Should the Olympic sprint skaters run the 500 meter twice?

Nils Lid Hjort*

ABSTRACT. The Olympic 500 meter sprint competition is the 'Formula One event' of speed skating, and is watched by millions of television viewers. A draw decides who should start in inner lane and who in outer lane. Many skaters dread the last inner lane, where they need to tackle heavier centrifugal forces than their companions in the last outer lane, at maximum speed around 55 km/hour, at a time when fatigue may set in. The aim of this article is to investigate this potential difference between last inner and last outer lane. For this purpose data from eleven Sprint World Championships 1984–1994 are exploited. A bivariate mixed effects model is used that in addition to the inner-outer lane information takes account of different ice and weather conditions on different days, unequal levels for different skaters, and the passing times for the first 100 meter. The underlying 'unfairness parameter', estimated with optimal precision, is about 0.05 seconds, and is indeed significantly different from zero; it is about three times as large as its estimated standard deviation.

Results from the work reported on here played a decisive role in leading the International Skating Union and the International Olympic Committee to change the rules for the 500 meter sprint event; as of the Nagano 1998 Olympic Games, the sprinters are to skate twice, with one start in inner lane and one in outer lane. The best average result determines the final list, and the best skaters from the first run are paired to skate last in the second run. It has also been decided that the same rules shall apply for the single distance 500 meter World Championships; these are arranged yearly from 1996 onwards.

KEY WORDS AND PHRASES: *combining data sources, Dan Jansen, mixed effects model, Olympic Games, speed skating, Sprint World Championships, unfairness parameter*

1. Background: Is the Olympic 500 meter unfair?

"He drew lane with anxious attentiveness — and could not conceal his disappointment: First outer lane! He threw his blue cap on the ice with a resigned movement, but quickly contained himself and picked it up again. With a start in inner lane he could have set a world record, perhaps be the first man in the world under 40 seconds. Now the record was hanging by a thin thread — possibly the gold medal too. At any rate he couldn't tolerate any further mishaps."

This book excerpt (from Bjørnsen, 1963, Ch. 9) illustrates what has been known for a long time, that even the most accomplished sprint skaters experience difficulties with the last inner lane. (Yevgeni Grishin indeed had his famous technical accident there and found himself stumbling into outer lane, caused by the leather on his left boot touching the ice as he had to lean over at the high speed; he miraculously went on to win the 1960 Squaw Valley Olympic gold medal at 40.2 seconds. Some days later he achieved 39.6, this time with last outer lane.) The last inner lane skaters have to fight a higher acceleration force,

* *Hjort is Professor of Statistics at the University of Oslo, Research Statistician with the Norwegian Computing Centre, and Member of the World's Speed Skating Statisticians' Association.*

since the inner track radius is about 25–26 meter and the outer track about 29–30 meter, at a time when speed is at peak and fatigue may set in; see Figure 1. The acceleration force formula is mv^2/r, where m is mass, v is velocity, and r is radius. Thus a 90 kg skater with top speed 400 meter by 27 seconds meets a force of about 80 kp in inner lane and about 70 kp in outer lane. As a consequence many skaters are not able to keep to the designated curve, glide out towards or even into the outer lane, and in such cases have to skate some extra distance. This phenomenon is particularly prominent on rinks where the ice is fast and the curvature radius minimal, as for the modern indoor rinks. The last inner lane skaters are also more prone to experiencing technical accidents.

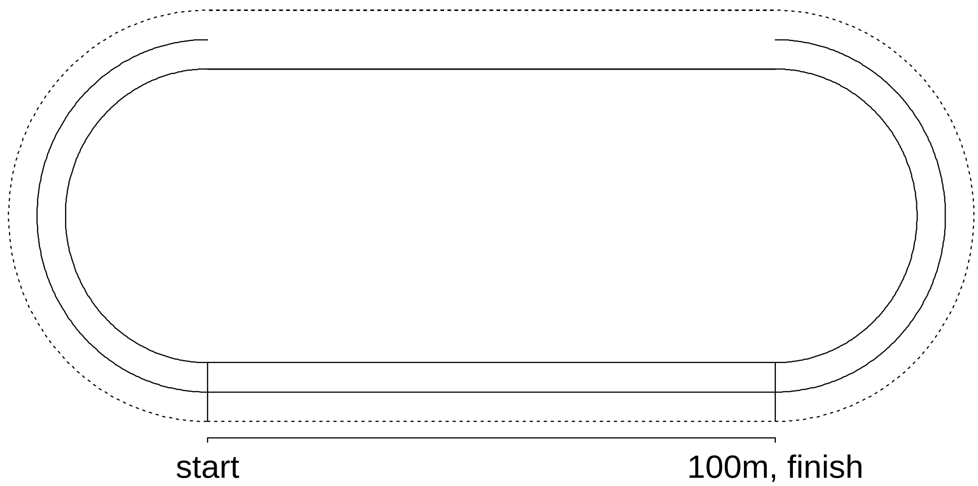

FIGURE 1. *The speed skating rink. The skaters exchange lanes on the back-straight. One lap, comprising one outer and one inner lane, is exactly 400 meter long.*

1.1. THE PRESENT INVESTIGATION. This problem has been recognised and solved in a satisfactory manner since 1975 when it comes to the annual Sprint World Championships events, in that an International Skating Union rule was enforced to make skaters have last inner and last outer lanes on alternate days. But the Olympic event has a potential unfairness built into it since the skaters run only once, and half of them are allotted last inner lane and the other half last outer lane, by chance.

My aim has been to estimate the potential 'unfairness' difference parameter in question as precisely as possible, and to test whether the unfairness is or is not statistically significant. The data I have used consist of the complete 500 meter lists from the eleven Sprint World Championships for Men Trondheim 1984–Calgary 1994. Each skater has a 100 meter passing time and a 500 meter result for the Saturday event, and similarly for the Sunday event, and skaters start in inner and outer tracks on alternative days. (The SWCs also include 1000 meter races on both Saturday and Sunday, and all four runs contribute to the final standing; our present concern lies however with the 500 meter only.)

Agree to define d as the average difference between a result reached by last inner track versus last outer track, for the average top sprinter:

$$d = \text{average difference (last inner track} - \text{last outer track)}. \tag{1.1}$$

This is the statistical parameter we wish to estimate with optimal precision. It is easy to put up some simple estimates based on individual differences, for example, but such estimates will typically lack in statistical precision. I have used a more sophisticated statistical approach, involving a model with ingredients

$$\begin{aligned} 500\,\mathrm{m}_{\mathrm{day}\,1} &= a_1 + b_1 \cdot 100\,\mathrm{m}_{\mathrm{day}\,1} + c_{\mathrm{skater}} \pm \tfrac{1}{2}d + \mathrm{variation}_{\mathrm{day}\,1}, \\ 500\,\mathrm{m}_{\mathrm{day}\,2} &= a_2 + b_2 \cdot 100\,\mathrm{m}_{\mathrm{day}\,2} + c_{\mathrm{skater}} \pm \tfrac{1}{2}d + \mathrm{variation}_{\mathrm{day}\,2}. \end{aligned} \tag{1.2}$$

The $\pm$ here is a plus if the skater has last inner track and a minus if the skater has last outer track. The model and methods used take account of the 100 meter passing times, the individual skater's ability, possible differences in day-to-day conditions, and the inner-outer information.

The model is more formally motivated and described in Section 2, and estimation methods along with analysis of their precision are developed in Section 3. I find it reasonable to interpret d in a somewhat conservative manner, and do not want it to be overly influenced by various extreme or too unusual results; it is meant to be the average difference for 'usual runs' where the skaters are up to their normal best. It is therefore necessary to decide on criteria to define outliers, run times that are subsequently to be removed from the final analysis. Three outlier tests are given in Section 4, which also contains material and analysis that support the fundamental statistical model used.

	$\widehat{d}$	st. error
1994 Calgary:	0.010	0.043
1993 Ikaho:	0.032	0.041
1992 Oslo:	−0.019	0.086
1991 Inzell:	0.023	0.040
1990 Tromsø:	0.096	0.087
1989 Heerenveen:	0.128	0.047
1988 West Allis:	−0.147	0.090
1987 Sainte Foy:	−0.151	0.080
1986 Karuizawa:	0.035	0.066
1985 Heerenveen:	0.090	0.058
1984 Trondheim:	0.131	0.038
grand average:	0.048	0.016

TABLE 1. *Estimates of the difference parameter d for eleven Sprint World Championships for men, along with standard errors (estimated standard deviation) for these.*

1.2. CONCLUSIONS. In view of the potentially serious implications (as measured on the Olympic scale) of the results and their interpretations I decided to gather enough data to reach a standard deviation of the final d estimate of about 0.02 seconds. Preliminary analysis based on a couple of SWCs indicated that this meant including about ten complete

SWC data sets. Section 5 reports more fully on analysis of the eleven SWCs 1984–1994; presently we exhibit an estimate of the d parameter for each, together with their standard errors (estimated standard deviations) and their grand (weighted) average.

The d estimates are seen to vary somewhat from competition to competition, as was expected in view of the sometimes quite different conditions regarding wind, temperature, gliding conditions, and the actual sizes of the inner and outer lane radii. In situations where there is a strong wind against the skater on the backstraight, 'the crossing stretch', as arguably was the case for many competitors in Sainte Foy 1987, West Allis 1988 and Oslo 1992, for example, it may actually be fortuitous to have the last inner course. The d estimates are indeed found to be negative, and the corresponding standard errors highest, for exactly these three events. The ds for other events have values in the range 0.09–0.13 (Trondheim 1984, Heerenveen outdoor 1985, Heerenveen indoor 1989, Tromsø 1990), and in the more modest range 0.01–0.04 (Karuizawa 1986, Inzell 1991, Ikaho 1993 and Calgary 1994). It is a visible feature of the data that the more statistically secure d estimates are those that are high, and typically reflect good and stable conditions with runners attaining their normal high standards. As a case in point, if one excludes 1987 Sainte Foy, 1988 West Allis, 1990 Tromsø and 1992 Oslo, which are the four events with standard errors for d estimates exceeding 0.08, then the grand average estimate for the remaining seven SWCs is 0.065, which is nearly four times its standard error 0.017 (95% confidence interval $[0.032, 0.098]$). The grand (weighted) average estimate for all eleven SWCs is 0.048 seconds, which is about three times its standard error 0.016, and hence statistically significant (p-value 0.001; 95% confidence interval $[0.174, 0.079]$). Further details, along with supplementary analysis and graphical evidence supporting the $d > 0$ conclusion, are given in Section 5.

Folklore knowledge in the stands has it that women are not as bothered with the last inner lane as men sometimes are; they are seen to keep much better to the designated curve, and also seem to have fewer slips and technical accidents than men. This is supported by data; see the discussion of Section 6. Thus the unfairness argument is much weaker in case of the women's event.

