seen to be bounded by

$$K^{(4)} P(W_m \geq 1+\epsilon) e^{m\alpha K}.$$

Application of Lemma B.2 completes the proof. $\square$

For the remainder of this appendix let Z be a standard normally distributed random variable which is independent of W_m for every $m \geq 1$, where W_m is as in Lemma B.2. Define

$$f_m^* \;=\; \frac{\mathbf{1}\{|Z-\tau_{0,m}| \geq d_0\}}{1-\Delta(\tau_{0,m}, d_0)}$$

and

$$f_m \;=\; \frac{E\left(\mathbf{1}\{|Z-\tau_{0,m}| \geq W_m d_0\} \prod_{i=1}^{l} \Delta(\tau_{i,m}, W_m d_i) \,\Big|\, Z\right)}{E\left((1-\Delta(\tau_{0,m}, W_m d_0)) \prod_{i=1}^{l} \Delta(\tau_{i,m}, W_m d_i)\right)} \tag{58}$$

where $\tau_{i,m}$, $i = 0, \ldots, l$, $l \geq 0$, are real numbers and $0 < d_i < \infty$, $i = 0, \ldots, l$ holds. (If $l = 0$, the product in (58) is to be interpreted as unity.)

Lemma B.4 *Assume that $\tau_{i,m} \to \tau_i \in \mathbf{R} \cup \{-\infty, \infty\}$ for $m \to \infty$, $i = 0, \ldots, l$, and that $\tau_{i,m} = O(m^{1/2})$ for each $i = 1, \ldots, l$. Then*

$$\lim_{m\to\infty} E\left((f_m - f_m^*)^+ \,\mathbf{1}\{|Z-\tau_{0,m}| \geq d_0\}\right) \;=\; 0.$$

Proof: Without loss of generality we may assume that all elements $\tau_{i,m}$ are non-negative. Introduce the random variables $A_m = \prod_{i=1}^{l} \Delta(\tau_{i,m}, W_m d_i)$ and $B_m = (1-\Delta(\tau_{0,m}, W_m d_0))A_m$ with the convention that $A_m = 1$ in case $l = 0$. We then observe that

$$f_m \;\leq\; \frac{EA_m}{EB_m}.$$

Since $f_m^* = (1-\Delta(\tau_{0,m}, d_0))^{-1}$ on the event $\{|Z-\tau_{0,m}| \geq d_0\}$ and since $\tau_{0,m} \to \tau_0 \in \mathbf{R} \cup \{\infty\}$ implies $\Delta(\tau_{0,m}, d_0) \to \Delta(\tau_0, d_0)$, the lemma is proven if we can establish that

$$\limsup_{m\to\infty} \frac{EA_m}{EB_m} \;\leq\; \frac{1}{1-\Delta(\tau_0, d_0)}. \tag{59}$$

Since $\Delta(\cdot,\cdot)$ is continuous on $(\mathbf{R} \cup \{-\infty, \infty\}) \times \mathbf{R}$, convergence in probability of $\Delta(\tau_{i,m}, W_m d_i)$ to $\Delta(\tau_i, d_i)$ follows from $W_m \to 1$ in probability. Since $\Delta(\cdot,\cdot)$ is bounded, it follows that EA_m converges to $\prod_{i=1}^{l} \Delta(\tau_i, d_i)$ while EB_m converges to $(1-\Delta(\tau_0, d_0)) \prod_{i=1}^{l} \Delta(\tau_i, d_i)$. In case $\tau_1, \ldots, \tau_l$ are all finite or $l = 0$, it is easy to see that the limit of EB_m is positive, and hence (59) immediately follows in this case.

In case at least one τ_i, $i = 1, \ldots, l$, is infinite, the limits of both EA_m and EB_m are zero and a more delicate argument is needed to establish (59): Without loss of generality we may assume that $\tau_i = \infty$ for $i = 1, \ldots, k$, $1 \leq k \leq l$, and that τ_i is finite for $i = k+1, \ldots, l$. Set $\tau_{*,m} = \min\{\tau_{1,m}, \ldots, \tau_{k,m}\}$. For arbitrary but fixed $0 < \epsilon < 1$, write $A_m = \sum_{i=1}^{4} A_m^{(i)}$ where

$$\begin{aligned}
A_m^{(1)} &= A_m \,\mathbf{1}\{W_m < 1-\epsilon\} \\
A_m^{(2)} &= A_m \,\mathbf{1}\{1-\epsilon \leq W_m < 1+\epsilon\} \\
A_m^{(3)} &= A_m \,\mathbf{1}\{1+\epsilon \leq W_m < \alpha\tau_{*,m}\} \\
A_m^{(4)} &= A_m \,\mathbf{1}\{\max(1+\epsilon, \alpha\tau_{*,m}) \leq W_m\}
\end{aligned}$$

where α satisfies $0 < \alpha < \min\{1/d_1, \ldots, 1/d_k\}$. We first show that $EA_m^{(1)}/EB_m \to 0$:

$$
\begin{aligned}
0 \le \frac{EA_m^{(1)}}{EB_m} &\le \frac{EA_m^{(1)}}{E(B_m\,\mathbf{1}\{1-\epsilon \le W_m < 1+\epsilon\})} \\
&\le \frac{\left(\prod_{i=1}^{l} \Delta(\tau_{i,m},(1-\epsilon)d_i)\right)P(W_m < 1-\epsilon)}{(1-\Delta(\tau_{0,m},(1+\epsilon)d_0))\left(\prod_{i=1}^{l}\Delta(\tau_{i,m},(1-\epsilon)d_i)\right)P(1-\epsilon \le W_m < 1+\epsilon)} \\
&= \frac{P(W_m < 1-\epsilon)}{(1-\Delta(\tau_{0,m},(1+\epsilon)d_0))P(1-\epsilon \le W_m < 1+\epsilon)} \to 0
\end{aligned}
$$

since $W_m \to 1$ in probability and $\Delta(\tau_0, (1+\epsilon)d_0) < 1$. Second, observe that

$$
\begin{aligned}
0 \le \frac{EA_m^{(2)}}{EB_m} &\le \frac{EA_m^{(2)}}{E(B_m\,\mathbf{1}\{1-\epsilon \le W_m < 1+\epsilon\})} \\
&\le \frac{EA_m^{(2)}}{(1-\Delta(\tau_{0,m},(1+\epsilon)d_0))EA_m^{(2)}} \to \frac{1}{1-\Delta(\tau_0,(1+\epsilon)d_0)} < \infty. \qquad (60)
\end{aligned}
$$

