The Statistician (2000)
49, *Part* 1, *pp.* 95–106

Focus on Sport

Batting strategy in limited overs cricket

Ian Preston and Jonathan Thomas

University College London, UK

[Received February 1999. Revised October 1999]

Summary. The paper attempts to understand batting strategies that are employed in limited overs cricket games. The question of the optimum batting strategy is posed in a simplified dynamic programming representation. We demonstrate that optimum strategies may be expected to differ fundamentally in the first and second innings, typically involving an increasing run rate when setting a target but a run rate which may decline over the course of an innings when chasing one. Data on English county level limited overs games are used to estimate a model of actual batting behaviour. The statistical framework takes the form of an interesting variant on conventional survival analysis models.

Keywords: Batting strategy; Cricket; Dynamic programming; Survival analysis

1. Introduction

Cricket is a game with many interesting strategic aspects. This paper attempts to understand batting strategies that are employed in limited overs cricket games. We investigate both those strategies that are most likely to lead to victory and those that are actually played. There are several reasons why these issues are of interest. An understanding of optimum strategies is of value in an obvious practical sense to those who play the game and want to win. Strategies played also determine the way in which probabilities of victory evolve over the course of games. An understanding of this may help in judging how teams stand at any point in a game and may therefore also be relevant to the design of rules for adjusting scores in rain-interrupted games. In principle, it could also be useful in the evaluation of the performance of players. Our results suggest that the optimum strategy depends on the way in which the likelihood of dismissal depends on the attempted scoring rate. Our empirical analysis of actual batting behaviour allows us to estimate the parameters of this relationship.

In this paper we consider the question of the optimum batting strategy in a simplified dynamic programming representation of a limited overs cricket game. Clarke (1988) has previously investigated optimum strategy in limited overs cricket by using dynamic programming methods and has simulated solutions numerically. We extend this by imposing a greater theoretical structure on the problem. This allows us to derive conclusions which can be generalized beyond his particular numerical specifications and which hopefully illuminate the source of his results. For example, his conclusion that the popular strategy of building up runs slowly has little justification, at least in

Address for correspondence: Ian Preston, Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK.
E-mail: i.preston@ucl.ac.uk

0039–0526/00/49095

the second innings, is mirrored in our findings. We suggest reasons, however, why such a strategy may have more justification in the first innings.

We also use data on English county level limited overs games to estimate a model of cricketers' behaviour. Bairam and Howells (1991a, b) and Bairam *et al.* (1990a, b) have investigated the empirical relationship between measures of performance, including batting and bowling inputs, and both probabilities of victory and the accumulation of points. This is a rather different exercise, however, from investigating the aspects of within-match strategy considered here. Our statistical framework takes the form of an interesting variant on conventional survival analysis models. Kimber and Hansford (1993) have previously applied survival analysis to data on individual innings. Our work goes further in its attempt to relate estimates to the effects of covariates on hazards within the context of a structural model of batting strategy.

Duckworth and Lewis (1998) have developed and estimated a target reduction rule based on ideas of neutrality between teams and balance in run scoring resources. This rule, adopted by several cricketing authorities, is undoubtedly a valuable improvement on crude proportional reduction rules and we hope that the sort of analysis in this paper can illuminate its possible foundations in the context of a model of cricketing behaviour.

The paper is structured as follows. Section 2 briefly outlines the main features of the game and considers optimal strategies. Several unresolved issues that are investigated through numerical dynamic programming and simulation are reported in Section 3. An empirical specification is described in Section 4 and Section 5 presents the results. Section 6 gives a summary and discusses directions for future research.

2. Batting strategy as a stochastic dynamic programming problem

Limited overs cricket is a game between two teams of 11 players. There are two 'innings' with each team 'batting' once and 'bowling' once. At any point in the game, the team batting has two players at the centre of the pitch, known as the 'wicket', facing balls delivered to them by the bowling side. Balls are delivered in groups of 6 known as 'overs'. The batting team scores 'runs' by attempting to strike the ball and to run between the ends of the wicket while the bowling side recovers and returns the ball. In doing this the batsmen expose themselves to the risk of being dismissed ('getting out' or 'losing a wicket') in several ways. We assume that all major forms of dismissal are made more likely by a higher run scoring rate. This seems reasonable if batsmen have any control over their scoring rate since otherwise it would be difficult to see why batsmen should show any inhibition in their rate of scoring. The innings ends after a set number of overs have been bowled or all but one of the members of the batting team have been dismissed, whichever happens sooner. The total runs accumulated by the team batting first is then the target to be achieved by the team batting second if they are to win the match. In some versions of the game, though none of the games analysed in the empirical section below, there are restrictions on the fielding side during certain overs of the innings, designed to encourage greater adventurousness by the batting side. The '15-overs rule' is an example of this. More detailed outlines of the game can be found in many books, such as Marylebone Cricket Club (1998) or Morrison (1998), and on several World Wide Web sites, such as the Cricinfo site at `http://www.cricket.org`.