How important is the 0.05 seconds difference? At top speed 26 seconds for 400 meters a skater manages an impressive 15.4 meter in 1 second, and about 0.75 meter in 0.05 seconds, which translates into roughly 0.15 percent. It would mean about 15 meters in a 10 000 m run, and about 65 meters in a marathon. So the difference matters! This is also borne out through comparison with real and simulated result lists for the 1994, 1992, 1988 Olympics, given in Appendix I. As alluded to above there are also reasons to believe that the real d number is *larger* for the modern indoor rinks that will host the future Olympic speed skating events, than the grand average value 0.05 arrived at here.

1.3. A PROPOSAL TO THE ISU AND THE IOC. A natural proposal to the International Skating Union and the International Olympic Committee, in view of these findings, is that the skaters should run the 500 meter twice, with one start in inner and one in outer lane, as in the Sprint World Championships. The most natural solution would then be as in alpine events and ski jumping, with a 'reversed starting list', with the best skaters from the first run starting the latest in the second run, subject to correct pairing with respect to inner and outer lane, and with the best average result defining the final ranking. This is a spectator-friendly regime too; large screens would at every stage inform viewers of the current ranking as well as the performances required by the next skaters to reach the top of the list. The average viewers would then easier understand when to be appropriately excited.

1.4. Coda: They said yes. Such a proposal was put forward to the ISU at their June 1992 congress in Davos, along with a brief summary of the work presented here (based on SWCs 1984–1992), through the Norwegian delegation. Also included as handout material to sell the argument were real and simulated 500 meter result lists from the two last Olympic events, where the simulated list in question was the the speculated outcome if the inner lane starters had started in outer lane and vice versa, and computed under the $d = 0.06$ assumption (which was the estimate based on 1984–1992 data; see Appendix I below for such lists for the 1988, 1992, 1994 events, using the $d = 0.05$ figure). It is fair to add that the attraction argument, that the proposed scenario would actually be (even) more spellbinding for spectators, was emphasised as much as the statistical questions related to the unfairness parameter. The meeting decided not to interfere with the already laid plans for the 1994 Lillehammer Olympics, and to reconsider the matter at their next congress in June 1994 in Boston. And at this meeting the representatives of the 34 attending countries (ISU members) after some debate unanimously voted yes to the new proposal, which was this time put forward by the ISU Technical Committee for Speed Skating, to be made effective from the 1998 Nagano Olympics onwards, for both men and women. The new 500 meter rule is also to be made effective for the annual World Championships for single distances, which are introduced as from 1996.

2. Sprint World Championships data and the statistical model

This section describes the data and motivates the statistical model which is used. A simpler model for the differences of finishing times is also considered.

2.1. The bivariate mixed effects model. For illustration, consider data* from the SWC 1994, held in the Olympic Oval, Calgary, Canada. These are of the form given in Table 2; a full listing of the 1984–1994 results is offered in Appendix II.
We represent the inner-outer information as

$$z_{1,i} = \begin{cases} -1 & \text{if no. } i \text{ starts in inner track on day 1,} \\ 1 & \text{if he starts in outer track on day 1,} \end{cases}$$

with a similar $z_{2,i}$ for day 2. Note that $z_{2,i}$ is always $-z_{1,i}$, by the ISU rules for these Championships. Let furthermore $x_{1,i}$ and $Y_{1,i}$ be 100 meter time and finishing 500 meter

* *Data source: SKØYTENYTT ('Speed Skating News'), 5/6/1984, 6/1985, 6/1986, 6/1987, 7 and 8/1988, 6/1989, 5/1990, 7/1991, 7/1992, 5/1993, 3 and 6/1994. This is an international bulletin issued by the World Speed Skating Statisticians' Association about ten times a year, and with contributing associate editors from about 20 countries. Chief editor is Magne Teigen, Veggli, Norway.*

time for skater i on day 1, and similarly $x_{2,i}$ and $Y_{2,i}$ are 100 meter time and finishing 500 meter result for the same skater on day 2.

		first day:			**second day:**	
		100	**500**		**100**	**500**
1. D. Jansen	o	9.82	35.96	i	9.75	35.76
2. S. Klevchena	o	9.78	36.39	i	9.82	36.27
3. J. Inoue	i	9.98	36.43	o	9.76	36.05
4. H. Shimizu	o	9.70	36.35	i	9.77	36.08
5. K. Scott	i	9.96	36.87	o	9.87	36.55
6. I. Zhelezovsky	i	10.08	36.90	o	10.22	36.99
7. T. Kuroiwa	i	10.11	36.83	o	10.07	36.75
8. Y.-M. Kim	o	9.81	36.56	i	9.76	36.53
... and so on ...						

TABLE 2. *Results for the best of the 33 skaters taking part in the Sprint World Championship in Calgary, Candada, January 29–30 1994. Here 'i' and 'o' signify that the skater started in respectively outer and inner lane.*

A natural starting point for the model building procedure is the representation

$$\begin{aligned} Y_{1,i} &= a_1 + b_1 x_{1,i} + c_i + \tfrac{1}{2} d z_{1,i} + e_{1,i}, \\ Y_{2,i} &= a_2 + b_2 x_{2,i} + c_i + \tfrac{1}{2} d z_{2,i} + e_{2,i} \end{aligned} \tag{2.1}$$

for $i = 1, \ldots, n$, where d is the quantity of primary interest, see (1.1); a skater needs the extra amount $\frac{1}{2}d$ if he starts in outer lane compared to the extra amount $-\frac{1}{2}d$ if he starts in inner lane. Further, $e_{i,1}$ and $e_{i,2}$ represent random statistical variation for the two runs, modelled as $2n$ independent terms distributed as $\mathcal{N}\{0, \sigma^2\}$, while c_i represents that particular skater's ability compared to the average level in the competition. Thus (2.1) is the formal version of the model that was hinted at in (1.2). Top skaters Dan Jansen and Igor Zhelezovsky have negative c_is, perhaps around -1, while those found in the second half or so on the final results list have positive c_is, perhaps around 1 for the slower ones. The c_is are not directly observable; they would be poorly estimated based on data from one competition, and although more accurate values for these could be assigned based on extensive data from several competitions, the intention presently is to treat them in the 'random effects' way. Assuming $c_1, \ldots, c_n$ to come from a normal $\{0, \kappa^2\}$, the model takes the form

$$\begin{pmatrix} Y_{1,i} \\ Y_{2,i} \end{pmatrix} \sim \mathcal{N}_2 \Big\{ \begin{pmatrix} a_1 + b_1 x_{1,i} + d w_i \\ a_2 + b_2 x_{2,i} - d w_i \end{pmatrix}, \begin{pmatrix} \sigma^2 + \kappa^2 & \kappa^2 \\ \kappa^2 & \sigma^2 + \kappa^2 \end{pmatrix} \Big\}. \tag{2.2}$$

Here, for convenience,

$$w_i = \tfrac{1}{2} z_{1,i} = -\tfrac{1}{2} z_{2,i} = \begin{cases} 1/2 & \text{if outer start on day 1 and inner start on day 2,} \\ -1/2 & \text{if inner start on day 1 and outer start on day 2.} \end{cases} \tag{2.3}$$

The 'intraclass correlation' $\rho = \kappa^2/(\sigma^2 + \kappa^2)$ parameter is here representing the stability for the average skater.

The Saturday parameters (a_1, b_1) and Sunday parameters (a_2, b_2) are a priori different since gliding and other conditions, like wind and temperature, are often different for the two days. There is an argument favouring $b_1 = b_2$, though, which means that the day-to-day difference in overall conditions can be well explained by the difference $a_2 - a_1$ alone. Suppose for example that the conditions are a bit worse on Sunday. This leads to somewhat larger x_2 values than x_1 values, and somewhat larger Y_2 values than Y_1 values, by about the same factor; the two slopes in question would be approximately the same. In the analysis I therefore put $b_1 = b_2$ in (2.2). Data support this assumption, see Section 4. This reduction from five to four mean parameters is not overly important; results based using all five parameters give very nearly the same results. I have similarly tested whether it is necessary to have different σ_1 and σ_2 parameters in (2.1)–(2.2), but data again have supported the simpler model with $\sigma_1 = \sigma_2$. Another suggestion, in the present quest to employ as few parameters as naturally possible, is to let $a_2 = ka_1$, $b_2 = kb_1$, and $\sigma_2 = k\sigma_1$, with a k factor envisaged to be quite close to 1. The $b_1 = b_2$ parameterisation is however easier and more effective.

2.2. A SIMPLER MODEL BASED ON DIFFERENCES. There is a rather simple alternative way of obtaining an estimate of d, based on the observed differences;

$$\begin{aligned} Y_{2,i} - Y_{1,i} &= a_2 - a_1 + b(x_{2,i} - x_{1,i}) + d(\tfrac{1}{2}z_{2,i} - \tfrac{1}{2}z_{1,i}) + \eta_{2,i} - \eta_{1.i} \\ &\sim \mathcal{N}\{a_2 - a_1 + b(x_{2,i} - x_{1,i}) - 2dw_i, 2\sigma^2\}. \end{aligned} \tag{2.4}$$

Here the individual c_i capability parameters disappear, and one circumvents the need for binormal analysis; ordinary linear regression gives $\widehat{d}_{\rm simple}$, say, in addition to $\widehat{a}_{0,\rm simple}$, $\widehat{b}_{\rm simple}$, and $\widehat{\sigma}_{\rm simple}$, where $a_0 = a_2 - a_1$. The d estimate based on all n pairs will be slightly more precise, however, as discussed in Section 3.3 below. To carry out outlier testing and a part of the final overall analysis, to be discussed in Section 5, it is necessary also to estimate κ, on which the (2.4) differences throw no light. In addition it is of course preferable to have as precise estimates of all parameters involved as possible, and b, in particular, is more precisely estimated in the full model than in the simpler differences model. We also note that the binormal (2.2) model, with $b_1 = b_2 = b$, is equivalent to stochastic independence between difference $Y_{2,i} - Y_{1,i}$ and average $\bar{Y}_i = (Y_{1,i} + Y_{2,i})/2$, and with

$$\begin{aligned} \bar{Y}_i &\sim \mathcal{N}\{\bar{a} + b\bar{x}_i, \kappa^2 + \sigma^2/2\}, \quad \text{where } \bar{x}_i = (x_{1,i} + x_{2,i})/2, \\ Y_{2,i} - Y_{1,i} &\sim \mathcal{N}\{a_2 - a_1 + b(x_{2,i} - x_{1,i}) - 2dw_i, 2\sigma^2\}. \end{aligned} \tag{2.5}$$

3. Parameter estimates and their precision

In this section the parameter estimation procedure for models of the type (2.2) is outlined. We also include analysis of the precision of the parameters, caring particularly about the d estimators.

