MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

"An Analysis of a proposed Model-free Approach to Linear Least Squares Regression"

verfasst von / submitted by

Manuel Hahn, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. oec.)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 951
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Statistik
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Hannes Leeb

1 Abstract

In their paper "A Model-free Approach to Linear Least Squares Regression with Exact Probabilities and Applications to Covariate Selection", Laurie Davies and Lutz Dümbgen propose a new approach to model-selection in a linear least squares regression setting that neither uses hypotheses nor requires a model for the distribution of the error term. Among other things they provide a way of calculating valid p-values and selecting a subset of regressors in a step-wise manner for any given real-valued data. The aim of this thesis is to provide an overview and commentary of their approach as well as of some of their stated theorems and corresponding proofs, focusing on the draft version of their paper from June 5 2019.

Contents

2 Introduction

2.1 Problem and general approach

Consider the standard least-squares regression problem:
Given data $y \in \mathbb{R}^n, x_j \in \mathbb{R}^n, j = 1, ..., q$ for some $n \in \mathbb{N}$ and $q < n$ we assume y to be a result of the relationship

$$y_i = \sum_{j=1}^{q} x_{j,i}\beta_j + \epsilon_i \;\; \forall i \in \{1, ..., n\}$$

where $\beta_j \in \mathbb{R}, j = 1, ..., q$ and $\epsilon_i, i = 1, ..., n$ are real-valued random variables. The goal of a model-selection approach in this setting is to, in a sense, find out which of the vectors x_j contribute to explaining the non-random part of y, which can be interpreted to mean $\beta_j \neq 0$, and which of the vectors x_j do not contribute, which can be interpreted to mean $\beta_j = 0$.

One typical way of addressing this problem is to assume a model for the random variables ϵ_i, e.g. assuming that they are i.i.d. and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ for some $\sigma \in \mathbb{R}^+$. To compare the viability of two candidate subsets of covariates $\mathcal{M}_0 \subset \mathcal{M}_1 \subseteq \{1, ..., q\}$ one would then formulate the null hypothesis $H_0 : \beta_j = 0 \; \forall j \in \mathcal{M}_1 \backslash \mathcal{M}_0$ for testing if the covariates that are included in subset $\mathcal{M}_1$ but not in subset $\mathcal{M}_0$ are relevant for explaining the value of y. The advantage of this approach is that a p-value for this hypothesis can explicitly be calculated using a test statistic distributed with Fisher's F-distribution. But this comes with the drawback of having to rely on rather strong assumptions, in particular the type of distribution of ϵ and the set of 'candidate' covariates being correct in some sense, in order not to introduce bias, as explored in for example [Berk et al., 2013].

In their paper, Laurie Davies and Lutz Dümbgen propose approaching this problem from a different angle which does not rely on assumptions to the nature of the error term ϵ_i, but nevertheless enables one to calculate valid p-values [Davies & Dümbgen, 2019]. Concretely their method asks the question of how likely it is that the covariates included in the subset $\mathcal{M}_1$ but not in $\mathcal{M}_0$ in a sense add more to the explanation of the value of y than an instance of the same number of white-noise covariates, in particular i.i.d. standard-normally distributed random variables, would. This 'goodness of explanation' so to speak is measured by the residual sum of squares. For a given set of covariates $\mathcal{M}$ the residual sum of squares is defined as follows:

$$ss(\mathcal{M}) = ||y - P_{\mathcal{M}}(y)||^2$$

where $||.||$ denotes the l^2-norm and $P_{\mathcal{M}}$ denotes the projection onto the space $\mathbb{V}_{\mathcal{M}} = span\{x_i : i \in \mathcal{M}\}$.
For a given set of covariates $\mathcal{M}$ and a natural number k the alternative residual sum of squares is defined as follows:

$$SS(\mathcal{M}, k) = ||y - P_{\mathcal{M}}^{(k)}(y)||^2$$

where $P_{\mathcal{M}}^{(k)}$ denotes the projection onto the (random) space
$\mathbb{V}_{\mathcal{M}}^{(k)} = span\{x_i : i \in \mathcal{M}, Z_1, ..., Z_k\}$ where $Z_1, ..., Z_k$ are i.i.d. random vectors with $Z_j \sim \mathcal{N}(0, I_n)$.
Define the concrete residual sums of squares

$$ss_0 := ss(\mathcal{M}_0)$$

$$ss_1 := ss(\mathcal{M}_1)$$

$$SS_1 := SS(\mathcal{M}_0, |\mathcal{M}_1 \setminus \mathcal{M}_0|).$$

As informally described above, the interesting probability in question is

$$P(SS_1 < ss_1).$$

Note that all randomness resides in SS_1 and stems from the introduction of the random vectors $Z_1, ..., Z_k$. Low values of this probability indicate that the covariates included in the subset $\mathcal{M}_1$ but not in $\mathcal{M}_0$ likely add more to the explanation of y, represented by the residual sum of squares, than an instance of the same number of mere Gaussian white noise covariates would. This can roughly be interpreted as the probability that some completely unrelated data simply by chance appears to explain y as well or better than the data that was actually collected. The authors define p-values based on this, which in turn can be used to formulate selection-procedures, some basic ones of which are covered below.
One of the key results in the paper is the fact that for $m_1 := |\mathcal{M}_1|$ and $m_0 := |\mathcal{M}_0|$ it holds, that

$$\begin{aligned} P(SS_1 < ss_1) &= F_{Beta(\frac{n-m_1}{2}, \frac{m_1-m_0}{2})}\left(\frac{ss_1}{ss_0}\right) \\ &= F_{F(m_1-m_0, n-m_1)}\left(\frac{ss_0 - ss_1}{m_1 - m_0} / \frac{ss_1}{n - m_1}\right) \end{aligned}$$

which is called Theorem 2.1 in the paper [Davies & Dümbgen, 2019, pg. 4] and is elaborated as Theorem 1 in this thesis. Note that F_D denotes the cumulative distribution function ('cdf') of a random variable distributed with the distribution D. This result not only gives us an explicit expression for the desired probability $P(SS_1 < ss_1)$ with a *Beta*-distribution but also shows us that the resulting p-values of the proposed approach are the same as the classical approach via hypotheses with an F-distribution, even though the latter relies on fairly strong assumptions, while the former does not. It should be noted however that the interpretations of the results of these two approaches are different.

2.2 Selection procedure: all subsets

Given the results from the section above the authors first specify a simple, theoretically optimal, selection procedure based on them [Davies & Dümbgen, 2019, pg. 4f]. The approach is as follows:

Select a significance threshold $\alpha \in (0,1)$. For any given subset of covariates $\mathcal{M} = \{j_1, j_2, ..., j_m\} \subseteq \{1, ..., q\}$ let the p-value of the covariate x_{j_l} be given by

$$P(SS(\mathcal{M} \setminus \{j_l\}, 1) < ss(\mathcal{M})) = F_{Beta(\frac{n-m}{2}, \frac{1}{2})}\left(\frac{ss(\mathcal{M})}{ss(\mathcal{M} \setminus \{j_l\})}\right)$$

and the p-value of the whole subset by

$$p(\mathcal{M}) = \max_{j_l \in \mathcal{M}} P(SS(\mathcal{M} \setminus \{j_l\}, 1) < ss(\mathcal{M})).$$

Note that these p-values are always correctly representing what they are intended to, independently of the 'real' structure of the data, since contrary to the classical approach they do not depend on the assumption that $\mathcal{M}$ is the 'correct' subset of covariates or that the error term ϵ has any specific distribution.
Now calculate this p-value $p(\mathcal{M})$ for every subset $\mathcal{M} \in \mathcal{P}(\{1, ..., q\})$. Any subset whose p-value exceeds the threshold α is discarded. From the remaining subsets, only maximal ones are kept, i.e. remaining sets which are strict subsets of another remaining set are discarded as well.

After this procedure is concluded the remaining subsets represent the 'maximal' subsets that can be obtained via this approach, since no additional covariate can be added to any of the subsets without causing the p-value to rise above the threshold α. If one wants to choose a single subset from possibly multiple remaining ones, a way to choose can be to select the one with the lowest residual sum of squares.

It should be noted that this procedure is mostly a theoretical construct of how to optimally use this Gaussian covariate approach for model selection, since in practice for the vast majority of problems it will be infeasible to calculate the p-value for every single one of the $\sum_{m=1}^{q} \binom{q}{m} m$ subset-covariate combinations.

2.3 Selection procedure: single stepwise

To avoid the problem of the exploding number of p-values that have to be calculated in the all subsets procedure, the authors describe a single stepwise procedure of covariate-selection which they aim to show with their Theorems 4.1, 4.2 and 4.3 to be asymptotically consistent in the i.i.d. Gaussian error-case under some assumptions [Davies & Dümbgen, 2019, pg. 5-10, 25-31]. In the 2019 draft version of their paper however these proofs omit some key arguments, which are being commented on in the elaboration of Hypothesis 2 and Theorem 3 in this thesis, which are analogous to Theorems 4.1 and 4.2 in the paper.

In order to describe the single stepwise procedure it suffices to describe a single step and a stopping criterion. Starting from an already selected subset $\mathcal{M}_0 \subset \{1, ..., q\}$ with $|\mathcal{M}_0| =: m_0$, a step either chooses a new covariate to be added to form a new subset or stops the procedure. We denote $ss_0 := ss(\mathcal{M}_0)$ and $ss_j := ss(\mathcal{M}_j)$ where $\mathcal{M}_j := \mathcal{M}_0 \cup \{j\}$ for $j \in \{1, ..., q\} \setminus \mathcal{M}_0$. Furthermore, we denote SS_l to represent the alternative residual sum of squares of the subset $\mathcal{M}_0$ extended by covariate x_l. In other words, each SS_l denotes an i.i.d. copy of $SS(\mathcal{M}_0, 1)$. By Theorem 1, we have for any $j, l \in \{1, ..., q\} \setminus \mathcal{M}_0$

$$P(ss_j < SS_l) = 1 - F_{Beta(\frac{n-m_0-1}{2}, \frac{1}{2})}\left(\frac{ss_j}{ss_0}\right) = F_{Beta(\frac{1}{2}, \frac{n-m_0-1}{2})}\left(\frac{ss_j}{ss_0}\right).$$

The p-value of a covariate x_j with $j \in \{1, ..., q\} \setminus \mathcal{M}_0$ is now being defined as the probability of adding any of the $q - m_0$ possible alternative white-noise covariates giving better information about y than adding the covariate x_j. In mathematical terms

$$p_j(\mathcal{M}_0) = 1 - P(ss_j < \min_{l \in \{1,...,q\} \setminus \mathcal{M}_0} SS_l).$$

Note that since the SS_l are i.i.d. we can simplify

$$\begin{aligned} p_j(\mathcal{M}_0) &= 1 - P(ss_j < \min_{l \in \{1,...,q\} \setminus \mathcal{M}_0} SS_l) = 1 - P(ss_j < SS_l)^{q-m_0} \\ &= 1 - F_{Beta(\frac{1}{2}, \frac{n-m_0-1}{2})}\left(\frac{ss_j}{ss_0}\right)^{q-m_0}. \end{aligned}$$

Now that we can calculate the p-value of adding any covariate x_j to the given subset $\mathcal{M}_0$, all we need to do is calculate this p-value $p_j(\mathcal{M}_0)$ for any $j \in \{1, ..., q\} \setminus \mathcal{M}_0$. If none of the p-values fulfil the relationship $p_j < \alpha$, then the procedure terminates and $\mathcal{M}_0$ is defined as the selected subset. If at least one p-value fulfils $p_j < \alpha$, then the procedure continues to the next step with the new subset $\mathcal{M}_\nu$ where ν denotes an index such that $p_\nu(\mathcal{M}_0) = \min_{j \in \{1,...,q\} \setminus \mathcal{M}_0} p_j(\mathcal{M}_0)$.

3 Statements and Proofs

3.1 Theorem 1

Let $x_j \in \mathbb{R}^n, j = 1, ..., q$. Let $\mathcal{M}_0 \subset \{1, ..., q\}$ such that $m_0 < n$, where $m_0 := |\mathcal{M}_0|$ and $x_j, j \in \mathcal{M}_0$ are linearly independent.
Let $\epsilon_j^{(0)} \in \mathbb{R}, j = 1, ..., n$. Let ss_0 be the RSS of the regression problem

$$y_i = \sum_{j \in \mathcal{M}_0} x_{ij}\beta_j^{(0)} + \epsilon_i^{(0)},$$

i.e. for $\mathbb{V}_0 := span(x_j : j \in \mathcal{M}_0)$ let

$$ss_0 := ||y - P_{\mathbb{V}_0}(y)||^2.$$

Let $\mathcal{M}_1$ be such, that $\mathcal{M}_0 \subset \mathcal{M}_1 \subseteq \{1, ..., q\}$, with $m_1 := |\mathcal{M}_1| < n$ and let SS_1 be the RSS of the regression problem

$$y_i = \sum_{j \in \mathcal{M}_0} x_{ij}\beta_j^{(1)} + \sum_{j \in \mathcal{M}_1 \setminus \mathcal{M}_0} z_{ij}\beta_j^{(1)} + \epsilon_i^{(1)}$$

where $z_{ij}, i = 1, ..., n, j \in \mathcal{M}_1 \setminus \mathcal{M}_0$ are iid $\mathcal{N}(0,1)$ random variables independent of $\epsilon^{(0)}$ and we write Z_k for the vector consisting of $(z_{ik})_{i \in \{1,...,n\}}$ and Z for the matrix consisting of the columns $(Z_k)_{k \in \mathcal{M}_1 \setminus \mathcal{M}_0}$.
That means for $\mathbb{V}_Z := span(x_j : j \in \mathcal{M}_0, Z_k : k \in \mathcal{M}_1 \setminus \mathcal{M}_0)$ we have

$$SS_1 := ||y - P_{\mathbb{V}_Z}(y)||^2.$$

Then the following holds for $ss_1 > 0$:

(i) $\frac{SS_1}{ss_0} \sim Beta(\frac{n-m_1}{2}, \frac{m_1-m_0}{2})$

(ii) $P(SS_1 < ss_1) = F_{Beta(\frac{n-m_1}{2}, \frac{m_1-m_0}{2})}(\frac{ss_1}{ss_0})$

(iii) $= F_{F(m_1-m_0, n-m_1)}(\frac{ss_0-ss_1}{m_1-m_0} / \frac{ss_1}{n-m_1})$.

