UNIVERSITETET I OSLO

Matematisk Institutt

Exam in:	**STK 4160 – Model Selection and Model Averaging**
	Part I of two parts: The project
With:	**Nils Lid Hjort**
Time for exam:	**6.–17.vi.2013**

This is the exam project set for STK 4160, spring semester 2013. It is made available on the course website as of *Thursday 6 June 12:00*, and candidates must submit their written reports by *Monday 17 June 14:00* (or earlier), to the reception office at the Department of Mathematics, in duplicate. The supplementary oral examinations take place *Friday June 21* (practical details concerning this are provided elsewhere). Reports may be written in nynorsk, bokmål, riksmål, English or Latin, and should preferably be text-processed (TeX, LaTeX, Word), but may also be hand-processed. Give your name on the first page. Write concisely (in der Beschränkung zeigt sich erst der Meister; brevity is the soul of wit; краткость – сестра таланта). Relevant figures need to be included in the report. Copies of machine programmes used (in `R`, or `matlab`, or similar) are also to be included, perhaps as an appendix to the report. Candidates are required to work on their own (i.e. without cooperation with any others). They are graciously allowed not to despair if they do not manage to answer all questions well.

Importantly, each student needs to submit *two special extra pages* with her or his report. *The first* (page A) is the 'erklæring' (self-declaration form), properly signed; it is available at the webpage as 'Exam Project, page A, declaration form'. *The second* (page B) is the student's one-page summary of the exam project report, which should also contain a brief self-assessment of its quality.

This exam set contains three exercises and comprises ten pages (including two pages with Appendices A, B, C).

Exercise 1

Consciousness itself is an infinite regression (says R.A. Wilson, and adds that this explains coincidences). Consider regression data of the familiar form (y_i, x_i) for $i = 1, \ldots, n$, where y_i is the response for individual or object i with covariate vector $x_i = (x_{i,1}, \ldots, x_{i,p})^{\mathrm{t}}$. The traditional linear normal regression model takes the y_i to be observations of random variables of the form

$$Y_i = x_i^{\mathrm{t}}\beta + \sigma\varepsilon_i \quad \text{for } i = 1, \ldots, n,$$

with the ε_i taken independent and standard normal, and where the $(p+1)$-dimensional model parameter is $\theta = (\sigma, \beta) = (\sigma, \beta_1, \ldots, \beta_p)$. We assume $n > p$ and that the variance matrix of the x_i has full rank p.

(a) Work out a convenient expression for the log-likelihood function

$$\ell_n(\theta) = \sum_{i=1}^n \log f(y_i \,|\, x_i, \theta),$$

and explain how the maximum likelihood estimates $\widehat\theta = (\widehat\sigma, \widehat\beta)$ may be found.

(b) Find a formula for $\ell_{n,\max}$, the maximised log-likelihood. Then demonstrate that ranking submodels (defined as selecting covariates among the p used in the full model, say $S \subset \{1, \ldots, p\}$) using the Akaike Information Criterion (AIC) is equivalent to ranking values of

$$a(S) = n \log \widehat\sigma_S + |S|$$

(with small values of $a(S)$ corresponding to large values of AIC). Here, $\widehat\sigma_S$ is the maximum likelihood estimate of σ in the model indexed by S, with $|S|$ the number of covariates included in S.

(c) Going back to the full model again (i.e. with all p covariates included), show that the matrix

$$\widehat J_n = -\frac{1}{n}\frac{\partial^2 \ell_n(\widehat\theta)}{\partial\theta\partial\theta^{\rm t}} \quad \text{may be written} \quad \frac{1}{\widehat\sigma^2}\begin{pmatrix} 2 & 0 \\ 0 & \Sigma_n \end{pmatrix},$$

where $\Sigma_n = n^{-1}\sum_{i=1}^n x_i x_i^{\rm t}$. (It is convenient to know that Σ_n also be expressed as $n^{-1}X^{\rm t}X$, where X is the $n \times p$ matrix having $x_i^{\rm t}$ as its ith row.)