Next we show that $EA_m^{(i)}/EB_m \to 0$ for $i = 3, 4$. Since $EA_m^{(2)}/EB_m$ is bounded as shown in (60), it suffices to establish $EA_m^{(i)}/EA_m^{(2)} \to 0$, $i = 3, 4$. Lemma B.3 implies that $EA_m^{(3)}/EA_m^{(2)} = O(\exp(-m(\epsilon^2/2 - K\alpha))$, where K is a constant independent of α. Choosing α sufficiently small establishes convergence to zero of $EA_m^{(3)}/EA_m^{(2)}$. Turning to $EA_m^{(4)}/EA_m^{(2)}$ observe that

$$
\begin{aligned}
0 \le \frac{EA_m^{(4)}}{EA_m^{(2)}} &\le \frac{P(W_m \ge \alpha\tau_{*,m})}{EA_m^{(2)}} \\
&\le \left(\prod_{i=1}^{l} \Delta(\tau_{i,m}, (1-\epsilon)d_i)\right)^{-1} \frac{P(W_m \ge \alpha\tau_{*,m})}{P(1-\epsilon \le W_m < 1+\epsilon)}, \qquad (61)
\end{aligned}
$$

where we have first made use of $\Delta(a, b) \le 1$ and then of monotonicity of $\Delta(\cdot, \cdot)$ in its second argument. Since τ_i is finite for $i = k+1, \ldots, l$, the limit of $\prod_{i=k+1}^{l} \Delta(\tau_{i,m}, (1-\epsilon)d_i)$ is positive. To prove convergence to zero of $EA_m^{(4)}/EA_m^{(2)}$ it hence suffices to establish that

$$
\left(\prod_{i=1}^{k} \Delta(\tau_{i,m}, (1-\epsilon)d_i)\right)^{-1} P(W_m \ge \alpha\tau_{*,m}) \qquad (62)
$$

converges to zero for $m \to \infty$. By Lemmata B.1 and B.2 the expression in (62) is bounded by a constant times

$$
\prod_{i=1}^{k} \left((\tau_{i,m} - (1-\epsilon)d_i) \exp\left(\frac{1}{2}(\tau_{i,m} - (1-\epsilon)d_i)^2\right)\right) \exp\left(-\frac{m}{2}(\alpha\tau_{*,m} - 1)^2\right). \qquad (63)
$$

Since $\tau_{i,m} = O(m^{1/2})$ and since $\tau_{*,m} \to \infty$ by construction, it follows that the expression in (63), and hence the one in (61), converges to zero.

Taken together we have now established that

$$
\limsup_{m\to\infty} \frac{EA_m}{EB_m} \le \frac{1}{1-\Delta(\tau_0, (1+\epsilon)d_0)}. \qquad (64)
$$

Since the r.h.s. of (64) converges to the r.h.s. of (59) for $\epsilon \to 0$, the proof is complete. □

Lemma B.5 *Assume that $\tau_{i,m} \to \tau_i \in \mathbf{R} \cup \{-\infty, \infty\}$ for $m \to \infty$, $i = 0, \ldots, l$, and that $\tau_{i,m} = O(m^{1/2})$ for $i = 1, \ldots, l$. Then*

$$
\lim_{m\to\infty} E\left((f_m - f_m^*)^+\,\mathbf{1}\{|Z - \tau_{0,m}| < d_0\}\right) = 0.
$$

Proof: Without loss of generality we may assume that all elements $\tau_{i,m}$ are non-negative. Since f_m^* is identically zero on the event $\{|Z - \tau_{0,m}| < d_0\}$ and since $f_m \geq 0$, it suffices to show that $E(f_m \mathbf{1}\{|Z - \tau_{0,m}| < d_0\})$ converges to zero. Observe that

$$0 \leq E(f_m \mathbf{1}\{|Z - \tau_{0,m}| < d_0\}) = E(A_m \mathbf{1}\{d_0 W_m \leq |Z - \tau_{0,m}| < d_0\})/EB_m \tag{65}$$

where A_m and B_m have been defined in the proof of Lemma B.4.

Consider first the case where all $\tau_1, \ldots, \tau_l$ are finite or $l = 0$. Then $\mathbf{1}\{d_0 W_m \leq |Z - \tau_{0,m}| < d_0\}$ converges to zero in probability since using independence of Z and W_m we have that

$$\begin{aligned}
&P(d_0 W_m \leq |Z - \tau_{0,m}| < d_0) \\
&\quad = P(d_0 W_m + \tau_{0,m} \leq Z < d_0 + \tau_{0,m},\, W_m < 1) + \\
&\qquad\quad P(\tau_{0,m} - d_0 < Z \leq \tau_{0,m} - d_0 W_m,\, W_m < 1) \\
&\quad = E\left((\Delta(\tau_{0,m}, d_0) - \Delta(\tau_{0,m}, d_0 W_m))\, \mathbf{1}\{W_m < 1\}\right) \\
&\quad \leq E\left((\Delta(\tau_{0,m}, d_0) - \Delta(\tau_{0,m}, d_0 W_m))\right)
\end{aligned} \tag{66}$$

and the upper bound in (66) converges to zero since $W_m \to 1$ in probability and $\Delta(\cdot, \cdot)$ is bounded. Furthermore, as shown in the proof of Lemma B.4, the random variables A_m and B_m converge to non-zero constants in the case under consideration. Since A_m and B_m are bounded, it follows that (65) converges to zero.