3.1. ESTIMATION IN THE MIXED EFFECTS MODEL. Suppose in general terms that we have n independent pairs of data

$$\begin{aligned} Y_{1,i} &= x_{1,i}'\beta + f_{1,i} = x_{1,i,1}\beta_1 + \cdots + x_{1,i,p}\beta_p + f_{1,i}, \\ Y_{2,i} &= x_{2,i}'\beta + f_{2,i} = x_{2,i,1}\beta_1 + \cdots + x_{2,i,p}\beta_p + f_{2,i}, \end{aligned}$$

where $\beta = (\beta_1, \ldots, \beta_p)'$ is the vector of regression coefficients, $x_{1,i}$ is a p-covariate vector for $Y_{1,i}$, and $x_{2,i}$ is a p-covariate vector for $Y_{2,i}$. Furthermore, $(f_{1,i}, f_{2,i})'$ is zero mean binormal with covariance κ^2 and variances $\sigma^2 + \kappa^2$; in the notation above $f_{1,i} = c_i + e_{1,i}$ and $f_{2,i} = c_i + e_{2,i}$. In other words,

$$\begin{pmatrix} Y_{1,i} \\ Y_{2,i} \end{pmatrix} \sim \mathcal{N}_2\{\begin{pmatrix} x'_{i,1}\beta \\ x'_{i,2}\beta \end{pmatrix}, \begin{pmatrix} \sigma^2+\kappa^2 & \kappa^2 \\ \kappa^2 & \sigma^2+\kappa^2 \end{pmatrix} = \frac{\sigma^2}{1-\rho}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\},$$

the latter matrix expression being in terms of the intraclass correlation parameter $\rho = \kappa^2/(\sigma^2+\kappa^2)$.

Aiming to find the maximum likelihood (ML) estimators, we start with the log-likelihood for data. It is

$$\log L(\beta,\sigma,\rho) = -2n\log\sigma - \frac{1}{2}n\log\frac{1+\rho}{1-\rho} - \frac{1}{2}\frac{1}{\sigma^2}\frac{Q(\beta)}{1+\rho}, \tag{3.1}$$

where

$$\begin{aligned} Q(\beta) &= \sum_{i=1}^n (Y_i - x'_i\beta)'\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}(Y_i - x'_i\beta) \\ &= \sum_{i=1}^n (Y_{1,i} - x'_{1,i}\beta)^2 + \sum_{i=1}^n (Y_{2,i} - x'_{2,i}\beta)^2 - 2\rho\sum_{i=1}^n (Y_{1,i} - x'_{1,i}\beta)(Y_{2,i} - x'_{2,i}\beta) \\ &= Q_1(\beta) + Q_2(\beta) - 2\rho Q_3(\beta). \end{aligned}$$

We are to find the maximisers $\widehat{\beta}$, $\widehat{\sigma}$, $\widehat{\rho}$ of (3.1). The one for β must be $\widehat{\beta} = \widehat{\beta}(\widehat{\rho})$, where $\widehat{\beta}(\rho)$ is found from minimisation of $Q(\beta)$, i.e.

$$\widehat{\beta}(\rho) = \{M_{11} + M_{22} - \rho(M_{12} + M_{21})\}^{-1}\{S_{11} + S_{22} - \rho(S_{12} + S_{21})\}, \tag{3.2}$$

in which

$$M_{uv} = \frac{1}{n}\sum_{i=1}^n x_{u,i}x'_{v,i} \quad \text{and} \quad S_{uv} = \frac{1}{n}\sum_{i=1}^n x_{u,i}Y_{v,i}, \quad u,v = 1,2.$$

And the ML estimators for σ and ρ are found from maximising $\log L(\widehat{\beta}(\rho),\sigma,\rho)$. Taking partial derivatives w.r.t. σ and ρ gives two equations that must be obeyed:

$$\widehat{\sigma}^2(\rho) = \frac{1}{1+\rho}\frac{Q(\widehat{\beta}(\rho))}{2n} \quad \text{and} \quad \widehat{\sigma}^2(\rho) = \frac{1-\rho}{1+\rho}\frac{Q_1(\widehat{\beta}(\rho)) + Q_2(\widehat{\beta}(\rho)) + 2Q_3(\widehat{\beta}(\rho))}{2n}. \tag{3.3}$$

Finding the ML estimator $\widehat{\rho}$ is accomplished either via maximisation of the log-likelihood profile function

$$\begin{aligned} \log L(\widehat{\beta}(\rho),\widehat{\sigma}(\rho),\rho) &= n\Big[-\log\Big\{\frac{1}{1+\rho}\frac{Q(\widehat{\beta}(\rho))}{2n}\Big\} + \frac{1}{2}\log\frac{1-\rho}{1+\rho} - 1\Big] \\ &= n\big[\tfrac{1}{2}\log(1-\rho^2) - \log\{Q(\widehat{\beta}(\rho))/2n\} - 1\big], \end{aligned} \tag{3.4}$$

or from solving for the two expressions in (3.3) to be equal. Note also that

$$\rho = \frac{2Q_3(\widehat\beta(\rho))}{Q_1(\widehat\beta(\rho)) + Q_2(\widehat\beta(\rho))} \tag{3.5}$$

at the parameter point that solves (3.3), i.e. for the ML solution. This fits the fact that $Q_1(\beta)$, $Q_2(\beta)$, $Q_3(\beta)$ have expected values $\sigma^2/(1-\rho)$, $\sigma^2/(1-\rho)$, and $\sigma^2\rho/(1-\rho)$, at the true model. Solving (3.5) for ρ is yet another method for carrying out the ML estimation.

We will actually use a sample-size modification for $\widehat\sigma$. Assuming for a moment that the value of ρ is known, one shows easily that $\widehat\beta(\rho)$ of (3.2) is normal, unbiased, and with covariance matrix $n^{-1}\sigma^2(1+\rho)M_\rho^{-1}$, where

$$M_\rho = M_{11} + M_{22} - \rho(M_{12} + M_{21}).$$

Furthermore arguments can be furnished, using orthogonalisation techniques and properties of the binormal distribution, to demonstrate that

$$2n\,\widehat\sigma^2(\rho)/(1+\rho) = Q(\widehat\beta(\rho))/(1+\rho) \sim 2\sigma^2\chi^2_{2n-p},$$

and that $Q(\widehat\beta(\rho))$ is statistically independent of $\widehat\beta(\rho)$. In particular this invites using the modified estimator

$$\widehat\sigma^2_{\rm un} = \frac{1}{1+\widehat\rho}\frac{Q(\widehat\beta(\widehat\rho))}{2n-p} = \frac{2n}{2n-p}\widehat\sigma^2\,; \tag{3.6}$$

it is slightly larger than the ML estimate $\widehat\sigma^2$, to take estimation variability of the p regression coefficients into account. We will also similarly use $\widehat\kappa^2_{\rm un} = \widehat\sigma^2_{\rm un}\widehat\rho/(1-\widehat\rho)$ as a sample-size corrected version of the ML estimator for κ^2. The argument is not disturbed by the insertion of $\widehat\rho$ for ρ, since we show in a minute that $\widehat\rho$ is approximately independent of $\widehat\beta = \widehat\beta(\widehat\rho)$.

3.2. Precision of estimates. Note that classic regression theory distributional results do not hold here, for example, $\widehat\beta$ is not quite normally distributed since the random $\widehat\rho$ is inserted. But traditional distributional approximations for ML estimators can be appealed to and implies that $(\widehat\beta, \widehat\sigma, \widehat\rho)$ is approximately jointly normally distributed with the correct mean vector and covariance matrix J^{-1}/n, where J is minus the mean of the normalised and twice differentiated log-likelihood, calculated at the true parameter values. With some efforts one finds

$$J = \begin{pmatrix} M_\rho/\{\sigma^2(1+\rho)\} & 0 & 0 \\ 0 & 4/\sigma^2 & 2/\{\sigma(1-\rho^2)\} \\ 0 & 2/\{\sigma(1-\rho^2)\} & 2/(1-\rho^2)^2 \end{pmatrix}.$$

This means that

$$\widehat\beta \approx \mathcal{N}_p\{\beta, n^{-1}\,\sigma^2(1+\rho)M_\rho^{-1}\}, \tag{3.7}$$

that

$$\begin{pmatrix}\widehat\sigma \\ \widehat\rho\end{pmatrix} \approx \mathcal{N}_2\Big\{\begin{pmatrix}\sigma \\ \rho\end{pmatrix}, \frac{1}{n}\begin{pmatrix} \frac12\sigma^2 & -\frac12\sigma(1-\rho^2) \\ -\frac12\sigma(1-\rho^2) & (1-\rho^2)^2 \end{pmatrix}\Big\},$$

and that $\widehat{\beta}$ and $(\widehat{\sigma}, \widehat{\rho})$ become independent for large n. Confidence intervals and tests can now be furnished in the usual fashion. We comment specifically on this below for the case of d.

The (3.7) result continues to be a very good approximation even if the exact conditions of the normal model (2.2) should be violated. This follows from standard large-sample theory. The distributional results for $\widehat{\sigma}$ and $\widehat{\rho}$ might have to be adjusted for nonnormality of $(Y_{1,i}, Y_{2,i})$; the more generally correct approximate variance of $\widehat{\sigma}$ can for example be shown to be $n^{-1}\sigma^2(\frac{1}{2} + \frac{1}{4}\mathrm{kurt})$, where kurt is the kurtosis of $Y_{1,i} - Y_{2,i}$. In any case this would not have seriously hampered our analysis, which is concerned primarily with d and the other mean parameters, but it is comforting that in fact no serious departure from normality could be detected in our data (as soon as the quite few outliers were removed), as further commented on in Section 4.

3.3. Comparison of two d estimators. The binormal mixed effects model used is that of (2.2), with mean parameter vector $(a_1, a_2, b, d)'$ and with covariate vectors $(1, 0, x_{1,i}, w_i)'$ and $(0, 1, x_{2,i}, -w_i)'$ associated with skater i. The (3.7) result is that $(\widehat{a}_1, \widehat{a}_2, \widehat{b}, \widehat{d})$ has covariance matrix approximately equal to $n^{-1}\sigma^2(1+\rho)M_\rho^{-1}$. In this situation,

$$M_{11} = \begin{pmatrix} 1 & 0 & \mathrm{ave}(x_1) & \mathrm{ave}(w) \\ 0 & 0 & 0 & 0 \\ \mathrm{ave}(x_1) & 0 & \mathrm{ave}(x_1^2) & \mathrm{ave}(x_1 w) \\ \mathrm{ave}(w) & 0 & \mathrm{ave}(x_1) & \mathrm{ave}(w^2) \end{pmatrix},$$

$$M_{22} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \mathrm{ave}(x_2) & \mathrm{ave}(-w) \\ 0 & \mathrm{ave}(x_2) & \mathrm{ave}(x_2^2) & \mathrm{ave}(-x_2 w) \\ 0 & \mathrm{ave}(-w) & \mathrm{ave}(-x_2 w) & \mathrm{ave}(w^2) \end{pmatrix},$$

$$M_{12} = \begin{pmatrix} 0 & 1 & \mathrm{ave}(x_2) & \mathrm{ave}(-w) \\ 0 & 0 & 0 & 0 \\ 0 & \mathrm{ave}(x_1) & \mathrm{ave}(x_1 x_2) & \mathrm{ave}(-x_1 w) \\ 0 & \mathrm{ave}(w) & \mathrm{ave}(x_2 w) & \mathrm{ave}(-w^2) \end{pmatrix},$$

$$M_{21} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & \mathrm{ave}(x_1) & \mathrm{ave}(w) \\ \mathrm{ave}(x_2) & 0 & \mathrm{ave}(x_1 x_2) & \mathrm{ave}(x_2 w) \\ \mathrm{ave}(-w) & 0 & \mathrm{ave}(-x_1 w) & \mathrm{ave}(-w^2) \end{pmatrix},$$

where $\mathrm{ave}(x_1)$ is the average $n^{-1}\sum_{i=1}^n x_{1,i}$, and so on. These can be computed for the given covariates x_1, x_2, w, and this is what is used to produce standard errors (estimated standard deviations) for $\widehat{d}$ in the SWCs; see Section 4 and Table 1 of Section 1.