(d) Set up a convenient expression for the score vector $u(y_i \,|\, x_i, \theta) = \partial \log f(y_i \,|\, x_i, \theta)/\partial\theta$, and show that

$$\widehat K_n = \frac{1}{n}\sum_{i=1}^n u(Y_i \,|\, x_i, \widehat\theta) u(Y_i \,|\, x_i, \widehat\theta)^{\rm t} \quad \text{may be written} \quad \frac{1}{\widehat\sigma^2}\begin{pmatrix} 2 + \widehat\kappa_4 & c_n^{\rm t} \\ c_n & M_n \end{pmatrix},$$

where $M_n = n^{-1}\sum_{i=1}^n \widehat\varepsilon_i^2 x_i x_i^{\rm t}$, in terms of estimated normalised residuals $\widehat\varepsilon_i = (y_i - x_i^{\rm t}\widehat\beta)/\widehat\sigma$; also, $\widehat\kappa_4 = n^{-1}\sum_{i=1}^n \widehat\varepsilon_i^4 - 3$ is their so-called kurtosis, and $c_n = n^{-1}\sum_{i=1}^n \widehat\varepsilon_i^3 x_i$.

(e) The formula almost always used for $\mathrm{Var}\,\widehat\beta$, in textbooks and software, is $\widehat\sigma^2\Sigma_n^{-1}/n$. When might $\widehat\sigma^2\Sigma_n^{-1}M_n\Sigma_n^{-1}/n$ be more appropriate?

(f) Show that ranking models using high scores of the Takeuchi Information Criterion (or model robust AIC) is equivalent to ranking models by small values of

$$a^*(S) = n \log \widehat\sigma_S + \mathrm{Tr}(\Sigma_{n,S}^{-1} M_{n,S}) + \tfrac{1}{2}\widehat\kappa_{4,S}.$$

Here $\Sigma_{n,S}$ and $M_{n,S}$ are constructed as above, but based on the covariate vectors corresponding to subset S, i.e. $\Sigma_{n,S} = n^{-1}\sum_{i=1}^n x_{i,S} x_{i,S}^{\rm t}$, etc., where $x_{i,S} = \pi_S x_i$ is the vector of length $|S|$ having components $x_{i,j}$ with $j \in S$. Similarly, $\widehat\kappa_{4,S}$ is the kurtosis of the normalised residuals $\widehat\varepsilon_{i,S} = (y_i - x_{i,S}^{\rm t}\widehat\beta_S)/\widehat\sigma_S$ emerging for submodel S.

Exercise 2

Hyller virkelighetens faktum: det at livet / lar seg abstrahere / til en sirkel, med indre / og ytre bane, for den som ser / lenge nok. In allround championships, speedskaters race four distances, with the men recording results t_1, t_2, t_3, t_4 for the 500 m, 1500 m, 5000 m, 10000 m. These results are converted into points on the 500 m scale, i.e. to

$$x_1 = t_1/1, \quad x_2 = t_2/3, \quad x_3 = t_3/10, \quad x_4 = t_4/20,$$

giving also a 'samalogue pointsum' of

$$\text{pointsum} = x_1 + x_2 + x_3 + x_4 = t_1/1 + t_2/3 + t_3/10 + t_4/20,$$

which then determines gold, silver, bronze, etc.; check the results list on page 9 for the 2013 World Allround Championships held at Vikingskipet, Hamar, 16–17 February. (Similarly, the female skaters, consistently termed 'ladies' by the International Skating Union [ISU], skate 500 m, 1500 m, 3000 m, 5000 m, with pointsum $t_1/1 + t_2/3 + t_3/6 + t_4/10$. For simplicity and concreteness we limit discussion here to the men's competitions.) Thus when Håvard Bøkko skated the 500 m 0.70 seconds faster than Sven Kramer, Kramer needed to skate 7.00 seconds faster on the 5000 m in order to match Bøkko after two distances, etc.

