CAN ONE ESTIMATE THE CONDITIONAL DISTRIBUTION OF POST-MODEL-SELECTION ESTIMATORS?

By

Hannes Leeb and Benedikt M. Pötscher

November 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1444

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Can One Estimate The Conditional Distribution of Post-Model-Selection Estimators?

Hannes Leeb
Department of Statistics, Yale University
and
Benedikt M. Pötscher
Department of Statistics, University of Vienna

October 2003

Abstract

We consider the problem of estimating the conditional distribution of a post-model-selection estimator where the conditioning is on the selected model. The notion of a post-model-selection estimator here refers to the combined procedure resulting from first selecting a model (e.g., by a model selection criterion like AIC or by a hypothesis testing procedure) and second estimating the parameters in the selected model (e.g., by least-squares or maximum likelihood), all based on the same data set. We show that it is impossible to estimate this distribution with reasonable accuracy even asymptotically. In particular, we show that no estimator for this distribution can be uniformly consistent (not even locally). This follows as a corollary to (local) minimax lower bounds on the performance of estimators for this distribution. Similar impossibility results are also obtained for the conditional distribution of linear functions (e.g., predictors) of the post-model-selection estimator.

AMS Mathematics Subject Classification 2000: 62F10, 62F12, 62J05, 62J07, 62C05.

Keywords: Inference after model selection, Post-model-selection estimator, Pre-test estimator, Selection of regressors, Akaike's information criterion AIC, Model uncertainty, Consistency, Uniform consistency, Lower risk bound.

Research of the first author was supported by the Max Kade Foundation and by the Austrian National Science Foundation (FWF), Grant No. P13868-MAT. A preliminary draft of the material in this paper was already written in 1999.

1 Introduction and Overview

In many statistical applications a data-based model selection step precedes the final parameter estimation and inference stage. For example, the specification of the model (choice of functional form, choice of regressors, number of lags, etc.) is often based on the data. In contrast, the traditional theory of statistical inference is concerned with the properties of estimators and inference procedures 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. As a consequence, the actual statistical properties of estimators or inference procedures following a data-driven model selection step are not described by the traditional theory which assumes an a priori given model; in fact, they may differ substantially from the properties predicted by this theory, cf., e.g., Danilov and Magnus (2001), Dijkstra and Veldkamp (1988), Pötscher (1991, Section 3.3), or Rao and Wu (2001, Section 12). Ignoring the additional uncertainty originating from the data-driven model selection step and (inappropriately) applying traditional theory can hence result in very misleading conclusions.

Investigations into the distributional properties of post-model-selection estimators, i.e., of estimators constructed after a data-driven model selection step, are relatively few and of recent vintage. Sen (1979) obtained the unconditional large-sample limit distribution of a post-model-selection estimator in an i.i.d. maximum likelihood framework, when selection is between two competing nested models. In Pötscher (1991) the asymptotic properties of a class of post-model-selection estimators (based on a sequence of hypothesis tests) were studied in a rather general setting covering non-linear models, dependent processes, and more than two competing models. In that paper, the large-sample limit distribution of the post-model-selection estimator was derived, both unconditional as well as conditional on having chosen a correct model (not necessarily the minimal one). See also Pötscher and Novak (1998) for further discussion and a simulation study. The finite-sample distribution of a post-model-selection estimator, both unconditional and conditional on having chosen a particular (possibly incorrect) model, was derived in Leeb and Pötscher (2003a) in a normal linear regression framework; this paper also studied asymptotic approximations that are in a certain sense superior to the asymptotic distribution derived in Pötscher (1991). The distributions of corresponding linear predictors constructed after model selection were studied in Leeb (2003a,b). Related work can also be found in Sen and Saleh (1987), Kabaila (1995), Pötscher (1995), Ahmed and Basu (2000), Kapetanios (2001), Hjort and Claeskens (2002), and Dukić and Peña (2002).

It transpires from the papers mentioned above that the finite-sample distributions (as well as the large-sample limit distributions) of post-model-selection estimators typically depend on the unknown model parameters, often in a complicated fashion. For inference purposes, e.g., for the construction of confidence sets, estimators for these distributions would be desirable. Consistent estimators for these distributions can typically be constructed quite easily, e.g., by suitably replacing unknown parameters in the large-sample limit distributions by estimators; cf. Section 4.1. However, the merits of such 'plug-in' estimators in small samples are questionable: It is known that the convergence of the finite-sample distributions to their large-sample limits is typically not uniform with respect to the underlying parameters (see Remark 6.6 below and Corollary 4.6 in Leeb and Pötscher (2003a)). Now, there is no reason to believe that this non-uniformity will disappear when unknown parameter values in the large-sample limit are replaced by estimators. This observation is the main motivation for the present paper to investigate the performance of estimators for the distribution of a post-model-selection estimator in general (not necessarily 'plug-in' estimators based on the limiting distribution). In particular, we ask whether estimators for the distribution function of post-model-selection estimators exist that do not suffer from the non-uniformity phenomenon mentioned above. As we show in this paper the answer in general is 'No'. We also show that these negative results extend to the problem of estimating the distribution of linear functions (e.g., linear predictors) of post-model-selection estimators.

To fix ideas consider for the moment the linear regression model

$$Y = V\chi + W\psi + u \tag{1}$$

where V and W, respectively, represent $n \times k$ and $n \times l$ non-stochastic regressor matrices ($k \geq 1, l \geq 1$), and the $n \times 1$ disturbance vector u is normally distributed with mean zero and variance-covariance matrix $\sigma^2 I_n$. We also assume for the moment that $(V, W)'(V, W)/n$ converges to a nonsingular matrix as the sample size n goes to infinity and that $\lim_{n\to\infty} V'W/n \neq 0$ (for a discussion of the case where this limit is zero see Example 1 in Section 4.2). Now suppose that the vector χ represents the parameters of interest, while the parameter vector ψ and the associated regressors in W have been entered into the model only to avoid possible misspecification. Suppose further that the necessity to include ψ or some of its components is then checked on the basis of the data, i.e., a model selection procedure is used to determine which components of ψ are to be retained in the model (the inclusion of χ not being disputed). The selected model is then used to obtain the final (post-model-selection) estimator $\tilde{\chi}$ for χ. We are now interested in the conditional finite-sample distribution of $\tilde{\chi}$ (appropriately scaled and centered) given the outcome of the model selection step. Denote this k-dimensional cumulative distribution function (cdf) by $G_{n,\theta,\sigma}(t|\hat{M})$, where $\hat{M}$ stands for the selected model, i.e., for the set of selected regressors. As indicated in the notation, this distribution function depends on the true parameters $\theta = (\chi', \psi')'$ and σ^2. For the sake of definiteness of discussion assume that the model selection procedure used here is the particular 'general-to-specific' procedure described in Section 2; we comment on the ramifications of our results for other model selection procedures below.

As mentioned above, it is not difficult to construct a consistent estimator for $G_{n,\theta,\sigma}(t|\hat{M})$, i.e., an estimator $\hat{G}_n(t|\hat{M})$ satisfying

$$P_{n,\theta,\sigma}\left(\left|\hat{G}_n(t|\hat{M}) - G_{n,\theta,\sigma}(t|\hat{M})\right| > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{2}$$

for each $\delta > 0$ and each θ, σ; see Section 4.1. However, it follows from the results in Section 4.2 that *any* estimator satisfying (2), i.e., *any consistent* estimator for $G_{n,\theta,\sigma}(t|\hat{M})$, necessarily also satisfies

$$\liminf_{n\to\infty} \sup_{||\theta||<R} P_{n,\theta,\sigma}\left(\left|\hat{G}_n(t|\hat{M}) - G_{n,\theta,\sigma}(t|\hat{M})\right| > \delta\right) \geq c > 0 \tag{3}$$

for suitable positive constants c, R, and δ. That is, while the probability in (2) converges to zero for every given θ by consistency, relation (3) shows that it does not do so uniformly in θ. It follows that $\hat{G}_n(t|\hat{M})$ can never be uniformly consistent (not even when restricting consideration to uniform consistency over all compact subsets of the parameter space). Hence, a large sample size does not guarantee a small estimation error with high probability when estimating the distribution function of a post-model-selection estimator. In this sense, reliably assessing the precision of post-model-selection estimators is an intrinsically hard problem. Apart from (3), we also provide minimax lower bounds for arbitrary (not necessarily consistent) estimators of the conditional distribution function $G_{n,\theta,\sigma}(t|\hat{M})$. For example, we provide results that imply that

$$\liminf_{n\to\infty} \inf_{\hat{G}_n(t|\hat{M})} \sup_{||\theta||<R} P_{n,\theta,\sigma}\left(\left|\hat{G}_n(t|\hat{M}) - G_{n,\theta,\sigma}(t|\hat{M})\right| > \delta\right) > 0 \tag{4}$$

holds for suitable positive constants R and δ, where the infimum extends over *all* estimators for $G_{n,\theta,\sigma}(t|\hat{M})$. The results in Section 4.2 in fact show that the balls $||\theta|| < R$ in (3) and (4) can be replaced by suitable balls (not necessarily centered at the origin) shrinking at the rate $n^{-1/2}$. This shows that the non-uniformity phenomenon described in (3)-(4) is a local, rather than a global, phenomenon. Moreover, relations (3)-(4) also hold conditionally, i.e., with the unconditional probability $P_{n,\theta,\sigma}(\cdot)$ in (3)-(4) replaced by the conditional probability given model M is selected, i.e., given the event $\hat{M} = M$. In Section 4.2 we further show that the non-uniformity phenomenon expressed in (3) and (4) typically also arises when the parameter of interest is not

χ, but some other linear transformation of $\theta = (\chi', \psi')'$. As discussed in Remark 6.3, the results also hold for randomized estimators of the conditional distribution function $G_{n,\theta,\sigma}(t|\hat{M})$. Hence no resampling procedure whatsoever can alleviate the problem.

The results mentioned above are proved in Section 4 for the particular 'general-to-specific' model selection procedure described in Section 2. Analogous results for a large class of model selection procedures, including Akaike's AIC, are then obtained in Section 5. In fact, it transpires from the proofs that the non-uniformity phenomenon expressed in (3)-(4) is not specific to the model selection procedures discussed in Sections 4 and 5 of the present paper, but will occur for most (if not all) model selection procedures, including consistent ones; cf. Remark 6.8.

One can also envisage a situation where one is more interested in the unconditional distribution of the post-model-selection estimator rather than in the conditional distribution. In this case similar results can be obtained and are reported in Leeb and Pötscher (2003b).

The plan of the paper is as follows: In Section 2 we introduce the basic framework and some notation, like the family of models M_p from which the particular 'general-to-specific' model selection procedure $\hat{p}$ selects as well as the post-model-selection estimator $\tilde{\theta}$. The conditional cdf $G_{n,\theta,\sigma}(t|p)$ of (a linear function of) the post-model-selection estimator $\tilde{\theta}$ given that $\hat{p}$ selects model M_p, is introduced and discussed in Section 3. Consistent estimation of $G_{n,\theta,\sigma}(t|p)$ and of $G_{n,\theta,\sigma}(t|\hat{p})$ (i.e., of the cdf conditional on the actual outcome of the model selection procedure) is discussed in Section 4.1. In Section 4.2 we provide a detailed analysis of the non-uniformity phenomenon encountered in (3)-(4), the main results of this section being given by Theorems 4.3, 4.5, and 4.8. In Section 5 the 'impossibility' results from Section 4 are extended to a large class of model selection procedures including Akaike's AIC and selection from a non-nested collection of models. Some remarks and extensions are collected in Section 6, and conclusions are drawn in Section 7. All proofs are collected into appendices. Finally a word on notation: The Euclidean norm is denoted by $\|\cdot\|$, and $\lambda_{\max}(E)$ denotes the largest eigenvalue of a symmetric matrix E. A prime denotes transposition of a matrix.

2 The Model and Estimators

Consider the linear regression model

$$Y = X\theta + u, \tag{5}$$

where X is a non-stochastic $n \times P$ matrix with $rank(X) = P$ and $u \sim N(0, \sigma^2 I_n)$, $\sigma^2 > 0$. Here n denotes the sample size and we assume $n > P \geq 1$. In addition, we assume that $Q = \lim_{n\to\infty} X'X/n$ exists and is non-singular. Except for Section 5, where also non-nested families of models will be considered, we shall – similar as in Pötscher (1991) – consider model selection from the collection of nested models $M_{\mathcal{O}} \subseteq M_{\mathcal{O}+1} \subseteq \cdots \subseteq M_P$, where for $0 \leq p \leq P$ the model M_p is given by

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

Clearly, the model M_p corresponds to the situation where only the first p regressors in (5) are included. For the most parsimonious model under consideration, i.e., for $M_{\mathcal{O}}$, we assume that $\mathcal{O}$ satisfies $0 \leq \mathcal{O} < P$; if $\mathcal{O} > 0$, this model contains as free parameters only those components of the parameter vector θ that are not subject to model selection. (In the notation used in connection with (1) we then have $\chi = (\theta_1, \ldots, \theta_{\mathcal{O}})'$ and $\psi =$

$(\theta_{\mathcal{O}+1},\ldots,\theta_P)'$.) Furthermore, note that $M_0 = \{(0,\ldots,0)'\}$ and that $M_P = \mathbf{R}^P$. We call M_p the regression model of order p.

The following notation will prove useful. For matrices B and C of the same row-dimension, the column-wise concatenation of B and C is denoted by $(B : C)$. If D is an $m \times P$ matrix, let $D[p]$ denote the $m \times p$ matrix consisting of the first p columns of D. Similarly, let $D[\neg p]$ denote the $m \times (P-p)$ matrix consisting of the last $P-p$ columns of D. 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 the above notation also in the 'boundary' cases $p = 0$ and $p = P$. It will always be clear from the context how expressions containing symbols like $D[0]$, $D[\neg P]$, $x[0]$, or $x[\neg P]$ are to be interpreted.) As usual, the i-th component of a vector x will be denoted by x_i, and the entry in the i-th row and j-th column of a matrix B is denoted by $B_{i,j}$.

The restricted least-squares estimator for θ under the restriction $\theta[\neg p] = 0$, i.e., under $\theta_{p+1} = \cdots = \theta_P = 0$, will be denoted by $\tilde{\theta}(p)$, $0 \le p \le P$ (in case $p = P$ the restriction being void). Note that $\tilde{\theta}(p)$ is given by the $P \times 1$ vector

$$\tilde{\theta}(p) = \begin{pmatrix} (X[p]'X[p])^{-1} X[p]'Y \\ (0,\ldots,0)' \end{pmatrix},$$

where the expressions $\tilde{\theta}(0)$ and $\tilde{\theta}(P)$, respectively, are to be interpreted as the zero-vector in $\mathbf{R}^P$ and as the unrestricted least-squares estimator for θ. Given a parameter vector θ in $\mathbf{R}^P$, the order of θ is defined as

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

Hence, if θ is the true parameter vector, only models M_p of order $p \ge p_0(\theta)$ are correct models. We stress that $p_0(\theta)$ is a property of a *single parameter*, 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 of parameters* M_p.

A model selection procedure is now nothing else than a data-driven (measurable) rule $\hat{p}$ that selects a value from $\{\mathcal{O},\ldots,P\}$ and thus selects a model from the list of candidate models $M_{\mathcal{O}},\ldots,M_P$. Except for Section 5, which treats general model selection procedures including Akaike's AIC, we shall consider in this paper as a leading case a model selection procedure based on a sequence of hypothesis tests. This procedure is given as follows: The sequence of hypotheses $H_0^p : \ p_0(\theta) < p$ is tested against the alternatives $H_1^p : \ p_0(\theta) = p$ in decreasing order starting at $p = P$. If, for some $p > \mathcal{O}$, H_0^p is the first hypothesis in the process that is rejected, we set $\hat{p} = p$. If no rejection occurs until even $H_0^{\mathcal{O}+1}$ is accepted, we set $\hat{p} = \mathcal{O}$. Each hypothesis in this sequence is tested by a kind of t-test where the error variance is always estimated from the overall model. More formally, we have

$$\hat{p} = \max\{p : \ |T_p| \ge c_p, \ 0 \le p \le P\}, \tag{6}$$

with $c_{\mathcal{O}} = 0$ in order to ensure a well-defined $\hat{p}$ in the range $\{\mathcal{O}, \mathcal{O}+1,\ldots,P\}$. The test-statistics are given by

$$T_p = \frac{\sqrt{n}\tilde{\theta}_p(p)}{\hat{\sigma}\xi_{n,p}} \qquad (0 \le p \le P)$$

with the convention that $T_0 = 0$. Furthermore,

$$\xi_{n,p} = \left(\left[\left(\frac{X[p]'X[p]}{n} \right)^{-1} \right]_{p,p} \right)^{\frac{1}{2}} \qquad (0 < p \le P)$$

denotes the square root of the p-th diagonal element of the matrix indicated, and $\hat{\sigma}^2$ is given by

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

The critical values c_p are independent of sample size and satisfy $0 < c_p < \infty$ for $\mathcal{O} < p \leq P$ (but see also Remark 6.2). Note that under the hypothesis H_0^p the statistic T_p is t-distributed with $n-P$ degrees of freedom for $0 < p \leq P$. It is also easy to see that the model selection procedure $\hat{p}$ has the property that the probability of selecting an incorrect model, i.e., the probability of the event $\{\hat{p} < p_0(\theta)\}$, converges to zero as the sample size increases. In contrast, the probability of selecting a correct (but possibly overparameterized) model, i.e., the probability of the event $\{\hat{p} = p\}$ for p satisfying $p \geq \max\{p_0(\theta), \mathcal{O}\}$, converges to a positive limit; cf., for example, Proposition 5.4 in Leeb (2003b).

The post-model-selection estimator $\tilde{\theta}$ can now be defined as follows: On the event $\hat{p} = p$, $\tilde{\theta}$ is given by the restricted least-squares estimator $\tilde{\theta}(p)$, i.e.,

$$\tilde{\theta} = \sum_{p=\mathcal{O}}^{P} \tilde{\theta}(p)\, \mathbf{1}(\hat{p} = p), \tag{7}$$

where $\mathbf{1}(\cdot)$ denotes an indicator function.

3 The Conditional Distribution of the Post-Model-Selection Estimator

In this section we introduce the distribution function of a linear transformation of $\tilde{\theta}$ conditional on the event $\hat{p} = p$ and summarize some of its properties that will be needed in the subsequent development. To this end, let A be a non-stochastic $k \times P$ matrix of rank k, $1 \leq k \leq P$. For $\mathcal{O} \leq p \leq P$ we then consider the conditional cdf

$$G_{n,\theta,\sigma}(t|p) = P_{n,\theta,\sigma}\left(\sqrt{n}A(\tilde{\theta} - \theta) \leq t \middle| \hat{p} = p\right) \qquad (t \in \mathbf{R}^k). \tag{8}$$

Here $P_{n,\theta,\sigma}(\cdot)$ denotes the probability measure corresponding to a sample of size n from (5), and $P_{n,\theta,\sigma}(\cdot|\hat{p} = p)$ denotes the associated conditional probability measure (the conditioning event always having positive probability; cf. (14) in Leeb (2003b) and the attending discussion). Note that on the event $\hat{p} = p$ the expression $A(\tilde{\theta} - \theta)$ equals $A(\tilde{\theta}(p) - \theta)$ in view of (7).