Several questions of strategy immediately arise. How fast should a batting team aim to score its runs at any point in either innings? How should the increased probability of dismissal be balanced against the value of accumulating more runs? First we consider these issues in the context of the second innings where the batting team is chasing a well-defined given target. The more complicated case of strategy in the first innings is then addressed.

2.1. The second-innings problem

The issue can be formally addressed within the context of a stochastic dynamic programming problem and it is natural to start with the second innings. Suppose that the team batting second must score a target of $\overline{R}$ runs in no more than $\overline{T}$ balls. Consider the kth batting partnership facing the tth ball of the innings with R_{t-1} runs having been accumulated. We suppose that the control variable is the *intended* run rate r_t^*. The effect of run scoring rates on the probability of dismissal is captured in a dismissal hazard $h_t = \eta_{k,t}(r_t^*)$. We assume $\eta_{k,t}(\cdot)$ to be positive, increasing and convex, which assures a well-behaved solution to the problem. This is clearly a simplification. The hazard may depend, for instance, on the number of balls since the beginning of the partnership as well as the number of balls since the beginning of the innings and the relationship between the two depends on the uncertain timing of dismissals. To capture this in the notation would be extremely cumbersome and we hope that nothing is lost in ignoring it.

Actual runs achieved if not dismissed are r_t. In the simplest case, there is no uncertainty in run scoring other than that associated with the possibility of dismissal and hence $r_t = r_t^*$. More generally, if we wish to allow for the runs scored to depend uncertainly on the intended run rate even if not dismissed then r_t can be a random variable, the distribution of which depends on r_t^*. The probability of eventually achieving the target, which is in this context the value function, is denoted $V_{k,t} = \Omega_{k,t}(R_{t-1})$.

By Bellman's equation (see, for instance, Bertsekas (1987))

$$V_{k,t} = \max_{r_t^*} [\{1 - \eta_{k,t}(r_t^*)\} E\{\Omega_{k,t+1}(R_t)\} + \eta_{k,t}(r_t^*)\, \Omega_{k+1,t+1}(R_{t-1})]. \tag{1}$$

Assuming for heuristic purposes that we can regard the scoring rate as a continuous variable and that first-order conditions characterize the optimum, then

$$0 = \{1 - \eta_{k,t}(r_t^*)\} \frac{\partial}{\partial r_t^*} E\{\Omega_{k,t+1}(R_t)\} - \eta'_{k,t}(r_t^*)[E\{\Omega_{k,t+1}(R_t)\} - \Omega_{k+1,t+1}(R_{t-1})]. \tag{2}$$

The first term is the expected marginal gain from run accumulation whereas the second is the expected marginal loss from the increased likelihood of dismissal. We can reformulate this in the form

$$\psi_t \equiv \frac{\partial}{\partial r_t^*} \ln\{1 - \eta_{k,t}(r_t^*)\} = \frac{\partial E\{\Omega_{k,t+1}(R_t)\}/\partial r_t^*}{E\{\Omega_{k,t+1}(R_t)\} - \Omega_{k+1,t+1}(R_{t-1})},$$

defining thereby ψ_t which can be seen to be increasing in the chosen run rate r_t^* and which might therefore be interpreted as an indicator of optimal lack of caution.

The case with no uncertainty is the easiest to analyse. For conciseness of exposition, let $V'_{k,t}$ denote the value of the derivative $\Omega'_{k,t}(R_{t-1})$ and let $V_{k+1,t}$ and $V'_{k+1,t}$ denote $\Omega_{k+1,t}(R_{t-2})$ and $\Omega'_{k+1,t}(R_{t-2})$ respectively, in each case as R_t evolves along the optimal path of run rates for the kth partnership. Then we can use equation (1) to show that

$$\begin{aligned}
\psi_t &= \frac{V'_{k,t+1}}{V_{k,t+1} - V_{k+1,t+1}} \\
&= \frac{(1 - h_{t+1})V'_{k,t+2} + h_{t+1} V'_{k+1,t+2}}{(1 - h_{t+1})V_{k,t+2} + h_{t+1} V_{k+1,t+2} - V_{k+1,t+1}} \\
&= \psi_{t+1}\beta_{t+1} + \frac{h_{t+1} V'_{k+1,t+2}}{\Delta V_{k+1,t+2}}(1 - \beta_{t+1})
\end{aligned}$$

where

$$\beta_{t+1} \equiv \frac{(1-h_{t+1})\,\Delta V_{k,t+2}}{(1-h_{t+1})\,\Delta V_{k,t+2} + h_{t+1}\,\Delta V_{k+1,t+2}}.$$

Thus

$$\Delta\psi_{t+1} = \left\{ \psi_{t+1} - \frac{h_{t+1}\,V'_{k+1,t+2}}{\Delta V_{k+1,t+2}} \right\} (1-\beta_{t+1}). \tag{3}$$