Note that this Theorem and its proof are a more detailed retelling of Theorem 2.1 from the original paper [Davies & Dümbgen, 2019, pg. 4, 23ff].

3.2 Proof of Theorem 1

Case $m_1 - m_0 = 1$:

Here we have $\mathcal{M}_1 \setminus \mathcal{M}_0 = \{\nu\}$. Let $\mathbb{V}_0^\perp$ be the orthogonal complement of $\mathbb{V}_0$, i.e. $\mathbb{V}_0^\perp := \{x \in \mathbb{R}^n : x \perp \eta \, \forall \eta \in \mathbb{V}_0\}$. Note that $\mathbb{V}_0^\perp$ is nonempty because $m_0 < n$. Let $b_i, i = 1, ..., n$ be an orthonormal basis of $\mathbb{R}^n$ such that

$$\mathbb{V}_0 = span(b_1, ..., b_{m_0}) \Rightarrow \mathbb{V}_0^\perp = span(b_{m_0+1}, ..., b_n)$$

and

$$b_{m_0+1} = \frac{y - P_{\mathbb{V}_0}(y)}{\sqrt{ss_0}}.$$

(Note that in the original paper the authors state that they define $b_{m_0+1} = \frac{y-P_{\mathbb{V}_0}(y)}{\sqrt{ss_0}^{-1}}$ [Davies & Dümbgen, 2019, pg. 23], which seems to be a mistake since we want b_{m_0+1} to be of unit-length and $||y - P_{\mathbb{V}_0}(y)|| = \sqrt{ss_0}$.)

We now define $\tilde{z}_j := b_j' Z_\nu$. By the rotational symmetry of the $\mathcal{N}(0,1)$-distribution these variables are again iid $\mathcal{N}(0,1)$-distributed. Further define

$$\tilde{Z}_\nu := P_{\mathbb{V}_0^\perp}(Z_\nu) = \sum_{j=m_0+1}^{n} Z_\nu' b_j \cdot b_j = \sum_{j=m_0+1}^{n} \tilde{z}_j \cdot b_j$$

and with that we have

$$\mathbb{V}_Z = span(b_1, ..., b_{m_0}, Z_\nu) = span(b_1, ..., b_{m_0}, P_{\mathbb{V}_0}(Z_\nu) + P_{\mathbb{V}_0^\perp}(Z_\nu))$$

$$= span(b_1, ..., b_{m_0}, P_{\mathbb{V}_0^\perp}(Z_\nu)) = span(b_1, ..., b_{m_0}, \tilde{Z}_\nu).$$

In other words, $b_1, ..., b_{m_0}, \tilde{Z}_\nu$ form an orthogonal basis for $\mathbb{V}_Z$. Writing the corresponding projection via this basis, we obtain (as shown below)

$$P_{\mathbb{V}_Z}(y) = P_{\mathbb{V}_0}(y) + \sqrt{ss_0} \frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2} \tilde{Z}_\nu. \tag{1}$$

(Note again that in the original paper this equation reads $P_{\mathbb{V}_Z}(y) = P_{\mathbb{V}_0}(y) - \sqrt{ss_0} \frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2} \tilde{Z}_\nu$ [Davies & Dümbgen, 2019, pg. 24] which seems to be a mistake.)

From this, it follows that

$$SS_1 = ss_0 - ss_0 \frac{(\tilde{Z}_\nu' b_{m_0+1})^2}{||\tilde{Z}_\nu||^2}. \tag{2}$$

And furthermore

$$\frac{SS_1}{ss_0} = 1 - \frac{(\tilde{Z}_\nu' b_{m_0+1})^2}{||\tilde{Z}_\nu||^2} = 1 - \frac{((\sum_{j=m_0+1}^n \tilde{z}_j b_j)' b_{m_0+1})^2}{||\tilde{Z}_\nu||^2} = \frac{||\tilde{Z}_\nu||^2}{||\tilde{Z}_\nu||^2} - \frac{\tilde{z}_{m_0+1}^2}{||\tilde{Z}_\nu||^2}$$

$$= \frac{\sum_{j=m_0+2}^n \tilde{z}_j^2}{\sum_{j=m_0+1}^n \tilde{z}_j^2} = \frac{\mathcal{X}_{n-m_0-1}^2}{\mathcal{X}_{n-m_0-1}^2 + \mathcal{X}_1^2} \sim Beta\left(\frac{n-m_0-1}{2}, \frac{1}{2}\right).$$

To show (1) and (2) first note the following:

$$\tilde{Z}_\nu' y = \tilde{Z}_\nu'(y - P_{\mathbb{V}_0}(y)) = \sqrt{ss_0}\tilde{Z}_\nu' \frac{y - P_{\mathbb{V}_0}(y)}{\sqrt{ss_0}} = \sqrt{ss_0}\tilde{Z}_\nu' b_{m_0+1} \tag{*}$$

Ad (1):

$$P_{\mathbb{V}_Z}(y) = P_{\mathbb{V}_0}(y) + \frac{\tilde{Z}_\nu' y}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu \overset{(*)}{=} P_{\mathbb{V}_0}(y) + \sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu \checkmark$$

Ad (2):

$$SS_1 = ||y - P_{\mathbb{V}_Z}(y)||^2 \overset{(1)}{=} ||(y - P_{\mathbb{V}_0}(y)) - \sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu||^2$$

$$= ||y - P_{\mathbb{V}_0}(y)||^2 - 2(y - P_{\mathbb{V}_0}(y))'(\sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu) + ||\sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu||^2$$

$$= ss_0 + ss_0\frac{(\tilde{Z}_\nu' b_{m_0+1})^2}{||\tilde{Z}_\nu||^2} - 2(y - P_{\mathbb{V}_0}(y))'(\sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu)$$

For the mixed term, we get

$$2(y - P_{\mathbb{V}_0}(y))'(\sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu) = 2\sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}(y - P_{\mathbb{V}_0}(y))'\tilde{Z}_\nu$$

$$= 2\sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}y'\tilde{Z}_\nu \overset{(*)}{=} 2\sqrt{ss_0}\frac{\tilde{Z}_\nu' b_{m_0+1}}{||\tilde{Z}_\nu||^2}\sqrt{ss_0}\tilde{Z}_\nu' b_{m_0+1} = 2ss_0\frac{(\tilde{Z}_\nu' b_{m_0+1})^2}{||\tilde{Z}_\nu||^2}$$

and therefore in total

$$SS_1 = ss_0 - ss_0\frac{(\tilde{Z}_\nu' b_{m_0+1})^2}{||\tilde{Z}_\nu||^2}. \checkmark$$

Case $m_1 - m_0 = k > 1$:
Applying the above case inductively we get (as shown below)

$$\frac{SS_1}{ss_0} \stackrel{d}{=} \prod_{l=1}^{k} U_l \tag{§}$$

where U_l are independent with $U_l \sim Beta(\frac{n-m_0-l}{2}, \frac{1}{2})$. If we can show $\prod_{l=1}^{k} U_l \sim Beta(\frac{n-m_1}{2}, \frac{m_1-m_0}{2})$ then (i) and (ii) hold.

We use the following result [Jambunathan, 1954]:
For U, V independent random variables:

$$U \sim Beta(a, b) \text{ and } V \sim Beta(a+b, c) \Rightarrow UV \sim Beta(a, b+c). \tag{**}$$

Induction: Let $V_i := \prod_{l=1}^{i} U_l$. By definition $V_1 = U_1 \sim Beta(\frac{n-m_0-1}{2}, \frac{1}{2})$. ✓

Now assume for arbitrary $1 \leq i < k$, that $V_i \sim Beta(\frac{n-m_0-i}{2}, \frac{i}{2})$ holds.
Using (**) with $a = \frac{n-m_0-(i+1)}{2}, b = \frac{1}{2}, c = \frac{i}{2}$ we have $U_{i+1} \sim Beta(a, b)$, $V_i \sim Beta(a+b, c)$ and

$$V_{i+1} = V_i U_{i+1} \sim Beta(a, b+c) = Beta\left(\frac{n-m_0-(i+1)}{2}, \frac{i+1}{2}\right) \checkmark$$

in particular

$$V_k = \prod_{l=1}^{k} U_l \sim Beta\left(\frac{n-m_0-k}{2}, \frac{k}{2}\right) = Beta\left(\frac{n-m_1}{2}, \frac{m_1-m_0}{2}\right)$$

so (i) and (ii) do indeed hold.

Ad. (§):
Induction: Assume for $k \in \{1, ..., q - m_0 - 1\}$ we have

$$\frac{SS_1}{ss_0} \stackrel{d}{=} \prod_{l=1}^{k} U_l$$

with $U_l \sim Beta(\frac{n-m_0-l}{2}, \frac{1}{2})$ independent, where $SS_1 = ||y - P_{\mathbb{V}_Z}(y)||^2$ with $\mathbb{V}_Z = span(x_1, ..., x_{m_0}, Z_{m_0+1}, ..., Z_{m_0+k})$ where Z_j are iid $\mathcal{N}(0, I_n)$-vectors. We need to show:

$$\frac{SS_2}{ss_0} \stackrel{d}{=} \prod_{l=1}^{k+1} U_l = \frac{SS_1}{ss_0} \cdot U_{k+1}$$

where $SS_2 = ||y - P_{\tilde{\mathbb{V}}_Z}(y)||^2$
with $\tilde{\mathbb{V}}_Z = span(x_1, ..., x_{m_0}, Z_{m_0+1}, ..., Z_{m_0+k}, Z_\nu)$ with Z_ν being an iid vector of n $\mathcal{N}(0, 1)$-random variables.

Note that it suffices to show that we can write $\frac{SS_2}{ss_0} = \frac{SS_1}{ss_0} \cdot U_{k+1}$ such that for any $z_{m_0+i} \in \mathbb{R}^n, i = 1, ..., k$ with $x_1, ..., x_{m_0}, z_{m_0+1}, ..., z_{m_0+k}$ linearly independent

$$U_{k+1}|Z_{m_0+1} = z_{m_0+1}, ..., Z_{m_0+k} = z_{m_0+k} \sim Beta(\frac{n - m_0 - k - 1}{2}, \frac{1}{2}).$$

This is because $P(x_1, ..., x_{m_0}, Z_{m_0+1}, ..., Z_{m_0+k}$ are linearly independent$) = 1$ and therefore the above statement implies that U_{k+1} is independent of $\frac{SS_1}{ss_0}$ and $U_{k+1} \sim Beta(\frac{n-m_0-k-1}{2}, \frac{1}{2})$.
So in the following we let $\mathbb{V}_Z = span(x_1, ..., x_{m_0}, z_{m_0+1}, ..., z_{m_0+k})$ and $\tilde{\mathbb{V}}_Z = span(x_1, ..., x_{m_0}, z_{m_0+1}, ..., z_{m_0+k}, Z_\nu)$ for some fixed $z_{m_0+i} \in \mathbb{R}^n, i = 1, ..., k$ as described above.

Now, let $b_1, ..., b_n$ be an orthonormal-basis of $\mathbb{R}^n$ such that

$$\mathbb{V}_Z = span(b_1, ..., b_{m_0+k})$$

and

$$b_{m_0+k+1} = \frac{y - P_{\mathbb{V}_Z}(y)}{\sqrt{SS_1}}.$$

We see that $Z_\nu' b_j \overset{iid}{\sim} \mathcal{N}(0,1)$ and

$$\tilde{\mathbb{V}}_Z = span(b_1, ..., b_{m_0+k}, \tilde{Z}_\nu)$$

where $\tilde{Z}_\nu = \sum_{j=m_0+k+1}^{n} Z_\nu' b_j \cdot b_j$.