The statistical analysis takes place conditional on covariates x_1, x_2, w, but it is also illuminating to study simpler approximations, based on 'typical behaviour' of these, whose values also can also be viewed as outcomes of random mechanisms. In such a framework it is clear that w behaves stochastically independent of x_1 and x_2, and that the symmetric $\pm\frac{1}{2}$ variable w has mean zero; in fact $\mathrm{ave}(w) = 0$ if n is even and $\pm\frac{1}{2}/n$ if n is odd. It follows that $\mathrm{ave}(x_1 w)$ and $\mathrm{ave}(x_2 w)$ will both be close to zero. Using these arguments it follows that

$$M_\rho \approx \begin{pmatrix} 1 & -\rho & \mathrm{ave}(x_1-\rho x_2) & 0 \\ -\rho & 1 & \mathrm{ave}(x_2-\rho x_1) & 0 \\ \mathrm{ave}(x_1-\rho x_2) & \mathrm{ave}(x_2-\rho x_1) & \mathrm{ave}(x_1^2+x_2^2-2\rho x_1x_2) & 0 \\ 0 & 0 & 0 & \frac12+\frac12\rho \end{pmatrix}.$$

In particular this demonstrates that

$$\mathrm{Var}\,\widehat{d} \approx n^{-1}\sigma^2(1+\rho)\,2/(1+\rho) = 2\sigma^2/n. \tag{3.8}$$

The simpler estimate $\widehat{d}_{\mathrm{simple}}$ based on differences, as in (2.4), can be analysed similarly. It emerges by ordinary linear regression of $Y_{2,i}-Y_{1,i}$ against $x_{2,i}-x_{1,i}$ and $-2w_i$. The covariance matrix for say $(\widehat{a}_{0,\mathrm{simple}},\widehat{b}_{\mathrm{simple}},\widehat{d}_{\mathrm{simple}})$ constructed in this fashion, where $a_0 = a_2-a_1$, is equal to

$$\begin{aligned} &\frac{2\sigma^2}{n}\begin{pmatrix} 1 & \mathrm{ave}(x_2-x_1) & \mathrm{ave}(-2w) \\ \mathrm{ave}(x_2-x_1) & \mathrm{ave}((x_2-x_1)^2) & \mathrm{ave}(-2w(x_2-x_1)) \\ \mathrm{ave}(-2w) & \mathrm{ave}(-2w(x_2-x_1)) & \mathrm{ave}(4w^2) \end{pmatrix}^{-1} \\ &\qquad \approx \frac{2\sigma^2}{n}\begin{pmatrix} 1 & \mathrm{ave}(x_2-x_1) & 0 \\ \mathrm{ave}(x_2-x_1) & \mathrm{ave}((x_2-x_1)^2) & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}. \end{aligned}$$

It follows that $\widehat{d}_{\mathrm{simple}}$ has approximately the very same precision as the more complicated $\widehat{d}$ that builds more explicitly on the binormal mixed effects model. These are approximations, however, and the $\widehat{d}$ will have a slight edge in each application, since the differences-based simpler method builds on less information.

We also note that the b parameter will typically be better estimated in the mixed model than in the differences model. Calculations as above show that $\widehat{b}$ has $(\sigma^2/2n)(1+\rho)/\tau^2$ while $\widehat{b}_{\mathrm{simple}}$ has $(\sigma^2/\tau^2)/n$ as approximate variances, where τ^2 is empirical variance of xs. Thus the ratio $\mathrm{Var}\,\widehat{b}_{\mathrm{simple}}/\mathrm{Var}\,\widehat{b}$ is $2/(1+\rho)$. In the same vein it can be shown that $\widehat{\sigma}_{\mathrm{un}}$ and the $\widehat{\sigma}_{\mathrm{simple}}$, available by regression on (2.4), have approximately the same precision.

4. Deciding on outliers, and validating the model

As mentioned in the introduction we would not wish to see d overly influenced by unusual results. Minor slips or accidents occur frequently in this technically demanding sport, and easily cause losses of tenths of a second; we envisage our d as the average difference for normal runs without such mishaps. This is one of several reasons favouring a robust analysis of the model and in particular a statistically robust estimate of d. While various robust procedures are available, I opt for the conceptually simple method of detecting outliers first and removing these from the final analysis. This is also quite reasonable in view of the fact that data convincingly support the parametric model (2.2), as discussed below.

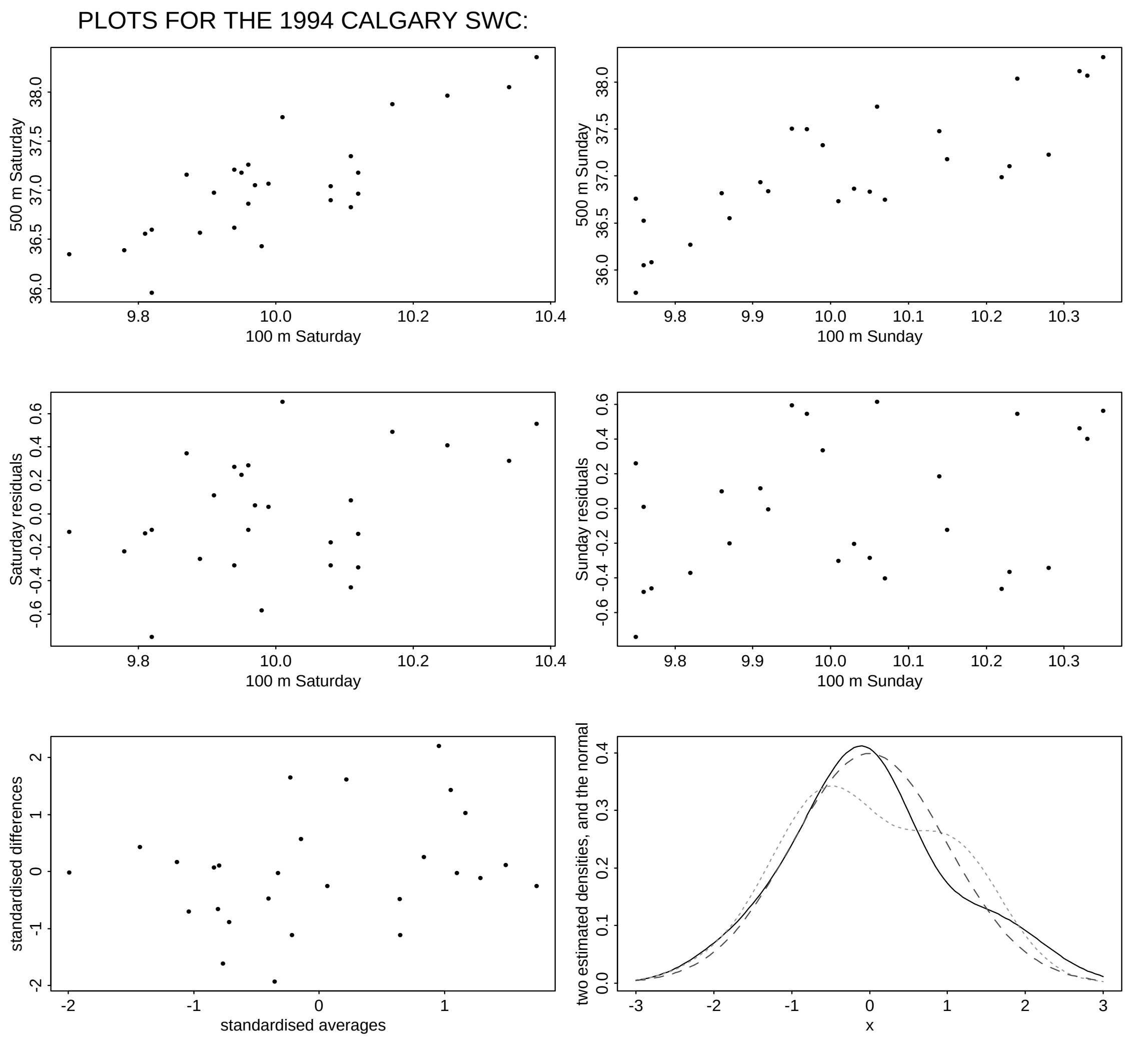

FIGURE 2 (a)–(f). *Plots to validate the mixed effects statistical model, using results from the 1994 Calgary Sprint World Championship. Plots (a) and (b) give Y_i versus x_i for each day and show acceptable linearity (as well as Dan Jansen's World Record 35.76 in plot (b) and Hiroyasu Shimizu's 9.70 in plot(a)). Plots (c) and (d) display Saturday and Sunday residuals against 100 meter passing times, and support the assumption of constant variability across x values. In (e) standardised differences diff_i^* are plotted against standardised averages ave_i^*, see (4.4), with non-significant correlation 0.249. Finally (f) gives nonparametrically estimated densities of the diff_i^*s (fully drawn curve) and of the ave_i^*s (dotted curve), using the traditional kernel method, together with the standard normal curve.*

4.1. OUTLIER CRITERIA. Let $r_{1,i}$ and $r_{2,i}$ the the lap times for the skater's last 400 meter. These can be expressed as

$$\begin{aligned} r_{1,i} &= Y_{1,i} - x_{1,i} = a_1 + (b-1)x_{1,i} + dw_i + c_i + e_{1,i}, \\ r_{2,i} &= Y_{2,i} - x_{2,i} = a_2 + (b-1)x_{2,i} - dw_i + c_i + e_{1,i}, \end{aligned} \tag{4.1}$$

It is natural to discard cases where one of these are too large, in comparison with expected normal behaviour, and also cases where the absolute difference $|r_{2,i} - r_{1,i}|$ between the skater's two lap times is too large. This statistical safety net will catch skaters who have had minor or not so minor technical slips in one of his or her two runs.