Depending on the choice of the matrix A, several important scenarios are covered by (8): The conditional cdf of $\sqrt{n}(\tilde{\theta} - \theta)$ is obtained by setting A equal to the $P \times P$ identity matrix I_P. The conditional cdf of the components of $\sqrt{n}(\tilde{\theta} - \theta)$ that are not restricted to zero in the selected model M_p, $p > 0$, is obtained by setting A to the $p \times P$ matrix $(I_p : 0)$. In case $\mathcal{O} > 0$, the conditional cdf of those components of $\sqrt{n}(\tilde{\theta} - \theta)$ which correspond to the parameter of interest χ in (1) can be studied by setting A to the $\mathcal{O} \times P$ matrix $(I_{\mathcal{O}} : 0)$ as we then have $A\theta = (\theta_1, \ldots, \theta_{\mathcal{O}})' = \chi$. Finally, if A is an $1 \times P$ vector, we obtain the conditional distribution of a linear predictor based on the post-model-selection estimator. See the examples at the end of Section 4.2 for more discussion.

The cdf $G_{n,\theta,\sigma}(t|p)$ and its properties have been analyzed in detail in Leeb and Pötscher (2003a) and Leeb (2003a). To be able to access these results we need some further notation. The expected value of the restricted least-squares estimator $\tilde{\theta}(p)$ will be denoted by $\eta_n(p)$ and is given by the $P \times 1$ vector

$$\eta_n(p) = \begin{pmatrix} \theta[p] \; + \; (X[p]'X[p])^{-1}X[p]'X[\neg p]\theta[\neg p] \\ (0,\ldots,0)' \end{pmatrix} \tag{9}$$

with the conventions that $\eta_n(0) = (0,\ldots,0)' \in \mathbf{R}^P$ and that $\eta_n(P) = \theta$. Furthermore, let $\Phi_{n,p}(\cdot)$ denote the cdf of $\sqrt{n}A(\tilde{\theta}(p) - \eta_n(p))$, i.e., the cdf of $\sqrt{n}A$ times the restricted least-squares estimator based on model M_p centered at its mean. Hence, $\Phi_{n,p}(\cdot)$ is the cdf of a k-variate Gaussian random vector with mean zero and variance-covariance matrix $\sigma^2 A[p](X[p]'X[p]/n)^{-1}A[p]'$ in case $p > 0$, and it is the cdf of point-mass at zero in $\mathbf{R}^k$ in case $p = 0$. If $p > 0$ and if the matrix $A[p]$ has full row rank k, then $\Phi_{n,p}(\cdot)$ has a density with respect to Lebesgue measure, and we shall denote this density by $\phi_{n,p}(\cdot)$. We note that $\eta_n(p)$ depends on θ and that $\Phi_{n,p}(\cdot)$ depends on σ (in case $p > 0$), although these dependencies are not shown explicitly in the notation.

For $p > 0$ we introduce

$$b_{n,p} = C_n^{(p)'}(A[p](X[p]'X[p]/n)^{-1}A[p]')^{-}, \tag{10}$$

and

$$\zeta_{n,p}^2 = \xi_{n,p}^2 - C_n^{(p)'}(A[p](X[p]'X[p]/n)^{-1}A[p]')^{-}C_n^{(p)}, \tag{11}$$

with $C_n^{(p)} = A[p](X[p]'X[p]/n)^{-1}e_p$, where e_p denotes the p-th standard basis vector in $\mathbf{R}^p$, and B^- denotes a generalized inverse of a matrix B. (Observe that $\zeta_{n,p}^2$ is invariant under the choice of the generalized inverse. The same is not necessarily true for $b_{n,p}$, but is true for $b_{n,p}z$ for all z in the column-space of $A[p]$. Also note that (13) below only depends on $b_{n,p}z$ with z in the column-space of $A[p]$.) We observe that the vector of covariances between $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ is precisely given by $\sigma^2 n^{-1} C_n^{(p)}$ (and hence does *not* depend on θ). Furthermore, observe that $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are uncorrelated if and only if $\zeta_{n,p}^2 = \xi_{n,p}^2$ if and only if $b_{n,p}z = 0$ for all z in the column-space of $A[p]$; cf. Lemma A.2 in Leeb (2003a).

Finally, for a univariate Gaussian random variable $\mathfrak{N}$ with zero mean and variance $s^2 \geq 0$, we write $\Delta_s(a,b)$ for $P(|\mathfrak{N} - a| < b)$, $a \in \mathbf{R} \cup \{-\infty, \infty\}$, $b \in \mathbf{R}$. Note that $\Delta_s(\cdot,\cdot)$ is symmetric around zero in its first argument, and that $\Delta_s(-\infty, b) = \Delta_s(\infty, b) = 0$ holds. In case $s = 0$, $\mathfrak{N}$ is to be interpreted as being equal to zero, hence $\Delta_0(a,b)$ reduces to the indicator function of the interval $(-b, b)$.

We are now in a position to present the explicit formulae for $G_{n,\theta,\sigma}(t|p)$ derived in Leeb (2003a). In case $p = \mathcal{O}$ we have

$$G_{n,\theta,\sigma}(t|\mathcal{O}) = \Phi_{n,\mathcal{O}}\left(t - \sqrt{n}A(\eta_n(\mathcal{O}) - \theta)\right), \tag{12}$$

i.e., the cdf of (a linear function of) the post-model-selection estimator $\tilde{\theta}$ conditional on $\hat{p} = \mathcal{O}$ coincides with the cdf of (this linear function of) the restricted least-squares estimator $\tilde{\theta}(\mathcal{O})$. However, in case $p > \mathcal{O}$ we have

$$G_{n,\theta,\sigma}(t|p) = \int_{z \leq t - \sqrt{n}A(\eta_n(p) - \theta)} m_{n,p,\theta,\sigma}(z)\Phi_{n,p}(dz). \tag{13}$$

In the above display, $\Phi_{n,p}(dz)$ denotes integration with respect to the measure induced by the normal cdf $\Phi_{n,p}(\cdot)$ on $\mathbf{R}^k$ and the integrand $m_{n,p,\theta,\sigma}(z)$ is given by

$$m_{n,p,\theta,\sigma}(z) = \Big[\int_0^\infty (1 - \Delta_{\sigma\zeta_{n,p}}(\sqrt{n}\eta_{n,p}(p) + b_{n,p}z, sc_p\sigma\xi_{n,p})) \prod_{q=p+1}^{P} \Delta_{\sigma\xi_{n,q}}(\sqrt{n}\eta_{n,q}(q), sc_q\sigma\xi_{n,q})h(s)ds \Big] / P_{n,\theta,\sigma}(\hat{p} = p), \tag{14}$$

where the model selection probability $P_{n,\theta,\sigma}(\hat{p}=p)$ is given by

$$P_{n,\theta,\sigma}(\hat{p}=p) = \Bigg[\int_0^\infty (1-\Delta_{\sigma\xi_{n,p}}(\sqrt{n}\eta_{n,p}(p), sc_p\sigma\xi_{n,p})) \prod_{q=p+1}^{P} \Delta_{\sigma\xi_{n,q}}(\sqrt{n}\eta_{n,q}(q), sc_q\sigma\xi_{n,q})h(s)ds\Bigg]. \tag{15}$$

In the two displays above, 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. The conditional finite-sample distribution of the post-model-selection estimator given in (13) is not normal, an exception being the case where $C_n^{(p)}=0$, i.e., when $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are uncorrelated; see Leeb (2003a, Section 3.3) for more discussion. On the other extreme, namely if $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are perfectly correlated, $\zeta_{n,p}=0$ holds and the function $\Delta_{\sigma\zeta_{n,p}}$ reduces to an indicator function. This is, e.g., the case if $A=I_P$ or if $A=(I_p:0)$.

To describe the large-sample limit of $G_{n,\theta,\sigma}(t|p)$, some further notation is necessary. For p satisfying $0<p\le P$, partition the matrix $Q=\lim_{n\to\infty} X'X/n$ 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 a $p\times p$ matrix. Let $\Phi_{\infty,p}(\cdot)$ be the cdf of a k-variate Gaussian random vector with mean zero and variance-covariance matrix $\sigma^2 A[p]Q[p:p]^{-1}A[p]'$, $0<p\le P$, and let $\Phi_{\infty,0}(\cdot)$ denote the cdf of point-mass at zero in $\mathbf{R}^k$. Note that $\Phi_{\infty,p}(\cdot)$ has a Lebesgue density if $p>0$ and the matrix $A[p]$ has full row rank k; in this case, we denote the Lebesgue density of $\Phi_{\infty,p}(\cdot)$ by $\phi_{\infty,p}(\cdot)$. Finally, for $p=1,\ldots,P$, define

$$\xi^2_{\infty,p} = (Q[p:p]^{-1})_{p,p},$$

$$\zeta^2_{\infty,p} = \xi^2_{\infty,p} - C_\infty^{(p)\prime}(A[p]Q[p:p]^{-1}A[p]')^{-}C_\infty^{(p)}, \tag{16}$$

$$b_{\infty,p} = C_\infty^{(p)\prime}(A[p]Q[p:p]^{-1}A[p]')^{-},$$

where $C_\infty^{(p)} = A[p]Q[p:p]^{-1}e_p$, with e_p denoting the p-th standard basis vector in $\mathbf{R}^p$. As the notation suggests, $\Phi_{\infty,p}(t)$ is the large-sample limit of $\Phi_{n,p}(t)$, and $C_\infty^{(p)}$, $\xi^2_{\infty,p}$, and $\zeta^2_{\infty,p}$ are the limits of $C_n^{(p)}$, $\xi^2_{n,p}$, and $\zeta^2_{n,p}$, respectively; moreover, $b_{n,p}z$ converges to $b_{\infty,p}z$ for each z in the column-space of $A[p]$. See Lemma A.2 in Leeb (2003a).

The next result is a special case of Corollary 5.4 in Leeb (2003a) and describes the large-sample limit of the conditional cdf under local alternatives to θ, under the assumption that the selected model M_p is a correct model for θ.

Proposition 3.1 *Let p satisfy $\mathcal{O}\le p\le P$. Suppose $\theta\in\mathbf{R}^P$ satisfies $\theta\in M_p$, i.e., $p_0(\theta)\le p$ holds. Moreover, let $\gamma\in\mathbf{R}^P$ and let $\sigma^{(n)}$ be a sequence of positive real numbers which converges to a (finite) limit $\sigma>0$ as*

[1]Recall that the total variation distance $||G-G^*||_{TV}$ between two cdfs G and G^* on $\mathbf{R}^k$ is defined as the total variation distance between the probability measures Q and Q^* induced by G and G^*, respectively. The total variation distance between Q and Q^* is defined as $||Q-Q^*||_{TV}=\sup_E |Q(E)-Q^*(E)|$, where the supremum is taken over all Borel sets E. Clearly, the relation $|G(t)-G^*(t)|\le ||G-G^*||_{TV}$ holds for all $t\in\mathbf{R}^k$. Thus, if G and G^* are close with respect to the total variation distance, then $G(t)$ is close to $G^*(t)$, uniformly in t.

$n \to \infty$. *Then the conditional cdf* $G_{n,\theta+\gamma/\sqrt{n},\sigma^{(n)}}(t|p)$ *converges to a limit* $G_{\infty,\theta,\sigma,\gamma}(t|p)$ *in total variation*[1] *, i.e.,*

$$\left|\left|G_{n,\theta+\gamma/\sqrt{n},\sigma^{(n)}}(\cdot|p) - G_{\infty,\theta,\sigma,\gamma}(\cdot|p)\right|\right|_{TV} \overset{n\to\infty}{\longrightarrow} 0. \tag{17}$$

The large-sample limit cdf $G_{\infty,\theta,\sigma,\gamma}(t|p)$ *is given as follows: In case* $p = \max\{p_0(\theta), \mathcal{O}\}$

$$G_{\infty,\theta,\sigma,\gamma}(t|p) = \Phi_{\infty,p}(t - \beta^{(p)}), \tag{18}$$

where

$$\beta^{(p)} = A \begin{pmatrix} Q[p:p]^{-1}Q[p:\neg p]\gamma[\neg p] \\ -\gamma[\neg p] \end{pmatrix},$$

with the convention that $\beta^{(p)} = -A\gamma$ *if* $p = 0$ *and that* $\beta^{(p)} = (0, \ldots, 0)'$ *if* $p = P$. *In case* $p > \max\{p_0(\theta), \mathcal{O}\}$

$$G_{\infty,\theta,\sigma,\gamma}(t|p) = \int_{z\le t-\beta^{(p)}} \frac{1 - \Delta_{\sigma\zeta_{\infty,p}}(\nu_p + b_{\infty,p}z, c_p\sigma\xi_{\infty,p})}{1 - \Delta_{\sigma\xi_{\infty,p}}(\nu_p, c_p\sigma\xi_{\infty,p})} \Phi_{\infty,p}(dz), \tag{19}$$

where $\nu_p = \gamma_p + (Q[p:p]^{-1}Q[p:\neg p]\gamma[\neg p])_p$. *(Note that* $\beta^{(p)} = \lim_{n\to\infty}\sqrt{n}A(\eta_n(p) - \theta - \gamma/\sqrt{n})$ *because* $\theta \in M_p$, *and that* $\nu_p = \lim_{n\to\infty}\sqrt{n}\eta_{n,p}(p)$ *in case* $\theta \in M_{p-1}$, *i.e.,* $p > p_0(\theta)$. *Here* $\eta_n(p)$ *is defined as in (9), but with* $\theta + \gamma/\sqrt{n}$ *replacing* θ.*)*

If $p > 0$ *and if the matrix* $A[p]$ *has full row rank* k, *then the Lebesgue density* $\phi_{\infty,p}(\cdot)$ *of* $\Phi_{\infty,p}(\cdot)$ *exists; the density of (18) is then given by* $\phi_{\infty,p}(t - \beta^{(p)})$, *while the density of (19) is given by the integrand in (19) times* $\phi_{\infty,p}(z)$, *evaluated at* $z = t - \beta^{(p)}$.

While the limiting cdf in (18) is Gaussian, the limiting cdf in (19) typically is not, an exception being the case where $C_\infty^{(p)} = 0$, i.e., when $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are asymptotically uncorrelated. In that case, the expressions in (18) and (19) coincide. Also note that the cdf $G_{\infty,\theta,\sigma,\gamma}(t|p)$ has been defined above only for $\theta \in M_p$ (and $\mathcal{O} \le p \le P$). If $\gamma = 0$, we write $G_{\infty,\theta,\sigma}(t|p)$ as shorthand for $G_{\infty,\theta,\sigma,0}(t|p)$ in the following.

Proposition 3.1 is restricted to sequences of parameters $\theta + \gamma/\sqrt{n}$ with $p_0(\theta) \le p$. The case where the selected model M_p is an incorrect model for θ, i.e., where we have $p_0(\theta) > p$, is analyzed in Corollary 5.4(c)-(d) of Leeb (2003a). For the results in the present paper, however, we shall only need to rely on the situation covered by Proposition 3.1. The reason essentially is that only over $1/\sqrt{n}$-'neighborhoods' of M_p the probability of actually selecting the model M_p is bounded away from zero. In contrast, for every fixed $\theta \notin M_p$ the probability of selecting the model M_p converges to zero as $n \to \infty$.

Proposition 3.2 *Let* p *satisfy* $\mathcal{O} \le p \le P$, *and let* r_n *be a sequence of positive real numbers.*

a. If $r_n = O(1/\sqrt{n})$ *as* $n \to \infty$, *then*

$$\liminf_{n\to\infty} \inf_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta[\neg p]||<r_n}} P_{n,\vartheta,\sigma}(\hat{p} = p) > 0 \tag{20}$$

holds for every σ, $0 < \sigma < \infty$. *(The infimum in the above display is to be interpreted as extending over* $\|\vartheta\| < r_n$ *if* $p = 0$ *and over all of* $\mathbf{R}^P$ *if* $p = P$*). In particular, it follows that* $\liminf_{n\to\infty} \inf_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<r_n}} P_{n,\vartheta,\sigma}(\hat{p} = p) > 0$ *for each* $\theta \in M_p$ *and* $0 < \sigma < \infty$.

b. Suppose $p < P$ holds. If $\sqrt{n}r_n \to \infty$ as $n \to \infty$, then

$$\lim_{n\to\infty} \inf_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<r_n}} P_{n,\vartheta,\sigma}(\hat{p} = p) = 0 \tag{21}$$

for each $\theta \in M_p$ and $0 < \sigma < \infty$.

c. If an infimum (resp. supremum) over $\sigma \in [\sigma_*, \sigma^*]$, $0 < \sigma_* \leq \sigma^* < \infty$, is inserted in (20) (resp. (21)) immediately after the limes inferior (resp. limes) operator, the result continues to hold.

4 Estimators for the Conditional Finite-Sample Distribution

For the purpose of inference after model selection the conditional finite-sample distribution of the post-model-selection-estimator is an object of particular interest. As we have seen, it depends on unknown parameters in a complicated manner, and hence one will have to be content with estimators for this cdf. The object we would primarily like to estimate is

$$G_{n,\theta,\sigma}(t|\hat{p}) = \sum_{p=\mathcal{O}}^{P} G_{n,\theta,\sigma}(t|p)\mathbf{1}(\hat{p} = p),$$

i.e., the conditional cdf after the model selection procedure has returned the model order $\hat{p}$. Estimation of $G_{n,\theta,\sigma}(t|p)$, i.e., of the conditional cdf given the event $\hat{p} = p$, is obviously closely related and will also be considered. As we shall see, it is not difficult to construct consistent estimators for $G_{n,\theta,\sigma}(t|\hat{p})$ and $G_{n,\theta,\sigma}(t|p)$ (over M_p in the latter case). However, we also find that *any* estimator for $G_{n,\theta,\sigma}(t|\hat{p})$ (or $G_{n,\theta,\sigma}(t|p)$) typically performs unsatisfactory, in that the estimation error can not become small uniformly over (subsets of) the parameter space even as sample size goes to infinity. In particular, no uniformly consistent estimators exist, not even locally.