Now consider the last partnership ($k = 10$). The problem takes a particularly special form since there are no future partnerships to achieve the target. Therefore $V_{k+1,t} \equiv 0$ for all t and $\beta_{t+1} = 1$. It follows immediately from equation (3) that ψ_t is constant throughout the partnership. If the hazard is not duration dependent, i.e. $\eta_{10,t}(\cdot) \equiv \eta_{10}(\cdot)$ for all t, then the optimum strategy is to aim at the required run rate throughout the duration of the partnership. For the final partnership the proper balance between accumulating runs and safeguarding the wicket in such circumstances implies neither speeding up nor slowing down in intended run rates.

Of course, it may be more complicated than this in practice for various reasons. Most particularly, certain forms of duration dependence in the hazard may provide batsmen with a good case to take time to assess the bowling conditions. Note, however, the form that this would need to take —a mere decline in the hazard with duration t would not suffice. Rather, a good case for 'playing oneself in' would have to involve a decline in the dependence of the hazard on the scoring rate. Variations in the quality of the bowling over the course of the innings and fielding restrictions such as the 15-overs rule could provide other arguments for a non-constant intended rate.

The solution to the last partnership problem can be treated as an input to the penultimate partnership problem and by backward recursion we can work back ultimately to a solution of the opening batsmen's problem. If, as suggested earlier, partnerships with no succeeding batsmen should aim at or close to the required rate, then intuition suggests that earlier partnerships ought to be aiming above it because of the security offered by the availability of further partnerships if early batsmen are dismissed. This can be confirmed for the case of the penultimate partnership ($k = 9$) on the penultimate ball of the innings. It is easy to see from equation (2) that this partnership should choose a higher run rate if a succeeding partnership is available than if not. Establishing the generality of this conjecture would require a more involved argument and we rely on simulations for corroboration.

How are these arguments affected by uncertainty over run scoring if batsmen are not dismissed? It is not obvious whether the response will be precautionary run scoring to build up a buffer of runs to guard against ill luck later in the innings or wicket retention to enable an onslaught on the target late in the innings if there is ill luck earlier. Considering only the effect of uncertainty on the optimal run rate on the penultimate ball makes plain that the answer depends on the concavity properties of the marginal hazard $\eta'_{k,t}(r_t)$. Again we address these issues through simulation.

2.2. The first-innings problem

The aim for the team batting first is to minimize the probability of victory for the second innings. This will correspond to the optimized probability derived in the previous section if, but only if, the opponent can be assumed to be behaving optimally.

Formally the problem has considerable similarity with that for the second innings with a similar Bellman's equation, for instance, governing the evolution of the probability of victory. The difference is only in the less precipitous shape of that probability at the cessation of the innings. The probability of victory is no longer equal to 0 or 1 depending on whether a target has been achieved. Rather it is a steadily increasing function of runs accumulated and therefore set as the target for the opponent.

There are strong grounds for thinking that there is no completely general result on whether optimal run trajectories should increase or decrease. This can be illustrated with a simple example. Suppose that the last batsman has two balls left in the innings, there is no run scoring uncertainty and

$$h_{k,t}(r) = \min\left(\frac{r^2}{100}, 1\right),$$
$$V_{k+1,t}(R) = \Phi(R - \gamma)$$

where $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function. The optimal run rate trajectory is increasing if $\gamma \leqslant 2.526$ and decreasing otherwise.

However, there are plausible arguments establishing a presumption in favour of an increasing trajectory. Suppose that the probability of victory at the end of the innings depends linearly on the number of runs scored, which may be a reasonable approximation over a fair range of accumulated runs. This is compatible with the approach in Clarke (1988), where maximization of runs scored is simply assumed to be the objective of the first innings. Consider the last partnership ($k = 10$) and adopt the same notation as that used for the second innings. Given the formal similarity to the second-innings problem, equation (3) continues to hold with $V_{k+1,t+1} \equiv R_t$. Hence

$$\Delta\psi_{t+1} = \left(\psi_{t+1} - \frac{h_{t+1}}{r_t}\right)(1 - \beta_{t+1}).$$

Assuming that $r_t \simeq r_{t+1}$, this will be positive given the assumptions on $\eta_{k,t}(\cdot)$, which imply that $\psi_{t+1} r_{t+1} - h_{t+1} > 0$. If there is no dependence on duration this implies a run rate increasing over the course of the partnership. It is easy to see why this should be. Under the assumed objective the loss from dismissal diminishes over the course of the innings since it depends on the potentially forgone overs if the innings ends early. This may give an incentive for pursuing a strategy of increasing recklessness over the course of the innings. The numbers in Table 2 of Clarke (1988) appear to support this contention, even though the interpretation in the text claims that 'the generally accepted view of scoring slowly at the beginning of the innings is not optimal'. The essence of this argument may be expected to have some validity even where the objective is not linear in runs.