The first outlier criterion is developed as follows. Based on all data from a two-day SWC event, excluding only those who fell or were disqualified, estimates are computed of (a_1, a_2, b, d) and of (σ, κ). Then we form the outlier test statistics

$$
\begin{aligned}
t_{1,i} &= \{Y_{1,i} - (\widehat{a}_1 + \widehat{b}x_{1,i} + \tfrac{1}{2}\widehat{d}z_{1,i})\}/(\widehat{\sigma}^2_{\rm un} + \widehat{\kappa}^2_{\rm un})^{1/2}, \\
t_{2,i} &= \{Y_{2,i} - (\widehat{a}_2 + \widehat{b}x_{2,i} + \tfrac{1}{2}\widehat{d}z_{2,i})\}/(\widehat{\sigma}^2_{\rm un} + \widehat{\kappa}^2_{\rm un})^{1/2},
\end{aligned}
\tag{4.2}
$$

both of which are approximately a standard normal if the skater had a 'normal' run without accidents. Slightly more accurate denominators could be constructed here, dividing by an estimate of the exact standard deviation rather than with the approximate one, but the difference is unsubstantial for the present purposes. We deem a case 'above normal bounds' if $t_{1,1}$ or $t_{2,i}$ exceed 2.75. Note that we allow cases with unusually *low* values of t, since these correspond to extraordinary *good* performances.

The second outlier criterion emerges by looking at normal behaviour of the difference between lap times. By (4.1) this leads to constructing

$$
\begin{aligned}
t_{3,i} &= \big[r_{2,i} - r_{1,i} - \{\widehat{a}_2 - \widehat{a}_1 + (\widehat{b} - 1)(x_{2,i} - x_{1,i}) - 2\widehat{d}w_i\}\big]/\sqrt{2}\widehat{\sigma}_{\rm un} \\
&= \{(Y_{2,i} - \widehat{a}_2 - \widehat{b}x_{2,i} - \widehat{d}w_i) - (Y_{1,i} - \widehat{a}_1 - \widehat{b}x_{2,i} + \widehat{d}w_i)\}/\sqrt{2}\widehat{\sigma}_{\rm un}.
\end{aligned}
\tag{4.3}
$$

Again these should be approximately distributed as standard normals if the skater's two runs were 'normal'. A case is deemed 'outside normal bounds' if $|t_{3,i}| \ge 2.75$.

4.2. Model validation. The adequacy of the basic model was checked against various potential violations, and, in short, no audible objections were raised by the data.

Plots of Y_i versus x_i for each day, and for each competition, agreed well with the assumed linearity. Plots of Saturday and Sunday residuals against respectively $x_{1,i}$ and $x_{2,i}$ revealed no departure from the assumed constant level of variability hypothesis. The simultaneous aspects of the (2.2) model were checked via the reformulated equivalent (2.5) model. Scatters were plotted and correlations computed to check for possible dependencies between averages and differences, and again these supported the mean and variance/covariance structure of the model. See Figure 2 (a)–(e) which illustrate these features for the case of the 1994 Calgary Championship.

Coming finally to checking the hypothesised bi-Gaußian distribution, this could be separated into one-dimensional normality assessment of averages and differences, as with (2.5). We have pointed out already, in Section 3.2, that deviations from normality is of no great concern as far as the d analysis is concerned, and that such deviations, specifically in the form of kurtosis values different from the zero predicted by normality, at most could cause mild concern for precision of the σ and κ estimates. However, the normal distribution fits nicely. For purposes of plotting and for comparison over different Championships it is convenient to standardise these, as

$$
\begin{aligned}
{\rm ave}^*_i &= \{\bar{Y}_i - (\widehat{\bar{a}} + \widehat{b}\bar{x}_i)\}/(\widehat{\kappa}^2_{\rm un} + \widehat{\sigma}^2_{\rm un}/2)^{1/2}, \\
{\rm diff}^*_i &= \{Y_{2,i} - Y_{1,i} - (\widehat{a}_2 - \widehat{a}_1 + \widehat{b}(x_{2,i} - x_{1,i}) - 2\widehat{d}w_i)\}/\sqrt{2}\widehat{\sigma}_{\rm un}.
\end{aligned}
\tag{4.4}
$$

These have zero mean and variance nearly equal to 1, and under the model hypothesis these should behave independently of each other and each be approximately standard normal. Figure 2(e) displays the 28 pairs of these variables for the 1994 SWC in Calgary (of the 33 participants, three had falls, and two were declared outliers, by the criteria above; see Appendix II), and Figure 2(f) shows smooth estimates of their densities, comparing them also with the standard normal. The fit is quite acceptable. The same was observed on the basis of quantile-quantile plots against the standard normal. The skewnesses were 0.324 and 0.269, and the kurtosis values were -0.149 and -0.703, all values well within normal range. (Approximate 90% ranges for the skewness and kurtosis, with 28 normal data points, are approximately ± 0.76 and ± 1.52.) The observed correlation is 0.249 and is also within the 90% expected range (which is approximately ± 0.31). Similar plots were made and coefficients computed for the other Championships as well, and again no challenges to the model assumptions were made.

4.3. How many SWCs should I analyse? The present investigation is a statistical detective search for a quite tiny parameter, in a way looking for the odd meter in 500 with a statistician's magnifying glass. We cannot expect to be able to declare a positive d with data from only one or two Championships. There is accordingly a question of gathering enough data to create a reasonable statistical power for the envisioned size of d, say about 0.05 seconds. From (3.8) it emerges that the optimally combined estimator $\widehat{d}_{\rm grand}$, taken over say K SWCs, would have a variance of the form $\{\sum_{j=1}^K n_j/(2\sigma_j^2)\}^{-1}$, with n_j skaters in the jth SWC. As an approximation this is the same as $2\sigma^2/N$, where $N = \sum_{j=1}^K n_j$, and analysis from the first couple of SWCs data sets I used suggested that $\sigma \approx 0.25$. As a rough guide, therefore, about 300 skaters' paired runs were necessary to gather, in order to form a grand estimate with standard error of size 0.02 seconds (95% confidence interval width around 0.08 seconds). Since around 30 skaters compete in each SWC I needed to go through about ten complete SWCs to achieve the necessary precision. A standard error of 0.02 seconds would give detection probability around 80% for a true $d = 0.05$, and around 90% for $d = 0.06$.

5. Analysing the Sprint World Championships 1984–1994

This section summarises analysis carried out for each of the eleven Championships 1984–1994, using methods developed in Section 3. Only skaters who passed all three outlier tests above were included.

First we give a table of all required parameter estimates, for each of the eleven situations. The d estimate column is identical to the list also given in Table 1 of Section 1, where also standard errors were given. Interpretation of the parameters is discussed in Sections 2 and 3. Again we note that the few cases where the d estimate actually is negative are also characterised by higher than normal values for σ, in particular, which means higher variability around each skater's normal level. On the occasions in question a tentative explanation lies with the partly severe and variable weather conditions that met the participants (Sainte Foy 1987, West Allis 1988, Oslo 1992). The cases with more than a tiny difference between a_1 and a_2, like for Sainte Foy 1987 and Tromsø 1990, are the events where Saturday and Sunday conditions differed markedly.

	a_1	a_2	b	d	ρ	σ	κ
1994:	16.984	16.938	2.007	0.010	0.838	0.156	0.355
1993:	18.861	18.998	1.892	0.032	0.799	0.158	0.315
1992:	13.541	13.369	2.494	−0.019	0.538	0.308	0.332
1991:	13.184	13.060	2.462	0.023	0.800	0.155	0.309
1990:	15.137	14.552	2.371	0.096	0.338	0.327	0.234
1989:	3.595	3.475	3.385	0.128	0.720	0.177	0.284
1988:	22.810	22.469	1.670	−0.147	0.701	0.335	0.513
1987:	18.461	17.823	2.056	−0.151	0.553	0.326	0.362
1986:	16.950	17.009	2.101	0.035	0.519	0.256	0.266
1985:	13.093	12.803	2.537	0.090	0.517	0.219	0.227
1984:	17.507	17.136	2.088	0.131	0.847	0.134	0.314

TABLE 3. *Estimates of the four mean value parameters a_1, a_2, b, d and of the three variance-covariance parameters ρ, σ, κ, for each of the eleven Sprint World Championships for men, 1984–1994.*

The primary method for testing $d = 0$ is simply to consider the natural overall estimate and compare its value to its standard error. If $\widehat{d}_j$ is the estimate in year j, with standard error se_j, then

$$\widehat{d} = \sum_{j=1984}^{1994} \widehat{d}_j/\mathrm{se}_j^2 \Big/ \sum_{j=1984}^{1994} 1/\mathrm{se}_j^2 \tag{5.1}$$

is the optimal combination, with precision given by $\mathrm{Var}\,\widehat{d} = (\sum_{j=1984}^{1994} 1/\mathrm{se}_j^2)^{-1}$. (This variance calculation is exactly valid in the case of fully known values for the se_js; that these are estimated rather than fully known does however only cause a very modest second order level increase in the variance.) These are the formulae that produced the grand estimate 0.048 with standard error 0.016 in Table 1 of Section 1. Again, this is significant with p-value 0.001.

It is also illuminating to present evidence against the $d = 0$ hypothesis in terms of direct as well as suitably corrected and standardised differences in lap times. The lap time difference for skater i is $r_{2,i} - r_{1,i} = a_2 - a_1 + (b-1)(x_{2,i} - x_{1,i}) + e_{2,i} - e_{1,i}$, if $d = 0$, so that the variable

$$\begin{aligned} D_i &= r_{2,i} - r_{1,i} - \{\widehat{a}_2 - \widehat{a}_1 + (\widehat{b} - 1)(x_{2,i} - x_{1,i})\} \\ &= (Y_{2,i} - \widehat{a}_2 - \widehat{b}x_{2,i}) - (Y_{1,i} - \widehat{a}_1 - \widehat{b}x_{1,i}) \end{aligned} \tag{5.2}$$

represents adjusted lap time difference for each skater, where ‘adjusted’ means relative to varying ice and weather conditions on the two days and information contained in the 100 meter passing times, but *not* adjusted for inner-outer lane information. The D_is are directly interpretable on the original time scale in seconds. Figure 3 gives density estimates for D_is observed for two groups of skaters, for each of the SWCs 1984–1994. The first group is the one with $(z_{1,i}, z_{2,i}) = (1, -1)$, or $w_i = \frac{1}{2}$, and has the presumed preferable last outer lane on the second day; the complimentary group has $(z_{1,i}, z_{2,i}) = (-1, 1)$, or $w_i = -\frac{1}{2}$, with last outer lane on the first day. Accordingly, *if* there is unfairness in the expected direction (d positive), then the D_is for the $w_i = \frac{1}{2}$ group can be expected to lean slightly to the left (mean value around $-d$), while the D_is for the $w_i = -\frac{1}{2}$ group would tend to lean slightly to the right (mean value around d). In Figure 3 skaters who fell or were disqualified were eliminated, as were those failing one or more of the three outlier tests described in Section 4. When presenting these pairs of densities I also went to the trouble of calculating the D_is with recomputed estimates $(\widehat{a}_1, \widehat{a}_2, \widehat{b})$ under the $d = 0$ hypothesis.