To the very brief explanation above must be added one crucial detail, which will also be the statistical focus of the present exercise – that not all participants are allowed the privilege of skating the fourth distance. 2013 actually marks what to many athletes and speedskating fans is an unfortunate discontinuity point, as from this year onwards only four pairs, i.e. *only eight skaters*, are allowed to skate the 10k in the European and World Allround Championships (similarly, only eight ladies can skate their 5k). From 1993 to 2012 there were twelve skaters (six pairs) on the longest distance, and up to 1992 there were sixteen skaters (eight pairs). Reasons given for this draconian cut-down include allusions to alleged wishes for making the televised events more 'compact'.

ISU's *New Rules* determining who among the skaters should be allowed to skate the 10k are given in this exam set's Appendix B. We do not necessarily need to go into their specifics here (essentially, those being among the top eight on both the 5k and samalogue after three distances lists are safe, and then it's about having a high rank on one of these lists in case one isn't on both), but they can perhaps be characterised as common sense but ad hoc rules, as opposed to statistically constructed rules. Their *intention* ought to be clear – in the contrafactual world, where all skaters race all four distances, the rule hopes to pick out the top eight skaters in the final samalogue list, i.e. those with the smallest values of $x_1 + x_2 + x_3 + x_4$. In speedskating discussion fora some have argued that the New Rules are a too simple modification of the previous rules; picking eight skaters based on top ranking on the two lists is a more vulnerable operation than picking twelve skaters based on the same lists (as was done from 1999 to 2012).

In the present exercise we shall therefore embark on a study of *statistically constructed rules*. The task is to predict x_4 (or, equivalently, the 10k time $t_4 = 20\,x_4$), and hence the toal pointsum $x_1 + x_2 + x_3 + x_4$, for each skater, after having witnessed and applauded his x_1, x_2, x_3. With such a prediction formula one can then handpick the eight skaters who should race the longest distance. To construct such rules we shall use the top $n = 250$ skaters from the Adelskalenderen; see the files `adelmen-april2013A` and `adelmen-april2013B` at the course website. Explanation and background for these data are otherwise as in the Claeskens and Hjort book (see e.g. pages 17 and 289), but the files now uploaded are the updated ones as of end-of-season 2012–2013.

(a) Read the Adelskalenderen data into your computer (see Appendix C for some helpful **R** tricks), and convert result times into points x_1, x_2, x_3 and $y = x_4$. For reasons of both numerical accuracy (when working with various regression models) and interpretation it is convenient to analyse data in terms of the transformed covariates

$$
\begin{aligned}
x_1^* &= x_1 - \bar{x}_1 = x_1 - 36.9333,\\
x_2^* &= x_2 - \bar{x}_2 = x_2 - 35.8764,\\
x_3^* &= x_3 - \bar{x}_3 = x_3 - 38.8718
\end{aligned}
$$

(with the bar as usual denoting the average of the variable in question). As a gentle start, carry out simple linear regression of y with respect to x_3^*; use this to form a prediction formula for $y = x_4$ based on x_3^* alone; apply this to predict $y = x_4$ for each of the 24 skaters of the 2013 World Allround Championships (see Appendix A, and the files `world2013-menA` and `world2013-menB` at the course website) based on their achievements on the 5k; and use this to produce a list of the top eight skaters (who are the ones who therefore ostensibly should have been allowed to skate the 10k).