4.1 Consistent Estimators

We construct a consistent estimator for $G_{n,\theta,\sigma}(t|\hat{p})$ and $G_{n,\theta,\sigma}(t|p)$ (over M_p in the latter case) by commencing from the asymptotic distribution. Specializing to the case $\gamma = 0$ and $\sigma^{(n)} = \sigma$ in Proposition 3.1, the large-sample limit of $G_{n,\theta,\sigma}(t|p)$ is given by $G_{\infty,\theta,\sigma}(t|p) = \Phi_{\infty,p}(t)$ in case $p = \max\{p_0(\theta), \mathcal{O}\}$, and by

$$G_{\infty,\theta,\sigma}(t|p) = \int_{\substack{z\in\mathbf{R}^k \\ z\leq t}} \frac{1 - \Delta_{\sigma\zeta_{\infty,p}}(b_{\infty,p}z, c_p\sigma\xi_{\infty,p})}{1 - \Delta_{\sigma\xi_{\infty,p}}(0, c_p\sigma\xi_{\infty,p})} \Phi_{\infty,p}(dz) \tag{22}$$

in case $p > \max\{p_0(\theta), \mathcal{O}\}$. Let $\hat{\Phi}_{n,p}(\cdot)$ denote the cdf of a k-variate Gaussian random vector with mean zero and variance-covariance matrix $\hat{\sigma}^2 A[p](X[p]'X[p]/n)^{-1}A[p]'$, $0 < p \leq P$, and let $\hat{\Phi}_{n,0}(\cdot)$ denote the cdf of point-mass at zero in $\mathbf{R}^k$. For given p, $\mathcal{O} \leq p \leq P$, an estimator $\check{G}_n(t|p)$ for $G_{n,\theta,\sigma}(t|p)$ is now defined as follows: For $p = \mathcal{O}$, we set $\check{G}_n(t|\mathcal{O}) = \hat{\Phi}_{n,\mathcal{O}}(t)$. For $p > \mathcal{O}$, we first employ an auxiliary procedure that consistently decides between $p_0(\theta) = p$ and $p_0(\theta) < p$, i.e., between $\theta \in M_p \diagdown M_{p-1}$ and $\theta \in M_{p-1}$, for every $\theta \in M_p$. (E.g., the procedure that decides for $p_0(\theta) = p$ whenever $|T_p| > s_{n,p}$ and for $p_0(\theta) < p$ otherwise, with $s_{n,p}$ satisfying $s_{n,p} \to \infty$, $s_{n,p} = o(n^{1/2})$ for $n \to \infty$ can be used. Alternatively, a consistent model selection procedure like BIC could be employed.) If the procedure decides for $p_0(\theta) = p$, we set $\check{G}_n(t|p) = \hat{\Phi}_{n,p}(t)$; otherwise we set $\check{G}_n(t|p)$ to the expression in (22) with $\hat{\sigma}$, $b_{n,p}$, $\zeta_{n,p}$, $\xi_{n,p}$, and $\hat{\Phi}_{n,p}(\cdot)$ replacing σ, $b_{\infty,p}$, $\zeta_{\infty,p}$, $\xi_{\infty,p}$, and $\Phi_{\infty,p}(\cdot)$, respectively. A little reflection shows that $\check{G}_n(t|p)$ is again a cdf (observe that $\check{G}_n(t|p)$

coincides with the conditional cdf $G^*_{n,\theta,\sigma}(t|p)$ given in (13) of Leeb (2003a) for $\sigma = \hat{\sigma}$). This gives an estimator $\check{G}_n(t|p)$ for $G_{n,\theta,\sigma}(t|p)$; as an estimator for $G_{n,\theta,\sigma}(t|\hat{p})$ we shall use $\check{G}_n(t|\hat{p}) = \sum_{p=\mathcal{O}}^{P} \check{G}_n(t|p)\mathbf{1}(\hat{p} = p)$. We have the following consistency results.

Proposition 4.1 *Let p satisfy $\mathcal{O} \leq p \leq P$. Then the estimator $\check{G}_n(t|p)$ is consistent (in the total variation distance) for $G_{n,\theta,\sigma}(t|p)$ and $G_{\infty,\theta,\sigma}(t|p)$ over the subset M_p (and over $0 < \sigma < \infty$). That is, for every $\delta > 0$*

$$P_{n,\theta,\sigma}\left(\left|\left|\check{G}_n(\cdot|p) - G_{n,\theta,\sigma}(\cdot|p)\right|\right|_{TV} > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{23}$$

$$P_{n,\theta,\sigma}\left(\left|\left|\check{G}_n(\cdot|p) - G_{\infty,\theta,\sigma}(\cdot|p)\right|\right|_{TV} > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{24}$$

for all $\theta \in M_p$ and all $\sigma > 0$. (The results (23) and (24) also hold with $P_{n,\theta,\sigma}(\cdot|\hat{p} = p)$ replacing $P_{n,\theta,\sigma}(\cdot)$.)

Corollary 4.2 *The estimator $\check{G}_n(t|\hat{p})$ is consistent (in the total variation distance) for $G_{n,\theta,\sigma}(t|\hat{p})$ over the entire parameter space, i.e., for every $\delta > 0$*

$$P_{n,\theta,\sigma}\left(\left|\left|\check{G}_n(\cdot|\hat{p}) - G_{n,\theta,\sigma}(\cdot|\hat{p})\right|\right|_{TV} > \delta\right) \overset{n\to\infty}{\longrightarrow} 0$$

for all $\theta \in \mathbf{R}^P$ and $\sigma > 0$.

While the estimators constructed above are consistent, they can be expected to perform poorly in finite samples when the true θ belongs to $M_p \backslash M_{p-1}$ but is 'close' to M_{p-1}, since the auxiliary decision procedure (although being consistent) will then have difficulties making the correct decision in finite samples and since $G_{n,\theta,\sigma}(\cdot|p)$ typically does not converge uniformly with respect to $\theta \in M_p \backslash M_{p-1}$ 'close' to M_{p-1} (cf. Remark 6.6 below and Corollary 4.6 in Leeb and Pötscher (2003a)). In the next section we show that this poor performance is not particular to the estimators constructed above, but is a genuine feature of the estimation problem under consideration.

4.2 Performance Limits for Estimators of the Conditional Distribution Function

In the following we provide lower bounds for the performance of estimators of the conditional cdf of the post-model-selection estimator $A\tilde{\theta}$; that is, we give lower bounds on the probability that the estimation error exceeds a certain threshold. In particular, these results imply that no uniformly consistent estimator for the conditional cdf exists, not even locally. We first provide results for estimation of $G_{n,\theta,\sigma}(t|p)$, which are then used as building blocks for the corresponding results for estimation of $G_{n,\theta,\sigma}(t|\hat{p})$. In all these results, the asymptotic correlation between $A[p]\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ as measured by $C_\infty^{(p)}$ will play an important rôle. Note that $C_\infty^{(p)}$ is given by $C_\infty^{(p)} = A[p]Q[p:p]^{-1}e_p$ and hence does *not* depend on the value of the unknown parameters θ or σ. We note that in the important special case discussed in the Introduction, cf. (1), the matrix A equals the $\mathcal{O} \times P$ matrix $(I_{\mathcal{O}} : 0)$, and the condition $C_\infty^{(p)} \neq 0$ reduces to the condition that the regressor corresponding to the p-th column of (V, W) is asymptotically correlated with at least one of the regressors corresponding to the columns of V. See Example 1 below for more discussion.

In the three results to follow we shall consider estimators for $G_{n,\theta,\sigma}(t|p)$ at a *fixed* value of the argument t. An estimator for $G_{n,\theta,\sigma}(t|p)$ is now nothing else than a real-valued random variable $\Gamma_n = \Gamma_n(Y, X)$. For mnemonic reasons we shall, however, use the symbol $\hat{G}_n(t|p)$ instead of Γ_n to denote an arbitrary estimator for $G_{n,\theta,\sigma}(t|p)$. This notation should not be taken as implying that the estimator is obtained by evaluating an estimated cdf at the argument t, or that it is constrained to lie between zero and one.

Theorem 4.3 *Let p satisfy $\mathcal{O} < p \le P$. Suppose that $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are asymptotically correlated, i.e., $C_\infty^{(p)} \ne 0$, and assume that p is the largest model order with this property. Then the following holds for each $\theta \in M_{p-1}$, $0 < \sigma < \infty$, and for each $t \in \mathbf{R}^k$:*

a. There exist $\delta_0 > 0$ and $\rho_0 > 0$ such that any estimator $\hat{G}_n(t|p)$ for $G_{n,\theta,\sigma}(t|p)$ satisfying

$$P_{n,\theta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\theta,\sigma}(t|p)\right| > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{25}$$

for each $\delta > 0$ (in particular, every estimator that is consistent over M_p) also satisfies

$$\sup_{\substack{\vartheta \in M_p \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\vartheta,\sigma}(t|p)\right| > \delta_0\right) \overset{n\to\infty}{\longrightarrow} 1. \tag{26}$$

The constants δ_0 and ρ_0 may be chosen in such a way that they depend only on t, Q, A, σ, and the critical value c_p. Moreover,

$$\liminf_{n\to\infty} \inf_{\hat{G}_n(t|p)} \sup_{\substack{\vartheta \in M_p \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\vartheta,\sigma}(t|p)\right| > \delta_0\right) > 0 \tag{27}$$

and

$$\sup_{\delta>0} \liminf_{n\to\infty} \inf_{\hat{G}_n(t|p)} \sup_{\substack{\vartheta \in M_p \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\vartheta,\sigma}(t|p)\right| > \delta\right) \ge \frac{1}{2} \tag{28}$$

hold, where the infima in (27) and (28) extend over all *estimators $\hat{G}_n(t|p)$ of $G_{n,\theta,\sigma}(t|p)$.*

b. The above continues to hold with $P_{n,\cdot,\sigma}(\cdot|\hat{p} = p)$ replacing $P_{n,\cdot,\sigma}(\cdot)$.

Proposition 4.4 *Suppose that the assumptions of Theorem 4.3 are satisfied, except that p is not the largest model order q for which $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically correlated. Then the results in Theorem 4.3 continue to hold for each $\theta \in M_{p-1}$, $0 < \sigma < \infty$, and for each $t \in \mathbf{R}^k$ with the property that the set $\{z \in \mathbf{R}^p : A[p]z \le t\}$ has positive Lebesgue measure in $\mathbf{R}^p$. (This condition is satisfied if $A[p]z < t$ holds for some $z \in \mathbf{R}^p$. In particular, the condition is satisfied for each $t \in \mathbf{R}^k$ if the matrix $A[p]$ has full row rank k.)*

As a point of interest we note that the non-uniformity phenomenon described in Theorem 4.3 and Proposition 4.4 occurs *within* the model M_p, which contains only parameters for which the selected model is correct; i.e., in (26)-(28) the suprema with respect to ϑ extend only over subsets of M_p. That is, it is typically even impossible to construct an estimator for $G_{n,\theta,\sigma}(t|p)$ which performs satisfactory for those local perturbations ϑ of the true parameter θ, for which the selected model is correct.

Consider next the case where neither Theorem 4.3 nor Proposition 4.4 apply (i.e., the model order p under consideration is such that either $p = \mathcal{O}$, or $C_\infty^{(p)} = 0$, or $C_\infty^{(p)} \ne 0$ but the set $\{z \in \mathbf{R}^p : A[p]z \le t\}$ has Lebesgue measure zero). In that case, it is indeed possible to construct an estimator for $G_{n,\theta,\sigma}(t|p)$ that is uniformly consistent over $\theta \in M_p$. However, this result provides little consolation, because the uniform consistency over $\theta \in M_p$ typically breaks down already in $1/\sqrt{n}$-'neighborhoods' of M_p, and results analogous to (26)-(28) can be established over such neighborhoods, even if Theorem 4.3 or Proposition 4.4 do not apply. This is of relevance as true parameter values in such $1/\sqrt{n}$-'neighborhoods' result in a positive probability of selecting the model M_p, cf. Proposition 3.2:

Theorem 4.5 *Let p satisfy $\mathcal{O} \leq p < P$. Suppose that $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically correlated, i.e., $C_\infty^{(q)} \neq 0$, for some q satisfying $p < q \leq P$, and let q^* denote the largest q with this property. Then the following holds for each $\theta \in M_p$, $0 < \sigma < \infty$, and for each $t \in \mathbf{R}^k$:*

a. There exist $\delta_0 > 0$ and $\rho_0 > 0$ such that any estimator $\hat{G}_n(t|p)$ for $G_{n,\theta,\sigma}(t|p)$ satisfying

$$P_{n,\theta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\theta,\sigma}(t|p)\right| > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{29}$$

for each $\delta > 0$ (in particular, every estimator that is consistent over M_p) also satisfies

$$\sup_{\substack{\vartheta \in M_{q^*} \\ ||\vartheta - \theta|| < \rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\vartheta,\sigma}(t|p)\right| > \delta_0\right) \overset{n\to\infty}{\longrightarrow} 1. \tag{30}$$

The constants δ_0 and ρ_0 may be chosen in such a way that they depend only on t, Q, A, σ, and the critical value c_p. Moreover,

$$\liminf_{n\to\infty} \inf_{\hat{G}_n(t|p)} \sup_{\substack{\vartheta \in M_{q^*} \\ ||\vartheta - \theta|| < \rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\vartheta,\sigma}(t|p)\right| > \delta_0\right) > 0 \tag{31}$$

and

$$\sup_{\delta>0} \liminf_{n\to\infty} \inf_{\hat{G}_n(t|p)} \sup_{\substack{\vartheta \in M_{q^*} \\ ||\vartheta - \theta|| < \rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|p) - G_{n,\vartheta,\sigma}(t|p)\right| > \delta\right) \geq \frac{1}{2} \tag{32}$$

hold, where the infima in (31) and (32) extend over all *estimators $\hat{G}_n(t|p)$ of $G_{n,\theta,\sigma}(t|p)$.*

b. The above continues to hold with $P_{n,\cdot,\sigma}(\cdot|\hat{p} = p)$ replacing $P_{n,\cdot,\sigma}(\cdot)$.

Summarizing so far we see that it is impossible to construct an estimator for $G_{n,\theta,\sigma}(t|p)$ which performs reasonably well in a neighborhood of the true parameter θ ($\theta \in M_p$), whenever the model order p considered has the property that $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically correlated for some q with $\max\{p, \mathcal{O}+1\} \leq q \leq P$, as then either Theorem 4.3 or Theorem 4.5 applies. In particular, no uniformly consistent estimator exists, not even locally.

In the remaining case, i.e., when $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically uncorrelated for each q in the range $\max\{p, \mathcal{O}+1\} \leq q \leq P$, it is indeed possible to construct an estimator for $G_{n,\theta,\sigma}(t|p)$ which is uniformly consistent (even in the total variation distance) over $1/\sqrt{n}$-'neighborhoods' of M_p as shown in Proposition 4.6 below. In fact, this estimator is given by $\hat{\Phi}_{n,p}(t)$, i.e., by an estimated version of the cdf of $\sqrt{n}A(\tilde{\theta}(p) - \eta_n(p))$, which has been introduced in Section 4.1. Note that under the assumptions of Proposition 4.6 the limiting distribution is normal and coincides with the limit distribution of the restricted least-squares estimator, i.e., of $\sqrt{n}A(\tilde{\theta}(p) - \eta_n(p))$, cf. Lemma C.2 in Appendix C. (The asymptotic uncorrelatedness assumption in Proposition 4.6 below thus can be viewed as describing the situation in which model selection has no 'effect' on the (conditional) distribution of the estimator asymptotically.) This suggests $\hat{\Phi}_{n,p}(t)$ as an estimator for $G_{n,\theta,\sigma}(t|p)$. However, inspection of the proof of Proposition 4.6 shows that the real reason why this proposition works is that under this asymptotic uncorrelatedness assumption the convergence of the finite-sample distribution to its limit can be shown to be uniform over $1/\sqrt{n}$-'neighborhoods' of M_p (see Lemma C.2), and hence the limit under local alternatives does not depend on the alternative.

Proposition 4.6 *Let p satisfy $\mathcal{O} \leq p \leq P$. Suppose that $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically uncorrelated, i.e., $C_\infty^{(q)} = 0$, for each $q = \max\{p, \mathcal{O}+1\}, \ldots, P$. Then*

$$\sup_{\substack{\theta \in \mathbf{R}^P \\ ||\theta[\neg p]|| < \rho/\sqrt{n}}} \sup_{\substack{\sigma \in \mathbf{R} \\ \sigma_* \leq \sigma \leq \sigma^*}} P_{n,\theta,\sigma}\left(\left|\left| \hat{\Phi}_{n,p}(\cdot) - G_{n,\theta,\sigma}(\cdot|p) \right|\right|_{TV} > \delta \right) \stackrel{n\to\infty}{\longrightarrow} 0 \tag{33}$$

holds for each $\delta > 0$, for each $\rho > 0$, and for any constants σ_* and σ^* satisfying $0 < \sigma_* \leq \sigma^* < \infty$. The result (33) also holds with $P_{n,\theta,\sigma}(\cdot|\hat{p} = p)$ replacing $P_{n,\theta,\sigma}(\cdot)$. (In case $p = P$, the first supremum in (33) is to be interpreted as extending over all $\theta \in \mathbf{R}^P$. Furthermore, the case $p = 0$ is impossible in view of Proposition 4.7 below.)

If the uncorrelatedness assumptions in the proposition even hold for all finite n, then the cdf $G_{n,\theta,\sigma}(\cdot|p)$ can be seen to reduce to the normal cdf $\Phi_{n,p}(\cdot)$ and hence can be estimated uniformly consistently over the larger space $M_P \times [\sigma_*, \sigma^*]$.

Clearly, the case to which Proposition 4.6 applies is quite exceptional. In fact, under the assumptions of Proposition 4.6 the restricted estimators $A\tilde{\theta}(q)$ for $q \geq \max\{p-1, \mathcal{O}\}$ perform asymptotically as well as the unrestricted estimator $A\tilde{\theta}(P)$. This is a consequence of the following result.

Proposition 4.7 *Let p satisfy $0 < p < P$. Then the following statements are equivalent:*

a. $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically uncorrelated, i.e., $C_\infty^{(q)} = (0, \ldots, 0)'$, for each $q = p+1, \ldots, P$.

b. $A\tilde{\theta}(p)$ is an asymptotically unbiased estimator for $A\theta$ ($\theta \in \mathbf{R}^P$).

c. The asymptotic variance-covariance matrices of $\sqrt{n}A\tilde{\theta}(p)$ and $\sqrt{n}A\tilde{\theta}(P)$ are identical.

In case $p = P$ the above statements are always trivially satisfied. In case $p = 0$, these statements are never satisfied.

It is easy to see that any of the above statements is equivalent to asymptotic unbiasedness of $A\tilde{\theta}(q)$ for all $q = p, \ldots, P$, and further also is equivalent to all the asymptotic variance-covariance matrices of $\sqrt{n}A\tilde{\theta}(q)$ for $q = p, \ldots, P$, being identical. Furthermore, a finite sample version of Proposition 4.7 can also easily be derived from the discussion following (19) in Leeb (2003a). In fact, it is shown in that reference for any given sample size that uncorrelatedness of $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ for $q = p+1, \ldots, P$ is equivalent to the estimators $A\tilde{\theta}(p)$ and $A\tilde{\theta}(P)$ being identical, which in turn is equivalent to all the estimators $A\tilde{\theta}(q)$ being identical for $q = p, \ldots, P$.

We next consider performance limits for estimators of $G_{n,\theta,\sigma}(t|\hat{p})$ at a *fixed* value of the argument t. Again, an estimator for $G_{n,\theta,\sigma}(t|\hat{p})$ is nothing else than a real-valued random variable $\Gamma_n = \Gamma_n(Y, X)$, but for convenience we shall use the symbol $\hat{G}_n(t|\hat{p})$ instead of Γ_n to denote an arbitrary estimator for $G_{n,\theta,\sigma}(t|p)$. This notation should again not be taken as implying that the estimator is obtained by evaluating an estimated cdf at the argument t, or that it is constrained to lie between zero and one. Regarding the non-uniformity phenomenon, we then have a dichotomy which is described in the following two results.