Next consider the case when at least one τ_i is infinite. Without loss of generality assume that the first k of the τ_i are infinite while the remaining $l - k$ are finite. Using arguments identical to the ones leading to (66) we may write the numerator of (65) as

$$\begin{aligned}
&E((\Delta(\tau_{0,m}, d_0) - \Delta(\tau_{0,m}, d_0 W_m)) A_m \, \mathbf{1}\{W_m < 1\}) \\
&\quad \leq E(A_m \, \mathbf{1}\{W_m < 1 - \epsilon\}) + \\
&\qquad\qquad E\left((\Delta(\tau_{0,m}, d_0) - \Delta(\tau_{0,m}, d_0 W_m)) A_m \, \mathbf{1}\{1 - \epsilon \leq W_m < 1\}\right) \\
&\quad \leq E(A_m^{(1)}) + (\Delta(\tau_{0,m}, d_0) - \Delta(\tau_{0,m}, (1-\epsilon) d_0)) E(A_m^{(2)}),
\end{aligned}$$

where $A_m^{(1)}$ and $A_m^{(2)}$ have been defined in the proof of Lemma B.4. In that proof, it was also shown that $EA_m^{(1)}/EB_m \to 0$ and $EA_m^{(2)}/EB_m \to (1 - \Delta(\tau_0, (1+\epsilon)d_0))^{-1}$ for $m \to \infty$. It follows that the lim sup of (65) is bounded by

$$\frac{\Delta(\tau_0, d_0) - \Delta(\tau_0, (1-\epsilon)d_0)}{1 - \Delta(\tau_0, (1+\epsilon)d_0)}$$

for every $0 < \epsilon < 1$. Letting $\epsilon \to 0$ completes the proof. □

C Proofs for Section 4

Consider a sequence $\theta^{(n)} \in \mathbf{R}^P$ and a sequence $\sigma^{(n)}$ of positive real numbers. For p satisfying $1 \leq p \leq P$ we will need to express

$$\left|\left| f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p) \; - \; f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p) \right|\right|_1 \tag{67}$$

in a different form. Define $\overline{\tau}_{j,n} = -n^{1/2} \mu_{n,j}(j) {\sigma^{(n)}}^{-1} \xi_{n,j}^{-1}$ for $j = p, \ldots, P$, where $\mu_n(j)$ is given by (6) with θ replaced by $\theta^{(n)}$. Observe that $f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and

$f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ both are a product of $\phi_{n,p}(\cdot)$ times a function that depends on t_p but not on $t_1,\ldots,t_{p-1}$. Decomposing $\phi_{n,p}(t)$ into the product of the marginal density of the last component and the conditional density of the first $p-1$ coordinates given the last one, the p-fold integral representing the L_1-norm in (67) can easily be reduced to

$$\int_{-\infty}^{\infty}\left|\frac{\int_0^\infty \mathbf{1}_{U(p,sc_p)}(t_p)\prod_{j=p+1}^{P}\Delta(\overline{\tau}_{j,n},sc_j)h(s)\,ds}{\int_0^\infty(1-\Delta(\overline{\tau}_{p,n},sc_p))\prod_{j=p+1}^{P}\Delta(\overline{\tau}_{j,n},sc_j)h(s)\,ds}-\frac{\mathbf{1}_{U(p,c_p)}(t_p)}{1-\Delta(\overline{\tau}_{p,n},c_p)}\right|\phi(t_p;(\sigma^{(n)})^2\xi_{n,p}^2)\,dt_p,$$

where we have used the representations of $f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ and $f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot\,|\,p)$ given in (12) and (23)-(24), respectively and where $\phi(\cdot;(\sigma^{(n)})^2\xi_{n,p}^2)$ denotes the density of a normally distributed random variable with mean zero and variance $(\sigma^{(n)})^2\xi_{n,p}^2$. Performing the substitution $r={\sigma^{(n)}}^{-1}\xi_{n,p}^{-1}t_p$ this can be rewritten as

$$\int_{-\infty}^{\infty}\left|\frac{\int_0^\infty \mathbf{1}\{|r-\overline{\tau}_{p,n}|\geq sc_p\}\prod_{j=p+1}^{P}\Delta(\overline{\tau}_{j,n},sc_j)h(s)\,ds}{\int_0^\infty(1-\Delta(\overline{\tau}_{p,n},sc_p))\prod_{j=p+1}^{P}\Delta(\overline{\tau}_{j,n},sc_j)h(s)\,ds}-\frac{\mathbf{1}\{|r-\overline{\tau}_{p,n}|\geq c_p\}}{1-\Delta(\overline{\tau}_{p,n},c_p)}\right|\phi(r)\,dr. \tag{68}$$