(b) Speedskaters differ, in size and shape and style and special talents; in particular, some are more 'sprinter type' and others are more 'stayer type'. Prediction rules taking this into account may start from the idea of making one regression for sprinters and another for stayers. Use the Adelskalenderen data to form such a rule, along the lines of

$$
\widehat{y}_i = \begin{cases} \widehat{\beta}_{0,A} + \widehat{\beta}_{1,A} x_{i,3}^* & \text{if } r_i > r_0,\\ \widehat{\beta}_{0,B} + \widehat{\beta}_{1,B} x_{i,3}^* & \text{if } r_i \le r_0, \end{cases}
$$

where $r_i = x_{i,3}/x_{i,1}$ and r_0 is a suitably chosen threshold parameter. Carry out such a scheme, first with the somewhat ad hoc value $r_0 = 1.035$ used for the illustration in Figure 1, and compute

$$
\text{crit} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y}_i|.
$$

Then attempt to find the r_0 value that minimises this crit value, and duplicate a version of Figure 1 for this optimal threshold value.

(c) An extension of the idea above is to think in terms of a regression structure $\beta_0 + \beta_1 x^*_{i,3}$ but where the parameters β_0 and β_1 depend on the ratio $r_i = x_{i,3}/x_{i,1}$. Argue that the regression model with mean structure

$$\beta_0 + \kappa_0 r_i + (\beta_1 + \kappa_1 r_i) x^*_{i,3}$$

is essentially the same as linear regression in the covariates $x^*_{i,3}, z^*_{i,1}, z^*_{i,2}$, where

$$z_{i,1} = x_{i,3}/x_{i,1} \quad \text{and} \quad z^*_{i,1} = z_{i,1} - \bar{z}_1 = z_{i,1} - 1.0531,$$
$$z_{i,2} = x^2_{i,3}/x_{i,1} \quad \text{and} \quad z^*_{i,2} = z_{i,2} - \bar{z}_2 = z_{i,2} - 40.9795.$$