Theorem 4.8 *Suppose that $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically correlated, i.e., $C_\infty^{(q)} \neq 0$, for some q satisfying $\mathcal{O} < q \leq P$, and let q^* denote the largest q with this property. Then the following holds for each $\theta \in M_{q^*-1}$,*

$0 < \sigma < \infty$*, and each* $t \in \mathbf{R}^k$*: There exist* $\delta_0 > 0$ *and* $\rho_0 > 0$ *such that any estimator* $\hat{G}_n(t|\hat{p})$ *for* $G_{n,\theta,\sigma}(t|\hat{p})$ *satisfying*

$$P_{n,\theta,\sigma}\left(\left|\hat{G}_n(t|\hat{p}) - G_{n,\theta,\sigma}(t|\hat{p})\right| > \delta\right) \stackrel{n\to\infty}{\longrightarrow} 0 \tag{34}$$

for each $\delta > 0$ *(in particular, every estimator that is consistent) also satisfies*

$$\begin{aligned} &\liminf_{n\to\infty} \sup_{\substack{\vartheta\in M_{q^*} \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|\hat{p}) - G_{n,\vartheta,\sigma}(t|\hat{p})\right| > \delta_0\right) \\ &\geq \lim_{n\to\infty} P_{n,\theta,\sigma}\left(\hat{p} = q^*\right) \\ &= (1 - \Delta_{\sigma\xi_{\infty,q^*}}(0, c_{q^*}\sigma\xi_{\infty,q^*})) \prod_{q=q^*+1}^{P} \Delta_{\sigma\xi_{\infty,q}}(0, c_q\sigma\xi_{\infty,q}) > 0. \end{aligned} \tag{35}$$

The constants δ_0 *and* ρ_0 *may be chosen in such a way that they depend only on* t, Q, A, σ*, and the critical value* c_p*. Moreover,*

$$\liminf_{n\to\infty} \inf_{\hat{G}_n(t|\hat{p})} \sup_{\substack{\vartheta\in M_{q^*} \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|\hat{p}) - G_{n,\vartheta,\sigma}(t|\hat{p})\right| > \delta_0\right) > 0 \tag{36}$$

and

$$\begin{aligned} &\sup_{\delta>0} \liminf_{n\to\infty} \inf_{\hat{G}_n(t|\hat{p})} \sup_{\substack{\vartheta\in M_{q^*} \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|\hat{p}) - G_{n,\vartheta,\sigma}(t|\hat{p})\right| > \delta\right) \\ &\geq \frac{1}{2} \lim_{n\to\infty} P_{n,\theta,\sigma}\left(\hat{p} = q^*\right) \\ &= \frac{1}{2}(1 - \Delta_{\sigma\xi_{\infty,q^*}}(0, c_{q^*}\sigma\xi_{\infty,q^*})) \prod_{q=q^*+1}^{P} \Delta_{\sigma\xi_{\infty,q}}(0, c_q\sigma\xi_{\infty,q}) > 0 \end{aligned} \tag{37}$$

hold, where the infima in (36) and (37) extend over all *estimators* $\hat{G}_n(t|\hat{p})$ *of* $G_{n,\theta,\sigma}(t|\hat{p})$*.*

Proposition 4.9 *Suppose that* $A\tilde{\theta}(q)$ *and* $\tilde{\theta}_q(q)$ *are asymptotically uncorrelated, i.e.,* $C_\infty^{(q)} = 0$*, for all* q *satisfying* $\mathcal{O} < q \leq P$*. Then*

$$\sup_{\theta\in\mathbf{R}^P} \sup_{\substack{\sigma\in\mathbf{R} \\ \sigma_*\leq\sigma\leq\sigma^*}} P_{n,\theta,\sigma}\left(\left|\left|\hat{\Phi}_{n,P}(\cdot) - G_{n,\theta,\sigma}(\cdot|\hat{p})\right|\right|_{TV} > \delta\right) \stackrel{n\to\infty}{\longrightarrow} 0 \tag{38}$$

holds for each $\delta > 0$*, and for any constants* σ_* *and* σ^* *satisfying* $0 < \sigma_* \leq \sigma^* < \infty$*.*

Inspection of the proof of Proposition 4.9 shows that (38) continues to hold if the estimator $\hat{\Phi}_{n,P}(\cdot)$ is replaced by any of the estimators $\hat{\Phi}_{n,p}(\cdot)$ for $\mathcal{O} \leq p \leq P$. We also note that in case $\mathcal{O} = 0$ the assumption of Proposition 4.9 is never satisfied (cf. Proposition 4.7), and hence Theorem 4.8 always applies in that case.

We conclude this section by illustrating the above results with some important examples.

Example 1: *(The conditional distribution of* $\tilde{\chi}$*)* Consider the model given in (1) with χ representing the parameter of interest. Using the general notation of the paper, this corresponds to the case $A\theta = (\theta_1, \ldots, \theta_{\mathcal{O}})' = \chi$ with A representing the $\mathcal{O} \times P$ matrix $(I_{\mathcal{O}} : 0)$. Here $k = \mathcal{O} > 0$. The cdf $G_{n,\theta,\sigma}(\cdot|p)$ then represents the cdf of $\sqrt{n}\,(\tilde{\chi} - \chi)$ conditional on the event $\hat{p} = p$. Assume first that

$$\lim_{n\to\infty} V'W/n \neq 0. \tag{39}$$

Then $C_\infty^{(r)} \neq 0$ holds for some $r > \mathcal{O}$. Consequently, for any such r the 'impossibility' results in Theorem 4.3 or Proposition 4.4 apply with $p = r$ (observe that $rank(A[p]) = \mathcal{O} = k$ always holds for $p = r > \mathcal{O}$). Furthermore, the 'impossibility' results in Theorem 4.5 apply for any p satisfying $\mathcal{O} \leq p < r$ for some r as above. Regarding 'impossibility' results for the estimation of $G_{n,\theta,\sigma}(t|\hat{p})$, we note that Theorem 4.8 always applies if (39) is satisfied. Next assume that $\lim_{n\to\infty} V'W/n = 0$. Then $C_\infty^{(r)} = 0$ for every $r > \mathcal{O}$. In this case Proposition 4.9 applies and a uniformly consistent estimator for $G_{n,\theta,\sigma}(t|\hat{p})$ indeed exists. (Of course, Proposition 4.6 then also holds for every $\mathcal{O} \leq p \leq P$.) Summarizing we note that any estimator for $G_{n,\theta,\sigma}(t|\hat{p})$ or for $G_{n,\theta,\sigma}(t|p)$ (for at least some p) suffers from the non-uniformity phenomenon except in the special case where the columns of V and W are asymptotically orthogonal in the sense that $\lim_{n\to\infty} V'W/n = 0$. But this is precisely the situation where inclusion or exclusion of the regressors in W has no effect on the (conditional) distribution of the estimator $\tilde{\chi}$ asymptotically; hence it is not surprising that also the model selection procedure does not have an effect on the estimation of the cdf of the post-model-selection estimator $\tilde{\chi}$. This observation may tempt one to enforce orthogonality between the columns of V and W by either replacing the columns of V by their residuals from the projection on the column space of W or vice versa. However, this is not helpful for the following reasons: In the first case one then in fact avoids model selection as all the restricted least-squares estimators for χ under consideration (and hence also the post-model selection estimator $\tilde{\chi}$) in the reparameterized model coincide with the unrestricted least-squares estimator. In the second case the coefficients of the columns of V in the reparameterized model no longer coincide with the parameter of interest χ (and again are estimated by one and the same estimator regardless of inclusion/exclusion of columns of the transformed W-matrix).

Example 2: *(The conditional distribution of $\tilde{\theta}$)* For A equal to I_P, the cdf $G_{n,\theta,\sigma}(t|p)$ is the conditional cdf of $\sqrt{n}(\tilde{\theta} - \theta)$ given $\hat{p} = p$. Here, $A\tilde{\theta}(q)$ reduces to the P-vector $(\tilde{\theta}(q)' : (0, \ldots, 0))'$, and hence $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are perfectly correlated for every $q > \mathcal{O}$. Therefore, Theorem 4.3 applies if p=P, and Theorem 4.5 applies in case $p < P$. (In the latter case, Proposition 4.4 applies as well.) Regarding estimation of $G_{n,\theta,\sigma}(t|\hat{p})$ observe that Theorem 4.8 here always applies. We therefore see that estimation of the conditional distribution of the entire parameter vector is always plagued by the non-uniformity phenomenon.

Example 3: *(The conditional distribution of the unrestricted components of $\tilde{\theta}$)* Let $r > 0$ be a given model order. Conditional on the event $\hat{p} = r$, the last $P - r$ components of $\tilde{\theta}$ are restricted to zero. If A is the $r \times P$ matrix $(I_r : 0)$, then the cdf $G_{n,\theta,\sigma}(t|r)$ is the conditional cdf of the first r (unrestricted) components of $\sqrt{n}(\tilde{\theta} - \theta)$ given the event $\hat{p} = r$. In this case $A\tilde{\theta}(r)$ equals $\tilde{\theta}(r)$, and hence $\tilde{\theta}_r(r)$ is perfectly correlated with $A\tilde{\theta}(r)$. For q satisfying $r < q \leq P$, $A\tilde{\theta}(q)$ reduces to the first r components of $\tilde{\theta}(q)$. It is now easy to see that $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically uncorrelated for each q satisfying $r < q \leq P$ if (and only if) the first r regressors and the last $P - r$ regressors are asymptotically mutually orthogonal in the sense that $\lim_{n\to\infty} X[r]'X[\neg r]/n = 0$. In this case and if $r > \mathcal{O}$ (or in case $r = P$), Theorem 4.3 applies with $p = r$. Otherwise, i.e., in case $\mathcal{O} \leq r < P$ and $\lim_{n\to\infty} X[r]'X[\neg r]/n \neq 0$, Theorem 4.5 applies with $p = r$. Also, the assumptions of Proposition 4.4 are always met by setting $p = r$, since $rank(A[r]) = r$. As a consequence, the non-uniformity phenomenon is always present when estimating this conditional cdf. Furthermore, Theorem 4.8 also always applies for some $q^* \geq r$.

Example 4: *(The conditional distribution of a linear predictor)* Suppose $A \neq 0$ is a $1 \times P$ vector and one is interested in estimating the conditional cdf $G_{n,\theta,\sigma}(t|\hat{p})$ of the linear predictor $A\tilde{\theta}$. Then Theorem 4.8 and the discussion following Proposition 4.9 shows that the non-uniformity phenomenon always arises in this estimation problem in case $\mathcal{O} = 0$. In case $\mathcal{O} > 0$, the non-uniformity problem is generically also present, except in the degenerate case where $C_\infty^{(q)} = 0$, for all q satisfying $\mathcal{O} < q \leq P$ (in which case Proposition 4.7 shows that the least-squares predictors from all models M_p, $\mathcal{O} \leq p \leq P$, perform asymptotically equally well).

5 Estimators for the Conditional Finite-Sample Distribution of Post-Model-Selection Estimators Based on AIC and Related Procedures

In this section we show that the 'impossibility' results obtained in the previous section for the 'general-to-specific' model selection procedure introduced in Section 2 carry over to a large class of model selection procedures, including the widely used Akaike's AIC. Again consider the linear regression model (5) with the same assumptions on the regressors and the errors as in Section 2. Let $\{0,1\}^P$ denote the set of all 0-1 sequences of length P. For each $\mathfrak{r} \in \{0,1\}^P$ let $M_\mathfrak{r}$ denote the set $\{\theta \in \mathbf{R}^P : \theta_i(1-\mathfrak{r}_i) = 0 \ for \ 1 \leq i \leq P\}$ where $\mathfrak{r}_i$ represents the i-th component of $\mathfrak{r}$. I.e., $M_\mathfrak{r}$ describes a linear submodel with those parameters θ_i restricted to zero for which $\mathfrak{r}_i = 0$. Now let $\mathfrak{R}$ be a user-supplied subset of $\{0,1\}^P$. We consider model selection procedures that select from the set $\mathfrak{R}$, or equivalently from the set of models $\{M_\mathfrak{r} : \mathfrak{r} \in \mathfrak{R}\}$. Note that there is now no assumption that the candidate models are nested (for example, if $\mathfrak{R} = \{0,1\}^P$ all possible submodels are potential candidates for selection). Also cases where the inclusion of a subset of regressors is undisputed on a priori grounds is obviously covered by this framework.

We shall assume throughout this section that $\mathfrak{R}$ contains $\mathfrak{r}_{full} = (1,\ldots,1)$ and also at least one element $\mathfrak{r}_*$ satisfying $|\mathfrak{r}_*| = P-1$, where $|\mathfrak{r}_*|$ represents the number of non-zero coordinates of $\mathfrak{r}_*$. Let $\hat{\mathfrak{r}}$ be an arbitrary model selection procedure, i.e., $\hat{\mathfrak{r}}$ is a measurable function of the data Y and X taking its values in $\mathfrak{R}$. We furthermore assume throughout this section that the model selection procedure $\hat{\mathfrak{r}}$ satisfies the following condition: For every $\mathfrak{r}_* \in \mathfrak{R}$ with $|\mathfrak{r}_*| = P-1$ there exists a positive finite constant c such that for every $\theta \in M_{\mathfrak{r}_*}$ which has exactly $P-1$ non-zero coordinates

$$\lim_{n\to\infty} P_{n,\theta,\sigma}\left(\{\hat{\mathfrak{r}} = \mathfrak{r}_{full}\} \blacktriangle \{|T_{\mathfrak{r}_*}| \geq c\}\right) = \lim_{n\to\infty} P_{n,\theta,\sigma}\left(\{\hat{\mathfrak{r}} = \mathfrak{r}_*\} \blacktriangle \{|T_{\mathfrak{r}_*}| < c\}\right) = 0 \tag{40}$$

holds for every $0 < \sigma < \infty$. Here $\blacktriangle$ denotes the symmetric difference operator and $T_{\mathfrak{r}_*}$ represents the usual t-statistic for testing the hypothesis $\theta_{i(\mathfrak{r}_*)} = 0$ in the full model, where $i(\mathfrak{r}_*)$ denotes the index of the unique coordinate of $\mathfrak{r}_*$ that equals zero.

This above condition is quite natural for the following reason: For $\theta \in M_{\mathfrak{r}_*}$ as in condition (40), every reasonable model selection procedure will – with probability approaching unity – decide only between $M_{\mathfrak{r}_*}$ and $M_{\mathfrak{r}_{full}}$; it is then quite natural that this decision will be based (at least asymptotically) on the likelihood ratio between these two models, which in turn boils down to the t-statistic. As will be shown below, condition (40) holds in particular for AIC-like procedures.

Let A be a non-stochastic $k \times P$ matrix of rank k, $1 \leq k \leq P$, as in Section 3. For every $\mathfrak{r} \in \mathfrak{R}$ we then consider the conditional cdf

$$K_{n,\theta,\sigma}(t|\mathfrak{r}) = P_{n,\theta,\sigma}\left(\sqrt{n}A(\bar{\theta}-\theta) \leq t \,\middle|\, \hat{\mathfrak{r}} = \mathfrak{r}\right) \qquad (t \in \mathbf{R}^k) \tag{41}$$

of the post-model-selection estimator $\bar{\theta}$ obtained from the model selection procedure $\hat{\mathfrak{r}}$, i.e.,

$$\bar{\theta} = \sum_{\mathfrak{r}\in\mathfrak{R}} \tilde{\theta}(\mathfrak{r})\,\mathbf{1}(\hat{\mathfrak{r}} = \mathfrak{r})$$

where the $P \times 1$ vector $\tilde{\theta}(\mathfrak{r})$ represents the restricted least-squares estimator obtained from model $M_{\mathfrak{r}}$, with the convention that $\tilde{\theta}(\mathfrak{r}) = 0 \in \mathbf{R}^P$ in case $\mathfrak{r} = (0, \ldots, 0)$. (In case $P_{n,\theta,\sigma}(\hat{\mathfrak{r}} = \mathfrak{r}) = 0$, we define $K_{n,\theta,\sigma}(t|p)$ equal to, say, the cdf of point-mass at zero in $\mathbf{R}^k$. This is done just for the sake of definiteness and has no effect on the results given below. For most model selection procedures the probability $P_{n,\theta,\sigma}(\hat{\mathfrak{r}} = \mathfrak{r})$ will be positive for any $\mathfrak{r} \in \mathfrak{R}$ anyway.) We also introduce

$$K_{n,\theta,\sigma}(t|\hat{\mathfrak{r}}) = \sum_{\mathfrak{r} \in \mathfrak{R}} K_{n,\theta,\sigma}(t|\mathfrak{r}) \, \mathbf{1}(\hat{\mathfrak{r}} = \mathfrak{r}) \qquad (t \in \mathbf{R}^k). \tag{42}$$

We then obtain the following 'impossibility' results which are analogous to the corresponding results in Section 4.2. Similar as in that section, an estimator for $K_{n,\theta,\sigma}(t|\mathfrak{r})$ at a *fixed* value of the argument t is nothing else than a real-valued random variable $\Gamma_n = \Gamma_n(Y, X)$, but for convenience we shall again write $\hat{K}_n(t|\mathfrak{r})$ instead of Γ_n to denote an arbitrary estimator for $K_{n,\theta,\sigma}(t|\mathfrak{r})$. Recall that we assume the model selection procedure $\hat{\mathfrak{r}}$ to satisfy condition (40) in the theorems to follow.

Theorem 5.1 *Let $\mathfrak{r}_* \in \mathfrak{R}$ satisfy $|\mathfrak{r}_*| = P - 1$, and let $i(\mathfrak{r}_*)$ denote the index of the unique coordinate of $\mathfrak{r}_*$ that equals zero. Suppose that $A\tilde{\theta}(\mathfrak{r}_{full})$ and $\tilde{\theta}_{i(\mathfrak{r}_*)}(\mathfrak{r}_{full})$ are asymptotically correlated, i.e., $AQ^{-1}e_{i(\mathfrak{r}_*)} \neq 0$, where $e_{i(\mathfrak{r}_*)}$ denotes the $i(\mathfrak{r}_*)$-th standard basis vector in $\mathbf{R}^P$. Then for every $\theta \in M_{\mathfrak{r}_*}$ which has exactly $P - 1$ non-zero coordinates, for each $0 < \sigma < \infty$, and for each $t \in \mathbf{R}^k$ the following holds with $\mathfrak{r} = \mathfrak{r}_*$ as well as with $\mathfrak{r} = \mathfrak{r}_{full}$:*

a. There exist $\delta_0 > 0$ and $\rho_0 > 0$ such that any estimator $\hat{K}_n(t|\mathfrak{r})$ for $K_{n,\theta,\sigma}(t|\mathfrak{r})$ satisfying

$$P_{n,\theta,\sigma}\left(\left|\hat{K}_n(t|\mathfrak{r}) - K_{n,\theta,\sigma}(t|\mathfrak{r})\right| > \delta\right) \stackrel{n\to\infty}{\longrightarrow} 0 \tag{43}$$

for each $\delta > 0$ (in particular, every estimator that is consistent over $M_{\mathfrak{r}}$) also satisfies

$$\sup_{\substack{\vartheta \in \mathbf{R}^P \\ ||\vartheta - \theta|| < \rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{K}_n(t|\mathfrak{r}) - K_{n,\vartheta,\sigma}(t|\mathfrak{r})\right| > \delta_0\right) \stackrel{n\to\infty}{\longrightarrow} 1. \tag{44}$$

The constants δ_0 and ρ_0 may be chosen in such a way that they depend only on t, Q, A, σ, and c. Moreover,

$$\liminf_{n\to\infty} \inf_{\hat{K}_n(t|\mathfrak{r})} \sup_{\substack{\vartheta \in \mathbf{R}^P \\ ||\vartheta - \theta|| < \rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{K}_n(t|\mathfrak{r}) - K_{n,\vartheta,\sigma}(t|\mathfrak{r})\right| > \delta_0\right) > 0 \tag{45}$$

and

$$\sup_{\delta > 0} \liminf_{n\to\infty} \inf_{\hat{K}_n(t|\mathfrak{r})} \sup_{\substack{\vartheta \in \mathbf{R}^P \\ ||\vartheta - \theta|| < \rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{K}_n(t|\mathfrak{r}) - K_{n,\vartheta,\sigma}(t|\mathfrak{r})\right| > \delta\right) \geq \frac{1}{2} \tag{46}$$

hold, where the infima in (45) and (46) extend over all *estimators $\hat{K}_n(t|\mathfrak{r})$ of $K_{n,\theta,\sigma}(t|\mathfrak{r})$.*

b. The above continues to hold with $P_{n,\cdot,\sigma}(\cdot|\hat{\mathfrak{r}} = \mathfrak{r})$ replacing $P_{n,\cdot,\sigma}(\cdot)$.