This conclusion may also shed light on batting behaviour in the second innings. Notice the implied conflict between the personal objective of a batsman interested in his own batting score and the objectives of the team if this is not simply to maximize expected runs, as in the second innings. Batsmen pursuing their own score would be expected to follow increasing run rates independently of whether that suited the interest of the team.

As regards earlier partnerships, the intuitive argument for presuming that earlier partnerships will aim ahead of the rates that would be chosen if there were no succeeding batsmen seems equally applicable and we would expect this to be evident in simulations.

3. Simulations

We explore some of the issues that were unresolved in the previous section through numerical solution. We assume that the match lasts for 40 overs and, for simplicity, take the over rather than the ball as the unit of decision.

We specify the hazard so that it is multiplicatively separable in components depending on the intended run rate and on the number of wickets lost. The former is assumed increasing and convex whereas the latter is stepwise increasing. The need to bound the dismissal probability from above at 1 leads to eventual non-convexity but, given chosen parameter values, this would come into play only for unrealistically high initial required run rates:

$$\eta_k(r^*) = \min\{f(k)\, g(r^*), 1\},$$

$$f(k) = \begin{cases} 0.5 & \text{if } k \leqslant 5, \\ 1 & \text{if } 6 \leqslant k \leqslant 7, \\ 1.5 & \text{if } 8 \leqslant k, \end{cases}$$

$$g(r^*) = \frac{1}{120} r^{*2}.$$

The assumption of declining skill down the batting order seems reasonable though we appreciate that, in practice, batsmen differ in more than one dimension of scoring ability. Our main conclusions above about the general shape of run rate trajectories were not dependent on this assumption, though the extent to which earlier batsmen should score differently from later batsmen will be affected by this sort of detail. We abstain completely from questions involving tactics on the bowling side.

To allow for uncertainty in run scoring we allow for a probability p that the batsmen fail to score in any particular over (independently of the intended scoring rate). By varying p from 0 upwards we can investigate numerically the importance of run scoring uncertainty. Runs scored if not dismissed are

$$\begin{aligned} r &= [r^*] && \text{with probability } (1-p)([r^*] - r^* + 1), \\ r &= [r^*] + 1 && \text{with probability } (1-p)(r^* - [r^*]), \\ r &= 0 && \text{with probability } p, \end{aligned}$$

where $[\cdot]$ denotes the greatest integer function, introduced as a convenient way of recognizing the discreteness that is inherent in run scoring while still allowing the intended run rate to vary continuously. One consequence of this is that there remains a small uncertainty for non-integer values of r^* even if $p = 0$. This could in principle upset the result that optimum scoring rates are flat for the final partnership and create a somewhat artificial preference for integer values of r^* in the optimum strategy. This does not seem important in the simulations reported here.

The optimal strategy in each innings was then found by dynamic programming methods for different values of the run scoring probability p. First-innings strategies are calculated as best responses to the assumed optimal behaviour by teams batting second. Run rate trajectories for each undismissed partnership were averaged over 100 simulations and run rate trajectories for whole innings, including dismissals, were averaged over 1000 simulations. Run targets for second-innings simulations were set at average achieved totals in the first innings. The results for individual partnerships given values $p = 0$ and $p = 0.5$ are illustrated in Fig. 1. These are arranged in a 2×2 array with first-innings partnerships in the first column and second-innings partnerships in the second. Figs 1(a) and 1(b) show results with no uncertainty in run scoring, $p = 0$, and Figs 1(c) and 1(d) show those with $p = 0.5$.

The earlier discussion is supported by the simulations for the case without uncertainty in run scoring. The last partnership in the second innings pursues a flat scoring rate and earlier partnerships do, as conjectured, aim ahead of the rate required and pursue a declining scoring rate. Note also that optimal scoring rates increase over the course of first-innings partnerships, as