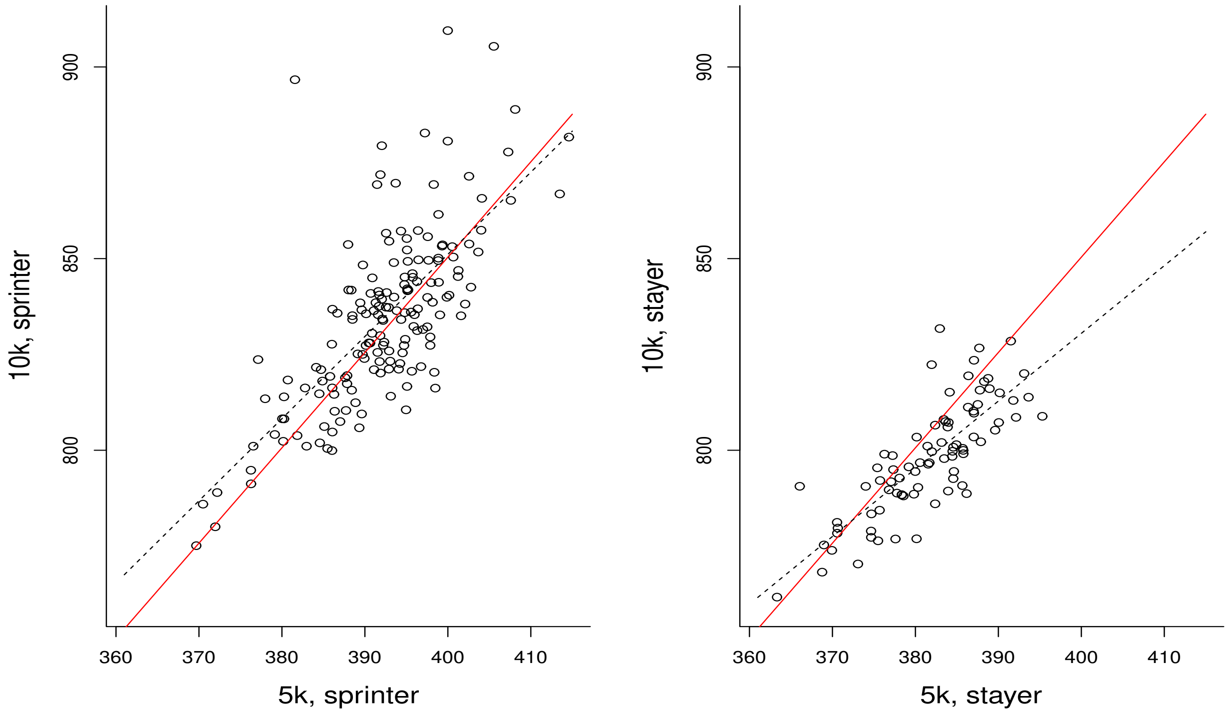

Figure 1: 5k and 10k personal bests are plotted for sprinters ($r > r_0$, to the left) and stayers ($r \le r_0$, to the right), where $r = x_3/x_1$, and with threshold for this illustration chosen as $r_0 = 1.035$. The full line is the overall regression line, and is the same in both parts of the figure, whereas the dotted lines indicate regression lines for respectively sprinters and stayers.

(d) The arguments above motivate studying the following linear regression model, along with various submodels:

$$Y_i = \beta_0 + \beta_1 x^*_{i,1} + \beta_2 x^*_{i,2} + \beta_3 x^*_{i,3} + \gamma_1 z^*_{i,1} + \gamma_2 z^*_{i,2} + \sigma \varepsilon_i,$$

with the ε_i taken independent and standard normal. We consider $x^*_{i,1}, x^*_{i,2}, x^*_{i,3}$ protected covariates and let models 0, 1, 2, 12 indicate those corresponding to exclusion or inclusion of the extra covariates $z^*_{i,1}, z^*_{i,2}$. Fit each of these four models, and for each compute three scores: the AIC; the TIC or model-robust AIC; and the cross-validated log-density score xv. Comment on your findings.

(e) Suppose a certain skater has results $t_{1,0}, t_{2,0}, t_{3,0}$ for the first three distances, leading to points $x_0 = (x_{1,0}, x_{2,0}, x_{3,0})$. Set up the necessary formulae for applying the Focussed Information Criterion (FIC) for selecting among models 0, 1, 2, 12 for the purpose of estimating the mean $\mu = \mu(x_0)$ of the distribution for $y = x_4$. Here we treat model 12 as the wide model and model 0 as the narrow model. Apply this FICology machinery for selecting the best model for predicting the 10k time of respectively Håvard Bøkko and Øystein Grødum (no. 4 and no. 64 on the Adelskalenderen). Provide details, perhaps in the form of a table and/or a plot. Also, comment both on your findings and on the assumptions underlying your analysis.

(f) Rather than focussing on one skater at a time, carry out a suitable average weighted FIC analysis (AFIC), where the task is to jointly estimate the mean values $\mu = \mu(x_0)$ for a total of six skaters from the 2013 World Allround Championships, namely Rotteveel, Silovs, Brodka, Verweij, Yuskov, Blokhuijsen, based on their achieved $x_0 = (x_{1,0}, x_{2,0}, x_{3,0})$ results during the first three distances in that competition (i.e. not their personal best times given in the Adelskalenderen). Carry out such an AFIC analysis, again comparing models 0, 1, 2, 12, giving equal weight of importance to these six skaters. Discuss why this might be a sensible approach in the present context of constructing alternatives to ISU's New Rules.

(g) Speedskaters are heteroskedastic creatures and it is useful to study models taking this suitably into account. In addition to models 0, 1, 2, 12 described in point (d), consider the eight-parameter model defined by

$$Y_i = \beta_0 + \beta_1 x_{i,1}^* + \beta_2 x_{i,2}^* + \beta_3 x_{i,3}^* + \gamma_1 z_{i,1}^* + \gamma_2 z_{i,2}^* + \sigma \exp(\phi x_{i,1}^*)\varepsilon_i,$$

where the ε_i again denote independent standard normals. Fit also this model to the Adelskalenderen data, and give a confidence interval for the heteroskedasticity parameter ϕ. Compute AIC and the log-density cross-validation scores, and comment on your findings.