We note that the probability of the conditioning event is eventually uniformly bounded away from zero, and hence the conditional probability is well-defined; cf. (71)-(72) in Appendix E.

We next consider performance limits for estimators of $K_{n,\theta,\sigma}(t|\hat{\mathfrak{r}})$ at a *fixed* value of the argument t. Again, an estimator for $K_{n,\theta,\sigma}(t|\hat{\mathfrak{r}})$ is nothing else than a real-valued random variable $\Gamma_n = \Gamma_n(Y, X)$, but for convenience we shall use $\hat{K}_n(t|\hat{\mathfrak{r}})$ instead of Γ_n to denote an arbitrary estimator for $K_{n,\theta,\sigma}(t|\hat{\mathfrak{r}})$.

Theorem 5.2 *Let $\mathfrak{r}_* \in \mathfrak{R}$ satisfy $|\mathfrak{r}_*| = P-1$, and let $i(\mathfrak{r}_*)$ denote the index of the unique coordinate of $\mathfrak{r}_*$ that equals zero. Suppose that $A\tilde{\theta}(\mathfrak{r}_{full})$ and $\tilde{\theta}_{i(\mathfrak{r}_*)}(\mathfrak{r}_{full})$ are asymptotically correlated, i.e., $AQ^{-1}e_{i(\mathfrak{r}_*)} \neq 0$, where $e_{i(\mathfrak{r}_*)}$ denotes the $i(\mathfrak{r}_*)$-th standard basis vector in $\mathbf{R}^P$. Then for every $\theta \in M_{\mathfrak{r}_*}$ which has exactly $P-1$ non-zero coordinates, for each $0 < \sigma < \infty$, and for each $t \in \mathbf{R}^k$ the following holds: There exist $\delta_0 > 0$ and $\rho_0 > 0$ such that any estimator $\hat{K}_n(t|\hat{\mathfrak{r}})$ for $K_{n,\theta,\sigma}(t|\hat{\mathfrak{r}})$ satisfying*

$$P_{n,\theta,\sigma}\left(\left|\hat{K}_n(t|\hat{\mathfrak{r}}) - K_{n,\theta,\sigma}(t|\hat{\mathfrak{r}})\right| > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{47}$$

for each $\delta > 0$ (in particular, every estimator that is consistent) also satisfies

$$\liminf_{n\to\infty} \sup_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{K}_n(t|\hat{\mathfrak{r}}) - K_{n,\vartheta,\sigma}(t|\hat{\mathfrak{r}})\right| > \delta_0\right)$$
$$\geq \liminf_{n\to\infty} P_{n,\theta,\sigma}\left(\hat{\mathfrak{r}} = \mathfrak{r}_{full}\right) = 1 - \Delta_{\sigma\xi_{\infty,i(\mathfrak{r}_*)}}(0, c\sigma\xi_{\infty,i(\mathfrak{r}_*)}) > 0. \tag{48}$$

The constants δ_0 and ρ_0 may be chosen in such a way that they depend only on t, Q, A, σ, and c. Moreover,

$$\liminf_{n\to\infty} \inf_{\hat{K}_n(t|\hat{\mathfrak{r}})} \sup_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{K}_n(t|\hat{\mathfrak{r}}) - K_{n,\vartheta,\sigma}(t|\hat{\mathfrak{r}})\right| > \delta_0\right) > 0 \tag{49}$$

and

$$\sup_{\delta>0} \liminf_{n\to\infty} \inf_{\hat{K}_n(t|\hat{\mathfrak{r}})} \sup_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{K}_n(t|\hat{\mathfrak{r}}) - K_{n,\vartheta,\sigma}(t|\hat{\mathfrak{r}})\right| > \delta\right)$$
$$\geq \frac{1}{2}\liminf_{n\to\infty} P_{n,\theta,\sigma}\left(\hat{\mathfrak{r}} = \mathfrak{r}_{full}\right) = \frac{1}{2}(1 - \Delta_{\sigma\xi_{\infty,i(\mathfrak{r}_*)}}(0, c\sigma\xi_{\infty,i(\mathfrak{r}_*)})) > 0 \tag{50}$$

hold, where the infima in (49) and (50) extend over all *estimators $\hat{K}_n(t|\hat{\mathfrak{r}})$ of $K_{n,\theta,\sigma}(t|\hat{\mathfrak{r}})$.*

The basic condition (40) on the model selection procedure employed in the above results will certainly hold for any hypothesis testing procedure that (i) asymptotically selects only correct models, (ii) employs a likelihood ratio test (or an asymptotically equivalent test) for testing $M_{\mathfrak{r}_{full}}$ versus smaller models (at least versus the models $M_{\mathfrak{r}_*}$ with $\mathfrak{r}_*$ as in condition (40)), and (iii) uses a critical value for the likelihood ratio test that converges to a finite positive constant. We next verify condition (40) for AIC-like procedures. Let

$$IC(\mathfrak{r}) = \log\left(RSS(\mathfrak{r})\right) + |\mathfrak{r}|\,\Upsilon_n/n \tag{51}$$

where $\Upsilon_n \geq 0$ denotes a sequence of real numbers satisfying $\lim_{n\to\infty}\Upsilon_n = \Upsilon$, where Υ is a positive real number. Of course, $IC(\mathfrak{r}) = AIC(\mathfrak{r})$ if $\Upsilon_n = 2$. The model selection procedure $\hat{\mathfrak{r}}_{IC}$ is then defined as a minimizer (more precisely, as a measurable selection from the set of minimizers) of $IC(\mathfrak{r})$ over $\mathfrak{R}$. It is well-known that the probability that $\hat{\mathfrak{r}}_{IC}$ selects an incorrect model converges to zero. Hence, elementary calculations show that condition (40) is satisfied for $c = \Upsilon^{1/2}$.

The analysis of post-model-selection estimators based on AIC-like model selection procedures given in this section proceeded by bringing this case under the umbrella of the results obtained in Section 4.2. Verification of condition (40) is the key that enables this approach. A complete analysis of post-model-selection estimators based on AIC-like model selection procedures, similar to the analysis in Section 4 for the 'general-to-specific' model selection procedure, is certainly possible by a direct and detailed analysis of the distribution of the post-model-selection estimator. (In fact, even the mild condition that $\mathfrak{R}$ contains $\mathfrak{r}_{full}$ and also at least one element $\mathfrak{r}_*$ satisfying $|\mathfrak{r}_*| = P-1$ can be relaxed in such an analysis.) We furthermore note that in the special case where $\mathfrak{R} = \{\mathfrak{r}_{full}, \mathfrak{r}_*\}$ and an AIC-type model selection procedure (51) is used, the results in the above two theorems in fact hold for all $\theta \in M_{\mathfrak{r}_*}$.

6 Remarks and Extensions

Remark 6.1 Although not emphasized in the notation, all results in the paper also hold if the elements of the design matrix X depend on sample size. Furthermore, note that all results are expressed solely in terms of the distributions $P_{n,\theta,\sigma}(\cdot)$ of Y, and hence they also apply if the elements of Y depend on sample size, including the case where the random vectors Y are defined on different probability spaces for different sample sizes.

Remark 6.2 The model selection procedure considered in Sections 2-4 is based on a sequence of tests which use critical values c_p that do not depend on sample size and satisfy $0 < c_p < \infty$ for $\mathcal{O} < p \leq P$. If these critical values are allowed to depend on sample size such that they now satisfy $c_{n,p} \to c_{\infty,p}$ as $n \to \infty$ with $0 < c_{\infty,p} < \infty$ for $\mathcal{O} < p \leq P$, the results in Leeb and Pötscher (2003a) as well as in Leeb (2003a,b) continue to hold; see Remark 6.2(i) in Leeb and Pötscher (2003a) and Remark 6.1(ii) in Leeb (2003a). As a consequence, the results in Sections 2-4 of the present paper can also be extended to this case quite easily.

Remark 6.3 The 'impossibility' results given in Theorems 4.3, 4.5, 4.8, and Proposition 4.4 (as well as the variants thereof discussed in the subsequent Remarks 6.4 and 6.5) also hold for the class of all randomized estimators (with $P^*_{n,\theta,\sigma}$ replacing $P_{n,\theta,\sigma}$ in those results, where $P^*_{n,\theta,\sigma}$ denotes the distribution of the randomized sample). This follows immediately from Lemma 3.6 and the attending discussion in Leeb and Pötscher (2002).

Remark 6.4 a. Let $F_{n,\theta,\sigma}(t|p) = P_{n,\theta,\sigma}(\sqrt{n}A(\tilde{\theta}(p) - \eta_n(p)) \leq t|\hat{p} = p)$ denote the conditional cdf of $A\tilde{\theta}$ given the event $\hat{p} = p$, now centered at the unconditional mean of $A\tilde{\theta}(p)$. This cdf is also studied in Leeb (2003a). Note that the cdfs $F_{n,\theta,\sigma}(t|p)$ and $G_{n,\theta,\sigma}(t|p)$ coincide for $\theta \in M_p$ since then $\eta_n(p) = \theta$ holds. Therefore, Proposition 3.1 continues to hold with $F_{n,\theta,\sigma}(t|p)$ replacing $G_{n,\theta,\sigma}(t|p)$, provided that $\gamma \in M_p$. Also, Proposition 4.1, Theorem 4.3, as well as Proposition 4.4 continue to hold with $F_{n,\theta,\sigma}(t|p)$ replacing $G_{n,\theta,\sigma}(t|p)$. Moreover, Proposition 4.6 holds with $F_{n,\theta,\sigma}(t|p)$ replacing $G_{n,\theta,\sigma}(t|p)$ under the weaker assumption that either $p = \mathcal{O}$ or that $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are asymptotically uncorrelated. (This is easy to see upon inspection of the proof of this proposition and upon using the uniform asymptotic approximation developed in Leeb (2003a).) Furthermore, analogous results can also be obtained for estimators of $F_{n,\theta,\sigma}(t|\hat{p})$.

b. In Leeb (2003a) also the cdfs $F^*_{n,\theta,\sigma}(t|p)$ and $G^*_{n,\theta,\sigma}(t|p)$ are analyzed, which correspond to a (typically infeasible) model selection procedure that makes use of knowledge of σ. Results completely analogous to the ones in the present paper can also be obtained for estimators of these cdfs.

c. Let $\psi_{n,\theta,\sigma}(p)$ denote the conditional mean of $\tilde{\theta}$ given $\hat{p} = p$, and consider the conditional cdf $H_{n,\theta,\sigma}(t|p) = P_{n,\theta,\sigma}(\sqrt{n}A(\tilde{\theta} - \psi_{n,\theta,\sigma}(p)) \leq t \,|\, \hat{p} = p)$. For this cdf a remark similar to part (a) above applies.

Remark 6.5 Results similar to the ones in Section 4.2 can also be obtained for estimation of the asymptotic cdf $G_{\infty,\theta,\sigma}(t|p)$ (or of the asymptotic cdfs corresponding to the variants discussed in the previous remark). Since these results are of limited interest, we omit them. In particular note that an 'impossibility' result for estimation of $G_{\infty,\theta,\sigma}(t|p)$ does *not* imply a corresponding 'impossibility' result for estimation of $G_{n,\theta,\sigma}(t|p)$, since $G_{n,\theta,\sigma}(t|p)$ does in general not converge uniformly to $G_{\infty,\theta,\sigma}(t|p)$ over the relevant subsets in the parameter space; cf. Remark 6.6 below.

Remark 6.6 The convergence of $G_{n,\theta,\sigma}(t|p)$ to $G_{\infty,\theta,\sigma}(t|p)$ is, in general, not uniform with respect to the parameter θ; cf. Leeb and Pötscher (2003a), Corollary 4.6. (A similar remark applies to the variants of the cdfs discussed in Remark 6.4.) More precisely, the following three cases can be shown to occur:

a. Assume $\mathcal{O} < p \leq P$ holds and $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are asymptotically correlated. Suppose $t \in \mathbf{R}^k$ is such that the set $\{z \in \mathbf{R}^p : A[p]z \leq t\}$ has positive Lebesgue measure. Then

$$\liminf_{n\to\infty} \sup_{\substack{\vartheta \in \mathbf{M}_p \\ ||\vartheta-\theta||<\rho/\sqrt{n}}} |G_{n,\vartheta,\sigma}(t|p) - G_{\infty,\vartheta,\sigma}(t|p)| > 0$$

for each $\theta \in M_{p-1}$ and for some constant $\rho > 0$.

b. Suppose the assumptions above are met with the exception that the set $\{z \in \mathbf{R}^p : A[p]z \leq t\}$ has Lebesgue measure zero. Then the cdf $G_{n,\theta,\sigma}(t|p)$ equals zero for each $\theta \in M_p$ (independently of n).

c. Assume $p = \mathcal{O}$ or that $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are asymptotically uncorrelated. Then it can be shown that $G_{n,\theta,\sigma}(t|p)$ converges to $G_{\infty,\theta,\sigma}(t|p)$ uniformly over M_p, even with respect to the total variation distance.

Remark 6.7 Remark 6.3 also applies to the 'impossibility' results in Section 5. Furthermore, Remarks 6.4-6.6 have appropriate analogues for the model selection procedures discussed in Section 5.

Remark 6.8 The model selection procedures considered in this paper are conservative procedures in the sense that even asymptotically the probability of selecting a model with too many parameters is positive. For consistent model selection procedures – like BIC or testing procedures with suitably diverging critical values c_p (cf. Bauer, Pötscher, and Hackl (1988)) – the (pointwise) asymptotic distribution is normal; this follows immediately from Lemma 1 in Pötscher (1991). However, this result is first of all highly misleading since the finite sample distribution can be far from a normal, the convergence of the finite-sample distribution to the asymptotic normal distribution not being uniform. Second, 'impossibility' results similar to the ones presented in Sections 4.2 and 5 of the present paper also hold when post-model-selection estimators based on consistent model selection procedures are considered. These will be discussed in detail elsewhere. For a simple special case such an 'impossibility' result is given in Section 2.3 of Leeb and Pötscher (2002).

7 Conclusion

While we have shown that consistent estimators for the conditional distribution of a post-model-selection estimator exist, we have also shown that no estimator for this conditional distribution can have satisfactory performance (locally) uniformly in the parameter space, even asymptotically. In particular, no (locally) uniformly consistent estimator for the conditional distribution exists. Hence, the answer to the question posed in the title has to be negative. The results in the present paper also cover the case of linear functions (e.g., predictors) of the post-model-selection estimator. Corresponding results for the unconditional distribution of the post-model-selection estimator are presented in a companion paper, Leeb and Pötscher (2003b).

The 'impossibility' results are derived in the framework of a normal linear regression model (and a fortiori these results continue to hold in any model which includes the normal linear regression as a special case). Furthermore, there is no reason to believe that the situation will get any better in more complex statistical models that allow, e.g., for nonlinearity or dependent data.

The results in the present paper are derived for a large class of conservative model selection procedures (i.e., procedures that select overparameterized model with positive probability asymptotically) including Akaike's AIC and typical 'general-to-specific' hypothesis testing procedures. However, similar 'impossibility' results also hold for consistent model selection procedures; cf. Remark 6.8.

The 'impossibility' of estimating the distribution of the post-model-selection estimator does not per se preclude the possibility of conducting valid inference after model selection, a topic that deserves further study. However, it certainly makes this a more challenging task.

A Proofs for Section 3

Proof of Proposition 3.2: Let $\vartheta^{(n)}$ be an arbitrary sequence of parameters in $\mathbf{R}^P$. Proposition 5.4 in Leeb (2003b) together with Remark 5.5 in that reference show that any accumulation point of $P_{n,\vartheta^{(n)},\sigma}(\hat{p}=p)$ is of the form

$$(1-\Delta_{\sigma\xi_{\infty,p}}(\upsilon_p, c_p\sigma\xi_{\infty,p}))\prod_{q=p+1}^{P}\Delta_{\sigma\xi_{\infty,q}}(\upsilon_q, c_q\sigma\xi_{\infty,q}) \tag{52}$$

in case $p > \mathcal{O}$, and of the form

$$\prod_{q=p+1}^{P}\Delta_{\sigma\xi_{\infty,q}}(\upsilon_q, c_q\sigma\xi_{\infty,q}) \tag{53}$$

in case $p = \mathcal{O}$. The quantities υ_q, $q = p, \dots, P$, in these displays are accumulation points of $\upsilon_q^{(n)} = \sqrt{n}\vartheta_q^{(n)} + \sqrt{n}((X[q]'X[q])^{-1}X[q]'X[\neg q]\vartheta^{(n)}[\neg q])_q$ in $\mathbf{R}\cup\{-\infty,\infty\}$. (In case $q = P$ this expression is to be interpreted as $\sqrt{n}\vartheta_P^{(n)}$ by our conventions.) Observe that the expression in (52) is positive if and only if $|\upsilon_q| < \infty$ holds for each $q = p+1, \dots, P$. The same is true for (53). (In case $p = P$, the expression in (52) is always positive).

To prove part (a) it suffices to show that any accumulation point of $P_{n,\vartheta^{(n)},\sigma}(\hat{p}=p)$ is positive whenever $\vartheta^{(n)}$ is a sequence satisfying $||\vartheta^{(n)}[\neg p]|| < r_n$. In case $p = P$ it is easy to see that (52) reduces to $1 - \Delta_{\sigma\xi_{\infty,P}}(\upsilon_P, c_P\sigma\xi_{\infty,P})$ which is bounded from below by the positive constant $1 - \Delta_{\sigma\xi_{\infty,P}}(0, c_P\sigma\xi_{\infty,P})$. In case $p < P$ note that $\sqrt{n}\vartheta^{(n)}[\neg p]$ is a bounded sequence and hence $\upsilon_q^{(n)}$ is bounded for each $q = p+1, \dots, P$. It follows that $|\upsilon_q| < \infty$ holds for $q = p+1, \dots, P$. This completes the proof of part (a).

To prove part (b) let $\vartheta^{(n)}$ be given by $\vartheta^{(n)}[P-1] = \theta[P-1]$ and $\vartheta_P^{(n)} = r_n/2$. Clearly then $||\vartheta^{(n)} - \theta|| < r_n$ is satisfied. Moreover, $\sqrt{n}\vartheta_P^{(n)} = \sqrt{n}r_n/2$ converges to $\upsilon_P = \infty$. It follows that $\lim_{n\to\infty} P_{n,\vartheta^{(n)},\sigma}(\hat{p}=p) = 0$ whenever $p < P$.

Part (c) is proved analogously. □

B Proofs for Section 4.1

Recall that $\Phi_{n,p}(t)$ denotes the cdf of a k-variate Gaussian random vector with mean zero and variance-covariance matrix $\sigma^2 A[p](X[p]'X[p]/n)^{-1}A[p]'$. In the proofs below it will be convenient to show the dependence of $\Phi_{n,p}(t)$ on σ in the notation. Thus, in the following we shall write $\Phi_{n,p,\sigma}(t)$ and $\Phi_{\infty,p,\sigma}(t)$, respectively, for the cdf of a k-variate Gaussian random vector with mean zero and variance-covariance matrix $\sigma^2 A[p](X[p]'X[p]/n)^{-1}A[p]'$ and $\sigma^2 A[p]Q[p:p]^{-1}A[p]'$, respectively. For convenience, let $\Phi_{n,0,\sigma}(t)$ and

$\Phi_{\infty,0,\sigma}(t)$ denote the cdf of point-mass at zero in $\mathbf{R}^k$. The following lemma is elementary to prove, if we observe that in case $rank(A[p]) = k$ the convergence $b_{n,p} \to b_{\infty,p}$ holds, since the generalized inverses in the definitions of these quantities reduce to the usual inverse.