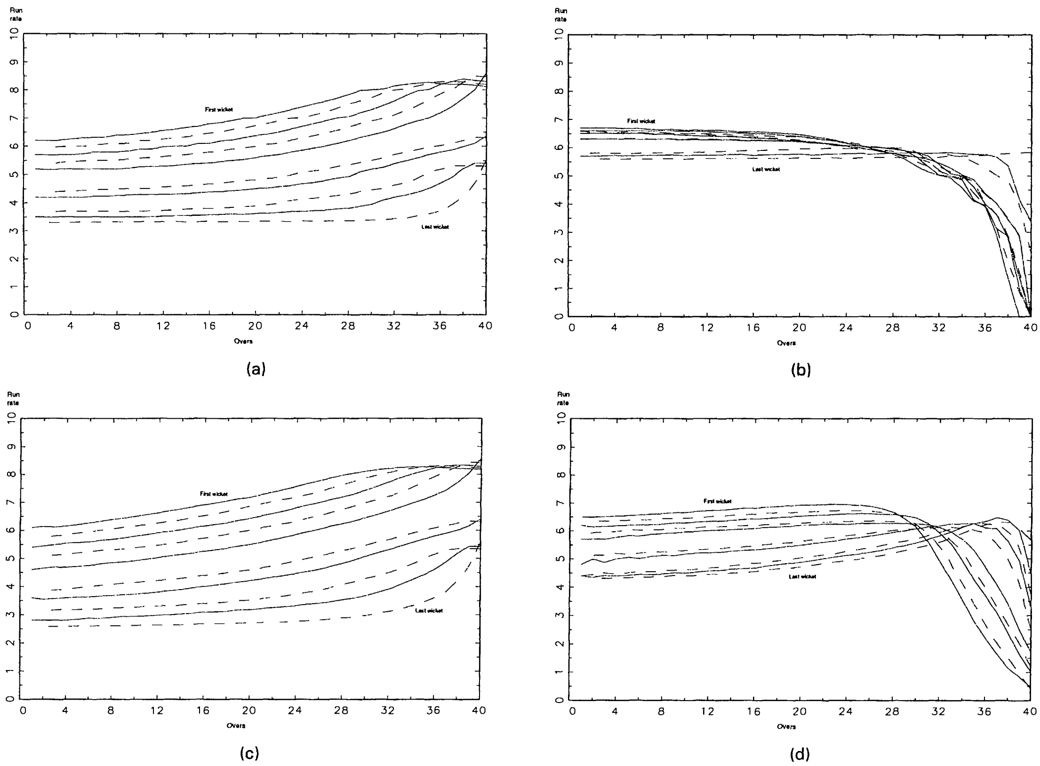

Fig. 1. Simulated partnerships: (a) first innings, $p = 0$; (b) second innings, $p = 0$; (c) first innings, $p = 0.5$; (d) second innings, $p = 0.5$

conjectured. All first-innings simulations start with 0 runs and 40 overs left. For later partnerships this would be an extremely bleak beginning and one in which early dismissal would lead to extremely likely defeat. In these sorts of circumstances linearity of the victory probability in runs scored might be a poor approximation.

The main unresolved issue regards the effect of uncertainty. In these simulations (and others unreported here) uncertainty seems to encourage the retention of wickets rather than precautionary run accumulation although this does not upset the conclusion that early partnerships in the second innings aim ahead of the rate required.

4. Empirical specification: the second innings

4.1. Model

We concentrate empirical attention on second-innings strategies. This is mainly because the need to achieve a known run target provides a useful empirical anchor on the level of intended run rates which is absent in the first innings. Although it would be possible to model an overall aspired run rate for the first innings there are no obvious covariates in the data and modelling the way in which this might change with the loss of wickets raises forbidding problems.

To simplify the notation consider the position at the beginning of a partnership. Let $\bar{r}$ be the required run rate, T be the remaining number of balls at the beginning of a partnership and Z be other factors which could include wickets lost, among other things.

The intended run rate after t balls (which need not but could be the optimum strategy) is

$$r^* = r^*(\bar{r}, T, Z, T-t, t).$$

It seems helpful to represent r^* as depending both on the number of balls faced t (because of initial nerves etc.) and on the potential number of balls remaining $T-t$ (for strategic reasons).

Without wishing to impose optimality, it still seems sensible to require that the strategy pursued involves eventually attaining the required target. Modelling, for convenience, in continuous time

$$\int_0^T r^*(\bar{r}, T, Z, T-t, t)\,\mathrm{d}t = \bar{R}.$$

If the hazard function is $\eta(r^*, Z, t)$ then the hazard along the chosen path is

$$\begin{aligned} h^* &= \eta\{r^*(\bar{r}, T, Z, T-t, t), Z, t\} \\ &= \eta^*(\bar{r}, T, Z, T-t, t). \end{aligned}$$

Note the potentially important restriction that $\bar{r}$ and T enter only through r^*.

Suppose that both the intended run rate and the hazard are multiplicatively separable (in the spirit of the proportional hazard models—see, for instance, Cox and Oakes (1985)) with

$$\begin{aligned} r^*(\bar{r}, T, Z, T-t, t) &= r_0(\bar{r}, T, Z)\,\phi(T-t, t), \\ \eta(r^*, Z, t) &= r^{*\alpha} f(Z)\,\psi(t). \end{aligned}$$

Then the proportional hazard feature is preserved in the chosen hazard specification

$$\eta^*(\bar{r}, T, Z, T-t, t) = r_0(\bar{r}, T, Z)^\alpha f(Z)\,\phi(T-t, t)^\alpha\,\psi(t).$$