Notice that as before we have

$$P_{\tilde{\mathbb{V}}_Z}(y) = P_{\mathbb{V}_Z}(y) + \frac{y'\tilde{Z}_\nu}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu = P_{\mathbb{V}_Z}(y) + \sqrt{SS_1}\frac{\tilde{Z}_\nu' b_{m_0+k+1}}{||\tilde{Z}_\nu||^2}\tilde{Z}_\nu$$

and therefore analogously to (2)

$$SS_2 = SS_1 - SS_1 \frac{(\tilde{Z}_\nu' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2}$$

$$\frac{SS_2}{ss_0} = \frac{SS_1}{ss_0} - \frac{SS_1}{ss_0}\frac{(\tilde{Z}_\nu' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2} = \frac{SS_1}{ss_0}(1 - \frac{(\tilde{Z}_\nu' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2})$$

so if we can show

$$U_{k+1} := 1 - \frac{(\tilde{Z}_\nu' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2} \sim Beta(\frac{n - m_0 - k - 1}{2}, \frac{1}{2})$$

then we are done.

$$1 - \frac{(\tilde{Z}_\nu' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2} = 1 - \frac{((\sum_{j=m_0+k+1}^{n} Z_\nu' b_j \cdot b_j)' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2}$$

$$= \frac{||\tilde{Z}_\nu||^2}{||\tilde{Z}_\nu||^2} - \frac{(Z_\nu' b_{m_0+k+1} \cdot b_{m_0+k+1}' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2}$$

$$= \frac{||\tilde{Z}_\nu||^2}{||\tilde{Z}_\nu||^2} - \frac{(Z_\nu' b_{m_0+k+1})^2}{||\tilde{Z}_\nu||^2} = \frac{\sum_{j=m_0+k+1}^{n}(Z_\nu' b_j)^2 - (Z_\nu' b_{m_0+k+1})^2}{\sum_{j=m_0+k+1}^{n}(Z_\nu' b_j)^2}$$

$$= \frac{\mathcal{X}^2_{n-m_0-k-1}}{\mathcal{X}^2_{n-m_0-k-1} + \mathcal{X}^2_1} \sim Beta\left(\frac{n - m_0 - k - 1}{2}, \frac{1}{2}\right). \checkmark$$

Ad. (iii):
First, note that for independent $\mathcal{X}_a^2$ and $\mathcal{X}_b^2$ we have

$$\frac{\mathcal{X}_a^2/a}{\mathcal{X}_b^2/b} \sim F(a,b)$$

$$\frac{\mathcal{X}_a^2}{\mathcal{X}_a^2+\mathcal{X}_b^2} \sim Beta\left(\frac{a}{2},\frac{b}{2}\right).$$

Now if $x = \frac{\mathcal{X}_a^2/a}{\mathcal{X}_b^2/b}$, then $x\frac{a}{b} = \frac{\mathcal{X}_a^2}{\mathcal{X}_b^2}$ and

$$\frac{x\frac{a}{b}}{x\frac{a}{b}+1} = \frac{\frac{\mathcal{X}_a^2}{\mathcal{X}_b^2}}{\frac{\mathcal{X}_a^2}{\mathcal{X}_b^2}+1} = \frac{\frac{\mathcal{X}_a^2}{\mathcal{X}_b^2}}{\frac{1}{\mathcal{X}_b^2}(\mathcal{X}_a^2+\mathcal{X}_b^2)} = \frac{\mathcal{X}_a^2}{\mathcal{X}_a^2+\mathcal{X}_b^2}.$$

I.e. for all $t \geq 0$ we have

$$F_{F(a,b)}(t) = F_{Beta(\frac{a}{2},\frac{b}{2})}\left(\frac{t\frac{a}{b}}{t\frac{a}{b}+1}\right).$$

We now simply apply this with

$$t = \frac{ss_0-ss_1}{m_1-m_0}/\frac{ss_1}{n-m_1}, a = m_1-m_0, b = n-m_1$$

and we obtain

$$F_{F(m_1-m_0,n-m_1)}\left(\frac{ss_0-ss_1}{m_1-m_0}/\frac{ss_1}{n-m_1}\right)$$

$$= F_{Beta(\frac{m_1-m_0}{2},\frac{n-m_1}{2})}\left(\frac{ss_0-ss_1}{ss_1}/(\frac{ss_0-ss_1}{ss_1}+1)\right)$$

$$= F_{Beta(\frac{m_1-m_0}{2},\frac{n-m_1}{2})}\left(\frac{ss_0-ss_1}{ss_1}/(\frac{1}{ss_1}(ss_0-ss_1+ss_1))\right)$$

$$= F_{Beta(\frac{m_1-m_0}{2},\frac{n-m_1}{2})}\left(\frac{ss_0-ss_1}{ss_0}\right) = F_{Beta(\frac{n-m_1}{2},\frac{m_1-m_0}{2})}\left(1-\frac{ss_0-ss_1}{ss_0}\right)$$

$$= F_{Beta(\frac{n-m_1}{2},\frac{m_1-m_0}{2})}\left(\frac{ss_0}{ss_0}-\frac{ss_0-ss_1}{ss_0}\right) = F_{Beta(\frac{n-m_1}{2},\frac{m_1-m_0}{2})}\left(\frac{ss_1}{ss_0}\right).$$

So (iii) does indeed hold.

$\square$

3.3 Lemma 1

Let a_n be a real-valued sequence converging to $a \in \mathbb{R} \cup \{-\infty, +\infty\}$ such that $\lim\limits_{n\to\infty} \frac{a_n}{n} > -2$. Then

$$\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$$

where we define $e^{-\infty} := 0$ and $e^{+\infty} := +\infty$.

If $\lim\limits_{n\to\infty} \frac{a_n}{n} = -2$, then either $\lim\limits_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = 0$ or the expression has no limit.

If a_n has no limit or $\lim\limits_{n\to\infty} \frac{a_n}{n} < -2$, then $\left(1 + \frac{a_n}{n}\right)^n$ has no limit.

Note that in the case $\lim\limits_{n\to\infty} \frac{a_n}{n} = -2$ one could try to further find sufficient conditions for convergence or divergence, but they are not required in the scope of this thesis.

3.4 Proof of Lemma 1

Case $a \in \mathbb{R}$:

Case $(a_n)_{n\in\mathbb{N}}$ does not contain a subsequence of all-zeroes:
It suffices to show that for the subsequence $(n)_{n\geq N}$ where $N \in \mathbb{N}$ such that $0 < |\frac{a_n}{n}| < 1 \; \forall n \geq N$ it holds, that

$$\log\left(\left(1 + \frac{a_n}{n}\right)^n\right) \xrightarrow[n\to\infty]{} a.$$

For $n \geq N$ we have

$$\log\left(\left(1 + \frac{a_n}{n}\right)^n\right) = n \log\left(1 + \frac{a_n}{n}\right) = a_n \frac{\log\left(1 + \frac{a_n}{n}\right)}{\frac{a_n}{n}}$$

and therefore

$$\lim_{n\to\infty} \log\left(\left(1 + \frac{a_n}{n}\right)^n\right) = \lim_{n\to\infty} a_n \frac{\log\left(1 + \frac{a_n}{n}\right)}{\frac{a_n}{n}} = a \lim_{x\to 0} \frac{\log(1+x)}{x} = a. \checkmark$$

Case $(a_n)_{n\in\mathbb{N}}$ does contain a subsequence of all-zeroes:

Here we have $a = 0$. Take an arbitrary subsequence n' of n. If $(a_{n'})$ is a subsequence containing a subsequence of all-zeroes, take n'' as a subsequence of n' such that $(a_{n''})$ is an all-zero subsequence. Then trivially $\lim\limits_{n''\to\infty} \left(1 + \frac{a_{n''}}{n''}\right)^{n''} = \lim\limits_{n''\to\infty} 1^{n''} = 1 = e^0 = e^a$. If $(a_{n'})$ is a subsequence containing finitely many zeroes, take n'' equal to n' and apply the first case to see $\lim\limits_{n''\to\infty} \left(1 + \frac{a_{n''}}{n''}\right)^{n''} = e^a$. ✓

Case $a = \infty$:

Let $b > 0$ and $b_n := \min(a_n, b) \xrightarrow[n\to\infty]{} b$.

Note that there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ we have
$a_n \geq b = b_n > 0$ and therefore

$$(1 + \frac{a_n}{n})^n \geq (1 + \frac{b_n}{n})^n.$$

Using the result from the case $a \in \mathbb{R}$ on b_n, it follows that

$$\liminf_{n\to\infty} (1 + \frac{a_n}{n})^n \geq \lim_{n\to\infty} (1 + \frac{b_n}{n})^n = e^b.$$

Since we can choose b arbitrarily large, this means

$$(1 + \frac{a_n}{n})^n \xrightarrow[n\to\infty]{} +\infty = e^{+\infty} \checkmark$$

Case $a = -\infty$:

Case $\lim\limits_{n\to\infty} \frac{a_n}{n} > -1$:

Let $b < 0$ and b_n a real-valued sequence such that $b_n \xrightarrow[n\to\infty]{} b$. Since $a_n \xrightarrow[n\to\infty]{} -\infty$ and by the additional assumption, we have $-1 < \lim\limits_{n\to\infty} \frac{a_n}{n} \leq 0$, and therefore there exists an $N \in \mathbb{N}$ such that $a_n \leq b_n < 0$ and $-1 \leq \frac{a_n}{n} \leq \frac{b_n}{n} < 0$ for all $n \geq N$ and therefore

$$(1 + \frac{a_n}{n})^n \leq (1 + \frac{b_n}{n})^n.$$

Using the first case, it follows that

$$\limsup_{n\to\infty} (1 + \frac{a_n}{n})^n \leq \lim_{n\to\infty} (1 + \frac{b_n}{n})^n = e^b.$$

Since we can choose b arbitrarily small, this means

$$(1 + \frac{a_n}{n})^n \xrightarrow[n\to\infty]{} 0 = e^{-\infty} \checkmark$$

Case $-2 < \lim\limits_{n\to\infty} \frac{a_n}{n} \leq -1$:

Note, that if $|(1+\frac{a_n}{n})^n| \xrightarrow[n\to\infty]{} 0$ then also $(1+\frac{a_n}{n})^n \xrightarrow[n\to\infty]{} 0$.

$$|(1+\frac{a_n}{n})^n| = |1+\frac{a_n}{n}|^n = (1+\frac{b_n}{n})^n$$

for the real-valued sequence b_n defined by $\frac{b_n}{n}+1 = |\frac{a_n}{n}+1|$. Note, that since $\lim\limits_{n\to\infty} |\frac{a_n}{n}+1|-1 = |\lim\limits_{n\to\infty} \frac{a_n}{n}+1|-1 < 0$ we have

$$b_n = n\left(|\frac{a_n}{n}+1|-1\right) \xrightarrow[n\to\infty]{} -\infty$$

and, since for all $n \geq N$ for some $N \in \mathbb{N}$ we have $-2 < \frac{a_n}{n} \leq 0$, there's also

$$-1 < \frac{b_n}{n} = |\frac{a_n}{n}+1|-1 \leq 0$$

for all $n \geq N$. Analogously to the above case we then get

$$(1+\frac{b_n}{n})^n \xrightarrow[n\to\infty]{} 0 \checkmark$$

Case $\lim\limits_{n\to\infty} \frac{a_n}{n} = -2$:

Note, that it might still be possible to have $(1+\frac{a_n}{n})^n \xrightarrow[n\to\infty]{} 0$ in this case, depending on the speed and direction of convergence of $1+\frac{a_n}{n}$ to -1. But if this is not the case, then $(1+\frac{a_n}{n})^n$ has no limit:

Note, that for $n \geq N$ for some $N \in \mathbb{N}$ we have $(1+\frac{a_{2n}}{2n})^{2n} \geq 0$ and $(1+\frac{a_{2n-1}}{2n-1})^{2n-1} \leq 0$ and therefore

$$\limsup_{n\to\infty}(1+\frac{a_n}{n})^n \geq 0 \text{ and } \liminf_{n\to\infty}(1+\frac{a_n}{n})^n \leq 0$$

and if $(1+\frac{a_n}{n})^n$ does not converge to 0, then we must have either

$$\limsup_{n\to\infty}(1+\frac{a_n}{n})^n > 0 \text{ or } \liminf_{n\to\infty}(1+\frac{a_n}{n})^n < 0$$

or both. Therefore we have

$$\liminf_{n\to\infty}(1+\frac{a_n}{n})^n < \limsup_{n\to\infty}(1+\frac{a_n}{n})^n$$

and the expression does not have a limit. ✓

Case a_n has no limit or $\lim\limits_{n\to\infty} \frac{a_n}{n} < -2$:

If $\lim\limits_{n\to\infty} \frac{a_n}{n} < -2$, then for some $\delta > 0$ and some $N \in \mathbb{N}$ we have for all $n \geq N$

$$(1 + \frac{a_{2n}}{2n})^{2n} \geq (-1 - \delta)^{2n}$$

and

$$(1 + \frac{a_{2n-1}}{2n-1})^{2n-1} \leq (-1 - \delta)^{2n-1}$$

which means, that

$$(1 + \frac{a_{2n}}{2n})^{2n} \xrightarrow[n\to\infty]{} \infty \text{ but } (1 + \frac{a_{2n-1}}{2n-1})^{2n-1} \xrightarrow[n\to\infty]{} -\infty$$

and therefore $(1 + \frac{a_n}{n})^n$ cannot have a limit. ✓

If a_n does not have a limit then

$$\liminf_{n\to\infty} a_n = i < s = \limsup_{n\to\infty} a_n.$$

Let n' be a subsequence, such that $\lim\limits_{n\to\infty} a_{n'} = i$ and n'' be a subsequence, such that $\lim\limits_{n\to\infty} a_{n''} = s$. Then either

$$\lim_{n\to\infty} (1 + \frac{a_{n'}}{n'})^{n'} = e^i < e^s = \lim_{n\to\infty} (1 + \frac{a_{n''}}{n''})^{n''}$$

or the limit on the left does not exist (we showed this to be possible if $i = -\infty$) and therefore $(1 + \frac{a_n}{n})^n$ cannot have a limit in n. ✓

$\square$

3.5 Lemma 2

Let q_n be a strictly monotonically increasing sequence of natural numbers. Let b_n a real-valued sequence such that

$$q_n b_n \xrightarrow[n\to\infty]{} a \in \mathbb{R} \cup \{-\infty, +\infty\}$$

and such that $\lim\limits_{n\to\infty} b_n > -2$. Then

$$(1+b_n)^{q_n} \xrightarrow[n\to\infty]{} e^a.$$

If $\lim\limits_{n\to\infty} b_n \leq -2$ or $\lim\limits_{n\to\infty} q_n b_n$ does not exist, then $\lim\limits_{n\to\infty} (1+b_n)^{q_n} \in \{-\infty, 0, +\infty\}$ or the limit does not exist.