Lemma B.1 *Suppose $p > 0$ and that $rank(A[p]) = k$. Define $S_{n,p}(z,\sigma) = \frac{1-\Delta_{\sigma\zeta_{n,p}}(b_{n,p}z, c_p\sigma\xi_{n,p})}{1-\Delta_{\sigma\xi_{n,p}}(0, c_p\sigma\xi_{n,p})}$ and $S_{\infty,p}(z,\sigma) = \frac{1-\Delta_{\sigma\zeta_{\infty,p}}(b_{\infty,p}z, c_p\sigma\xi_{\infty,p})}{1-\Delta_{\sigma\xi_{\infty,p}}(0, c_p\sigma\xi_{\infty,p})}$ for $z \in \mathbf{R}^k$, $0 < \sigma < \infty$. Let $\sigma^{(n)}$ converge to σ, $0 < \sigma < \infty$. Then $S_{n,p}(z,\sigma^{(n)})$ converges to $S_{\infty,p}(z,\sigma)$ for every $z \in \mathbf{R}^k$ if $\zeta_{\infty,p} \neq 0$, and for every $z \in \mathbf{R}^k$ except possibly for z satisfying $|z| = c_p\sigma\xi_{\infty,p}/|b_{\infty,p}|$ if $\zeta_{\infty,p} = 0$. (Note that $b_{\infty,p} \neq 0$ in case $\zeta_{\infty,p} = 0$.)*

Lemma B.2 *Let $(\Omega, \mathcal{A})$ and $(\Xi, \mathcal{B})$ be measurable spaces and let $\Psi : \Omega \to \Xi$ be a measurable function. Suppose μ_n and μ are probability measures on $(\Omega, \mathcal{A})$ satisfying $\|\mu_n - \mu\|_{TV} \to 0$. Let ρ_n be the probability measure induced by μ_n and Ψ, i.e., $\rho_n(B) = \mu_n(\Psi^{-1}(B))$ for $B \in \mathcal{B}$. Then ρ_n converges to a probability measure ρ with respect to the total variation distance and ρ is the measure induced by μ and Ψ.*

Lemma B.2 follows immediately from $\|\rho_n - \rho\|_{TV} \leq \|\mu_n - \mu\|_{TV}$. The following observation is useful in the proof of Proposition 4.1: Since the proposition depends on Y only through its distribution, we may assume without loss of generality that the errors in (5) are given by $u_t = \sigma\varepsilon_t$, $t \in \mathbf{N}$, with i.i.d. ε_t. In particular, all random variables involved are then defined on the same probability space; cf. Remark 6.1.

Proof of Proposition 4.1: We consider first the case $p > \mathcal{O}$ and assume for the moment that the matrix $A[p]$ has full row rank k. Then $\Phi_{n,p,\sigma}(\cdot)$ and $\Phi_{\infty,p,\sigma}(\cdot)$ possess densities $\phi_{n,p,\sigma}(\cdot)$ and $\phi_{\infty,p,\sigma}(\cdot)$, respectively, with respect to Lebesgue measure on $\mathbf{R}^k$. Since $\hat{\sigma} \to \sigma$ in $P_{n,\theta,\sigma}$-probability, each subsequence contains a further subsequence along which $\hat{\sigma} \to \sigma$ almost surely (with respect to the probability measure on the common probability space supporting all random variables involved), and we restrict ourselves to this further subsequence for the moment. In particular, we write $\{\hat{\sigma} \to \sigma\}$ for the event that $\hat{\sigma}$ converges to σ along the subsequence under consideration; clearly, the event $\{\hat{\sigma} \to \sigma\}$ has probability one. Lemma B.1 now shows that on the event $\{\hat{\sigma} \to \sigma\}$ the function $S_{n,p}(z,\hat{\sigma})\phi_{n,p,\hat{\sigma}}(z)$ converges to $S_{\infty,p}(z,\sigma)\phi_{\infty,p,\sigma}(z)$ for every z except for a set of Lebesgue measure zero. Observe that both functions are probability densities with respect to Lebesgue measure on $\mathbf{R}^k$, cf. the discussion prior to Proposition 4.1. In view of Scheffè's Lemma they hence converge in absolute mean. By the same argument, also $\phi_{n,p,\hat{\sigma}}(\cdot)$ converges to $\phi_{\infty,p,\sigma}(\cdot)$ in absolute mean. Note that the absolute mean convergence of the densities translates into convergence in total variation for the corresponding cdfs. Now $\check{G}_n(t|p) = \hat{\Phi}_{n,p,\hat{\sigma}}(t)$ in case the auxiliary procedure decides for $p_0(\theta) = p$, and $\check{G}_n(t|p) = \int_{z \in \mathbf{R}^k, z \leq t} S_{n,p}(z,\hat{\sigma})\phi_{n,p,\hat{\sigma}}(z)dz$ otherwise. Since the auxiliary procedure decides consistently between $p_0(\theta) = p$ and $p_0(\theta) < p$ for every $\theta \in M_p$, it follows that (24) holds along the subsequence under consideration in case $p > \mathcal{O}$ and if $A[p]$ has rank k. This proves (24) in case $p > \mathcal{O}$ and $A[p]$ has rank k.

In case $p > \mathcal{O}$ but where the matrix $A[p]$ does not have full row rank k, let $G^I_{n,\theta,\sigma}(t|p)$, $\check{G}^I_n(t|p)$, and $G^I_{\infty,\theta,\sigma}(t|p)$, respectively, be defined in exactly the same way as $G_{n,\theta,\sigma}(t|p)$, $\check{G}_n(t|p)$, and $G_{\infty,\theta,\sigma}(t|p)$, except that the $p \times P$ matrix $(I_p : 0)$ replaces A. Note that then I_p replaces $A[p]$ (and that the value of k changes to p). Since the matrix I_p has full row rank p, the preceding paragraph shows that (24) holds with $\check{G}^I_n(t|p)$ and $G^I_{\infty,\theta,\sigma}(t|p)$ replacing $\check{G}_n(t|p)$ and $G_{\infty,\theta,\sigma}(t|p)$, respectively. But $\check{G}_n(t|p)$ and $G_{\infty,\theta,\sigma}(t|p)$, respectively, are the cdfs of the image measures of $\check{G}^I_n(t|p)$ and $G^I_{\infty,\theta,\sigma}(t|p)$ induced by the linear mapping $x \mapsto A[p]x$, $x \in \mathbf{R}^p$. (This is obvious for $\check{G}_n(t|p)$ because of its interpretation as the conditional cdf $G^*_{n,\theta,\sigma}(t|p)$ in (13) of Leeb (2003a) with $\sigma = \hat{\sigma}$. Observe further that $G_{n,\theta,\sigma}(t|p)$ is clearly the cdf of the induced measure obtained from the cdf $G^I_{n,\theta,\sigma}(t|p)$. Since $G^I_{n,\theta,\sigma}(t|p) \to G^I_{\infty,\theta,\sigma}(t|p)$ and $G_{n,\theta,\sigma}(t|p) \to G_{\infty,\theta,\sigma}(t|p)$ with respect to total variation distance for $\theta \in M_p$ by Proposition 3.1, an application of Lemma B.2 shows that $G_{\infty,\theta,\sigma}(t|p)$ is indeed the cdf of the induced measure obtained from the cdf $G^I_{\infty,\theta,\sigma}(t|p)$.) Therefore, the total variation distance of $\check{G}_n(t|p)$ and $G_{\infty,\theta,\sigma}(t|p)$ is bounded from above by that of $\check{G}^I_n(t|p)$ and $G^I_{\infty,\theta,\sigma}(t|p)$. This proves (24) also in this case.

The case where $p = \mathcal{O} > 0$ follows in a similar way since (again after passing to appropriate subsequences) $\phi_{n,p,\hat{\sigma}}(\cdot)$ converges to $\phi_{\infty,p,\sigma}(\cdot)$ in absolute mean on the event $\{\hat{\sigma} \to \sigma\}$ as defined above. The case where $p = \mathcal{O} = 0$ is trivial, because both cdfs in (24) coincide and are equal to the cdf of point-mass at zero in $\mathbf{R}^k$. This completes the proof of (24).

The validity of (23) now follows for $\theta \in M_p$ since $G_{\infty,\theta,\sigma}(t|p)$ is then the limit of $G_{n,\theta,\sigma}(t|p)$ with respect to the total variation distance; see Proposition 3.1. Finally, the claim in parenthesis follows from (23) and (24) in view of Proposition 3.2(a). □

Proof of Corollary 4.2: Observe that

$$\begin{aligned} & P_{n,\theta,\sigma}\left(\left|\left|\check{G}_n(\cdot|\hat{p}) - G_{n,\theta,\sigma}(\cdot|\hat{p})\right|\right|_{TV} > \delta\right) \\ = & \sum_{p=\mathcal{O}}^{P} P_{n,\theta,\sigma}\left(\left|\left|\check{G}_n(\cdot|p) - G_{n,\theta,\sigma}(\cdot|p)\right|\right|_{TV} > \delta,\, \hat{p} = p\right) \\ \leq & \sum_{\mathcal{O} \leq p < p_0(\theta)} P_{n,\theta,\sigma}\left(\hat{p} = p\right) + \sum_{p \geq p_0(\theta)} P_{n,\theta,\sigma}\left(\left|\left|\check{G}_n(\cdot|p) - G_{n,\theta,\sigma}(\cdot|p)\right|\right|_{TV} > \delta\right). \end{aligned}$$

Each term in the first sum on the far right-hand side of the above display now obviously converges to zero (cf. Leeb (2003b), Corollary 5.6), whereas every term in the second sum converges to zero by Proposition 4.1. □

C Proofs for Section 4.2

We start with some preparatory remarks. The total variation distance between $P_{n,\theta,\sigma}$ and $P_{n,\vartheta,\sigma}$ satisfies $||P_{n,\theta,\sigma} - P_{n,\vartheta,\sigma}||_{TV} \leq 2\Phi(\|\theta - \vartheta\|\, \lambda^{1/2}_{\max}(X'X)/2\sigma) - 1$. Furthermore, if $\theta^{(n)}$ and $\vartheta^{(n)}$ satisfy $\left\|\theta^{(n)} - \vartheta^{(n)}\right\| = O(n^{-1/2})$, the sequence $P_{n,\vartheta^{(n)},\sigma}$ is contiguous with respect to the sequence $P_{n,\theta^{(n)},\sigma}$. This follows exactly in the same way as Lemma A.1 in Leeb and Pötscher (2002). We also need the following lemma.

Lemma C.1 *Let p satisfy $\mathcal{O} < p \leq P$. Suppose $\theta \in M_{p-1}$, $0 < \sigma < \infty$, and $\rho > 0$. Then*

$$\liminf_{n\to\infty} \inf_{\substack{\vartheta \in M_p \\ ||\vartheta - \theta|| < \rho/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\hat{p} = p\right) = (1 - \Delta_{\sigma\xi_{\infty,p}}(0, c_p\sigma\xi_{\infty,p})) \prod_{q=p+1}^{P} \Delta_{\sigma\xi_{\infty,q}}(0, c_q\sigma\xi_{\infty,q}) \tag{54}$$

$$= \lim_{n\to\infty} P_{n,\theta,\sigma}\left(\hat{p} = p\right) > 0. \tag{55}$$

Proof of Lemma C.1: We proceed similarly as in the proof of Proposition 3.2, observing that now the quantities υ_q, $q > p$, are all equal to zero since $\vartheta^{(n)} \in M_p$. Since $(1 - \Delta_{\sigma\xi_{\infty,p}}(\upsilon_p, c_p\sigma\xi_{\infty,p}))$ is minimal for $\upsilon_p = 0$, we see that the right-hand side of (54), which obviously is positive, is a lower bound for the left-hand side. Using (24) in Leeb (2003b) and observing that $\theta \in M_{p-1}$ completes the proof. □

Proof of Theorem 4.3: We first prove (26) and (27). For this purpose we make use of Lemma 3.1 in Leeb and Pötscher (2002) with $\alpha = \theta \in M_{p-1}$, $B = M_p$, $B_n = \{\vartheta \in M_p : \|\vartheta - \theta\| < \rho_0 n^{-1/2}\}$, $\beta = \vartheta$, $\varphi_n(\beta) = G_{n,\vartheta,\sigma}(t|p)$, $\widehat{\varphi}_n = \hat{G}_n(t|p)$, where $\rho_0 > 0$ will be chosen shortly (and σ is held fixed). The contiguity assumption of this lemma is satisfied in view of the preparatory remark above. It hence remains only to show that there exists a value of $\rho_0 > 0$ such that δ^* in Lemma 3.1 of Leeb and Pötscher (2002) (which represents the limit inferior of the oscillation of $\varphi_n(\cdot)$ over B_n) is positive. Applying Lemma 3.5(a) of Leeb and Pötscher (2002) with $\zeta_n = \rho_0 n^{-1/2}$ and the set G_0 equal to the set G, it remains, in light of Proposition 3.1, to show that there exists a $\rho_0 > 0$ such that $G_{\infty,\theta,\sigma,\gamma}(t|p)$ as a function of γ is non-constant on the set $\{\gamma \in M_p : \|\gamma\| < \rho_0\}$. The corresponding δ_0 can then be chosen as any positive number less than one-half of the oscillation of $G_{\infty,\theta,\sigma,\gamma}(t|p)$ over this set. That such a $\rho_0 > 0$ indeed exists follows from Lemma D.1 in Appendix D. Note that this lemma is applicable here since $rank(A[p]) = rank(A) = k$, either trivially in case $p = P$ or as a consequence of Proposition 4.7(a),(c) and of the assumption on the model order p in case $p < P$. Furthermore, observe that $G_{\infty,\theta,\sigma,\cdot}(t|p)$ is given by (19) for $\theta \in M_{p-1}$ and hence does not depend on θ, but only on t, Q, A, σ, and c_p. As a consequence, ρ_0 and δ_0 can be chosen such that they also depend only on these quantities. This completes the proof of (26) and (27).

To prove (28) we use Corollary 3.4 in Leeb and Pötscher (2002) with the same identification of notation as above and with $V = M_p$ (viewed as a vector space isomorphic to $\mathbf{R}^p$). Note that then $\zeta_n = \rho_0 n^{-1/2}$. The asymptotic uniform equicontinuity condition in that corollary is then satisfied in view of $||P_{n,\theta,\sigma} - P_{n,\vartheta,\sigma}||_{TV} \leq 2\Phi(\|\theta - \vartheta\| \lambda_{\max}^{1/2}(X'X)/2\sigma) - 1$. Applying Corollary 3.4 in Leeb and Pötscher (2002) then establishes (28). This completes the proof of part (a).

Part (b) is proved exactly as part (a) making additional use of Corollary C.2 and Remark C.1 in Leeb and Pötscher (2002). The events E_n appearing in this corollary are given here by $\{\hat{p} = p\}$. Clearly, $P_{n,\vartheta,\sigma}(\hat{p} = p)$ is always positive. The constant M in Corollary C.2 of Leeb and Pötscher (2002) is now given by the right-hand side of (54) above. □

Proof of Proposition 4.4: Completely analogous to the proof of Theorem 4.3. □

Proof of Theorem 4.5: We again use results from Leeb and Pötscher (2002) this time with the identification $\alpha = \theta \in M_p$, $B = M_{q^*}$, $B_n = \{\vartheta \in M_{q^*} : \|\vartheta - \theta\| < \rho_0 n^{-1/2}\}$, $\beta = \vartheta$, $\varphi_n(\beta) = G_{n,\vartheta,\sigma}(t|p)$, $\hat{\varphi}_n = \hat{G}_n(t|p)$, $V = M_{q^*}$, and $\zeta_n = \rho_0 n^{-1/2}$ (again σ is held fixed). The proof of part (a) is then similar to the proof of part (a) of Theorem 4.3, except for using Lemma D.2 instead of Lemma D.1 and except for the fact that the argument that ρ_0 and δ_0 only depend on t, Q, A, σ, and c_p is now slightly more complex, since $G_{\infty,\theta,\sigma,\cdot}(t|p)$ for $\theta \in M_{q^*}$ depends on θ. However, observe that $G_{\infty,\theta,\sigma,\cdot}(t|p)$ as a function of $\theta \in M_{q^*}$ follows only two different formulae which themselves do not depend on θ, cf. (18), (19).

Part (b) is proved exactly as the corresponding part of Theorem 4.3, except that positivity of the constant $M = \liminf_{n\to\infty} \inf_{\substack{\vartheta \in M_{q^*} \\ ||\vartheta - \theta|| < \rho/\sqrt{n}}} P_{n,\vartheta,\sigma}(\hat{p} = p)$ follows now since M is bounded from below by the expression in part (a) of Proposition 3.2. □

Lemma C.2 *Let p satisfy $\mathcal{O} \leq p \leq P$. Suppose that $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically uncorrelated, i.e., $C_\infty^{(q)} = 0$, for each $q = \max\{p, \mathcal{O}+1\}, \ldots, P$. Then*

$$\sup_{\substack{\theta \in \mathbf{R}^P \\ ||\theta[\neg p]|| < \rho/\sqrt{n}}} \sup_{\substack{\sigma \in \mathbf{R} \\ \sigma_* \leq \sigma \leq \sigma^*}} ||G_{n,\theta,\sigma}(\cdot|p) - \Phi_{\infty,p,\sigma}(\cdot)||_{TV} \stackrel{n\to\infty}{\longrightarrow} 0, \tag{56}$$

for every $\rho > 0$, and for any constants σ_* and σ^* satisfying $0 < \sigma_* \leq \sigma^* < \infty$, where the cdf $\Phi_{\infty,p,\sigma}(\cdot)$ has been defined in Appendix B. (In case $p = P$, the first supremum in (56) is to be interpreted as extending over all $\theta \in \mathbf{R}^P$. Furthermore, the case $p = 0$ is impossible in view of Proposition 4.7).