In the spirit of the Weibull model, we set

$$\begin{aligned} \phi(T-t, t) &= (T-t+1)^\lambda t^\mu, \\ \psi(t) &= \kappa\rho t^{\rho-1} \end{aligned}$$

so that

$$\begin{aligned} r^*(\bar{r}, T, Z, T-t, t) &= r_0(\bar{r}, T, Z)(T-t+1)^\lambda t^\mu, \\ \eta^*(\bar{r}, T, Z, T-t, t) &= r_0(\bar{r}, T, Z)^\alpha f(Z)\kappa\rho(T-t+1)^{\lambda\alpha} t^{\mu\alpha+\rho-1}. \end{aligned}$$

By compatibility with the target we have that

$$r_0(\bar{r}, T, Z)\int_0^T (T-t+1)^\lambda t^\mu\,\mathrm{d}t = \bar{r}T$$

and therefore

$$r_0(\bar{r}, T, Z) = \frac{\bar{r}}{(T+1)^{\lambda+\mu} B(\mu+1, \lambda+1)},$$

where $B(\cdot, \cdot)$ denotes the beta function and we have assumed that $\mu+1, \lambda+1 > 0$ and that T is large.

Then the survival function S is given by

$$-\ln\{S(\bar{r}, T, Z, s)\} = \int_0^s \eta^*(\bar{r}, T, Z, T-t, t)\,\mathrm{d}t$$
$$= r_0(\bar{r}, T, Z)^\alpha f(Z)\kappa\rho \int_0^s (T-t+1)^{\lambda\alpha} t^{\mu\alpha+\rho-1}\,\mathrm{d}t$$
$$= \bar{r}^\alpha f(Z)\kappa\rho(T+1)^\rho \frac{B_{s/(T+1)}(\mu\alpha+\rho, \lambda\alpha+1)}{B(\mu+1, \lambda+1)^\alpha}$$

where $B_x(\cdot, \cdot) \equiv I_x(\cdot, \cdot)B(\cdot, \cdot)$ and $I_x(\cdot, \cdot)$ is the incomplete beta function and we have assumed that $\mu\alpha+\rho, \lambda\alpha+1>0$.

In recognition of the uncertainty in run scoring, assume that the actual run rate r follows $r^*+\epsilon$ where ϵ is a mean zero independent and identically distributed error term, assumed to be distributed independently of the event of dismissal. Then the average achieved run rate over the duration of the partnership r_s is

$$r_s = \frac{1}{s}\int_0^s r^*(\bar{r}, T, Z, T-t, t)\,\mathrm{d}t + \frac{1}{s}\int_0^s \epsilon\,\mathrm{d}t$$
$$= \frac{1}{s}\, r_0(\bar{r}, T, Z) \int_0^s (T-t+1)^\lambda t^\mu\,\mathrm{d}t + v_s$$
$$= \frac{T+1}{s}\,\bar{r}\, I_{s/(T+1)}(\mu+1, \lambda+1) + v_s.$$

v_s must be heteroscedastic with a variance which diminishes with the length of spell. Thus it would be feasible to estimate an achieved run rate equation by, say, non-linear least squares on the homoscedastic equation

$$s^{1/2} r_s = s^{1/2}(T+1)\bar{r}\, I_{s/(T+1)}(\mu+1, \lambda+1) + s^{1/2} v_s.$$

4.2. Estimation

We estimate the model by maximum likelihood assuming the error term ϵ to be Gaussian. Given the assumption that ϵ is distributed independently of dismissal we can factorize the likelihood for a partnership surviving s balls and achieving a run rate of r_s given $\bar{r}$, T and Z into a form

$$l_1(s, \bar{r}, T, Z; \mu, \lambda, \alpha, \rho, \theta_1)\, l_2(r_s, \bar{r}, T, Z; \mu, \lambda, \sigma, \theta_2),$$

where θ_1 and θ_2 denote parameters other than those already defined. μ, λ and σ are thus identified from the information on runs scored alone, whereas durations of partnerships provide identifying information on the remaining parameters as well as further information on μ and λ. Our estimates maximize the complete likelihood over the whole parameter space but we exploit this structure to derive starting values, firstly estimating μ, λ, σ and θ_2 by maximizing $l_2(\cdot)$ alone, then fixing those parameters at the estimated values and estimating α, ρ and θ_1 by maximizing $l_1(\cdot)$. The theory suggested that the slope of optimal trajectories should differ for differing partnerships so in certain estimations we allowed λ to differ between upper, middle and lower order batsmen.

There is a potentially serious source of bias due to omitted variables if we do not allow for match conditions, since a poor pitch is likely to lead to low targets and high second-innings hazards for reasons which are not connected with the choice of strategy. We choose to do this by including a full set of match effects (in the logarithm of the hazard) when estimating—i.e. we estimate a different intercept for each match. This, of course, precludes the inclusion of any other variables which do not vary within matches, such as team effects, though this should not cause worries since these effects will be completely absorbed in estimation (by the match effects).