Again, sufficient conditions for different results in the $\lim\limits_{n\to\infty} b_n \leq -2$ case could be found, but are omitted since they are not required in the scope of this thesis. Note that in the following proof a sequence of natural numbers $(x_n)_{n\in\mathbb{N}}$ is called 'even' if $x_n \mod 2 = 0$ for all $n \in \mathbb{N}$ and is called 'odd' if $x_n \mod 2 = 1$ for all $n \in \mathbb{N}$.

3.6 Proof of Lemma 2

If $q_n b_n$ has a limit a and $\lim\limits_{n\to\infty} b_n > -2$:

First, define $c_n := q_n b_n \xrightarrow[n\to\infty]{} a$ and then construct the sequence a_n such that $a_{q_n} = c_n$ and $a_n = d_n$ for all n not contained in the sequence q_n, where d_n is an arbitrary sequence with $d_n \xrightarrow[n\to\infty]{} a$ and such that $\lim\limits_{n\to\infty} \frac{d_n}{n} > -2$.

We see that $a_n \xrightarrow[n\to\infty]{} a$ and $\lim\limits_{n\to\infty} \frac{a_n}{n} > -2$ and that $(c_n)_{n\in\mathbb{N}}$ is a subsequence of $(a_n)_{n\in\mathbb{N}}$ in such a way, that

$$((1+b_n)^{q_n})_{n\in\mathbb{N}} = \left(\left(1+\frac{c_n}{q_n}\right)^{q_n}\right)_{n\in\mathbb{N}} = \left(\left(1+\frac{a_{q_n}}{q_n}\right)^{q_n}\right)_{n\in\mathbb{N}}$$

is a subsequence of $\left(\left(1+\frac{a_n}{n}\right)^n\right)_{n\in\mathbb{N}}$ and therefore has the same limit (if the latter sequence has one). Using this, we get

$$\lim_{n\to\infty} (1+b_n)^{q_n} = \lim_{n\to\infty} \left(1+\frac{a_n}{n}\right)^n = e^a$$

where the last equality follows from Lemma 1. ✓

If $\lim_{n\to\infty} b_n = -2$:

If q_n contains even and odd subsequences:

With an analogous argument to the $\lim_{n\to\infty} \frac{a_n}{n} = -2$ case in the proof of Lemma 1, we see that in this case either $\lim_{n\to\infty}(1+b_n)^{q_n} = 0$ or there is no limit.

If q_n contains no even or no odd subsequences:

Note, that we have $1 + b_n = -1 + c_n$ for some c_n with $c_n \xrightarrow[n\to\infty]{} 0$. In particular, $c_n = b_n + 2$. Note further, that $q_n c_n \xrightarrow[n\to\infty]{} -\infty$. If q_n contains no odd subsequences, then

$$\lim_{n\to\infty}(1+b_n)^{q_n} = \lim_{n\to\infty}|1+b_n|^{q_n} = \lim_{n\to\infty}|-1+c_n|^{q_n} = \lim_{n\to\infty}(1-c_n)^{q_n} = e^{+\infty} = +\infty$$

and if q_n contains no even subsequences, then

$$\lim_{n\to\infty}(1+b_n)^{q_n} = -\lim_{n\to\infty}|1+b_n|^{q_n} = -\lim_{n\to\infty}(1-c_n)^{q_n} = -e^{+\infty} = -\infty \checkmark$$

Other cases:

If $\lim_{n\to\infty} b_n < -2$, then for some $\delta > 0$ and some $N \in \mathbb{N}$ we have for all $n \geq N$ such that q_n is even

$$(1+b_n)^{q_n} \geq (-1-\delta)^{q_n}$$

and for all $n \geq N$ such that q_n is not even

$$(1+b_n)^{q_n} \leq (-1-\delta)^{q_n}.$$

This means that if there exists an even subsequence $q_{n'}$ of q_n, then $\lim_{n\to\infty}(1+b_{n'})^{q_{n'}} = +\infty$ and if there exists an odd subsequence $q_{n''}$ of q_n, then $\lim_{n\to\infty}(1+b_{n''})^{q_{n''}} = -\infty$. So if both exist, the sequence $(1+b_n)^{q_n}$ has no limit. If no odd subsequences exist, then $\lim_{n\to\infty}(1+b_n)^{q_n} = +\infty$ and if no even subsequences exist, then $\lim_{n\to\infty}(1+b_n)^{q_n} = -\infty$. ✓

It remains to show that if $q_n b_n$ has no limit, then $(1+b_n)^{q_n}$ doesn't either. In this case, we have $\liminf_{n\to\infty} q_n b_n = i < s = \limsup_{n\to\infty} q_n b_n$. Let n' be a subsequence of n such that $q_{n'} b_{n'} \xrightarrow[n'\to\infty]{} i$ and let n'' be a subsequence of n such that $q_{n''} b_{n''} \xrightarrow[n''\to\infty]{} s$. Then we either have

$$\lim_{n\to\infty}(1+b_{n'})^{q_{n'}} = e^i < e^s = \lim_{n\to\infty}(1+b_{n''})^{q_{n''}}$$

or the limit on the left does not exist, and therefore $(1+b_n)^{q_n}$ cannot have a limit. ✓

$\square$

3.7 Hypothesis 2

Let $\alpha \in (0,1), m_0 \in \mathbb{N}$, q_n a strictly monotonically increasing sequence of natural numbers. Let $x_n(\alpha)$ be defined by

$$\alpha = 1 - F_{Beta(\frac{1}{2}, \frac{n-m_0-1}{2})}(x_n(\alpha))^{q_n - m_0}.$$

Then

$$x_n(\alpha) \asymp \frac{2\log q_n - \log\log q_n - 2\log(-\log(1-\alpha)) - \log\pi}{n}$$

where for two sequences a_n and b_n the relationship $a_n \asymp b_n$, verbally 'a_n and b_n are asymptotically equal' is defined as

$$\frac{a_n}{b_n} \xrightarrow[n\to\infty]{} 1 \text{ or analogously } a_n = b_n(1+o(1)).$$

This hypothesis is based on Theorem 4.1 of the 2019 draft version of the original paper [Davies & Dümbgen, 2019, pg. 7, 25f] and the proof follows the same steps with more detailed argumentation. The reason as to why it is not called a theorem but a hypothesis in this thesis is that following the same chain of argument as the proof in the original paper requires additional results that are neither standard arguments nor are argued in the original proof. Specifically, due to q_n not being a constant but growing in n, the asymptotic equality (4) in the following proof does not trivially hold but requires results about the speed of convergence of the two previous asymptotic equations. We presume these to hold in the following proof. These results are:

For a real-valued sequence x_n such that $nx_n \xrightarrow[n\to\infty]{} +\infty$ it holds, that

$$F_{Beta(\frac{1}{2}, \frac{n-1}{2})}(x_n) = F_{\chi_1^2}(nx_n)(1+a_n) \tag{A1}$$

where $a_n = o(\exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n))$ and

$$F_{\chi_1^2}(nx_n) = \left(1 - \exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi})\right)(1+b_n) \tag{A2}$$

where $b_n = o(\exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n))$.

Note, that the similar but weaker results

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n) \asymp F_{\mathcal{X}_1^2}(nx_n)$$

and

$$F_{\mathcal{X}_1^2}(nx_n) \asymp 1 - exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi})$$

are stated in the original proof and will be derived in the following proof, but do not suffice to complete the proof with the present line of argument.

3.8 Proof of Hypothesis 2 (under A1 and A2)

Let $(Z_i)_{i\in\mathbb{N}}$ be iid $\mathcal{N}(0,1)$-distributed random variables. For any $n \in \mathbb{N}$, $n \geq 2$ we have

$$\frac{Z_1^2}{\sum_{i=1}^n Z_i^2} = \frac{Z_1^2}{Z_1^2 + \sum_{i=2}^n Z_i^2} = \frac{\mathcal{X}_1^2}{\mathcal{X}_1^2 + \mathcal{X}_{n-1}^2} \sim Beta(\frac{1}{2}, \frac{n-1}{2}).$$

Using this, we can show that we have

$$\frac{\mathcal{X}_1^2}{n} \approx Beta(\frac{1}{2}, \frac{n-1}{2}) \tag{1}$$

meaning $nBeta(\frac{1}{2}, \frac{n-1}{2}) \xrightarrow{d} \mathcal{X}_1^2$ and therefore also

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{t}{n}) \xrightarrow[n\to\infty]{} F_{\mathcal{X}_1^2}(t) \ \forall t > 0.$$

ad. (1):
It suffices to show:

$$n\frac{Z_1^2}{\sum_{i=1}^n Z_i^2} \xrightarrow{d} \mathcal{X}_1^2.$$

And this holds, because

$$n\frac{Z_1^2}{\sum_{i=1}^n Z_i^2} = \frac{Z_1^2}{\frac{1}{n}\sum_{i=1}^n Z_i^2} \xrightarrow{d} Z_1^2 \sim \mathcal{X}_1^2$$

since

$$\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \xrightarrow{a.s.} 1$$

which follows from $E[Z_1^2] = 1$ and the law of large numbers. ✓

Using (1), we can show that

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n) \asymp F_{\mathcal{X}_1^2}(nx_n) \tag{2}$$

for any real-valued sequence $x_n, n \in \mathbb{N}$ such that $nx_n \xrightarrow[n\to\infty]{} +\infty$.

<u>ad. (2):</u>
We need to show

$$\frac{F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n)}{F_{\mathcal{X}_1^2}(nx_n)} \xrightarrow[n\to\infty]{} 1.$$

For this it suffices to show that both the nominator and the denominator converge to 1.

$$F_{\mathcal{X}_1^2}(nx_n) \xrightarrow[n\to\infty]{} 1$$

since $F_{\mathcal{X}_1^2}$ is a cdf and $nx_n \xrightarrow[n\to\infty]{} \infty$ by assumption. ✓

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n) \xrightarrow[n\to\infty]{} 1$$

because: First note that

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n) = F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{nx_n}{n}).$$

Let $k > 0$. Since $nx_n \xrightarrow[n\to\infty]{} \infty$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $k < nx_n$. Therefore for all $n \geq N$

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{nx_n}{n}) > F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{k}{n}) \xrightarrow[n\to\infty]{(1)} F_{\mathcal{X}_1^2}(k)$$

and therefore

$$\liminf_{n\to\infty} F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{nx_n}{n}) \geq F_{\mathcal{X}_1^2}(k) \xrightarrow[k\to\infty]{} 1$$

and because we can choose k arbitrarily large

$$1 \geq \limsup_{n\to\infty} F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{nx_n}{n}) \geq \liminf_{n\to\infty} F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{nx_n}{n}) \geq 1$$

meaning

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n) = F_{Beta(\frac{1}{2},\frac{n-1}{2})}(\frac{nx_n}{n}) \xrightarrow[n\to\infty]{} 1. \checkmark$$

And with that, for any real-valued sequence $x_n, n \in \mathbb{N}$ such that $nx_n \xrightarrow[n\to\infty]{} \infty$

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n) \overset{(2)}{\asymp} F_{\mathcal{X}_1^2}(nx_n) \overset{(3)}{\asymp} 1 - \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{nx_n}}exp(-\frac{nx_n}{2})$$

$$= 1 - (\frac{2}{\pi})^{\frac{1}{2}}(nx_n)^{-\frac{1}{2}}exp(-\frac{nx_n}{2}) = 1 - exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi}).$$

ad. (3):
We want to show

$$F_{\mathcal{X}_1^2}(nx_n) \asymp 1 - \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{nx_n}}\exp(-\frac{nx_n}{2}).$$

This is equivalent to

$$F_{\mathcal{X}_1^2}(nx_n) = \left(1 - \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{nx_n}}\exp(-\frac{nx_n}{2})\right)(1+o(1)).$$

First, note that Lemma 2 from [Feller, 1957a, pg. 166ff] gives us

$$1 - \Phi(t) = \frac{\phi(t)}{t}(1 + o(\frac{1}{t})) \tag{*}$$

where $\Phi(t) = F_{\mathcal{N}(0,1)}(t)$ and $\phi(t) = f_{\mathcal{N}(0,1)}(t)$.