The following probabilistic reinterpretation of (68) will prove useful. Let Z be a standard normally distributed random variable and let $\overline{W}_n$ for $n>P$ be a non-negative random variable, such that $(n-P)\overline{W}_n^2$ is chi-square distributed with $n-P$ degrees of freedom. Furthermore Z and $\overline{W}_n$ are assumed to be independent for every $n>P$. Define random variables $\overline{f}_n$ and $\overline{f}_n^*$ by

$$\overline{f}_n^* \;=\; \frac{\mathbf{1}\{|Z-\overline{\tau}_{p,n}|\geq c_p\}}{1-\Delta(\overline{\tau}_{p,n},c_p)}$$

and

$$\overline{f}_n \;=\; \frac{E\left(\mathbf{1}\{|Z-\overline{\tau}_{p,n}|\geq \overline{W}_nc_p\}\prod_{j=p+1}^{P}\Delta(\overline{\tau}_{j,n},\overline{W}_nc_j)\;\middle|\;Z\right)}{E\left((1-\Delta(\overline{\tau}_{p,n},\overline{W}_nc_p))\prod_{j=p+1}^{P}\Delta(\overline{\tau}_{j,n},\overline{W}_nc_j)\right)}$$

where $E(\cdot)$ denotes the expectation operator w.r.t. the joint distribution of Z and $\{\overline{W}_n : n>P\}$. It is now easy to see that (68), and hence (67), is equal to

$$E\left|\overline{f}_n\;-\;\overline{f}_n^*\right|.$$

Observe that $E\overline{f}_n=E\overline{f}_n^*=1$ holds in view of the definition of $\Delta(\cdot,\cdot)$ and independence of Z and $\overline{W}_n$. Consequently,

$$0\;=\;E\left(\overline{f}_n-\overline{f}_n^*\right)\;=\;E\left(\overline{f}_n-\overline{f}_n^*\right)^+\;-\;E\left(\overline{f}_n-\overline{f}_n^*\right)^-$$

and therefore

$$E\left|\overline{f}_n-\overline{f}_n^*\right|\;=\;2E\left(\overline{f}_n-\overline{f}_n^*\right)^+. \tag{69}$$

Proof of Proposition 4.1: It suffices to show for every given p, $1\leq p\leq P$, that (67), and hence (69), converges to zero for $n\to\infty$ for any sequence $\theta^{(n)}\in M_p$

and $\sigma^{(n)} > 0$. Since $\theta^{(n)} \in M_p$, it follows that $\mu_{n,j}(j) = 0$ for $j = p+1, \ldots, P$ and hence $\overline{\tau}_{j,n} = 0$ for $j = p+1, \ldots, P$. To verify convergence of (69) to zero, it suffices to show that every subsequence (n_i) contains a further subsequence $(n_{i(k)})$ such that the convergence occurs along this subsequence $(n_{i(k)})$. Since $\mathbf{R} \cup \{-\infty, \infty\}$ is compact, every subsequence contains a further subsequence along which $\overline{\tau}_{p,n}$ converges to some $\overline{\tau}_p \in \mathbf{R} \cup \{-\infty, \infty\}$. Hence, for verifying convergence of (69) we may assume without loss of generality that $\overline{\tau}_{p,n}$ converges to $\overline{\tau}_p$. Set $\overline{\tau}_j = 0$ for $j = p+1, \ldots, P$. Decompose the expectation on the r.h.s of (69) as follows:

$$\begin{aligned} E(\overline{f}_n - \overline{f}_n^*)^+ &= E\left((\overline{f}_n - \overline{f}_n^*)^+ \mathbf{1}\{|Z - \overline{\tau}_{p,n}| \geq c_p\}\right) + \\ &\quad E\left((\overline{f}_n - \overline{f}_n^*)^+ \mathbf{1}\{|Z - \overline{\tau}_{p,n}| < c_p\}\right). \end{aligned}$$

Upon making the identification $m = n - P$, $l = P - p$, $\tau_{i,m} = \overline{\tau}_{i+p,m+P}$, $\tau_i = \overline{\tau}_{i+p}$, $W_m = \overline{W}_{m+P}$, $f_m^* = \overline{f}_{m+P}^*$, $f_m = \overline{f}_{m+P}$, $d_i = c_{i+p}$ for $i = 0, \ldots, l$, and noting that $\tau_{i,m} = 0$ for $i = 1, \ldots, l$, an application of Lemmata B.4 and B.5 establishes that (69) converges to zero. $\square$

Proof of Theorem 4.2: Again it suffices to show for every given p, $1 \leq p \leq P$, that (69) converges to zero for $n \to \infty$, where now $\theta^{(n)}$ is any sequence in C and $\sigma^{(n)} \in \mathbf{R}$ satisfying $\sigma^{(n)} \geq \sigma_* > 0$. By a reasoning similar to the one in the proof of Proposition 4.1, we may assume without loss of generality that $\overline{\tau}_{j,n}$ converges to some $\overline{\tau}_j \in \mathbf{R} \cup \{-\infty, \infty\}$ for every $j = p, \ldots, P$. Since $\theta^{(n)} \in C$, a bounded set, boundedness of $\mu_{n,j}(j)$ follows from $X'X/n \to Q > 0$. Since $\xi_{n,j}$ converges to a positive constant and since $\sigma^{(n)} \geq \sigma_* > 0$, the relation $\overline{\tau}_{j,n} = O(n^{1/2})$ for $j = p, \ldots, P$ follows. Decomposing (69) as in the proof of Proposition 4.1, we may again apply Lemmata B.4 and B.5 to conclude convergence to zero of (69). $\square$

As a point of interest we note the following: The proof of Proposition 4.1 makes use of Lemmata B.4 and B.5. However, as shown in the proof, $\tau_{i,m} = 0$ for $i = 1, \ldots, l$ holds due to the assumption $\theta^{(n)} \in M_p$, and thus this corresponds to the trivial case discussed in the proofs of Lemmata B.4 and B.5. As a consequence, the proof of Proposition 4.1 does not rely on Lemmata B.1 – B.3 at all. However, these lemmata as well as the full complexity of the proofs of Lemmata B.4 and B.5 are required for the proof of Theorem 4.2, due to the fact that $\theta^{(n)} \notin M_p$ (implying the presence of unbounded $\tau_{i,m}$) is possible in the framework of this theorem.