(h) Explain how the FIC and AFIC analyses of points (e) and (f) must be amended if we take the extended model of point (g) as the wide model, i.e. rather than model 12. If you have time, carry out such FIC and AFIC analyses, now comparing five models, i.e. models 0, 1, 2, 12, and the wider model of (g).

(i) Feel free to explore also one or two alternative models for the Adelskalenderen data, with the hope of achieving more correct predictions than those stemming from the models utilised above. Based on your explorations in this exercise, formulate your favourite statistical rule for how to determine which eight skaters should be allowed to skate the 10k in the World Allround Championships in Heerenveen, 22–23 March 2014. Also, check which skaters this rule of yours would have selected for the 10k at the Hamar event 16–17 February 2013.

Exercise 3

EVERY EXTENSION OF KNOWLEDGE arises from making the conscious the unconscious. Let this be our Nietzschean motivation for initially studying a somewhat simple but nevertheless useful model for data on the unit interval $(0,1)$, with density $\theta y^{\theta-1}$, where θ is a positive parameter. In the present exercise we shall consider a two-parameter extension of this, defined by the density

$$f(y,\theta,\gamma) = \theta y^{\theta-1}\gamma(1-y^\theta)^{\gamma-1} \quad \text{for } y \in (0,1),$$

with γ being this second positive parameter. This model may of course be utilised in a general fashion, but it is of special interest to examine its behaviour and uses when γ is in the vicinity of $\gamma_0 = 1$.

(a) Show that the density indeed integrates to 1, and that its cumulative may be expressed as

$$F(y,\theta,\gamma) = 1-(1-y^\theta)^\gamma \quad \text{for } y \in (0,1).$$

Find also an explicit expression for the p_0 quantile

$$\mu = \mu(\theta,\gamma) = F^{-1}(p_0,\theta,\gamma),$$

where p_0 is a given level.

(b) Assuming for the moment that the narrow model corresponding to $\gamma = 1$ is correct, find an expression for the maximum likelihood estimator $\widehat\theta_{\rm narr}$, and the associated limit distribution for $\sqrt{n}(\widehat\theta_{\rm narr}-\theta)$. Find also the limit distribution for

$$\sqrt{n}(\widehat\mu_{\rm narr}-\mu), \quad \text{where} \quad \widehat\mu_{\rm narr} = p_0^{1/\widehat\theta_{\rm narr}} \text{ and } \mu = p_0^{1/\theta}.$$

(c) The model has a Fisher information matrix $J = J(\theta,\gamma)$, which we here shall need to study carefully only at the narrow null model, i.e. for $\gamma = 1$. Show that $J(\theta,1)$ can be expressed as

$$J = \begin{pmatrix} 1/\theta^2 & k/\theta \\ k/\theta & 1 \end{pmatrix},$$

where $k = -0.64493$ (actually, $k = 1-\pi^2/6$, which you may try to prove, but it is sufficient to find its numerical value).

(d) For estimating the p_0 quantile, there are at the outset two estimators to consider,

$$\widehat\mu_{\rm narr} = p_0^{1/\widehat\theta_{\rm narr}} \quad \text{and} \quad \widehat\mu_{\rm wide} = \{1-(1-p_0)^{1/\widehat\gamma}\}^{1/\widehat\theta},$$

where $(\widehat\theta,\widehat\gamma)$ are the maximum likelihood estimators in the wider two-parameter model. How much can γ be allowed to differ from 1, before $\widehat\mu_{\rm wide}$ becomes more precise than $\widehat\mu_{\rm narr}$? Briefly discuss your result.