Figure 3 with its eleven parts is found on pages 24

FIGURE 3. *Nonparametric kernel-method density estimates for the modified difference variables D_i, for the $w_i = 1/2$ group (fully drawn) and for the $w_i = -1/2$ group (dotted line), for each of the eleven SWCs 1984–1994. The advantage of having last outer lane is arguably prominent for the 1991, 1990, 1989, 1986, 1985, 1984 occasions, undecided for the 1994, 1993, 1992 events, while the advantage seems to have been with the last inner lane in 1988 and 1987. The number of skaters contributing to the eleven sub-figures are respectively 26, 29, 30, 33, 28, 29, 28, 32, 26, 30, 27, for the years 1984–1994.*

These figures give good visual impressions of the tentative differences between the inner and outer lane situations, for each particular SWC 1984–1994. They also reveal information about variability level, corresponding to the $\widehat{\sigma}_{\mathrm{un}}$ values given in Table 3 above; Tromsø 1990, West Allis 1988 and Sainte Foy 1987 were quite variable occasions (again, explainable by weather and gliding conditions), whereas Trondheim 1994, Heerenveen indoor 1989 and Inzell 1991, for example, were 'cleaner' occasions with less variability around each skater's normal capacity level. Figure 4 presents essentially the same information in another way, by ordinary data dot plots.

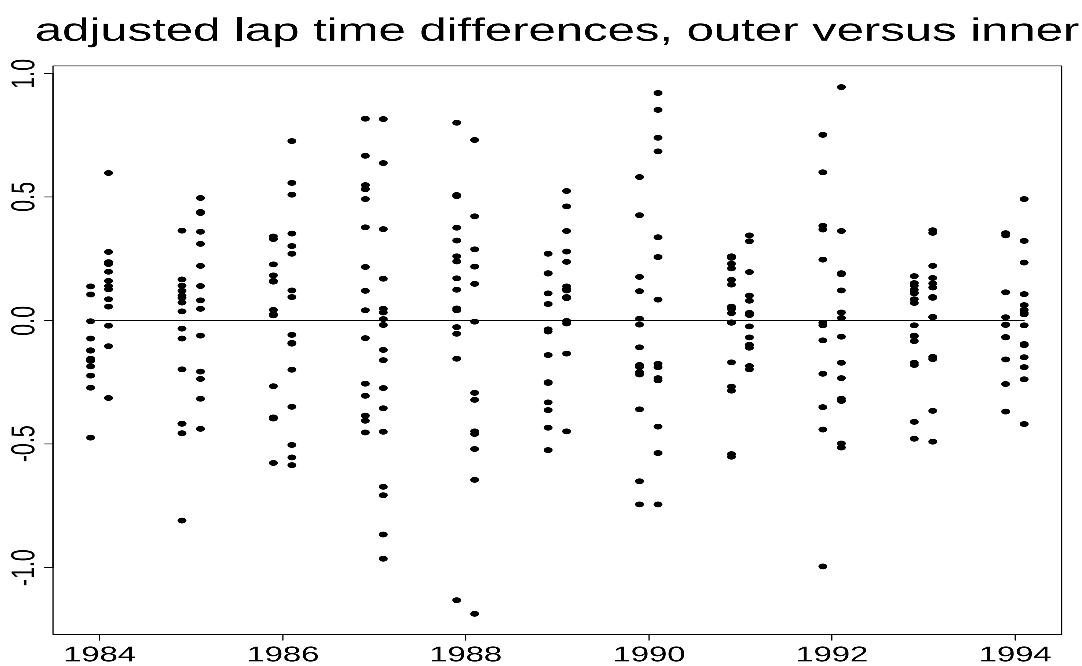

FIGURE 4. *Plots of the modified difference variables D_i, for the $w_i = 1/2$ group (to the left) and for the $w_i = -1/2$ group (to the right), for each of the eleven SWCs 1984–1994. See also the legend to Figure 3.*

A quite informative overall picture can also be formed by comparing the all in all 159 skaters who had $w_i = \frac{1}{2}$ with the 159 skaters who had $w_i = -\frac{1}{2}$. The comparison is most meaningful if the D_is are all scaled with the appropriate estimated standard deviation, which varies from year to year. Figure 5 therefore presents densities for the two groups, of the adjusted and standardised variables

$$\begin{aligned} D_i^* = D_i/\sqrt{2}\widehat{\sigma}_{\rm un} &= \big[r_{2,i} - r_{1,i} - \{\widehat{a}_2 - \widehat{a}_1 + (\widehat{b} - 1)(x_{2,i} - x_{1,i})\}\big]/\sqrt{2}\widehat{\sigma}_{\rm un} \\ &= \{(Y_{2,i} - \widehat{a}_2 - \widehat{b}x_{2,i}) - (Y_{1,i} - \widehat{a}_1 - \widehat{b}x_{1,i})\}/\sqrt{2}\widehat{\sigma}_{\rm un}, \end{aligned} \tag{5.3}$$

with the appropriate year-specific $\widehat{\sigma}_{\rm un}$. Again the D_i^*s that were used employed re-estimated versions of $\widehat{a}_1$, $\widehat{a}_2$, $\widehat{b}$, $\widehat{\sigma}_{\rm un}$, all arrived at under the $d = 0$ hypothesis. The D_i^*s with $w_i = \frac{1}{2}$ should lean somewhat to the left (mean value around $-d/\sqrt{2}\sigma$) while those with $w_i = -\frac{1}{2}$ should lean the other way (mean value around $d/\sqrt{2}\sigma$), again, *if* d indeed is positive. The figure seems to support this, in view of the large sample sizes for the two group. If my reader agrees that the $w_i = \frac{1}{2}$ curve gives slightly but markedly more probability mass to the left hand side than does the $w_i = -\frac{1}{2}$ curve, then he or she might endorse the change in rule towards asking the skaters to sprint twice.

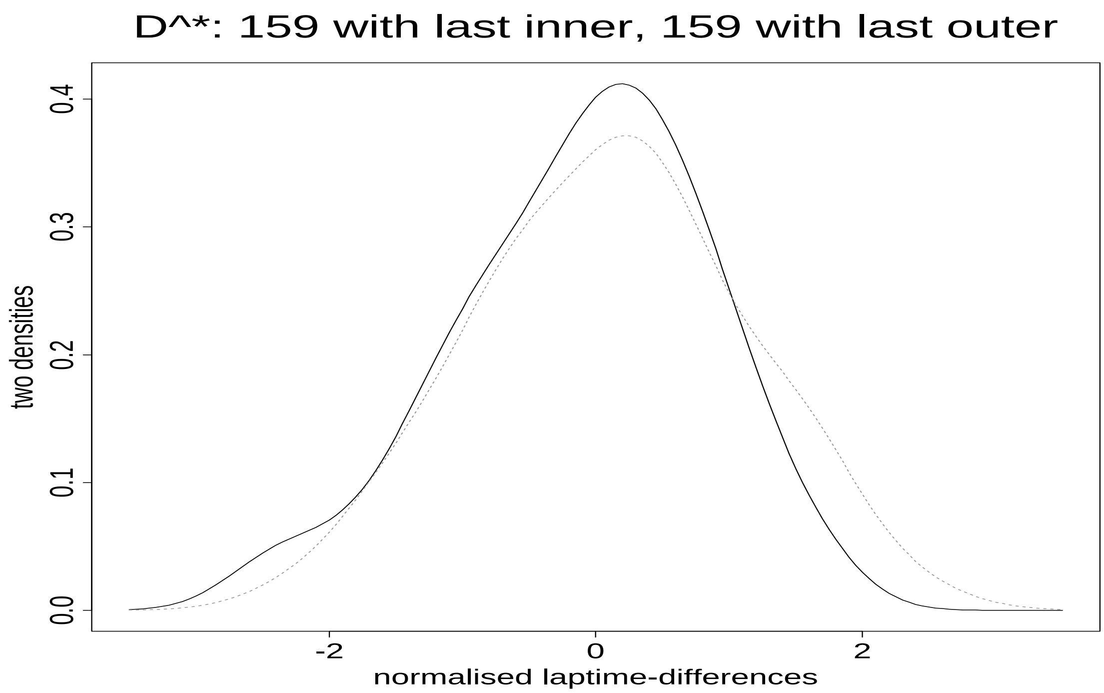

FIGURE 5. *Density estimates for the two groups of adjusted and standardised difference variables D_i^* of (5.3). The fully drawn curve is that of the 159 skaters who had last outer lane at the Sunday event while the dotted curve is that of the complimentary 159 skaters who had last outer lane on Saturday. The former gives more probability mass to the left, indicating that the last outer lane is advantageous.*

6. Women

Women are different from men, aerodynamically speaking. As hinted at in the introduction section the best women sprinters do not seem to be hampered as much as many male sprinters with high speed in the last inner lane, and also offer spectators far fewer spectacular high speed falls (there were only three 500 meter-falls in eleven SWCs among the women, including Christa Rothenburger's in 1985, where she still managed to win the Championship, but 16 in the same period among the men). I nevertheless went ahead to collect and analyse the same amount of data on the women's runs, adhering to the Equal Statistics for Women commandment, to estimate also their seemingly less significant d parameter. The results were as follows, as far as the d estimation is concerned. Note that the SWCs are always held jointly for men and women, on the same days, so the same weather and ice quality conditions reign over both.