And further note that for $Z \sim \mathcal{N}(0,1)$

$$F_{\mathcal{X}_1^2}(nx_n) = P(Z^2 < nx_n) = P(|Z| < \sqrt{nx_n}) = 2\left(\Phi(\sqrt{nx_n}) - \frac{1}{2}\right).$$

Using the result of the above Lemma, we get (with some $a_n = o(\frac{1}{\sqrt{nx_n}})$)

$$2\left(\Phi(\sqrt{nx_n}) - \frac{1}{2}\right) = -2\left(\frac{1}{2} - \Phi(\sqrt{nx_n})\right) = -2\left(1 - \Phi(\sqrt{nx_n}) - \frac{1}{2}\right)$$

$$\overset{(*)}{=} -2\left(\frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}(1+a_n) - \frac{1}{2}\right) = 2\left(\frac{1}{2} - \frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}(1+a_n)\right)$$

$$= 2\left(\frac{1}{2} + \frac{1}{2}a_n - \frac{1}{2}a_n - \frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}(1+a_n)\right)$$

$$= 2\left(\frac{1}{2}(1+a_n) - \frac{1}{2}a_n - \frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}(1+a_n)\right)$$

$$= 2\left(\frac{1}{2} - \frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}\right)(1+a_n) - a_n = \left(1 - 2\frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}\right)(1+a_n) - a_n$$

$$= \left(1 - 2\frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}\right)\left(1 + a_n - \frac{a_n}{1 - 2\frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}}\right)$$

$$= \left(1 - 2\frac{\phi(\sqrt{nx_n})}{\sqrt{nx_n}}\right)\left(1 + o(\frac{1}{\sqrt{nx_n}}) - \frac{o(\frac{1}{\sqrt{nx_n}})}{1 - 2o(1)}\right)$$

$$= \left(1 - \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{nx_n}}\exp(-\frac{nx_n}{2})\right)(1 + o(\frac{1}{\sqrt{nx_n}}))$$

$$= \left(1 - \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{nx_n}}\exp(-\frac{nx_n}{2})\right)(1 + o(1)) \checkmark$$

Using this result (or rather the stronger assumed (A2) and (A1)) we can show, for real-valued sequence x_n with $nx_n \xrightarrow[n\to\infty]{} \infty$ and such that $q_n \exp\left(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n\right) \xrightarrow[n\to\infty]{} c$ for some $c \in \mathbb{R}$

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n)^{q_n} \overset{(4)}{\asymp} (1 - \exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi}))^{q_n}$$

$$\overset{(5)}{\asymp} \exp(-q_n \exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi})).$$

ad. (4):
From (2) and (3) we get

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n)^{q_n} = (1 - \exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi}))^{q_n}(1+c_n)^{q_n}$$

for some $c_n = o(1)$. In order to have

$$(1+c_n)^{q_n} = (1 + o(1))$$

and therefore the desired asymptotic equality, by Lemma 2 we need to show that $q_n c_n \xrightarrow[n\to\infty]{} 0$, i.e. $c_n = o(\frac{1}{q_n})$.

First note, that we need $c_n = o(\exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n))$, because then

$$q_n c_n = q_n \exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n)\frac{c_n}{\exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n)} \xrightarrow[n\to\infty]{} c \cdot 0 = 0.$$

Note further, that from the calculations for (2) and (3) we see, that for some $a_n = o(1)$, $b_n = o(\frac{1}{\sqrt{nx_n}})$

$$F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n)^{q_n} \overset{(2)}{=} F_{\mathcal{X}_1^2}(nx_n)^{q_n}(1+a_n)^{q_n}$$

$$\overset{(3)}{=} \left(1 - \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{nx_n}}\exp(-\frac{nx_n}{2})\right)^{q_n}(1+b_n)^{q_n}(1+a_n)^{q_n}$$

$$= \left(1 - \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{nx_n}}\exp(-\frac{nx_n}{2})\right)^{q_n}(1+a_n+b_n+a_nb_n)^{q_n}.$$

We will now use assumptions (A1) and (A2). With these, we have

$$a_n = o(\exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n)) \text{ and } b_n = o(\exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n))$$

and therefore

$$c_n = a_n + b_n + a_nb_n = o(\exp(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n)). \checkmark$$

ad. (5):
First define $d_n := \exp\left(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi}\right)$.
We want to show

$$\frac{(1-d_n)^{q_n}}{\exp(-q_nd_n)} \xrightarrow[n\to\infty]{} 1.$$

For this it suffices to show that for some constant $a \in \mathbb{R}$

$$q_nd_n \xrightarrow[n\to\infty]{} a$$

because then $\exp(-q_nd_n) \xrightarrow[n\to\infty]{} \exp(-a)$ and $(1-d_n)^{q_n} \xrightarrow[n\to\infty]{} \exp(-a)$ according to Lemma 2.

$$q_nd_n = q_n\exp\left(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n + \frac{1}{2}\log\frac{2}{\pi}\right)$$

$$= q_n\exp\left(-\frac{1}{2}nx_n - \frac{1}{2}\log nx_n\right)\exp\left(\frac{1}{2}\log\frac{2}{\pi}\right)$$

$$\xrightarrow[n\to\infty]{} c\exp\left(\frac{1}{2}\log\frac{2}{\pi}\right) \in \mathbb{R} \checkmark$$

Finally, our goal is to asymptotically solve the equation

$$1 - \alpha = F_{Beta(\frac{1}{2},\frac{n-1}{2})}(x_n(\alpha))^{q_n}.$$

In order for $\hat{x}_n(\alpha)$ to be an asymptotic solution to this equation, we need to set it such that it fulfils the prerequisites for (3), (4) and (5):

$$n\hat{x}_n(\alpha) \xrightarrow[n\to\infty]{} +\infty, \tag{R1}$$

$$q_n \exp\left(-\frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha)\right) \xrightarrow[n\to\infty]{} c \tag{R2}$$

for some $c \in \mathbb{R}$, and such that the following asymptotic equation holds:

$$1 - \alpha \asymp \exp\left(-q_n \exp\left(-\frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi}\right)\right). \tag{R3}$$

So once we show that the proposed solution

$$\hat{x}_n(\alpha) := \frac{2\log q_n - \log\log q_n - 2\log(-\log(1-\alpha)) - \log\pi}{n}$$

fulfils these three properties, then we are done.

Note, that

$$n\hat{x}_n(\alpha) = 2\log q_n - \log\log q_n - 2\log(-\log(1-\alpha)) - \log\pi \xrightarrow[n\to\infty]{} +\infty$$

so (R1) is fulfilled. ✓

Note further, that

$$\exp\left(\frac{1}{2}n\hat{x}_n(\alpha)+\frac{1}{2}\log n\hat{x}_n(\alpha)\right)$$

$$=\exp\left(\frac{1}{2}[2\log q_n-\log\log q_n-2\log(-\log(1-\alpha))-\log\pi]+\frac{1}{2}\log n\hat{x}_n(\alpha)\right)$$

$$=\exp\left(\log q_n-\frac{1}{2}\log\log q_n-\log(-\log(1-\alpha))-\frac{1}{2}\log\pi\right)\sqrt{n\hat{x}_n(\alpha)}$$

$$=\frac{-q_n}{\sqrt{\pi\log q_n}\log(1-\alpha)}\sqrt{n\hat{x}_n(\alpha)}$$

and therefore

$$q_n\exp\left(-\frac{1}{2}n\hat{x}_n(\alpha)-\frac{1}{2}\log n\hat{x}_n(\alpha)\right)=q_n/\left(\frac{-q_n}{\sqrt{\pi\log q_n}\log(1-\alpha)}\sqrt{n\hat{x}_n(\alpha)}\right)$$

$$=-\sqrt{\pi}\log(1-\alpha)\sqrt{\frac{\log q_n}{n\hat{x}_n(\alpha)}}\xrightarrow[n\to\infty]{}-\frac{\sqrt{\pi}\log(1-\alpha)}{\sqrt{2}}\in\mathbb{R}$$

because

$$\frac{\log q_n}{n\hat{x}_n(\alpha)}=\frac{\log q_n}{2\log q_n-\log\log q_n-2\log(-\log(1-\alpha))-\log\pi}$$

and using the fact that q_n is a subsequence of n and de L'Hospital's theorem:

$$\lim_{n\to\infty}\frac{\log q_n}{n\hat{x}_n(\alpha)}=\lim_{n\to\infty}\frac{\frac{d}{dn}\log n}{\frac{d}{dn}[2\log n-\log\log n-2\log(-\log(1-\alpha))-\log\pi]}$$

$$=\lim_{n\to\infty}\frac{\frac{1}{n}}{\frac{2}{n}-\frac{1}{\log n}\frac{1}{n}}=\lim_{n\to\infty}\frac{1}{2-\frac{1}{\log n}}=\frac{1}{2}$$

so (R2) is fulfilled as well. ✓

The following expressions are equivalent to (R3):

$$1-\alpha = \exp\left(-q_n \exp\left(-\frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi}\right)\right)(1+o(1))$$

$$\log(1-\alpha) = -q_n \exp\left(-\frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi}\right) + \log(1+o(1))$$

$$q_n \exp\left(-\frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi}\right) = -\log(1-\alpha) - \log(1+o(1))$$

$$\log\left(q_n \exp\left(-\frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi}\right)\right) = \log(-\log(1-\alpha) - o(1))$$

$$\log q_n - \frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi} = \log(-\log(1-\alpha)) + o(1).$$

And the given expression for $\hat{x}_n(\alpha)$ indeed fulfils this:

$$\begin{aligned}
&\log q_n - \frac{1}{2}n\hat{x}_n(\alpha) - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi}\\
&= \log q_n - [\log q_n - \frac{1}{2}\log\log q_n - \log(-\log(1-\alpha)) - \frac{1}{2}\log\pi] - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log\frac{2}{\pi}\\
&= \frac{1}{2}\log\log q_n + \log(-\log(1-\alpha)) + \frac{1}{2}\log\pi - \frac{1}{2}\log n\hat{x}_n(\alpha) + \frac{1}{2}\log 2 - \frac{1}{2}\log\pi\\
&= \log(-\log(1-\alpha)) + \frac{1}{2}[\log\log q_n - \log n\hat{x}_n(\alpha) + \log 2]\\
&= \log(-\log(1-\alpha)) + o(1)
\end{aligned}$$

because

$$\log\log q_n - \log n\hat{x}_n(\alpha) + \log 2 = \log[2\frac{\log q_n}{n\hat{x}_n(\alpha)}]$$

and as we already showed above

$$\frac{\log q_n}{n\hat{x}_n(\alpha)} \xrightarrow[n\to\infty]{} \frac{1}{2}$$

and therefore

$$\log[2\frac{\log q_n}{n\hat{x}_n(\alpha)}] \xrightarrow[n\to\infty]{} \log 1 = 0. \checkmark$$

So $\hat{x}_n(\alpha)$ is indeed an asymptotic solution to the problem.

$\square$

3.9 Theorem 3

The following holds, if Hypothesis 2 holds.
Let $k \in \mathbb{N}, \sigma \in \mathbb{R}^+$. Let q_n be a strictly monotonically increasing sequence of natural numbers with $k < q_n$. For any $n \geq k$: Let $x_j \in \mathbb{R}^n, j \in \{1, ..., q_n\}$ such that for any j with $k < j \leq q_n$ for the set $\{1, ..., k\} \cup \{j\}$, the corresponding x_i are linearly independent. Let ϵ be a vector of n iid $\mathcal{N}(0,1)$-distributed variables. Let $\beta \in \mathbb{R}^k$. Let

$$y = \sum_{i=1}^{k} \beta_j x_j + \sigma\epsilon,$$

called the 'true' model. Let $\mathbb{M} = \mathcal{P}(\{1, ..., k\})$, i.e. the set of all models which only contain 'true' regressors. For any $\mathcal{M} \in \mathbb{M}$ let $\mathcal{M}_\nu$ with $\nu \in \{1, ..., q_n\} \setminus \mathcal{M}$ denote the set $\mathcal{M} \cup \{\nu\}$ and let $P_\mathcal{M}$ denote the projection onto $span(x_i, i \in \mathcal{M})$. Furthermore, let

$$u = u(\mathcal{M}) := \sum_{j \in \{1,...,k\} \setminus \mathcal{M}} \beta_j x_j, \ \tilde{x}_\nu = \tilde{x}_\nu(\mathcal{M}) := \frac{x_\nu - P_\mathcal{M}(x_\nu)}{||x_\nu - P_\mathcal{M}(x_\nu)||}.$$

The following assumptions shall hold (where $m = m(\mathcal{M}) := |\mathcal{M}|$):

$$\lim_{n\to\infty} \min_{\mathcal{M} \in \mathbb{M} \setminus \{1,...,k\}} \frac{\max_{\nu \in \{1,...,k\} \setminus \mathcal{M}} |u'\tilde{x}_\nu|}{\sqrt{(k-m+3)\log q_n}} = \infty \tag{ASS. 1}$$

$$\lim_{n\to\infty} \min_{\mathcal{M} \in \mathbb{M} \setminus \{1,...,k\}} \frac{||u - P_\mathcal{M}(u)||}{\sqrt{(k-m+3)\log q_n}} = \infty \tag{ASS. 2}$$

$$\exists \tau > 2 : \exists N \in \mathbb{N} : \forall n \geq N : \min_{\mathcal{M} \in \mathbb{M}} \frac{\max_{\nu \in \{1,...,k\} \setminus \mathcal{M}} (u'\tilde{x}_\nu)^2}{||u - P_\mathcal{M}(u)||^2 + \sigma^2 n} > \frac{\tau \log q_n}{n} \tag{ASS. 3}$$

$$\exists \tau' < 2 : \exists N \in \mathbb{N} : \forall n \geq N : \max_{\mathcal{M} \in \mathbb{M}} \frac{\max_{\nu \in \{k+1,...,q\}} (u'\tilde{x}_\nu)^2}{||u - P_\mathcal{M}(u)||^2 + \sigma^2 n} < \frac{\tau' \log q_n}{n}. \tag{ASS. 4}$$

And further assume that for all $\nu \geq k+1$ (i.e. ν not in the 'true' model)

$$\frac{|u'\tilde{x}_\nu|\sqrt{\log q_n}}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} = o(1).$$

Then the Gaussian stepwise procedure is asymptotically consistent, meaning that for any $\mathcal{M} \in \mathbb{M} \setminus \{1, ..., k\}, \alpha \in (0,1)$

$$P\left(\exists \nu \in \{1, ..., k\} \setminus \mathcal{M} : 1 - F_{Beta}(1 - \frac{ss_\nu}{ss_0})^{q_n - m} < \alpha\right) \xrightarrow[n \to \infty]{} 1$$

and for any $\mathcal{M} \in \mathbb{M}, \alpha \in (0,1)$

$$P\left(\exists \nu \in \{k+1, ..., q_n\} : 1 - F_{Beta}(1 - \frac{ss_\nu}{ss_0})^{q_n - m} < \alpha\right) \xrightarrow[n \to \infty]{} 0$$

where $m := |\mathcal{M}|$, F_{Beta} is the cdf of a $Beta(\frac{1}{2}, \frac{n-m-1}{2})$-distribution, ss_0 is the RSS of the model $\mathcal{M}$ and ss_ν is the RSS of the model $\mathcal{M}_\nu$, as defined in Theorem 1.