Proof of Lemma C.2: Consider first the case $p > \mathcal{O}$. In view of Theorem 4.1 of Leeb (2003a) it suffices to establish (56) with $G^*_{n,\theta,\sigma}(t|p)$ in place of $G_{n,\theta,\sigma}(t|p)$, where $G^*_{n,\theta,\sigma}(t|p)$ denotes the cdf

$$G^*_{n,\theta,\sigma}(t|p) = \int_{z \le t - \sqrt{n}A(\eta_n(p)-\theta)} \frac{1 - \Delta_{\sigma\zeta_{n,p}}(\sqrt{n}\eta_{n,p}(p) + b_{n,p}z, c_p\sigma\xi_{n,p})}{1 - \Delta_{\sigma\xi_{n,p}}(\sqrt{n}\eta_{n,p}(p), c_p\sigma\xi_{n,p})} \Phi_{n,p,\sigma}(dz).$$

Note that $rank(A[p]) = rank(A) = k$ holds, as a consequence of Proposition 4.7(a),(c) and the uncorrelatedness assumption in the lemma. Therefore, the cdfs $\Phi_{n,p,\sigma}(\cdot)$ and $\Phi_{\infty,p,\sigma}(\cdot)$ have densities with respect to Lebesgue measure, denoted by $\phi_{n,p,\sigma}(\cdot)$ and $\phi_{\infty,p,\sigma}(\cdot)$, respectively. The Lebesgue density of $G^*_{n,\theta,\sigma}(t|p)$ is then given by

$$g^*_{n,\theta,\sigma}(z|p) = \frac{1 - \Delta_{\sigma\zeta_{n,p}}(\sqrt{n}\eta_{n,p}(p) + b_{n,p}(z - \sqrt{n}A(\eta_n(p)-\theta)), c_p\sigma\xi_{n,p})}{1 - \Delta_{\sigma\xi_{n,p}}(\sqrt{n}\eta_{n,p}(p), c_p\sigma\xi_{n,p})} \phi_{n,p,\sigma}(z - \sqrt{n}A(\eta_n(p)-\theta)). \quad (57)$$

It suffices now to show that

$$\left|\left|G^*_{n,\theta^{(n)},\sigma^{(n)}}(\cdot|p) - \Phi_{\infty,p,\sigma^{(n)}}(\cdot)\right|\right|_{TV} = 2^{-1}\left|\left|g^*_{n,\theta^{(n)},\sigma^{(n)}}(z|p) - \phi_{\infty,p,\sigma^{(n)}}(z)\right|\right|_1 \overset{n\to\infty}{\longrightarrow} 0$$

for every sequence $(\theta^{(n)}, \sigma^{(n)})$ satisfying $||\theta^{(n)}[\neg p]|| < \rho/\sqrt{n}$ and $\sigma_* \le \sigma^{(n)} \le \sigma^*$. (In case $p = P$, $\theta^{(n)}$ can be completely arbitrary.) By passing to subsequences if necessary, we can assume without loss of generality that $\sqrt{n}\theta^{(n)}_p$ converges to a limit in $\mathbf{R} \cup \{-\infty, \infty\}$, that $\sqrt{n}\theta^{(n)}[\neg p]$ converges to a limit $\alpha \in \mathbf{R}^{P-p}$ satisfying $||\alpha|| \le \rho$ (in case $p < P$), and that $\sigma^{(n)}$ converges to a limit σ satisfying $\sigma_* \le \sigma \le \sigma^*$. It follows that $\sqrt{n}\eta_{n,p}(p)$ (constructed from $\theta^{(n)}$ rather than θ) converges to a constant $\nu \in \mathbf{R} \cup \{-\infty, \infty\}$. Since $C^{(p)}_\infty = 0$, it follows from (10) that $b_{n,p}$ converges to 0. From (20) in Leeb (2003a) we have $\sqrt{n}A(\eta_n(p) - \theta^{(n)}) = -\sum_{q=p+1}^{P} \xi^{-2}_{n,q} C^{(q)}_n \sqrt{n}\eta_{n,q}(q)$. Observe that $\eta_{n,q}(q)$ for $q > p$ depends on $\theta^{(n)}$ only through $\theta^{(n)}[\neg p]$; hence $\sqrt{n}\eta_{n,q}(q)$ is bounded. Since $C^{(q)}_n \to C^{(q)}_\infty = 0$, $\xi_{n,q} \to \xi_{\infty,q} > 0$, we conclude that $\sqrt{n}A(\eta_n(p) - \theta^{(n)}) \to 0$. Furthermore, observe that $\zeta_{n,p} \to \zeta_{\infty,p}$ and that $\zeta_{\infty,p} = \xi_{\infty,p} > 0$ because of (16) and $C^{(p)}_\infty = 0$. It is thus easy to see that (57) converges to $\phi_{\infty,p,\sigma}(z)$ for every $z \in \mathbf{R}^k$. The result now follows from Scheffè's lemma.

In case $p = \mathcal{O} > 0$, the argument is similar, but even simpler, as now the density of $G^*_{n,\theta,\sigma}(t|\mathcal{O})$ is given only by the normal density appearing in (57), with the first factor replaced by the constant one. The case $p = \mathcal{O} = 0$ is not possible given the assumptions of the lemma, as noted in Proposition 4.7. □

Proof of Proposition 4.6: In view of Lemma C.2 and the fact that $\hat{\Phi}_{n,p}(\cdot) = \Phi_{n,p,\hat{\sigma}}(\cdot)$ it suffices to show that

$$\sup_{\substack{\sigma\in\mathbf{R} \\ \sigma_* \le \sigma \le \sigma^*}} P_{n,\theta,\sigma}\left(||\Phi_{n,p,\hat{\sigma}}(\cdot) - \Phi_{n,p,\sigma}(\cdot)||_{TV} > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \quad (58)$$

holds for each $\delta > 0$, and for any constants σ_* and σ^* satisfying $0 < \sigma_* \le \sigma^* < \infty$. (Note that the probability in (58) does in fact not depend on θ.) By the assumptions and Proposition 4.7 we have $p > 0$ and $rank(A[p]) = rank(A) = k$. As a consequence, the cdfs in (58) have densities with respect to Lebesgue measure. Note that Scheffè's lemma implies that the map attaching the k-variate normal density to its mean μ and variance-covariance matrix Σ is continuous for all μ and all positive-definite Σ, when the space of densities is equipped with the L_1-norm. Consequently, the k-variate normal cdf depends continuously on μ and Σ on the same domain, when the space of cdfs is equipped with the total variation distance. Now the mean of $\Phi_{n,p,\sigma}(\cdot)$ is identically zero and its variance covariance matrix is $\sigma^2 A[p](X[p]'X[p]/n)^{-1}A[p]'$. Since the matrix $A[p](X[p]'X[p]/n)^{-1}A[p]'$ converges to the positive definite matrix $A[p](Q[p:p])^{-1}A[p]'$ as $n \to \infty$, and since σ varies in the compact set $[\sigma_*, \sigma^*]$, it follows that the sequence $||\Phi_{n,p,\tau}(\cdot) - \Phi_{n,p,\sigma}(\cdot)||_{TV}$ is a uniformly equicontinuous sequence of functions of τ on $[\sigma_*, \sigma^*]$. Consequently, the probability in (58) is bounded by

$P_{n,\theta,\sigma}\left(|\hat{\sigma}-\sigma|>\eta\right) \leq P_{n,\theta,\sigma}\left(|\hat{\sigma}/\sigma-1|>\eta/\sigma^*\right)$ for some $\eta=\eta(\delta)>0$ that is independent of n. Since $\hat{\sigma}/\sigma$ is distributed as $(n-P)^{1/2}$ times the square root of a chi-square distributed random variable with $n-P$ degrees of freedom, the last bound is seen to be independent of θ and σ and to converge to zero.

The result regarding the conditional probability follows immediately from the unconditional result and Proposition 3.2(c). □

Proof of Proposition 4.7: That part (a) implies part (b) follows from (20) in Leeb (2003a), observing that $C_n^{(q)} \to C_\infty^{(q)}$ and that $\eta_{n,q}(q)$ converges to a finite limit. The reverse implication follows by passing to the limit in (20) of Leeb (2003a) and observing that by suitable choice of $\theta \in \mathbf{R}^P$ the limit of $(\eta_{n,p+1}(p+1),\ldots,\eta_{n,P}(P))'$ can take on the value of every standard basis vector in $\mathbf{R}^{P-p}$. To prove the equivalence of parts (a) and (c), we use Proposition 3.1 in Leeb (2003a) and equation (19) in that paper to obtain $\sum_{r=1}^{q}\sigma^2\xi_{\infty,r}^{-2}C_\infty^{(r)}C_\infty^{(r)\prime}$ as the formula for the asymptotic variance-covariance matrix of $\sqrt{n}A\tilde{\theta}(q)$. Since the terms in this sum are nonnegative definite, the equivalence follows. The final claims regarding the cases $p=P$ and $p=0$ are either obvious or follow immediately from the representation of the asymptotic variance-covariance matrix of $\sqrt{n}A\tilde{\theta}(q)$ just given. □

Proof of Theorem 4.8: In view of the definition of $G_{n,\vartheta,\sigma}(t|\hat{p})$ we have

$$\left|\hat{G}_n(t|\hat{p})-G_{n,\vartheta,\sigma}(t|\hat{p})\right| = \sum_{p=\mathcal{O}}^{P}\left|\hat{G}_n(t|\hat{p})-G_{n,\vartheta,\sigma}(t|p)\right|\mathbf{1}(\hat{p}=p) \geq \left|\hat{G}_n(t|\hat{p})-G_{n,\vartheta,\sigma}(t|q^*)\right|\mathbf{1}(\hat{p}=q^*).$$

Hence, for every $\vartheta \in \mathbf{R}^P$ and every $\delta>0$

$$P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|\hat{p})-G_{n,\vartheta,\sigma}(t|\hat{p})\right|>\delta\right) \geq P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|\hat{p})-G_{n,\vartheta,\sigma}(t|q^*)\right|>\delta \mid \hat{p}=q^*\right)P_{n,\vartheta,\sigma}(\hat{p}=q^*),$$

observing that the conditional probabilities are well-defined since $P_{n,\vartheta,\sigma}(\hat{p}=q^*)$ is always positive (cf. Leeb (2003b), p.10). This implies

$$\begin{aligned}&\liminf_{n\to\infty}\sup_{\substack{\vartheta\in M_{q^*}\\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|\hat{p})-G_{n,\vartheta,\sigma}(t|\hat{p})\right|>\delta\right)\\ &\geq \liminf_{n\to\infty}\sup_{\substack{\vartheta\in M_{q^*}\\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}\left(\left|\hat{G}_n(t|\hat{p})-G_{n,\vartheta,\sigma}(t|q^*)\right|>\delta \mid \hat{p}=q^*\right)\liminf_{n\to\infty}\inf_{\substack{\vartheta\in M_{q^*}\\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}(\hat{p}=q^*). \quad (59)\end{aligned}$$

Lemma C.1 above shows that

$$\liminf_{n\to\infty}\inf_{\substack{\vartheta\in M_{q^*}\\ ||\vartheta-\theta||<\rho_0/\sqrt{n}}} P_{n,\vartheta,\sigma}(\hat{p}=q^*) = (1-\Delta_{\sigma\xi_{\infty,q^*}}(0,c_{q^*}\sigma\xi_{\infty,q^*}))\prod_{q=q^*+1}^{P}\Delta_{\sigma\xi_{\infty,q}}(0,c_q\sigma\xi_{\infty,q}) = \lim_{n\to\infty}P_{n,\theta,\sigma}(\hat{p}=q^*)$$

which obviously is positive. Suppose now that $\hat{G}_n(t|\hat{p})$ satisfies (34). Then it also satisfies $P_{n,\theta,\sigma}(|\hat{G}_n(t|\hat{p})-G_{n,\theta,\sigma}(t|q^*)|>\delta|\hat{p}=q^*)\overset{n\to\infty}{\longrightarrow}0$, since the probability $P_{n,\theta,\sigma}\left(\hat{p}=q^*\right)$ of the conditioning event is bounded away from zero as just shown. We may now apply Theorem 4.3(b) with $p=q^*$ to the first term in the product on the right-hand side of (59) since $\hat{G}_n(t|\hat{p})$ can certainly also be viewed as an estimator for $G_{n,\theta,\sigma}(t|q^*)$. This establishes (35). Furthermore, note that (59) remains valid if an infimum extending over all estimators is inserted between the limes inferior and the supremum on both sides of (59). Again applying Theorem 4.3(b) with $p=q^*$ completes the proof of (36)-(37). □

Lemma C.3 *For every p satisfying $\mathcal{O} \leq p < P$ and every $0 < \sigma_* \leq \sigma^* < \infty$ we have*

$$\lim_{\rho\to\infty} \limsup_{n\to\infty} \sup_{\substack{\sigma\in\mathbf{R} \\ \sigma_*\leq\sigma\leq\sigma^*}} \sup_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta[\neg p]||\geq\rho/\sqrt{n}}} P_{n,\vartheta,\sigma}(\hat{p} = p) = 0. \tag{60}$$

(The inner supremum in the above display is to be interpreted as extending over $\|\vartheta\| \geq \rho/\sqrt{n}$ if $p = 0$.)

Proof of Lemma C.3: Let $(\vartheta^{(n)}, \sigma^{(n)})$ be an arbitrary sequence satisfying $\vartheta^{(n)} \in \mathbf{R}^P$ and $\sigma_* \leq \sigma^{(n)} \leq \sigma^*$. Then similar as in the proof of Proposition 3.2 any accumulation point of $P_{n,\vartheta^{(n)},\sigma^{(n)}}(\hat{p} = p)$ is given by (52)-(53), with υ_q defined there and σ now representing an accumulation point of $\sigma^{(n)}$. Note that $0 < \sigma_* \leq \sigma \leq \sigma^* < \infty$ holds. By definition of $\upsilon_q^{(n)}$ we can express $(\upsilon_{p+1}^{(n)}, \ldots, \upsilon_P^{(n)})'$ as a linear function of $\sqrt{n}\vartheta^{(n)}[\neg p]$. In particular, $(\upsilon_{p+1}^{(n)}, \ldots, \upsilon_P^{(n)})' = T_n\sqrt{n}\vartheta^{(n)}[\neg p]$, where T_n is an upper triangular square matrix of dimension $P - p$ with all diagonal elements equal to unity. Note that T_n converges to a non-singular matrix T, say, since the elements above the diagonal can be expressed in terms of elements of $X'X/n$. Consequently, $c = \inf_n \lambda_{\min}^{1/2}(T_n'T_n)$ is positive. It follows that $||(\upsilon_{p+1}^{(n)}, \ldots, \upsilon_P^{(n)})'|| \geq c\sqrt{n}||\vartheta^{(n)}[\neg p]||$. Now, for given $\rho > 0$, the expression inside the outer limit operator in (60) is an accumulation point of $P_{n,\vartheta^{(n)},\sigma^{(n)}}(\hat{p} = p)$ for a suitable sequence $(\vartheta^{(n)}, \sigma^{(n)})$ satisfying $||\vartheta^{(n)}[\neg p]|| \geq \rho/\sqrt{n}$ and $\sigma_* \leq \sigma^{(n)} \leq \sigma^*$; it hence is of the form (52)-(53) and $||(\upsilon_{p+1}, \ldots, \upsilon_P)'|| \geq c\rho$ holds. Since all terms in (52)-(53) are bounded by one, since $0 < \sigma < \infty$, and since each of the terms $\Delta_{\sigma\xi_{\infty,q}}(\upsilon_q, c_q\sigma\xi_{\infty,q})$, $p < q \leq P$, converges to zero if $|\upsilon_q|$ diverges to infinity, the expression inside the outer limit operator in (60) converges to zero for $\rho \to \infty$, if we observe that $\max_{p<q\leq P}|\upsilon_q| \geq (c\rho/(P-p))^{1/2}$. □

Proof of Proposition 4.9: Since

$$P_{n,\vartheta,\sigma}\left(\left\|\hat{\Phi}_{n,P}(\cdot) - G_{n,\vartheta,\sigma}(\cdot|\hat{p})\right\|_{TV} > \delta\right) = \sum_{p=\mathcal{O}}^{P} P_{n,\vartheta,\sigma}\left(\left\|\hat{\Phi}_{n,P}(\cdot) - G_{n,\vartheta,\sigma}(\cdot|p)\right\|_{TV} > \delta, \hat{p} = p\right)$$

it suffices to show for every p, $\mathcal{O} \leq p \leq P$, that

$$\sup_{\theta\in\mathbf{R}^P} \sup_{\substack{\sigma\in\mathbf{R} \\ \sigma_*\leq\sigma\leq\sigma^*}} P_{n,\theta,\sigma}\left(\left\|\hat{\Phi}_{n,P}(\cdot) - G_{n,\theta,\sigma}(\cdot|p)\right\|_{TV} > \delta, \hat{p} = p\right) \overset{n\to\infty}{\longrightarrow} 0. \tag{61}$$

In case $p = P$, (61) follows immediately from Proposition 4.6 using the relation $P(A \cap B) \leq P(A)$. Hence, assume $p < P$. Using the relation $P(A \cap B) \leq \min(P(A), P(B))$, (61) follows if

$$\sup_{\substack{\theta\in\mathbf{R}^P \\ ||\theta[\neg p]||<\rho/\sqrt{n}}} \sup_{\substack{\sigma\in\mathbf{R} \\ \sigma_*\leq\sigma\leq\sigma^*}} P_{n,\theta,\sigma}\left(\left\|\hat{\Phi}_{n,P}(\cdot) - G_{n,\theta,\sigma}(\cdot|p)\right\|_{TV} > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{62}$$

holds for every p, $\mathcal{O} \leq p < P$, and for every $\rho > 0$, and if

$$\lim_{\rho\to\infty} \limsup_{n\to\infty} \sup_{\substack{\theta\in\mathbf{R}^P \\ ||\theta[\neg p]||\geq\rho/\sqrt{n}}} \sup_{\substack{\sigma\in\mathbf{R} \\ \sigma_*\leq\sigma\leq\sigma^*}} P_{n,\theta,\sigma}(\hat{p} = p) = 0 \tag{63}$$

holds for every $\mathcal{O} \leq p < P$. Now, (63) follows immediately from Lemma C.3. Furthermore, (62) follows from Proposition 4.6, if we can show that

$$\sup_{\theta\in\mathbf{R}^P} \sup_{\substack{\sigma\in\mathbf{R} \\ \sigma_*\leq\sigma\leq\sigma^*}} P_{n,\theta,\sigma}\left(\left\|\hat{\Phi}_{n,P}(\cdot) - \hat{\Phi}_{n,p}(\cdot)\right\|_{TV} > \delta\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{64}$$

holds for every p satisfying $\mathcal{O} \leq p \leq P$. First observe that the probability in (64) does not depend on θ. Second, Proposition 4.7 shows that $\mathcal{O} > 0$, and also that the matrices $A[p](X[p]'X[p]/n)^{-1}A[p]'$ converge to one and the same positive-definite matrix for $\mathcal{O} \leq p \leq P$. Arguing as in the proof of Proposition 4.6 establishes (64). □

D Auxiliary Lemmata for Appendix C

Lemma D.1 *Let p satisfy $\mathcal{O} < p \leq P$, and assume that $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are asymptotically correlated, i.e., $C_\infty^{(p)} \neq 0$. Moreover, let $\theta \in M_{p-1}$, and let $t \in \mathbf{R}^k$ be such that the set $\{z \in \mathbf{R}^p : A[p]z \leq t\}$ has positive Lebesgue measure (which is satisfied for all $t \in \mathbf{R}^k$ if, e.g., $rank(A[p]) = k$). Then $G_{\infty,\theta,\sigma,\gamma}(t|p)$ is non-constant as a function of $\gamma \in M_p$.*

Lemma D.2 *Let p satisfy $\mathcal{O} \leq p < P$, assume that $A\tilde{\theta}(q)$ and $\tilde{\theta}_q(q)$ are asymptotically correlated, i.e., $C_\infty^{(q)} \neq 0$, for some q satisfying $p < q \leq P$, and let q^* denote the largest q with this property. Moreover, let $t \in \mathbf{R}^k$ and let $\theta \in M_p$. Then $G_{\infty,\theta,\sigma,\gamma}(t|p)$ is non-constant as a function of $\gamma \in M_{q^*}$.*