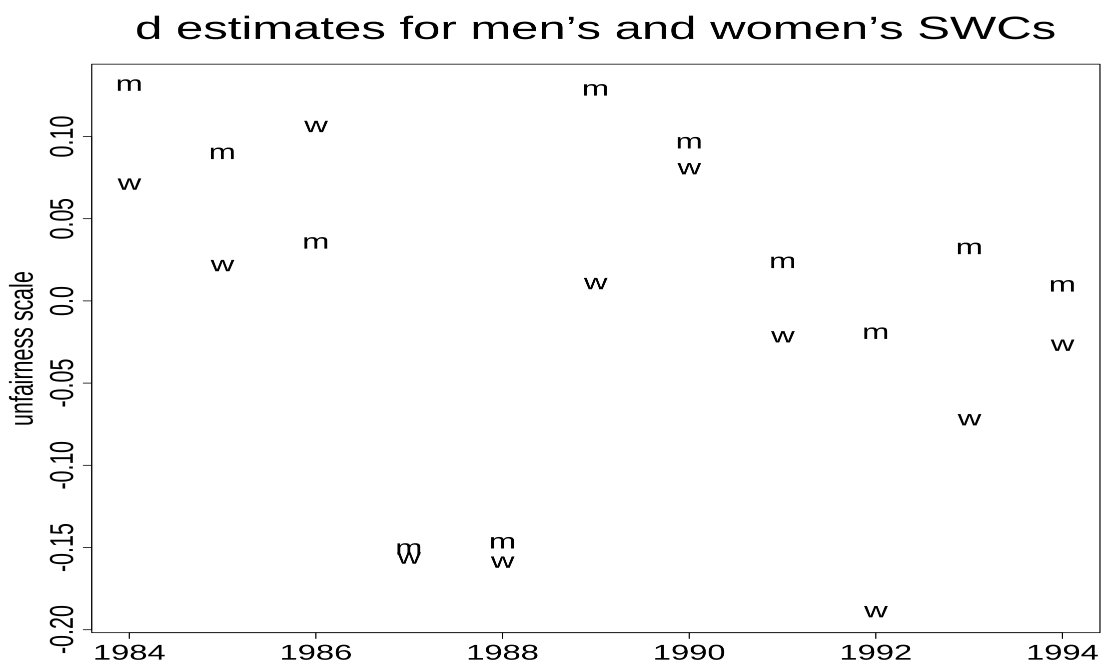

FIGURE 6. *Estimates of the d parameter, for the women's and the men's events, for eleven SWCs 1984–1994. The correlation is 0.792, and indicates that the real d parameter changes somewhat from event to event.*

	$\widehat{d}$	st. error
1994 Calgary:	−0.027	0.039
1993 Ikaho:	−0.072	0.067
1992 Oslo:	−0.189	0.090
1991 Inzell:	−0.022	0.079
1990 Tromsø:	0.080	0.063
1989 Heerenveen:	0.010	0.046
1988 West Allis:	−0.159	0.086
1987 Sainte Foy:	−0.157	0.159
1986 Karuizawa:	0.106	0.095
1985 Heerenveen:	0.021	0.079
1984 Trondheim:	0.071	0.081
grand average:	−0.015	0.020

TABLE 4. *Estimates of the difference parameter d for eleven Sprint World Championships for women, along with standard errors (estimated standard deviation) for these.*

The first thing to note is that it is not possible to reject the null hypothesis $d = 0$ based on these data, and any reasonable confidence interval will cover zero. A second prominent feature is that the d estimates for women are highly correlated with those for the men (the correlation coefficient is 0.792); see Figure 6. This supports the notion that each SWC event has its own characteristics, depending on wind, weather and gliding conditions; in addition two skating rinks might differ somewhat with respect to the actual radii used for inner and outer lanes. This also lends support to the appropriateness of the new ISU/IOC rules, that also the women skaters are to run the Olympic 500 meter twice from Nagano 1998 onwards, as well as in the annual single distance World Championships that are introduced in 1996.

7. Supplementing remarks

Remark A. The inner-outer question can also be asked for the 1000 meter competition. In this event skaters the first 200 meter, in particular, are quite different stretches for the two skaters. A potential difference is not likely to be statistically visible, however; the distance is long enough to diminish any tiny discrepancies. My student Siri Størmer is considering SWC results from several 1000 meters, looking for inner-outer significances as well as other aspects, and indeed preliminary investigations have not found any significant inner-outer differences.

Remark B. It is natural to think that the d parameter is not quite one and only one constant, but that it changes slightly from event to event, depending on weather and the rink itself. The comments regarding the similarity between the men's and the women's d estimates above supports this notion, cf. Figure 6. A simple model for this is to postulate that $d_j \sim \mathcal{N}\{d_0, \omega_0^2\}$, where d_0 is some presumed grand mean over many events and ω_0 the level of variation. Our $\widehat{d}_j$ is approximately a $\mathcal{N}\{d_j, \mathrm{se}_j^2\}$, conditionally on d_j. One may now show that

$$\sum_{j=1984}^{1994} (\widehat{d}_j - \widehat{d})^2/\mathrm{se}_j^2 \quad \text{has mean value} \quad 10 + \omega_0^2(A_2 - A_4/A_2),$$

where $\widehat{d}$ is as in (5.1) and $A_q = \sum_{j=1984}^{1994} 1/\mathrm{se}_j^q$ for $q = 2, 4$. This gives us the opportunity to estimate the variability between events parameter ω_0; I find 0.057 for the men's events and 0.042 for the women's. In other words, according to this model, the d_js can be expected to wander from about -0.05 up to about 0.14, in 90% of such competitions for the men. The corresponding figures for the women are -0.08 to 0.05.

That the underlying parameters of the model, from (a_1, a_2, b, d) to (σ, ρ, κ), vary from event to event in a suitably regular fashion can be modelled as well, as with the d_j parameters. This may lead us to empirical Bayes modelling and estimation. Such a framework yields refined estimators of the individual parameters that in an overall sense would be more precise than for example the $\widehat{d}_j$s of Section 1's Table 1. Our main task has been to estimate the grand average of these, however, and it seems more natural to let the parameters of each competition speak for themselves, without weighing in similar information from other years.

Remark C. One may ask whether there is a difference between the very best skaters in the world and the not quite as excellent ones, regarding their ability to tackle the last inner lane. The answer would depend on the selected party with which one wishes to compare the very best. The level of the skaters being allowed to compete at the SWCs is now uniformly very high, however, and significant differences would have been surprising. To investigate this matter, I divided each of the 1984–1994 sets into two halves, the best part (decided on by the average 500 meter result) and the remaining ones. For each half one can fit the (2.2) model (with $b_1 = b_2$), and in particular compare the $\widehat{d}_{\rm best}$ estimate for the very best skaters in the world with the corresponding $\widehat{d}_{\rm rest}$ valid for the not quite as spectacular skaters. Overall there seems to be a certain tendency towards $d_{\rm best} < d_{\rm rest}$, that is, the very best skaters are better at handling also the last inner lane problems. This discrepancy is not significant, however; there are instances where the opposite happens, and the best combined estimate of $d_{\rm best} - d_{\rm rest}$ is -0.044 with standard error 0.029.

Acknowledgements. I am indebted to all the participants of the statistical experiments described here for their eager contributions. I am also grateful for the interest shown in this project by Tron Espeli at the International Skating Union's Technical Committee for Speed Skating.

Reference

Bjørnsen, K. (1963). *13 år med Kuppern & Co.* Nasjonalforlaget, Oslo. ['Kuppern' is Knut Johannesen, Olympic medalist in 1956, 1960, 1964.]

Appendix I

The left hand list given here gives the real results from the Olympic 500 meter sprint event. The supplementary faked list is meant to show what presumably would have happened if the inner-outer lane draw had ended in the opposite way. For example, 1988 silver medalist Jan Ykema would have ended fourth while fourth ranked Sergei Fokitshev would have grabbed the bronze, and so on.

Olympic Games, Lillehammer 1994

Real list:				Speculative list:	
1.	Aleksandr Golubyev	i	36.33	1. Aleksandr Golubyev	36.38
2.	Sergei Klevtshena	i	36.39	2. Sergei Klevtshena	36.44
3.	Manabu Horii	o	36.53	3. Manabu Horii	36.48
4.	Hongbo Liu	i	36.54	4. Hiroyasu Shimizu	36.55
5.	Hiroyasu Shimizu	o	36.60	5. Hongbo Liu	36.59
6.	Junichi Inoue	i	36.63	6. Grunde Njøs	36.61
7.	Grunde Njøs	o	36.66	7. Yasunori Miyabe	36.67
8.	Dan Jansen	i	36.68	8. Junichi Inoue	36.68
9.	Yasunori Miyabe	o	36.72	Igor Zhelezovsky	36.68
10.	Igor Zhelezovsky	o	36.73	10. Dan Jansen	36.73
11.	Sylvain Bouchard	o	37.01	11. Sylvain Bouchard	36.96
12.	Patrick Kelly	i	37.07	12. Yoon-Man Kim	37.05

	Vadim Shakshakbayev	i	37.07	13.	Patrick Kelly	37.12
14.	Yoon-Man Kim	o	37.10		Vadim Shakshakbayev	37.12
15.	Mikhail Vostroknutov	i	37.15	15.	Mikhail Vostroknutov	37.20
16.	Andrei Bakhvalov	i	37.24	16.	Sean Ireland	37.25
17.	Sean Ireland	o	37.30	17.	Andrei Bakhvalov	37.29
18.	Peter Adeberg	o	37.35	18.	Peter Adeberg	37.30
19.	David Cruikshank	i	37.37	19.	Nathaniel Mills	37.36
20.	Nathaniel Mills	o	37.41	20.	Gerard van Velde	37.40
21.	Gerard van Velde	o	37.45	21.	David Cruikshank	37.42
22.	Roland Brunner	i	37.47	22.	Arie Loef	37.47
23.	Oleg Kostromitin	i	37.50	23.	Hans Markström	37.48
24.	Arie Loef	o	37.52	24.	Roland Brunner	37.52
25.	Hans Markström	o	37.53	25.	Oleg Kostromitin	37.55
26.	Michael Ireland	i	37.67	26.	Michael Ireland	37.72
27.	David Besteman	i	37.68	27.	David Besteman	37.73
28.	Lars Funke	o	37.80	28.	Lars Funke	37.75
29.	Alessandro de Taddei	i	37.87	29.	Sung-Yeol Jaegal	37.85
30.	Sung-Yeol Jaegal	o	37.90	30.	Nico van der Vlies	37.89
31.	Nico van der Vlies	o	37.94	31.	Alessandro de Taddei	37.92
32.	Davide Carta	o	37.98	32.	Davide Carta	37.93
33.	Vladimir Klepinin	o	38.09	33.	Vladimir Klepinin	38.02
34.	Magnus Enfeldt	i	38.10	34.	Jae-Shik Lee	38.05
	Jae-Shik Lee	o	38.10	35.	Kyou-Hyuk Lee	38.08
36.	Kyou-Hyuk Lee	o	38.13	36.	Magnus Enfeldt	38.15
37.	Arjan Schreuder	i	38.33	37.	Arjan Schreuder	38.38
38.	Zsolt Balo	i	38.56	38.	Zsolt Balo	38.61
39.	Michael Spielmann	i	38.58	39.	Michael Spielmann	38.63
–	Roger Strøm	i	dnf		Roger Strøm	—