(e) Under the local asymptotics scenario $\gamma = 1+\delta/\sqrt{n}$, so that in particular the p_0 quantile we focus on may be expressed as $\mu_n = \mu(\theta_0, 1+\delta/\sqrt{n})$, use general results from Claeskens & Hjort (Ch. 7) to give a clear representation of the joint limit distribution of

$$\begin{pmatrix} \sqrt{n}(\widehat\mu_{\rm narr} - \mu_n) \\ \sqrt{n}(\widehat\mu_{\rm wide} - \mu_n) \\ D_n \end{pmatrix},$$

where $D_n = \sqrt{n}(\widehat\gamma - 1)$. Use this to exhibit the limiting correlation between $\widehat\mu_{\rm narr}$ and $\widehat\mu_{\rm wide}$. Also, give formulae for

$$\begin{aligned} {\rm risk}_{\rm narr}(\delta) &= \lim_{n\to\infty} n\,{\rm E}(\widehat\mu_{\rm narr} - \mu_n)^2, \\ {\rm risk}_{\rm wide}(\delta) &= \lim_{n\to\infty} n\,{\rm E}(\widehat\mu_{\rm wide} - \mu_n)^2, \end{aligned}$$

and draw these in a diagram, for the case of $\theta_0 = 3.333$ and the focus parameter being the median.

(f) For any function $m(D_n)$, either continuous or with a finite number of discontinuity points, use the above to give a representation of the limit distribution of

$$\sqrt{n}(\widehat\mu^* - \mu_n), \quad \text{where} \quad \widehat\mu^* = \{1 - m(D_n)\}\widehat\mu_{\rm narr} + m(D_n)\widehat\mu_{\rm wide}.$$

For the particular case of wide model preference function

$$m(D_n) = \frac{\exp(\tfrac12 D_n^2/\kappa^2 - 1)}{1 + \exp(\tfrac12 D_n^2/\kappa^2 - 1)},$$

where $\kappa^2 = 1/(1-k^2)$, give a suitable expression for

$${\rm risk}_{\rm smooth}(\delta) = \lim_{n\to\infty} n\,{\rm E}(\widehat\mu^* - \mu_n)^2$$

(perhaps as an integral), and if you find time, compute and display also this curve, alongside those you found in point (e) (again for $\theta_0 = 3.333$ and the focus being the median). How can this particular $m(D_n)$ function be motivated?

Appendix A: World Allround Championships results 2013

1	Sven Kramer NED	36.71 (9)	6:13.42 (1)	1:46.75 (4)	13:11.86 (1)	149.228
2	Håvard Bøkko NOR	36.01 (2)	6:22.00 (5)	1:46.34 (1)	13:15.83 (3)	149.447
3	Bart Swings BEL	36.73 (10)	6:19.72 (3)	1:46.51 (3)	13:11.91 (2)	149.800
4	Sverre L. Pedersen NOR	36.66 (7)	6:20.06 (4)	1:47.15 (8)	13:25.65 (4)	150.664
5	Ivan Skobrev RUS	36.79 (12)	6:19.06 (2)	1:46.92 (6)	13:35.90 (6)	151.131
6	Renz Rotteveel NED	36.92 (14)	6:25.12 (7)	1:48.39 (11)	13:35.84 (5)	152.354
7	Haralds Silovs LAT	36.20 (3)	6:32.52 (12)	1:47.11 (7)	13:47.38 (7)	152.524
8	Zbigniew Brodka POL	35.80 (1)	6:35.88 (15)	1:46.49 (2)	14:09.29 (8)	153.348
9	Koen Verweij NED	36.65 (6)	6:24.35 (6)	1:47.51 (9)		110.921
10	Denis Yuskov RUS	36.97 (16)	6:28.23 (10)	1:46.77 (5)		111.383
11	Jan Blokhuijsen NED	36.68 (8)	6:25.89 (8)	1:48.59 (13)		111.465
12	Jan Szymanski POL	36.39 (4)	6:35.43 (14)	1:48.02 (10)		111.939
13	Jonathan Kuck USA	37.41 (18)	6:27.62 (9)	1:48.44 (12)		112.318
14	Dmitry Babenko KAZ	37.22 (17)	6:34.93 (13)	1:49.57 (16)		113.236
15	Simen S. Nilsen NOR	36.74 (11)	6:42.07 (19)	1:49.71 (17)		113.517
16	Roland Cieslak POL	37.41 (18)	6:37.71 (16)	1:49.15 (14)		113.564
17	Lucas Makowsky CAN	36.81 (13)	6:46.75 (21)	1:49.15 (14)		113.868
18	Moritz Geisreiter GER	38.55 (24)	6:31.13 (11)	1:49.93 (18)		114.306
19	Hiroki Abe JPN	37.42 (20)	6:41.94 (18)	1:51.51 (20)		114.784
20	Alec Janssens CAN	38.14 (22)	6:39.87 (17)	1:49.98 (19)		114.787
21	Longjiang Sun CHN	36.50 (5)	6:50.93 (22)	1:52.75 (22)		115.176
22	Joey Mantia USA	36.96 (15)	6:55.19 (23)	1:52.32 (21)		115.919
23	Marco Cignini ITA	38.22 (23)	6:42.76 (20)	1:53.37 (23)		116.286
24	Bram Smallenbroek AUT	37.66 (21)	WDR (24)			