Before we prove the above lemmata, we provide a representation of $G_{\infty,\theta,\sigma,\gamma}(t|p)$ for $p > 0$ that will be useful in the following. For $0 < p \leq P$ define $Z_p = \sum_{r=1}^{p} \xi_{\infty,r}^{-2} C_\infty^{(r)} W_r$, where $C_\infty^{(r)}$ has been defined after (16) and the random variables W_r are independent normally distributed with mean zero and variances $\sigma^2 \xi_{\infty,r}^2$; for convenience, let Z_0 denote the zero vector in $\mathbf{R}^k$. Observe that Z_p, $p > 0$, is normally distributed with mean zero and variance-covariance matrix $\sigma^2 A[p]Q[p:p]^{-1}A[p]'$ (note that we have shown in the proof of Proposition 4.7 that the asymptotic variance-covariance matrix of $\sqrt{n}A\tilde{\theta}(p)$ can be expressed as $\sum_{r=1}^{p} \sigma^2 \xi_{\infty,r}^{-2} C_\infty^{(r)} C_\infty^{(r)\prime}$). Also the joint distribution of Z_p and the W_r is normal, where the covariance vector between Z_p and W_r is given by $\sigma^2 C_\infty^{(r)}$ in case $r \leq p$; otherwise Z_p and W_r are independent. Define the constants $\nu_r = \gamma_r + (Q[r:r]^{-1}Q[r:\neg r]\gamma[\neg r])_r$ for $0 < r \leq P$. It is now easy to see that $\beta^{(p)}$ defined in Proposition 3.1 equals $-\sum_{r=p+1}^{P} \xi_{\infty,r}^{-2} C_\infty^{(r)} \nu_r$. (This is seen as follows: it was noted in Proposition 3.1 that $\beta^{(p)} = \lim_{n\to\infty} \sqrt{n}A(\eta_n(p) - \theta - \gamma/\sqrt{n})$ for $\theta \in M_p$, when $\eta_n(p)$ is defined as in (9), but with $\theta + \gamma/\sqrt{n}$ replacing θ. Using the representation (20) of Leeb (2003a) and taking limits the result follows if we observe that $\sqrt{n}\eta_{n,r}(r) \longrightarrow \nu_r$ for $\theta \in M_p$.) In view of (18), the cdf $G_{\infty,\theta,\sigma,\gamma}(t|p)$ can now equivalently be written as

$$G_{\infty,\theta,\sigma,\gamma}(t|p) = P\left(Z_p \leq t + \sum_{r=p+1}^{P} \xi_{\infty,r}^{-2} C_\infty^{(r)} \nu_r\right) \tag{65}$$

in case $p = \max\{p_0(\theta), \mathcal{O}\} > 0$, and (65) trivially holds in case $p = 0$. In case $p > \max\{p_0(\theta), \mathcal{O}\}$ the cdf $G_{\infty,\theta,\sigma,\gamma}(t|p)$ is given by (19), and it is elementary but tedious to show, following the steps in Section 3.1 of Leeb (2003a), that this is equivalent to

$$G_{\infty,\theta,\sigma,\gamma}(t|p) = P\left(Z_p \leq t + \sum_{r=p+1}^{P} \xi_{\infty,r}^{-2} C_\infty^{(r)} \nu_r \,\middle|\, |W_p + \nu_p| \geq c_p \sigma \xi_{\infty,p}\right). \tag{66}$$

(This can also be derived from the fact that the distribution of $(Z_p', W_p + \nu_p, \ldots, W_P + \nu_P)'$ represents the limiting distribution of $\sqrt{n}(A(\tilde{\theta}(p) - \eta_n(p))', \tilde{\theta}_p(p), \ldots, \tilde{\theta}_P(P))'$ under $P_{n,\theta+\gamma/\sqrt{n},\sigma}$ with $\theta \in M_{p-1}$ (and $\eta_n(p)$ defined as in (9), but with $\theta + \gamma/\sqrt{n}$ replacing θ.))

Proof of Lemma D.1: Since $\theta \in M_{p-1}$, $G_{\infty,\theta,\sigma,\gamma}(t|p)$ is given by (66). For $\gamma \in M_p$ the quantities $\nu_{p+1}, \ldots, \nu_P$ are easily seen to be zero, while ν_p equals γ_p. This leads to

$$G_{\infty,\theta,\sigma,\gamma}(t|p) = P\left(Z_p \leq t \,|\, |W_p + \gamma_p| \geq c_p \sigma \xi_{\infty,p}\right)$$

for $\gamma \in M_p$. Since $Z_p = Z_{p-1} + \xi_{\infty,p}^{-2} C_\infty^{(p)} W_p$, we obtain

$$G_{\infty,\theta,\sigma,\gamma}(t|p) = P\left(Z_{p-1} + \xi_{\infty,p}^{-2} C_\infty^{(p)} W_p \leq t \,\middle|\, |W_p + \gamma_p| \geq c_p \sigma \xi_{\infty,p}\right). \tag{67}$$

Assume now that (67) is constant in $\gamma_p \in \mathbf{R}$. Using Lemma D.3 below with $Z_{p-1} - t$, W_p, $-\xi_{\infty,p}^{-2} C_\infty^{(p)}$, $-\gamma_p$, and $c_p \sigma \xi_{\infty,p}$ replacing Z, W, C, x, and δ, respectively, we obtain that either $P(Z_p \le t) = 0$ or that $\xi_{\infty,p}^{-2} C_\infty^{(p)} = 0$. By assumption of the lemma the set $\{z \in \mathbf{R}^p : A[p]z \le t\}$ has positive Lebesgue measure. Hence, $P(Z_p \le t)$ must be positive. (To see why, note that Z_p is concentrated in the column space of $A[p]$, and that Z_p is non-degenerate *within* the column-space of $A[p]$.) It would follow that $\xi_{\infty,p}^{-2} C_\infty^{(p)} = 0$, contradicting the assumption that $A\tilde{\theta}(p)$ and $\tilde{\theta}_p(p)$ are asymptotically correlated. □

Proof of Lemma D.2: By the assumptions on q^*, note that either $q^* = P$ or that $C_\infty^{(r)} = 0$ for each $r = q^* + 1, \ldots, P$. Consider first the case $p = \max\{p_0(\theta), \mathcal{O}\}$. By (65) we have $G_{\infty,\theta,\sigma,\gamma}(t|p) = P\left(Z_p \le t + \sum_{r=p+1}^{q^*} \xi_{\infty,r}^{-2} C_\infty^{(r)} \nu_r\right)$. Observe that $(\nu_{p+1}, \ldots, \nu_{q^*})'$ varies in all of $\mathbf{R}^{q^*-p}$ when γ varies in M_{q^*}. Hence, the last mentioned probability goes to zero along an appropriate sequence of $(\nu_{p+1}, \ldots, \nu_{q^*})'$ (namely a sequence along which at least one coordinate of $t + \sum_{r=p+1}^{q^*} \xi_{\infty,r}^{-2} C_\infty^{(r)} \nu_r$ goes to $-\infty$). Since $Z_{q^*} = Z_p + \sum_{r=p+1}^{q^*} \xi_{\infty,r}^{-2} C_\infty^{(r)} W_r$ and since the W_r, $r = p+1, \ldots, P$, are independent of Z_p, the cdf $G_{\infty,\theta,\sigma,\gamma}(t|p)$ can also be written as a (regular) conditional probability

$$G_{\infty,\theta,\sigma,\gamma}(t|p) = P\left(Z_{q^*} \le t \,|\, W_{p+1} = -\nu_{p+1}, \ldots, W_{q^*} = -\nu_{q^*}\right). \tag{68}$$

Suppose now that $G_{\infty,\theta,\sigma,\gamma}(t|p)$ is constant in $\gamma \in M_{q^*}$, or equivalently, is constant when $(\nu_{p+1}, \ldots, \nu_{q^*})'$ varies in all of $\mathbf{R}^{q^*-p}$. It follows from the above discussion that the conditional probability in (68) is then zero for all $(\nu_{p+1}, \ldots, \nu_{q^*})' \in \mathbf{R}^{q^*-p}$. By integration with respect to the distribution of $(W_{p+1}, \ldots, W_{q^*})$ we obtain that $P(Z_{q^*} \le t) = 0$. From Proposition 4.7(c) it follows that Z_{q^*} has a non-singular normal distribution on $\mathbf{R}^k$, which contradicts $P(Z_{q^*} \le t) = 0$. This proves the lemma in case $p = \max\{p_0(\theta), \mathcal{O}\}$. Consider next the case $p > \max\{p_0(\theta), \mathcal{O}\}$ and assume that $G_{\infty,\theta,\sigma,\gamma}(t|p)$ is constant in $\gamma \in M_{q^*}$. Now $G_{\infty,\theta,\sigma,\gamma}(t|p)$ is given by (66). Letting $\gamma_p \to \infty$, ν_p converges to ∞ as well, and the expression in (66) converges to that in (65). Hence, (65) would have to be constant as a function of $(\nu_{p+1}, \ldots, \nu_{q^*})'$ (note that $(\nu_{p+1}, \ldots, \nu_{q^*})'$ depends only on $\gamma[\neg p]$ but not on γ_p), which already has been shown to lead to a contradiction. □

Lemma D.3 *Let Z be a random vector with values in $\mathbf{R}^k$, let W be a univariate random variable independent of Z, and assume that W has a Lebesgue density which is positive almost everywhere. Furthermore, let $C \in \mathbf{R}^k$ and let $\delta > 0$. Then $P(Z \le CW \,|\, |W - x| \ge \delta)$ is constant in $x \in \mathbf{R}$ if and only if $P(Z \le CW) = 0$ or $C = 0$.*

Proof of Lemma D.3: If $C = 0$, then $P(Z \le CW \,|\, |W - x| \ge \delta)$ equals $P(Z \le 0)$, which is constant in x. If $P(Z \le CW) = 0$, obviously also $P(Z \le CW \,|\, |W - x| \ge \delta) = 0$, and hence is constant in x. Conversely, assume that $P(Z \le CW \,|\, |W - x| \ge \delta) = P(Z \le CW \,|\, |W - x'| \ge \delta)$ for each $x, x' \in \mathbf{R}$. Letting $x' \to \infty$ implies that

$$\frac{P(Z \le CW,\, |W - x| \ge \delta)}{P(|W - x| \ge \delta)} = P(Z \le CW)$$

holds for each $x \in \mathbf{R}$. This is equivalent to

$$P(Z \le CW,\, W \in B) = P(Z \le CW)\, P(W \in B), \tag{69}$$

for all sets B of the form $B = (x - \delta, x + \delta)$ with $x \in \mathbf{R}$. Since both sides in (69) are sigma-additive set functions and since W is absolutely continuous with respect to Lebesgue measure, both set functions also agree on all sets of the form $(-\infty, x + \delta]$, and hence on the entire Borel sigma-field on $\mathbf{R}$. This implies independence of $\{Z \le CW\}$ and W. In particular, we have

$$P(Z \le CW) = P(Z \le CW \,|\, W = w)$$

for almost all $w \in \mathbf{R}$. Furthermore, by the assumed independence of Z and W, we have

$$P(Z \leq CW) = P(Z \leq CW \mid W = w) = P(Z \leq Cw)$$

for almost all $w \in \mathbf{R}$. Now if $C \neq 0$, the right-hand side of the above display goes to zero either for $w \to \infty$ or for $w \to -\infty$, implying that $P(Z \leq CW) = 0$. $\square$

Remark D.4 Suppose that W is independent of Z, that W has a density with respect to Lebesgue measure, and that $P(|W - x| \geq \delta)$ is positive for each $x \in \mathbf{R}$. Then Lemma D.3 continues to hold under this weaker condition on W. If W is as above but is not absolutely continuous with respect to Lebesgue measure, then Lemma D.3 continues to hold if $P(Z \leq CW \mid |W - x| \geq \delta)$ is replaced either by $P(Z \leq CW \mid W < x - \delta \text{ or } W \geq x + \delta)$ or by $P(Z \leq CW \mid W \leq x - \delta \text{ or } W > x + \delta)$. If W is as above but $P(|W - x| \geq \delta) = 0$ for some $x \in \mathbf{R}$, Lemma D.3 continues to hold if the condition that $P(Z \leq CW \mid |W - x| \geq \delta)$ be constant for all $x \in \mathbf{R}$ is replaced by the condition that this holds only for those x satisfying $P(|W - x| \geq \delta) > 0$. (Since $\lim_{x \to \infty} P(|W - x| \geq \delta) = 1$, this condition is not empty.)

E Proofs for Section 5

Proof of Theorem 5.1: After rearranging the elements of θ (and hence the regressors) and correspondingly rearranging the rows of the matrix A if necessary, we may assume without loss of generality that $\mathfrak{r}_* = (1, \ldots, 1, 0)$, and hence that $i(\mathfrak{r}_*) = P$. That is, $M_{\mathfrak{r}_*} = M_{P-1}$ and $M_{\mathfrak{r}_{full}} = M_P$. Furthermore, note that after this arrangement $C_\infty^{(P)} \neq 0$. Let $\hat{p}$ be the model selection procedure introduced in Section 2 with $\mathcal{O} = P - 1$ and $c_P = c$. Let $\tilde{\theta}$ be the corresponding post-model-selection estimator and let $G_{n,\theta,\sigma}(t|p)$ be as defined in Section 3. Condition (40) can now equivalently be written as follows: For every $\theta \in M_{P-1}$ which has exactly $P-1$ non-zero coordinates

$$\lim_{n\to\infty} P_{n,\theta,\sigma}\left(\{\hat{\mathfrak{r}} = \mathfrak{r}_{full}\} \blacktriangle \{\hat{p} = P\}\right) = \lim_{n\to\infty} P_{n,\theta,\sigma}\left(\{\hat{\mathfrak{r}} = \mathfrak{r}_*\} \blacktriangle \{\hat{p} = P-1\}\right) = 0 \tag{70}$$

holds for every $0 < \sigma < \infty$. Since the sequences $P_{n,\vartheta^{(n)},\sigma}$ and $P_{n,\theta,\sigma}$ are contiguous for $\vartheta^{(n)}$ satisfying $\left\|\theta - \vartheta^{(n)}\right\| = O(n^{-1/2})$ as remarked at the beginning of Appendix C, it follows that condition (70) continues to hold with $P_{n,\vartheta^{(n)},\sigma}$ replacing $P_{n,\theta,\sigma}$. This implies that for every sequence of positive real numbers s_n with $s_n = O(n^{-1/2})$, for every σ, $0 < \sigma < \infty$, and for every $\theta \in M_{P-1}$ which has exactly $P-1$ non-zero coordinates

$$\liminf_{n\to\infty} \inf_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<s_n}} P_{n,\vartheta,\sigma}(\hat{\mathfrak{r}} = \mathfrak{r}_{full}) = \liminf_{n\to\infty} \inf_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<s_n}} P_{n,\vartheta,\sigma}(\hat{p} = P) > 0, \tag{71}$$

$$\liminf_{n\to\infty} \inf_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<s_n}} P_{n,\vartheta,\sigma}(\hat{\mathfrak{r}} = \mathfrak{r}_*) = \liminf_{n\to\infty} \inf_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<s_n}} P_{n,\vartheta,\sigma}(\hat{p} = P-1) > 0, \tag{72}$$

hold, the positivity following from Proposition 3.2. A further consequence is that

$$\sup_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<s_n}} \|K_{n,\vartheta,\sigma}(t|\mathfrak{r}_{full}) - G_{n,\vartheta,\sigma}(t|P)\|_{TV} \to 0, \tag{73}$$

$$\sup_{\substack{\vartheta\in\mathbf{R}^P \\ ||\vartheta-\theta||<s_n}} \|K_{n,\vartheta,\sigma}(t|\mathfrak{r}_*) - G_{n,\vartheta,\sigma}(t|P-1)\|_{TV} \to 0 \tag{74}$$

as $n \to \infty$. From (73)-(74) we conclude that the limit of $K_{n,\theta,\sigma}(t|\mathfrak{r}_{full})$ (with respect to total variation distance) under local alternatives $\theta + \gamma/\sqrt{n}$ exists, and coincides with $G_{\infty,\theta,\sigma,\gamma}(t|P)$. The same is true for the limit of $K_{n,\theta,\sigma}(t|\mathfrak{r}_*)$ under local alternatives and $G_{\infty,\theta,\sigma,\gamma}(t|P-1)$. Because of (71)-(72) we may assume that all relevant probabilities are positive (at least from a certain n_0 onwards). Repeating the proof of Theorem 4.3 with $p = P$ and where $K_{n,\vartheta,\sigma}(t|\mathfrak{r}_{full})$ replaces $G_{n,\vartheta,\sigma}(t|P)$, as well as repeating the proof of Theorem 4.5 with $p = P - 1$, $q^* = P$ and where $K_{n,\vartheta,\sigma}(t|\mathfrak{r}_*)$ replaces $G_{n,\vartheta,\sigma}(t|P-1)$, gives the desired result. □

Proof of Theorem 5.2: Observe that (70)-(74) again hold and that

$$\lim_{n\to\infty} P_{n,\theta,\sigma}(\hat{\mathfrak{r}} = \mathfrak{r}_{full}) = \lim_{n\to\infty} P_{n,\theta,\sigma}(\hat{p} = P) > 0,$$

$$\lim_{n\to\infty} P_{n,\theta,\sigma}(\hat{\mathfrak{r}} = \mathfrak{r}_*) = \lim_{n\to\infty} P_{n,\theta,\sigma}(\hat{p} = P-1) > 0.$$

Repeating the proof of Theorem 4.8 with $q^* = P$, with $K_{n,\vartheta,\sigma}(t|\hat{\mathfrak{r}})$ replacing $G_{n,\vartheta,\sigma}(t|\hat{p})$, and with using Theorem 5.1(b) instead of Theorem 4.3(b) gives the desired result. □

F References

Ahmed, S. E. & A. K. Basu (2000): Least squares, preliminary test and Stein-type estimation in general vector AR(p) models. *Statistica Neerlandica* 54, 47–66.

Bauer, P., Pötscher, B. M. & P. Hackl (1988): Model selection by multiple test procedures. *Statistics* 19, 39–44.

Danilov, D. L. & J. R. Magnus (2001): On the harm that pre-testing does. Working Paper, Center for Economic Research, Tilburg University.

Dijkstra, T. K. & J. H. Veldkamp (1988): 'Data-driven selection of regressors and the bootstrap'. *Lecture Notes in Economics and Mathematical Systems* 307, 17–38.

Dukić, V. M. & E. A Peña (2002): Estimation after model selection in a gaussian model. Manuscript.

Hjort, N. L. & G. Claeskens (2002): Frequentist model average estimators. Manuscript.

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

Kapetanios, G. (2001): Incorporating lag order selection uncertainty in parameter inference for AR models, *Economics Letters* 72, 137–144.

Leeb, H. (2003a): The distribution of a linear predictor after model selection: conditional finite-sample distributions and asymptotic approximations. Working Paper, Department of Statistics, University of Vienna.

Leeb, H. (2003b): The distribution of a linear predictor after model selection: unconditional finite-sample distributions and asymptotic approximations. Working Paper, Department of Statistics, University of Vienna.

Leeb, H. & B. M. Pötscher (2002): Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk bound results. Working Paper, Department of Statistics, University of Vienna.

Leeb, H. & B. M. Pötscher (2003a): The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory* 19, 100–142.

Leeb, H. & B. M. Pötscher (2003b): Can one estimate the unconditional distribution of post-model-selection estimators? Working Paper, Department of Statistics, University of Vienna.

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

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

Pötscher, B. M. & A. J. Novak (1998): The distribution of estimators after model selection: large and small sample results. *Journal of Statistical Computation and Simulation* 60, 19–56.

Rao, C. R. & Y. Wu (2001): 'On model selection,' *IMS Lecture Notes Monograph Series* 83, 1–57.

Sen, P. K (1979): Asymptotic properties of maximum likelihood estimators based on conditional specification. *Annals of Statistics* 7, 1019–1033.

Sen P. K & A. K. M. E. Saleh (1987): On preliminary test and shrinkage M-estimation in linear models. *Annals of Statistics* 15, 1580–1592.