Proof of Proposition 4.3: (a) Pointwise convergence of $\phi_{n,p}$ to $\phi_{\infty,p}$ is obvious since the variance-covariance matrix ${\sigma^{(n)}}^2(X[p]'X[p]/n)^{-1}$ corresponding to the former density converges to the variance-covariance matrix corresponding to the latter density. The arguments of the $\Delta(\cdot,\cdot)$-function in (12) converge to the arguments of the $\Delta(\cdot,\cdot)$-function in (36). By continuity of $\Delta(\cdot,\cdot)$ on $(\mathbf{R} \cup \{-\infty, \infty\}) \times \mathbf{R}$, the norming constant in (12) converges to the norming constant in (36). The endpoints of the excision interval in (12) converge to the endpoints of the excision interval of (36) under the assumptions of the proposition. Hence the indicator function in (12) converges to the one in (36) at least for all t_p not equal to the endpoints of the excision interval in (36). Since the set of all $t \in \mathbf{R}^p$ such that the last coordinate is equal to a given value is obviously a Lebesgue null set the λ-almost sure convergence of (12) to (36) follows. Convergence in the L_1-norm then follows from Scheffe's theorem.
(b) Follows immediately from part (a) and Theorem 4.2. $\square$

Proof of Corollary 4.5: In light of Proposition 4.3 it only remains to show that the limit ν_p of $\sqrt{n}\mu_{n,p}(p)$ satisfies (37) under the assumptions of part (a) and equals $\pm\infty$ under the assumptions of parts (b) and (c). Note that $\sqrt{n}\mu_{n,p}(p)$, defined by (6) with $\theta^{(n)}$ replacing θ, is given by (39). Under the assumptions of part (a) we have $\theta[\neg p] = 0$ and $\theta_p = 0$, which shows that ν_p is indeed given by (37). For part (b) note that $\theta_p \neq 0$ while $\theta[\neg p] = 0$. This implies divergence of $\sqrt{n}\mu_{n,p}(p)$ to $\nu_p = \infty$ or $\nu_p = -\infty$ in this case. Part (c) follows since condition (38) again implies divergence of $\sqrt{n}\mu_{n,p}(p)$. □

Proof of Corollary 4.6: We first prove (40) with $F_{n,\overline{\theta},\sigma}(t\,|\,p)$ replaced by $F^*_{n,\overline{\theta},\sigma}(t\,|\,p)$. For given $t \in \mathbf{R}^p$ and $\sigma > 0$ choose $\alpha \neq 0$ smaller than $-t_p - \sigma\xi_{\infty,p}c_p$ and let ρ be a real number larger than $|\alpha|$. For any given $\theta \in M_{p-1}$, define $\theta^{(n)} = \theta + (\alpha/\sqrt{n})e^{(p)}$ where $e^{(p)}$ is the p-th standard basis vector in $\mathbf{R}^P$. By Corollary 4.5(a) the cdf $F^*_{n,\theta^{(n)},\sigma}(\cdot\,|\,p)$ converges to the cdf of an excised normal distribution with density (36), where $\nu_p = \alpha$. By definition of α we have $t_p < -\alpha - \sigma\xi_{\infty,p}c_p$, the lower endpoint of the excision interval in (36). Consequently, $F^*_{n,\theta^{(n)},\sigma}(t\,|\,p)$ converges to $\Phi_{\infty,p}(t)\,(1 - \Delta(-\alpha\sigma^{-1}\xi^{-1}_{\infty,p}, c_p))^{-1}$. Since $\theta^{(n)} \in M_p \setminus M_{p-1}$, it follows from Corollary 4.5(b) that $F_{\infty,\theta^{(n)},\sigma}(t\,|\,p) = \Phi_{\infty,p}(t)$. This implies

$$\left| F^*_{n,\theta^{(n)},\sigma}(t\,|\,p) \;-\; F_{\infty,\theta^{(n)},\sigma}(t\,|\,p) \right| \;\to\; \Phi_{\infty,p}(t)\,\left| 1 - \frac{1}{1 - \Delta(-\alpha\sigma^{-1}\xi^{-1}_{\infty,p}, c_p)} \right| \qquad (70)$$

as $n \to \infty$. Since $\alpha\sigma^{-1}\xi^{-1}_{\infty,p}$ is finite and c_p is positive, we see that the limit in (70) is positive. Observing that $\theta^{(n)} \in B_p(\theta, \rho/\sqrt{n})$ since $|\alpha| < \rho$, the result (40) with $F^*_{n,\theta^{(n)},\sigma}(\cdot\,|\,p)$ replacing $F_{n,\theta^{(n)},\sigma}(\cdot\,|\,p)$ follows. Proposition 4.1 then immediately delivers (40). □

D Proofs for Section 5

Proof of Lemma 5.1: To prove the lemma it suffices to show for every $1 \leq p \leq P$ that

$$\left|\left| f_{n,\theta^{(n)},\sigma^{(n)}}(\cdot|p)\pi_{n,\theta^{(n)},\sigma^{(n)}}(p) - f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot|p)\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p) \right|\right|_1 \longrightarrow 0 \qquad (71)$$

for $n \to \infty$ for any sequence $\theta^{(n)} \in \mathbf{R}^P$ and any sequence of positive real numbers $\sigma^{(n)}$. Arguing similarly as at the beginning of Appendix C, it is easy to see that the L_1-norm in (71) can be written as

$$\begin{aligned} \int_{-\infty}^{\infty} \Bigg| \int_0^{\infty} \mathbf{1}\{|r - \overline{\tau}_{p,n}| \geq sc_p\} \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, sc_j)h(s)ds \, - \\ \mathbf{1}\{|r - \overline{\tau}_{p,n}| \geq c_p\} \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) \Bigg| \phi(r)dr \end{aligned} \qquad (72)$$

where $\overline{\tau}_{j,n}$ has been defined in Appendix C. With random variables Z and $\overline{W}_n$ as in Appendix C, the expression in (72) can be rewritten as

$$\begin{aligned} E\Bigg| E\left(\mathbf{1}\{|Z - \overline{\tau}_{p,n}| \geq \overline{W}_n c_p\} \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, \overline{W}_n c_j) \,\Bigg|\, Z \right) \, - \\ \mathbf{1}\{|Z - \overline{\tau}_{p,n}| \geq c_p\} \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) \Bigg|. \end{aligned} \qquad (73)$$