This Theorem and its proof correspond to Theorem 4.2 of the original paper [Davies & Dümbgen, 2019, pg. 8, 26-29]. Note that the interpretation of the Theorem's statement is, that if the given assumptions hold then the probability of the single stepwise procedure having selected the 'true' model at termination tends to 1 for $n \to \infty$. This is because at each step, at least one of the 'true' regressors is added with probability tending to 1, while none of the 'wrong' regressors are added, with probability tending to 1 as well. This results in the probability of the model at any given step being a subset of the 'true' model tending to 1, while the probability of terminating at any step that doesn't include all regressors of the 'true' model tends to 0.
Note that in the very last argument of the proof the authors simply state that in the case of unboundedness of $|u'\tilde{x}_\nu|$ the term $C|u'\tilde{x}_\nu|\sqrt{\log q_n}$ can be ignored because of its lower rate of growth compared to $(u'\tilde{x}_\nu)^2$. This seems to overlook the dependency of q on n, which does not let one simply ignore this term without an additional assumption on the rate of growth of q. Therefore in the above formulation I added the assumption that for all $\nu \geq k+1$

$$\frac{|u'\tilde{x}_\nu|\sqrt{\log q_n}}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} = o(1).$$

Also, (ASS. 3) and (ASS. 4) have been reformulated to be consistent with standard mathematical notation, since in the original paper they are formulated as inequalities between terms not dependent on n and terms dependent on n.

3.10 Proof of Theorem 3

In the following, let $\nu \in \{1, ..., k\} \setminus \mathcal{M}$, i.e. part of the 'true' model, but not yet selected.

Let $t_n(\alpha)$ be defined by the equation

$$1 - \alpha = F_{Beta(\frac{1}{2}, \frac{n-m_0-1}{2})}(t_n(\alpha))^{q_n - m_0}.$$

For x_ν to be selected into the model, the inequality $1 - \frac{ss_\nu}{ss_0} > t_n(\alpha)$ must hold. We see that, by Hypothesis 2, asymptotically x_ν is a candidate to be selected if

$$1 - \frac{ss_\nu}{ss_0} = 1 - \frac{||y - P_{\mathcal{M}_\nu}||^2}{||y - P_{\mathcal{M}}||^2} = \frac{||y - P_{\mathcal{M}}||^2 - ||y - P_{\mathcal{M}_\nu}||^2}{||y - P_{\mathcal{M}}||^2} > \frac{\tau \log q_n}{n}$$

since

$$\frac{\tau \log q_n}{n} \overset{(*)}{\gtrsim} \frac{2\log q_n - \log\log q_n - 2\log(-\log(1-\alpha)) - \log \pi}{n} \asymp t_n(\alpha)$$

for any $\tau > 2$. So we need to show that the first inequality holds with probability tending to 1 for any $\mathcal{M} \in \mathbb{M}$ and some $\nu \in \{1, ..., k\} \setminus \mathcal{M}$.

ad. (*):

Let $\hat{t}_n$ be defined as $\hat{x}_n$ in Hypothesis 2. The relation (*) is defined by

$$\frac{\tau \log q_n}{n} \asymp \max\{\frac{\tau \log q_n}{n}, \frac{2\log q_n - \log\log q_n - 2\log(-\log(1-\alpha)) - \log \pi}{n}\}.$$

So we look at

$$\begin{aligned}&\frac{\frac{\tau \log q_n}{n}}{\max\{\frac{\tau \log q_n}{n}, \frac{2\log q_n - \log\log q_n - 2\log(-\log(1-\alpha)) - \log \pi}{n}\}}\\ &= \frac{\tau \log q_n}{\max\{\tau \log q_n, 2\log q_n - \log\log q_n - 2\log(-\log(1-\alpha)) - \log \pi\}}\\ &= 1\mathbb{1}_{\{\tau \log q_n \geq n\hat{x}_n\}} + \frac{\tau \log q_n}{n\hat{t}_n}\mathbb{1}_{\{\tau \log q_n < n\hat{t}_n\}}.\end{aligned}$$

If $\frac{\tau \log q_n}{n\hat{t}_n}\mathbb{1}_{\{\tau \log q_n < n\hat{t}_n\}} \xrightarrow[n\to\infty]{} 1$ or $\tau \log q_n \geq n\hat{t}_n$ for all $n \geq N$ for some $N \in \mathbb{N}$ then the relation (*) holds. Note that the limit for the given expression has already been calculated in the proof for Hypothesis 2. Using the result from there, we see

$$\lim_{n\to\infty} \frac{\tau \log q_n}{n\hat{t}_n} = \tau \lim_{n\to\infty} \frac{\log q_n}{n\hat{t}_n} = \frac{\tau}{2}$$

So if $\tau > 2$, then indeed for some $N \in \mathbb{N}$ we have for all $n \geq N$

$$\frac{\tau \log q_n}{n\hat{t}_n} > 1$$

and therefore

$$\tau \log q_n > n\hat{t}_n$$

and the relation (*) holds. ✓ Note that analogously the relation (*) becomes $\lesssim$ if $\tau > 2$ is replaced by $\tau' < 2$.

Now, note the general result that, given a fixed $n > k$, for any $\mathcal{M} \in \mathbb{M}$, $\nu \in \{1, ..., q_n\}$ and $v \in \mathbb{R}^n$ the following holds:

$$P_{\mathcal{M}_\nu}(v) = P_{\mathcal{M}}(v) + \frac{v'(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}(x_\nu - P_{\mathcal{M}}(x_\nu)) \tag{1}$$

and

$$||v - P_{\mathcal{M}_\nu}(v)||^2 = ||v - P_{\mathcal{M}}(v)||^2 - \frac{|v'(x_\nu - P_{\mathcal{M}}(x_\nu))|^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}. \tag{2}$$

(1) holds, since for $b_1, ..., b_m$ being an orthogonal basis of $\mathbb{V}_{\mathcal{M}}$ we can write $\mathbb{V}_{\mathcal{M}_\nu} = span(b_j, j \in \mathcal{M}, (x_\nu - P_{\mathcal{M}}(x_\nu)))$, where the vectors in the span are an orthogonal basis for $\mathbb{V}_{\mathcal{M}_\nu}$ and therefore we can write the projection onto $\mathbb{V}_{\mathcal{M}_\nu}$ by simply adding the projection of v onto the vector $(x_\nu - P_{\mathcal{M}}(x_\nu))$ to the projection onto $\mathbb{V}_{\mathcal{M}}$.

(2) then holds, because

$$||v - P_{\mathcal{M}_\nu}(v)||^2 \overset{(1)}{=} ||(v - P_{\mathcal{M}}(v)) - \frac{v'(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}(x_\nu - P_{\mathcal{M}}(x_\nu))||^2$$

$$\begin{aligned}
= & \quad ||(v - P_{\mathcal{M}}(v))||^2 \\
- & \quad 2(v - P_{\mathcal{M}}(v))' \frac{v'(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}(x_\nu - P_{\mathcal{M}}(x_\nu)) \\
+ & \quad ||\frac{v'(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}(x_\nu - P_{\mathcal{M}}(x_\nu))||^2
\end{aligned}$$

$$\begin{aligned}
= & \quad ||(v - P_{\mathcal{M}}(v))||^2 \\
- & \quad 2\frac{v'(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}(v - P_{\mathcal{M}}(v))'(x_\nu - P_{\mathcal{M}}(x_\nu)) \\
+ & \quad \frac{(v'(x_\nu - P_{\mathcal{M}}(x_\nu)))^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}
\end{aligned}$$

and since $P_{\mathcal{M}}(v)$ and $(x_\nu - P_{\mathcal{M}}(x_\nu))$ are orthogonal

$$\begin{aligned}
= & \quad ||(v - P_{\mathcal{M}}(v))||^2 \\
- & \quad 2\frac{v'(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}v'(x_\nu - P_{\mathcal{M}}(x_\nu)) \\
+ & \quad \frac{(v'(x_\nu - P_{\mathcal{M}}(x_\nu)))^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}
\end{aligned}$$

$$= ||v - P_{\mathcal{M}}(v)||^2 - 2\frac{(v'(x_\nu - P_{\mathcal{M}}(x_\nu)))^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2} + \frac{(v'(x_\nu - P_{\mathcal{M}}(x_\nu)))^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}$$

$$= ||v - P_{\mathcal{M}}(v)||^2 - \frac{(v'(x_\nu - P_{\mathcal{M}}(x_\nu)))^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}. \checkmark$$

Notice further, that for any $\mathcal{M} \in \mathbb{M}, \nu \in \{1, ..., q_n\} \setminus \mathcal{M}$ we have

$$||y - P_{\mathcal{M}}(y)||^2 = ||u - P_{\mathcal{M}}(u)||^2 + 2\sigma u'(\epsilon - P_{\mathcal{M}}(\epsilon)) + \sigma^2||\epsilon - P_{\mathcal{M}}(\epsilon)||^2 \quad (3)$$

and

$$||y - P_{\mathcal{M}_\nu}(y)||^2 = ||u - P_{\mathcal{M}_\nu}(u)||^2 + 2\sigma u'(\epsilon - P_{\mathcal{M}_\nu}(\epsilon)) + \sigma^2||\epsilon - P_{\mathcal{M}_\nu}(\epsilon)||^2. \quad (4)$$

ad. (3)/(4):
If (3) holds then trivially (4) holds analogously. Remember, that

$$u = \sum_{j \in \{1,...,k\} \setminus \mathcal{M}} \beta_j x_j$$

and therefore

$$\sum_{j=1}^{k} \beta_j x_j = \sum_{j \in \mathcal{M}} \beta_j x_j + u.$$

Using this we see

$$||y - P_{\mathcal{M}}(y)||^2 = ||\sum_{i=1}^{k} \beta_j x_j + \sigma\epsilon - P_{\mathcal{M}}(\sum_{j=1}^{k} \beta_j x_j + \sigma\epsilon)||^2$$

$$= ||\sum_{j=1}^{k} \beta_j x_j - P_{\mathcal{M}}(\sum_{j=1}^{k} \beta_j x_j) + \sigma\epsilon - P_{\mathcal{M}}(\sigma\epsilon)||^2$$
$$= ||\sum_{j \in \mathcal{M}} \beta_j x_j - P_{\mathcal{M}}(\sum_{j \in \mathcal{M}} \beta_j x_j) + (u - P_{\mathcal{M}}(u)) + \sigma(\epsilon - P_{\mathcal{M}}(\epsilon))||^2$$
$$= ||(u - P_{\mathcal{M}}(u)) + \sigma(\epsilon - P_{\mathcal{M}}(\epsilon))||^2$$
$$= ||u - P_{\mathcal{M}}(u)||^2 + 2\sigma u'(\epsilon - P_{\mathcal{M}}(\epsilon)) + \sigma^2||\epsilon - P_{\mathcal{M}}(\epsilon)||^2. \checkmark$$

Remember that $\tilde{x}_\nu = \frac{x_\nu - P_{\mathcal{M}}(x_\nu)}{||x_\nu - P_{\mathcal{M}}(x_\nu)||}$. With the use of (1)-(4) we obtain

$$||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2 = |u'\tilde{x}_\nu|^2 + 2\sigma u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) + \sigma^2(\epsilon'\tilde{x}_\nu)^2. \quad (5)$$

ad. (5):
First, (3) and (4) give us

$$\begin{aligned}
&||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2 = \\
&||u - P_{\mathcal{M}}(u)||^2 - ||u - P_{\mathcal{M}_\nu}(u)||^2 \\
&+2\sigma(u'(\epsilon - P_{\mathcal{M}}(\epsilon)) - u'(\epsilon - P_{\mathcal{M}_\nu}(\epsilon))) \\
&+\sigma^2(||\epsilon - P_{\mathcal{M}}(\epsilon)||^2 - ||\epsilon - P_{\mathcal{M}_\nu}(\epsilon)||^2).
\end{aligned}$$

Using (2) we see

$$||u - P_{\mathcal{M}}(u)||^2 - ||u - P_{\mathcal{M}_\nu}(u)||^2 \stackrel{(2)}{=} \frac{|u'(x_\nu - P_{\mathcal{M}}(x_\nu))|^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2} = |u'\tilde{x}_\nu|^2 \checkmark$$

$$||\epsilon - P_{\mathcal{M}}(\epsilon)||^2 - ||\epsilon - P_{\mathcal{M}_\nu}(\epsilon)||^2 \stackrel{(2)}{=} \frac{|\epsilon'(x_\nu - P_{\mathcal{M}}(x_\nu))|^2}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2} = |\epsilon'\tilde{x}_\nu|^2 \checkmark$$

and using (1) we see

$$\begin{aligned}
&u'(\epsilon - P_{\mathcal{M}}(\epsilon)) - u'(\epsilon - P_{\mathcal{M}_\nu}(\epsilon)) = u'(\epsilon - P_{\mathcal{M}}(\epsilon) - \epsilon + P_{\mathcal{M}_\nu}(\epsilon)) \\
\stackrel{(1)}{=}\ & u'(\frac{\epsilon'(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||^2}(x_\nu - P_{\mathcal{M}}(x_\nu))) = \epsilon' \frac{(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||} u' \frac{(x_\nu - P_{\mathcal{M}}(x_\nu))}{||x_\nu - P_{\mathcal{M}}(x_\nu)||} \\
&= u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) \checkmark
\end{aligned}$$

So (5) does indeed hold.