Table 1. *Results from the World Allround Championships 2013 held at Hamar, February 16-17. The columns give the times achieved for the four distances 500 m, 5000 m, 1500 m, 10000 m (with numbers in parentheses indicating ranking), and the pointsum. The new ISU rules allow as of 2013 only eight skaters to start at the 10000 m, so for the remaining skaters the pointsum given is that after three distances. (Note: Koen Verweij was qualified for the 10k, but withdrew.)*

Appendix B: Current ISU Rules

From the ISU Rules, as detailed in *Special Regulations & Technical Rules for Speed Skating and Short Track Speed Skating 2012* I excerpt the following (page 49):

Qualification for the fourth distance at World Allround Championships, Rule 4:

(a) In the fourth distance only 8 Competitors shall start. The selection of qualified Competitors are made among Skaters ranked among the top 16 after 3 distances, and will be based on two different ranking lists: The final classification in the longest of the three skated distances (i.e. 3000 m Ladies and 5000 m Men, respectively), and the classification in total points after three distances. Competitors who are among the 8 best in both of these ranking lists are directly qualified for the fourth distance. Among the Competitors who are placed among the 8 best in only one of these ranking lists, the next to qualify is the Competitor with the best position in either of the two lists. If two Competitors have equal position in the two ranking lists, the Competitor in the classification in total points after three distances is the first of them to qualify. If two Competitors share the same position in one of the ranking lists, the Competitor who is better placed in the other ranking list, will qualify first.

Long is the list of excellent 10000 m races of the past which would never have taken place had the New Rules been in place, e.g. Geir Karlstad's world record race 14:12.14. A rather drastic illustration of how the New Rules would have worked out for various past events is the following, from the European Championships in Helsinki 1998. Then *three of the top six* (including bronze medallist Vadim Saiutin) would not have qualified for the 10k and hence not have been on the final result list.

Appendix C: Some useful R tricks

Here I list just a few potentially useful **R** programming details.

1. To read the Adelskalenderen data (as of April 2013) into your **R** session, use e.g.
   `dataall <- matrix(scan("adelmen-april2013B",skip=3),byrow=T,ncol=12)`
   from which you may then go on with `data <- dataall[1:250, ]`.
2. To easily find parameter estimates in standard regression models, without necessarily programming the log-likelihood function etc., one may use
   `look <- glm(y ~ X + Z, family = gaussian)`
   followed by `look$coef`. Here `X` and `Z` may be matrices with several columns each. Also, `look$res` will give the residuals.
3. To carry out cross-validation one needs to delete one line at a time from a data matrix. If `A` is such a matrix, then `A[-7, ]` is the reduced matrix having line 7 pushed out.