The expression in (73) is bounded from above by

$$E\left| \mathbf{1}\{|Z - \overline{\tau}_{p,n}| \geq \overline{W}_n c_p\} \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, \overline{W}_n c_j) \;-\; \mathbf{1}\{|Z - \overline{\tau}_{p,n}| \geq c_p\} \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) \right|. \qquad (74)$$

Using a similar subsequence argument as in the proof of Theorem 4.2, we may assume without loss of generality that $\overline{\tau}_{j,n}$ converges to a limit $\overline{\tau}_j \in \mathbf{R} \cup \{-\infty, \infty\}$ for all j. Recalling that $\overline{W}_n \to 1$ in probability and that Z and $\overline{W}_n$ are independent, it is easy to see that both indicator functions in (74) converge to the same limit $\mathbf{1}\{|Z - \overline{\tau}_p| \geq c_p\}$ in probability, where the latter indicator function is to be interpreted as the constant 1 in case $\overline{\tau}_p = \pm\infty$. Since $\Delta(\cdot, \cdot)$ is continuous on $(\mathbf{R} \cup \{-\infty, \infty\}) \times \mathbf{R}$, convergence of the expression inside the expectation operator in (74) follows. Convergence of (74) to zero now follows from the dominated convergence theorem, since the integrand in (74) is bounded by 2. □

Proof of Theorem 5.2: We prove part (b) first. It suffices to show for every $0 \leq p \leq P$ that

$$\left|\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p) - \pi_{n,\theta^{(n)},\sigma^{(n)}}(p)\right| \longrightarrow 0 \tag{75}$$

for $n \to \infty$ for any sequence $\theta^{(n)} \in \mathbf{R}^P$ and any sequence of positive real numbers $\sigma^{(n)}$. Arguing as in the proof of Lemma 5.1 we may assume that $\overline{\tau}_{j,n}$ converges to a limit $\overline{\tau}_j \in \mathbf{R} \cup \{-\infty, \infty\}$ for all j. It is easy to see that the absolute value in (75) can be rewritten as

$$\begin{aligned} &\left|(1 - \Delta(\overline{\tau}_{p,n}, c_p)) \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) - \int_{-\infty}^{\infty} (1 - \Delta(\overline{\tau}_{p,n}, sc_p)) \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, sc_j) h(s) ds\right| \\ &= \left|(1 - \Delta(\overline{\tau}_{p,n}, c_p)) \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) - E\left((1 - \Delta(\overline{\tau}_{p,n}, \overline{W}_n c_p)) \prod_{j=p+1}^{P} \Delta(\overline{\tau}_{j,n}, \overline{W}_n c_j)\right)\right| \end{aligned} \tag{76}$$

in case $p \geq 1$ and as

$$\begin{aligned} &\left|\prod_{j=1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) - \int_{-\infty}^{\infty} \prod_{j=1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) h(s) ds\right| \\ &= \left|\prod_{j=1}^{P} \Delta(\overline{\tau}_{j,n}, c_j) - E\left(\prod_{j=1}^{P} \Delta(\overline{\tau}_{j,n}, \overline{W}_n c_j)\right)\right| \end{aligned} \tag{77}$$

in case $p = 0$. Since $\overline{W}_n \to 1$ in probability, since $\Delta(\cdot, \cdot)$ is continuous on $(\mathbf{R} \cup \{-\infty, \infty\}) \times \mathbf{R}$ and bounded, convergence of (76) and (77) to zero for $n \to \infty$ follows immediately. This completes the proof of part (b).

To prove part (a) observe that, using (14), (26) and the triangle inequality, the total variation distance between $F_{n,\theta,\sigma}$ and $F^*_{n,\theta,\sigma}$ can be bounded from above by the sum over $p = 0, \ldots, P$ of

$$\begin{aligned} \Bigg\|\mathbf{1}_{\mathbf{R}_+^{P-p}}(u[\neg p] + \sqrt{n}\theta[\neg p]) \Big[&F_{n,\theta,\sigma}(u[p] + \sqrt{n}(\theta[p] - \mu_n(p))|p)\pi_{n,\theta,\sigma}(p) - \\ &F^*_{n,\theta,\sigma}(u[p] + \sqrt{n}(\theta[p] - \mu_n(p))|p)\pi^*_{n,\theta,\sigma}(p)\Big]\Bigg\|_{TV}. \end{aligned} \tag{78}$$

By our conventions, this reduces to $|\pi_{n,\theta,\sigma}(0) - \pi^*_{n,\theta,\sigma}(0)|$ in case $p = 0$, which goes to zero uniformly in θ, σ as already shown. For $p \geq 1$ observe that a translation of the argument of the distribution function does not affect the total variation distance and that the cdfs in (76) are both concentrated on the same subspace. It is then easy to see that (78) is bounded from above by the L_1-distance between $f_{n,\theta,\sigma}(\cdot|p)\pi_{n,\theta,\sigma}(p)$ and $f^*_{n,\theta,\sigma}(\cdot|p)\pi^*_{n,\theta,\sigma}(p)$ w.r.t. the Lebesgue measure on that subspace. An application of Lemma 5.1 then completes the proof. □