In the following, we write $c(a,b)$ for the binomial factor $\binom{a}{b}$. Notice that for a given m there are at most $q_n c(k,k-m)$ subsets $\mathcal{M}_\nu$ of size $m+1$, since for each of the $c(k,k-m)$ possible models $\mathcal{M} \in \mathbb{M}$ of size m, there are $q_n - m \leq q_n$ possible regressors ν to be added. Or in other terms:

$$|\{\mathcal{M}_\nu : m = i\}| \leq q_n c(k,k-i).$$

From this it follows that for a fixed $i \in \mathbb{N}$ such that $0 \leq i < k$ the inequality

$$\max_{\mathcal{M}_\nu : m=i} |\epsilon' \tilde{x}_\nu(\mathcal{M})| \geq \sqrt{2(\log c(k,k-i) + 3\log q)}$$

holds with a probability smaller than

$$\begin{aligned}
&q_n c(k,k-i)\exp(-(\log c(k,k-i) + 3\log q_n)) \\
&= \exp(\log q_n)\exp(\log c(k,k-i))\exp(-\log c(k,k-i) - 3\log q_n) \\
&= \exp(-2\log q_n) = \exp(\log q_n^{-2}) = q_n^{-2}.
\end{aligned}$$

<u>This is, because:</u>

First, note that $\epsilon' \tilde{x}_\nu = \tilde{x}_\nu' \epsilon \sim \mathcal{N}(0,1)$ since $||\tilde{x}_\nu|| = 1$. Therefore, $|\epsilon' \tilde{x}_\nu|$ are Half-normally-distributed with cdf $F_{HN}(x) = \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)$, where erf denotes the Gauss error function, defined as $\operatorname{erf}(x) = 2\Phi(x\sqrt{2}) - 1$ for $x > 0$. Furthermore, erfc denotes the complementary Gauss error function, defined as $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$.

We write $b(k,i,q_n) := \log c(k,k-i) + 3\log q_n$. Then, with HN denoting a Half-normally distributed variable

$$\begin{aligned}
&P\left(\max_{\mathcal{M}_\nu : m=i} |\epsilon' \tilde{x}_\nu| \geq \sqrt{2(\log c(k,k-i) + 3\log q_n)}\right) \\
&= P\left(\max_{\mathcal{M}_\nu : m=i} |\epsilon' \tilde{x}_\nu| \geq \sqrt{2b(k,i,q_n)}\right) \\
&\leq \sum_{\mathcal{M}_\nu : m=i} P\left(HN \geq \sqrt{2b(k,i,q_n)}\right) \\
&= |\{\mathcal{M}_\nu : m = i\}| \cdot P\left(HN \geq \sqrt{2b(k,i,q_n)}\right) \\
&= |\{\mathcal{M}_\nu : m = i\}| \cdot \left(1 - F_{HN}(\sqrt{2b(k,i,q_n)})\right) \\
&\leq q_n c(k,k-i) \cdot \left(1 - F_{HN}(\sqrt{2b(k,i,q_n)})\right)
\end{aligned}$$

$$= q_n c(k, k-i) \cdot \left(1 - \operatorname{erf}\left(\frac{\sqrt{2}\sqrt{b(k,i,q_n)}}{\sqrt{2}}\right)\right)$$

$$= q_n c(k, k-i) \cdot \operatorname{erfc}\left(\sqrt{b(k,i,q_n)}\right)$$

$$\overset{(+)}{\leq} q_n c(k, k-i) \exp(-b(k,i,q_n))$$

where the inequality $(+)$ holds, because for $x > 0$ we have [Chiani, Dardari, 2002]

$$\operatorname{erfc}(x) \leq \exp(-x^2). \checkmark$$

This means, that we have

$$P(\exists \mathcal{M} \in \mathbb{M} : |\mathcal{M}| = i : \exists \nu \in \{1, ..., q_n\} \setminus \mathcal{M} : |\epsilon' \tilde{x}_\nu| \geq \sqrt{2b(k,i,q_n)}) < q_n^{-2}.$$

If we fix a specific $\mathcal{M}$, the probability can only become smaller, therefore, for any given $\mathcal{M} \in \mathbb{M}$ with size i:

$$P(\exists \nu \in \{1, ..., q_n\} \setminus \mathcal{M} : |\epsilon' \tilde{x}_\nu| \geq \sqrt{2b(k,i,q_n)}) < q_n^{-2}.$$

Analogously, if we fix a specific ν, the probability can only become smaller, therefore, for any given $\mathcal{M} \in \mathbb{M}$ with size i and $\nu \in \{1, ..., q_n\} \setminus \mathcal{M}$:

$$P(|\epsilon' \tilde{x}_\nu| \geq \sqrt{2b(k,i,q_n)}) < q_n^{-2}.$$

And therefore, for any given $\mathcal{M} \in \mathbb{M}$ with size i:

$$P(\exists \nu \in \{1, ..., k\} \setminus \mathcal{M} : |\epsilon' \tilde{x}_\nu| \geq \sqrt{2b(k,i,q_n)}) < k q_n^{-2} < q_n^{-1}.$$

or equivalently, for any given $\mathcal{M} \in \mathbb{M}$ with size i

$$P(\forall \nu \in \{1, ..., k\} \setminus \mathcal{M} : |\epsilon' \tilde{x}_\nu| < \sqrt{2b(k,i,q_n)}) \geq 1 - q_n^{-1}. \qquad (*)$$

Notice that this inequality still holds if we use $\forall \nu \in \{1, ..., q_n\} \setminus \mathcal{M}$ instead.

Using this fact together with equation (5), we can deduce that for a given $\mathcal{M} \in \mathbb{M}$ the following inequality holds for all $\nu \in \{1, ..., k\} \setminus \mathcal{M}$ with probability tending to 1 for $n \to \infty$:

$$||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2 \geq (u'\tilde{x}_\nu)^2 - 2\sigma|u'\tilde{x}_\nu|\sqrt{2(\log c(k, k-i) + 3\log q_n)}. \tag{6}$$

ad. (6):

For this, because of (5) and (*), it suffices to show that, if the inequality $|\epsilon'\tilde{x}_\nu| < \sqrt{2(\log c(k, k-i) + 3\log q_n)}$ holds, then also

$$(u'\tilde{x}_\nu)^2 + 2\sigma u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) + \sigma^2(\epsilon'\tilde{x}_\nu)^2 \geq (u'\tilde{x}_\nu)^2 - 2\sigma|u'\tilde{x}_\nu|\sqrt{2(\log c(k, k-i) + 3\log q_n)}.$$

The following statements are equivalent to this

$$2\sigma u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) + \sigma^2(\epsilon'\tilde{x}_\nu)^2 \geq -2\sigma|u'\tilde{x}_\nu|\sqrt{2(\log c(k, k-i) + 3\log q_n)}$$

$$u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) + \frac{\sigma}{2}(\epsilon'\tilde{x}_\nu)^2 \geq -|u'\tilde{x}_\nu|\sqrt{2(\log c(k, k-i) + 3\log q_n)}$$

And since $\frac{\sigma}{2}(\epsilon'\tilde{x}_\nu)^2 \geq 0$, for the last statement, and therefore also the desired statement to be true it suffices to show

$$u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) \geq -|u'\tilde{x}_\nu|\sqrt{2(\log c(k, k-i) + 3\log q_n)}.$$

The case $\epsilon'\tilde{x}_\nu = 0$ is trivial.

Case $\epsilon'\tilde{x}_\nu < 0$:

Then the following is equivalent to the given statement

$$u'\tilde{x}_\nu \leq -|u'\tilde{x}_\nu|\frac{\sqrt{2(\log c(k, k-i) + 3\log q_n)}}{\epsilon'\tilde{x}_\nu}$$

and this a true statement, since

$$\frac{\sqrt{2(\log c(k, k-i) + 3\log q_n)}}{\epsilon'\tilde{x}_\nu} < -1. \checkmark$$

Case $\epsilon'\tilde{x}_\nu > 0$:

Then the following is equivalent to the given statement

$$u'\tilde{x}_\nu \geq -|u'\tilde{x}_\nu|\frac{\sqrt{2(\log c(k, k-i) + 3\log q_n)}}{\epsilon'\tilde{x}_\nu}$$

and this a true statement, since

$$\frac{\sqrt{2(\log c(k, k-i) + 3\log q_n)}}{\epsilon'\tilde{x}_\nu} > 1. \checkmark$$

Notice, that

$$c(k,k-m) \leq k^{k-m} < q_n^{k-m}.$$

Using this it follows from (ASS. 1), that for any given $\mathcal{M} \in \mathbb{M} \setminus \{1,...,k\}$ there exists $\nu = \nu(\mathcal{M})$ with $\nu \in \{1,...,k\} \setminus \mathcal{M}$ such that with probability tending to 1:

$$||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2 \geq (u'\tilde{x}_\nu)^2(1+a_n). \tag{7}$$

for some $a_n = o(1)$.

Ad. (7):

The following holds with probability tending to 1:

$$\begin{aligned}
&||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2 \\
&\overset{(6)}{\geq} (u'\tilde{x}_\nu)^2 - 2\sigma u'\tilde{x}_\nu \sqrt{2(\log c(k,k-i) + 3\log q_n)} \\
&= (u'\tilde{x}_\nu)^2 (1 - \frac{2\sigma\sqrt{2(\log c(k,k-i) + 3\log q_n)}}{u'\tilde{x}_\nu}) \\
&= (u'\tilde{x}_\nu)^2 (1 - 1/\frac{u'\tilde{x}_\nu}{2\sigma\sqrt{2(\log c(k,k-i) + 3\log q_n)}})
\end{aligned}$$

where

$$\begin{aligned}
&\frac{u'\tilde{x}_\nu}{2\sigma\sqrt{2(\log c(k,k-i) + 3\log q_n)}} \geq \frac{u'\tilde{x}_\nu}{2\sigma\sqrt{2(\log q_n^{k-m} + 3\log q_n)}} \\
&= \frac{1}{2\sigma\sqrt{2}} \frac{u'\tilde{x}_\nu}{\sqrt{(k-m)\log q_n + 3\log q_n}} = \frac{1}{2\sigma\sqrt{2}} \frac{u'\tilde{x}_\nu}{\sqrt{(k-m+3)\log q_n}} \xrightarrow[n\to\infty]{} \infty
\end{aligned}$$

for some $\nu \in \{1,...,k\} \setminus \mathcal{M}$, according to (ASS. 1), so indeed

$$1/\frac{u'\tilde{x}_\nu}{2\sigma\sqrt{2(\log c(k,k-i) + 3\log q_n)}} = o(1)$$

and therefore (7) holds. ✓

Also, it holds that for any given $\mathcal{M} \in \mathbb{M}$, with probability tending to 1: (8*)

$$||y - P_{\mathcal{M}}(y)||^2 \leq ||u - P_{\mathcal{M}}(u)||^2 + 2\sigma||u - P_{\mathcal{M}}(u)||\sqrt{\log c(k, k-m) + 3\log q_n} + \sigma^2||\epsilon||^2$$

<u>because:</u>

$$||y - P_{\mathcal{M}}(y)||^2 \stackrel{(3)}{=} ||u - P_{\mathcal{M}}(u)||^2 + 2\sigma u'(\epsilon - P_{\mathcal{M}}(\epsilon)) + \sigma^2||\epsilon - P_{\mathcal{M}}(\epsilon)||^2$$

I.e. the given inequality is equivalent to

$$2\sigma(u - P_{\mathcal{M}}(u))'\epsilon + \sigma^2||\epsilon - P_{\mathcal{M}}(\epsilon)||^2 \leq 2\sigma||u - P_{\mathcal{M}}(u)||\sqrt{b(k, i, q_n)} + \sigma^2||\epsilon||^2.$$

First note, that $||\epsilon - P_{\mathcal{M}}(\epsilon)||^2 \asymp ||\epsilon||^2$ almost surely, because we have

$$\frac{1}{n}||\epsilon||^2 \sim \frac{1}{n}\mathcal{X}_n^2 \xrightarrow{a.s.} 1$$

and for $\{b_1, ..., b_n\}$ an orthonormal basis of $\mathbb{R}^n$ such that $span(x_j : j \in \mathcal{M}) = span(b_j : j \in \mathcal{M})$ we have

$$\frac{1}{n}||\epsilon - P_{\mathcal{M}}(\epsilon)||^2 = \frac{1}{n}||P_{\mathcal{M}}^{\perp}(\epsilon)||^2 = \frac{1}{n}||\sum_{j \notin \mathcal{M}} b_j'\epsilon \cdot b_j||^2$$

and since b_j are orthonormal, $b_j'\epsilon$ are i.i.d. $\mathcal{N}(0, 1)$ and by the Pythagorean theorem

$$\frac{1}{n}||\sum_{j \notin \mathcal{M}} b_j'\epsilon \cdot b_j||^2 = \frac{1}{n}\sum_{j \notin \mathcal{M}} (b_j'\epsilon)^2 \sim \frac{1}{n}\mathcal{X}_{n-m}^2 = \frac{n-m}{n}\frac{1}{n-m}\mathcal{X}_{n-m}^2 \xrightarrow{a.s.} 1.$$

So if

$$(u - P_{\mathcal{M}}(u))'\epsilon < ||u - P_{\mathcal{M}}(u)||\sqrt{b(k, i, q_n)}$$

with probability tending to 1, then the desired inequality also holds with probability tending to 1. Note, that

$$(u - P_{\mathcal{M}}(u))'\epsilon \stackrel{d}{=} ||u - P_{\mathcal{M}}(u)||Z$$

for $Z \sim \mathcal{N}(0, 1)$. So the above inequality holds with probability tending to 1 if

$$P(Z < \sqrt{b(k, i, q_n)}) \xrightarrow[n \to \infty]{} 1$$

which holds, since $\sqrt{b(k, i, q_n)} \xrightarrow[n \to \infty]{} \infty$.