5. Data and results

We have data from all 145 Axa Equity & Law Sunday League matches in the 1996 season. The 18 competing county teams would each have played each other once if the weather had allowed, but eight matches for various reasons lack adequate data for the purpose. The data cover 959 second-innings partnerships. Data on lengths of partnerships and runs scored were typed in from the original score-cards, because the conventional abbreviated published score-cards do not contain all the detail required. Sunday league innings last for 40 overs and conveniently there are no fielding restrictions which apply early in the innings.

Results with λ constant across partnerships are reported in Table 1. We produce columns for estimates without and with match effects, though we see strong grounds to prefer the latter. Match effects are jointly strongly significant and the increased coefficient on the required run rate when they are included accords with the expected direction of bias from their omission.

The estimated value of 1.591 for α, significantly different from 1 at the 5% level, offers evidence of a convex hazard as assumed in the theory. The estimated values of μ and λ, which are significantly different from 0 only in the latter case, suggest a strategy of increasing run rates. As the earlier simulations showed, this might be optimal with uncertainty in run scoring but not without. The positive estimated value for ρ suggests dismissal hazards which increase with the

Table 1. Model estimates with fixed λ†

Parameter	*Estimates for model without match effects*		*Estimates for model with match effects*	
	Coefficient	*Estimated standard error*	*Coefficient*	*Estimated standard error*
Constant	−3.526	−14.874	−4.705	−8.254
Required run rate	0.450	4.608	1.597	7.548
1st wicket	−0.818	−3.112	0.794	2.050
2nd wicket	−0.888	−3.375	0.835	2.163
3rd wicket	−0.694	−2.646	0.814	2.166
4th wicket	−0.534	−2.074	0.782	2.172
5th wicket	−0.529	−1.999	0.588	1.667
6th wicket	−0.245	−0.973	0.922	2.762
7th wicket	−0.113	−0.449	0.728	2.221
8th wicket	0.204	0.815	0.908	2.777
9th wicket	−0.002	−0.010	0.532	1.697
$\ln(\rho + \mu\alpha)$	0.201	6.439	0.292	6.898
$\ln(\sigma)$	1.747	50.594	1.766	48.973
$\ln(1 + \mu)$	−0.044	−1.082	0.070	1.678
$\ln(1 + \lambda)$	−0.484	−8.946	−0.304	−5.319
μ	−0.044	−1.107	0.073	1.620
λ	−0.383	−11.505	−0.262	−6.21
α	0.450	4.608	1.597	7.548
ρ	1.242	13.030	1.222	5.110
σ	5.736	28.966	5.845	27.737
Log-likelihood	−5129.630		−5029.524	
Sample size	959		959	

†Likelihood ratio test for significance of match effects $\chi^2_{139} = 200.212$, $p = 0.001$.

duration of partnerships (even given the intended run rate) and is suggestive of increasing carelessness or adventurousness as the duration of an innings is prolonged.

Table 2 reports results from allowing λ to vary with the number of wickets lost. The gain in the likelihood is sufficiently large to suggest a rejection of constancy for λ. Furthermore the differences accord with what theory suggests to be reasonable in an optimization context where the slopes to run trajectories are steeper for later partnerships. Indeed the picture of a negligible slope for high order partnerships and a more pronounced upward slope for those in the lower order, illustrated in Fig. 2, has some similarity with what is seen in the Fig. 1(d), particularly when it is remembered that the data for higher order partnerships are concentrated in early overs and for lower order partnerships in later overs.

6. Discussion

We have established several results concerning the nature of the optimum batting strategy for limited overs cricket and pointed to fundamental and possibly poorly appreciated differences in the optimum strategies for the two innings. We have also estimated the behaviour of actual cricketers in the second innings of such matches. Interestingly, actual trajectories share some of the features of what we believe to be optimal behaviour though we would be reluctant to suggest that we have any strong evidence for saying that cricketers are actually optimizing. It might be a

Table 2. Model estimates with wicket-dependent λ†

Parameter	*Estimates for model without match effects*		*Estimates for model with match effects*	
	Coefficient	*Estimated standard error*	*Coefficient*	*Estimated standard error*
Constant	−3.043	−12.415	−3.541	−5.839
Required run rate	0.512	6.030	1.099	7.227
1st wicket	−1.135	−4.473	−0.672	−2.125
2nd wicket	−1.209	−4.764	−0.606	−1.894
3rd wicket	−1.034	−4.070	−0.608	−1.921
4th wicket	−0.894	−3.536	−0.620	−2.006
5th wicket	−0.807	−3.095	−0.530	−1.666
6th wicket	−0.531	−2.099	−0.181	−0.590
7th wicket	−0.436	−1.668	−0.314	−0.962
8th wicket	0.303	1.200	0.872	2.689
9th wicket	0.034	0.150	0.507	1.636
$\ln(\rho + \mu\alpha)$	0.226	7.555	0.324	8.639
$\ln(\sigma)$	1.423	50.167	1.427	46.924
$\ln(1 + \mu)$	0.147	3.789	0.173	3.880
$\ln(1 + \lambda_{1-4})$	0.051	0.830	0.083	1.133
$\ln(1 + \lambda_{5-7})$	−0.211	−3.763	−0.185	−2.984
$\ln(1 + \lambda_{8-10})$	−1.265	−21.869	−1.213	−18.591
μ	0.159	3.523	0.188	3.563
λ_{1-4}	0.052	0.809	0.087	1.087
λ_{5-7}	−0.190	−4.188	−0.169	−3.279
λ_{8-10}	−0.718	−43.952	−0.703	−36.221
α	0.512	6.030	1.099	7.227
ρ	1.172	11.208	1.175	5.923
σ	4.150	35.253	4.165	32.890
Log-likelihood	−4965.867		−4868.359	
Sample size	959		959	