Proof of Proposition 5.3: We prove part (b) first. Since $\Delta(\cdot,\cdot)$ is continuous on $(\mathbf{R} \cup \{-\infty, \infty\}) \times \mathbf{R}$, convergence of $\Delta(-n^{1/2}\mu_{n,j}(j)\sigma^{-1}\xi_{n,j}^{-1}, c_j)$ to $\Delta(-\nu_j\sigma^{-1}\xi_{\infty,j}^{-1}, c_j)$ for $j = \max(1,p),\ldots,P$ follows immediately. Parts (b) and (c) for $\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p)$ now follow from (15) and (16) observing for part (c) that one factor in (15) converges to zero and the others are bounded by one. Applying Theorem 5.2(b) delivers parts (b) and (c) also for $\pi_{n,\theta^{(n)},\sigma^{(n)}}(p)$. Part (a) for $f^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot|p)\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p)$ now follows from part (b) and Proposition 4.3(a). An application of Lemma 5.1 then completes the proof of the proposition. □

Lemma D.1 *Let H_n be a sequence of distribution functions of finite mass (not necessarily equal to one) on $\mathbf{R}^k$, which converges to H in total variation distance. Let $a_n \in \mathbf{R}^k$ be a sequence with limit $a \in \mathbf{R}^k$. Assume that H_n and H are absolutely continuous w.r.t Lebesgue measure and that the density h of H is continuous almost everywhere. Then $H_n(\cdot + a_n)$ converges to $H(\cdot + a)$ in the total variation distance.*

Proof: Observe that

$$\begin{aligned}
&||H_n(\cdot + a_n) - H(\cdot + a)||_{TV} \\
&\leq \; ||H_n(\cdot + a_n) - H(\cdot + a_n)||_{TV} + ||H(\cdot + a_n) - H(\cdot + a)||_{TV} \\
&= \; ||H_n(\cdot) - H(\cdot)||_{TV} + ||H(\cdot + a_n) - H(\cdot + a)||_{TV}.
\end{aligned}$$

It hence suffices to show that $H(\cdot + a_n)$ converges to $H(\cdot + a)$ in the total variation distance. Since the density h is continuous almost everywhere, the density $h(\cdot + a_n)$ of $H(\cdot + a_n)$ converges almost everywhere to the density $h(\cdot + a)$ of $H(\cdot + a)$. Convergence of $h(\cdot + a_n)$ to $h(\cdot + a)$ in the L_1-sense now follows from Scheffe's theorem. The proof is completed by observing that $||H(\cdot + a_n) - H(\cdot + a)||_{TV} \leq ||h(\cdot + a_n) - h(\cdot + a)||_1$. □

Proof of Corollary 5.4: For $\theta^{(n)}$ as in the corollary and $1 \leq j \leq P$ recall that $\sqrt{n}\mu_{n,j}(j)$ is given by (39) with j replacing p. Since θ_j and $\theta[\neg j]$ are zero for $j > p_0(\theta)$, one immediately sees that $\sqrt{n}\mu_{n,j}(j)$ converges to υ_j given in (46) for $j > p_0(\theta)$. If $p_0(\theta) > 0$, then for $j = p_0(\theta)$ one also sees that $\sqrt{n}\mu_{n,j}(j)$ converges to $\nu_j = \pm\infty$, since $\theta_j \neq 0$ and $\theta[\neg j] = 0$. Hence the assumptions of Proposition 5.3 are satisfied for any $p \geq p_0(\theta)$. Part (a) of the corollary is now an immediate consequence of Proposition 5.3(b). To prove part (b), expand $F^*_{n,\theta^{(n)},\sigma^{(n)}}$ as in (14) and observe that the terms for $p < p_0(\theta)$ converge to zero in total variation distance since the probabilities $\pi^*_{n,\theta^{(n)},\sigma^{(n)}}(p)$ vanish in the limit as shown in part (a) and since the conditional cdfs have total mass one. For $p \geq p_0(\theta)$ observe that $\sqrt{n}(\theta^{(n)}[p] - \mu_n(p))$ converges to $-(Q[p:p])^{-1}Q[p:\neg p]\gamma[\neg p]$ and that $\sqrt{n}\theta[\neg p] = \gamma[\neg p]$. Part (b) for $F^*_{n,\theta^{(n)},\sigma^{(n)}}$ then follows from an application of Proposition 5.3 together with Lemma D.1. The analogous result for $F_{n,\theta^{(n)},\sigma^{(n)}}$ now follows from Theorem 5.2(a). □

Proof of Corollary 5.5: We first prove (49) with $F^*_{n,\overline{\theta},\sigma}$ replacing $F_{n,\overline{\theta},\sigma}$. For given $u \in \mathbf{R}^P$ and $\sigma > 0$ choose $\alpha \neq 0$ such that

$$-\sigma\xi_{\infty,P}c_P \; < \; u_P + \alpha < 0 \tag{79}$$

holds. Let ρ be a real number larger that $|\alpha|$ and set $\theta^{(n)} = \theta + (\alpha/\sqrt{n})e^{(P)}$ where $e^{(P)}$ is the P-th standard basis vector in $\mathbf{R}^P$. Note that $\theta^{(n)} \in B_P(\theta, \rho/\sqrt{n})$. Since $\theta \in M_{P-1}$ and $\alpha \neq 0$, it follows that $p_0(\theta^{(n)}) = P$. Hence,

$$F_{\infty,\theta^{(n)},\sigma}(u) \; = \; F_{\infty,\theta^{(n)},\sigma}(u|P) \; = \; \Phi_{\infty,P}(u) \tag{80}$$

in view of Corollaries 4.5 and 5.4. On the other hand, $F^*_{n,\theta^{(n)},\sigma}(u)$ converges to $F_{\infty,\theta,\sigma,\alpha e^{(P)}}(u)$ given by (47) with $\gamma = \alpha e^{(P)}$. Since $u_P + \alpha < 0$, the indicator functions in (47) are zero for $p < P$ and hence $F_{\infty,\theta,\sigma,\alpha e^{(P)}}(u)$ reduces to