Olympic Games, Albertville 1992

	Real list:				**Speculative list:**	
1.	Uwe-Jens Mey	o	37.14	1.	Uwe-Jens Mey	37.09
2.	Toshiyuki Kuroiwa	i	37.18	2.	Toshiyuki Kuroiwa	37.23
3.	Junichi Inoue	i	37.26	3.	Junichi Inoue	37.31
4.	Dan Jansen	o	37.46	4.	Dan Jansen	37.41
5.	Yasunori Miyabe	i	37.49	5.	Gerard van Velde	37.44
	Gerard van Velde	o	37.49	6.	Aleksandr Golubyev	37.46
7.	Aleksandr Golubyev	o	37.51	7.	Chen Song	37.53
8.	Igor Zhelezovsky	i	37.57	8.	Yasunori Miyabe	37.54
9.	Chen Song	o	37.58	9.	Yoon-Man Kim	37.55
10.	Yoon-Man Kim	o	37.60	10.	Igor Zhelezovsky	37.62
11.	Hongbo Liu	i	37.66	11.	Sung-Yul Yegal	37.66
12.	Sung-Yul Yegal	o	37.71	12.	Hongbo Liu	37.71
13.	Nick Thometz	i	37.83	13.	Vadim Shakshakbayev	37.81
14.	Robert Dubreuil	i	37.86	14.	Guy Thibault	37.84
	Vadim Shakshakbayev	o	37.86	15.	Nick Thometz	37.88
16.	Guy Thibault	o	37.89	16.	Robert Dubreuil	37.91
17.	Kevin Scott	i	38.02	17.	Kevin Scott	38.07
18.	Yukihiro Miyabe	i	38.12	18.	Yukihiro Miyabe	38.17
19.	Marty Pierce	i	38.15	19.	Marty Pierce	38.20

20.	Björn Forslund	i	38.24
21.	Sergei Klevtshena	i	38.26
22.	David Cruikshank	i	38.28
23.	Peter Adeberg	o	38.33
24.	Yong-Cho Li	o	38.38
25.	Olaf Zinke	o	38.40
26.	Harri Ilkka	i	38.48
27.	Jun Dai	i	38.51
	Rintje Ritsma	o	38.51
29.	Arie Loef	o	38.61
30.	Sean Ireland	i	38.70
31.	Pawel Abratkiewicz	o	38.74
	In-Hoon Lee	o	38.74
33.	Hans Markström	o	38.89
34.	Bo König	i	39.06
35.	In-Hol Choi	o	39.59
36.	Zsolt Ballo	i	39.70
37.	Csaba Madarasz	i	40.41
38.	Joakim Karlberg	o	40.71
39.	Jiri Kyncl	i	40.92
40.	Roland Brunner	o	42.18
41.	Jiri Musil	o	42.20
42.	Bajro Cenanovic	i	43.09
43.	Slavenko Likic	o	43.81

20.	Peter Adeberg	38.28
21.	Björn Forslund	38.29
22.	Sergei Klevtshena	38.31
23.	Yong-Cho Li	38.33
	David Cruikshank	38.33
25.	Olaf Zinke	38.35
26.	Rintje Ritsma	38.46
27.	Harri Ilkka	38.53
28.	Arie Loef	38.56
	Jun Dai	38.56
30.	Pawel Abratkiewicz	38.69
	In-Hoon Lee	38.69
32.	Sean Ireland	38.75
33.	Hans Markström	38.84
34.	Bo König	39.11
35.	In-Hol Choi	39.54
36.	Zsolt Ballo	39.75
37.	Csaba Madarasz	40.46
38.	Joakim Karlberg	40.66
39.	Jiri Kyncl	40.97
40.	Roland Brunner	42.13
41.	Jiri Musil	42.15
42.	Bajro Cenanovic	43.14
43.	Slavenko Likic	43.76

Olympic Games, Calgary 1988

Real list:

1.	Uwe-Jens Mey	i	36.45
2.	Jan Ykema	i	36.76
3.	Akira Kuroiwa	o	36.77
4.	Sergei Fokitchev	o	36.82
5.	Ki-Tae Bae	o	36.90
6.	Igor Zhelezovsky	o	36.94
7.	Guy Thibault	o	36.96
8.	Nick Thometz	i	37.16
9.	Yasumitsu Kanehama	i	37.25
10.	Frode Rønning	i	37.31
11.	Yasushi Kuroiwa	o	37.34
12.	Vitali Makovetski	o	37.35
13.	Kimihiro Hamaya	o	37.38
14.	Gaetan Boucher	i	37.47
15.	Erik Henriksen	i	37.50
16.	Menno Boelsma	o	37.52
17.	Daniel Turcotte	i	37.60
18.	Göran Johansson	i	37.69
	Bjørn Hagen	o	37.69
20.	Hanspeter Oberhuber	i	37.73
21.	André Hoffmann	i	37.75
22.	Marty Pierce	o	37.76
23.	Michael Richmond	o	37.77

Speculative list:

1.	Uwe-Jens Mey	36.50
2.	Akira Kuroiwa	36.72
3.	Sergei Fokitchev	36.77
4.	Jan Ykema	36.81
5.	Ki-Tae Bae	36.85
6.	Igor Zhelezovsky	36.89
7.	Guy Thibault	36.91
8.	Nick Thometz	37.21
9.	Yasushi Kuroiwa	37.29
10.	Vitali Makovetski	37.30
	Yasumitsu Kanehama	37.30
12.	Kimihiro Hamaya	37.33
13.	Frode Rønning	37.36
14.	Menno Boelsma	37.47
15.	Gaetan Boucher	37.52
16.	Erik Henriksen	37.55
17.	Bjørn Hagen	37.64
18.	Daniel Turcotte	37.65
19.	Marty Pierce	37.71
20.	Michael Richmond	37.72
21.	Göran Johansson	37.74
22.	Hein Vergeer	37.75
23.	Jerzy Dominik	37.78

24.	Hein Vergeer	o	37.80		Hanspeter Oberhuber	37.78
25.	Jerzy Dominik	o	37.83	25.	André Hoffmann	37.80
26.	Michael Hadschieff	i	37.90	26.	Michael Hadschieff	37.95
27.	Uwe Streb	i	38.03	27.	Uwe Streb	38.08
28.	Peter Adeberg	i	38.11	28.	Peter Adeberg	38.16
29.	Robert Tremblay	i	38.34	29.	Robert Tremblay	38.39
30.	Hans Magnusson	o	38.60	30.	Hans Magnusson	38.55
31.	Claude Nicoleau	i	38.63	31.	Claes Bengtsson	38.61
32.	Claes Bengtsson	o	38.66	32.	Claude Nicoleau	38.68
33.	Hans van Helden	i	39.05	33.	Hans van Helden	39.10
34.	Christian Eminger	o	39.70	34.	Christian Eminger	39.65
35.	Behudin Merdovic	o	fell	–	Behudin Merdovic	—
36.	Nikolai Gulyayev	i	fell	–	Nikolai Gulyayev	—
–	Dan Jansen	i	dnf	–	Dan Jansen	—

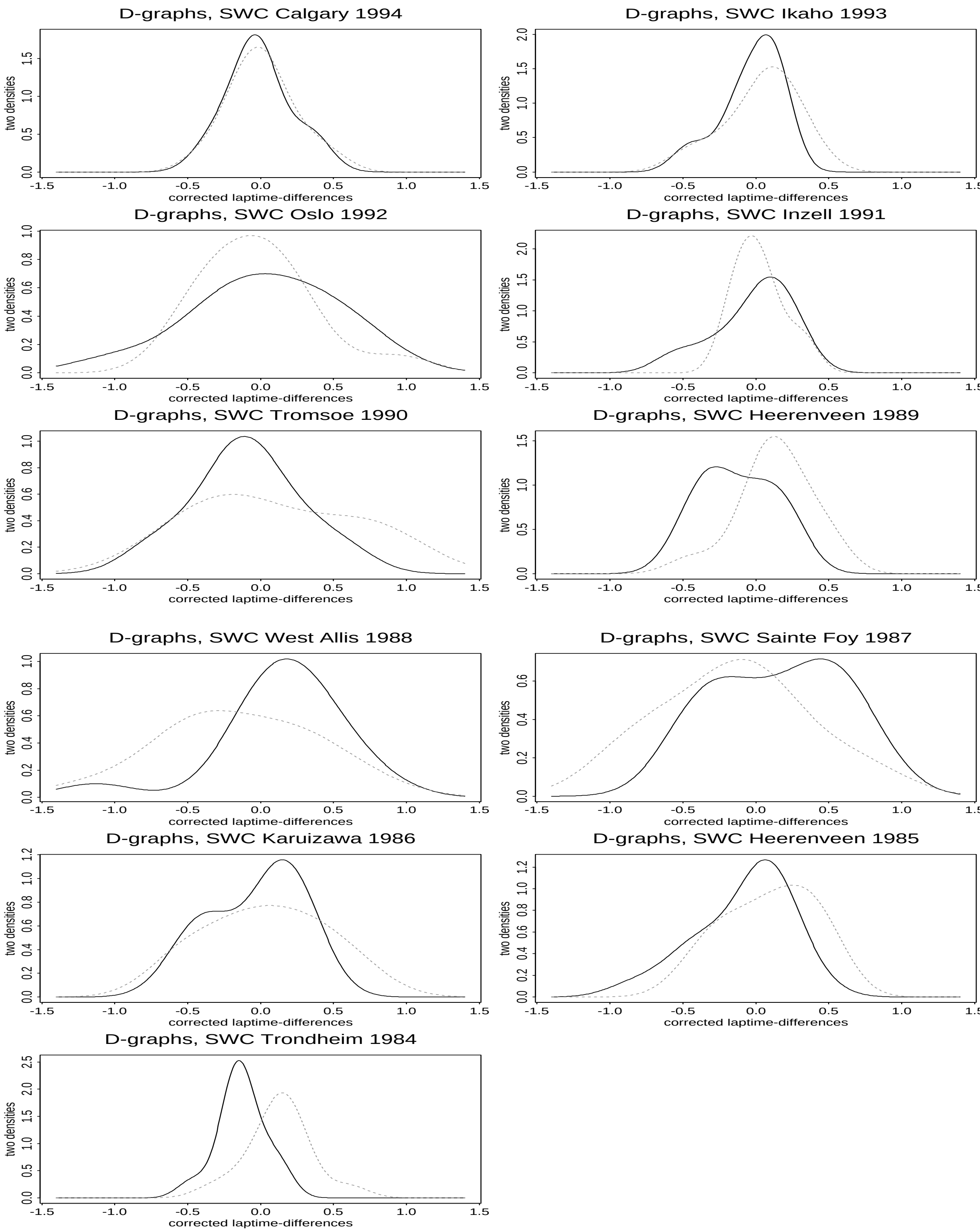

FIGURE 3. *Nonparametric kernel-method density estimates for the modified difference variables D_i, for the $w_i = 1/2$ group (fully drawn) and for the $w_i = -1/2$ group (dotted line), for each of the eleven SWCs 1984–1994. The advantage of having last outer lane is arguably prominent for the 1991, 1990, 1989, 1986, 1985, 1984 occasions, undecided for the 1994, 1993, 1992 events, while the advantage seems to have been with the last inner lane in 1988 and 1987. The number of skaters contributing to the eleven sub-figures are respectively 26, 29, 30, 33, 28, 29, 28, 32, 26, 30, 27, for the years 1984–1994.*