†Likelihood ratio test for significance of match effects $\chi^2_{139} = 195.015$, $p = 0.001$.

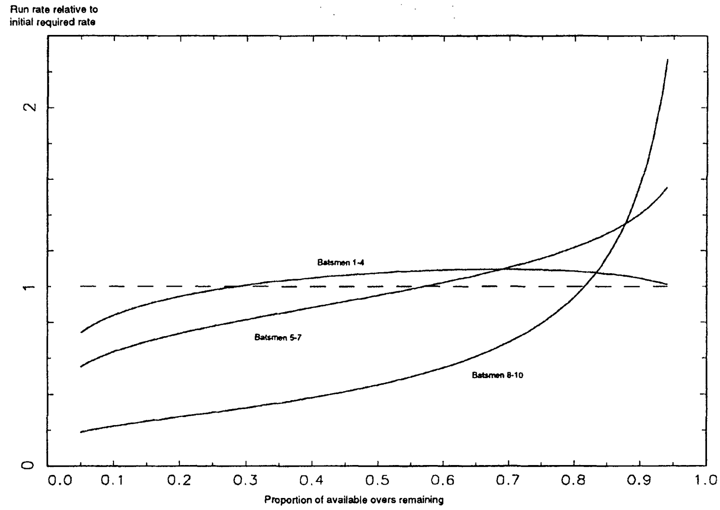

Fig. 2. Estimated batting strategies

worthwhile exercise to apply simulation-based methods to estimate an explicitly optimizing model using data from the same score-cards (in the fashion of Rust (1995, 1996)). Other strategic aspects in cricket may also deserve and repay attention, including bowling strategy.

Acknowledgements

We are grateful for the assistance of Tim Lamb and Richard Lockwood in gaining access to the data which were made available by the England and Wales Cricket Board and the Press Association. The responsibility for the interpretation of the data rests solely with the authors. We have benefited from discussions with or suggestions from the Joint Editor, Frank Duckworth, Jeremy Edwards, Tony Lewis, Hamish Low, Malcolm Pemberton, James Symons, Richard Vaughan, Frank Windmeijer, participants in seminars in Birmingham, Glasgow and London and a referee.

References

Bairam, E. I. and Howells, J. M. (1991a) The probability of winning and maximising behaviour for one-day international cricket. *Economics Discussion Paper 9120*. University of Otago, Dunedin.

———(1991b) The probability of winning and maximising behaviour in cricket: the New Zealand experience. *Economics Discussion Paper 9121*. University of Otago, Dunedin.

Bairam, E. I., Howells, J. M. and Turner, G. M. (1990a) Production functions in cricket: the Australian and New Zealand experience. *Appl. Econ.*, **22**, 871–879.

———(1990b) Production functions in cricket: one-day and three-day games in New Zealand. *J. Aust. Soc. Sports Hist.*, **6**, 203–217.

Bertsekas, D. (1987) *Dynamic Programming: Deterministic and Stochastic Models*. New York: Prentice Hall.

Clarke, S. R. (1988) Dynamic programming in one-day cricket—optimal scoring rates. *J. Oper. Res. Soc.*, **39**, 331–337.

Cox, D. and Oakes, D. (1985) *Analysis of Survival Data*. New York: Chapman and Hall.

Duckworth, F. C. and Lewis, A. J. (1998) A fair method of resetting the target in interrupted one-day cricket matches. *J. Oper. Res. Soc.*, **49**, 220–227.

Kimber, A. C. and Hansford, A. R. (1993) A statistical analysis of batting in cricket. *J. R. Statist. Soc.* A, **156**, 443–455.

Marylebone Cricket Club (1998) *Cricket: Know the Game*. London: Black.

Morrison, I. (1998) *Play the Game: Cricket*. London: Ward Lock.

Rust, J. (1995) Do people behave according to Bellman's principle of optimality? University of Wisconsin, Madison.

———(1996) Numerical dynamic programming in economics. In *Handbook of Computational Economics* (eds H. Amman, D. Kendrick and J. Rust). Amsterdam: North-Holland.