$$F_{\infty,\theta,\sigma,\alpha e^{(P)}}(u|P)\pi_{\infty,\theta,\sigma,\alpha e^{(P)}}(P) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_P} \phi_{\infty,P}(r)\mathbf{1}_{(-\alpha-\sigma\xi_{\infty,P}c_P,-\alpha+\sigma\xi_{\infty,P}c_P)}(r_P)dr \tag{81}$$

in view of Corollary 4.5(a). Because of (79), the upper integration limit u_P in (81) is larger than the left endpoint of the excision interval. Hence the expressions in (80) and (81) differ, showing that the limits of $F_{\infty,\theta^{(n)},\sigma}(u)$ and $F^*_{n,\theta^{(n)},\sigma}(u)$ are different. This proves the result for $F^*_{n,\theta,\sigma}$. The result for $F_{n,\theta,\sigma}$ now follows from Theorem 5.2(a). □

Proof of Corollary 5.6: (a) Let p satisfy $1 \leq p \leq P$, let $\rho > 0$ and $\sigma > 0$ be arbitrary and choose an arbitrary $\theta \in M_{p-1}$. Define $\theta^{(n)} = \theta + \frac{1}{\sqrt{n}}\alpha e^{(p)}$ with $0 < |\alpha| < \rho$ and where $e^{(p)}$ denotes the p-th standard basis vector in $\mathbf{R}^P$. Observe that $\theta^{(n)} \in M_p \setminus M_{p-1}$ and that $\theta^{(n)} \in B_p(\theta, \rho/\sqrt{n})$. Now, (45) (with $\gamma = 0$) implies

$$\pi_{\infty,\theta^{(n)},\sigma}(p) \;=\; \prod_{j=p+1}^{P} \Delta(0, c_j)$$

since $p_0(\theta^{(n)}) = p$. Corollary 5.4 (with $\gamma = \frac{\alpha}{\sqrt{n}}e^{(p)}$) implies

$$\pi^*_{n,\theta^{(n)},\sigma}(p) \;\longrightarrow\; (1 - \Delta(-\alpha\sigma^{-1}\xi^{-1}_{\infty,p}, c_p)) \prod_{j=p+1}^{P} \Delta(0, c_j)$$

for $n \to \infty$. Since $\alpha \neq 0$, it follows that $|\pi^*_{n,\theta^{(n)},\sigma}(p) - \pi_{\infty,\theta^{(n)},\sigma}(p)|$ converges to a positive constant.

(b) For p satisfying $0 \leq p < P$ and arbitrary $\rho > 0$, $\sigma > 0$ and $\theta \in M_p$ define $\theta^{(n)} = \theta + \frac{1}{\sqrt{n}}\alpha e^{(p+1)}$, where $0 < |\alpha| < \rho$. Observe that $\theta^{(n)} \in M_{p+1} \setminus M_p$ and that $\theta^{(n)} \in B_{p+1}(\theta, \rho/\sqrt{n})$. Applying (45) (with $\gamma = 0$) implies

$$\pi_{\infty,\theta^{(n)},\sigma}(p) \;=\; 0$$

since $p_0(\theta^{(n)}) = p + 1 > p$. Corollary 5.4 (with $\gamma = \frac{\alpha}{\sqrt{n}}e^{(p+1)}$) implies

$$\pi^*_{n,\theta^{(n)},\sigma}(p) \;\longrightarrow\; \Delta(-\alpha\sigma^{-1}\xi^{-1}_{\infty,p+1}, c_{p+1}) \prod_{j=p+2}^{P} \Delta(0, c_j)$$

since $\upsilon_{p+1} = \alpha$ and $\upsilon_j = 0$ for $j > p + 1$. It follows that $|\pi^*_{n,\theta^{(n)},\sigma}(p) - \pi_{\infty,\theta^{(n)},\sigma}(p)|$ converges to a positive constant.

The above proves the corollary for $\pi^*_{n,\theta,\sigma}(p)$. The result for $\pi_{n,\theta,\sigma}(p)$ now follows from an application of Theorem 5.2(b). □

References

BILLINGSLEY, P., AND F. TOPSOE (1967): "Uniformity in weak convergence," *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **7**, 1–16.

CHANDRA, T. K. (1989): "Multidimensional Polya's theorem.," *Bull. Calcutta Math. Soc.*, **81**, 227–231.

ELSINGER, H. (1994): "'Pre-Test' Schätzer und Modellselektion im linearen Regressionsmodell," Master's thesis, Sozial- und Wirtschaftswissenschaftliche Fakultät der Universität Wien.

GAENSSLER, P., AND W. STUTE (1977): *Wahrscheinlichkeitstheorie.* Springer, Berlin.

GILES, D., AND V. SRIVASTAVA (1993): "The exact distribution of a least squares regression coefficient estimator after a preliminary t-test," *Statist. Prob. Lett.*, **16**, 59–64.

GRADSTEJN, I. S., AND I. M. RYZIK (1985): *Table of Integrals, Series, and Products.* Academic Press, Inc., New York, corrected and enlarged edn.

KABAILA, P. (1995): "The effect of model selection on confidence regions and prediction regions," *Econometric Theory*, **11**, 537–549.

PÖTSCHER, B. M. (1991): "Effects of model selection on inference," *Econometric Theory*, **7**, 163–185.

——— (1995): "Comment on "The effect of model selection on confidence regions and prediction regions"," *Econometric Theory*, **11**, 550–559.

PÖTSCHER, B. M., AND A. J. NOVAK (1998): "The distribution of estimators after model selection: large and small sample results," *J. Statist. Comput. Simul.*, **60**, 19–56.

SEN, P. K. (1979): "Asymptotic properties of maximum likelihood estimators based on conditional specification," *Ann. Stat.*, **7**, 1019–1033.