From there it follows from (ASS. 2), that for any given $\mathcal{M} \in \mathbb{M}$, with probability tending to 1:

$$||y - P_{\mathcal{M}}(y)||^2 \leq ||u - P_{\mathcal{M}}(u)||^2(1 + o(1)) + \sigma^2 n(1 + o(1)) \tag{8}$$

<u>Ad. (8):</u>

$$||u - P_{\mathcal{M}}(u)||^2 + 2\sigma||u - P_{\mathcal{M}}(u)||\sqrt{\log c(k, k-m) + 3\log q_n} + \sigma^2||\epsilon||^2$$

$$= ||u - P_{\mathcal{M}}(u)||^2(1 + 2\sigma\frac{\sqrt{\log c(k, k-m) + 3\log q_n}}{||u - P_{\mathcal{M}}(u)||}) + \sigma^2||\epsilon||^2$$

where

$$\frac{\sqrt{\log c(k, k-m) + 3\log q_n}}{||u - P_{\mathcal{M}}(u)||} = 1/\frac{||u - P_{\mathcal{M}}(u)||}{\sqrt{\log c(k, k-m) + 3\log q_n}}$$

$$\leq 1/\frac{||u - P_{\mathcal{M}}(u)||}{\sqrt{\log q_n^{k-m} + 3\log q_n}} = 1/\frac{||u - P_{\mathcal{M}}(u)||}{\sqrt{(k - m + 3)\log q_n}} \xrightarrow[n\to\infty]{} 0$$

by (ASS. 2) and

$$||\epsilon||^2 = n\frac{1}{n}||\epsilon||^2 = n(1 + o(1))$$

since as shown above we have

$$\frac{1}{n}||\epsilon||^2 \xrightarrow{a.s.} 1.$$

Combining (7) and (8), by (ASS. 1) and (ASS. 2), we see that with probability tending to 1 there exists a $\nu \in \{1, ..., k\} \setminus \mathcal{M}$ such that

$$1 - \frac{||y - P_{\mathcal{M}_\nu}(y)||^2}{||y - P_{\mathcal{M}}(y)||^2} > \frac{(u'\tilde{x}_\nu)^2}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n}(1 + o(1)). \tag{9}$$

<u>Ad. (9):</u>

$$1 - \frac{||y - P_{\mathcal{M}_\nu}(y)||^2}{||y - P_{\mathcal{M}}(y)||^2} = \frac{||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2}{||y - P_{\mathcal{M}}(y)||^2}$$

$$\stackrel{(7)}{\geq} \frac{(u'\tilde{x}_\nu)^2}{||y - P_{\mathcal{M}}(y)||^2}(1 + o(1))$$

$$\stackrel{(8)}{\geq} \frac{(u'\tilde{x}_\nu)^2}{||u - P_{\mathcal{M}}(u)||^2(1 + o(1)) + \sigma^2 n(1 + o(1))}(1 + o(1))$$

$$= \frac{(u'\tilde{x}_\nu)^2}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n}(1 + o(1)). \checkmark$$

So with probability tending to 1 there exists a ν which is part of the true model and is a candidate for being added to the current model if asymptotically

$$\frac{(u'\tilde{x}_\nu)^2}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} \geq \frac{\tau \log q}{n}.$$

And according to (ASS. 3), this is the case for at least one ν in the 'true' model. That means if $n \to \infty$ then at least one of the regressors from the 'true' model must be a candidate to be added to the model. ✓

It remains to show that for $n \to \infty$, if ν is not part of the 'true' model, it will not be a candidate to be added to the model.
So let $\mathcal{M} \in \mathbb{M}$, $\nu \in \{k+1, ..., q_n\}$. We obtain

$$||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2 \leq (u'\tilde{x}_\nu)^2 + C|u'\tilde{x}_\nu|\sqrt{\log q_n} \tag{10}$$

with probability tending to 1 for some constant $C > 0$.

<u>Ad. (10):</u>

$$||y - P_{\mathcal{M}}(y)||^2 - ||y - P_{\mathcal{M}_\nu}(y)||^2 \overset{(5)}{=} (u'\tilde{x}_\nu)^2 + 2\sigma u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) + \sigma^2(\epsilon'\tilde{x}_\nu)^2$$

therefore the desired inequality is equivalent to

$$(u'\tilde{x}_\nu)^2 + 2\sigma u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) + \sigma^2(\epsilon'\tilde{x}_\nu)^2 \leq (u'\tilde{x}_\nu)^2 + C|u'\tilde{x}_\nu|\sqrt{\log q_n}$$

$$2\sigma u'\tilde{x}_\nu(\epsilon'\tilde{x}_\nu) + \sigma^2(\epsilon'\tilde{x}_\nu)^2 \leq C|u'\tilde{x}_\nu|\sqrt{\log q_n}$$

$$2\sigma u'\tilde{x}_\nu \frac{\epsilon'\tilde{x}_\nu}{\sqrt{\log q_n}} + \sigma^2 \frac{(\epsilon'\tilde{x}_\nu)^2}{\sqrt{\log q_n}} \leq C|u'\tilde{x}_\nu|$$

$$u'\tilde{x}_\nu \frac{\epsilon'\tilde{x}_\nu}{\sqrt{\log q_n}} + \frac{\sigma}{2}\epsilon'\tilde{x}_\nu \frac{\epsilon'\tilde{x}_\nu}{\sqrt{\log q_n}} \leq \frac{C}{2\sigma}|u'\tilde{x}_\nu|$$

$$\frac{1}{n}u'\tilde{x}_\nu \frac{\epsilon'\tilde{x}_\nu}{\sqrt{\log q_n}} + \frac{\sigma}{2}\frac{1}{n}\epsilon'\tilde{x}_\nu \frac{\epsilon'\tilde{x}_\nu}{\sqrt{\log q_n}} \leq \frac{C}{2\sigma}\frac{1}{n}|u'\tilde{x}_\nu|$$

$$\frac{1}{n}u'\tilde{x}_\nu \frac{\epsilon'\tilde{x}_\nu}{\sqrt{2(k-m+3)\log q_n}} + \frac{\sigma}{2}\frac{1}{n}\epsilon'\tilde{x}_\nu \frac{\epsilon'\tilde{x}_\nu}{\sqrt{2(k-m+3)\log q_n}}$$

$$\leq \frac{C}{\sqrt{2(k-m+3)}2\sigma}\frac{1}{n}|u'\tilde{x}_\nu|$$

where $\epsilon'\tilde{x}_\nu \sim \mathcal{N}(0,1)$, so

$$\frac{1}{n}\epsilon'\tilde{x}_\nu \xrightarrow{a.s.} 0$$

and $|\epsilon'\tilde{x}_\nu| \leq \sqrt{2(\log c(k, k-i) + 3\log q_n)} < \sqrt{2(k-m+3)\log q_n}$ with probability tending to 1 according to (*). Therefore

$$\frac{\epsilon'\tilde{x}_\nu}{\sqrt{2(k-m+3)\log q_n}} \leq 1$$

with probability tending to 1 and so the above inequality indeed holds with probability tending to 1, for any constant $C \geq \sqrt{2(k-m+3)}2\sigma$. ✓

And, similar to (8), we have with probability tending to 1

$$||y - P_{\mathcal{M}}(y)||^2 \geq ||u - P_{\mathcal{M}}(u)||^2(1 + o(1)) + \sigma^2 n(1 + o(1)). \tag{11}$$

<u>ad. (11):</u>
First note that analogously to (8*) we obtain

$$||y - P_{\mathcal{M}}(y)||^2 \geq ||u - P_{\mathcal{M}}(u)||^2 - 2\sigma||u - P_{\mathcal{M}}(u)||\sqrt{b(k, i, q_n)} + \sigma^2||\epsilon||^2$$

with probability tending to 1 because

$$(u - P_{\mathcal{M}}(u))'\epsilon > -||u - P_{\mathcal{M}}(u)||\sqrt{b(k, i, q_n)}$$

with probability tending to 1, since

$$(u - P_{\mathcal{M}}(u))'\epsilon \stackrel{d}{=} ||u - P_{\mathcal{M}}(u)||Z$$

and

$$P(Z > -\sqrt{b(k, i, q_n)}) \xrightarrow[n\to\infty]{} 1.$$

from which (11) follows analogously to the argument in (8).

Combining (10) and (11), we have analogously to (9) with probability tending to 1:

$$1 - \frac{||y - P_{\mathcal{M}_\nu}(y)||^2}{||y - P_{\mathcal{M}}(y)||^2} < \frac{(u'\tilde{x}_\nu)^2 + C|u'\tilde{x}_\nu|\sqrt{\log q_n}}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n}(1 + o(1)).$$

Therefore $\tilde{x}_v$ is asymptotically not a candidate to be added if

$$\frac{(u'\tilde{x}_\nu)^2 + C|u'\tilde{x}_\nu|\sqrt{\log q_n}}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} \leq \frac{\tau' \log q_n}{n}$$

for some $\tau' < 2$.

Note, that

$$\frac{(u'\tilde{x}_\nu)^2 + C|u'\tilde{x}_\nu|\sqrt{\log q_n}}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} = \frac{(u'\tilde{x}_\nu)^2}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} + \frac{C|u'\tilde{x}_\nu|\sqrt{\log q_n}}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n}$$

and by (ASS. 4), for some $N \in \mathbb{N}$ we have for all $n \geq N$:

$$\frac{(u'\tilde{x}_\nu)^2}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} < \frac{\tau' \log q_n}{n}.$$

That means if

$$\frac{C|u'\tilde{x}_\nu|\sqrt{\log q_n}}{||u - P_{\mathcal{M}}(u)||^2 + \sigma^2 n} \xrightarrow[n\to\infty]{} 0$$

then indeed x_ν is asymptotically not a candidate and we are done. But this is the case, by assumption.

$\square$

4 Discussion of the results

With their proposed novel approach to linear least squares regression, Davies and Dümbgen might be on track to introducing a useful model-free alternative to the classical method of frequentist inference. Their result on the distribution of the ratio of residual sum of squares $\frac{SS_1}{ss_0}$ in their defined setting (elaborated as Theorem 1 above) is simple, yet provides a powerful basis for further exploration of the possibilities and properties lying within this approach. Given the fact that the result also includes a statement about the equality of the p-values calculated with this assumption-free approach and the p-values calculated from the typical approach with standard assumptions, it might also be seen as a way to justify results obtained by that approach, by simply shifting their interpretation to this assumption-free setting.

The 2019 draft version of their paper, which the bulk of this thesis is based upon, however appears to overlook some important details in the included proofs about the consistency of the single stepwise procedure which are discussed above. (Theorems 4.1 and 4.2 in the 2019 version of the paper, Hypothesis 2 and Theorem 3 in this thesis). Notably they refrain from arguing why the implicit dependence of q on n does not interfere in key steps of the argumentation in both of these proofs. The authors address this problem in the second draft version of their paper from May 19, 2020, in which they replace the formulation and proofs of these theorems.

5 Appendix

5.1 German Abstract

In ihrem Paper "A Model-free Approach to Linear Least Squares Regression with Exact Probabilities and Applications to Covariate Selection" ("Ein modellfreier Ansatz zu linearer Kleinste-Quadrate Regression mit exakten Wahrscheinlichkeiten und Anwendungen zu Kovariablenselektion") schlagen Laurie Davies und Lutz Dümbgen einen neuen Ansatz zur Modellselektion in einem linearen Kleinst-Quadrate Setting vor, welches weder Hypothesen benutzt noch ein Modell für die Verteilung des Fehlerterms voraussetzt. Unter Anderem liefern sie eine Möglichkeit, für beliebige gegebene reellwertige Daten gültige p-Werte zu berechnen und einen Satz von Regressoren mittels einer schrittweisen Methode zu selektieren. Das Ziel dieser Arbeit ist es, einen Überblick und einen Kommentar zu ihrem Ansatz und weiters auch zu einigen ihrer Theoreme und der entsprechenden Beweise abzugeben. Hierbei liegt die Betrachtung auf der Entwurfsversion ihres Papers vom 5. Juni 2019.

References

[Berk et al., 2013] Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. *Valid post-selection inference.* Annals of Statistics, 41(2):802–837, 2013

[Chiani, Dardari, 2002] Chiani, M. and Dardari, D., *Improved exponential bounds and approximation for the Q-function with application to average error probability computation*, Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE, Taipei, Taiwan, 2002, pp. 1399-1402 vol.2, doi: 10.1109/GLOCOM.2002.1188428.

[Davies & Dümbgen, 2019] Davies, Laurie and Dümbgen, Lutz *A Model-free Approach to Linear Least Squares Regression with Exact Probabilities and Applications to Covariate Selection.* `https://arxiv.org/abs/1906.01990`, 2019/2020

[Feller, 1957a] Feller, William. *An Introduction to Probability Theory and its Applications, Second Edition, Volume 1.* John Wiley & Sons, Inc., Publishers, 1957

[Jambunathan, 1954] Jambunathan, M. V.. *Some Properties of Beta and Gamma Distribution.* The Annals of Mathematical Statistics, vol. 25, no. 2, 1954, pp. 401–405. JSTOR, `www.jstor.org/stable/2236745`.